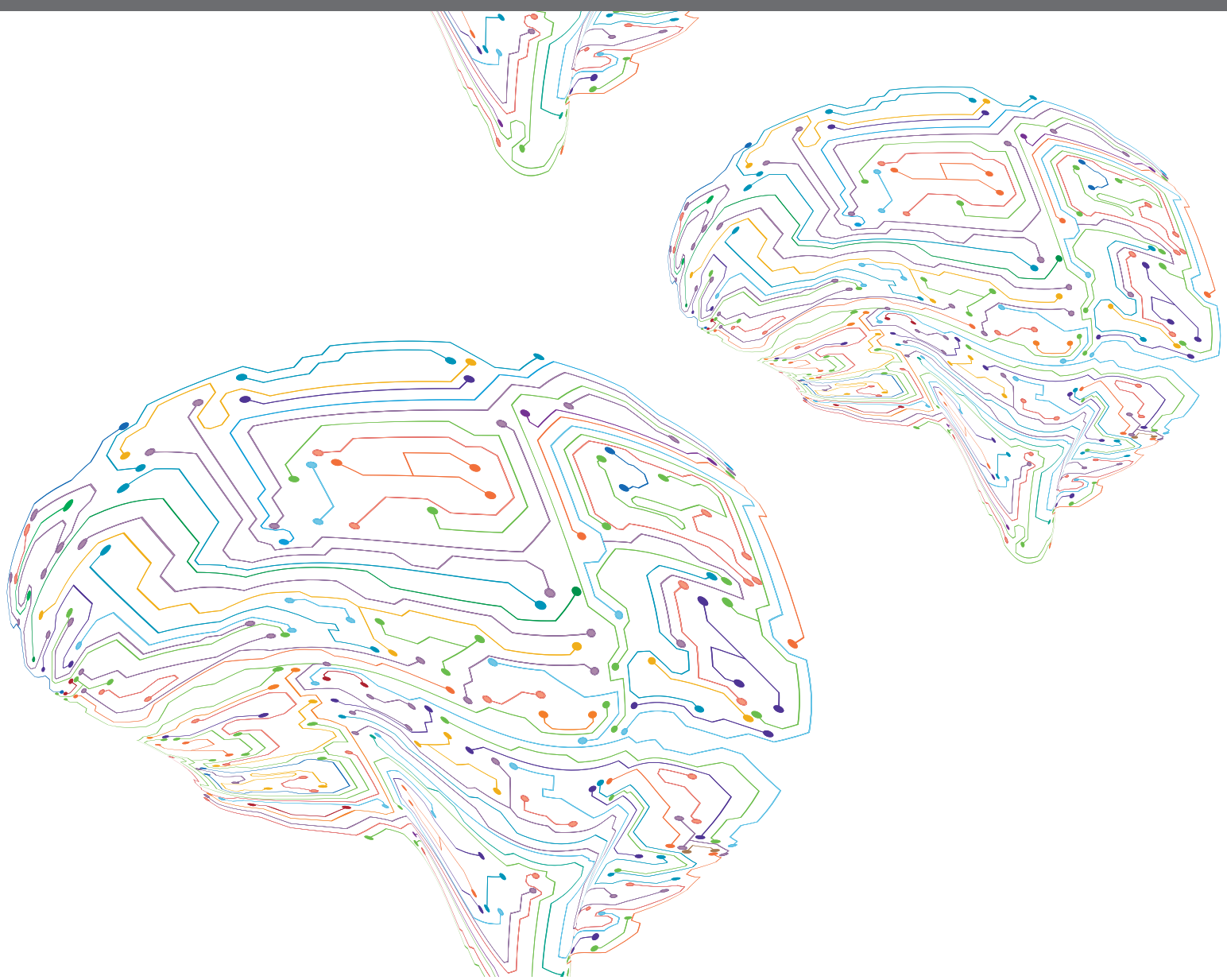
PERSISTENT ACTIVITY IN THE BRAIN – FUNCTIONS AND ORIGIN

EDITED BY: Shintaro Funahashi and Emmanuel Procyk

PUBLISHED IN: Frontiers in Neural Circuits



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-600-2
DOI 10.3389/978-2-88974-600-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

PERSISTENT ACTIVITY IN THE BRAIN – FUNCTIONS AND ORIGIN

Topic Editors:
Shintaro Funahashi, Kyoto University, Japan
Emmanuel Procyk, Institut National de la Santé et de la Recherche Médicale (INSERM), France

Citation: Funahashi, S., Procyk, E., eds. (2022). Persistent Activity in the Brain – Functions and Origin. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88974-600-2

Table of Contents

frontiers in Neural Circuits

EDITORIAL
published: 08 February 2022
doi: 10.3389/fncir.2021.841451

Editorial: Persistent Activity in the Brain – Functions and Origin

Shintaro Funahashi[1,2]* *and Emmanuel Procyk*[3]*

[1] Advanced Research Institute for Multidisciplinary Science, Beijing Institute of Technology, Beijing, China, [2] Shenzhen Institute of Advanced Technology, Chinese Academy of Science, Shenzhen, China, [3] Univ Lyon, Université Lyon 1, Inserm, Stem Cell and Brain Research Institute U1208, Bron, France

Keywords: prefrontal cortex, dynamic code, working memory, network, model

Editorial on the Research Topic

Persistent Activity in the Brain – Functions and Origin

Since Fuster and Alexander first found persistent firing in the prefrontal cortex during delayed-response performance (Fuster and Alexander, 1971), persistent neuronal activity has been observed in multiple brain areas and species, in relation to a variety of cognitive activities that require sustained processing or maintenance of states to bridge different events separated in time or to control or monitor specific functions. Persistent delay-period activity observed in the prefrontal cortex has been shown to play important roles to temporarily maintain either retrospective or prospective information necessary to perform a variety of cognitive behaviors during the delay interval, which corresponds to the function of working memory. Therefore, persistent activity has been considered as a neural correlate of working memory. Although it is an important phenomenon to understand neural mechanisms of working memory, planning or cognitive control, debates have been continued among researchers regarding its meaning, sources, and even relevance. In this Research Topic, we offered an opportunity to present research outputs carried out with various approaches including electrophysiology, behavior, and computational modeling, and to further discuss the origin and functional significance of persistent activity.

The brain uses different timescales to process information. Cavanagh et al. explored the features of persistent activity based on intrinsic timescales in single neurons for information processing. As primary sensory areas process momentary sensory inputs regardless of their context, short neural timescales in turn contribute to processing and representing such information. However, neural activity in association areas that integrate various information to achieve distant goals, are characterized by longer timescale. Single neurons display different individual timescales, which predict the strength of mnemonic encoding. Neurons exhibiting longer timescales play greater roles for stable maintenance of mnemonic information and for integrating multiple pieces of task-relevant information. Although neurons exhibiting shorter timescales are also present within higher cortical areas, the presence of neurons exhibiting heterogeneous timescales could be important for adapting environmental demands dynamically.

Although persistent activity has been frequently observed in the prefrontal cortex, this activity has also been observed in many other brain areas and the function and the information represented by this activity are not always the same across different brain areas, because different brain areas participate in different information processing. Roussy et al. dissociated neural processes for visual perception from those of visual working memory based on whether the brain produces mental representations of the visual stimulus when its physical signals are available or not. In fact, neurons in early visual areas exhibit persistent activity representing perceptual signals, while neurons in association areas exhibit the activity representing working memory signals. Persistent activity has also been observed in human brain imaging studies, although not in all frontal regions observed

OPEN ACCESS

Edited and reviewed by:
Edward S. Ruthazer, McGill University, Canada

***Correspondence:**
Emmanuel Procyk
emmanuel.procyk@inserm.fr
Shintaro Funahashi
funahashi.shintaro.35e@st.kyoto-u.ac.jp

Received: *22 December 2021*
Accepted: *31 December 2021*
Published: *08 February 2022*

Citation:
Funahashi S and Procyk E (2022) Editorial: Persistent Activity in the Brain – Functions and Origin. Front. Neural Circuits 15:841451. doi: 10.3389/fncir.2021.841451

in monkey studies. Curtis and Sprague reviewed these results obtained with fMRI signals, and showed that the content of working memory can be decoded from the pattern of neural activations in several brain areas, suggesting that these approaches find which of these brain areas contribute to working memory and what relevant features they represent.

Li et al. showed that persistent activity is observed even in naïve untrained monkeys, indicating that specific local neural circuits supporting persistent activity are innately present in the prefrontal cortex. However, the operation of these circuits is flexibly modulated by the task demand, the experience, and the age and also by the effect of neuromodulators. Systemic muscarinic blockade disrupts working memory performances. Vijayraghavan and Everling reviewed such neuromodulatory effects of acetylcholine on persistent activity in the prefrontal cortex. They showed that muscarinic blockade by local iontophoretic application caused pronounced suppressive effect of persistent activity representing remembered information and that the suppressive effect of persistent activity is dose-dependent and monotonically through muscarinic M1 receptors. The observed heterogeneity of muscarinic actions also outlines unexpected modulatory effects in primate prefrontal cortex when compared with rodent studies.

Recurrent neural circuit models have been proposed to explain how sustained representations are generated. Barbosa et al., Novikov et al., and Sarazin et al. have used such models of uni—or—multi-area networks to examine features of persistent or dynamic activity. Models are often informed by known biological elements and properties of recurrent circuits reviewed by Li et al., Curtis and Sprague, Roussy et al., Sarazin et al. used a recurrent neural circuit model with spike timing-dependent synaptic plasticity and showed that this model can learn, memorize, and replay a large-size of continuous dynamical sequences of spiking activity under asynchronous irregular nature at different timescales, which replicates the dual dynamical and persistent aspects of working memory representations observed in the prefrontal cortex.

Although persistent activity has been observed in single-neuron studies while monkeys performed delay tasks, such activity is often unstable along the delay period and heterogeneous within and across trials, which raises questions on how it contributes to working memory or to other processes putatively supported by tonic neural processing. In fact, dynamic coding and multiplexing during memory delays have been observed and are proposed to maximize the dimensionality of neural representations (Amengual and Ben Hamed; Cavanagh et al.; Curtis and Sprague; Sarazin et al.). Amengual and Ben Hamed suggested that heterogeneous features of persistent activity are caused by intrinsic oscillatory dynamics working at multiple timescales and that these features allow to dynamically incorporate multiple sources of information.

When subjects perform working memory tasks, an increase of gamma-band activity has been observed and is considered to reflect activation of neural populations representing the content of working memory. Novikov et al. examined functions of gamma-band oscillation on persistent activity by investigating joint effects of gamma-band oscillatory inputs and noise on the dynamics of the neural circuit with a metastable active condition. They showed that gamma-band oscillations are able to preferentially stabilize the active condition of the circuit in which information is retained in working memory, and that the synchronization of gamma oscillators affects the ability of the gamma inputs to stabilize the retention of working memory, indicating an importance of gamma-band oscillation for maintaining information in working memory. Importance of gamma-band oscillation is also shown for maintaining multiple items in working memory simultaneously. Barbosa et al. hypothesized that different features of an object (e.g., color and position) stored in different cortical areas are bound in memory through synchrony across feature-specific neural populations and tested this hypothesis using a neural network model composed of two one-dimensional attractor networks (one for color and one for position). They found that different memorized items are held at different phases of the network oscillation, that binding is accomplished through the synchronization of parts of bumps across the brain areas, and that encoding and decoding of object features are accomplished through rate coding.

Thus, although the debate regarding the meaning, source and relevance of persistent activity is continued, the maintenance of information and of neural patterns is still a central phenomenon to understand the neural mechanisms of working memory. Further insights might come from understanding whether and how intrinsic oscillatory dynamics working at multiple timescales contribute to generating persistent information. Also, as Curtis and Sprague suggested, decoding information represented in persistent activity observed in various brain areas may help decipher which brain areas are necessary for working memory and which features are represented and maintained in these brain areas. Pushing the line even further, Fuster observed in his opinion paper that "working memory consists in the temporary activation of an updated cortical network of long-term memory for the attainment of an objective," and conceptualize it as a *cognit*, an operational memory network. The *cognit* is distributed by nature, and thus, a network level approach is crucial to fully comprehend its functioning.

AUTHOR CONTRIBUTIONS

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

FUNDING

EP was supported by NORAD ANR-19-CE37-0008-01 and performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon. EP is employed by the Centre National de la Recherche Scientifique.

REFERENCES

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

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neural Circuits

REVIEW
published: 07 May 2020
doi: 10.3389/fncir.2020.00015

Plasticity of Persistent Activity and Its Constraints

Sihai Li[1], *Xin Zhou*[1,2], *Christos Constantinidis*[1] *and Xue-Lian Qi*[1]*

[1]*Department of Neurobiology and Anatomy, Wake Forest School of Medicine, Winston Salem, NC, United States,*
[2]*Department of Computer Science, Stanford University, Stanford, CA, United States*

Stimulus information is maintained in working memory by action potentials that persist after the stimulus is no longer physically present. The prefrontal cortex is a critical brain area that maintains such persistent activity due to an intrinsic network with unique synaptic connectivity, NMDA receptors, and interneuron types. Persistent activity can be highly plastic depending on task demands but it also appears in naïve subjects, not trained or required to perform a task at all. Here, we review what aspects of persistent activity remain constant and what factors can modify it, focusing primarily on neurophysiological results from non-human primate studies. Changes in persistent activity are constrained by anatomical location, with more ventral and more anterior prefrontal areas exhibiting the greatest capacity for plasticity, as opposed to posterior and dorsal areas, which change relatively little with training. Learning to perform a cognitive task for the first time, further practicing the task, and switching between learned tasks can modify persistent activity. The ability of the prefrontal cortex to generate persistent activity also depends on age, with changes noted between adolescence, adulthood, and old age. Mean firing rates, variability and correlation of persistent discharges, but also time-varying firing rate dynamics are altered by these factors. Plastic changes in the strength of intrinsic network connections can be revealed by the analysis of synchronous spiking between neurons. These results are essential for understanding how the prefrontal cortex mediates working memory and intelligent behavior.

Keywords: working memory, prefrontal cortex, training, monkey, neurophysiology

OPEN ACCESS

Edited by:
Emmanuel Procyk, Institut National de la Santé et de la Recherche Médicale (INSERM), France

Reviewed by:
Guillermo Gonzalez-Burgos, University of Pittsburgh, United States
Matthew L. Leavitt, Facebook (United States), United States

***Correspondence:**
Xue-Lian Qi
xqi@wakehealth.edu

Received: *10 December 2019*
Accepted: *26 March 2020*
Published: *07 May 2020*

Citation:
Li S, Zhou X, Constantinidis C and Qi X-L (2020) Plasticity of Persistent Activity and Its Constraints. Front. Neural Circuits 14:15. doi: 10.3389/fncir.2020.00015

INTRODUCTION

Working memory, the ability to maintain and manipulate information in mind over seconds, is one of the key components of higher cognitive functions (Baddeley, 2012). Early neurophysiological studies identified neurons in the lateral prefrontal cortex that generate persistent activity during working memory tasks (Fuster and Alexander, 1971; Kubota and Niki, 1971). Furthermore, the activity of individual prefrontal neurons was shown to be sensitive to the identity and location of remembered stimuli (Fuster and Alexander, 1971; Funahashi et al., 1989; Constantinidis et al., 2001b), as well as task variables, quantities, and categorical judgments (Freedman et al., 2001; Crowe et al., 2013; Blackman et al., 2016). As a result, information about all of these variables can be decoded from the activity of ensembles of prefrontal neurons (Meyers et al., 2008, 2012).

Working memory is not the only cognitive domain that persistent neural activity seems to predict (Constantinidis and Luna, 2019). For example, activity elicited during the preparatory period of an antisaccade task is correlated with the levels of working memory activity, on a neuron by neuron basis (Zhou et al., 2016a). Response preparation is a critical parameter of inhibitory control (DeSouza et al., 2003; Ordaz et al., 2010) and baseline activity may thus be tied to working memory, encoding advance preparation for the upcoming requirement to resist the stimulus appearance.

In recent years, alternative models have been proposed for working memory that do not rely on persistent activity, such as ones that rely on short-term modification of synaptic properties to maintain information, instead (Stokes, 2015; Mi et al., 2017; Lundqvist et al., 2018). It has also been suggested that the rhythmicity of activity generated during working memory is the critical neural variable for maintenance rather than the rate of persistent discharges. The magnitude, frequency and the phase of neural oscillations have indeed been demonstrated to be modulated as a function of stimuli and task information (Lundqvist et al., 2016, 2018). While more than one mechanism may play a role in the representation of information in working memory, these findings do not contradict the storage of working memory information in persistent neural activity (Riley and Constantinidis, 2016; Constantinidis et al., 2018). Modeling studies in which changes in synaptic plasticity are sufficient to maintain information in working memory in some tasks also reveal that persistent discharges are necessary for other, more complex tasks (Bouchacourt and Buschman, 2019; Masse et al., 2019). Only measures of persistent activity are strongly predictive of behavior in working memory tasks (Constantinidis et al., 2001b; Wimmer et al., 2014). We therefore focus exclusively on persistent activity in this review.

Although persistent activity maintains stimulus representations, it is also subject to change, which appears as a result of learning and development. Such plasticity is necessary for and provides a foundation for intelligent behavior. In recent years, neurophysiological and imaging studies have provided new insights into the effects of training in working memory tasks on the prefrontal cortex (Qi and Constantinidis, 2013; Constantinidis and Klingberg, 2016). Human and animal studies have made it possible to investigate how the prefrontal cortex responds to visual stimuli before and after behavioral training in a cognitive task, and how new information is integrated into neural circuits that are simultaneously maintaining information about the stimuli (Olesen et al., 2004; Meyer et al., 2011). Plasticity also occurs at different life stages, for example in adolescence, when the improvement of behavioral performance is associated with changes in prefrontal cortical activity (Constantinidis and Luna, 2019).

To understand the mechanisms of plasticity related to working memory it is necessary to first consider the neural circuits that generate persistent activity. Neural activity is thought to be sustained by reverberations of discharges in a network of neurons with reciprocal and recurrent connections (Wang, 2001; Wimmer et al., 2014; Riley and Constantinidis, 2016; Zylberberg and Strowbridge, 2017). The past decade has seen significant gains in our understanding of how persistent neural activity may change over time. In the current review, we aim to examine the latest insights on this topic. We focus mainly on visual-spatial working memory, the ability to maintain the spatial location of visual stimuli in mind, as this model provides us with a parametric variable, whose representation in neural activity is well understood (Riley and Constantinidis, 2016). We also focus on the lateral prefrontal cortex, the brain region most intricately implicated in this function, in non-human primates (Constantinidis and Procyk, 2004). The following sections review the mechanisms and circuits of persistent activity generation, how and to what extent these are plastic, and the open questions in the field, to be addressed in future studies.

MECHANISMS AND MODELS OF PERSISTENT ACTIVITY GENERATION

Intrinsic Circuits

Persistent activity depends simultaneously upon the properties of single neurons, the properties of neural networks within a cortical area, and the properties of long-distance networks between cortical areas. The influence of intrinsic prefrontal networks (schematically illustrated in **Figures 1A–C**) on neuronal activity can be investigated by physiological means. Nearby cortical neurons tend to generate near-synchronous spikes, within 0–2 ms of each other, significantly more often than would be expected by chance (Constantinidis et al., 2001a; Zick et al., 2018). These neurons also tend to be positively correlated at slower time scales, as evidenced by discharge rates averaged over periods in the order of 0.5–1 s (Constantinidis et al., 2001a; Kiani et al., 2015; Leavitt et al., 2017b). Cross-correlation analysis (**Figure 1D**), quantifying the relative timing of spiking of two neurons at the millisecond scale reveals that, when present, millisecond-scale cross-correlation peaks are most often centered at time 0, indicating synchronous firing (Constantinidis and Goldman-Rakic, 2002; Zhou et al., 2014). This pattern of cross-correlation peak is consistent with two neurons receiving input from common synaptic sources and provides a measure of the strength of intrinsic connections. The degree of synchronization is higher for neurons with similar spatial tuning and neurons active in the same epochs of the behavioral task, as would be predicted for neurons receiving shared input, which results in similar functional properties (Constantinidis et al., 2001a). We rely on cross-correlation measures to make inferences on circuit organization, and plasticity, below.

Axonal Projections

Reverberating activity through layer II/III horizontal excitatory connections between neurons with similar stimulus tuning is currently believed to be the primary mechanism of persistent discharge generation (Constantinidis and Wang, 2004). The basic circuit is illustrated in **Figure 1A**. Anatomical studies identified that prefrontal neurons receive horizontal connections from clusters of cells arranged in 0.2–0.8 mm wide stripes of the cortex, providing an anatomical substrate for such reverberation (Goldman-Rakic, 1984; Levitt et al., 1993; Kritzer and Goldman-Rakic, 1995; Pucak et al., 1996).

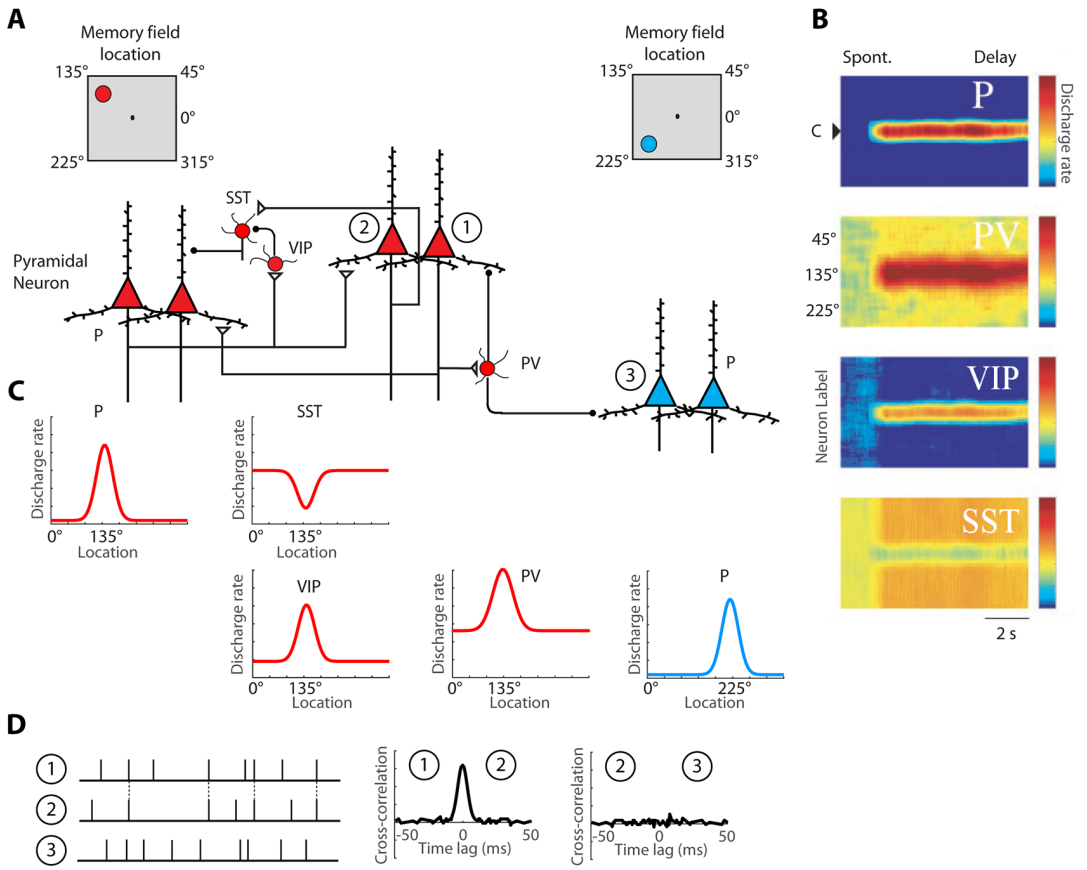


FIGURE 1 | Schematic illustration of the basic intrinsic circuit that maintains persistent activity in the prefrontal cortex. **(A)** Different types of neurons are indicated as follows. P, Pyramidal Neuron; PV, Parvalbumin Interneuron; VIP, Vasoactive Intestinal Polypeptide expressing Interneuron; SST, Somatostatin expressing Interneuron. Open triangles denote excitatory synapses; black circles indicate inhibitory synapses. 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 a preference for stimulus location (y-axis), as a function of time (x-axis). **(C)** Tuning curves of the same neuronal population, during the delay period. **(D)** Schematic illustration of cross-correlation analysis for neurons 1, 2, and 3, indicated in panel **(A)**. Raster plots represent spike time series of each neuron, obtained during a baseline period, before the appearance of stimuli. Synchronous spikes between neurons 1 and 2 result in a cross-correlation peak, centered at 0-lag. Adapted with permission from Zhou et al. (2012) and Wang et al. (2004) Copyright 2004 National Academy of Sciences.

Intrinsic connectivity is quantitatively enhanced within the prefrontal cortex compared to other cortical areas. Prefrontal pyramidal neurons exhibit the most extensive dendritic trees and the largest number of spines among cortical neurons (Elston, 2000, 2003). Physiological signatures of this greater extent of synaptic inputs into prefrontal neurons have been found in comparative cross-correlation studies, contrasting different cortical areas. Prefrontal neurons appear to receive a greater percentage of their inputs from neurons located at greater distances (>1 mm), and consequently to share a greater proportion of their inputs with neurons located at longer distances; in contrast, the spatial spread of inputs to posterior parietal neurons is much more limited and neurons located at shorter distances of each other (in the order of 0.2–0.5 mm) share a greater proportion of their inputs (Katsuki et al., 2014).

Other systematic differences between cortical areas in terms of axonal projections have also been identified recently, such as the MRI-based T1-weighted/T2-weighted ratio (Burt et al., 2018). This ratio is indicative of the extent of myelin presence within gray matter and provides a measure of convergence of axonal projections (Glasser and Van Essen, 2011; Huntenburg et al., 2017). This ratio is highest in the primary visual cortex and lowest (indicating most sparse connections) in the prefrontal cortex (Burt et al., 2018).

NMDA Receptors

NMDA receptors are critical in any neural circuit that generates persistent activity (Constantinidis and Wang, 2004). The relatively slow decay time constant of NMDA receptor-mediated synaptic currents allows post-synaptic neurons to remain in a depolarized state for a longer time (Wang, 2001).

If a network of excitatory neurons contained only AMPA synaptic receptors, which produce synaptic currents with very fast decay time constant, unrealistically high firing rates would be necessary to sustain neural activity during the delay period of a memory task (Wang, 1999). Experimental results also support the role of NMDA receptors in the generation of persistent activity, as NMDA antagonists greatly degrade persistent activity (Wang et al., 2013; Wang and Arnsten, 2015). For example, systemic administration of ketamine, a non-specific NMDA antagonist, decreases the strength of effective connectivity between prefrontal neurons, as evidenced by a decrease in the synchronous spiking between simultaneously recorded neurons (Zick et al., 2018).

NMDA expression is also area-specific. Among the different subunits that compose NMDARs in the adult brain, GluN2B has the slowest decay time constant. A gradient of GluN2B expression exists in the primate brain, with highest levels of expression observed in the prefrontal cortex (Burt et al., 2018), consistent with the ideas originally proposed by Wang (1999) and Wang (2001), that the slow decay constant of synaptic NMDARs is important in models of persistent activity.

Finally, NMDA represents one of the main mechanisms through which dopamine affects persistent activity. Iontophoretic application of dopamine agonists onto prefrontal neurons active during working memory affects firing rate in an inverted U fashion; at moderate doses, they increase activity for preferred stimuli and suppress non-preferred responses (Vijayraghavan et al., 2007; Ott et al., 2014). These agonists enhance the representation of actively remembered stimuli and suppress distractors (Jacob et al., 2016). Computational and experimental studies suggest that dopamine improves the signal-to-noise ratio of persistent activity mainly *via* enhancement of NMDAR currents (Yang and Seamans, 1996; Durstewitz et al., 2000; Seamans et al., 2001; Chen et al., 2004).

Interneuron Specialization

Inhibitory neurons in the prefrontal cortex exhibit persistent activity as pyramidal neurons do (Rao et al., 1999, 2000; Constantinidis and Goldman-Rakic, 2002; Constantinidis et al., 2002). Computational models suggest that inhibition is essential for creating stimulus-selective persistent activity (Compte et al., 2000), and both computational and experimental results suggest that prefrontal interneurons generally exhibit higher baseline firing rates and broader tuning than pyramidal neurons (Constantinidis and Goldman-Rakic, 2002).

A division of labor among cortical interneurons has been hypothesized, in which multiple types of GABAergic neurons form a specialized network, to facilitate stimulus-specific persistent activity (Wang et al., 2004), as illustrated in **Figure 1A**. In this scheme, pyramidal neurons would recruit Parvalbumin (PV) expressing inhibitory interneurons to suppress the activation of other pyramidal neurons, with different spatial turning, since PV cells target the cell bodies of pyramidal neurons. Anatomical evidence that suggests that PFC neurons with similar memory fields are grouped in clusters that may be the anatomical substrate for recurrent excitation (Goldman-Rakic, 1984; Levitt et al., 1993; Kritzer and Goldman-Rakic, 1995; Pucak et al., 1996) and in such a scheme, PV interneurons could provide lateral inhibition by inhibiting neurons in different clusters, as depicted in the model. Alternatively, PV cells may provide feedback inhibition to adjacent pyramidal cells that reciprocally excite the PV cells, as has been demonstrated experimentally in the rodent cortex (Adesnik et al., 2012; Atallah et al., 2012; Wilson et al., 2012). Primate interneurons exhibit broader tuning curves than pyramidal neurons (Constantinidis and Goldman-Rakic, 2002) and in such as scheme, PV neurons would facilitate stimulus-specific working memory by sharpening the tuning function of adjacent pyramidal neurons and contributing to Excitatory/Inhibitory (E/I) balance. Without feedback inhibition, recurrent excitation may shift the E/I balance and bring the network into an unstable, hyper-excited state, which would also be deleterious for the maintenance of working memory (Constantinidis and Wang, 2004).

The second class of inhibitory interneurons, expressing Vasoactive Intestinal Peptide (VIP), 80% of which also express Calretinin (Gabbott and Bacon, 1997), would inhibit a third class of interneurons, those expressing Somatostatin (SST) and likely Calbindin. VIP neurons are interneuron-targeting cells and when activated, they would inhibit SST neurons, which are peridendritic-targeting cells and they tonically inhibit pyramidal neurons (Pi et al., 2013; Dienel and Lewis, 2019). The model predicts that SST neurons exhibit a high spontaneous rate (**Figures 1B,C**), which during the baseline period, before a stimulus appearance, inhibits tonically all pyramidal neurons. The properties of SST inputs have not been investigated in detail in the primate cortex, but in the rodent cortex, SST neurons are strongly modulated by acetylcholine (Chen et al., 2015; Urban-Ciecko et al., 2018). After a stimulus is maintained in working memory, SST neurons would effectively release from inhibition pyramidal neurons that have already attained a state of excitation by the same stimulus. Other populations of SST neurons, not recruited by the stimulus held in memory would continue to inhibit non-activated pyramidal neurons, thus suppressing background noise as well as potential activation by subsequent, distracting stimuli (Wang et al., 2004).

The activation profiles of these three classes of interneurons and tuning curves relative to the tuning of pyramidal neurons they are linked to are schematically depicted in **Figures 1B,C**. Direct experimental evidence for the disinhibitory role of VIP cells has been provided by rodent studies (Pi et al., 2013). The model is simplified, in that VIP neurons also inhibit PV neurons, at least in rodent visual cortex. VIP-to-SST and VIP-PV synapses also show strong short-term synaptic depression, which suggests that synaptic output from VIP neurons is best fit to briefly inhibit other interneurons, possibly suppressing the phasic effect of distracting stimuli, rather than being a continuous input during the entire delay period (Pi et al., 2013). Finally, VIP neurons in the mouse barrel cortex are not well-tuned to stimulus properties, suggesting distant inputs (Yu et al., 2019).

Nonetheless, the basic circuit of **Figure 1** appears to be conserved in primates. A subset of primate Calretinin

interneurons preferentially targets Calbindin interneurons (Meskenaite, 1997; Melchitzky and Lewis, 2008; Fish et al., 2018), thus creating an analogous circuit. Furthermore, interneuron-targeting cells are more abundant in association cortices, and particularly in the prefrontal cortex, compared to the sensory cortex (Defelipe et al., 1999; Elston and González-Albo, 2003). At least indirect evidence supports the idea that a disinhibiting circuit is more pronounced in the prefrontal cortex: interneurons with high baseline firing rate and inverted tuning (consistent with the profile of disinhibiting neurons) are more numerous in the prefrontal cortex than in the posterior parietal cortex (Zhou et al., 2012). While the basic circuit of **Figure 1A** appears to be present across species and cortical areas, the intrinsic prefrontal circuit is more capable of generating and sustaining persistent activity than its afferent areas.

Long-Distance Circuits

Although the prefrontal cortex may be the primary source of persistent activity in working memory, the generation of persistent activity is not exclusive to the prefrontal cortex alone. Neurons exhibiting persistent activity have been identified in several additional brain areas, including the posterior, parietal, and inferior temporal cortex, thalamic nuclei, particularly the mediodorsal nucleus of the thalamus, and also the basal ganglia (Constantinidis and Procyk, 2004). This is not to say that persistent activity is entirely distributed across areas, either; it was found to be absent in visual cortical area MT and to emerge *de novo* in area MST, in one well-studied paradigm (Mendoza-Halliday et al., 2014). Long-distance connections between these areas and the prefrontal cortex have been hypothesized to provide a larger scale circuit to generate persistent activity during working memory. Recent modeling efforts suggest that long-range inter-area reverberation may support the emergence of persistent activity in areas whose local circuit organization is not sufficient for its maintenance (Mejias and Wang, 2019). Direct evidence for the necessity of thalamocortical connections for the generation of persistent activity has been provided by rodent studies (Guo et al., 2017). Ultimately, this means that long-distance connections may be essential for the generation of persistent activity both within and outside of the prefrontal cortex.

The existence of persistent activity in multiple brain areas does not necessarily mean that all aspects of working memory are distributed, either (Leavitt et al., 2017a). Instead, different areas appear to be involved with during aspects of working memory (Riley and Constantinidis, 2016). The prefrontal cortex is uniquely equipped to represent information about the spatial location of an initial stimulus after distracting information has been presented, whereas the posterior parietal cortex seems to track the most recent stimulus (Qi et al., 2010). Neuronal activity related to executive control of information maintained in memory is similarly thought to originate in the prefrontal cortex and be transmitted to the posterior parietal cortex (Crowe et al., 2013). The distinction between the patterns of activity in the posterior parietal and prefrontal cortex, however, depends on the parameters of the specific working memory task that is being performed. Under some tasks, the posterior parietal and prefrontal cortex may represent different types of information, encoding either the initial or subsequent stimuli (Jacob and Nieder, 2014; Qi et al., 2015; Masse et al., 2017).

Linking Circuit Models With Behavior

Persistent activity recorded in the prefrontal cortex is predictive of behavior in working memory tasks. Trials in which the preferred stimulus of a recorded neuron elicits less activity than average are more likely to result in errors (Funahashi et al., 1989; Zhou et al., 2013). As a result, a near-linear relationship between behavioral performance and persistent neural activity has been revealed in tasks that parametrically modulate the properties of stimuli held in working memory (Constantinidis et al., 2001b). Choice probability analysis, comparing the distributions of firing rates in the delay period of correct and error trials, also reveals a stronger relationship between persistent activity in the prefrontal cortex and behavioral outcomes, compared to other areas (Mendoza-Halliday et al., 2014).

Computational models provide a link between persistent activity and behavioral performance in working memory tasks. Persistent activity is sustained in these models by recurrent connections between neurons with similar tuning for stimulus properties, thus allowing activation to be maintained past the presence of the afferent input (Compte et al., 2000; Murray et al., 2017). The system can be thought of as a continuous attractor. Drifts in neuronal activity across the network of prefrontal neurons predict precisely the relationship between firing rate and the endpoint of the saccade (the spatial location being recalled by the monkey) in the ODR task (Wimmer et al., 2014). For example, 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, as would be expected if the location recalled was determined by the peak of activity at the end of the delay period.

PLASTICITY

The plasticity of neural activity is essential for intelligent behavior. Persistent activity can be highly plastic and is influenced by several factors that also impact working memory performance. This is not to say however that there are no limits in plasticity. The following sections examine plastic changes and their constraints as a result of training and age (section 3.1), and the circuit changes that likely mediate them ("Cellular Substrates of Plasticity" section).

Initial Working Memory Training

Persistent activity appears to be generated automatically, in subjects not required or even trained to perform a task (Meyer et al., 2007; Riley et al., 2017). When naïve monkeys are passively viewing stimuli, some prefrontal neurons become activated and continue to discharge after the stimuli are no longer present. Working memory has sometimes been thought to require willful effort, and/or training in specific working memory tasks (Postle, 2006). In our everyday experience, however, we can track our environment and recall information even when not explicitly

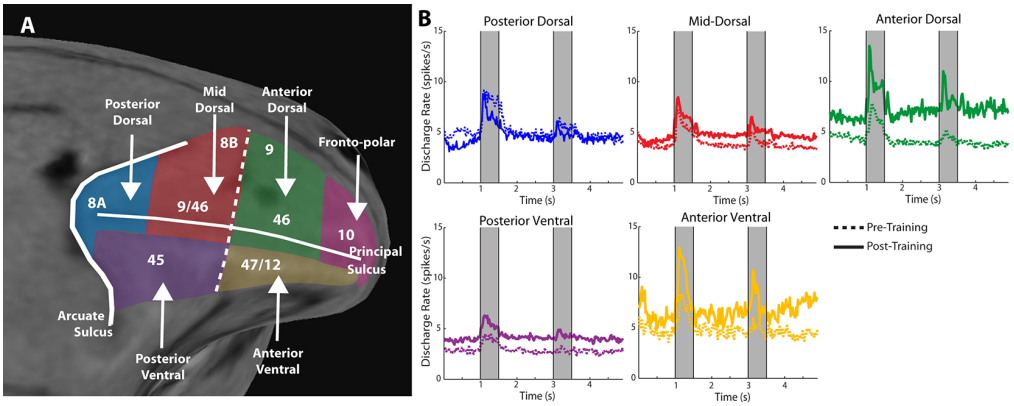


FIGURE 2 | (A) Anatomical MRI of the monkey lateral prefrontal cortex with anterior/posterior and dorsal/ventral subdivisions indicated, relative to the Principle and Arcuate Sulci. **(B)** Mean firing rate of neurons recorded in these subdivisions in monkeys both before and after they were trained to perform spatial working memory tasks. Gray bars represent stimulus presentations. Data are shown separately for each prefrontal region. Adapted with permission from Riley et al. (2018).

prompted to do so ahead of time, implying that working memory may be an automatically generated process. In agreement with this intuition, a proportion of prefrontal neurons are persistently active even when the subjects were not assigned to remember any stimuli (Meyer et al., 2007; Riley et al., 2017). Furthermore, the rate of persistent discharges in this population is selective for properties of the stimuli, including spatial location, color, and shape. It therefore appears that a prefrontal circuit is hardwired to automatically generate persistent activity once activated by sensory stimuli. However, there are limits to the information that may be represented automatically. For example, the identity of a stimulus generally did not survive a second stimulus presentation in the experiments discussed above, and information about whether the shape of two stimuli was the same or not was largely absent in naïve animals (Meyer et al., 2007; Meyers et al., 2012; Riley et al., 2017).

Training to perform a working memory task for the first time does elicit plastic changes in persistent activity (Mendoza-Halliday and Martinez-Trujillo, 2017; Riley et al., 2018). More neurons are active after such training and generate a higher level of persistent activity (Meyer et al., 2007; Riley et al., 2017). The circuit changes that training induces appear to be lasting and the difference in firing rate between naïve and trained animals are evident even when the trained monkeys are tested with the passive presentation of stimuli, in the same fashion they did before training. The mean firing rate of persistent discharges is also higher in the trained than that in naïve monkeys, though the execution of the task further amplifies persistent activity compared to the passive viewing of stimuli (Riley et al., 2017). Plasticity is not all-or-none in terms of exposure to training. Increases in firing rate tended to accrue with cumulative training and are reflective of the level of performance in the working memory task at each point in time (Qi et al., 2011; Tang et al., 2019). These effects represent average changes in neuronal activity, sampled from different groups of neurons at different stages of training. It will be interesting for future experiments to track the activity of individual neurons as learning of a new task takes place.

Anatomical Constraints on Plasticity

Anatomical position is an important constraint on the plasticity of persistent activity. Within the lateral prefrontal cortex, levels of persistent activity depend on position across the dorsoventral (Kadohisa et al., 2015; Constantinidis and Qi, 2018) and anterior-posterior axes (Riley et al., 2017, 2018). The lateral aspect of the prefrontal cortex is subdivided into areas 8a, 46, 8b, and 9 in its dorsal aspect, areas 12 and 45 in its ventral aspect, and area 10 covering the frontal pole (Walker, 1940). There is also evidence of a specialization in the anterior-posterior aspect, with the caudal aspect of area 46 shown to be functionally dissociable from the anterior aspect; the former is referred to as area 9/46, whereas the most anterior area is called area 46, in this nomenclature (Petrides, 2000). Division in more areas has also been proposed, based on the evidence provided by fMRI studies probing functional connectivity at rest (Goulas et al., 2017). Based on physiological evidence, we have recently proposed dividing the lateral PFC into subdivisions as follows (**Figure 2A**): a posterior, mid-, and anterior-dorsal region, a posterior- and anterior-ventral region, and a frontopolar region (Riley et al., 2017).

Neurons in different prefrontal subdivisions exhibit different properties and aptitudes for plasticity (**Figure 2B**). The posterior aspect of the prefrontal cortex is the most specialized for stimulus location (posterior-dorsal) and object information (posterior-ventral) but is affected relatively little by training (Constantinidis and Qi, 2018). Little difference in mean persistent firing rate is observed in the posterior-dorsal prefrontal cortex before and after training, though more neurons become active (Meyer et al., 2011; Riley et al., 2018). Instead, most of the plasticity in persistent activity occurs in the mid-, and anterior-dorsal areas of the prefrontal cortex (area 46). Across the medio-lateral axis of the dorsal prefrontal cortex, little or no changes in plasticity are seen in the most dorsal areas (areas 8b and 9), whereas plasticity of persistent discharges is evident in the principal sulcus region (area 46), and more so in the ventrolateral prefrontal cortex (Meyer et al., 2011). The organization of the prefrontal cortex has been a matter of debate, with at least some studies failing to

identify dissociable responses of neurons in different prefrontal subdivisions (Rao et al., 1997; Lara and Wallis, 2014). In terms of plasticity, however, there seems to be more agreement, and lesion studies support the idea of ventral areas being more essential for the acquisition of new tasks, which implies greater capacity for plasticity (Buckley et al., 2009).

Plasticity Changes Beyond Firing Rate

Encoding of information in neuronal firing depends not only on the mean firing rate of neuronal responses, but also on how variable these responses are from trial to trial, and on whether firing rates of neurons are positively correlated with each other, which limits how much information can be stored in their collective discharges (Moreno-Bote et al., 2014). The effects of plasticity similarly affect not only mean firing rate but also the variability of persistent activity (Qi and Constantinidis, 2012b) and the correlation of firing rate between simultaneously recorded neurons (Qi and Constantinidis, 2012a). The Fano factor of spike counts, a measure of variability, generally decreases after practicing the task, with the greatest decreases observed in neurons that exhibit persistent activity, compared to neurons that do not. This decrease in trial-to-trial variability may be responsible for increasing the reliability of stimulus property representation after training. Similarly, the spike-count correlation of persistent firing rates between pairs of neurons (known as noise correlation) also decreases after training, which improves the information that can be decoded from simultaneously active neurons (Qi and Constantinidis, 2012a).

Task training also alters the time course and dynamics of persistent activity (Kobak et al., 2014; Tang et al., 2019). Prefrontal neurons are known to exhibit dynamics during working memory tasks. For example, the firing rate of some neurons is known to "ramp up" or decrease during the trial, so that information about the stimulus is encoded dynamically at different time points (Romo et al., 1999; Meyers et al., 2012; Stokes et al., 2013). However, the existence of dynamics does not undermine the representation of information in working memory. Recent work (Murray et al., 2017) has revealed a stable subspace, where information can be maintained in an invariant fashion (**Figures 3A,B**). Similar subspaces have been identified across a variety of working memory tasks (Murray et al., 2017; Spaak et al., 2017; Parthasarathy et al., 2019). Training in a working memory task does alter the dynamics of persistent activity. Neuronal responses recorded in animals trained to perform a working memory task exhibit more pronounced increases and decreases of activity during the time course of the trial than animals passively viewing (Kobak et al., 2014; Tang et al., 2019). The consequence of this change is that a greater percentage of firing rate variance is accounted for by components unrelated to the remembered stimulus location or identity.

Plasticity When Learning to Perform Additional Tasks

After monkeys have been trained to perform basic cognitive tasks, it is possible to train them in more complex tasks, including ones requiring working memory for multiple stimuli. Training in tasks with multiple-stimuli can improve working memory capacity (at least in the task trained) and induces plastic changes in prefrontal activity (Tang et al., 2019): more neurons become activated, their baseline firing rate decreases, and although persistent activity may not change appreciably, the rate of persistent activity relative to baseline is enhanced after training. A debate exists in the human imaging literature, with some studies revealing decreases in activity after training in complex tasks (Schneiders et al., 2011; Kühn et al., 2013; Schweizer et al., 2013; Takeuchi et al., 2013) and the decreases are often interpreted as improvements in efficiency, or strategy (Constantinidis and Klingberg, 2016). The decrease in baseline activity observed in the neurophysiological studies may be partially responsible for such results, particularly when activity is averaged over long periods, as in fMRI studies. Acquiring data during training in a variety of tasks will be essential for understanding the full repertoire of plastic changes.

Like humans, monkeys are known to develop strategies when attempting to master complex tasks, e.g., suggestive of a grouping of multiple stimuli in memory based on their geometric arrangement (Tang et al., 2017). The selectivity of prefrontal responses for remembered displays containing multiple stimuli, or sequences of remembered stimuli, is often very different than for single, identical stimuli (Konecky et al., 2017; Tang et al., 2019) depending on the corresponding mental operation performed.

An important finding of the training studies with multiple stimuli was changed in the dynamics of neuronal activity (Tang et al., 2019). As was the case with the initial working memory training, once subjects practiced a new task requiring memory for multiple stimuli and improved their performance, a greater percentage of activity could be explained by "condition-independent" components, not related to the stimuli being remembered (**Figures 3C,D**).

Learning to perform multiple tasks also alters the levels of persistent activity representing the newly acquired information (Sarma et al., 2016). Modulation of persistent activity depending on what information needs to be maintained in memory can take place very rapidly, e.g., within a few trials, when a subject learns a new sensory-motor association (Asaad et al., 1998) or on a trial-to-trial basis, when the subject is cued to remember a particular feature of the stimulus to perform judgment and ignore others (Mante et al., 2013). The persistent activity can also be modulated in the course of a single trial during the execution of dual-task paradigms when the subject temporarily focuses on the representation of one stimulus in memory before resuming a task requiring representation of another stimulus (Watanabe and Funahashi, 2014).

Age

The normal developmental and aging process provides another opportunity to study plastic changes in the ability of the prefrontal cortex to generate persistent discharges, regardless of training and life experiences. Behavioral performance and neural activity in working memory tasks change markedly around the time of puberty, a developmental event associated with the release of sex hormones and significant neurological

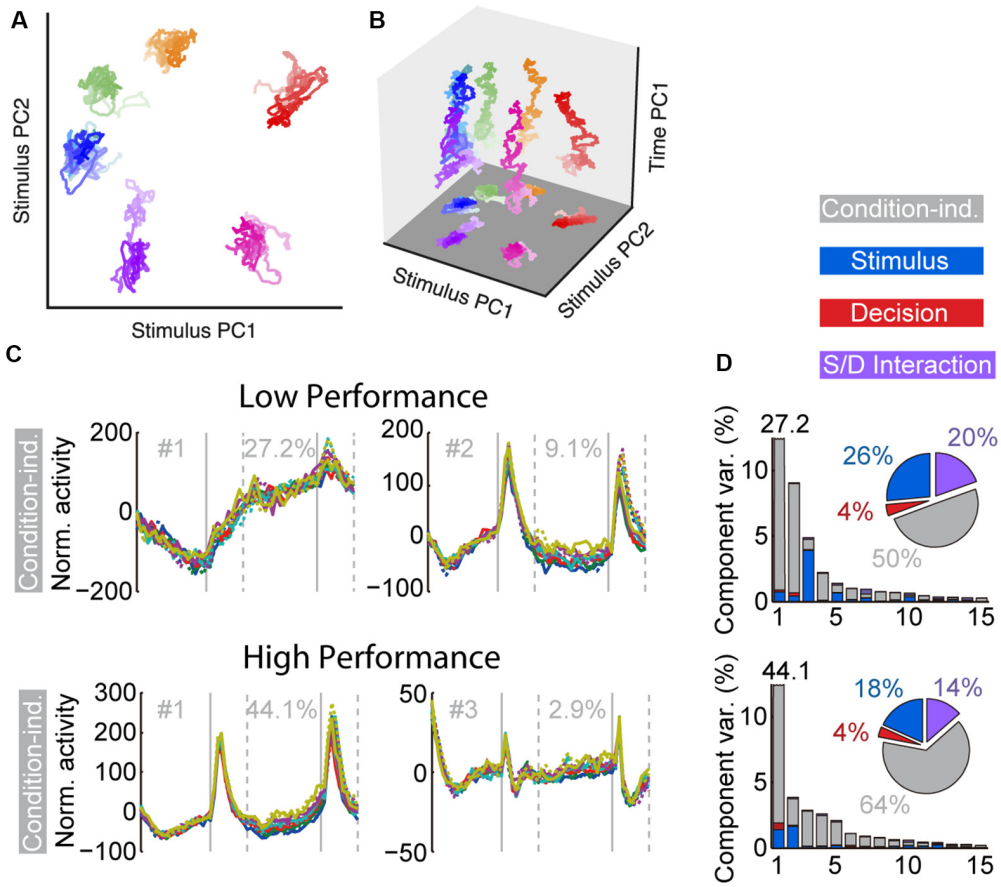


FIGURE 3 | (A) Population trajectories during the delay period projected into the mnemonic subspace, defined *via* PCA on time-averaged delay activity. Here the x and y axes display the first and second principal components (PC1 and PC2) of the subspace, respectively. Each trace corresponds to a stimulus condition, roughly corresponding with the actual stimulus location on the screen. The shading of the traces marks the time during the delay, from early (light) to late (dark). **(B)** Same data as in **(A)**, with the z-axis denoting time. **(C)** Demixed PCA Analysis. The first two condition-independent components of dPCA analysis from low and high-performance sessions are plotted, based on an experiment requiring monkeys to maintain multiple stimuli in working memory. **(D)** Histogram representing the amount of variance accounted by different condition-independent, stimulus, decision, and interaction components. Reproduced with permission from Murray et al. (2017; **A,B**) and Tang et al. (2019; **C,D**).

change (Zhou et al., 2013, 2016b). The performance of working memory tasks is subtly but significantly higher in adult monkeys compared to adolescent monkeys that have entered puberty, just as it improves in humans between these two developmental stages (Montez et al., 2019). Persistent activity is also higher in adult animals than adolescent ones (**Figures 4A–C**). 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. Furthermore, the adult prefrontal cortex can more effectively filter distracting stimuli during working memory.

In the other end of the life spectrum, advanced age in monkeys is marked by a significant loss of persistent firing in the prefrontal cortex. 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). The persistent activity can be partially restored to more youthful levels by inhibiting cAMP signaling, or by blocking HCN or KCNQ channels (Wang et al., 2011). Notably, both in the adolescent and aged monkeys it was the persistent activity that differed from that recorded in adults (Wang et al., 2011; Zhou et al., 2016b). Prefrontal activity during the stimulus presentation differed little between age groups, suggesting that stimulus-driven neuronal responses were fully mature in puberty and resistant to the effects of aging in old monkeys. These studies provide some brief snapshots of neuronal activity at two critical life stages. It will be important for future studies to reveal the full-time course of persistent discharges from young to old age.

Cellular Substrates of Plasticity

The observed changes in neural activity through training and age suggest plasticity in the circuit that generates persistent discharges. Local-circuit differences are evident between adolescent and adult monkeys that could explain the

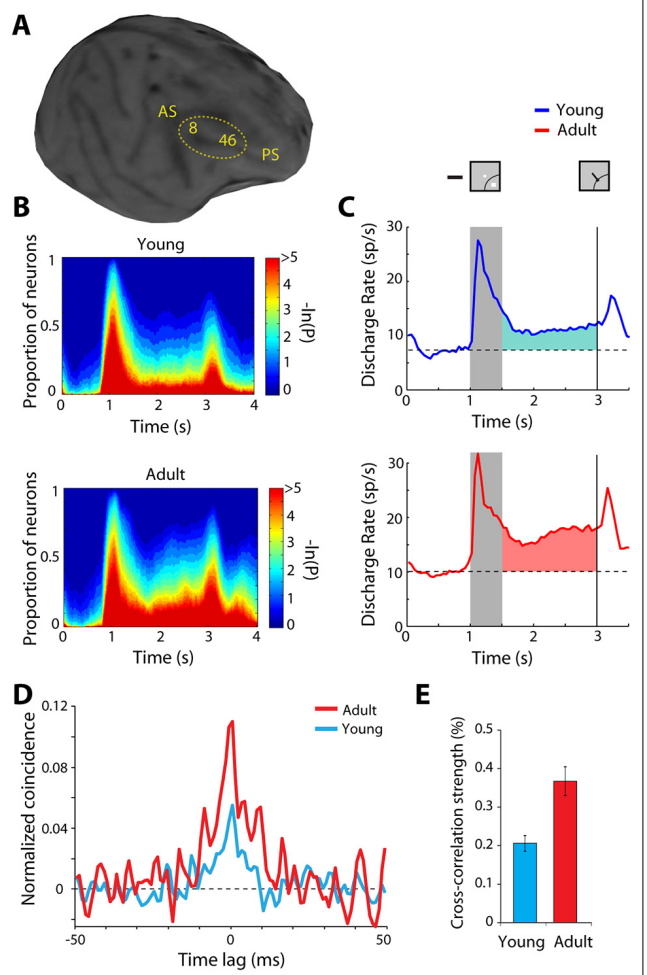


FIGURE 4 | (A) Anatomical MRI of an adolescent monkey, with the dorsolateral cortex (areas 8 and 46) indicated. AS, Arcuate Sulcus; PS, Principal Sulcus. **(B)** The proportion of neurons with significant differences in firing rates compared to baseline fixation, at different time points of the ODR task (color scale represents the probability of a paired *t*-test), for the adolescent and adult groups. **(C)** Average population peri-stimulus time histogram for neurons that responded to the visual stimulus and were recorded during the ODR task in the adolescent stage (top) or the adult stage (bottom). Responses are shown for a stimulus in each neuron's receptive field. Gray bar represents an interval of stimulus presentation; the vertical line represents the time of the fixation offset. Insets schematically display the stimulus location and direction of eye movement relative to the receptive field (arc), which varied for each neuron. **(D)** Averaged, normalized cross-correlation histogram for adolescent and adult monkeys. **(E)** Average value (strength) calculated in the center 5 ms of the cross-correlation histogram for adolescent and adult monkeys. Adapted with permission from Zhou et al. (2014, 2016b).

decreased ability of the immature prefrontal cortex to generate persistent discharges. Zero-lag spiking synchronization based on cross-correlation analysis of nearby neurons (recorded at distances between 0.5–1 mm from each other) is markedly lower in adolescent than in adult monkeys (**Figures 4D,E**). This difference is primarily the effect of changes in inhibitory interactions (Zhou et al., 2014), possibly due to decreases in the connectivity strength of pyramidal neurons onto interneurons, which lessens the net output of inhibitory connections as the prefrontal cortex matures (Gonzalez-Burgos et al., 2015). Interestingly, a decrease in zero-lag synchrony of prefrontal neurons has been recently implicated in schizophrenia (Zick et al., 2018), a condition that, among other pathological symptoms, compromises working memory.

Neuromodulators have also been implicated in dynamic changes of persistent activity and are likely to be involved during learning or the selection of stimulus features. Most notably, cholinergic stimulation through the iontophoretic application of cholinergic agonists (Yang et al., 2013; Sun et al., 2017; Dasilva et al., 2019), or the stimulation of the cholinergic basal forebrain (Qi et al., 2019) leads to a general increase in activity of neurons in the prefrontal cortex. Conversely, systemic administration of the muscarinic antagonist scopolamine (Zhou et al., 2011) or iontophoresis of muscarinic and nicotinic-α7 inhibitors seems to depress prefrontal persistent activity (Yang et al., 2013; Major et al., 2015; Dasilva et al., 2019). Cholinergic stimulation elicits not only direct changes in neural activity but also long-term neuroplasticity effects, suggestive of circuit reorganization (Brzosko et al., 2019).

The reason that plasticity differs between PFC subdivisions can also be traced to systematic differences between anatomical connections and cellular mechanisms. Anatomical studies point to relative segregation of projections from the posterior parietal cortex, which terminate mostly to the posterior dorsal PFC (areas 8 and 46, including both banks of the principal sulcus), and from the inferior temporal cortex, which terminate on the posterior ventral PFC (Petrides and Pandya, 1984; Selemon and Goldman-Rakic, 1988; Cavada and Goldman-Rakic, 1989). Areas higher in the sensory and limbic hierarchies projecting to more anterior prefrontal subdivisions (Gerbella et al., 2013; Barbas, 2015; Borra et al., 2019). Increasingly anterior prefrontal areas integrate inputs from more posterior ones, being activated by higher-order cognitive operations in a rostral-caudal axis of cognitive control (Petrides, 2005).

A greater capacity for plasticity after training in a cognitive task may also point to the specialization of underlying cellular and molecular mechanisms (Kuboshima-Amemori and Sawaguchi, 2007), which may vary between prefrontal subdivisions. Indeed, direct evidence of systematic variation of plasticity markers between limbic and eulaminate areas has been recently documented in the prefrontal cortex (García-Cabezas et al., 2017). Calcium/calmodulin-dependent protein kinase II (CaMKII), which is essential for plasticity, is more impoverished in area 46d compared to more anterior limbic areas, whereas makers of cortical stability, including intracortical myelin, perineuronal nets, and PV show the reverse 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 (Laroche et al., 2000; McEwen and Morrison, 2013), and may differ between areas. Finally, short-term synaptic plasticity, depression or facilitation, has been documented in the prefrontal cortex, and this too may be critical, particularly for task-related plasticity (Hempel et al., 2000). Tying these cellular and

molecular mechanisms to actual changes in neuronal activity and capacity for plasticity will be an important goal for future studies.

CONCLUSIONS AND OPEN QUESTIONS

This review summarized the current state of knowledge on the generation and plasticity of persistent activity during working memory. Some conclusions emerge from this review. We conclude that although the prefrontal cortex is not the only area where persistent discharges are evident, its unique cellular and circuit organization makes it essential for the generation of persistent activity. Training in working memory tasks greatly affects the neuronal circuit of the prefrontal cortex, causing more neurons to exhibit persistent activity and to have this activity reach higher discharge rates. Different prefrontal subdivisions have different capacities for plasticity, with most plastic changes being evident in anterior and ventral areas. Plasticity may manifest itself in a variety of ways. Higher firing rate during the delay period is the most obvious effect of training and the adult stage of maturation, compared to adolescence and old age. However, changes in firing rate variability, correlation between firing rates of different neurons, decreases in baseline firing rate, and changes in neuronal dynamics have all been identified as markers of plasticity.

Many questions related to the generation and plasticity of persistent activity remain open for future research: first, we hypothesized that the functional circuit of **Figure 1** is most developed in the primate prefrontal cortex. Testing of the role of identified interneuron populations in different cortical areas in the context of working memory could provide direct evidence that this is the case. Second, what are the actual synaptic changes that occur at the level of intrinsic circuits, within the prefrontal cortex, as well as in long-range connections between the prefrontal cortex and other areas when subjects learn and practice working memory tasks? Addressing this question will require the interrogation of circuits in subjects while they learn to perform working memory tasks. Recent technical developments have brought this aim within reach. Next, what is the role of different neurotransmitter systems during learning? It is well understood that dopamine and acetylcholine play an essential role in neuroplasticity but there is a gap regarding how these factors affect persistent activity during training. A final area of unanswered questions has to do with the generation of object memory. Although spatial memory can be manipulated parametrically and modeled in a neural circuit, object memory has proven more elusive. We therefore ask how objects are maintained in working memory and the factors that govern plasticity for object memory. These questions will have to be addressed in future studies.

AUTHOR CONTRIBUTIONS

CC and X-LQ conceived and organized the article. SL, XZ, CC, and X-LQ authored the text jointly.

FUNDING

This work was supported by the National Institutes of Health under award numbers R01 MH116675 and R01 EY017077.

ACKNOWLEDGMENTS

We wish to thank Russell Jaffe and Balbir Singh for valuable comments on the manuscript.

REFERENCES

Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z. J., and Scanziani, M. (2012). A neural circuit for spatial summation in visual cortex. *Nature* 490, 226–231. doi: 10.1038/nature11526

Asaad, W. F., Rainer, G., and Miller, E. K. (1998). Neural activity in the primate prefrontal cortex during associative learning. *Neuron* 21, 1399–1407. doi: 10.1016/s0896-6273(00)80658-3

Atallah, B. V., Bruns, W., Carandini, M., and Scanziani, M. (2012). Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* 73, 159–170. doi: 10.1016/j.neuron.2011.12.013

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* 63, 1–29. doi: 10.1146/annurev-psych-120710-100422

Barbas, H. (2015). General cortical and special prefrontal connections: principles from structure to function. *Annu. Rev. Neurosci.* 38, 269–289. doi: 10.1146/annurev-neuro-071714-033936

Blackman, R. K., Crowe, D. A., Denicola, A. L., Sakellaridi, S., Macdonald, A. W. III., and Chafee, M. V. (2016). Monkey prefrontal neurons reflect logical operations for cognitive control in a variant of the ax continuous performance task (AX-CPT). *J. Neurosci.* 36, 4067–4079. doi: 10.1523/JNEUROSCI.3578-15.2016

Borra, E., Ferroni, C. G., Gerbella, M., Giorgetti, V., Mangiaracina, C., Rozzi, S., et al. (2019). Rostro-caudal connectional heterogeneity of the dorsal part of the macaque prefrontal area 46. *Cereb. Cortex* 29, 485–504. doi: 10.1093/cercor/bhx332

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

Brzosko, Z., Mierau, S. B., and Paulsen, O. (2019). Neuromodulation of spike-timing-dependent plasticity: past, present, and future. *Neuron* 103, 563–581. doi: 10.1016/j.neuron.2019.05.041

Buckley, M. J., Mansouri, F. A., Hoda, H., Mahboubi, M., Browning, P. G., Kwok, S. C., et al. (2009). Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. *Science* 325, 52–58. doi: 10.1126/science.1172377

Burt, J. B., Demirtas, M., Eckner, W. J., Navejar, N. M., Ji, J. L., Martin, W. J., et al. (2018). Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* 21, 1251–1259. doi: 10.1038/s41593-018-0195-0

Cavada, C., and Goldman-Rakic, P. S. (1989). Posterior parietal cortex in rhesus monkey: I. Parcellation of areas based on distinctive limbic and sensory corticocortical connections. *J. Comp. Neurol.* 287, 393–421. doi: 10.1002/cne.902870402

Chen, G., Greengard, P., and Yan, Z. (2004). Potentiation of NMDA receptor currents by dopamine D1 receptors in prefrontal cortex. *Proc. Natl. Acad. Sci. U S A* 101, 2596–2600. doi: 10.1073/pnas.0308618100

Chen, N., Sugihara, H., and Sur, M. (2015). An acetylcholine-activated microcircuit drives temporal dynamics of cortical activity. *Nat. Neurosci.* 18, 892–902. doi: 10.1038/nn.4002

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

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

Constantinidis, C., Franowicz, M. N., and Goldman-Rakic, P. S. (2001b). The sensory nature of mnemonic representation in the primate prefrontal cortex. *Nat. Neurosci.* 4, 311–316. doi: 10.1038/85179

Constantinidis, C., Funahashi, S., Lee, D., Murray, J. D., Qi, X. L., Wang, M., et al. (2018). Persistent spiking activity underlies working memory. *J. Neurosci.* 38, 7020–7028. doi: 10.1523/JNEUROSCI.2486-17.2018

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

Constantinidis, C., and Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nat. Rev. Neurosci.* 17, 438–449. doi: 10.1038/nrn.2016.43

Constantinidis, C., and Luna, B. (2019). Neural substrates of inhibitory control maturation in adolescence. *Trends Neurosci.* 42, 604–616. doi: 10.1016/j.tins.2019.07.004

Constantinidis, C., and Procyk, E. (2004). The primate working memory networks. *Cogn. Affect. Behav. Neurosci.* 4, 444–465. doi: 10.3758/cabn.4.4.444

Constantinidis, C., and Qi, X. L. (2018). Representation of spatial and feature information in the monkey dorsal and ventral prefrontal cortex. *Front. Integr. Neurosci.* 12:31. doi: 10.3389/fnint.2018.00031

Constantinidis, C., and Wang, X. J. (2004). A neural circuit basis for spatial working memory. *Neuroscientist* 10, 553–565. doi: 10.1177/1073858404268742

Constantinidis, C., Williams, G. V., and Goldman-Rakic, P. S. (2002). A role for inhibition in shaping the temporal flow of information in prefrontal cortex. *Nat. Neurosci.* 5, 175–180. doi: 10.1038/nn799

Crowe, D. A., Goodwin, S. J., Blackman, R. K., Sakellaridi, S., Sponheim, S. R., Macdonald, A. W., et al. (2013). Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. *Nat. Neurosci.* 16, 1484–1491. doi: 10.1038/nn.3509

Dasilva, M., Brandt, C., Gotthardt, S., Gieselmann, M. A., Distler, C., and Thiele, A. (2019). Cell class-specific modulation of attentional signals by acetylcholine in macaque frontal eye field. *Proc. Natl. Acad. Sci. U S A* 116, 20180–20189. doi: 10.1073/pnas.1905413116

Defelipe, J., González-Albo, M. C., Del Río, M. R., and Elston, G. N. (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, 515–526. doi: 10.1002/(sici)1096-9861(19990927)412:3<515::aid-cne10>3.0.co;2-1

DeSouza, J. F., Menon, R. S., and Everling, S. (2003). Preparatory set associated with pro-saccades and anti-saccades in humans investigated with event-related FMRI. *J. Neurophysiol.* 89, 1016–1023. doi: 10.1152/jn.00562.2002

Dienel, S. J., and Lewis, D. A. (2019). Alterations in cortical interneurons and cognitive function in schizophrenia. *Neurobiol. Dis.* 131:104208. doi: 10.1016/j.nbd.2018.06.020

Durstewitz, D., Seamans, J. K., and Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nat. Neurosci.* 3, 1184–1191. doi: 10.1038/81460

Elston, G. N. (2000). Pyramidal cells of the frontal lobe: all the more spinous to think with. *J. Neurosci.* 20:RC95. doi: 10.1523/JNEUROSCI.20-18-j0002.2000

Elston, G. N. (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, 1313–1318. doi: 10.1046/j.1460-9568.2003.02552.x

Elston, G. N., and González-Albo, M. C. (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, 19–30. doi: 10.1159/000071957

Fish, K. N., Rocco, B. R., and Lewis, D. A. (2018). Laminar distribution of subsets of GABAergic axon terminals in human prefrontal cortex. *Front. Neuroanat.* 12:9. doi: 10.3389/fnana.2018.00009

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

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

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

Gabbott, P. L., and Bacon, S. J. (1997). Vasoactive intestinal polypeptide containing neurones in monkey medial prefrontal cortex (mPFC): colocalisation with calretinin. *Brain Res.* 744, 179–184. doi: 10.1016/s0006-8993(96)01232-2

García-Cabezas, M. A., Joyce, M. K. P., John, Y. J., Zikopoulos, B., and Barbas, H. (2017). Mirror trends of plasticity and stability indicators in primate prefrontal cortex. *Eur. J. Neurosci.* 46, 2392–2405. doi: 10.1111/ejn.13706

Gerbella, M., Borra, E., Tonelli, S., Rozzi, S., and Luppino, G. (2013). Connectional heterogeneity of the ventral part of the macaque area 46. *Cereb. Cortex* 23, 967–987. doi: 10.1093/cercor/bhs096

Glasser, M. F., and Van Essen, D. C. (2011). Mapping human cortical areas *in vivo* based on myelin content as revealed by T1- and T2-weighted MRI. *J. Neurosci.* 31, 11597–11616. doi: 10.1523/JNEUROSCI.2180-11.2011

Goldman-Rakic, P. S. (1984). Modular organization of prefrontal cortex. *Trends Neurosci.* 7, 419–424. doi: 10.1016/s0166-2236(84)80146-0

Gonzalez-Burgos, G., Miyamae, T., Pafundo, D. E., Yoshino, H., Rotaru, D. C., Hoftman, G., et al. (2015). Functional maturation of GABA synapses during postnatal development of the monkey dorsolateral prefrontal cortex. *Cereb. Cortex* 25, 4076–4093. doi: 10.1093/cercor/bhu122

Goulas, A., Stiers, P., Hutchison, R. M., Everling, S., Petrides, M., and Margulies, D. S. (2017). Intrinsic functional architecture of the macaque dorsal and ventral lateral frontal cortex. *J. Neurophysiol.* 117, 1084–1099. doi: 10.1152/jn.00486.2016

Guo, Z. V., Inagaki, H. K., Daie, K., Druckmann, S., Gerfen, C. R., and Svoboda, K. (2017). Maintenance of persistent activity in a frontal thalamocortical loop. *Nature* 545, 181–186. doi: 10.1038/nature22324

Hempel, C. M., Hartman, K. H., Wang, X. J., Turrigiano, G. G., and Nelson, S. B. (2000). Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex. *J. Neurophysiol.* 83, 3031–3041. doi: 10.1152/jn.2000.83.5.3031

Huntenburg, J. M., Bazin, P. L., Goulas, A., Tardif, C. L., Villringer, A., and Margulies, D. S. (2017). A systematic relationship between functional connectivity and intracortical myelin in the human cerebral cortex. *Cereb. Cortex* 27, 981–997. doi: 10.1093/cercor/bhx030

Jacob, S. N., and Nieder, A. (2014). Complementary roles for primate frontal and parietal cortex in guarding working memory from distractor stimuli. *Neuron* 83, 226–237. doi: 10.1016/j.neuron.2014.05.009

Jacob, S. N., Stalter, M., and Nieder, A. (2016). Cell-type-specific modulation of targets and distractors by dopamine D1 receptors in primate prefrontal cortex. *Nat. Commun.* 7:13218. doi: 10.1038/ncomms13218

Kadohisa, M., Kusunoki, M., Petrov, P., Sigala, N., Buckley, M. J., Gaffan, D., et al. (2015). Spatial and temporal distribution of visual information coding in lateral prefrontal cortex. *Eur. J. Neurosci.* 41, 89–96. doi: 10.1111/ejn.12754

Katsuki, F., Qi, X. L., Meyer, T., Kostelic, P. M., Salinas, E., and Constantinidis, C. (2014). Differences in intrinsic functional organization between dorsolateral prefrontal and posterior parietal cortex. *Cereb. Cortex* 24, 2334–2349. doi: 10.1093/cercor/bht087

Kiani, R., Cueva, C. J., Reppas, J. B., Peixoto, D., Ryu, S. I., and Newsome, W. T. (2015). Natural grouping of neural responses reveals spatially segregated clusters in prearcuate cortex. *Neuron* 85, 1359–1373. doi: 10.1016/j.neuron.2015.02.014

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C. E., Kepecs, A., Mainen, Z. F., et al. (2014). Demixed principal component analysis of population activity in higher cortical areas reveals independent representation of task parameters. *arXiv*:1410.6031 [Preprint]. https://doi.org/10.7554/eLife.10989.001

Konecky, R. O., Smith, M. A., and Olson, C. R. (2017). Monkey prefrontal neurons during Sternberg task performance: full contents of working memory or most recent item? *Psychon. Bull. Rev.* 117, 2269–2281. doi: 10.1152/jn.00541.2016

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

Kuboshima-Amemori, S., and Sawaguchi, T. (2007). Plasticity of the primate prefrontal cortex. *Neuroscientist* 13, 229–240. doi: 10.1177/1073858406298554

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

Kühn, S., Schmiedek, F., Noack, H., Wenger, E., Bodammer, N. C., Lindenberger, U., et al. (2013). The dynamics of change in striatal activity following updating training. *Hum. Brain Mapp.* 34, 1530–1541. doi: 10.1002/hbm.22007

Lara, A. H., and Wallis, J. D. (2014). Executive control processes underlying multi-item working memory. *Nat. Neurosci.* 17, 876–883. doi: 10.1038/nn.3702

Laroche, S., Davis, S., and Jay, T. M. (2000). Plasticity at hippocampal to prefrontal cortex synapses: dual roles in working memory and consolidation. *Hippocampus* 10, 438–446. doi: 10.1002/1098-1063(2000)10:4<438::aid-hipo10>3.0.co;2-3

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

Leavitt, M. L., Pieper, F., Sachs, A. J., and Martinez-Trujillo, J. C. (2017b). Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles. *Proc. Natl. Acad. Sci. U S A* 114, E2494–E2503. doi: 10.1073/pnas.1619949114

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

Lundqvist, M., Herman, P., and Miller, E. K. (2018). Working memory: delay activity, yes! persistent activity? Maybe not. *J. Neurosci.* 38, 7013–7019. doi: 10.1523/JNEUROSCI.2485-17.2018

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

Major, A. J., Vijayraghavan, S., and Everling, S. (2015). Muscarinic attenuation of mnemonic rule representation in macaque dorsolateral prefrontal cortex during a pro- and anti-saccade task. *J. Neurosci.* 35, 16064–16076. doi: 10.1523/JNEUROSCI.2454-15.2015

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

Masse, N. Y., Hodnefield, J. M., and Freedman, D. J. (2017). Mnemonic encoding and cortical organization in parietal and prefrontal cortices. *J. Neurosci.* 37, 6098–6112. doi: 10.1523/JNEUROSCI.3903-16.2017

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

McEwen, B. S., and Morrison, J. H. (2013). The brain on stress: vulnerability and plasticity of the prefrontal cortex over the life course. *Neuron* 79, 16–29. doi: 10.1016/j.neuron.2013.06.028

Mejias, J. F., and Wang, X. J. (2019). Mechanisms of distributed working memory in a large-scale model of the macaque neocortex. *BioRxiv* [Preprint]. doi: 10.1101/760231

Melchitzky, D. S., and Lewis, D. A. (2008). Dendritic-targeting GABA neurons in monkey prefrontal cortex: comparison of somatostatin- and calretinin-immunoreactive axon terminals. *Synapse* 62, 456–465. doi: 10.1002/syn.20514

Mendoza-Halliday, D., and Martinez-Trujillo, J. C. (2017). Neuronal population coding of perceived and memorized visual features in the lateral prefrontal cortex. *Nat. Commun.* 8:15471. doi: 10.1038/ncomms15471

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

Meskenaite, V. (1997). Calretinin-immunoreactive local circuit neurons in area 17 of the cynomolgus monkey, Macaca fascicularis. *J. Comp. Neurol.* 379, 113–132. doi: 10.1002/(sici)1096-9861(19970303)379:1<113::aid-cne8>3.0.co;2-7

Meyer, T., Qi, X. L., and Constantinidis, C. (2007). Persistent discharges in the prefrontal cortex of monkeys naive to working memory tasks. *Cereb. Cortex* 17, i70–i76. doi: 10.1093/cercor/bhm063

Meyer, T., Qi, X. L., Stanford, T. R., and Constantinidis, C. (2011). Stimulus selectivity in dorsal and ventral prefrontal cortex after training in working memory tasks. *J. Neurosci.* 31, 6266–6276. doi: 10.1523/JNEUROSCI.6798-10.2011

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., and Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J. Neurophysiol.* 100, 1407–1419. doi: 10.1152/jn.90248.2008

Meyers, E. M., Qi, X. L., and Constantinidis, C. (2012). Incorporation of new information into prefrontal cortical activity after learning working memory tasks. *Proc. Natl. Acad. Sci. U S A* 109, 4651–4656. doi: 10.1073/pnas.1201022109

Mi, Y., Katkov, M., and Tsodyks, M. (2017). Synaptic correlates of working memory capacity. *Neuron* 93, 323–330. doi: 10.1016/j.neuron.2016.12.004

Montez, D. F., Calabro, F. J., and Luna, B. (2019). Working memory improves developmentally as neural processes stabilize. *PLoS One* 14:e0213010. doi: 10.1371/journal.pone.0213010

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. *Nat. Neurosci.* 17, 1410–1417. doi: 10.1038/nn.3807

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

Olesen, P. J., Westerberg, H., and Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nat. Neurosci.* 7, 75–79. doi: 10.1038/nn1165

Ordaz, S., Davis, S., and Luna, B. (2010). Effects of response preparation on developmental improvements in inhibitory control. *Acta Psychol. Amst.* 134, 253–263. doi: 10.1016/j.actpsy.2010.02.007

Ott, T., Jacob, S. N., and Nieder, A. (2014). Dopamine receptors differentially enhance rule coding in primate prefrontal cortex neurons. *Neuron* 84, 1317–1328. doi: 10.1016/j.neuron.2014.11.012

Parthasarathy, A., Tang, C., Herikstad, R., Cheong, L. F., Yen, S. C., and Libedinsky, C. (2019). Time-invariant working memory representations in the presence of code-morphing in the lateral prefrontal cortex. *Nat. Commun.* 10:4995. doi: 10.1038/s41467-019-12841-y

Petrides, M. (2000). Dissociable roles of mid-dorsolateral prefrontal and anterior inferotemporal cortex in visual working memory. *J. Neurosci.* 20, 7496–7503. doi: 10.1523/JNEUROSCI.20-19-07496.2000

Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 781–795. doi: 10.1098/rstb.2005.1631

Petrides, M., and Pandya, D. N. (1984). Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *J. Comp. Neurol.* 228, 105–116. doi: 10.1002/cne.902280110

Pi, H. J., Hangya, B., Kvitsiani, D., Sanders, J. I., Huang, Z. J., and Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control. *Nature* 503, 521–524. doi: 10.1038/nature12676

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

Pucak, M. L., Levitt, J. B., Lund, J. S., and Lewis, D. A. (1996). Patterns of intrinsic and associational circuitry in monkey prefrontal cortex. *J. Comp. Neurol.* 376, 614–630. doi: 10.1002/(sici)1096-9861(19961223)376:4<614::aid-cne9>3.0.co;2-4

Qi, X. L., and Constantinidis, C. (2012a). Correlated discharges in the primate prefrontal cortex before and after working memory training. *Eur. J. Neurosci.* 36, 3538–3548. doi: 10.1111/j.1460-9568.2012.08267.x

Qi, X. L., and Constantinidis, C. (2012b). Variability of prefrontal neuronal discharges before and after training in a working memory task. *PLoS One* 7:e41053. doi: 10.1371/journal.pone.0041053

Qi, X. L., and Constantinidis, C. (2013). Neural changes after training to perform cognitive tasks. *Behav. Brain Res.* 241, 235–243. doi: 10.1016/j.bbr.2012.12.017

Qi, X. L., Katsuki, F., Meyer, T., Rawley, J. B., Zhou, X., Douglas, K. L., et al. (2010). Comparison of neural activity related to working memory in primate dorsolateral prefrontal and posterior parietal cortex. *Front. Syst. Neurosci.* 4:12. doi: 10.3389/fnsys.2010.00012

Qi, X. L., Liu, R., Vazdarjanova, A. I., Blake, D. T., and Constantinidis, C. (2019). Nucleus basalis stimulation enhances working memory and stabilizes attractor networks in prefrontal cortex. *bioRxiv* [Preprint]. doi: 10.1101/674465

Qi, X. L., Meyer, T., Stanford, T. R., and Constantinidis, C. (2011). Changes in prefrontal neuronal activity after learning to perform a spatial working memory task. *Cereb. Cortex* 21, 2722–2732. doi: 10.1093/cercor/bhr058

Qi, X. L., Zhou, X., and Constantinidis, C. (2015). "Neurophysiological mechanisms of working memory: cortical specialization and plasticity," in *Attention and Performance XXV*, eds P. Jolicoeur, C. Lefebre and J. C. Martinez-Trujillo (London: Academic Press), 171–186.

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

Rao, S. G., Williams, G. V., and Goldman-Rakic, P. S. (1999). Isodirectional tuning of adjacent interneurons and pyramidal cells during working memory: evidence for microcolumnar organization in PFC. *J. Neurophysiol.* 81, 1903–1916. doi: 10.1152/jn.1999.81.4.1903

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

Riley, M. R., and Constantinidis, C. (2016). Role of prefrontal persistent activity in working memory. *Front. Syst. Neurosci.* 9:181. doi: 10.3389/fnsys.2015.00181

Riley, M. R., Qi, X. L., and Constantinidis, C. (2017). Functional specialization of areas along the anterior-posterior axis of the primate prefrontal cortex. *Cereb. Cortex* 27, 3683–3697. doi: 10.1093/cercor/bhw190

Riley, M. R., Qi, X. L., Zhou, X., and Constantinidis, C. (2018). Anterior-posterior gradient of plasticity in primate prefrontal cortex. *Nat. Commun.* 9:3790. doi: 10.1038/s41467-018-06226-w

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

Sarma, A., Masse, N. Y., Wang, X. J., and Freedman, D. J. (2016). Task-specific versus generalized mnemonic representations in parietal and prefrontal cortices. *Nat. Neurosci.* 19, 143–149. doi: 10.1038/nn.4168

Schneiders, J. A., Opitz, B., Krick, C. M., and Mecklinger, A. (2011). Separating intra-modal and across-modal training effects in visual working memory: an fMRI investigation. *Cereb. Cortex* 21, 2555–2564. doi: 10.1093/cercor/bhr037

Schweizer, S., Grahn, J., Hampshire, A., Mobbs, D., and Dalgleish, T. (2013). Training the emotional brain: improving affective control through emotional working memory training. *J. Neurosci.* 33, 5301–5311. doi: 10.1523/jneurosci.2593-12.2013

Seamans, J. K., Durstewitz, D., Christie, B. R., Stevens, C. F., and Sejnowski, T. J. (2001). Dopamine D1/D5 receptor modulation of excitatory synaptic inputs to layer V prefrontal cortex neurons. *Proc. Natl. Acad. Sci. U S A* 98, 301–306. doi: 10.1073/pnas.98.1.301

Selemon, L. D., and Goldman-Rakic, P. S. (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, 4049–4068. doi: 10.1523/jneurosci.08-11-04049.1988

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

Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* 19, 394–405. doi: 10.1016/j.tics.2015.05.004

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

Sun, Y., Yang, Y., Galvin, V. C., Yang, S., Arnsten, A. F., and Wang, M. (2017). Nicotinic α4β2 cholinergic receptor influences on dorsolateral prefrontal cortical neuronal firing during a working memory task. *J. Neurosci.* 37, 5366–5377. doi: 10.1523/jneurosci.0364-17.2017

Takeuchi, H., Taki, Y., Nouchi, R., Hashizume, H., Sekiguchi, A., Kotozaki, Y., et al. (2013). Effects of working memory training on functional connectivity and cerebral blood flow during rest. *Cortex* 49, 2106–2125. doi: 10.1016/j.cortex.2012.09.007

Tang, H., Qi, X. L., Riley, M. R., and Constantinidis, C. (2019). Working memory capacity is enhanced by distributed prefrontal activation and invariant temporal dynamics. *Proc. Natl. Acad. Sci. U S A* 116, 7095–7100. doi: 10.1073/pnas.1817278116

Tang, H., Riley, M. R., and Constantinidis, C. (2017). Lateralization of executive function: working memory advantage for same hemifield stimuli in the monkey. *Front. Neurosci.* 11:532. doi: 10.3389/fnins.2017.00532

Urban-Ciecko, J., Jouhanneau, J. S., Myal, S. E., Poulet, J. F. A., and Barth, A. L. (2018). Precisely timed nicotinic activation drives SST inhibition in neocortical circuits. *Neuron* 97, 611.e5–625.e5. doi: 10.1016/j.neuron.2018.01.037

Vijayraghavan, S., Wang, M., Birnbaum, S. G., Williams, G. V., and Arnsten, A. F. (2007). Inverted-U dopamine D1 receptor actions on prefrontal neurons engaged in working memory. *Nat. Neurosci.* 10, 376–384. doi: 10.1038/nn1846

Walker, A. E. (1940). A cytoarchitectural study of the prefrontal area of the macaque monkey. *J. Comp. Neurol.* 73, 59–86. doi: 10.1002/cne.900730106

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

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

Wang, M., and Arnsten, A. F. (2015). Contribution of NMDA receptors to dorsolateral prefrontal cortical networks in primates. *Neurosci. Bull.* 31, 191–197. doi: 10.1007/s12264-014-1504-6

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

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

Wang, X. J., Tegner, J., Constantinidis, C., and Goldman-Rakic, P. S. (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, 1368–1373. doi: 10.1073/pnas.0305337101

Watanabe, K., and Funahashi, S. (2014). Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nat. Neurosci.* 17, 601–611. doi: 10.1038/nn.3667

Wilson, N. R., Runyan, C. A., Wang, F. L., and Sur, M. (2012). Division and subtraction by distinct cortical inhibitory networks *in vivo*. *Nature* 488, 343–348. doi: 10.1038/nature11347

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

Yang, Y., Paspalas, C. D., Jin, L. E., Picciotto, M. R., Arnsten, A. F., and Wang, M. (2013). Nicotinic α7 receptors enhance NMDA cognitive circuits in dorsolateral prefrontal cortex. *Proc. Natl. Acad. Sci. U S A* 110, 12078–12083. doi: 10.1073/pnas.1307849110

Yang, C. R., and Seamans, J. K. (1996). Dopamine D1 receptor actions in layers V-VI rat prefrontal cortex neurons *in vitro*: modulation of dendritic-somatic signal integration. *J. Neurosci.* 16, 1922–1935. doi: 10.1523/jneurosci.16-05-01922.1996

Yu, J., Hu, H., Agmon, A., and Svoboda, K. (2019). Recruitment of GABAergic interneurons in the barrel cortex during active tactile behavior. *Neuron* 104, 412.e4–427.e4. doi: 10.1016/j.neuron.2019.07.027

Zhou, X., Katsuki, F., Qi, X. L., and 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, 31–38. doi: 10.1152/jn.01151.2011

Zhou, X., Qi, X. L., and Constantinidis, C. (2016a). Distinct roles of the prefrontal and posterior parietal cortices in response inhibition. *Cell Rep.* 14, 2765–2773. doi: 10.1016/j.celrep.2016.02.072

Zhou, X., Zhu, D., Qi, X. L., Li, S., King, S. G., Salinas, E., et al. (2016b). Neural correlates of working memory development in adolescent primates. *Nat. Commun.* 7:13423. doi: 10.1038/ncomms13423

Zhou, X., Qi, X. L., Douglas, K., Palaninathan, K., Kang, H. S., Buccafusco, J. J., et al. (2011). Cholinergic modulation of working memory activity in primate prefrontal cortex. *J. Neurophysiol.* 106, 2180–2188. doi: 10.1152/jn.00148.2011

Zhou, X., Zhu, D., Katsuki, F., Qi, X. L., Lees, C. J., Bennett, A. J., et al. (2014). Age-dependent changes in prefrontal intrinsic connectivity. *Proc. Natl. Acad. Sci. U S A* 111, 3853–3858. doi: 10.1073/pnas.1316594111

Zhou, X., Zhu, D., Qi, X. L., Lees, C. J., Bennett, A. J., Salinas, E., et al. (2013). Working memory performance and neural activity in the prefrontal cortex of peri-pubertal monkeys. *J. Neurophysiol.* 110, 2648–2660. doi: 10.1152/jn.00370.2013

Zick, J. L., Blackman, R. K., Crowe, D. A., Amirikian, B., Denicola, A. L., Netoff, T. I., et al. (2018). Blocking NMDAR disrupts spike timing and decouples monkey prefrontal circuits: implications for activity-dependent disconnection in schizophrenia. *Neuron* 98, 1243.e5–1255.e5. doi: 10.1016/j.neuron.2018.05.010

Zylberberg, J., and Strowbridge, B. W. (2017). Mechanisms of persistent activity in cortical circuits: possible neural substrates for working memory. *Annu. Rev. Neurosci.* 40, 603–627. doi: 10.1146/annurev-neuro-070815-014006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neural Circuits

REVIEW
published: 21 December 2020
doi: 10.3389/fncir.2020.615626

A Diversity of Intrinsic Timescales Underlie Neural Computations

Sean E. Cavanagh[1]*, *Laurence T. Hunt*[1,2,3,4]*† *and Steven W. Kennerley*[1]*†

[1]*Department of Clinical and Movement Neurosciences, University College London, London, United Kingdom,* [2]*Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom,* [3]*Max Planck-UCL Centre for Computational Psychiatry and Aging, University College London, London, United Kingdom,* [4]*Department of Psychiatry, Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, United Kingdom*

Neural processing occurs across a range of temporal scales. To facilitate this, the brain uses fast-changing representations reflecting momentary sensory input alongside more temporally extended representations, which integrate across both short and long temporal windows. The temporal flexibility of these representations allows animals to behave adaptively. Short temporal windows facilitate adaptive responding in dynamic environments, while longer temporal windows promote the gradual integration of information across time. In the cognitive and motor domains, the brain sets overarching goals to be achieved within a long temporal window, which must be broken down into sequences of actions and precise movement control processed across much shorter temporal windows. Previous human neuroimaging studies and large-scale artificial network models have ascribed different processing timescales to different cortical regions, linking this to each region's position in an anatomical hierarchy determined by patterns of inter-regional connectivity. However, even within cortical regions, there is variability in responses when studied with single-neuron electrophysiology. Here, we review a series of recent electrophysiology experiments that demonstrate the heterogeneity of temporal receptive fields at the level of single neurons within a cortical region. This heterogeneity appears functionally relevant for the computations that neurons perform during decision-making and working memory. We consider anatomical and biophysical mechanisms that may give rise to a heterogeneity of timescales, including recurrent connectivity, cortical layer distribution, and neurotransmitter receptor expression. Finally, we reflect on the computational relevance of each brain region possessing a heterogeneity of neuronal timescales. We argue that this architecture is of particular importance for sensory, motor, and cognitive computations.

Keywords: neuronal timescale, autocorrelation, time constant, decision-making, working memory

OPEN ACCESS

Edited by:
Emmanuel Procyk, Institut National de la Santé et de la Recherche Médicale (INSERM), France

Reviewed by:
Ken-ichi Amemori, Kyoto University, Japan
Timothy D. Hanks, University of California, Davis, United States

***Correspondence:**
Sean E. Cavanagh
sean.cavanagh.12@ucl.ac.uk
Laurence T. Hunt
laurence.hunt@psych.ox.ac.uk
Steven W. Kennerley
s.kennerley@ucl.ac.uk

†These authors have contributed equally to this work

Received: *09 October 2020*
Accepted: *18 November 2020*
Published: *21 December 2020*

Citation:
Cavanagh SE, Hunt LT and Kennerley SW (2020) A Diversity of Intrinsic Timescales Underlie Neural Computations. Front. Neural Circuits 14:615626. doi: 10.3389/fncir.2020.615626

INTRODUCTION

Imagine you are listening to Beethoven's 9th symphony. As you listen, neurons in the auditory cortex are responding to the momentary pitch of the music. In isolation, these momentary pitches are meaningless. The notes must be contextualized across bars (seconds), melodies (tens of seconds), and movements (minutes) for the music to be appreciated and understood. The beauty of the music depends upon melodic expectations that are established over both long and short timescales. The neural processing of information across a diversity of timescales is not only key to many aspects of perception, but also cognition and motor control.

TIMESCALES OF CORTICAL REGIONS REFLECT HIERARCHY

More formally, we will consider the neural temporal receptive field as the length of time over which inputs can be integrated by a neural substrate. Previous work has established the notion of temporal receptive fields, and characterized the temporal properties of neural activity in response to sensory stimuli (Sen et al., 2001; Kiebel et al., 2008; Chen et al., 2015; Hasson et al., 2015). This work has revealed many relevant parallels with the more established concept of spatial receptive fields. It has long been known that neurons in different cortical areas process sensory information across different spatial scales. As information becomes more highly processed, neural representations are based upon larger physical areas and contain more abstract representations which require the integration of multiple sources of sensory information. As a general principle, the size and complexity of spatial receptive fields increase along a visual hierarchy (Lennie, 1998). For example, neurons in the early visual cortex encode the presence of simple features of stimuli (e.g., orientation) in a small, specific area of the visual field (Hubel and Wiesel, 1962, 1968). At the other end of the ventral visual stream, neurons in the inferotemporal cortex encode high-level information about object identity, independent of its location in the visual field (Tanaka, 1996; Brincat and Connor, 2004; Chang and Tsao, 2017). A similar pattern of representational hierarchies is also present in the motor domain, with receptive field sizes increasing and more complex motor representations becoming evident, such as selectivity for sequences of actions, as you move from the primary motor cortex more anteriorly to premotor and prefrontal regions (Luppino et al., 1991; Picard and Strick, 1996; Shima and Tanji, 2000; Nachev et al., 2008; Russo et al., 2020; but see also, Yokoi and Diedrichsen, 2019). Furthermore, in the cognitive domain, complex representations are evident mainly in the prefrontal cortex (Wallis et al., 2001), which also exhibits a hierarchical anatomical organization of abstract representations (Koechlin et al., 2003; Badre, 2008; Nee and D'Esposito, 2016).

When reviewing *temporal* receptive fields, we will initially apply a similar framework and consider how representation size and complexity could vary across neural substrates. In the temporal domain, as a possible equivalent to the neuronal diversity in representing spatial scale, neurons may signal an event (e.g., a sensory stimulus, action, or goal) for varying lengths of time after it occurs. Some neurons may represent this information with a fixed pattern of activity, invariant of how long ago it occurred, within a set temporal window (e.g., 5 s). The length of this window may vary across neurons, and the representation carried by other neurons may be restricted to when the event initially occurs. In higher cortical areas, the temporal receptive window may also be task-dependent, as demonstrated (for example) in working memory tasks with variable delays (Funahashi et al., 1989) or in time estimation tasks with variable durations (Wang et al., 2018). By possessing a spectrum of these representations concurrently, it would allow the brain to hold salient information in working memory while continuing to monitor fluctuations in the environment.

We can also consider the complexity of information in the temporal domain. Stimuli often vary across time, and information must be temporally integrated to enable perception. Further to the musical symphony analogy presented at the start of this piece, another good example is language comprehension (Hasson et al., 2015). To understand speech, the brain integrates auditory information over tens of milliseconds to detect words, which in turn are combined over several seconds to form sentences, which are then integrated across minutes to facilitate the understanding of discourse. Another example would be when we take a journey. The overarching goal of the journey, across many minutes, is to reach a destination. But in order to reach this goal, we set subgoals which are achieved through sequences of actions (across seconds). These action sequences in turn require the precise co-ordination of muscle groups at a timescale of milliseconds. In both of these examples, different neural substrates likely underly the processing of information across different timescales. Therefore, a temporal receptive field may also constitute the length of time over which inputs can be combined or outputs organized—with higher complexity associated with longer integration times.

Recent work has begun to address how different *cortical regions* process information across different temporal scales. Several studies by Hasson and colleagues have utilized an innovative protocol to demonstrate this with human neuroimaging (see Hasson et al., 2015 for an in depth review; **Figure 1A**). Human subjects passively experienced a complex stimulus (e.g., listening to a story) across several minutes, before the stimulus was "scrambled" and presented again. For the scrambled versions, the original stimulus was fragmented to different degrees. For instance, some versions only reorganized the paragraphs of the story, whilst others shuffled the order of all of the words. Regardless of the degree of shuffling, fMRI activity recorded in early auditory cortices showed a high degree of inter-subject reliability. However, in higher cortical areas, reliable responses were only observed when scrambled stimuli preserved the structure of paragraphs (Lerner et al., 2011). The interpretation of these results was that early cortical regions processed momentary input regardless of its context, whereas in higher cortical regions information was processed across a much longer timescale. These findings have been demonstrated with various sensory modalities (i.e., auditory, visual, and audio-visual) and with different neuroimaging techniques (Hasson et al., 2008; Lerner et al., 2011; Honey et al., 2012). More recent studies have built on this work to directly infer the timescale over which activity is structured by applying a Hidden Markov Model to the time course of neural activity during movie-watching (Baldassano et al., 2017, 2018). This again reveals a nested hierarchy of timescales from lower to higher cortical areas, with responses in higher areas generalising to an audio description of the same story, while hippocampal activity demarcates high-level boundaries between distinct episodes in the movie. It is notable that a similar hierarchy of timescales can also be found by examining data acquired

during the resting state (Stephens et al., 2013), linking these findings to the rich literature on slow timescale interactions between large-scale brain regions while at rest (reviewed in Buckner et al., 2013).

In another line of work, researchers have indexed temporal scales of cortical regions by measuring the spike-count autocorrelation of single neuron activity recorded from macaque monkeys (Ogawa and Komatsu, 2010; Murray et al., 2014). Utilising task independent neural activity recorded during short (∼1,000 ms) pre-trial fixation periods, the decay rate of autocorrelation can be captured with an exponential equation and used to define the cortical region's intrinsic timescale (**Figure 1B**). When results from a large number of electrophysiological datasets were collated, there was a strong relationship between a region's position in the anatomical hierarchy (Felleman and Van Essen, 1991; Barbas and Rempel-Clower, 1997) and its intrinsic timescale (Murray et al., 2014; **Figure 1C**). Moreover, the potential functional relevance of resting spike-count autocorrelation was suggested such that regions with longer intrinsic timescales also contained neurons with longer task-related maintenance of reward information across trials (Bernacchia et al., 2011; Spitmaan et al., 2020).

A large-scale network model of interconnected regions, guided by anatomical data on hierarchical connectivity (Markov et al., 2014a) and local recurrent connectivity (Elston et al., 2011), was sufficient to reproduce this variation in intrinsic timescales (Chaudhuri et al., 2015; **Figure 1D**). In the model, individual neurons are embedded within densely interconnected networks. Areas of the frontal cortex are densely connected with multiple areas, whereas sensory areas have lower, and typically more local, connection densities (Chaudhuri et al., 2015; Wang and Kennedy, 2016). These connection patterns form cortical hierarchies, defined by asymmetric local (interlaminar) and extrinsic (long-range) connections (Bastos et al., 2015; Chaudhuri et al., 2015; Wang and Kennedy, 2016). These anatomical hierarchies result in long integrative timescales of neurons in frontal cortex, contrasted with short timescales of neurons in sensory areas (Romo et al., 1999; Wang, 2001, 2020; Kiebel et al., 2008; Benucci et al., 2009; Chaudhuri et al., 2015; Wang and Kennedy, 2016).

Although in this initial work variation in intrinsic timescales had only been assigned to brain regions as a whole, perhaps individual neurons within those regions were also capable of processing information across a diversity of timescales. For example, previous research on spatial receptive fields has demonstrated that although there is a general trend of higher cortical regions exhibiting larger spatial receptive fields (Lennie, 1998), in studies where larger numbers of visual neurons were recorded, a significant amount of within-region heterogeneity is also found (Blasdel and Fitzpatrick, 1984; Gur et al., 2005; Nauhaus et al., 2016; Siegle et al., 2019). It was therefore crucial to test whether single neurons had individual timescales, which varied *within* cortical regions.

If single neurons did indeed have their own temporal receptive fields, what would this imply for their roles in cognitive function? It is already established that there is a large degree of heterogeneity with which neurons in higher brain regions are involved in cognitive computations (Shafi et al., 2007; Jun et al., 2010; Wallis and Kennerley, 2010; Meister et al., 2013). It might therefore be the case that a neuron's intrinsic timescale determines its functional role in extended cognitive processes, such as decision-making and working-memory—specifically the neuron's strength and dynamics of information encoding. This could be examined by relating an individual neuron's encoding properties with its own intrinsic timescale, as opposed to the broader timescale of the brain region it inhabited.

Several studies have begun to address these questions with single neuron electrophysiology experiments in macaque monkeys. Here, we will review this work and consider its significant implications—specifically what it may tell us about how neural circuits are organized, how they compute information, and how we should go about studying them in future.

A DIVERSITY OF TIMESCALES AT THE SINGLE NEURON LEVEL

One of the first studies to examine single neuron intrinsic timescales using spike-count autocorrelation was (Cavanagh et al., 2016), which utilized electrophysiology data recorded from macaque monkeys during a value-based decision-making task (Hosokawa et al., 2013). Before the monkey began to make a choice, there was a 1,000 ms fixation period on each trial. The same spike-count autocorrelation analysis was applied (Murray et al., 2014; **Figure 1B**), with one important difference. Instead of pooling the autocorrelograms of all neurons within a brain region, a timescale was fitted for each individual neuron. Although this inevitably made the fitting process more noisy, and some neurons were poorly described by a simple exponential decay, the majority of neurons exhibited a decay in autocorrelation structure reliably quantified by an exponential function (see **Figure 2** for examples). Importantly, this analysis highlighted a striking degree of *within-region* variability (**Figure 2**)—even within the anterior cingulate cortex (ACC), which sat at the apex of the hierarchy identified in Murray et al. (2014); there was a spectrum of timescales including many neurons with short timescales. Further studies then applied the same single-neuron analysis to many different brain regions throughout the cortical hierarchy (**Figure 3**), including posterior parietal cortex (Wasmuht et al., 2018), lateral prefrontal cortex (lPFC; Cavanagh et al., 2016, 2018; Wasmuht et al., 2018; Fascianelli et al., 2019; Fontanier et al., 2020; Kim and Sejnowski, 2020), orbitofrontal cortex (OFC; Cavanagh et al., 2016, 2018; Fascianelli et al., 2019), cingulate cortex (Cavanagh et al., 2016, 2018; Fontanier et al., 2020), premotor cortex (Cirillo et al., 2018), and frontopolar cortex (Fascianelli et al., 2019). All of these regions contained single neurons with a diversity of timescales, suggesting that brain regions possessing a heterogonous distribution of timescales is a generalized feature of cortical organization. Despite this work predominantly focussing on higher-level cortical regions, it would be reasonable to predict that lower-level sensory regions may also contain

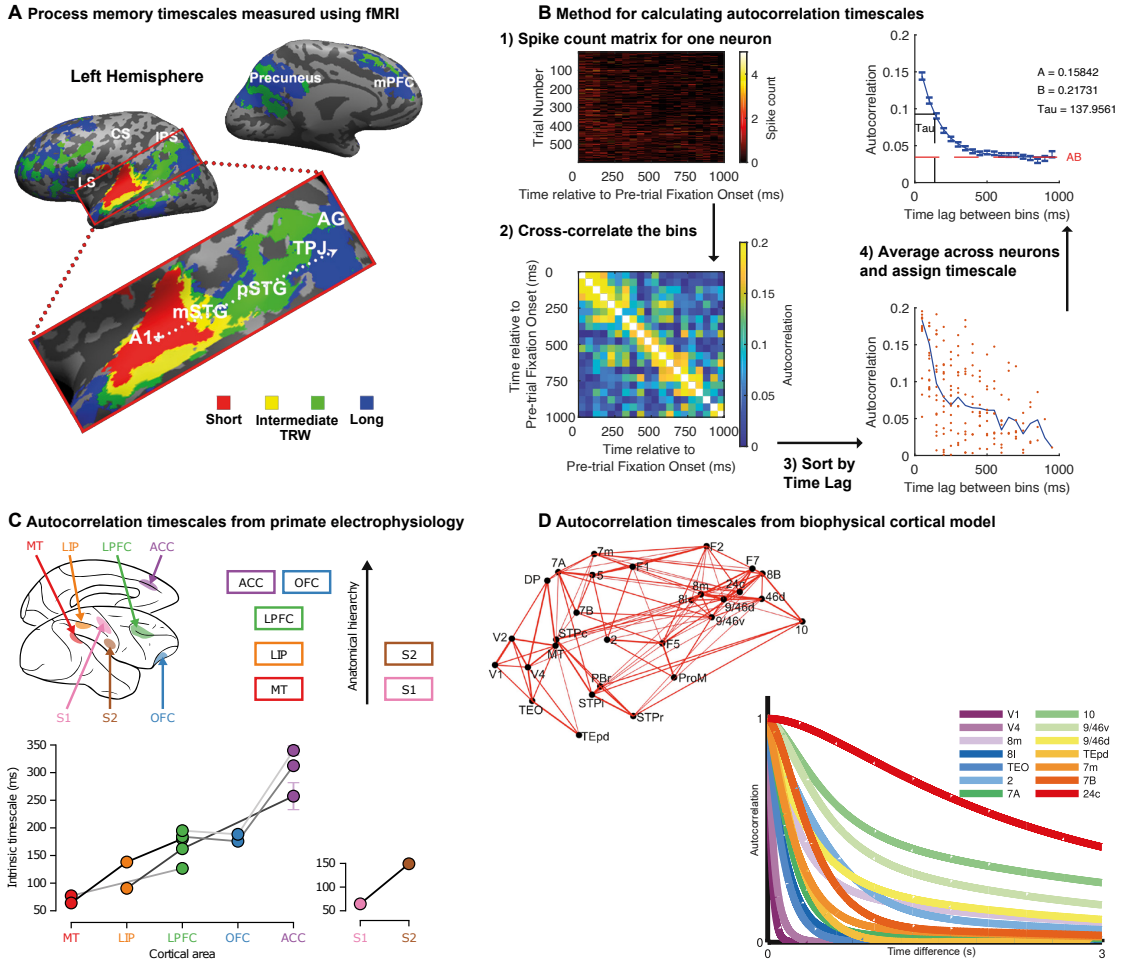


FIGURE 1 | Temporal receptive fields vary across cortical brain regions. **(A)** Topography of temporal receptive fields defined using fMRI data recorded during the passive listening to stories. The color in the voxel heatmap depicts the shortest period to which the original auditory story stimulus could be scrambled and a reliable inter-subject correlation still be obtained (red: story played backward; yellow: a story with word-order scrambled; green: a story with sentence-order scrambled; blue: a story with paragraph-order scrambled). Early auditory areas (A1+) were reliable across subjects even on the most scrambled stimuli, whereas activity in higher regions such as the temporal-parietal junction (TPJ) responded reliably only in the least scrambled condition. There was a gradual hierarchical progression of timescales along the temporal-parietal axis. Data originally published in Lerner et al. (2011), figure reproduced from Hasson et al. (2015), with permission. **(B)** Spike-count autocorrelation method for assigning neuronal timescales. Spike counts for each neuron during the pre-trial fixation periods are subdivided into non-overlapping 50 ms bins. This data from this matrix is correlated across trials to produce a measure of autocorrelation as a function of time-lag between bins. The data is averaged across all neurons recorded in a cortical region before the rate of decay is captured with an exponential fit. The tau parameter determines the intrinsic timescale of the cortical region. **(C)** Intrinsic timescales of seven cortical regions as a function of their position in the anatomical hierarchy. Regions further up in the hierarchy have longer timescales. Each of the different data points (circles) from each brain region were collected by a different research lab, with the lines between datapoints indicating multiple brain areas collected by the same research lab. Reproduced with permission from Murray et al. (2014). **(D)** A large-scale biophysically-realistic neural network simulation shows a hierarchy of timescales in response to visual input. An important feature of the model is the inclusion of anatomical data regarding inter-regional connectivity shown here. In the graph, as in panel **(B)**, autocorrelation is plotted as a function of time lag. Brain regions with more prolonged, stable autocorrelation functions are those at the apex of the cortical hierarchy. Reproduced with permission from Chaudhuri et al. (2015).

heterogenous timescales. This appears likely from single-neuron autocorrelograms (Murray et al., 2014), and is supported by a previous study which observed heterogeneity of autocorrelation decay when recording intracellularly from neurons in cat striate cortex (Azouz and Gray, 1999). Variability in single-neuron timescales has also been observed in several early visual areas of the mouse brain (Siegle et al., 2019), and it will be interesting to explore this more conclusively in future studies of primate cortex.

Once it had been established single neurons possessed different individual timescales, these studies next investigated whether this variation was functionally significant. A simple way to approach this question was to test the relationship between timescales quantified during the resting (fixation) period of the task with the strength of a neuron's subsequent task-related activity. For example, encoding of the value of the chosen option during a decision making task may arise as a consequence of the process of evidence integration during a temporally extended

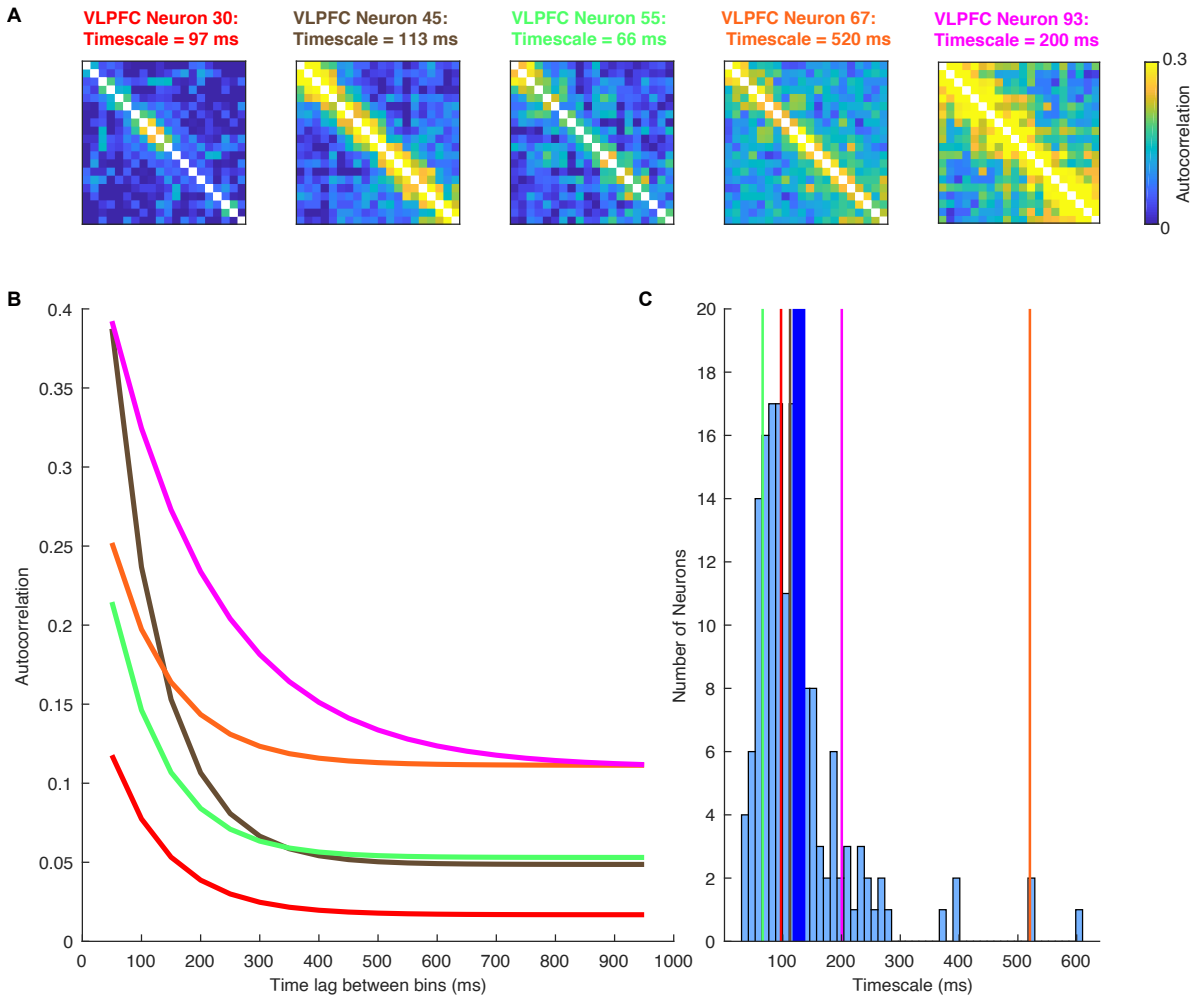


FIGURE 2 | A Heterogeneity of single-neuron timescales exist within a brain region. Data recorded from the ventrolateral prefrontal cortex (VLPFC) during a working memory task (Cavanagh et al., 2018). **(A)** Autocorrelation structure of five VLPFC neurons, plotted as a function of time within the pre-trial fixation period. As in **Figure 1B**, these are calculated by correlating the spike count autocorrelation across trials. Despite being recorded in the same brain region, there is a large degree of diversity. **(B)** Autocorrelation structure of VLPFC neurons, plotted as a function of the time lag between bins. As in **Figure 1B**, the data from above have been sorted by the time lag. Each of the lines corresponds to an exponential fit of the decaying autocorrelation of one of the neurons' heatmaps above (corresponding color). There is substantial heterogeneity in the individual neurons making up the whole region average. Each neuron has an exponential decay reasonably distinct from the population average. **(C)** Histogram showing the single neuron exponential decay time constant assigned to all neurons within VLPFC. The vertical lines mark the example neurons shown in this figure. The thicker blue line marks the population mean.

decision process (Hunt et al., 2012, 2015), or may also support maintaining value information until later in the trial when learning can occur by assigning credit to the chosen option (Rangel and Hare, 2010; Jocham et al., 2016; Enel et al., 2020). Because both evidence integration and working memory for value are temporally extended processes, it might be expected that single neurons with longer intrinsic timescales are more involved in these cognitive processes. This relationship could be explored by correlating timescales with the coding strength of each neuron, or alternatively by subdividing a brain region's entire population by a median split of timescales and comparing the task-related activity in the two groups (Cavanagh et al., 2016). Together, these analyses revealed that prefrontal neurons with longer timescales exhibited stronger chosen value coding when a decision was being made. Moreover, long timescale neurons in OFC continued to signal the chosen value until an outcome was received (**Figure 4A**). Neurons with longer timescales were therefore more involved in both choice and maintaining a representation of the expected outcome across delays which could support credit assignment processes.

The apparent relationship between intrinsic timescales and task-related processing can also be extended to another cognitive process—working memory. Whereas decision-making requires the gradual integration of evidence across time (Gold and Shadlen, 2007), working memory involves the maintenance of task-relevant information in the absence of direct sensory input

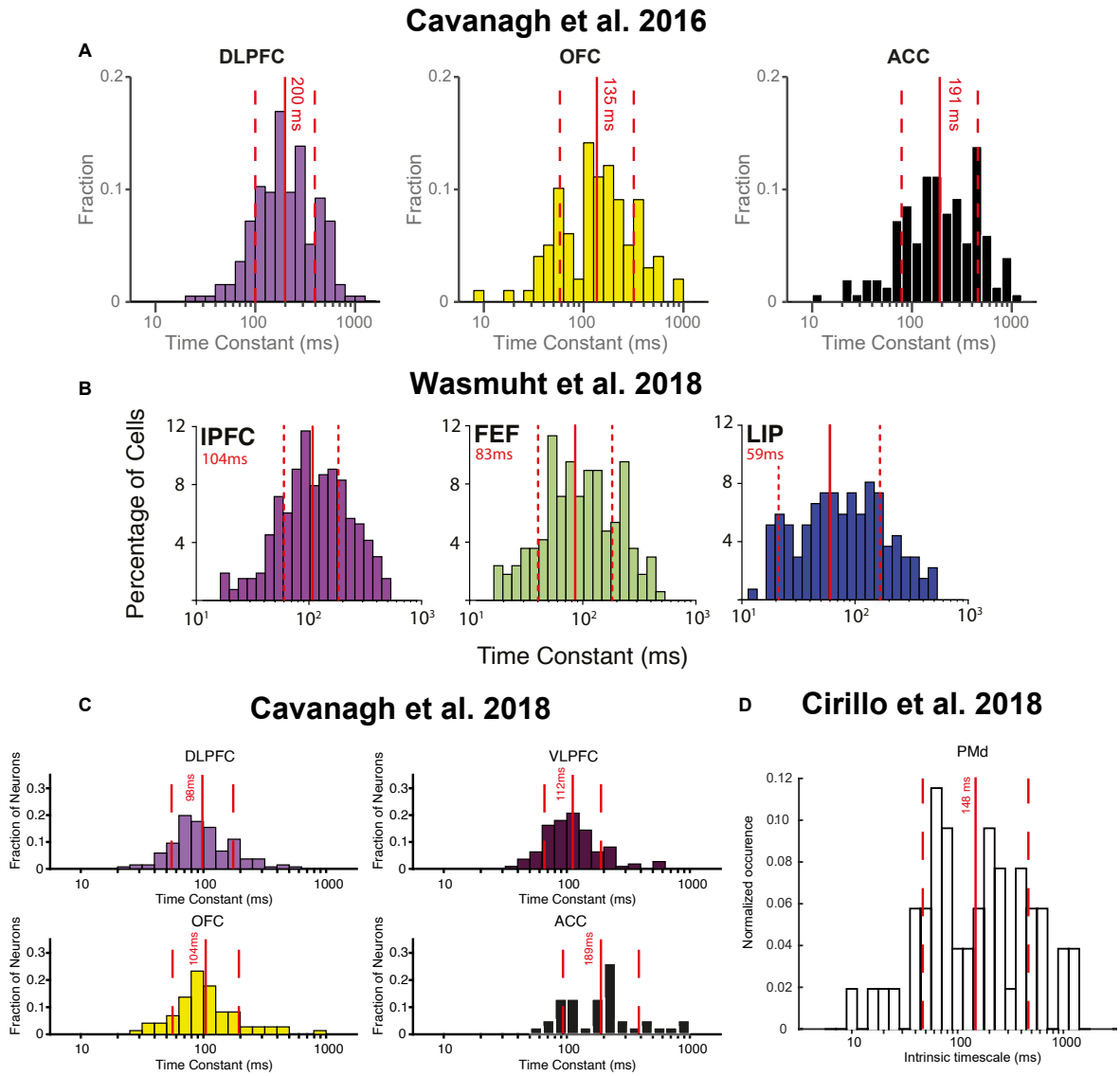


FIGURE 3 | Heterogeneity of single-neuron timescales exist within multiple brain regions across several studies. **(A)** Histograms of single-neuron timescales for dorsolateral prefrontal cortex (DLPFC), orbitofrontal cortex (OFC), and anterior cingulate cortex (ACC). Data were recorded during a value-based decision-making task. Adapted from Cavanagh et al. (2016), where originally published with a CCBY4 licence. **(B)** Histograms of single-neuron timescales for lateral prefrontal cortex (lPFC), frontal eye field (FEF), and lateral intraparietal area (LIP). Data were recorded during a change-detection working memory task. Adapted from Wasmuht et al. (2018), where originally published with a CCBY4 licence. **(C)** Histogram of single-neuron timescales for DLPFC, VLPFC, OFC, and ACC. Data were recorded during an oculomotor delayed working memory task. Adapted from Cavanagh et al. (2018), where originally published with a CCBY4 licence. **(D)** Histogram of single-neuron timescales for dorsal premotor cortex (PMd), recorded during a rule-based working memory task. Adapted from Cirillo et al. (2018), where originally published with a CCBY4 licence.

(Goldman-Rakic, 1995). A number of studies demonstrated that neuronal timescales predicted the strength of mnemonic encoding on a variety of different working-memory paradigms, with longer timescale neurons again playing a greater role in the maintenance of mnemonic information (Nishida et al., 2014; Cavanagh et al., 2018; Cirillo et al., 2018; Wasmuht et al., 2018; Fascianelli et al., 2019; Fontanier et al., 2020; Kim and Sejnowski, 2020; **Figure 4B**). This effect was present in multiple brain regions [lPFC, cingulate cortex, frontal eye field (FEF), premotor cortex], and for multiple different modalities of mnemonic information (spatial location/response direction, expected reward size, stimulus color).

While the results discussed so far had uncovered the relationship between neuronal timescales and the *strength* of encoding, they did not address another important computational property: the *pattern* with which this information was encoded. The temporal dynamics of population encoding has become of increasing interest and controversy in both decision-making

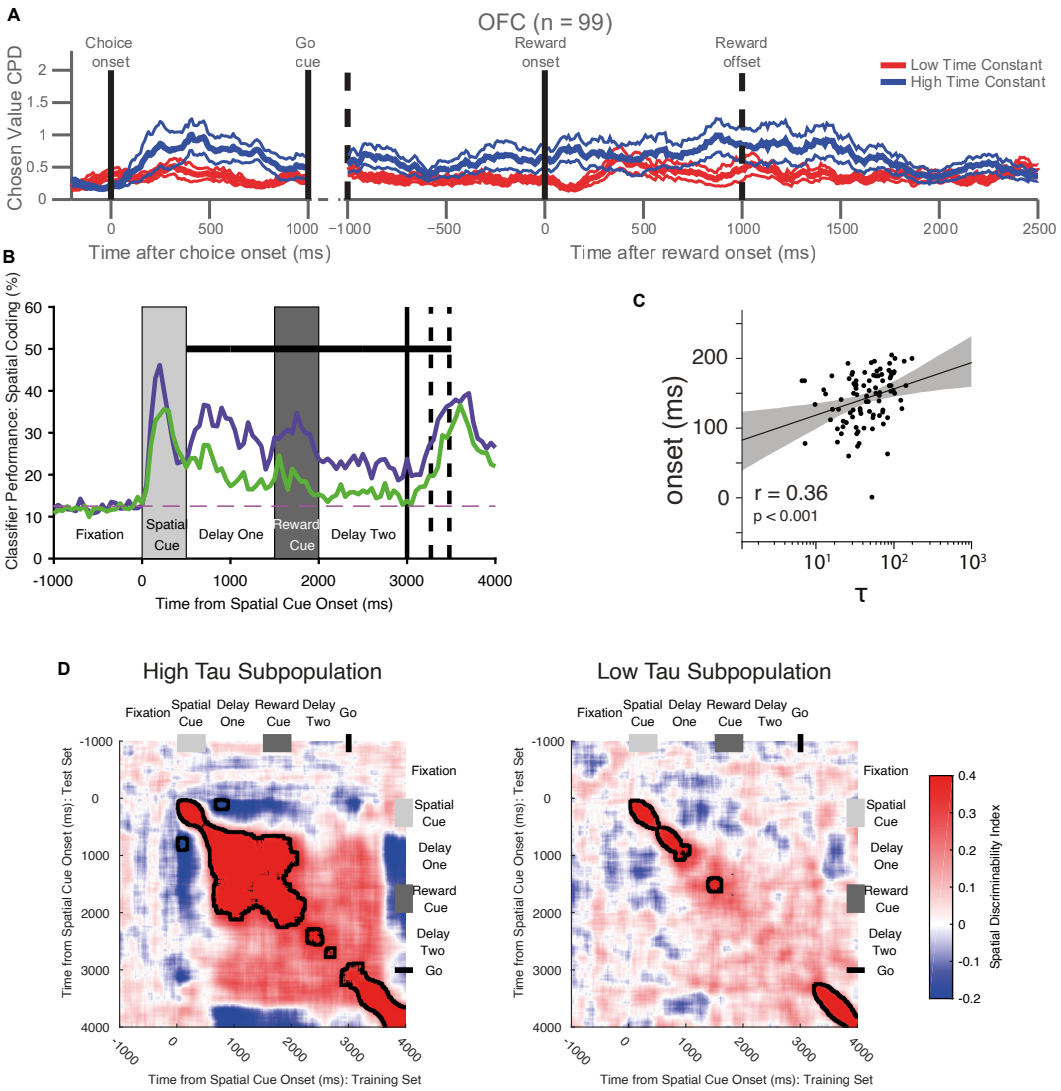


FIGURE 4 | Functional roles of single-neuron timescales during cognitive tasks. **(A)** Long timescale neurons in the orbitofrontal cortex (OFC) are more involved in decision-making and the maintenance of value information until the outcome. The graph shows the coefficient of partial determination (CPD) of chosen value coding for long timescale (blue) and short timescale (red) neurons within OFC. Long timescale neurons have stronger value coding at the time of choice, then throughout the trial until the end of the outcome period. Adapted from Cavanagh et al. (2016) where originally published with a CCBY4 licence. **(B)** Long timescale neurons in the ventrolateral prefrontal cortex (VLPFC) are more involved in the maintenance of spatial working memory information. The graph shows the accuracy with which a linear classifier could decode the remembered spatial location from a subpopulation of neurons with long (purple) and short (timescales). The neural population with longer timescales shows stronger signaling of working memory information—specifically during the delay period. The dashed horizontal line shows chance-level classifier performance. The black horizontal bar shows a significant difference between the two populations. Adapted from Cavanagh et al. (2018) where originally published with a CCBY4 licence. **(C)** Correlation between the onset latency of significant stimulus encoding with the intrinsic timescale in the lateral prefrontal cortex. There is a significant correlation—neurons with shorter timescales encode information more quickly following stimulus onset. Each dot represents one neuron, the black line indicates a linear fit to the data with the shaded area depicting the 95% confidence interval of the fit. Adapted from Wasmuht et al. (2018) where originally published with a CCBY4 licence. **(D)** VLPFC long-timescale neurons have more stable working memory encoding than VLPFC short timescale neurons. The heatmaps show the cross-temporal stability of spatial coding in the two populations. In the long timescale subpopulation, there is greater stability of spatial coding: the off-diagonal elements are warm in color, meaning that the same population code persists throughout the delay epoch following the spatial cue. Although a stable state is reached during delay-one, this is disrupted by the presentation of the distracting reward cue, and there is only a weak non-significant cross-temporal generalization between delay-one and delay-two. In the low time-constant population, coding is always dynamic (i.e., on diagonal heat), so no stable state is established. Adapted from Cavanagh et al. (2018) where originally published with a CCBY4 licence.

and working memory research fields (Latimer et al., 2015; Constantinidis et al., 2018; Lundqvist et al., 2018). Competing explanations propose that the pattern of neural encoding is either stable (Constantinidis et al., 2018), time-varying (Lundqvist et al., 2018) or even activity-silent (Stokes, 2015), during working memory. While these discrepancies may

relate to the task paradigm studied, or the brain region recorded from, it also possibly reflects the inherent neuronal properties—such as intrinsic timescales—of the cells sampled. To explore this, two studies compared the cross-temporal encoding dynamics of short and long timescale neurons (Cavanagh et al., 2018; Wasmuht et al., 2018). When the population of lateral prefrontal neurons were split according to their timescale, the group with longer timescales exhibited stable mnemonic coding whereas those with shorter timescales displayed dynamic coding (**Figure 4D**). These results reveal that in addition to the strength of encoding, intrinsic timescales can also explain computational dynamics, which here has proven useful in reconciling stable and time-varying working memory theories. While in these two studies the target of working memory was an object or spatial array, a similar separation of stable and dynamic subspaces has recently been found for value-coding neurons in OFC and ACC as well, suggesting this may be a general property of PFC coding (Enel et al., 2020).

Surprisingly, although long timescale neurons exhibited a stable pattern of encoding, this was disrupted by the presentation of a salient distractor (Cavanagh et al., 2018; **Figure 4D**). One may have predicted that the maintenance of stable encoding would be important to shield mnemonic information from distraction, and that long timescale neurons would be essential for this process. This result may instead indicate that the function of these long timescale neurons is the *integration* of multiple pieces of task-relevant information, rather than the stable *maintenance* of individual pieces of information. These alternative hypotheses tie directly in to the ideas proposed at the start of this review: is the function of a temporal receptive field to maintain information for a fixed time window, or to integrate all of the information occurring within that window? Unfortunately, all of the cognitive paradigms reviewed so far have been unable to arbitrate between these two hypotheses because the task-relevant stimuli do not vary sufficiently across time. Although the decision-making task discussed earlier (Cavanagh et al., 2016) involves the gradual integration of implicit, noisy value estimates across time (Gold and Shadlen, 2007; Hunt and Hayden, 2017), these internal estimates are not accessible to the experimenter. Future work could utilize a decision-making paradigm with experimenter controlled time-varying evidence (Kira et al., 2015; Cavanagh et al., 2020), which requires the combination of many different stimuli. A paradigm such as this dissociates individual information from the integrated total, and would help to determine whether intrinsic timescales better predict a functional role in information integration or maintenance.

So far, most of the research in this area has focussed on how long timescale neurons may be more functionally important for extended cognitive processes. However, there has been less evidence presented regarding the possible roles of short timescale neurons. This has been addressed by a recent study which demonstrated that during an inter-trial period neurons with short timescales encoded momentary feedback information more strongly (Fontanier et al., 2020). This contrasted with long timescale neurons, which at this point of the task preferentially encoded information which was relevant for future decisions which would occur in subsequent trials. Additionally, there has also been some evidence to suggest that neurons with shorter timescales may encode information at a shorter latency (Wasmuht et al., 2018; **Figure 4C**). It is unknown whether neuronal timescale varies as a function of cortical layer, but as we discuss further below, this result would be consistent with shorter timescale neurons residing in layer IV (and so receiving earlier input), and longer timescale neurons residing in layers II and III (where local recurrent excitation would allow temporally extended computation to occur). Furthermore, it has been suggested that short timescale neurons may utilize a time-varying dynamic representation in order to increase coding dimensionality (Wasmuht et al., 2018)—a computational feature which may be crucial for complex behavior (Rigotti et al., 2013). However, these ideas will have to explored more specifically in future studies (see also section on "Computational Advantages of a Diversity of Within-Region Neuronal Timescales").

In addition to quantifying the rate of exponential decay of spike-count autocorrelation, it is important to consider other features of the autocorrelograms. Single-neuron autocorrelograms also significantly vary in their offset, and the importance of this parameter has yet to be explored. One recent study also identified important heterogeneity in the initial time-lag before autocorrelation begins to decay as a function of time (Fontanier et al., 2020), a feature which was particularly prominent in cingulate cortex (Murray et al., 2014; Cavanagh et al., 2016, 2018; Fontanier et al., 2020). Related to the time-lag of autocorrelograms, other studies in rodents have demonstrated diversity in the time-lag of stimulus representations (Harvey et al., 2012; Morcos and Harvey, 2016; Scott et al., 2017). This pattern of activity could arise from network architectures facilitating the sequential activation of individual neurons, and may be a mechanism through which a dynamic population code could underlie the retention of information in working memory (Goldman, 2009; Rajan et al., 2016).

Aside from analyzing resting spike-count autocorrelation, other researchers have devised different methods to quantify single neuron temporal receptive fields (Bernacchia et al., 2011; Scott et al., 2017; Dragomir et al., 2020; Hart and Huk, 2020; Spitmaan et al., 2020). The majority of these have focussed on the temporal dynamics of task-related encoding—highlighting a heterogeneity for the duration of information maintenance across neurons within the same brain region. An advantage of the autocorrelation approach is that by considering resting activity, it can quantify the intrinsic properties of the neuron, and then determine how these intrinsic properties influence the neuron's role in computations. Hence, this approach can provide broader insights about the underlying cortical architecture (see also later section on "The Anatomical and Biophysical Basis of Single Neuron Timescales"). Furthermore, during the pre-trial fixation period, the subjects are in a controlled, attentive state without eye movements or knowledge of the forthcoming task stimuli. This minimizes the potential confounds of any task-related responses, and facilitates an analysis method that

can be applied and compared across many datasets. However, there are also important advantages to quantifying timescales using patterns of task selectivity—such as having access to a greater amount of data than that limited to the fixation period. This may help to identify neurons with timescales much longer than can be captured with an exponential limited to a 1,000 ms fixation window. A further advantage of this method is being able to relate the extracted timescales more directly to behavior. These timescales may be far longer than those quantified using resting autocorrelation (Bernacchia et al., 2011; Spitmaan et al., 2020), and the two may or may not be directly related (Spitmaan et al., 2020). While quantifying timescales directly using task selectivity has provided interesting results, a more detailed discussion of these is outside the scope of this review.

THE ANATOMICAL AND BIOPHYSICAL BASIS OF SINGLE NEURON TIMESCALES

This section will set out to address what factors contribute to the diversity of single neuron timescales (**Figure 5A**). As a starting point, it will consider factors which have already been suggested to contribute to the diversity of timescales at the level of cortical regions, and try to apply these at a more local level.

Local Connection Patterns

When addressing the biophysical basis of intrinsic timescales at the single neuron level, it is helpful to first consider existing work probing the determinants of timescales at the level of cortical regions (Chaudhuri et al., 2015). Chaudhuri et al. (2015) developed a large-scale dynamical model of macaque neocortex where each brain area is described by a recurrent network (**Figure 1D**). Both local and inter-regional circuit mechanisms contributed to a hierarchy of timescales across cortical areas (**Figure 5B**), which closely resembled the experimental timescales derived from autocorrelation (Murray et al., 2014). A particularly important feature of the model was that regions higher in the cortical hierarchy were endowed with stronger local excitatory connection strength, motivated by the empirical observation that pyramidal neurons in these regions possess a greater number of dendritic spines (Elston, 2000, 2003). However, the experimental evidence suggests there is widespread heterogeneity in spine density within cortical regions (Elston, 2003)—mirroring the variability in single neuron timescales presented in this review. It is therefore important to consider whether local, within region, differences in excitatory connection strength contribute to single neuronal timescales.

By extending the inferences made at the level of cortical regions (Chaudhuri et al., 2015), it is likely that neurons with the longest timescales have the strongest levels of local recurrent connections. One way to examine this hypothesis is to consider noise correlations—the spike count correlation between pairs of simultaneously recorded neurons—as an indirect measure of connection strengths (Cohen and Kohn, 2011). Intriguingly, initial analyses suggest longer timescale cells exhibit higher noise correlations—and hence stronger local connection strengths (Wasmuht et al., 2018; **Figure 5C**). There is also evidence suggesting that the stable population codes generally observed in higher cortical areas are supported by stronger coupling between neurons (Runyan et al., 2017). In addition to the strength of local connectivity, the architecture of those connections may be of relevance. Aside from the temporal domain, it has been shown directly that neurons in mouse primary visual cortex with stronger connectivity share more similar spatial receptive fields (Cossell et al., 2015; Lee et al., 2016). Future studies could examine whether there is a similar association for temporal receptive fields. Interestingly, theoretical work has proposed classes of network architecture that could facilitate a diversity of timescales differentially concentrated in separate parts of the wider network. This can be realised through localized eigenvectors in the network's connectivity matrix (Chaudhuri et al., 2014).

Significant insights into the role of local connectivity in single neuron timescales have also been provided by computational modeling. Computational accounts have stressed the importance of heterogeneous local connection weights for producing a diversity of single neuron timescales (Bernacchia et al., 2011; Chaudhuri et al., 2014). A separate body of theoretical work has also investigated how a closely related temporal feature of neural activity, population sequences where individual neurons have dynamic responses with heterogenous latencies, may arise (Goldman, 2009; Harvey et al., 2012; Rajan et al., 2016). This may be through a highly structured feedforward architecture (Goldman, 2009), or a random network with minimally adjusted connections (Rajan et al., 2016). It is plausible that such architectures could also account for a heterogeneity of timescales. For instance, neurons at different positions in a feedforward network may have a different timescale as well as latency, although this has yet to be explored.

Recent work employing artificial spiking recurrent neural networks (RNN) has also provided further evidence of the importance of local connection patterns in determining neuronal timescales (Kim and Sejnowski, 2020). RNNs trained to perform working memory tasks were shown to contain neurons with a heterogeneity of timescales (**Figure 5D**). As in the electrophysiological data, neurons with longer timescales exhibited stronger and more stable encoding. Interestingly, a heterogeneity of timescales only emerged once RNNs had been trained to perform a temporally extended task (as opposed to a task not requiring the maintenance of information across time; **Figure 5D**), and was most dependent upon local connection strengths. Surprisingly, the connection strengths between pairs of inhibitory neurons were particularly important. Despite the fact that the networks were trained with a biologically implausible gradient descent learning algorithm, it will be important to explore these insights with more biophysically realistic network architectures along with experimental data.

Inter-regional Connection Patterns and Cortical Layer

In addition to local connectivity, the other vital architectural feature facilitating heterogenous timescales across brain areas

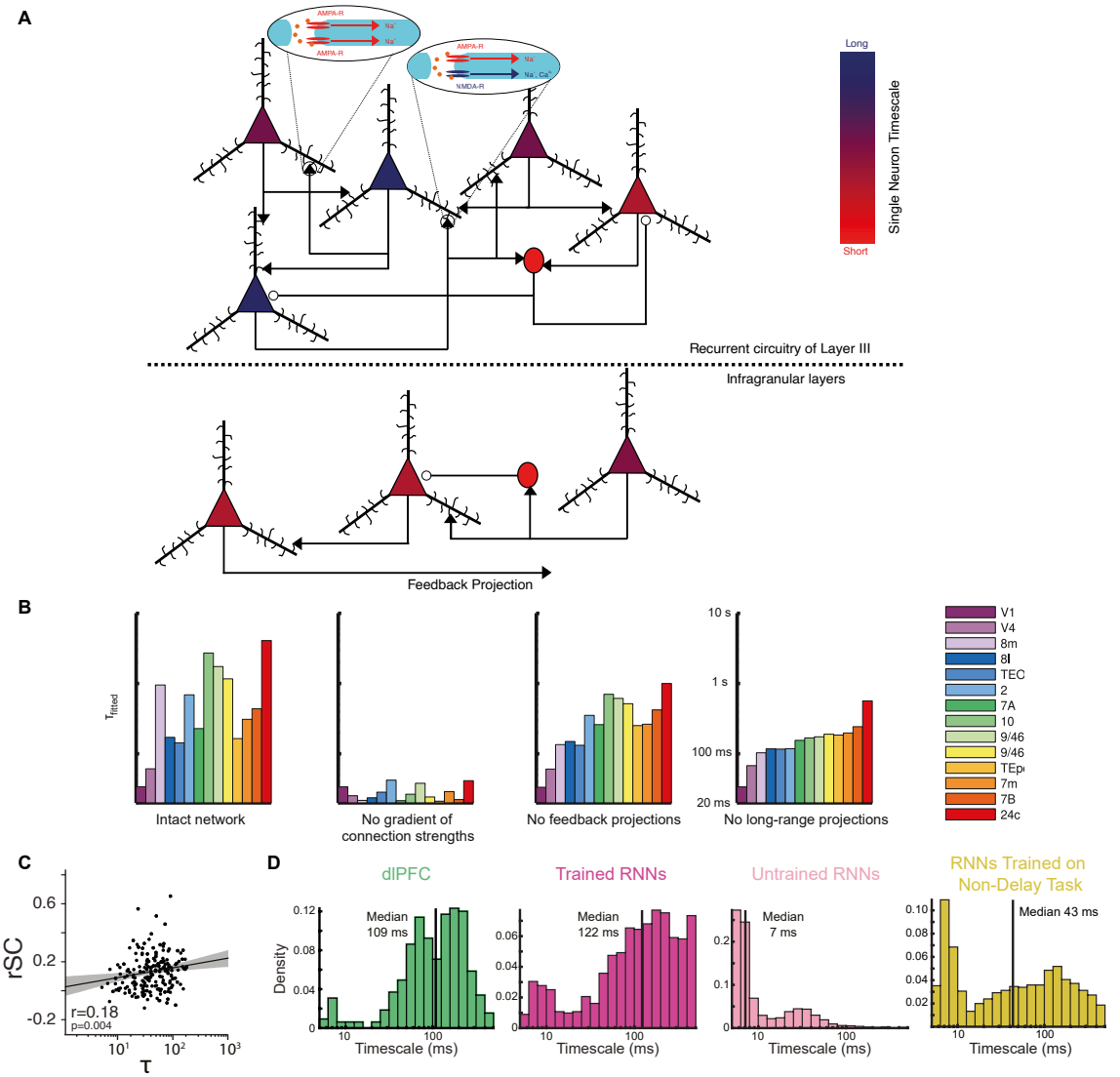


FIGURE 5 | Anatomical and biophysical basis of neuronal timescales. **(A)** Schematic illustration of the circuit properties hypothesized to determine neuronal timescales. The putative timescale of each example neuron is indicated by the color of their cell bodies—with a diverse range spanning from short timescales in bright red to long timescales in dark blue (see also color bar). Excitatory pyramidal neurons have triangular cell bodies and contribute to glutamatergic synapses indicated by black arrowheads. These synapses may contain different relative proportions of AMPA and NMDA receptors [depicted here as a synapse with only AMPA receptors (left oval inset), or a 1:1 AMPA:NMDA receptor ratio at a different synapse (right oval inset) for simplicity]. Inhibitory interneurons have circular cell bodies and contribute to GABAergic synapses indicated by open circles. Above the dashed line is a model circuit of layer III in the primate prefrontal cortex, below the dashed line is a model circuit of an infragranular layer. Several anatomical features are proposed to determine a neuron's timescale. First, we propose that neurons in layer III generally have longer timescales than those in the infragranular layers. Second, we propose that the longest timescales are observed in excitatory neurons with a greater number of recurrent connections and expressing a higher proportion of NMDA receptors relative to AMPA receptors, and with a greater proportion of NR2B subunits. The shortest timescales are found in the inhibitory interneurons, although more realistic models would include a diversity of interneuron subtypes with potentially differing intrinsic time constants. **(B)** Both local and long-range projections are important in determining the timescales of cortical regions in a biophysical network model of the primate cortex. The histograms show the autocorrelation timescales (T_{Fitted}) calculated from the activity in the intact network (far left), then for various perturbations. When the gradient of increasing numbers of intra-region excitatory synapses further up the cortical hierarchy is removed (center left), the diversity of neuronal timescales is lost entirely. When feedback (center right) and all long-range projections (far right) are removed, there is a reduction in timescale diversity from the intact network. Reproduced with permission from Chaudhuri et al. (2015). **(C)** Evidence that local recurrent connection strengths determine single neuronal timescales. The spike count noise correlation with simultaneously recorded neurons (rSC), an estimate of local connectivity, is positively correlated with neuronal timescale (τ). Each dot depicts a cell recorded from either the lateral prefrontal cortex, the FEF, or the lateral intraparietal area of the macaque cortex. The black line and shaded region denote a regression line with a 95% confidence interval. Adapted from Wasmuht et al. (2018) where originally published with a CCBY4 licence. **(D)** Evidence of single-neuron timescale heterogeneity in trained recurrent neural networks (RNNs). Histograms show the single neuronal timescales from an experimental dataset recorded from the DLPFC (Constantinidis et al., 2016), an RNN trained to perform a delayed match to sample task, an untrained RNN, and an RNN trained to perform a similar choice task without a working memory delay. The RNN trained to perform the temporally extended cognitive process exhibits a diversity of single unit timescales most in keeping with the experimental data. Reproduced from Kim and Sejnowski (2020).

in a biophysical circuit model was the pattern of inter-regional connectivity (Chaudhuri et al., 2015; **Figures 1D**, **5B**). The specific constellation of inputs and outputs to a brain region was critically important in determining its timescale. In short, the higher a brain region is located in an anatomical hierarchy, as defined by its inter-areal connections (Markov et al., 2014a), is strongly predictive of its average neuronal time-constants (reviewed in Wang, 2020). When applying this insight to single neurons, we should consider connectivity profiles at the level of cortical regions merely a helpful sketch of an infinitely more detailed neural architecture. Within a given brain region, the incoming and outgoing projections from each neuron are inevitably varied. A minority of neurons may receive direct projections from other regions, or be closely connected with other cells that do, whereas further neurons may be relatively distant from extra-regional input. Therefore, it is possible this heterogeneity in inter-areal projections may be another contributor to determining single neuron timescales.

The inter-region connectivity profile may relate to the cortical layer within which the neuron is situated. For instance, neurons with feedforward connections typically reside in supragranular layers, while those with feedback connections inhabit the infragranular layers (Felleman and Van Essen, 1991; Markov et al., 2014b). Interestingly, the cortical layer may also determine the degree of local connectivity, with neurons in layer III of prefrontal cortex thought to have particularly strong recurrent connections (Goldman-Rakic, 1995; Kritzer and Goldman-Rakic, 1995) reflected by an increase in spine density in prefrontal and parietal cortices relative to early sensory areas (Elston, 2003; Elston et al., 2011; Gilman et al., 2017). Recent studies have leveraged new technologies to demonstrate that task-related working memory activity mainly resides in supragranular layers (Markowitz et al., 2015; Bastos et al., 2018; Finn et al., 2019), providing experimental evidence that recurrent circuitry may be important for generating persistent activity. In future studies, laminar electrode probes may also provide insight into the relationship between neural timescales and cortical layer.

Beyond single neuron electrophysiology studies, recent work has shown that the functional connectivity between brain regions, as determined by resting state fMRI BOLD signal or magnetoencephalography (MEG), is also closely related to the hierarchical heterogeneity in local circuit properties (Demirtaş et al., 2019). A large-scale biophysical model of cortex, with the intrinsic properties such as the levels of excitation and inhibition of individual brain regions varied according to their hierarchical position (Burt et al., 2018), was able to closely mirror human resting state functional connectivity measures (Demirtaş et al., 2019). It was also able to predict a hierarchical topography of spectral features of resting-state MEG. An important advance of this study was that it accounted for heterogenous circuit properties between regions (although not within them). This suggests that at a more local level, the within-region heterogeneity we have discussed in this article (which we posited to be important in determining timescale) may also have an important influence on functional connectivity and oscillatory activity. This provides a link between neuronal timescales and large scale brain networks.

Cell Type and Receptor Expression

The neuron type, for instance whether it is excitatory or inhibitory, likely has an impact on a cell's timescale. In prominent spiking circuit models for extended cognitive processes, such as decision-making and working memory, pyramidal cells and interneurons play different functional roles (Brunel and Wang, 2001; Wang, 2002). Subgroups of pyramidal cells exhibit stimulus-specific persistent activity for particular choice options or memoranda, while interneurons provide non-selective inhibition. If this architecture is indeed present in primate cortex, it is likely excitatory neurons embedded within richly reverberant pools should have longer timescales than interneurons, as well as other non-selective pyramidal neurons. Some recent experimental evidence using neuronal spike width as a proxy for cell type suggests that the ratio of putative pyramidal to inhibitory neurons increases progressively up the cortical hierarchy, possibly facilitating stronger persistent dynamics (Torres-Gomez et al., 2020). Although using this technique could reveal information about the biophysical basis of single neuron timescales, a more reliable investigation of the role of different cell types may require experimental techniques currently only available in rodents. This may also uncover dissociable timescales in different types of GABAergic interneurons which are hypothesized to play distinctive roles in persistent activity (Wang et al., 2004).

Another important determinant of neuronal timescales may be neurotransmitter receptor expression. Slow decaying NMDA receptor (NMDA-R) synaptic currents, which allow post-synaptic neurons to remain depolarized for a greater length of time, are thought to be critical for the stability of neural activity (Wang, 2001). NMDA-R expression is variable across neurons, and given its importance for persistent activity, likely contributes to a neuron's timescale. This could be tested empirically using iontophoresis of NMDA-R antagonists (Wang et al., 2013). The specific subunit combination of the NMDA-R may also be of relevance. NMDA-R are heterotetramers, meaning they are the assembly of four distinct subunits. Each NMDA-R typically consists of two NR1 subunits, together with two NR2 subunits. While the eight possible splice variants of the NR1 subunit are relatively similar, the four varieties of NR2 subunits (NR2A, NR2B, NR2C, NR2D) are more heterogeneous. The NR2 subunit expressed in each receptor is therefore important in determining its kinetic properties (Monyer et al., 1994; Vicini et al., 1998). The NR2 subunits are differentially expressed across different cell types, brain regions, and at different stages of development (Watanabe et al., 1992; Monyer et al., 1994). Interestingly, recent work has shown that the NR2B subunit is increasingly expressed further up the anatomical hierarchy, with the greatest expression in prefrontal cortex (Wang et al., 2008; Burt et al., 2018). Therefore, at a more local level, the degree to which a neuron expresses the NR2B subunit may be important in determining its timescale. In addition to the NMDA-R, other neurotransmitter systems may be relevant, particularly those exerting ascending

neuromodulation such as dopamine (Arnsten et al., 2015) and acetylcholine (Croxson et al., 2011).

In summary, we have discussed a broad range of factors which may contribute to a neuron's timescale—namely cellular morphology, connectivity profile, and receptor expression (**Figure 5A**). When randomly sampling neurons within the macaque prefrontal cortex, the morphology, cell-type, cortical layer and synaptic features are unknown. Recorded neurons are therefore likely sampled from separate subnetworks with differing underlying properties. This may explain the heterogeneity observed in neuronal timescales.

COMPUTATIONAL ADVANTAGES OF A DIVERSITY OF WITHIN-REGION NEURONAL TIMESCALES

The general advantages of processing information across a range of temporal scales at the whole brain level are clear. Short timescales allow one to respond rapidly to important changes in the environment, while long timescales facilitate the integration of information to improve the signal-to-noise of working-memory and decision-making computations. Previous perspectives have addressed the computational advantages of a diversity of processing timescales in detail, and suggested these processes may occur in different brain areas. However, here we will specifically consider why it would be computationally advantageous for *individual brain regions* to also possess their own diversity of timescales.

While the distinction of neuronal timescales at the level of cortical regions has proven important, this has most commonly been framed in the context of processing simple sensory stimuli. In reality, the brain must also process much more complex features of the environment across a range of timescales. The computation of many of these complex features is limited to cortical association areas, as neural computations are constrained by a region's anatomy. These computations often require the integration of many different attributes, but not necessarily across time. To compute the value of a reward—its probability, magnitude, any delay before receiving it, and the acquisition costs must be integrated. Like transient sensory stimuli, values can also evolve sporadically. It is thus important that values are dynamically tracked to facilitate rapid responses to the sudden appearance of a highly rewarding stimulus that is too good to miss, but also integrated gradually across time to improve signal-to-noise ratio and maximize decision-making accuracy. By extension, if the computation of complex features such as value are limited to higher cortical regions, it would be advantageous if neural populations *within* cortical association regions also had a range of diverse timescales for processing value. There is now some experimental evidence showing how this may occur—with neurons in cingulate cortex showing different responses according to their timescale. During an inter-trial period, short timescale neurons signalled the outcome from the immediately preceding trial, whereas long time scale neurons encoded a separate piece of information which was relevant to future decisions on subsequent trials (Fontanier et al., 2020).

Another useful implementation for this diversity of timescales would be in reinforcement learning (Sutton and Barto, 1998). Here, agents compute reward expectation by using a temporal filter to weigh previous outcomes. The optimal timescale for the filter is dependent upon the volatility of the environment; in a stable setting a long temporal filter allows more accurate predictions, whereas in a dynamic setting a short temporal filter should be employed to track changing payoffs (Behrens et al., 2007). Through applying a differential weighting to neurons with different reward timescales in response to changes in environment volatility, efficient reward expectations could be estimated. There is already experimental evidence for heterogenous reward timescales, with neurons integrating to different degrees across previous outcomes (Bernacchia et al., 2011). A similar concept has been explored in a recent neuroimaging study, where ACC was shown to possess a range of learning rates when humans made decisions in a volatile environment (Meder et al., 2017). It would be interesting for future studies to explore how these timescales are utilized. Specifically, whether the outputs of neurons with different timescales are indeed weighted differently by a decoder somewhere within the brain according to the current environmental volatility. This would be in line with similar previous observations of how neural population activity can be flexibly weighted according to current behavioral demands (Raposo et al., 2014), and shed light on how a brain region may utilize its diversity of timescales.

A brain region potentially capable of implementing these ideas is the ACC. ACC neurons not only encode choice and reward history (Seo and Lee, 2007), ACC activity encodes reward information and learning rates over diverse temporal scales (Bernacchia et al., 2011; Meder et al., 2017). Moreover, in the case of both the anatomical connection density patterns (Chaudhuri et al., 2015) and intrinsic neuronal time constants (Murray et al., 2014; Cavanagh et al., 2016), ACC is at the top of the cortical hierarchy, potentially organized in local anatomical gradients (Meder et al., 2017). The simultaneous representation of multiple time constants in ACC may allow the computation of reward trajectories by comparing estimates of recent and past reward rates.

In addition to adaptively weighting neurons according to their timescales, the temporal dimensionality of neural representations is also relevant for decoding. When encoding an item in working-memory, one computational perspective suggests that a stable pattern of neural population activity is preferable—as irrespective of the passage of time, a downstream decoder can utilize the same readout weights for the interpretation of a mnemonic representation (Murray et al., 2017). As we have shown earlier in this piece, neurons with long timescales may be particularly adapted to perform this stable maintenance function (Cavanagh et al., 2016, 2018). A recently emerging, and highly influential, concept in computational neuroscience has been the importance of mixed selectivity in maximizing the dimensionality of neural representations (Rigotti et al., 2013; Fusi et al., 2016; Stringer et al., 2019). In tasks with multiple features, prefrontal neurons generally encode these features with non-linear interactions, and this in turn maximizes

the number of different available linear classifiers which could be utilized for readout. In addition to mixing the activity across neurons, varying the activity across time would further increase the dimensionality (Wasmuht et al., 2018). Therefore, while the stable coding schema offers some advantages, this is at the expense of minimizing the possible dimensionality of encoding relative to a population whose activity varies across time. By possessing subpopulations of neurons with different timescales, the prefrontal cortex is simultaneously providing easily-interpretable readout, as well as a high dimensional one—dependent upon which neurons a downstream decoder chooses to listen to at any given time. This would appear an important advantage of individual brain regions being capable of processing information across different timescales.

In addition to determining the value of stimuli, flexibly applying abstract rules is another important aspect of higher-level cognition (Miller and Cohen, 2001). It requires an agent to modify their response to a stimulus according to dynamically changing contexts or goals. Similar to value computations, experimental evidence suggests the neural substrates for rule based processing reside within higher cortical areas such as prefrontal cortex (Buckley et al., 2009); with neurons encoding abstract rules and rapidly altering how stimulus features are mapped onto actions (Wallis et al., 2001; Buschman et al., 2012; Mante et al., 2013). These rules are often implemented in a hierarchical fashion (Botvinick et al., 2009) which naturally necessitates the organization of behavior at a range of different timescales. Such behaviors often need to be applied rapidly based upon a single salient piece of information, and this would not be possible if prefrontal cortex was only capable of processing information across long timescales as suggested from previous studies (Murray et al., 2014; Hasson et al., 2015).

Another cognitive process which may involve a diversity of neuronal timescales is evidence accumulation. Evidence accumulation refers to the process by which information favoring alternative hypotheses is gradually integrated over time, and has been proposed to underlie perceptual, value-based, and many other forms of decision (Gold and Shadlen, 2007; Krajbich et al., 2010; Shadlen and Kiani, 2013). A series of recent behavioral studies have revealed that the timescale across which evidence is accumulated can be flexibly adjusted according to features of the stimulus or environment (Ossmy et al., 2013; Glaze et al., 2015; Bronfman et al., 2016; Levi et al., 2018; Piet et al., 2018; Ganupuru et al., 2019). For instance, in change detection tasks, humans weigh evidence differently according to how long they expect the intervening "change" in a noisy background stimulus to last (Ossmy et al., 2013). By adopting a shorter accumulation timescale for expected signals with a briefer duration, humans can perform this challenging task effectively. One mechanism by which the decision timescales could be adjusted is through individual brain regions having access to neural representations accumulating evidence across a diversity of timescales- as we have proposed in this review. This solution would provide a flexibility which could solve many more complex problems faced in the real world. For instance, Ossmy et al. (2013) contemplate a real-world example whereby a radar operator must interpret whether signal fluctuations may represent a missile, a passenger plane, or noise. In this problem, the brain must simultaneously accumulate evidence to detect the two important features (missile and plane), which may have different signal patterns/durations. A brain region utilizing a heterogeneity of timescales and applying them to integrate the same visual signal would be well suited to solve this problem. Another example of where this may be useful is situations where different types of decisions, which may use different criteria, must be made based on the same stimulus. A recent study suggested that during a similar change detection task, humans used separate timescales for the initial decision that they had detected a stimulus change, and a second decision to gauge their confidence (Ganupuru et al., 2019). This provides further evidence to suggest that the brain simultaneously has access to multiple neural representations of accumulated evidence across different timescales. These concepts will need to be explored further in future neurophysiological studies probing flexible timescales in evidence accumulation.

Interestingly, work from computational modeling studies suggests that a heterogeneity of timescales is not a default property of neural networks (Kim and Sejnowski, 2020). This heterogeneity only begins to emerge after the network is trained to perform a temporally extended task. Networks trained to perform a simpler response-based task, without any temporal component, had shorter and less heterogenous timescales. This is further evidence that this heterogeneity is present to support the computations discussed in this review: decision-making and working-memory.

Although many of the studies above focus on the processing of task-relevant stimuli, it is also likely that a similarly broad range of timescales of operation may be needed when performing motor control; in particular for temporally extended sequences of complex actions. For example, a recent theoretical account of motor cortex dynamics used a network model with balanced excitation and inhibition to generate "stability-optimized circuits" (SOCs) that could generate complex movements (Hennequin et al., 2014). The authors found that in order to generate such movements, the time constant of membrane and synaptic dynamics in the SOCs (∼200 ms) had to be set to match the dominant timescale in the data they were trying to model (Churchland et al., 2012), giving these connections a slower time-constant than other randomly connected synapses in the same model. They argued that such segregation of fast and slow time-constants may even arise within the same neuron, *via* the respective contribution of proximal and distal synapses in the dendritic arbor. Further evidence supporting a diversity of timescales within a single motor control region comes from functional MRI studies of sequential skilled motor performance in humans (Yokoi and Diedrichsen, 2019). Here, contrary to the common hypothesis of an anatomical division of labor between different levels of a motor control hierarchy, it was instead found that the representation of (short-timescale) movement "chunks" and (long-timescale) movement "sequences" can be spatially overlapping in premotor and parietal areas.

One important consideration is how the resting autocorrelation time constants of individual neurons (generally

ranging from tens to hundreds of milliseconds) can be related with behaviors that occur across timescales which are orders of magnitude longer. Many of the behaviors discussed in this section occur across timescales much greater than the longest individual neuronal timescale measured. This likely reflects that these behaviors are generated by network-level states, to which the contribution of individual neurons is at least somewhat redundant. Furthermore, as the timescales are assigned during a short window of resting activity, their values likely reflect only a fraction of the duration of persistent activity which could potentially be supported. It is also important to remember this same challenge applies to current circuit models which have assigned timescales of similar magnitudes to cortical regions (Chaudhuri et al., 2015).

IMPLICATIONS FOR FUTURE ELECTROPHYSIOLOGY STUDIES

In this review, we have demonstrated that an individual neuron's intrinsic timescale while at rest provides insight into its functional properties and roles during cognitive tasks. This has important implications for how neurophysiological datasets are collected and analyzed. For instance, one commonly employed tactic in neurophysiology recordings in areas such as the lateral intraparietal sulcus has been to preselect which neurons to record from based upon their properties during a memory-guided saccade task (Gnadt and Andersen, 1988), in order to establish that neuron's receptive field. When using this technique, investigators select neurons which exhibit stable, persistent activity before examining their properties during a cognitive task of interest. It is therefore likely that they are predominantly sampling neurons with longer timescales. While this approach has proven fruitful, and it is understandable given the technological challenges of recording from sufficient numbers of neurons, it has likely led to a biased perspective of the overall neural dynamics. For instance, it may have overstated the proportion of neurons exhibiting stable activity during working-memory tasks and gradual ramping activity during perceptual decision-making (Goldman-Rakic, 1995; Gold and Shadlen, 2007). This is important because it entirely overlooks the roles of neurons with non-classical response profiles. It also has arguably led to an over-emphasis of the capabilities of individual neurons, supported by idealised examples, and the disregarding of more sophisticated population-level solutions to computational problems (Rigotti et al., 2013). A more complete understanding of neural computations requires us to understand the roles of all of the neurons in these cognitive processes, and the recording of as representative a sample as possible in order to appreciate how neurons function together as a population. Fortunately, new technologies are becoming increasingly available that will allow investigators to record from many neurons simultaneously across each cell layer (Sofroniew et al., 2016; Jun et al., 2017). This should hopefully facilitate a more unbiased characterization of the heterogeneity of neuronal responses. If researchers are particularly interested in a certain subpopulation of neurons with stable activity, they will still be able to find these neurons *post hoc* using the timescale method discussed in this review. However, they will also have access to a plethora of extra information about what neurons with other timescales are doing, and how the population as a whole behaves.

Although this review has primarily focussed on macaque neurophysiology studies, future work may also seek to apply the timescale analyses to high-density electrophysiological recordings collected from rodents (Siegle et al., 2019); where cognitive processes are being studied with an increasingly sophisticated repertoire of techniques. These include the precise perturbation of neural circuits, recording from genetically identifiable neurons and the implementation of neuropsychiatric disease models. These experiments would provide some more concrete insights into the determinants of single neuron timescales.

One important limitation of the majority of neurophysiological datasets considered in this review is that they study a behavior which requires the prolonged maintenance/integration of information. It therefore makes sense that a prominent role for long timescale neurons in these computations was established. However, to fully explore the functional importance of a diversity of timescales, tasks which require tracking information (and modulating behavior) over both short and long timescales should be explored (Behrens et al., 2007; Daw et al., 2011; Massi et al., 2018). Future work should try to establish if there is an important role for neurons with shorter timescales in such tasks. Another important consideration is to study a task which simultaneously requires the dynamic tracking of complex information, as well as its gradual integration. As suggested earlier in this piece, one possibility would be to record neurophysiological activity on an evidence integration task, where subjects must combine many samples with unique characteristics. This would make clearly dissociable predictions for the neural representations to expect in shorter timescale (momentary evidence) and longer timescale (integrated evidence) neurons.

Furthermore, while the partition of neurons into short and long timescale provides intuition and is necessary when analyzing the patterns of coding at the population level, it is a relatively coarse simplification of the underlying concept of a heterogeneity of single-neuron timescales. Ideally, a task design would demonstrate the utility of a diverse continuum of timescales. For instance, subjects could be trained to temporally filter previous rewards across a different number of trials according to a cue presented each trial. This should require the processing of reward across a range of timescales, and the trial-wise adaptive weighting of each timescale population dependent on the behavioral cue.

CONCLUSION

In summary, we have reviewed important electrophysiological evidence from a series of recent studies that convincingly demonstrate the heterogeneity of timescales at the level of single neurons within a cortical region. This heterogeneity is functionally relevant for the computations that neurons perform during decision-making and working memory. A neuron's

timescale is likely determined by the neurotransmitter it releases, its local connectivity pattern, receptor expression, and cortical layer. It is important for individual brain regions to have neurons with a heterogeneity of timescales, as many high-level cognitive processes such as learning, planning, and rule-based behavior require making adaptive decisions to changing environmental demands. These computations generally occur in higher cortical regions which have a long timescale when considered as a whole-brain region, but individual neurons in these areas display a diversity of timescales. A heterogeneity of timescales also offers a compromise between robust stable representations that are easy to read out and those which are most efficient and high dimensional. Future experimental work further demonstrating some of the advantages of short timescale neurons in higher cortical areas, and how a population may effectively utilize a whole distribution of timescales, will further strengthen our arguments about their computational role. The method we have outlined has already provided important computational insights and will prove an increasingly valuable tool as researchers start to record from more neurons simultaneously.

AUTHOR CONTRIBUTIONS

SC: conceptualization, writing—original draft, writing—review and editing. LH and SK: conceptualization, supervision, funding acquisition, writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

SC was supported by the Middlesex Hospital Medical School General Charitable Trust. LH was supported by a Henry Wellcome Fellowship (098830/Z/12/Z) and Henry Dale Fellowship (208789/Z/17/Z) from the Wellcome Trust, a National Alliance for Research on Schizophrenia and Depression Young Investigator Grant from the Brain and Behavior Research Foundation, and the National Institute for Health Research Oxford Health Biomedical Research Centre. SK was supported by Wellcome Trust Investigator Awards (096689/Z/11/Z and 220296/Z/20/Z). The publication fees were funded by UCL's Wellcome Trust open access grant.

REFERENCES

Arnsten, A. F., Wang, M., and Paspalas, C. D. (2015). Dopamine's actions in primate prefrontal cortex: challenges for treating cognitive disorders. *Pharmacol. Rev.* 67, 681–696. doi: 10.1124/pr.115.010512

Azouz, R., and Gray, C. M. (1999). Cellular mechanisms contributing to response variability of cortical neurons *in vivo*. *J. Neurosci.* 19, 2209–2223. doi: 10.1523/JNEUROSCI.19-06-02209.1999

Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cogn. Sci.* 12, 193–200. doi: 10.1016/j.tics.2008.02.004

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., and Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron* 95, 709.e5–721.e5. doi: 10.1016/j.neuron.2017.06.041

Baldassano, C., Hasson, U., and Norman, K. A. (2018). Representation of real-world event schemas during narrative perception. *J. Neurosci.* 38, 9689–9699. doi: 10.1523/JNEUROSCI.0251-18.2018

Barbas, H., and Rempel-Clower, N. (1997). Cortical structure predicts the pattern of corticocortical connections. *Cereb. Cortex* 7, 635–646. doi: 10.1093/cercor/7.7.635

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

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J. M., Oostenveld, R., Dowdall, J. R., et al. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* 85, 390–401. doi: 10.1016/j.neuron.2014.12.018

Behrens, T. E., Woolrich, M. W., Walton, M. E., and Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.* 10, 1214–1221. doi: 10.1038/nn1954

Benucci, A., Ringach, D. L., and Carandini, M. (2009). Coding of stimulus sequences by population responses in visual cortex. *Nat. Neurosci.* 12, 1317–1324. doi: 10.1038/nn.2398

Bernacchia, A., Seo, H., Lee, D., and Wang, X. J. (2011). A reservoir of time constants for memory traces in cortical neurons. *Nat. Neurosci.* 14, 366–372. doi: 10.1038/nn.2752

Blasdel, G. G., and Fitzpatrick, D. (1984). Physiological organization of layer 4 in macaque striate cortex. *J. Neurosci.* 4, 880–895. doi: 10.1523/JNEUROSCI.04-03-00880.1984

Botvinick, M. M., Niv, Y., and Barto, A. G. (2009). Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition* 113, 262–280. doi: 10.1016/j.cognition.2008.08.011

Brincat, S. L., and Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat. Neurosci.* 7, 880–886. doi: 10.1038/nn1278

Bronfman, Z. Z., Brezis, N., and Usher, M. (2016). Non-monotonic temporal-weighting indicates a dynamically modulated evidence-integration mechanism. *PLoS Comput. Biol.* 12:e1004667. doi: 10.1371/journal.pcbi.1004667

Brunel, N., and Wang, X. J. (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J. Comput. Neurosci.* 11, 63–85. doi: 10.1023/a:1011204814320

Buckley, M. J., Mansouri, F. A., Hoda, H., Mahboubi, M., Browning, P. G., Kwok, S. C., et al. (2009). Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. *Science* 325, 52–58. doi: 10.1126/science.1172377

Buckner, R. L., Krienen, F. M., and Yeo, B. T. (2013). Opportunities and limitations of intrinsic functional connectivity MRI. *Nat. Neurosci.* 16, 832–837. doi: 10.1038/nn.3423

Burt, J. B., Demirtaş, M., Eckner, W. J., Navejar, N. M., Ji, J. L., Martin, W. J., et al. (2018). Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* 21, 1251–1259. doi: 10.1038/s41593-018-0195-0

Buschman, T. J., Denovellis, E. L., Diogo, C., Bullock, D., and Miller, E. K. (2012). Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. *Neuron* 76, 838–846. doi: 10.1016/j.neuron.2012.09.029

Cavanagh, S. E., Lam, N. H., Murray, J. D., Hunt, L. T., and Kennerley, S. W. (2020). A circuit mechanism for decision-making biases and NMDA receptor hypofunction. *eLife* 9:e53664. doi: 10.7554/eLife.53664

Cavanagh, S. E., Towers, J. P., Wallis, J. D., Hunt, L. T., and Kennerley, S. W. (2018). Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex. *Nat. Commun.* 9:3498. doi: 10.1038/s41467-018-05873-3

Cavanagh, S. E., Wallis, J. D., Kennerley, S. W., and Hunt, L. T. (2016). Autocorrelation structure at rest predicts value correlates of single neurons during reward-guided choice. *eLife* 5:e18937. doi: 10.7554/eLife.18937

Chang, L., and Tsao, D. Y. (2017). The code for facial identity in the primate brain. *Cell* 169, 1013.e14–1028.e14. doi: 10.1016/j.cell.2017.05.011

Chaudhuri, R., Bernacchia, A., and Wang, X. J. (2014). A diversity of localized timescales in network activity. *eLife* 3:e01239. doi: 10.7554/eLife.01239

Chaudhuri, R., Knoblauch, K., Gariel, M. A., Kennedy, H., and Wang, X. J. (2015). A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron* 88, 419–431. doi: 10.1016/j.neuron.2015.09.008

Chen, J., Hasson, U., and Honey, C. J. (2015). Processing timescales as an organizing principle for primate cortex. *Neuron* 88, 244–246. doi: 10.1016/j.neuron.2015.10.010

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., et al. (2012). Neural population dynamics during reaching. *Nature* 487, 51–56. doi: 10.1038/nature11129

Cirillo, R., Fascianelli, V., Ferrucci, L., and Genovesio, A. (2018). Neural intrinsic timescales in the macaque dorsal premotor cortex predict the strength of spatial response coding. *iScience* 10, 203–210. doi: 10.1016/j.isci.2018.11.033

Cohen, M. R., and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nat. Neurosci.* 14, 811–819. doi: 10.1038/nn.2842

Constantinidis, C., Funahashi, S., Lee, D., Murray, J. D., Qi, X. L., Wang, M., et al. (2018). Persistent spiking activity underlies working memory. *J. Neurosci.* 38, 7020–7028. doi: 10.1523/JNEUROSCI.2486-17.2018

Constantinidis, C., Qi, X.-L., and Meyer, T. (2016). *Single-Neuron Spike Train Recordings from Macaque Prefrontal Cortex During a Visual Working Memory Task Before and After Training.* doi: 10.6080/K0ZW1HVD Available online at: CRCNS.org. doi: 10.6080/K0ZW1HVD

Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., et al. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. *Nature* 518, 399–403. doi: 10.1038/nature14182

Croxson, P. L., Kyriazis, D. A., and Baxter, M. G. (2011). Cholinergic modulation of a specific memory function of prefrontal cortex. *Nat. Neurosci.* 14, 1510–1512. doi: 10.1038/nn.2971

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69, 1204–1215. doi: 10.1016/j.neuron.2011.02.027

Demirtaş, M., Burt, J. B., Helmer, M., Ji, J. L., Adkinson, B. D., Glasser, M. F., et al. (2019). Hierarchical heterogeneity across human cortex shapes large-scale neural dynamics. *Neuron* 101, 1181.e13–1194.e13. doi: 10.1016/j.neuron.2019.01.017

Dragomir, E. I., Štih, V., and Portugues, R. (2020). Evidence accumulation during a sensorimotor decision task revealed by whole-brain imaging. *Nat. Neurosci.* 23, 85–93. doi: 10.1038/s41593-019-0535-8

Elston, G. N. (2000). Pyramidal cells of the frontal lobe: all the more spinous to think with. *J. Neurosci.* 20:RC95. doi: 10.1523/JNEUROSCI.20-18-j0002.2000

Elston, G. N. (2003). Cortex, cognition and the cell: new insights into the pyramidal neuron and prefrontal function. *Cereb. Cortex* 13, 1124–1138. doi: 10.1093/cercor/bhg093

Elston, G. N., Benavides-Piccione, R., Elston, A., Manger, P. R., and Defelipe, J. (2011). Pyramidal cells in prefrontal cortex of primates: marked differences in neuronal structure among species. *Front. Neuroanat.* 5:2. doi: 10.3389/fnana.2011.00002

Enel, P., Wallis, J. D., and Rich, E. L. (2020). Stable and dynamic representations of value in the prefrontal cortex. *eLife* 9:e54313. doi: 10.7554/eLife.54313

Fascianelli, V., Tsujimoto, S., Marcos, E., and Genovesio, A. (2019). Autocorrelation structure in the macaque dorsolateral, but not orbital or polar, prefrontal cortex predicts response-coding strength in a visually cued strategy task. *Cereb. Cortex* 29, 230–241. doi: 10.1093/cercor/bhx321

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1

Finn, E. S., Huber, L., Jangraw, D. C., Molfese, P. J., and Bandettini, P. A. (2019). Layer-dependent activity in human prefrontal cortex during working memory. *Nat. Neurosci.* 22, 1687–1695. doi: 10.1038/s41593-019-0487-z

Fontanier, V., Sarazin, M., Stoll, F. M., Delord, B., and Procyk, E. (2020). Local inhibitory control of frontal network metastability underlies the temporal signature of cognitive states. *bioRxiv* [Preprint]. doi: 10.1101/2020.08.20.259192

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

Fusi, S., Miller, E. K., and Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* 37, 66–74. doi: 10.1016/j.conb.2016.01.010

Ganupuru, P., Goldring, A. B., Harun, R., and Hanks, T. D. (2019). Flexibility of timescales of evidence evaluation for decision making. *Curr. Biol.* 29, 2091.e4–2097.e4. doi: 10.1016/j.cub.2019.05.037

Gilman, J. P., Medalla, M., and Luebke, J. I. (2017). Area-specific features of pyramidal neurons-a comparative study in mouse and rhesus monkey. *Cereb. Cortex* 27, 2078–2094. doi: 10.1093/cercor/bhw062

Glaze, C. M., Kable, J. W., and Gold, J. I. (2015). Normative evidence accumulation in unpredictable environments. *eLife* 4:e08825. doi: 10.7554/eLife.08825

Gnadt, J. W., and Andersen, R. A. (1988). Memory related motor planning activity in posterior parietal cortex of macaque. *Exp. Brain Res.* 70, 216–220. doi: 10.1007/BF00271862

Gold, J. I., and Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574. doi: 10.1146/annurev.neuro.29.051605.113038

Goldman, M. S. (2009). Memory without feedback in a neural network. *Neuron* 61, 621–634. doi: 10.1016/j.neuron.2008.12.012.

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

Gur, M., Kagan, I., and Snodderly, D. M. (2005). Orientation and direction selectivity of neurons in V1 of alert monkeys: functional relationships and laminar distributions. *Cereb. Cortex* 15, 1207–1221. doi: 10.1093/cercor/bhi003

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

Harvey, C. D., Coen, P., and Tank, D. W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* 484, 62–68. doi: 10.1038/nature10918

Hasson, U., Chen, J., and Honey, C. J. (2015). Hierarchical process memory: memory as an integral component of information processing. *Trends Cogn. Sci.* 19, 304–313. doi: 10.1016/j.tics.2015.04.006

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* 28, 2539–2550. doi: 10.1523/JNEUROSCI.5487-07.2008

Hennequin, G., Vogels, T. P., and Gerstner, W. (2014). Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron* 82, 1394–1406. doi: 10.1016/j.neuron.2014.04.045

Honey, C. J., Thesen, T., Donner, T. H., Silbert, L. J., Carlson, C. E., Devinsky, O., et al. (2012). Slow cortical dynamics and the accumulation of information over long timescales. *Neuron* 76, 423–434. doi: 10.1016/j.neuron.2012.08.011

Hosokawa, T., Kennerley, S. W., Sloan, J., and Wallis, J. D. (2013). Single-neuron mechanisms underlying cost-benefit analysis in frontal cortex. *J. Neurosci.* 33, 17385–17397. doi: 10.1523/JNEUROSCI.2221-13.2013

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154. doi: 10.1113/jphysiol.1962.sp006837

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243. doi: 10.1113/jphysiol.1968.sp008455

Hunt, L. T., Behrens, T. E., Hosokawa, T., Wallis, J. D., and Kennerley, S. W. (2015). Capturing the temporal evolution of choice across prefrontal cortex. *eLife* 4:e11945. doi: 10.7554/eLife.11945

Hunt, L. T., and Hayden, B. Y. (2017). A distributed, hierarchical and recurrent framework for reward-based choice. *Nat. Rev. Neurosci.* 18, 172–182. doi: 10.1038/nrn.2017.7

Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F., and Behrens, T. E. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nat. Neurosci.* 15, 470–476, S471–S473. doi: 10.1038/nn.3017

Jocham, G., Brodersen, K. H., Constantinescu, A. O., Kahn, M. C., Ianni, A. M., Walton, M. E., et al. (2016). Reward-guided learning with and without causal attribution. *Neuron* 90, 177–190. doi: 10.1016/j.neuron.2016.02.018

Jun, J. K., Miller, P., Hernández, A., Zainos, A., Lemus, L., Brody, C. D., et al. (2010). Heterogenous population coding of a short-term memory and decision task. *J. Neurosci.* 30, 916–929. doi: 10.1523/JNEUROSCI.2062-09.2010

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551, 232–236. doi: 10.1038/nature24636

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4:e1000209. doi: 10.1371/journal.pcbi.1000209

Kim, R., and Sejnowski, T. J. (2020). Strong inhibitory signaling underlies stable temporal dynamics and working memory in spiking neural networks. *bioRxiv* [Preprint]. doi: 10.1101/2020.02.11.944751

Kira, S., Yang, T., and Shadlen, M. N. (2015). A neural implementation of Wald's sequential probability ratio test. *Neuron* 85, 861–873. doi: 10.1016/j.neuron.2015.01.007

Koechlin, E., Ody, C., and Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science* 302, 1181–1185. doi: 10.1126/science.1088545

Krajbich, I., Armel, C., and Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nat. Neurosci.* 13, 1292–1298. doi: 10.1038/nn.2635

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

Latimer, K. W., Yates, J. L., Meister, M. L., Huk, A. C., and Pillow, J. W. (2015). NEURONAL MODELING. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science* 349, 184–187. doi: 10.1126/science.aaa4056

Lee, W. C., Bonin, V., Reed, M., Graham, B. J., Hood, G., Glattfelder, K., et al. (2016). Anatomy and function of an excitatory network in the visual cortex. *Nature* 532, 370–374. doi: 10.1038/nature17192

Lennie, P. (1998). Single units and visual cortical organization. *Perception* 27, 889–935. doi: 10.1068/p270889

Lerner, Y., Honey, C. J., Silbert, L. J., and Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci.* 31, 2906–2915. doi: 10.1523/JNEUROSCI.3684-10.2011

Levi, A. J., Yates, J. L., Huk, A. C., and Katz, L. N. (2018). Strategic and dynamic temporal weighting for perceptual decisions in humans and macaques. *eNeuro* 5:ENEURO.0169-18.2018. doi: 10.1523/ENEURO.0169-18.2018

Lundqvist, M., Herman, P., and Miller, E. K. (2018). Working memory: delay activity, yes! persistent activity? Maybe not. *J. Neurosci.* 38, 7013–7019. doi: 10.1523/JNEUROSCI.2485-17.2018

Luppino, G., Matelli, M., Camarda, R. M., Gallese, V., and Rizzolatti, G. (1991). Multiple representations of body movements in mesial area 6 and the adjacent cingulate cortex: an intracortical microstimulation study in the macaque monkey. *J. Comp. Neurol.* 311, 463–482. doi: 10.1002/cne.903110403

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

Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., et al. (2014a). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cereb. Cortex* 24, 17–36. doi: 10.1093/cercor/bhs270

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., et al. (2014b). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *J. Comp. Neurol.* 522, 225–259. doi: 10.1002/cne.23458

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

Massi, B., Donahue, C. H., and Lee, D. (2018). Volatility facilitates value updating in the prefrontal cortex. *Neuron* 99, 598.e4–608.e4. doi: 10.1016/j.neuron.2018.06.033

Meder, D., Kolling, N., Verhagen, L., Wittmann, M. K., Scholl, J., Madsen, K. H., et al. (2017). Simultaneous representation of a spectrum of dynamically changing value estimates during decision making. *Nat. Commun.* 8:1942. doi: 10.1038/s41467-017-02169-w

Meister, M. L., Hennig, J. A., and Huk, A. C. (2013). Signal multiplexing and single-neuron computations in lateral intraparietal area during decision-making. *J. Neurosci.* 33, 2254–2267. doi: 10.1523/JNEUROSCI.2984-12.2013

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167

Monyer, H., Burnashev, N., Laurie, D. J., Sakmann, B., and Seeburg, P. H. (1994). Developmental and regional expression in the rat brain and functional properties of four NMDA receptors. *Neuron* 12, 529–540. doi: 10.1016/0896-6273(94)90210-0

Morcos, A. S., and Harvey, C. D. (2016). History-dependent variability in population dynamics during evidence accumulation in cortex. *Nat. Neurosci.* 19, 1672–1681. doi: 10.1038/nn.4403

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., et al. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* 17, 1661–1663. doi: 10.1038/nn.3862

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

Nachev, P., Kennard, C., and Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nat. Rev. Neurosci.* 9, 856–869. doi: 10.1038/nrn2478

Nauhaus, I., Nielsen, K. J., and Callaway, E. M. (2016). Efficient receptive field tiling in primate V1. *Neuron* 91, 893–904. doi: 10.1016/j.neuron.2016.07.015

Nee, D. E., and D'Esposito, M. (2016). The hierarchical organization of the lateral prefrontal cortex. *eLife* 5:e12112. doi: 10.7554/eLife.12112

Nishida, S., Tanaka, T., Shibata, T., Ikeda, K., Aso, T., and Ogawa, T. (2014). Discharge-rate persistence of baseline activity during fixation reflects maintenance of memory-period activity in the macaque posterior parietal cortex. *Cereb. Cortex* 24, 1671–1685. doi: 10.1093/cercor/bht031

Ogawa, T., and Komatsu, H. (2010). Differential temporal storage capacity in the baseline activity of neurons in macaque frontal eye field and area V4. *J. Neurophysiol.* 103, 2433–2445. doi: 10.1152/jn.01066.2009

Ossmy, O., Moran, R., Pfeffer, T., Tsetsos, K., Usher, M., and Donner, T. H. (2013). The timescale of perceptual evidence integration can be adapted to the environment. *Curr. Biol.* 23, 981–986. doi: 10.1016/j.cub.2013.04.039

Picard, N., and Strick, P. L. (1996). Motor areas of the medial wall: a review of their location and functional activation. *Cereb. Cortex* 6, 342–353. doi: 10.1093/cercor/6.3.342

Piet, A. T., El Hady, A., and Brody, C. D. (2018). Rats adopt the optimal timescale for evidence integration in a dynamic environment. *Nat. Commun.* 9:4265. doi: 10.1038/s41467-018-06561-y

Rajan, K., Harvey, C. D., and Tank, D. W. (2016). Recurrent network models of sequence generation and memory. *Neuron* 90, 128–142. doi: 10.1016/j.neuron.2016.02.009

Rangel, A., and Hare, T. (2010). Neural computations associated with goal-directed choice. *Curr. Opin. Neurobiol.* 20, 262–270. doi: 10.1016/j.conb.2010.03.001

Raposo, D., Kaufman, M. T., and Churchland, A. K. (2014). A category-free neural population supports evolving demands during decision-making. *Nat. Neurosci.* 17, 1784–1792. doi: 10.1038/nn.3865

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

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

Runyan, C. A., Piasini, E., Panzeri, S., and Harvey, C. D. (2017). Distinct timescales of population coding across cortex. *Nature* 548, 92–96. doi: 10.1038/nature23020

Russo, A. A., Khajeh, R., Bittner, S. R., Perkins, S. M., Cunningham, J. P., Abbott, L. F., et al. (2020). Neural trajectories in the supplementary motor

area and motor cortex exhibit distinct geometries, compatible with different classes of computation. *Neuron* 107, 745.e6–758.e6. doi: 10.1016/j.neuron.2020.05.020

Scott, B. B., Constantinople, C. M., Akrami, A., Hanks, T. D., Brody, C. D., and Tank, D. W. (2017). Fronto-parietal cortical circuits encode accumulated evidence with a diversity of timescales. *Neuron* 95, 385.e5–398.e5. doi: 10.1016/j.neuron.2017.06.013

Sen, K., Theunissen, F. E., and Doupe, A. J. (2001). Feature analysis of natural sounds in the songbird auditory forebrain. *J. Neurophysiol.* 86, 1445–1458. doi: 10.1152/jn.2001.86.3.1445

Seo, H., and Lee, D. (2007). Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* 27, 8366–8377. doi: 10.1523/JNEUROSCI.2369-07.2007

Shadlen, M. N., and Kiani, R. (2013). Decision making as a window on cognition. *Neuron* 80, 791–806. doi: 10.1016/j.neuron.2013.10.047

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

Shima, K., and Tanji, J. (2000). Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. *J. Neurophysiol.* 84, 2148–2160. doi: 10.1152/jn.2000.84.4.2148

Siegle, J. H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., et al. (2019). A survey of spiking activity reveals a functional hierarchy of mouse corticothalamic visual areas. *bioRxiv* [Preprint]. doi: 10.1101/805010

Sofroniew, N. J., Flickinger, D., King, J., and Svoboda, K. (2016). A large field of view two-photon mesoscope with subcellular resolution for *in vivo* imaging. *eLife* 5:e14472. doi: 10.7554/eLife.14472

Spitmaan, M., Seo, H., Lee, D., and Soltani, A. (2020). "Multiple timescales of neural dynamics and integration of task-relevant signals across cortex," in *Proceedings of the National Academy of Sciences*, ed M. E. Goldberg (New York, NY: Columbia University), 22522–22531. doi: 10.1073/pnas.2005993117

Stephens, G. J., Honey, C. J., and Hasson, U. (2013). A place for time: the spatiotemporal structure of neural dynamics during natural audition. *J. Neurophysiol.* 110, 2019–2026. doi: 10.1152/jn.00268.2013

Stokes, M. G. (2015). "Activity-silent" working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* 19, 394–405. doi: 10.1016/j.tics.2015.05.004

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature* 571, 361–365. doi: 10.1038/s41586-019-1346-5

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139. doi: 10.1146/annurev.ne.19.030196.000545

Torres-Gomez, S., Blonde, J. D., Mendoza-Halliday, D., Kuebler, E., Everest, M., Wang, X. J., 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, 4544–4562. doi: 10.1093/cercor/bhaa056

Vicini, S., Wang, J. F., Li, J. H., Zhu, W. J., Wang, Y. H., Luo, J. H., et al. (1998). Functional and pharmacological differences between recombinant *N*-methyl-D-aspartate receptors. *J. Neurophysiol.* 79, 555–566. doi: 10.1152/jn.1998.79.2.555

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

Wallis, J. D., and Kennerley, S. W. (2010). Heterogeneous reward signals in prefrontal cortex. *Curr. Opin. Neurobiol.* 20, 191–198. doi: 10.1016/j.conb.2010.02.009

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

Wang, X.-J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 36, 955–968. doi: 10.1016/s0896-6273(02)01092-9

Wang, X.-J. (2020). Macroscopic gradients of synaptic excitation and inhibition in the neocortex. *Nat. Rev. Neurosci.* 21, 169–178. doi: 10.1038/s41583-020-0262-x

Wang, X.-J., and Kennedy, H. (2016). Brain structure and dynamics across scales: in search of rules. *Curr. Opin. Neurobiol.* 37, 92–98. doi: 10.1016/j.conb.2015.12.010

Wang, J., Narain, D., Hosseini, E. A., and Jazayeri, M. (2018). Flexible timing by temporal scaling of cortical responses. *Nat. Neurosci.* 21, 102–110. doi: 10.1038/s41593-017-0028-6

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

Wang, X.-J., Tegnér, J., Constantinidis, C., and Goldman-Rakic, P. S. (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, 1368–1373. doi: 10.1073/pnas.0305337101

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

Wasmuht, D. F., Spaak, E., Buschman, T. J., Miller, E. K., and Stokes, M. G. (2018). Intrinsic neuronal dynamics predict distinct functional roles during working memory. *Nat. Commun.* 9:3499. doi: 10.1038/s41467-018-05961-4

Watanabe, M., Inoue, Y., Sakimura, K., and Mishina, M. (1992). Developmental changes in distribution of NMDA receptor channel subunit mRNAs. *NeuroReport* 3, 1138–1140. doi: 10.1097/00001756-199212000-00027

Yokoi, A., and Diedrichsen, J. (2019). Neural organization of hierarchical motor sequence representations in the human neocortex. *Neuron* 103, 1178.e7–1190.e7. doi: 10.1016/j.neuron.2019.06.017

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neural Circuits

REVIEW
published: 15 March 2021
doi: 10.3389/fncir.2021.648624

Neuromodulation of Persistent Activity and Working Memory Circuitry in Primate Prefrontal Cortex by Muscarinic Receptors

Susheel Vijayraghavan[1] *and Stefan Everling*[1,2]*

[1] Department of Physiology and Pharmacology, The University of Western Ontario, London, ON, Canada, [2] Robarts Research Institute, The University of Western Ontario, London, ON, Canada

Neuromodulation by acetylcholine plays a vital role in shaping the physiology and functions of cerebral cortex. Cholinergic neuromodulation influences brain-state transitions, controls the gating of cortical sensory stimulus responses, and has been shown to influence the generation and maintenance of persistent activity in prefrontal cortex. Here we review our current understanding of the role of muscarinic cholinergic receptors in primate prefrontal cortex during its engagement in the performance of working memory tasks. We summarize the localization of muscarinic receptors in prefrontal cortex, review the effects of muscarinic neuromodulation on arousal, working memory and cognitive control tasks, and describe the effects of muscarinic M1 receptor stimulation and blockade on the generation and maintenance of persistent activity of prefrontal neurons encoding working memory representations. Recent studies describing the pharmacological effects of M1 receptors on prefrontal persistent activity demonstrate the heterogeneity of muscarinic actions and delineate unexpected modulatory effects discovered in primate prefrontal cortex when compared with studies in rodents. Understanding the underlying mechanisms by which muscarinic receptors regulate prefrontal cognitive control circuitry will inform the search of muscarinic-based therapeutic targets in the treatment of neuropsychiatric disorders.

Keywords: muscarinic acetylcholine receptor, M1 receptor, M2 receptor, working memory, persistent activity, prefrontal cortex, antisaccade, primate

OPEN ACCESS

Edited by:
Shintaro Funahashi,
Kyoto University, Japan

Reviewed by:
Shengtao Yang,
Yale University, United States
Andreas Nieder,
University of Tübingen, Germany

***Correspondence:**
Stefan Everling
severlin@uwo.ca

Received: *01 January 2021*
Accepted: *11 February 2021*
Published: *15 March 2021*

Citation:
Vijayraghavan S and Everling S (2021) Neuromodulation of Persistent Activity and Working Memory Circuitry in Primate Prefrontal Cortex by Muscarinic Receptors. Front. Neural Circuits 15:648624. doi: 10.3389/fncir.2021.648624

INTRODUCTION

The ability to maintain and manipulate information about the sensory world, motor actions, and previously learned experience is central to cognition and flexible behavior. Persistent, short-term elevated activity in cortical circuits has been proposed to Fuster and Alexander (1971) and Goldman-Rakic (1995) underlie the capacity to actively maintain such knowledge, or "working memory" (WM). The prefrontal cortex (PFC) in primates plays a pivotal role in the neural circuitry that processes such behaviorally relevant mental representations that are deployed to guide imminent choices and actions (Fuster and Alexander, 1971; Fuster, 1992, 1993; Miller and Cohen, 2001).

All major ascending neuromodulatory systems innervate the cerebral cortex, including the PFC and influence the dynamics of persistent activity and cortical WM circuitry (Arnsten et al., 2012). The modulatory actions of acetylcholine (ACh) on cortical function have been of long-standing interest partly because cholinergic dysfunction has been implicated in cognitive and WM deficits that manifest in psychiatric and neurological disorders including Alzheimer's disease (Hampel et al., 2018, 2019), dementia associated with Parkinson's disease (Noufi et al., 2019), major depressive disorder (Dagytë et al., 2011), and schizophrenia (Sarter and Bruno, 1998; Dean et al., 2003). Progressive cortical cholinergic deafferentation is a hallmark of Alzheimer's dementia and cholinergic pathology accompanies the cognitive disruption that manifests in the disease (Hampel et al., 2018). Inhibitors of acetylcholinesterase, which breaks down released acetylcholine, is a standard component of the treatment regimen in Alzheimer's dementia, although its efficacy in ameliorating cognitive deficits in patients has been questioned (Marucci et al., 2020). Decreased muscarinic receptor density has been reported in patients with schizophrenia (Dean et al., 2002; for detailed review see Dean et al., 2003). Further, xanomeline, a muscarinic agonist has shown clinical promise and improves short-term memory and other cognitive functions in schizophrenic patients (Shekhar et al., 2008).

Acetylcholine mediates its neuromodulatory influence via the ionotropic nicotinic and metabotropic G-protein coupled muscarinic receptor families (Picciotto et al., 2012). Subtypes from both cholinergic receptor families function in cortical WM circuitry, including in the PFC. There has recently been considerable interest in how ACh, acting through these receptors, influences neurophysiology of primate PFC during the performance of WM tasks (Baxter and Crimins, 2018; Galvin et al., 2018, 2020a; Vijayraghavan et al., 2018). Here, we will review studies of cortical muscarinic neuromodulation of WM performance and recapitulate recent work from our laboratory and others exploring muscarinic neuropharmacology of persistent activity and WM representations in primate PFC. Whereas there are several excellent published synopses regarding the functions of cortical ACh (McCormick, 1993; Steriade, 2004; Picciotto et al., 2012; Venkatesan et al., 2020), nicotinic and muscarinic neuromodulation of cognition and WM (Sarter and Bruno, 1997; Robbins and Arnsten, 2009; Klinkenberg and Blokland, 2010; Wallace and Bertrand, 2013), we will primarily focus on neurophysiological and pharmacological studies in dorsolateral PFC of non-human primates in this review.

LOCALIZATION OF MUSCARINIC RECEPTORS IN PRIMATE PFC

Among several cholinergic nuclei in the brainstem and basal forebrain, the nucleus basalis of Meynert is the principal source of ACh in the primate cerebral cortex (Mesulam et al., 1983; Lewis, 1991; Smiley et al., 1997). Additionally, in rodents, a fraction of cortical interneurons that express vasoactive intestinal peptide, also coexpress choline acetyltransferase, an enzyme that synthesizes ACh (Eckenstein and Baughman, 1984). However, hitherto such putatively cholinergic and GABAergic interneurons have not been shown in primate cerebral cortex (Mesulam et al., 1983), and the basal forebrain appears to be the only source of cholinergic innervation of cortex in primates. Corticopetal cholinergic afferents from the nucleus basalis innervate superficial and deep layers of macaque PFC, forming both symmetric synapses and boutons in proximity to symmetric and asymmetric synapses near dendritic spines (Mrzljak et al., 1995). Interestingly, the fraction of cortical cholinergic varicosities exhibiting synaptic specializations increases in primates when compared with rodents and is further augmented in humans (Smiley et al., 1997) versus monkeys (Mrzljak et al., 1995). Thus, ACh innervation of primate PFC has the capacity to act through both synaptic specialization and volume transmission (Mrzljak et al., 1995).

The prominent nicotinic receptor subtypes expressed in macaque cortex, including PFC, are α4β2 and α7 receptors (Wallace and Bertrand, 2013; Galvin et al., 2018). The muscarinic ACh family is comprised of G_q-coupled M1, M3, and M5 receptors and $G_{i/o}$-coupled M2, and M4 receptor families (Caulfield and Birdsall, 1998; Brown, 2010; Jones et al., 2012). Of these, M1 and M2 receptors are prominently expressed in PFC in primates. M1 receptors (M1Rs) and M2 receptors (M2Rs) are present in PFC in rodents (Levey et al., 1991), primates (Mrzljak et al., 1993; Medalla and Barbas, 2012) and humans (Scarr et al., 2009; Bubser et al., 2011; Dean and Scarr, 2016). M3 receptor expression has been examined in rats and is mainly expressed in the hippocampus and to a lesser extent in cerebral cortex, where it is absent in cortical layer III/IV (Levey et al., 1994; Bubser et al., 2011). However, M3 receptor expression has not been examined thus far in primate cerebral cortex owing to lack of selective immunohistochemical tools in primates. Expression patterns of other muscarinic receptor subtypes in monkey PFC are hitherto unknown.

In rodents, cortical M1 receptors are predominantly expressed postsynaptically (Levey, 1996) and are presumed to mediate the excitatory effect of muscarinic agonists on cortical activity in brain slices (McCormick, 1989, 1993; McCormick et al., 1993). Autoradiography using M1R- and M2R-preferring compounds suggests that M1R laminar expression in monkey PFC is present in all layers with strong bands of expression in layers III and V, while M2Rs are enriched in layer III and V/VI in the PFC, with the exception of Walker's area 46, where the expression is predominantly in layer V (Lidow et al., 1989; Mrzljak et al., 1993).

M1 receptor expression in monkey cortex was examined with immunohistochemistry and mRNA expression in a series of studies by Mrzljak and colleagues (Lidow et al., 1989; Mrzljak et al., 1993, 1996, 1998). In area 46 (dorsolateral PFC) of the macaque, M1Rs were expressed throughout all layers with greatest expression found in supragranular layer III and infragranular layers V/VI. Conspicuous M1R expression was found in the soma, dendrites, dendritic spines, and in close association with both asymmetric (presumably glutamatergic) and symmetric synapses (presumably GABAergic or cholinergic). The presence of M1Rs in conjunction with asymmetric synapses in PFC circuitry points to a role in modulating thalamocortical and corticocortical excitatory transmission. Early

studies indicated that M2Rs serve as autoreceptors on cholinergic efferents, inhibiting ACh release from terminals (Dudar and Szerb, 1969; Mash et al., 1985; Mash and Potter, 1986), and muscarinic inhibition of the release of tritiated ACh does not occur in M2R knock-out mice (Zhang et al., 2002). Further decline in cortical M2R expression in Alzheimer's disease has been attributed to the degeneration of cholinergic afferents from the basal forebrain (Flynn et al., 1995; Levey, 1996). Remarkably, immunohistochemical examination of the cortical M2Rs in monkey PFC (Mrzljak et al., 1993, 1998) and primary visual cortex (Mrzljak et al., 1996; Disney et al., 2006; Disney and Aoki, 2008) revealed a more complex profile of M2R expression. PFC M2R expression was found in both pre-and postsynaptic specializations, and in both cases, the expression was associated with both symmetric and asymmetric synapses in pyramidal and non-pyramidal neurons (Mrzljak et al., 1993). In primary visual cortex, M2R expression forms interdigitated patches of dense and sparse expression that coincide, respectively, with interblobs and blobs defined by cytochrome oxidase staining (Mrzljak et al., 1996; Disney et al., 2006). M2R expression was more prominent in the parvocellular inferotemporal channel of the visual stream wherein neurons possess orientation tuning but lack color opponency. Thus, M2R expression in primary visual cortex constitutes an intriguing example of convergence between neuromodulatory specialization and functional segregation. Associational and cross-callosal projections also form interdigitated stripes in dorsolateral PFC (Goldman-Rakic and Schwartz, 1982; Schwartz and Goldman-Rakic, 1984; Pucak et al., 1996). However, it is hitherto unknown if muscarinic receptor expression demonstrates congruence with these hodological features in PFC. Mrzljak et al. (1993) did not comment on whether anisotropy was observed in the larger scale distribution of M1Rs or M2Rs in their reports on muscarinic receptors in dorsolateral PFC.

However, other elegant work from the Barbas group demonstrated hodological specificity in M2R expression in dorsolateral PFC (Medalla and Barbas, 2012). The cholinergic arousal system remains active during waking and rapid eye movement sleep (REM, "paradoxical sleep"), but not during the slow wave non-REM phase of sleep, in contrast to norepinephrine, the other major neuromodulatory system involved in arousal (Aston-Jones and Bloom, 1981; Lee and Dan, 2012). Additionally, while the cerebral cortex is in a deactivated state during non-REM sleep, positron emission tomography studies have shown that certain limbic prefrontal areas are reactivated earlier upon the transition to REM sleep, while dorsolateral PFC remains deactivated (Muzur et al., 2002). Muzur et al. (2002) proposed that this was due to selective cholinergic inhibition of dorsolateral PFC. Medalla and Barbas (2012) tested this hypothesis in the context of the anterior cingulate cortex (ACC; area 32) and areas 9 and 46 of the dorsolateral PFC. Cholinergic innervation of the ACC is dense in comparison to dorsolateral PFC, and the ACC sends a substantial glutamatergic projection to the latter (Johnston et al., 2007). The ACC is reactivated earlier during REM sleep, and the question remains as to how the dorsolateral PFC does not also get activated concomitantly, given the strong excitatory projection from ACC and restored cholinergic tone in the dorsolateral PFC during REM sleep. Medalla and Barbas (2012) hypothesized that the distribution of M2Rs on ACC projections and their postsynaptic targets could account for the lack of REM sleep activation of dorsolateral PFC. Using serial electron microscopy, fluorescent immunohistochemistry and pathway tracing, Medalla and Barbas (2012) examined M2R expression in area 9 of the PFC. They found that presynaptic M2R expression was enriched in the glutamatergic afferents from ACC to area 9 when compared with associational fibers from PFC area 46. These presynaptic M2Rs, when activated by ACh release in PFC area 9, would lead to presynaptic suppression of glutamate release (Kimura and Baughman, 1997), thus nullifying the strong excitatory drive from the ACC during REM sleep when ACh is being released in the cortical mantle. M2Rs were also found the dendritic shafts of putative inhibitory neurons targeted by the ACC projection, while M2Rs were localized primarily on dendritic spines of pyramidal neurons that were targets of the associational fibers from neighboring PFC area 46. PFC areas 9 and 46 share functional congruence in the generation and maintenance of persistent activity and WM in the cognitive circuitry of dorsolateral PFC, and therefore M2R postsynaptic expression in area 9 pyramidal neurons receiving inputs from area 46 may have facilitatory physiological effects. Similarly, cholinergic suppression in the rat piriform cortex occurs only in synapses associated with intrinsic projections, and not afferent inputs (Hasselmo and Bower, 1992). Thus, M2R expression shows remarkable specificity and correspondence with functional and hodological attributes.

NEUROPHARMACOLOGY OF CORTICAL MUSCARINIC RECEPTORS IN AROUSAL

The ascending reticular activating system, including the cholinergic and noradrenergic systems, regulate the sleep-arousal cycle and transitions between the different brain states (Moruzzi and Magoun, 1949; Lee and Dan, 2012). During slow-wave sleep (Non-REM sleep), thalamocortical circuitry is in a relatively quiescent slow oscillation with synchronized transitions between silent and active states (Steriade et al., 2001; Constantinople and Bruno, 2011). Signatures of this oscillation are manifest across different physiological scales: the activity of individual neurons and their membrane potential, the local field potential (LFP) and in the scalp electroencephalogram (EEG). Upon transitioning to REM sleep and the wake state, overall activity of individual neurons becomes desynchronized, with concomitant desynchronization in the other electrophysiological signatures of brain states, including the LFP and scalp EEG. The bistable activity seen during slow wave sleep, comprising of brief interludes of quiescence and activity have been termed down- and up-states (Contreras and Steriade, 1995). Isolated cortical slices can spontaneously replicate this mode of bistable activity, termed the slow oscillation (Sanchez-Vives and McCormick, 2000). Indeed, this low frequency (0.1–0.5 Hz) oscillation can be expressed in cultured random cortical networks and deafferented cortical slabs (Sanchez-Vives et al., 2017) and is a unifying feature

of cortical activity under different anesthetic regimes (Lewis et al., 2012). It has been proposed that this is a default activity pattern and emergent property of cortical networks (Sanchez-Vives et al., 2017). It has also been speculated that the wake state may be akin to a persistent up-state (Constantinople and Bruno, 2011), where oscillatory transitions to quiescence do not occur. *In vivo*, the dynamics of ascending neuromodulatory arousal systems engender these transitions (Moruzzi and Magoun, 1949; Constantinople and Bruno, 2011; Jones, 2020).

The ascending brainstem and basal forebrain cholinergic systems causally contribute to regulating these physiological transitions accompanying brain state transitions (Metherate et al., 1992; Jones, 2008). A brief discussion of this subject may be useful here, since the physiological mechanisms that are at play in cholinergic modulation of arousal may have commonalities with cholinergic neuromodulation of cortical neurophysiology during cognitive control and WM. Interestingly, as noted previously, cholinergic nuclei are active during REM sleep, when cortical activity, LFP and the EEG are indistinguishable from the wake state, which is not the case for the monoamine arousal systems, which are very active during wakefulness, possess low activity during non-REM sleep, but are completely quiescent during REM sleep. Basal forebrain stimulation in rodents can depolarize auditory cortical neurons, desynchronize their activity with shifts in subthreshold membrane potential oscillations from low ($\leq$5 Hz) to high (20–40 Hz) oscillations (Metherate et al., 1992). Optogenetic activation of basal forebrain cholinergic neurons, or of terminal projections thereof in the primary visual cortex, desynchronize the LFP and visual cortical neuronal activity. Lee and Dan (2012) proposed that two effects of muscarinic modulation, viz., muscarinic suppression of intracortical synaptic transmission (Gil et al., 1997; Hsieh et al., 2000; Medalla and Barbas, 2012) and muscarinic depolarization of cortical neurons may be instrumental in ACh-induced depolarization during brain state transitions. In mice, ACh neuromodulation of somatostatin-positive interneurons, that participate in a disynaptic disinhibitory relay through their inhibition of parvalbumin positive interneurons, has been shown to be necessary for the desynchronization of neuronal firing and the LFP in somatosensory cortex (Chen et al., 2015). Muscarinic agonists increase REM state duration and decrease the onset latency of REM sleep induction (Sitaram et al., 1976; Hohagen et al., 1993), while muscarinic receptor antagonists decrease the duration of REM sleep (Velazquez-Moctezuma et al., 1990; Gillin et al., 1991; Kim and Jeong, 1999). A recent study showed that a double knockout of M1Rs and M3Rs in mice almost completely abolishes REM sleep, indicating that G_q coupled muscarinic receptors were essential in regulating this epoch of the sleep rhythm (Niwa et al., 2018).

There are tonic and phasic components to the activity of cholinergic neurons and ACh fluctuations *in vivo* can be measured by amperometric methods (Parikh et al., 2004, 2007). Amperometric monitoring of ACh has revealed that tonic ACh release is coordinated across multiple brain areas during REM sleep, while phasic ACh release is synchronized during performance of a WM task (Ruivo et al., 2017).

To summarize, the cholinergic system is essential in the regulation of states of cortical arousal, and neuromodulation by muscarinic receptors, particularly of the M1R family, are essential in the generation of desynchronized states during REM sleep and awake behavior. Some of the physiological mechanisms that elicit the transitions to wake-like desynchronized cortical activity may also be involved in the neuromodulation of persistent activity in awake cortical circuits engaged in active behavior.

MUSCARINIC MODULATION OF WM AND COGNITIVE CONTROL CIRCUITRY

A substantial body of work has examined the role of the cholinergic system in cognitive performance in rodents, monkeys, and humans (Fibiger, 1991). Early clues about the importance of muscarinic receptor function in WM performance came from Bartus and Johnson (1976), who found that systemic muscarinic blockade with muscarinic antagonist scopolamine caused a delay-dependent deficit in a match-to-sample WM task, wherein the deficits were pronounced only at longer delays. Thereafter, many studies have replicated and elaborated upon this deficit in WM performance in monkeys (Rusted and Warburton, 1988; Rupniak et al., 1991; Rusted et al., 1991; Spinelli et al., 2006). Interestingly, Rupniak et al. (1991) found that the cholinesterase inhibitor, physostigmine, could reverse pro-amnestic deficits caused by systemic scopolamine, but physostigmine could not improve WM performance in aged monkeys or when distractor load was increased in the task. Moreover, they reported that stimulus luminance did not interact with scopolamine-induced deficits in delayed response performance, whereas increasing attentional load by reducing stimulus presentation time exacerbated scopolamine's effects independent of the length of the delay, leading the authors to argue that systemic muscarinic blockade may affect attentional aspects of task performance, instead of WM. Nevertheless, a gamut of studies have shown that systemic muscarinic blockade caused WM, delayed match-to-sample, recognition memory and other cognitive deficits and pro-psychotic states (Penetar and McDonough, 1983; Aigner et al., 1987; Aigner and Mishkin, 1993; Barak and Weiner, 2006, 2009; Buccafusco et al., 2008; Plakke et al., 2008; Barak, 2009). This has led to the identification of a collection of cognitive deficits and psychosis-like symptoms termed the anti-muscarinic syndrome (Yeomans, 1995). Scopolamine has been shown to affect sensory discrimination, acoustic startle reflex, prepulse inhibition, recognition memory, short-term memory, delayed match-to-sample, set shifting, and attentional performance in rodents (Dunnett et al., 1990; Jones and Shannon, 2000; Ukai et al., 2004; Barak and Weiner, 2006). Systemic muscarinic blockade causes deficits in the WM performance and executive function in humans (Green et al., 2005; Ellis et al., 2006). Muscarinic blockade also produces deficits in the learning of rules specifying outcome-based odor discrimination in rodents (Saar et al., 2001). Lesions of the basal forebrain cholinergic system in monkeys have reported conflicting effects on cognitive tasks. In one study basal forebrain ibotenic acid lesions cause recognition memory and delayed non-match-to-sample performance deficits

(Aigner et al., 1991). Contrastingly, in another study, lesions of the nucleus basalis of Meynert in monkeys appear to spare delayed response performance with short delays and instead cause attentional performance deficits (Voytko et al., 1994). Since systemic drug administration or lesions cholinergic nuclei innervating cerebral cortex could have manifold effects on the distributed brain circuitry that subserves various components of these behavioral tasks, deficits in cognitive control and WM highlighted by the studies summarized above do not necessarily indicate deficits in WM circuitry in the PFC or alterations in PFC persistent activity that maintains WM representations.

However, other reports have addressed the role of cholinergic innervation locally in the PFC in WM and other PFC-dependent cognitive tasks. Baxter's group reported an intriguing finding: cholinergic deafferentation, by injection of an immunotoxin based on saporin in the PFC of rhesus monkeys, resulted in a specific and selective deficit in spatial delayed response, but not in other demanding tasks that engaged attention but did not require WM, such as strategy implementation, object-in-place scene learning, or reward-based decision-making as assessed by reinforcer devaluation (Croxson et al., 2011). Other evidence about the role of ACh neuromodulation during delayed response tasks comes from the neurophysiology of the nucleus basalis. In monkeys engaged in spatial delayed responses, most nucleus basalis neurons were active only during the choice and reward phases of the task, and the proportion of neurons responding during the delay period was far less prominent (Richardson and DeLong, 1986). Richardson and DeLong further reported that nucleus basalis neuronal activity in response to stimuli depended on the task context, but that the activity of these neurons was related to rewarding or aversive stimuli and cues that predict them (Richardson and DeLong, 1991). Interestingly, in macaques, intermittent stimulation of the nucleus basalis improves performance in a WM task, while continuous stimulation degrades performance (Liu et al., 2017). A brief stimulation of cholinergic fibers has been shown to cause long lasting modulation of hippocampal and cortical activity stimulation (Krnjević et al., 1981; Cole and Nicoll, 1983; McCormick and Prince, 1986), mediated by the modulation of the M-current, so named because of its inhibition by muscarinic stimulation (Brown and Adams, 1980). The M-current is generated by voltage-gated KCNQ potassium channels and it counteracts overexcitability upon neuronal depolarization; its inhibition by muscarinic receptors causes an increase in excitability of neurons.

MUSCARINIC MODULATION OF PFC NEUROPHYSIOLOGY AND PERSISTENT ACTIVITY

There have been fewer studies describing the neurophysiological effects of muscarinic modulation in primates. Muscarinic blockade disrupts attentional modulation of cortical activity in primary visual cortex but nicotinic blockade, while reducing V1 excitability, did not affect attentional modulation (Herrero et al., 2008). Herrero et al. (2008) also reported that muscarinic antagonist scopolamine, in addition to reducing the overall activity of PFC neurons, reduced the attentional component of V1 neuronal firing. Herrero et al. (2008) also found that ACh increased the activity of V1 neurons, and at low doses, enhanced neuronal attentional selectivity. However, at higher doses ceiling effects appeared due to non-specific increase in activity that disrupted attentional modulation. One point that emerges from this is that the actions of ACh in V1 appear to be uniformly excitatory, an observation that will be pertinent to our discussion of ACh actions in PFC later. Consistent with this effect of muscarinic receptors in attentional modulation, muscarinic receptor blockade by systemic and local infusion of scopolamine in macaque intraparietal cortex, including the lateral intraparietal area and area 7a, produced a deficit in covert orienting in a cued stimulus detection task (Davidson et al., 1999; Davidson and Marrocco, 2000). A salient stimulus appeared in one of two previously cued locations and monkeys were trained to manually respond to the stimulus onset for reward. Attention was captured by changing the luminance of one of the cues prior to the appearance of the stimulus. Analysis of reaction times after scopolamine infusion demonstrated that covert attentional orienting was compromised.

Miller and Desimone (1992) recorded neuronal activity with systemic muscarinic blockade in the inferotemporal cortex during the performance of a delayed match-to-sample recency memory task with sequential delayed presentations of multiple stimuli after a test stimulus. The rewarded response was a lever release when a succeeding stimulus matched the test stimulus. Systemic scopolamine administration was deleterious to task performance, demonstrating the ACh actions through muscarinic receptors promoted recency memory. Surprisingly, this performance deficit was not accompanied by commensurate changes in the stimulus related activity of inferotemporal neurons. The number of neurons that showed selectivity for match vs. non-match stimuli was not significantly affected by muscarinic blockade. However, when compared with placebo, scopolamine administration caused paradoxical increases in stimulus responsive neuronal activity compared with baseline activity (Miller and Desimone, 1992). This increased stimulus responsivity did not change quantifiable task-related information in the neuronal activity. Thus, it appears that the effects of muscarinic blockade on task performance were not explained by changes in inferotemporal cortical activity.

Another study examined the effects of systemic injections of scopolamine on the activity of macaque PFC neurons (Zhou et al., 2011) engaged in an oculomotor delayed response task (Hikosaka and Wurtz, 1983; Funahashi et al., 1989) with varying distractor load (**Figure 2**). Delay period activity was suppressed after scopolamine administration, and behavioral performance degraded (**Figure 2A**) with small increases in saccade reaction times and saccade end-point dispersion. Interestingly, the performance degradation due to scopolamine was independent of distractor load during the trial (**Figure 2A**). The effects of muscarinic blockade were contingent upon the presence of a delay, whereby visually guided saccade performance (zero second delay) was unaffected. While scopolamine had modest effects on overall neuronal activity in the PFC, it significantly affected

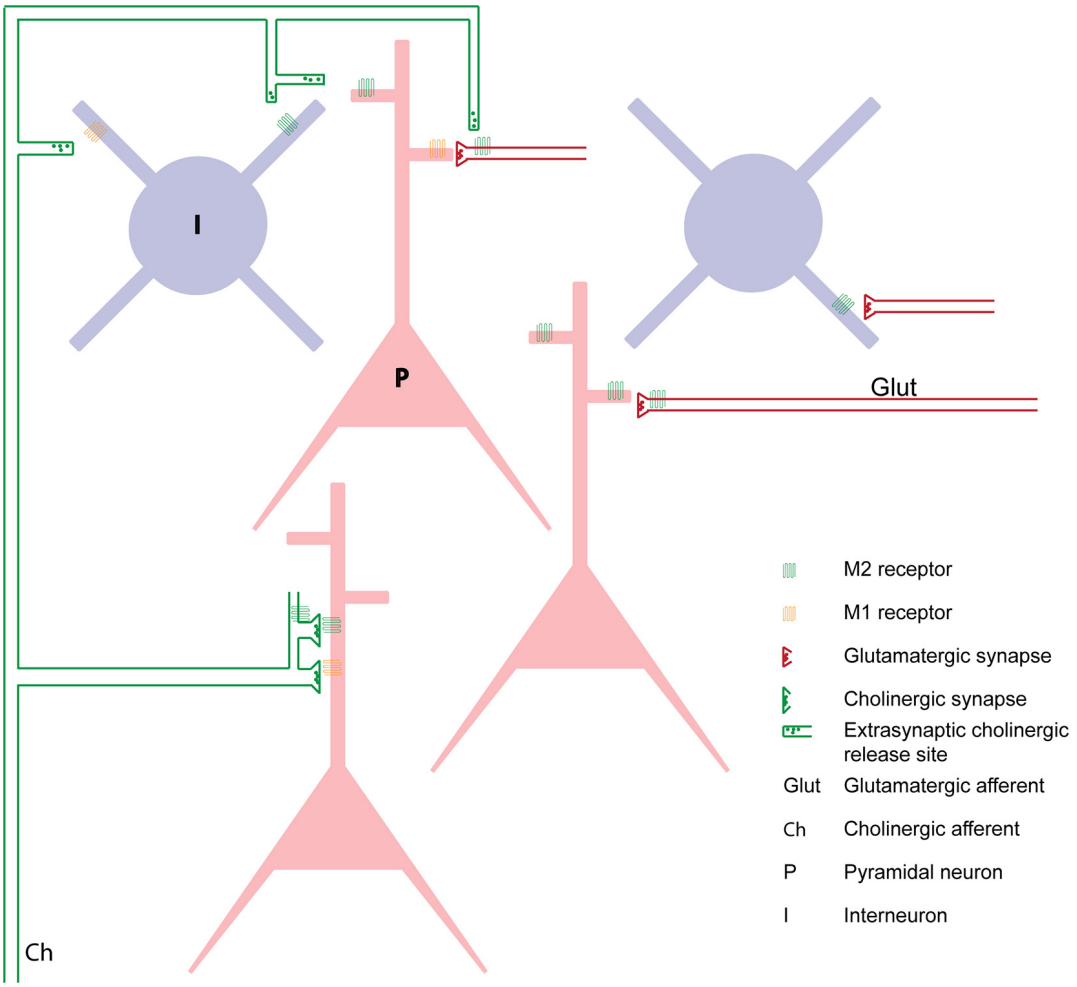


FIGURE 1 | Schematic showing the localization of muscarinic receptors on PFC neurons and their relationship to cholinergic and glutamatergic innervation based on data from Mrzljak et al. (1993) and Medalla and Barbas (2012). M1Rs are localized on dendritic shafts of pyramidal and interneurons, and on the spines of pyramidal neurons where they are apposed to glutamatergic synapses. Cholinergic synapses are mainly found on dendritic shafts, while extrasynaptic cholinergic release sites on cholinergic axons result in diffuse volume transmission to influence muscarinic receptors on PFC neurons. M2Rs are found on postsynaptic dendritic spines of pyramidal neurons and on dendritic shafts of interneurons. M2Rs are also found on presynaptic terminals of glutamatergic afferents in PFC and as autoreceptors on cholinergic terminals.

the delay period activity after visual stimulus presentation (**Figure 2B**). The authors also found that stimulus-related activity during the presentation of the peripheral cue in the delayed response task was comparatively unaffected, indicating that sensory stimulus processing in the PFC was not affected. Zhou et al. (2011) also tested the effects of scopolamine on PFC activity and performance in a delayed match/non-match-to-position task. The monkeys reported whether a second cue, presented after a short delay, was at the same or different location with respect to the first cue. They found that scopolamine's effects were not idiosyncratic to the oculomotor delayed response paradigm and manifested in the delayed match-to- position task also, whereby memory period persistent activity encoding the position of the first cue was diminished by systemic muscarinic blockade. Further, in this task, the authors found that, when the first stimulus was presented outside the neuron's response field and the second stimulus appeared within the neuron's response field, there was elevated activity in the delay period prior to the appearance of the second stimulus which reflected covert anticipation of the onset of the second stimulus in the response field. Muscarinic blockade diminished this anticipatory memory period activity. Thus, systemic muscarinic blockade had pronounced effects on PFC persistent activity representing the remembered location of a target and disrupted mnemonic performance.

The results of Zhou et al. (2011) indicated that muscarinic blockade reduces activity in PFC, but since that study employed systemic injections, it could not be determined if this effect was due to blockade of local PFC muscarinic receptors or due to network consequences of muscarinic blockade elsewhere in the brain. Our group has conducted experiments on the effects of local muscarinic blockade using microiontophoresis

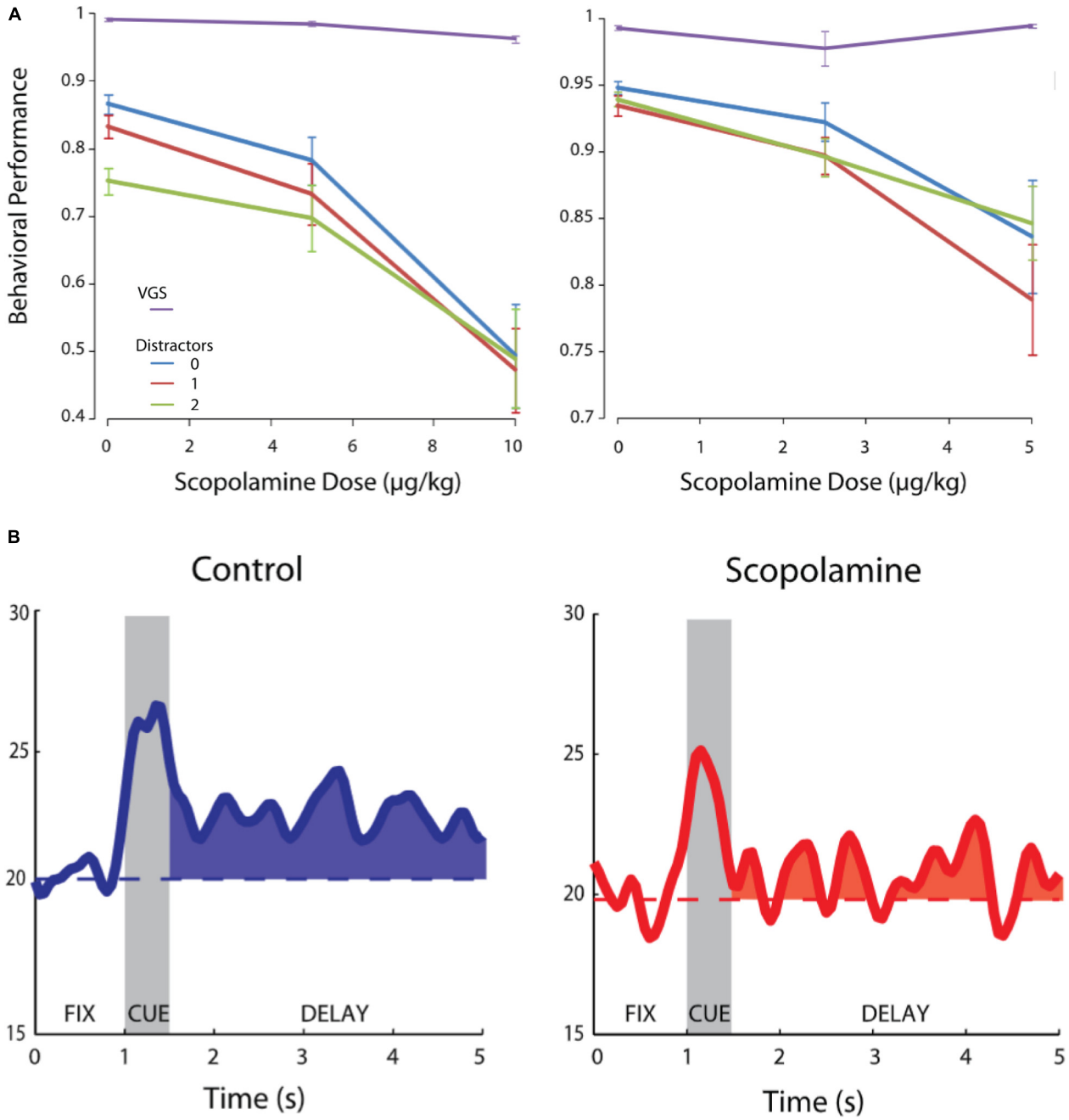


FIGURE 2 | Effects of systemic scopolamine administration on oculomotor delayed response performance and delay activity of PFC units. Adapted from Zhou et al. (2011). **(A)** Behavioral performance in the oculomotor delayed response task for two monkeys after scopolamine infusion at various doses. Purple, visually guided saccades (0 s delay); Blue, 0 distractors during delay; Red, 1 distractor during delay; Green, 2 distractors during delay. **(B)** Systemic scopolamine administration reduces delay period persistent activity of PFC neurons. Modified with permission from Journal of Neurophysiology.

on PFC neuronal activity while monkeys performed randomly interleaved pro- and antisaccades (**Figure 3A**), where the current trial rule had to be maintained in WM (Major et al., 2015). The pro- and antisaccade task is dependent of the integrity of dorsolateral PFC (Condy et al., 2007; Koval et al., 2011), and deficits in antisaccade performance are diagnostic indicator of the integrity of the PFC (Everling and Fischer, 1998). In the version of the task employed by Major et al. (2015), the rule cue was briefly presented, and had to be remembered through the memory period (**Figure 3A**). Persistent activity of PFC neurons encodes the task rule through this memory period (Skoblenick and Everling, 2012; Vijayraghavan et al., 2017). Microiontophoresis employs small electrical currents to eject charged moieties and drugs from the recording electrode

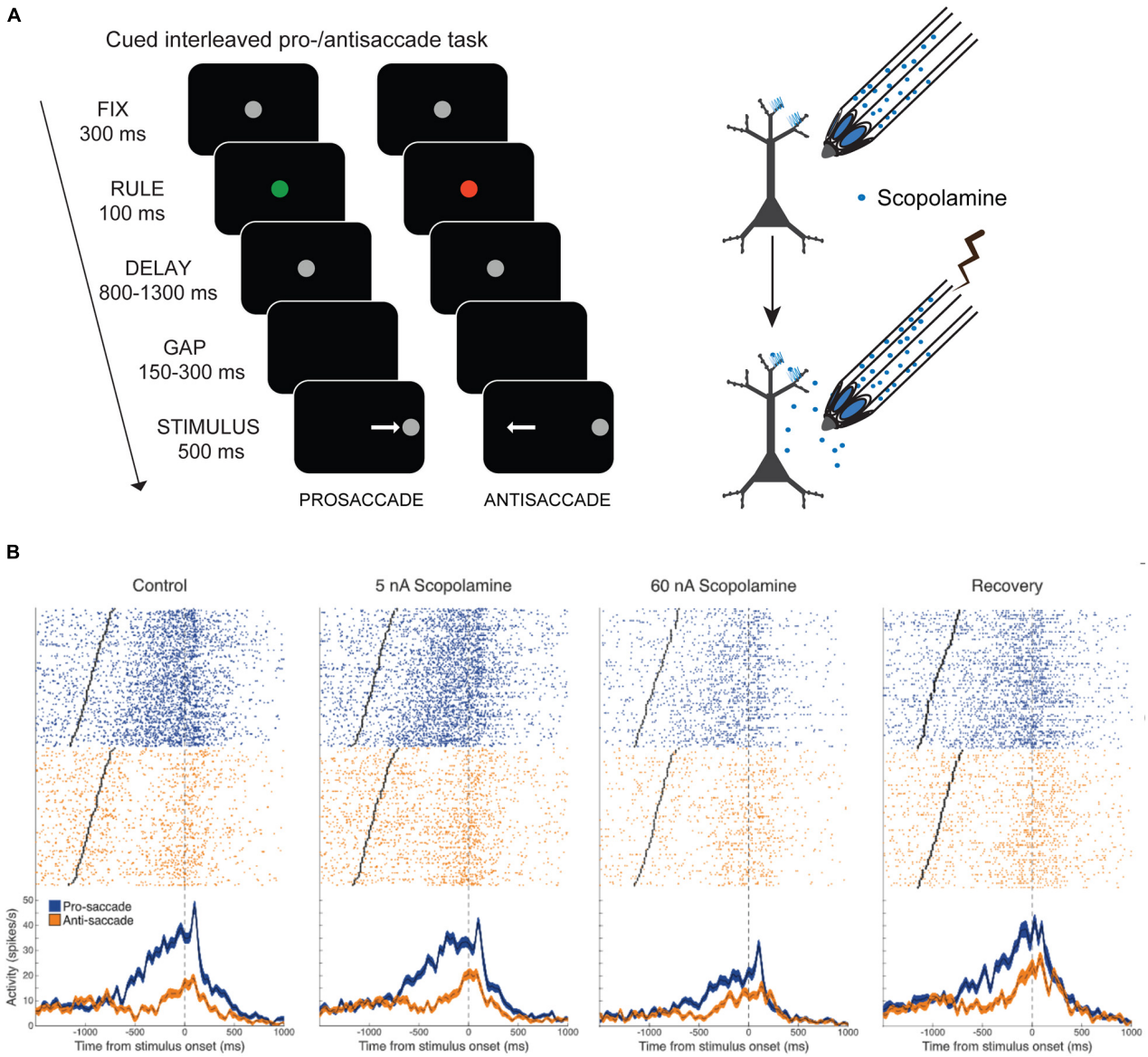


FIGURE 3 | Effects of local delivery of muscarinic antagonist scopolamine by microiontophoresis on PFC activity during pro- and antisaccade task. Adapted from Major et al. (2015). **(A)** Pro- and antisaccade task structure is shown. After central fixation, the fixation spot changes to a colored rule cue which flashes briefly. Subsequently the spot becomes white again. After a delay wherein the trial rule, based on the color of the rule cue, is maintained in WM, the fixation spot disappears, and after a brief gap, the peripheral stimulus appears left or right of the fixation spot. The subject makes a saccade toward (prosaccade) or away (antisaccade) from the spot, based on the current trial rule. Trial temporal structure are also shown. Illustration on the right shows the recording and iontophoresis technique. **(B)** Microiontophoresis of increasing doses of scopolamine cause increasing suppression of the activity of a PFC neuron that has delay-period activity preferring the pro-saccade rule. Right panel shows recovery after cessation of drug application.

(Hicks, 1984). The currents employed in these *in vivo* studies are on the order of ∼100 nA and are not expected to elicit extraneous electrophysiological effects on the recorded neurons. Moreover, usually, the quantities of drugs ejected are not enough to elicit behavioral effects. Major et al. (2015) found that local stimulation of muscarinic receptors dose-dependently and monotonically suppressed the activity of a majority of PFC neurons recorded during the rule WM task performance, and concomitantly degraded all forms of task-related neuronal selectivity, including WM for the rule (**Figure 3B**), peripheral stimulus selectivity and perisaccadic activity. Thus, some of the effects of systemic muscarinic blockade described by Zhou et al. (2011) would appear to be explained by local blockade of muscarinic receptors in PFC. In contrast to Zhou et al. (2011), we found that peripheral visual stimulus selectivity was also reduced upon local scopolamine application. These differences could be due to concentration differences due to systemic application versus local drug ejection or differences in the behavioral task

structure. In the interleaved rule-based antisaccade task, activity of PFC neurons is differentially modulated by the rule prior to the onset of the peripheral stimulus. These prestimulus activity differences are, perhaps in some respects, analogous to the anticipatory activity observed in the delayed match-to-position task in Zhou et al. (2011) and may convolve with visual stimulus responsivity accordingly.

Recently, stimulation of nicotinic α4β2 receptors was also examined in monkey PFC during spatial delayed response task performance (Sun et al., 2017). The authors found α4β2 receptor stimulation also augmented PFC delay period persistent activity in the oculomotor delayed response task and improved the memory period spatial tuning of these neurons in that period. Moreover, in a variant of that task where a distractor was presented during the delay period, α4β2 receptor stimulation shielded neuronal spatial tuning during the delay period from the effects of the distractor. α4β2 receptor stimulation did not affect sensory activity related to the peripheral visual stimulus or response-related perisaccadic activity at the end of the trial. However, α4β2 agonism also enhanced the activity of neurons that had activity related to central gaze fixation. Another report from the Arnsten group showed that iontophoretic stimulation of α7 nicotinic receptors enhanced NMDA-dependent persistent activity of PFC neurons (Yang et al., 2013). An α7 receptor agonist augmented delay period activity and spatial tuning of monkey PFC neurons during spatial delayed response, an effect which could be reversed by an α7 receptor antagonist. Interestingly, α7 receptor stimulation did not have appreciable effects on peripheral stimulus-related activity and instead facilitated and synergized with the actions of NMDA NR2B receptors on PFC neurons to influence delay period persistent activity.

Thus, the findings regarding muscarinic blockade of neuronal selectivity for the peripheral visual stimulus and perisaccadic selectivity in Major et al. (2015) contrast with analysis of the effects of α4β2 nicotinic receptor stimulation on stimulus-selective and perisaccadic neurons in Sun et al. (2017) and stimulus-selective activity after α7 receptor stimulation from Yang et al. (2013). We found that muscarinic blockade reduces selectivity for all task attributes in PFC neurons, including visual stimulus and saccade direction selectivity, having a comprehensive disruptive effect on PFC neuronal task engagement. Thus, nicotinic receptor subtypes in PFC appear to be more specialized in their actions on prefrontal circuitry that generates and maintains persistent delay activity, whereas general muscarinic receptor modulation appears to affect the gamut of observable PFC task-related activity.

Since muscarinic antagonism engendered such pronounced suppression of the activity of PFC neurons during WM, it would be expected that muscarinic and cholinergic agonists may enhance persistent activity and WM representations. There is some evidence that muscarinic stimulation can sustain persistent activity through intrinsic mechanisms (Egorov et al., 2002). Rat entorhinal cortical neurons, in the presence of the cholinergic agonist carbachol, respond to current pulse stimulation with long lasting activity that is reminiscent of persistent activity displayed by cortical neurons in WM tasks (**Figure 4A**). This carbachol-induced response is graded with increasing discharge rate after successive stimulations. The persistent responses could be blocked by general muscarinic blockade (**Figure 4A**) or by pirenzepine, an antagonist preferentially blocking M1Rs. Synaptic stimulation in concert with carbachol application also generated persistent spiking accompanied by the generation of nifedipine-sensitive Ca^{2+} plateau potentials. Thus, ACh, through muscarinic mechanisms, could facilitate persistent activity in cortical neurons through cell-autonomous intrinsic mechanisms that, in concert with stimulus evoked responses, could engender WM representations. However, traditionally, WM persistent activity is thought to be a network phenomenon, generated by slow reverberatory synaptic activity in a network of neurons (Wang, 2001). Whether this intriguing phenomenon that manifested in rodent entorhinal cortical slices would also occur *in vivo* in primate PFC was not clear.

To clarify whether carbachol could induce or augment persistent activity in PFC during WM task performance, our group conducted experiments where we microiontophoretically applied carbachol on PFC neurons in rhesus monkeys performing the rule-memory guided pro- and antisaccade task (Major et al., 2018). Surprisingly, we found that carbachol had mixed effects on neuronal physiology and persistent activity (**Figure 4B**). Carbachol application significantly excited roughly half the PFC neurons recorded, while ∼40% of neurons were inhibited. Moreover, carbachol increased the activity of broad-spiking presumed excitatory pyramidal neurons, while effects on excitability of narrow-spiking presumed mainly parvalbumin-positive interneurons were more varied. Rule encoding in the persistent activity during the delay epoch was diminished by carbachol application, especially at higher doses. This decrease in rule selectivity occurred notwithstanding the direction of changes in excitability of the neurons. It is noteworthy that carbachol is a general cholinergic agonist that has agonist activity at both muscarinic and nicotinic receptors. However, as discussed earlier, studies heretofore report that stimulation of the major nicotinic receptor subtypes in PFC appear to be generally excitatory (Yang et al., 2013; Sun et al., 2017), suggesting that the physiological actions of carbachol in Major et al. (2018) were mediated by muscarinic receptors.

NEUROMODULATION OF PFC PERSISTENT ACTIVITY BY MUSCARINIC RECEPTOR SUBTYPES

Given the pervasiveness of scopolamine-induced suppression of PFC neurons described above (Major et al., 2015), and that some or all of the effects of PFC carbachol stimulation were mediated by muscarinic receptors (Major et al., 2018), the question arises as to which muscarinic receptor subtypes contributed to the various physiological effects on persistent activity and task-selectivity changes caused by these cholinergic manipulations.

As discussed previously, M1Rs are the dominant muscarinic receptor subtype expressed in PFC. Since they are localized postsynaptically at asymmetric synapses on dendritic spines of pyramidal neurons, M1R constitutes an attractive candidate for mediating the general suppression of PFC by muscarinic

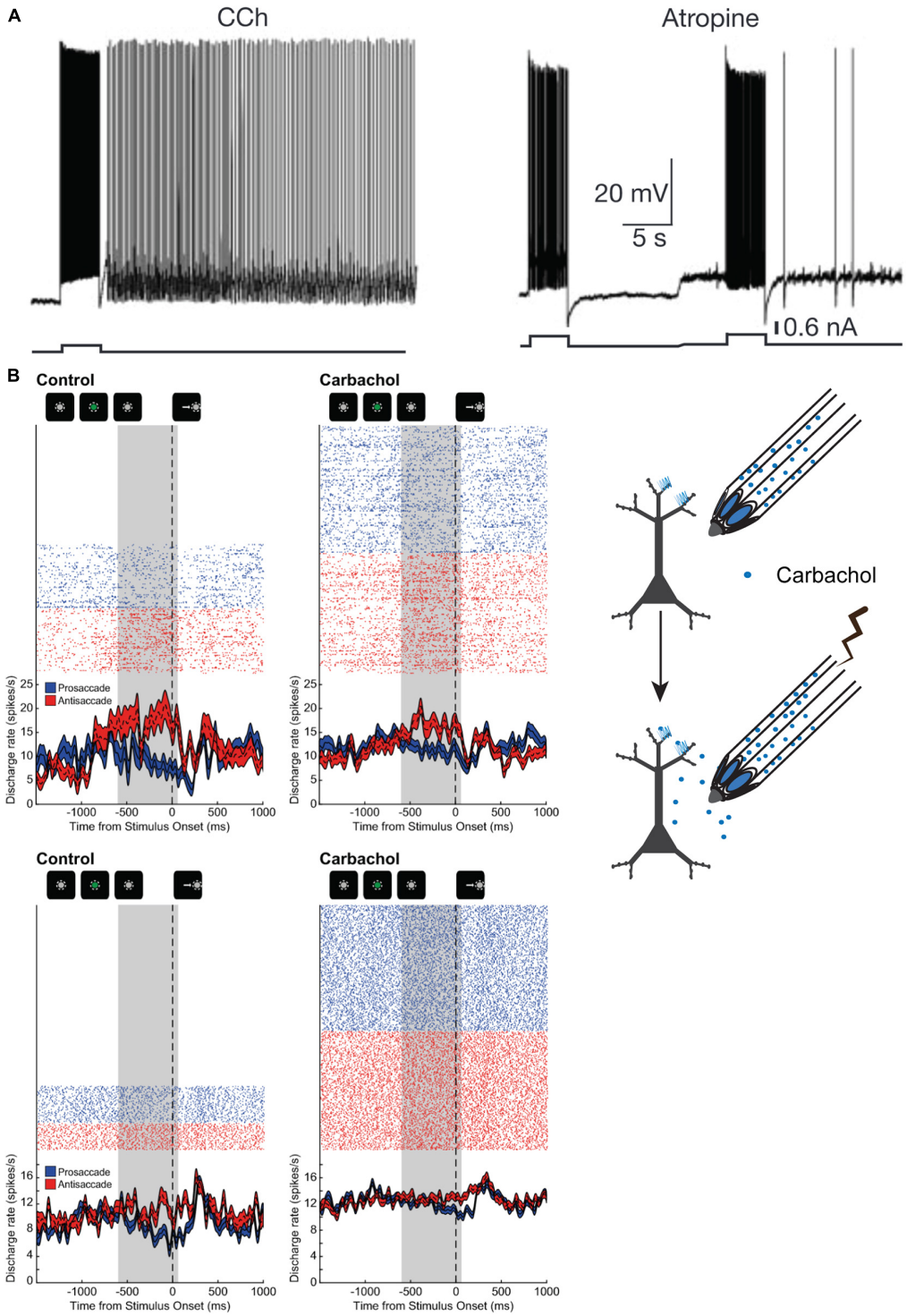


FIGURE 4 | Influence of cholinergic agonist carbachol on persistent activity. **(A)** adapted from Egorov et al. (2002); **(B)** adapted from Major et al. (2018). **(A)** Example of persistent activity evoked in a rat entorhinal cortical neuron by a current pulse in the presence of carbachol. Neuronal discharge persists after cessation of stimulus, and after muscarinic blockade, and after muscarinic blockade. Neuronal discharge persists after cessation of stimulus in the presence of carbachol, but not during blockade of muscarinic receptors. **(B)** Illustration on the right shows experimental design of neuronal recording and carbachol iontophoresis. Shown on the left is the activity of two PFC neurons **(top and bottom panels)** with persistent rule-selective activity during the delay period is shown during control **(left)** and during carbachol **(right)** application. Carbachol attenuated WM activity for the antisaccade rule in the neuron shown in the **top panel**, while the activity of the neuron in the **bottom panel** was augmented by carbachol, but selectivity for the trial rule in the delay period was nevertheless diminished. Gray area shows the last 600 ms of the delay period prior to fixation offset. Reproduced with permission from Nature Publishing group.

blockade. M1R stimulation inhibits the M-current and can thereby increase cortical neuronal excitability (McCormick and Prince, 1986; Marrion, 1997; Shirey et al., 2009; Young and Thomas, 2014). M1Rs and KCNQ channels are both expressed on dendritic spines and dendrites in layer III pyramidal neurons of PFC (Galvin et al., 2020b). An allosteric potentiator of M1R signaling increased the activity of medial PFC neurons in rodents *in vivo* and restored reversal learning in a transgenic model of Alzheimer's disease (Shirey et al., 2009). A selective M1R antagonist and scopolamine both produce antidepressant actions in rodents due to actions in medial PFC (Navarria et al., 2015). M1R knockout mice have been found to have selective deficits in non-match-to-sample tasks while, surprisingly, showing performance enhancement in match-to-sample tasks, with a reduction in theta burst stimulation, and long-term potentiation in mice (Anagnostaras et al., 2003). An M1R positive allosteric modulator was found to enhance cognitive task performance in macaques, including self-ordered spatial search, and an object retrieval detour task (Uslaner et al., 2013). M1R also mediates long-term excitability changes in striatal neurons (Lv et al., 2017). KCNQ channels that generate the M-current are active near the action potential threshold (Brown and Adams, 1980), and inhibition of the M-current by pharmacological blockade of KCNQ channels increases PFC delay period activity during oculomotor delayed response (Wang et al., 2011). The M-current is dependent on PIP2 levels, which are regulated by phospholipase C, downstream of G_q signaling (Suh and Hille, 2002, 2005, 2007; Suh et al., 2006). Since M1R is coupled to G_q signaling and the inositol phosphate pathway (Popiolek et al., 2016; Maeda et al., 2019), it may be the main conduit for increasing neuronal excitability in primate PFC by inhibiting the M-current.

On the other hand, M1R could have inhibitory influences by direct activation of parvalbumin-positive interneurons (Yi et al., 2014). In rhesus macaque areas V1 and MT, the majority of parvalbumin-positive interneurons are found to express M1Rs (Disney and Aoki, 2008; Disney and Reynolds, 2014), although it is not clear if this is also the case in PFC. M1R activation leads to Ca^{2+} mobilization from intracellular stores through the IP3 receptor and this release of Ca^{2+} can transiently hyperpolarize neocortical neurons through the activation of calcium-activated SK potassium channels (Gulledge and Stuart, 2005). The transient suppression is usually followed by long lasting depolarization of the neuron. Metabotropic glutamate receptors also mobilize this IP3-receptor and SK channel-dependent mechanism to cause transient suppression of cortical neurons as well (Hagenston et al., 2008). One confound in the interrogation of subtype-selective muscarinic actions has been the lack of subtype selectivity of orthosteric muscarinic agonists and antagonists, the ACh binding motif is conserved among the receptor subtypes (Jones et al., 2012; Jiang et al., 2014). Among the older generation of orthosteric compounds, some, such as the agonist McN-A-343 and the antagonist pirenzepine have a pharmacological preference for M1Rs (Mitchelson, 2012) but are not highly subtype-selective (Giachetti et al., 1986; Davies et al., 2001). Recently, however, a new class of M1R agents have been synthesized that show pharmacological activity by binding at allosteric sites on the receptor and show considerable subtype selectivity and clinical promise (Bubser et al., 2011). These comprise allosteric agonists and antagonists, that act on non-ACh receptor sites and activate or inhibit the receptor directly, and positive allosteric modulators, that do not activate the receptor alone, but in concert with endogenous ACh can augment the ACh response.

Recently, our group has tested the effects of M1Rs on persistent WM activity for rules in monkey PFC (Vijayraghavan et al., 2018). Microiontophoresis of a selective M1R allosteric agonist, VU0357017 (Lebois et al., 2010; Digby et al., 2012), M1R-preferring agonist McN-A-343 and M1R-preferring antagonist pirenzepine were performed on PFC neurons engaged in the rule-memory guided pro- and antisaccade task described earlier (**Figure 5A**). Surprisingly, we found that M1R-selective allosteric agonist VU0357017 dose-dependently and strongly suppressed PFC neurons during task performance (**Figure 5A**). At lower dose ranges, about half of the PFC neurons tested were inhibited by the allosteric agonist, while at higher dose ranges, almost all (81%) of neurons tested were inhibited. Application of the orthosteric agonist, McN-A-343, also induced substantial suppression of ∼60% of PFC neurons. Furthermore, application of the allosteric M1R agonist disrupted the rule selectivity of the persistent delay activity of many PFC neurons (**Figure 5B**), while a few neurons showing increases in persistent activity and increase in rule representation (Vijayraghavan et al., 2018). Interestingly, M1R blockade with pirenzepine (**Figure 5B**) also suppressed the activity of many PFC neurons (Vijayraghavan et al., 2018). However, the proportion of neurons that displayed suppression did not increase with higher doses and, at the population level was pirenzepine induced suppression was milder than that observed previously with general antagonist scopolamine and milder than the suppression with the high doses of the M1R-selective allosteric agonist. Moreover, in contrast to the effects of scopolamine, although application of the M1R-selective antagonist altered the rule-selectivity in the delay period activity in some individual PFC neurons, the rule selectivity at the level of the population was not significantly altered. M1R stimulation did not differentially affect narrow-spiking putative interneurons and regular-spiking putative pyramidal neurons, indicating that increased inhibition from parvalbumin-positive interneurons could not explain the physiological suppression caused by the agonist or antagonist.

In summary, these results indicated that M1R blockade could not account for the pervasive neuronal suppression and general disruption of task selectivity that was observed with general muscarinic blockade with scopolamine (Major et al., 2015). Further, M1R overstimulation unexpectedly has strong suppressive effects on WM activity in PFC.

Another recent study examined M1R modulation of PFC WM activity during oculomotor delayed response performance in aged monkeys (Galvin et al., 2020b). This study also found that high doses of the same allosteric M1R agonist, VU0357017, suppressed PFC WM activity, but in contrast with Vijayraghavan et al. (2018), low doses of the allosteric agonist enhanced PFC persistent activity (**Figures 6A,C**). Galvin et al. (2020b) also reported that systemic administration of another M1R positive

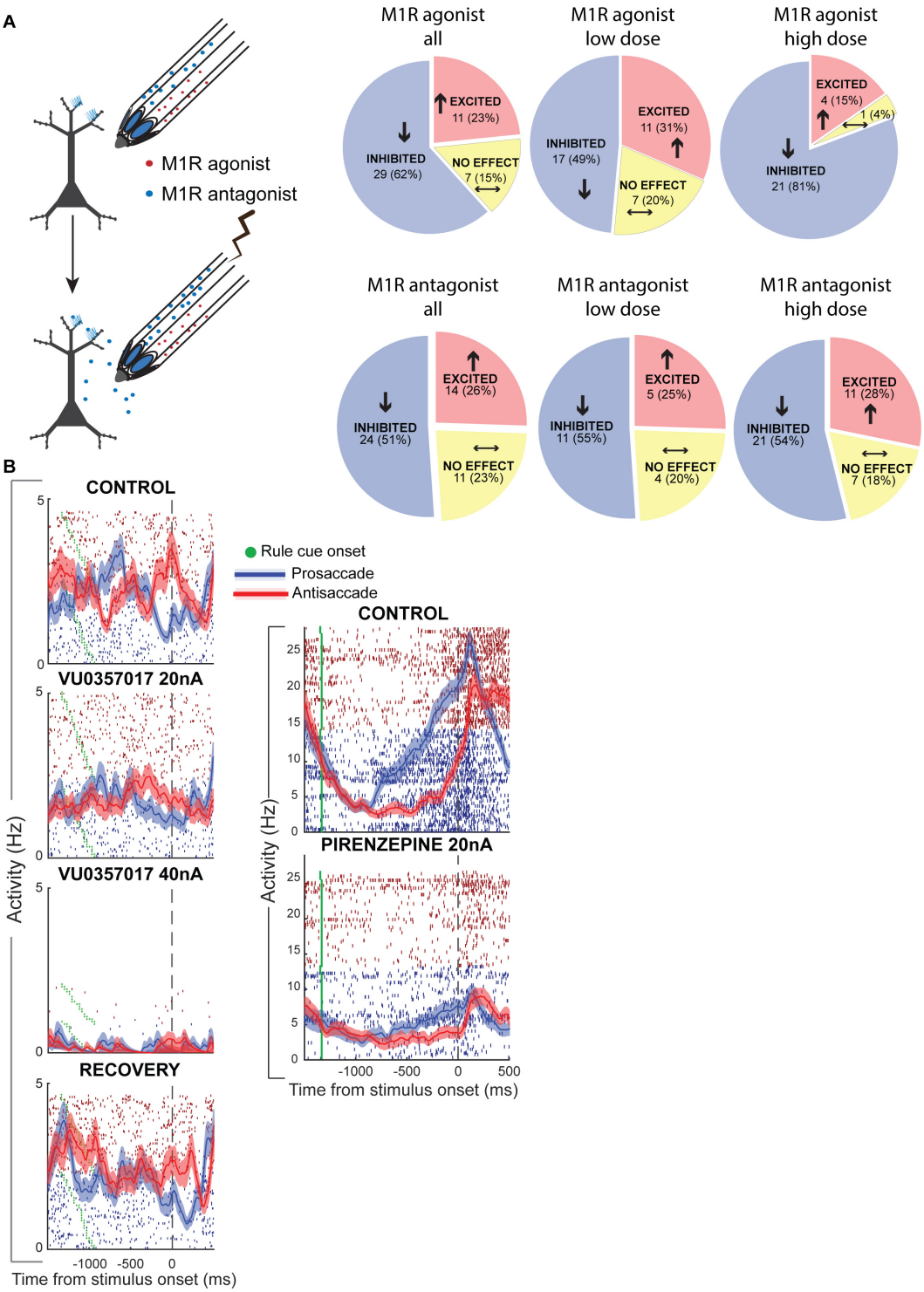


FIGURE 5 | Muscarinic M1R modulation of rule WM in monkey PFC. Adapted from Vijayraghavan et al. (2018). **(A)** Experimental design of iontophoresis and recording experiments shown on the left. Effects of M1R allosteric agonist, VU0357017 **(top panel)** and M1R antagonist pirenzepine on neuronal physiology in PFC. Pie-charts show number of neurons in the population that were significantly inhibited, excited or unaffected by drug application. Left-most panel shows the net drug effect on neurons tested at any (both low and high) doses of the M1R agonist. **Middle panel**, low doses; **Right panel**, High doses. **(B) Left panel** shows the effects of two doses of the M1R agonist on a PFC neuron with delay period activity selective for antisaccades over prosaccades. High dose of the M1R agonist strongly suppresses the neuron and disrupts rule selectivity in the delay period. Recovery shown in **bottom left panel**. **Right panel** shows the activity of a PFC neuron before and during application of M1R antagonist pirenzepine. This neuron showed ramping persistent activity during the delay period that was selective for prosaccades. M1R blockade inhibited this neuron and also diminished rule selectivity. Modified with permission from Elsevier (Neuron).

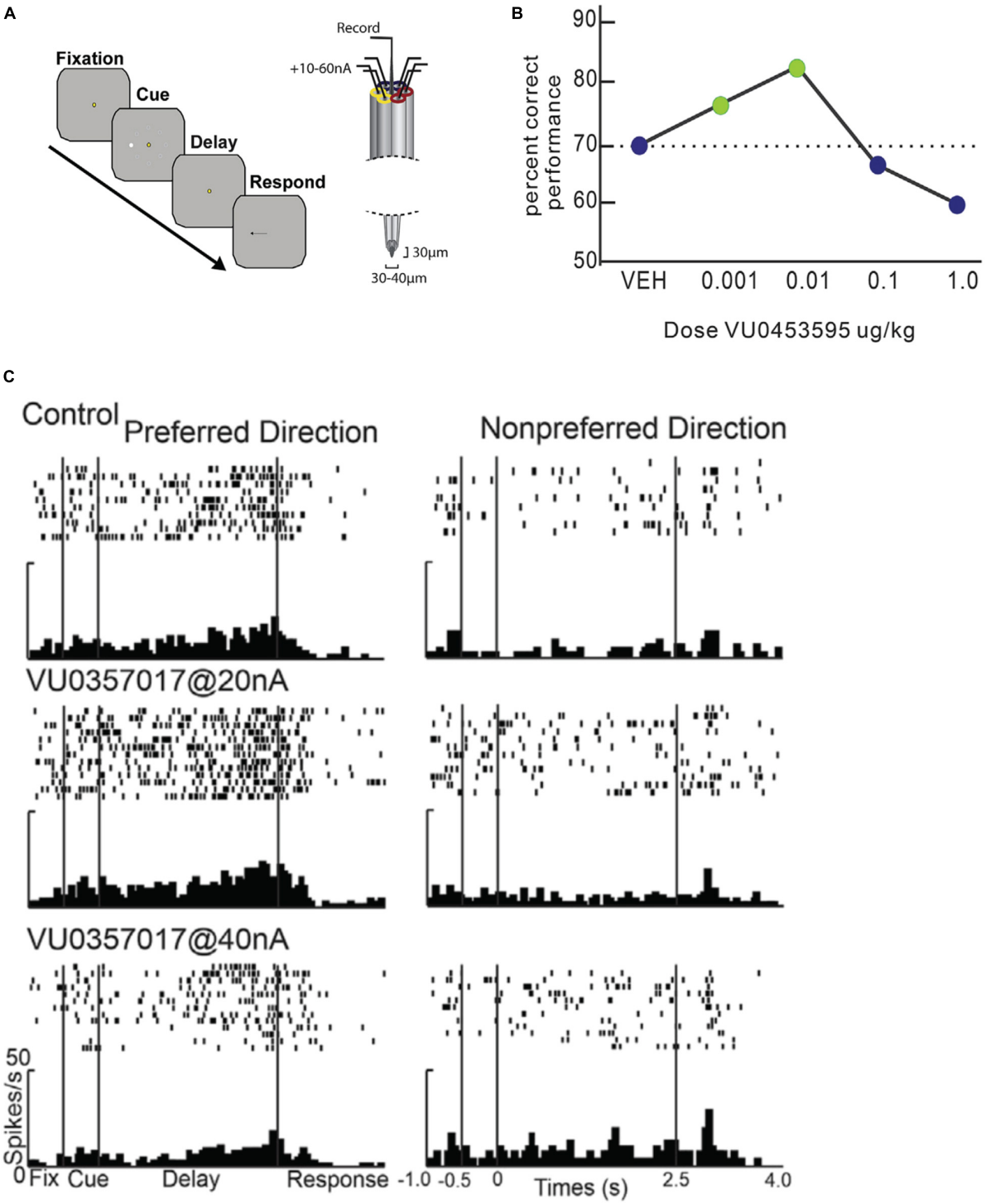


FIGURE 6 | Effects of M1R stimulation on spatial delayed response performance and delay period persistent activity. Adapted from Galvin et al., 2020b. **(A)** Schematic of trial structure of oculomotor delayed response and iontophoresis technique from Galvin et al. (2020b). After central fixation, a peripheral cue briefly flashes at one of eight locations. The cue location is maintained in WM during the delay period, when central fixation continues to be maintained. At the end of the delay indicated by fixation spot offset, the subject makes a saccade to the remembered location. **(B)** Behavioral dose response curves for systemic administration of M1R positive allosteric modulator VU0453595 during spatial delayed response performance in by an aged monkey. M1R stimulation has an inverted-U effect on WM performance. WM performance degrades at doses higher than the optimal dose. **(C)** Microiontophoresis of increasing doses of M1 allosteric agonist, VU0357017 on persistent spatially tuned delay period activity of a PFC neuron. **Left panel** shows rasters and histograms for neurons preferred direction. **Right panel** shows rasters and histograms for neurons non-preferred direction. M1R agonist application at low dose enhances WM activity, while higher dose application suppresses the neuron. Reproduced with permission from Elsevier (Neuron).

allosteric modulator in aged monkeys improved WM behavioral performance at low doses but disrupted performance at high doses (**Figure 6B**). They further reported that inhibiting the M-current could restore delay-related firing which had been suppressed by selective M1R antagonist, telenzepine.

The effects of different muscarinic actions on neuronal physiology in the PFC from the studies discussed above have been summarized in **Table 1**. These surprising results with M1R agonists point to the possibility that M1R overstimulation in primate PFC may trigger signaling mechanisms that lead to neuronal suppression. Thus, the actions of ACh in PFC in alert behaving primates may involve mechanisms that engender non-trivial suppression of cortical activity through M1Rs. Further the results in Vijayraghavan et al. (2018) suggest that the actions of ACh on M1R do not completely account for the suppressive effects of general muscarinic blockade on PFC neurons.

Several mechanisms may account for the suppression due to M1R overstimulation. One possibility is that M1R excitation of interneurons at high doses of stimulation leads to a suppression of PFC neurons. However, this is unlikely, as noted above because Vijayraghavan et al. (2018) reported that narrow-spiking putative parvalbumin positive interneurons were also equally suppressed by the agonist. This, of course, does not account for other classes of interneurons which are not narrow spiking, the increase in activity of which may well have caused suppression of the pyramidal neurons. Another possible mechanism for the inhibition may be SK potassium channel activation by intracellular Ca^{2+} mobilization due to M1R stimulation, as discussed elsewhere in this review (Gulledge and Stuart, 2005). It is noteworthy, that previous iontophoretic studies examining G_q protein-coupled receptors have found that stimulating these receptors has inhibitory effects on PFC neurons in primates and in some rodent studies. a1 adrenergic receptor stimulation suppresses delay period activity in a spatial delayed response task (Birnbaum et al., 2004) and G_q metabotropic glutamate receptor 1 was shown to increase inhibitory transmission in rat medial PFC, impairing decision-making (Sun and Neugebauer, 2011).

Galvin et al. (2020b) propose that suppression due to overstimulation of M1Rs could be the result of membrane hyperpolarization due to increase in the open state of KCNQ2 channels (Jentsch, 2000) due to M1R-mediated protein kinase C-cyclic AMP-protein kinase A signaling. Indeed, they show that retigabine, a positive allosteric modulator that preferentially targets KCNQ2 channels and increases the open state of the channels reduces persistent activity of PFC neurons. Future experiments must address the underlying mechanism involved in the suppression of persistent activity in the PFC by M1R overstimulation.

Vijayraghavan et al. (2018) also reported that M1R antagonist application suppressed roughly half of the PFC neurons tested even at the highest doses tested and did not systematically alter WM rule selectivity, in contrast with the uniform neuronal suppression and loss of task selectivity due to scopolamine (Major et al., 2015). This suggests that there are other muscarinic excitatory mechanisms independent of M1Rs active in the PFC. Vijayraghavan et al. (2018) proposed that there may be excitatory mechanisms based on M2R activation which may

TABLE 1 | Qualitative comparison of the physiological effects of local muscarinic receptor manipulation on PFC WM activity from various reports discussed in this review.

Study	Muscarinic manipulation	Species	Behavioral task	Proportion of neurons significantly suppressed or excited	Notes
Major et al. (2015)	General muscarinic blockade (Scopolamine)	*Macaca mulatta*	Rule WM pro- and antisaccade task	57% – – – 15% +++	General suppression, disruption of rule, stimulus location and saccade direction selectivity
Major et al. (2018)	Cholinergic stimulation (Carbachol)	*Macaca mulatta*	Rule WM pro- and antisaccade task	39% – – – 49% +++	Rule selectivity degradation due to non-specific increase in activity
Vijayraghavan et al. (2018)	M1R stimulation (VU0357017, McN-A-343)	*Macaca mulatta*	Rule WM pro- and antisaccade task	23% – – – 62% +++	81% neurons inhibited at higher doses, with disruption of rule selectivity
	M1R blockade (Pirenzepine)			51% – – – 26% +++	Suppression does not increase with dose, and population rule selectivity not affected.
Galvin et al. (2020b)	M1R stimulation (VU0357017, cevimeline)	*Macaca mulatta* (aged)	Oculomotor delayed response task	+++	Low doses augment PFC WM activity in aged macaques
	M1R blockade (Pirenzepine, telenzepine)			– – –	Pirenzepine and Telenzepine suppress PFC WM activity

explain why M1R blockade does not replicate the efficacy of general muscarinic blockade in neuronal suppression and task selectivity. As discussed in this review, M2R is present postsynaptically in both pyramidal neuron dendritic spines and in the dendrites of interneurons. M2Rs are $G_{i/o}$-coupled receptors, and previous microiontophoretic studies in monkey PFC have shown that stimulation of the dopamine D2 receptor, which is also coupled to $G_{i/o}$, can augment the activity of specific classes of PFC neurons during WM tasks (Wang et al., 2004; Vijayraghavan et al., 2016, 2017; recently reviewed by Ott and Nieder, 2019). Thus, in addition to their documented role in autoinhibition and heteroinhibition as presynaptic receptors (Murakoshi, 1995), post-synaptic M2R signaling may lead to increase in PFC neuronal excitability and augmentation of persistent activity. In support of this hypothesis, preliminary data from our group suggests that M2R antagonism suppresses the delay activity of PFC cells engaged in the rule-memory guided pro- and antisaccade task. Future studies with local application of M1R-selective allosteric antagonists, like VU0255035, and M2R-selective agonists and antagonists in PFC will help resolve these apparent paradoxes of muscarinic actions on PFC WM circuits.

These results with M1R compounds in monkey PFC are of particular interest, because M1R-selective agents are being actively investigated for cognitive enhancement and amelioration of cognitive deficits in neuropsychiatric disorders (Bubser et al., 2011; Thiele, 2013; Carruthers et al., 2015). M1R based therapeutics, such as KarXT, a coformulation of M1R agonist xanomeline and trospium, a peripheral muscarinic M2R antagonist that ameliorates non-target side effects of xanomeline, are showing promising results in clinical trials for the treatment of schizophrenia (Brannan et al., 2019).

CONCLUSION

In this review, we have discussed the neuromodulatory influence of the corticopetal cholinergic system through muscarinic receptors on primate PFC WM circuits that manifest persistent memory-related activity. The anatomical localization of these receptors shows exquisite specificity and correspondence with network connections within the PFC. Cortical muscarinic receptors play a pivotal role in arousal and brain state transitions, and their activation is necessary for the proper functioning of recurrent circuits in the PFC that generate persistent activity in WM tasks. Recent work shows that their role in primate PFC may be quite different from what would be expected from prior studies in other model systems like rodents and moreover, diverges from their role in sensory cortical areas. Further elucidation of muscarinic neuromodulation of PFC cognitive circuitry promises to be a rewarding endeavor for translational research and the development of new targets for the treatment of neuropsychiatric and neurological disorders.

AUTHOR CONTRIBUTIONS

SV and SE wrote and edited the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the CIHR grant FRN148365 to SE and the Canada First Research Excellence Fund to BrainsCAN.

REFERENCES

Aigner, T. G., and Mishkin, M. (1993). Scopolamine impairs recall of one-trial stimulus-reward association in monkeys. *Behav. Brain Res.* 54, 133–136. doi: 10.1016/0166-4328(93)90071-w

Aigner, T. G., Mitchell, S. J., Aggleton, J. P., DeLong, M. R., Struble, R. G., Price, D. L., et al. (1987). Effects of scopolamine and physostigmine on recognition memory in monkeys with ibotenic-acid lesions of the nucleus basalis of Meynert. *Psychopharmacology* 92, 292–300. doi: 10.1007/bf00210833

Aigner, T. G., Mitchell, S. J., Aggleton, J. P., DeLong, M. R., Struble, R. G., Price, D. L., et al. (1991). Transient impairment of recognition memory following ibotenic-acid lesions of the basal forebrain in macaques. *Exp. Brain Res.* 86, 18–26. doi: 10.1007/bf00231036

Anagnostaras, S. G., Murphy, G. G., Hamilton, S. E., Mitchell, S. L., Rahnama, N. P., Nathanson, N. M., et al. (2003). Selective cognitive dysfunction in acetylcholine M1 muscarinic receptor mutant mice. *Nat. Neurosci.* 6, 51–58. doi: 10.1038/nn992

Arnsten, A. F. T., Wang, M. J., and Paspalas, C. D. (2012). Neuromodulation of thought: flexibilities and vulnerabilities in prefrontal cortical network synapses. *Neuron* 76, 223–239. doi: 10.1016/j.neuron.2012.08.038

Aston-Jones, G., and Bloom, F. (1981). Activity of norepinephrine-containing locus coeruleus neurons in behaving rats anticipates fluctuations in the sleep-waking cycle. *J. Neurosci.* 1, 876–886. doi: 10.1523/jneurosci.01-08-00876.1981

Barak, S. (2009). Modeling cholinergic aspects of schizophrenia: focus on the antimuscarinic syndrome. *Behav. Brain Res.* 204, 335–351. doi: 10.1016/j.bbr.2009.04.006

Barak, S., and Weiner, I. (2006). Scopolamine induces disruption of latent inhibition which is prevented by antipsychotic drugs and an acetylcholinesterase inhibitor. *Neuropsychopharmacology* 32, 989–999. doi: 10.1038/sj.npp.1301208

Barak, S., and Weiner, I. (2009). Towards an animal model of an antipsychotic drug-resistant cognitive impairment in schizophrenia: scopolamine induces abnormally persistent latent inhibition, which can be reversed by cognitive enhancers but not by antipsychotic drugs. *Int. J. Neuropsychoph.* 12, 227–241. doi: 10.1017/s1461145708009176

Bartus, R. T., and Johnson, H. R. (1976). Short-term memory in the rhesus monkey: disruption from the anti-cholinergic scopolamine. *Pharmacol. Biochem. Be.* 5, 39–46. doi: 10.1016/0091-3057(76)90286-0

Baxter, M. G., and Crimins, J. L. (2018). Acetylcholine receptor stimulation for cognitive enhancement: better the devil you know? *Neuron* 98, 1064–1066. doi: 10.1016/j.neuron.2018.06.018

Birnbaum, S. G., Yuan, P. X., Wang, M., Vijayraghavan, S., Bloom, A. K., Davis, D. J., et al. (2004). Protein Kinase C overactivity impairs prefrontal cortical regulation of working memory. *Science* 306, 882–884. doi: 10.1126/science.1100021

Brannan, S., Miller, A., Felder, C., Paul, S., and Breier, A. (2019). T106. KARXT: A M1/M4 preferring muscarinic agonist for the treatment of schizophrenia. *Schizophrenia Bull.* 45, S244–S245. doi: 10.1093/schbul/sbz019.386

Brown, D. A. (2010). Muscarinic acetylcholine receptors (mAChRs) in the nervous system: some functions and mechanisms. *J. Mol. Neurosci.* 41, 340–346. doi: 10.1007/s12031-010-9377-2

Brown, D. A., and Adams, P. R. (1980). Muscarinic suppression of a novel voltage-sensitive K+ current in a vertebrate neurone. *Nature* 283, 673–676. doi: 10.1038/283673a0

Bubser, M., Byun, N., Wood, M. R., and Jones, C. K. (2011). Muscarinic receptor pharmacology and circuitry for the modulation of cognition. *Handb. Exp. Pharmacol.* 208, 121–166. doi: 10.1007/978-3-642-23274-9_7

Buccafusco, J. J., Terry, A. V., Webster, S. J., Martin, D., Hohnadel, E. J., Bouchard, K. A., et al. (2008). The scopolamine-reversal paradigm in rats and monkeys: the importance of computer-assisted operant-conditioning memory tasks for screening drug candidates. *Psychopharmacology* 199, 481–494. doi: 10.1007/s00213-007-0887-8

Carruthers, S. P., Gurvich, C. T., and Rossell, S. L. (2015). The muscarinic system, cognition and schizophrenia. *Neurosci. Biobehav. Rev.* 55, 393–402. doi: 10.1016/j.neubiorev.2015.05.011

Caulfield, M. P., and Birdsall, N. J. (1998). International Union of Pharmacology. XVII. Classification of muscarinic acetylcholine receptors. *Pharmacol. Rev.* 50, 279–290.

Chen, N., Sugihara, H., and Sur, M. (2015). An acetylcholine-activated microcircuit drives temporal dynamics of cortical activity. *Nat. Neurosci.* 18, 892–902. doi: 10.1038/nn.4002

Cole, A., and Nicoll, R. (1983). Acetylcholine mediates a slow synaptic potential in hippocampal pyramidal cells. *Science* 221, 1299–1301. doi: 10.1126/science.6612345

Condy, C., Wattiez, N., Rivaud-Péchoux, S., Tremblay, L., and Gaymard, B. (2007). Antisaccade deficit after inactivation of the principal sulcus in monkeys. *Cereb. Cortex* 17, 221–229. doi: 10.1093/cercor/bhj140

Constantinople, C. M., and Bruno, R. M. (2011). Effects and mechanisms of wakefulness on local cortical networks. *Neuron* 69, 1061–1068. doi: 10.1016/j.neuron.2011.02.040

Contreras, D., and Steriade, M. (1995). Cellular basis of EEG slow rhythms: a study of dynamic corticothalamic relationships. *J. Neurosci.* 15, 604–622. doi: 10.1523/jneurosci.15-01-00604.1995

Croxson, P. L., Kyriazis, D. A., and Baxter, M. G. (2011). Cholinergic modulation of a specific memory function of prefrontal cortex. *Nat. Neurosci.* 14, 1510–1512. doi: 10.1038/nn.2971

Dagytë, G., Boer, J. A. D., and Trentani, A. (2011). The cholinergic system and depression. *Behav. Brain Res.* 221, 574–582. doi: 10.1016/j.bbr.2010.02.023

Davidson, M., and Marrocco, R. (2000). Local infusion of scopolamine into intraparietal cortex slows covert orienting in rhesus monkeys. *J. Neurophysiol.* 83, 1536–1549.

Davidson, M. C., Cutrell, E. B., and Marrocco, R. T. (1999). Scopolamine slows the orienting of attention in primates to cued visual targets. *Psychopharmacology* 142, 1–8. doi: 10.1007/s002130050855

Davies, R. H., Scholes, H. E., Virdi, S., and Broadley, K. J. (2001). Inhibition of field stimulation-induced contractions of rabbit vas deferens by muscarinic receptor agonists: selectivity of McN-A-343 for M1 receptors. *J. Pharm. Pharmacol.* 53, 487–496.

Dean, B., Bymaster, F., and Scarr, E. (2003). Muscarinic receptors in schizophrenia. *Curr. Mol. Med.* 3, 419–426. doi: 10.2174/1566524033479654

Dean, B., McLeod, M., Keriakous, D., McKenzie, J., and Scarr, E. (2002). Decreased muscarinic1 receptors in the dorsolateral prefrontal cortex of subjects with schizophrenia. *Mol. Psychiatr.* 7, 1083–1091. doi: 10.1038/sj.mp.4001199

Dean, B., and Scarr, E. (2016). COMT genotype is associated with differential expression of muscarinic M1 receptors in human cortex. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 171, 784–789. doi: 10.1002/ajmg.b.32440

Digby, G. J., Noetzel, M. J., Bubser, M., Utley, T. J., Walker, A. G., Byun, N. E., et al. (2012). Novel allosteric agonists of M1 muscarinic acetylcholine receptors induce brain region-specific responses that correspond with behavioral effects in animal models. *J. Neurosci.* 32, 8532–8544. doi: 10.1523/JNEUROSCI.0337-12.2012

Disney, A. A., and Aoki, C. (2008). Muscarinic acetylcholine receptors in macaque V1 are most frequently expressed by parvalbumin-immunoreactive neurons. *J. Comp. Neurol.* 507, 1748–1762. doi: 10.1002/cne.21616

Disney, A. A., Domakonda, K. V., and Aoki, C. (2006). Differential expression of muscarinic acetylcholine receptors across excitatory and inhibitory cells in visual cortical areas V1 and V2 of the macaque monkey. *J. Comp. Neurol.* 499, 49–63. doi: 10.1002/cne.21096

Disney, A. A., and Reynolds, J. H. (2014). Expression of m1−type muscarinic acetylcholine receptors by parvalbumin−immunoreactive neurons in the primary visual cortex: a comparative study of rat, guinea pig, ferret, macaque, and human. *J. Comp. Neurol.* 522, 986–1003. doi: 10.1002/cne.23456

Dudar, J. D., and Szerb, J. C. (1969). The effect of topically applied atropine on resting and evoked cortical acetylcholine release. *J. Physiol.* 203, 741–762. doi: 10.1113/jphysiol.1969.sp008890

Dunnett, S. B., Wareham, A. T., and Torres, E. M. (1990). Cholinergic blockade in prefrontal cortex and hippocampus disrupts short-term memory in rats. *Neuroreport* 1, 61–64. doi: 10.1097/00001756-199009000-00017

Eckenstein, F., and Baughman, R. W. (1984). Two types of cholinergic innervation in cortex, one co-localized with vasoactive intestinal polypeptide. *Nature* 309, 153–155. doi: 10.1038/309153a0

Egorov, A. V., Hamam, B. N., Fransén, E., Hasselmo, M. E., and Alonso, A. A. (2002). Graded persistent activity in entorhinal cortex neurons. *Nature* 420, 173–178. doi: 10.1038/nature01171

Ellis, J. R., Ellis, K. A., Bartholomeusz, C. F., Harrison, B. J., Wesnes, K. A., Erskine, F. F., et al. (2006). Muscarinic and nicotinic receptors synergistically modulate working memory and attention in humans. *Int. J. Neuropsychopharmacol.* 9, 175–189. doi: 10.1017/S1461145705005407

Everling, S., and Fischer, B. (1998). The antisaccade: a review of basic research and clinical studies. *Neuropsychologia* 36, 885–899. doi: 10.1016/S0028-3932(98)00020-7

Fibiger, H. C. (1991). Cholinergic mechanisms in learning, memory and dementia: a review of recent evidence. *Trends Neurosci.* 14, 220–223. doi: 10.1016/0166-2236(91)90117-D

Flynn, D. D., Ferrari-DiLeo, G., Levey, A. I., and Mash, D. C. (1995). Differential alterations in muscarinic receptor subtypes in Alzheimer's disease: implications for cholinergic-based therapies. *Life Sci.* 56, 869–876. doi: 10.1016/0024-3205(95)00022-x

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

Fuster, J. M. (1992). Prefrontal neurons and the cognitive foundation of motor action. *Adv. Neurol.* 57, 351–360.

Fuster, J. M. (1993). Frontal lobes. *Curr. Opin. Neurobiol.* 3, 160–165. doi: 10.1016/0959-4388(93)90204-c

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

Galvin, V. C., Arnsten, A. F., and Wang, M. (2018). Evolution in neuromodulation—the differential roles of acetylcholine in higher order association vs. Primary visual cortices. *Front. Neural Circ.* 12:67. doi: 10.3389/fncir.2018.00067

Galvin, V. C., Arnsten, A. F. T., and Wang, M. (2020a). Behavioral pharmacology of the cholinergic system. *Curr. Top. Behav. Neurosci.* 45, 89–99. doi: 10.1007/7854_2020_142

Galvin, V. C., Yang, S. T., Paspalas, C. D., Yang, Y., Jin, L. E., Datta, D., et al. (2020b). Muscarinic M1 receptors modulate working memory performance and activity via KCNQ potassium channels in the primate prefrontal cortex. *Neuron* 106, 649–661e4. doi: 10.1016/j.neuron.2020.02.030

Giachetti, A., Giraldo, E., Ladinsky, H., and Montagna, E. (1986). Binding and functional profiles of the selective M1 muscarinic receptor antagonists trihexyphenidyl and dicyclomine. *Brit. J. Pharmacol.* 89, 83–90. doi: 10.1111/j.1476-5381.1986.tb11123.x

Gil, Z., Connors, B. W., and Amitai, Y. (1997). Differential regulation of neocortical synapses by neuromodulators and activity. *Neuron* 19, 679–686. doi: 10.1016/s0896-6273(00)80380-3

Gillin, J. C., Sutton, L., Ruiz, C., Golshan, S., Hirsch, S., Warmann, C., et al. (1991). Dose dependent inhibition of REM sleep in normal volunteers by biperiden, a muscarinic antagonist. *Biol. Psychiat.* 30, 151–156. doi: 10.1016/0006-3223(91)90169-m

Goldman-Rakic, P., and Schwartz, M. (1982). Interdigitation of contralateral and ipsilateral columnar projections to frontal association cortex in primates. *Science* 216, 755–757. doi: 10.1126/science.6177037

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

Green, A., Ellis, K. A., Ellis, J., Bartholomeusz, C. F., Ilic, S., Croft, R. J., et al. (2005). Muscarinic and nicotinic receptor modulation of object and spatial n-back working memory in humans. *Pharmacol. Biochem. Behav.* 81, 575–584. doi: 10.1016/j.pbb.2005.04.010

Gulledge, A. T., and Stuart, G. J. (2005). Cholinergic inhibition of neocortical pyramidal neurons. *J. Neurosci.* 25, 10308–10320. doi: 10.1523/JNEUROSCI.2697-05.2005

Hagenston, A. M., Fitzpatrick, J. S., and Yeckel, M. F. (2008). MGluR-mediated calcium waves that invade the soma regulate firing in layer V medial prefrontal cortical pyramidal neurons. *Cereb. Cortex* 18, 407–423. doi: 10.1093/cercor/bhm075

Hampel, H., Mesulam, M.-M., Cuello, A. C., Khachaturian, A. S., Vergallo, A., Farlow, M. R., et al. (2019). Revisiting the cholinergic hypothesis in Alzheimer's disease: emerging evidence from translational and clinical research. *J. Prev. Alzheimers Dis.* 6, 2–15. doi: 10.14283/jpad.2018.43

Hampel, H., Mesulam, M.-M., Cuello, C. A., Farlow, M. R., Giacobini, E., Grossberg, G. T., et al. (2018). The cholinergic system in the pathophysiology and treatment of Alzheimer's disease. *Brain* 141, 1917–1933. doi: 10.1093/brain/awy132

Hasselmo, M. E., and Bower, J. M. (1992). Cholinergic suppression specific to intrinsic not afferent fiber synapses in rat piriform (olfactory) cortex. *J. Neurophysiol.* 67, 1222–1229. doi: 10.1152/jn.1992.67.5.1222

Herrero, J. L., Roberts, M. J., Delicato, L. S., Gieselmann, M. A., Dayan, P., and Thiele, A. (2008). Acetylcholine contributes through muscarinic receptors to attentional modulation in V1. *Nature* 454, 1110–1114. doi: 10.1038/nature07141

Hicks, T. P. (1984). The history and development of microiontophoresis in experimental neurobiology. *Prog. Neurobiol.* 22, 185–240.

Hikosaka, O., and Wurtz, R. H. (1983). Visual and oculomotor functions of monkey substantia nigra pars reticulata. III. Memory-contingent visual and saccade responses. *J. Neurophysiol.* 49, 1268–1284. doi: 10.1152/jn.1983.49.5.1268

Hohagen, F., Riemann, D., Spiegel, R., Holzhauer, M., and Berger, M. (1993). Influence of the cholinergic agonist SDZ 210-086 on sleep in healthy subjects. *Neuropsychopharmacol* 9, 225–232. doi: 10.1038/npp.1993.58

Hsieh, C. Y., Cruikshank, S. J., and Metherate, R. (2000). Differential modulation of auditory thalamocortical and intracortical synaptic transmission by cholinergic agonist. *Brain Res.* 880, 51–64. doi: 10.1016/s0006-8993(00)02766-9

Jentsch, T. J. (2000). Neuronal KCNQ potassium channels: physiology and role in disease. *Nat. Rev. Neurosci.* 1, 21–30. doi: 10.1038/35036198

Jiang, S., Li, Y., Zhang, C., Zhao, Y., Bu, G., Xu, H., et al. (2014). M1 muscarinic acetylcholine receptor in Alzheimer's disease. *Neurosci. Bull.* 30, 295–307. doi: 10.1007/s12264-013-1406-z

Johnston, K., Levin, H. M., Koval, M. J., and Everling, S. (2007). Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. *Neuron* 53, 453–462. doi: 10.1016/j.neuron.2006.12.023

Jones, B. E. (2008). Modulation of cortical activation and behavioral arousal by cholinergic and orexinergic systems. *Ann. N. Y. Acad. Sci.* 1129, 26–34. doi: 10.1196/annals.1417.026

Jones, B. E. (2020). Arousal and sleep circuits. *Neuropsychopharmacology* 45, 6–20. doi: 10.1038/s41386-019-0444-2

Jones, C. K., Byun, N., and Bubser, M. (2012). Muscarinic and nicotinic acetylcholine receptor agonists and allosteric modulators for the treatment of schizophrenia. *Neuropsychopharmacology* 37, 16–42. doi: 10.1038/npp.2011.199

Jones, C. K., and Shannon, H. E. (2000). Muscarinic cholinergic modulation of prepulse inhibition of the acoustic startle reflex. *J. Pharmacol. Exp. Ther.* 294, 1017–1023.

Kim, E.-J., and Jeong, D.-U. (1999). Transdermal scopolamine alters phasic REM activity in normal young adults. *Sleep* 22, 515–520. doi: 10.1093/sleep/22.4.515

Kimura, F., and Baughman, R. W. (1997). Distinct muscarinic receptor subtypes suppress excitatory and inhibitory synaptic responses in cortical neurons. *J. Neurophysiol.* 77, 709–716. doi: 10.1152/jn.1997.77.2.709

Klinkenberg, I., and Blokland, A. (2010). The validity of scopolamine as a pharmacological model for cognitive impairment: a review of animal behavioral studies. *Neurosci. Biobehav. Rev.* 34, 1307–1350. doi: 10.1016/j.neubiorev.2010.04.001

Koval, M. J., Lomber, S. G., and Everling, S. (2011). Prefrontal cortex deactivation in macaques alters activity in the superior colliculus and impairs voluntary control of saccades. *J. Neurosci.* 31, 8659–8668. doi: 10.1523/JNEUROSCI.1258-11.2011

Krnjević, K., Reiffenstein, R. J., and Ropert, N. (1981). Disinhibitory action of acetylcholine in the rat's hippocampus: extracellular observations. *Neuroscience* 6, 2465–2474. doi: 10.1016/0306-4522(81)90092-0

Lebois, E. P., Bridges, T. M., Lewis, L. M., Dawson, E. S., Kane, A. S., Xiang, Z., et al. (2010). Discovery and characterization of novel subtype-selective allosteric agonists for the investigation of M(1) receptor function in the central nervous system. *ACS Chem. Neurosci.* 1, 104–121. doi: 10.1021/cn900003h

Lee, S.-H., and Dan, Y. (2012). Neuromodulation of brain states. *Neuron* 76, 209–222. doi: 10.1016/j.neuron.2012.09.012

Levey, A. I. (1996). Muscarinic acetylcholine receptor expression in memory circuits: implications for treatment of Alzheimer?disease. *Proc. Natl. Acad. Sci. U.S.A.* 93, 13541–13546. doi: 10.1073/pnas.93.24.13541

Levey, A. I., Edmunds, S. M., Heilman, C. J., Desmond, T. J., and Frey, K. A. (1994). Localization of muscarinic M3 receptor protein and M3 receptor binding in rat brain. *Neuroscience* 63, 207–221. doi: 10.1016/0306-4522(94)90017-5

Levey, A. I., Kitt, C. A., Simonds, W. F., Price, D. L., and Brann, M. R. (1991). Identification and localization of muscarinic acetylcholine receptor proteins in brain with subtype-specific antibodies. *J. Neurosci.* 11, 3218–3226.

Lewis, D. A. (1991). Distribution of choline acetyltransferase-immunoreactive axons in monkey frontal cortex. *Neuroscience* 40, 363–374. doi: 10.1016/0306-4522(91)90126-9

Lewis, L. D., Weiner, V. S., Mukamel, E. A., Donoghue, J. A., Eskandar, E. N., Madsen, J. R., et al. (2012). Rapid fragmentation of neuronal networks at the onset of propofol-induced unconsciousness. *Proc. Natl. Acad. Sci. U.S.A.* 109, E3377–E3386. doi: 10.1073/pnas.1210907109

Lidow, M. S., Gallager, D. W., Rakic, P., and Goldman−Rakic, P. S. (1989). Regional differences in the distribution of muscarinic cholinergic receptors in the macaque cerebral cortex. *J. Comp. Neurol.* 289, 247–259. doi: 10.1002/cne.902890206

Liu, R., Crawford, J., Callahan, P. M., Terry, A. V., Constantinidis, C., and Blake, D. T. (2017). Intermittent stimulation of the nucleus basalis of meynert improves working memory in adult monkeys. *Curr. Biol.* 27, 2640–2646.e4. doi: 10.1016/j.cub.2017.07.021

Lv, X., Dickerson, J. W., Rook, J. M., Lindsley, C. W., Conn, P. J., and Xiang, Z. (2017). M1 muscarinic activation induces long-lasting increase in intrinsic excitability of striatal projection neurons. *Neuropharmacology* 118, 209–222. doi: 10.1016/j.neuropharm.2017.03.017

Maeda, S., Qu, Q., Robertson, M. J., Skiniotis, G., and Kobilka, B. K. (2019). Structures of the M1 and M2 muscarinic acetylcholine receptor/G-protein complexes. *Science* 364, 552–557. doi: 10.1126/science.aaw5188

Major, A. J., Vijayraghavan, S., and Everling, S. (2015). Muscarinic attenuation of mnemonic rule representation in macaque dorsolateral prefrontal cortex during a pro- and anti-saccade task. *J. Neurosci.* 35, 16064–16076. doi: 10.1523/jneurosci.2454-15.2015

Major, A. J., Vijayraghavan, S., and Everling, S. (2018). Cholinergic overstimulation attenuates rule selectivity in macaque prefrontal cortex. *J. Neurosci.* 38, 1137–1150. doi: 10.1523/jneurosci.3198-17.2017

Marrion, N. V. (1997). Control of M-current. *Annu. Rev. Physiol.* 59, 483–504. doi: 10.1146/annurev.physiol.59.1.483

Marucci, G., Buccioni, M., Ben, D. D., Lambertucci, C., Volpini, R., and Amenta, F. (2020). Efficacy of acetylcholinesterase inhibitors in Alzheimer's disease. *Neuropharmacology* doi: 10.1016/j.neuropharm.2020.108352 [Epub ahead of print],

Mash, D., Flynn, D., and Potter, L. (1985). Loss of M2 muscarine receptors in the cerebral cortex in Alzheimer's disease and experimental cholinergic denervation. *Science* 228, 1115–1117. doi: 10.1126/science.3992249

Mash, D. C., and Potter, L. T. (1986). Autoradiographic localization of M1 and M2 muscarine receptors in the rat brain. *Neuroscience* 19, 551–564. doi: 10.1016/0306-4522(86)90280-0

McCormick, D. A. (1989). Cholinergic and noradrenergic modulation of thalamocortical processing. *Trends Neurosci.* 12, 215–221. doi: 10.1016/0166-2236(89)90125-2

McCormick, D. A. (1993). Chapter 36: actions of acetylcholine in the cerebral cortex and thalamus and implications for function. *Prog. Brain Res.* 98, 303–308. doi: 10.1016/s0079-6123(08)62412-7

McCormick, D. A., and Prince, D. A. (1986). Mechanisms of action of acetylcholine in the guinea−pig cerebral cortex in vitro. *J. Physiol.* 375, 169–194. doi: 10.1113/jphysiol.1986.sp016112

McCormick, D. A., Wang, Z., and Huguenard, J. (1993). Neurotransmitter control of neocortical neuronal activity and excitability. *Cereb. Cortex* 3, 387–398.

Medalla, M., and Barbas, H. (2012). The anterior cingulate cortex may enhance inhibition of lateral prefrontal cortex via m2 cholinergic receptors at dual synaptic sites. *J. Neurosci.* 32, 15611–15625. doi: 10.1523/jneurosci.2339-12.2012

Mesulam, M.-M., Mufson, E. J., Wainer, B. H., and Levey, A. I. (1983). Central cholinergic pathways in the rat: an overview based on an alternative nomenclature (Ch1–Ch6). *Neuroscience* 10, 1185–1201. doi: 10.1016/0306-4522(83)90108-2

Metherate, R., Cox, C., and Ashe, J. (1992). Cellular bases of neocortical activation: modulation of neural oscillations by the nucleus basalis and endogenous acetylcholine. *J. Neurosci.* 12, 4701–4711. doi: 10.1523/jneurosci.12-12-04701.1992

Miller, E., and Cohen, J. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167

Miller, E., and Desimone, R. (1992). Scopolamine affects short-term memory but not inferior temporal neurons. *Neuroreport* 4, 81–84.

Mitchelson, F. J. (2012). The pharmacology of McN-A-343. *Pharmacol. Ther.* 135, 216–245. doi: 10.1016/j.pharmthera.2012.05.008

Moruzzi, G., and Magoun, H. W. (1949). Brain stem reticular formation and activation of the EEG. *Electroencephalogr. Clin. Neurophysiol.* 1, 455–473. doi: 10.1016/0013-4694(49)90219-9

Mrzljak, L., Levey, A. I, Belcher, S., and Goldman−Rakic, P. S. (1998). Localization of the m2 muscarinic acetylcholine receptor protein and mRNA in cortical neurons of the normal and cholinergically deafferented rhesus monkey. *J. Comp. Neurol.* 390, 112–132.

Mrzljak, L., Levey, A. I., and Goldman-Rakic, P. S. (1993). Association of m1 and m2 muscarinic receptor proteins with asymmetric synapses in the primate cerebral cortex: morphological evidence for cholinergic modulation of excitatory neurotransmission. *Proc. Natl. Acad. Sci. U.S.A.* 90, 5194–5198. doi: 10.1073/pnas.90.11.5194

Mrzljak, L., Levey, A. I., and Rakic, P. (1996). Selective expression of m2 muscarinic receptor in the parvocellular channel of the primate visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 93, 7337–7340. doi: 10.1073/pnas.93.14.7337

Mrzljak, L., Pappy, M., Leranth, C., and Goldman-Rakic, P. (1995). Cholinergic synaptic circuitry in the macaque prefrontal cortex. *J. Comp. Neurol.* 357, 603–617. doi: 10.1002/cne.903570409

Murakoshi, T. (1995). Cholinergic modulation of synaptic transmission in the rat visual cortex in vitro. *Vis. Res.* 35, 25–35. doi: 10.1016/0042-6989(94)e0056-q

Muzur, A., Pace-Schott, E. F., and Hobson, J. A. (2002). The prefrontal cortex in sleep. *Trends Cogn. Sci.* 6, 475–481. doi: 10.1016/s1364-6613(02)01992-7

Navarria, A., Wohleb, E. S., Voleti, B., Ota, K. T., Dutheil, S., Lepack, A. E., et al. (2015). Rapid antidepressant actions of scopolamine: Role of medial prefrontal cortex and M1-subtype muscarinic acetylcholine receptors. *Neurobiol. Dis.* 82, 254–261. doi: 10.1016/j.nbd.2015.06.012

Niwa, Y., Kanda, G. N., Yamada, R. G., Shi, S., Sunagawa, G. A., Ukai-Tadenuma, M., et al. (2018). Muscarinic acetylcholine receptors Chrm1 and Chrm3 are essential for REM sleep. *Cell Rep.* 24, 2231–2247.e7. doi: 10.1016/j.celrep.2018.07.082

Noufi, P., Khoury, R., Jeyakumar, S., and Grossberg, G. T. (2019). Use of cholinesterase inhibitors in non-Alzheimer's dementias. *Drug Aging* 36, 719–731. doi: 10.1007/s40266-019-00685-6

Ott, T., and Nieder, A. (2019). Dopamine and cognitive control in prefrontal cortex. *Trends Cogn. Sci.* 23, 213–234. doi: 10.1016/j.tics.2018.12.006

Parikh, V., Kozak, R., Martinez, V., and Sarter, M. (2007). Prefrontal acetylcholine release controls cue detection on multiple timescales. *Neuron* 56, 141–154. doi: 10.1016/j.neuron.2007.08.025

Parikh, V., Pomerleau, F., Huettl, P., Gerhardt, G. A., Sarter, M., and Bruno, J. P. (2004). Rapid assessment of in vivo cholinergic transmission by amperometric detection of changes in extracellular choline levels. *Eur. J. Neurosci.* 20, 1545–1554. doi: 10.1111/j.1460-9568.2004.03614.x

Penetar, D. M., and McDonough, J. H. (1983). Effects of cholinergic drugs on delayed match-to-sample performance of rhesus monkeys. *Pharmacol. Biochem. Behav.* 19, 963–967. doi: 10.1016/0091-3057(83)90399-4

Picciotto, M. R., Higley, M. J., and Mineur, Y. S. (2012). Acetylcholine as a neuromodulator: cholinergic signaling shapes nervous system function and behavior. *Neuron* 76, 116–129. doi: 10.1016/j.neuron.2012.08.036

Plakke, B., Ng, C.-W., and Poremba, A. (2008). Scopolamine impairs auditory delayed matching-to-sample performance in monkeys. *Neurosci. Lett.* 438, 126–130. doi: 10.1016/j.neulet.2008.04.015

Popiolek, M., Nguyen, D. P., Reinhart, V., Edgerton, J. R., Harms, J., Lotarski, S. M., et al. (2016). Inositol phosphate accumulation in vivo provides a measure of muscarinic M1 receptor activation. *Biochemistry* 55, 7073–7085. doi: 10.1021/acs.biochem.6b00688

Pucak, M., Levitt, J., Lund, J., and Lewis, D. (1996). Patterns of intrinsic and associational circuitry in monkey prefrontal cortex. *J. Comp. Neurol.* 376, 614–630. doi: 10.1002/(SICI)1096-9861(19961223)376

Richardson, R., and DeLong, M. (1986). Nucleus basalis of Meynert neuronal activity during a delayed response task in monkey. *Brain Res.* 399, 364–368. doi: 10.1016/0006-8993(86)91529-5

Richardson, R. T., and DeLong, M. R. (1991). The Basal forebrain. *Anat. Funct. Adv. Exp. Med. Biol.* 295, 233–252. doi: 10.1007/978-1-4757-0145-6_12

Robbins, T., and Arnsten, A. (2009). The neuropsychopharmacology of fronto-executive function: monoaminergic modulation. *Annu. Rev. Neurosci.* 32, 267–287. doi: 10.1146/annurev.neuro.051508.135535

Ruivo, L. M. T.-G., Baker, K. L., Conway, M. W., Kinsley, P. J., Gilmour, G., Phillips, K. G., et al. (2017). Coordinated acetylcholine release in prefrontal cortex and hippocampus is associated with arousal and reward on distinct timescales. *Cell Rep.* 18, 905–917. doi: 10.1016/j.celrep.2016.12.085

Rupniak, N., Samson, N., Tye, S., Field, M., and Iversen, S. (1991). Evidence against a specific effect of cholinergic drugs on spatial memory in primates. *Behav. Brain Res.* 43, 1–6. doi: 10.1016/S0166-4328(05)80047-6

Rusted, J. M., Eaton-Williams, P., and Warburton, D. M. (1991). A comparison of the effects of scopolamine and diazepam on working memory. *Psychopharmacology* 105, 442–445. doi: 10.1007/bf02244443

Rusted, J. M., and Warburton, D. M. (1988). The effects of scopolamine on working memory in healthy young volunteers. *Psychopharmacology* 96, 145–152. doi: 10.1007/bf00177553

Saar, D., Grossman, Y., and Barkai, E. (2001). Long-lasting cholinergic modulation underlies rule learning in rats. *J. Neurosci.* 21, 1385–1392. doi: 10.1523/jneurosci.21-04-01385.2001

Sanchez-Vives, M. V., Massimini, M., and Mattia, M. (2017). Shaping the default activity pattern of the cortical network. *Neuron* 94, 993–1001. doi: 10.1016/j.neuron.2017.05.015

Sanchez-Vives, M. V., and McCormick, D. A. (2000). Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nat. Neurosci.* 3, 1027–1034. doi: 10.1038/79848

Sarter, M., and Bruno, J. P. (1997). Cognitive functions of cortical acetylcholine: toward a unifying hypothesis. *Brain Res. Rev.* 23, 28–46. doi: 10.1016/s0165-0173(96)00009-4

Sarter, M., and Bruno, J. P. (1998). Cortical acetylcholine, reality distortion, schizophrenia, and lewy body dementia: too much or too little cortical acetylcholine? *Brain Cogn.* 38, 297–316. doi: 10.1006/brcg.1998.1035

Scarr, E., Cowie, T. F., Kanellakis, S., Sundram, S., Pantelis, C., and Dean, B. (2009). Decreased cortical muscarinic receptors define a subgroup of subjects with schizophrenia. *Mol. Psychiatry* 14, 1017–1023. doi: 10.1038/mp.2008.28

Schwartz, M. L., and Goldman−Rakic, P. S. (1984). Callosal and intrahemispheric connectivity of the prefrontal association cortex in rhesus monkey: relation between intraparietal and principal sulcal cortex. *J. Comp. Neurol.* 226, 403–420. doi: 10.1002/cne.902260309

Shekhar, A., Potter, W. Z., Lightfoot, J., Lienemann, J., Dub, S., Mallinckrodt, C., et al. (2008). Selective muscarinic receptor agonist xanomeline as a novel treatment approach for schizophrenia. *Am. J. Psychiatr.* 165, 1033–1039. doi: 10.1176/appi.ajp.2008.06091591

Shirey, J. K., Brady, A. E., Jones, P. J., Davis, A. A., Bridges, T. M., Kennedy, J. P., et al. (2009). A selective allosteric potentiator of the M1 muscarinic acetylcholine receptor increases activity of medial prefrontal cortical neurons and restores impairments in reversal learning. *J. Neurosci.* 29, 14271–14286. doi: 10.1523/JNEUROSCI.3930-09.2009

Sitaram, N., Wyatt, R., Dawson, S., and Gillin, J. (1976). REM sleep induction by physostigmine infusion during sleep. *Science* 191, 1281–1283. doi: 10.1126/science.176724

Skoblenick, K., and Everling, S. (2012). NMDA antagonist ketamine reduces task selectivity in macaque dorsolateral prefrontal neurons and impairs performance

of randomly interleaved prosaccades and antisaccades. *J. Neurosci.* 32, 12018–12027. doi: 10.1523/JNEUROSCI.1510-12.2012

Smiley, J. F., Morrell, F., and Mesulam, M. M. (1997). Cholinergic synapses in human cerebral cortex: an ultrastructural study in serial sections. *Exp. Neurol.* 144, 361–368. doi: 10.1006/exnr.1997.6413

Spinelli, S., Ballard, T., Feldon, J., Higgins, G., and Pryce, C. (2006). Enhancing effects of nicotine and impairing effects of scopolamine on distinct aspects of performance in computerized attention and working memory tasks in marmoset monkeys. *Neuropharmacology* 51, 238–250. doi: 10.1016/j.neuropharm.2006.03.012

Steriade, M. (2004). Acetylcholine systems and rhythmic activities during the waking–sleep cycle. *Prog. Brain Res.* 145, 179–196. doi: 10.1016/s0079-6123(03)45013-9

Steriade, M., Timofeev, I., and Grenier, F. (2001). Natural waking and sleep states: a view from inside neocortical neurons. *J. Neurophysiol.* 85, 1969–1985. doi: 10.1152/jn.2001.85.5.1969

Suh, B., and Hille, B. (2007). Regulation of KCNQ channels by manipulation of phosphoinositides. *J. Physiol.* 582, 911–916. doi: 10.1113/jphysiol.2007.132647

Suh, B.-C., and Hille, B. (2002). Recovery from muscarinic modulation of M current channels requires Phosphatidylinositol 4,5-Bisphosphate synthesis. *Neuron* 35, 507–520. doi: 10.1016/s0896-6273(02)00790-0

Suh, B.-C., and Hille, B. (2005). Regulation of ion channels by phosphatidylinositol 4,5-bisphosphate. *Curr. Opin. Neurobiol.* 15, 370–378. doi: 10.1016/j.conb.2005.05.005

Suh, B.-C., Inoue, T., Meyer, T., and Hille, B. (2006). Rapid chemically induced changes of PtdIns(4,5)P2 Gate KCNQ ion channels. *Science* 314, 1454–1457. doi: 10.1126/science.1131163

Sun, H., and Neugebauer, V. (2011). mGluR1, but not mGluR5, activates feed-forward inhibition in the medial prefrontal cortex to impair decision making. *J. Neurophysiol.* 106, 960–973. doi: 10.1152/jn.00762.2010

Sun, Y., Yang, Y., Galvin, V. C., Yang, S., Arnsten, A. F., and Wang, M. (2017). Nicotinic α4β2 cholinergic receptor influences on dorsolateral prefrontal cortical neuronal firing during a working memory task. *J. Neurosci.* 37, 5366–5377. doi: 10.1523/JNEUROSCI.0364-17.2017

Thiele, A. (2013). Muscarinic signaling in the brain. *Annu. Rev. Neurosci.* 36, 271–294. doi: 10.1146/annurev-neuro-062012-170433

Ukai, M., Okuda, A., and Mamiya, T. (2004). Effects of anticholinergic drugs selective for muscarinic receptor subtypes on prepulse inhibition in mice. *Eur. J. Pharmacol.* 492, 183–187. doi: 10.1016/j.ejphar.2004.03.066

Uslaner, J. M., Eddins, D., Puri, V., Cannon, C. E., Sutcliffe, J., Chew, C. S., et al. (2013). The muscarinic M1 receptor positive allosteric modulator PQCA improves cognitive measures in rat, cynomolgus macaque, and rhesus macaque. *Psychopharmacology* 225, 21–30. doi: 10.1007/s00213-012-2788-8

Velazquez-Moctezuma, J., Shalauta, M. D., Gillin, J. C., and Shiromani, P. J. (1990). Differential effects of cholinergic antagonists on REM sleep components. *Psychopharmacol. Bull.* 26, 349–353.

Venkatesan, S., Jeoung, H.-S., Chen, T., Power, S. K., Liu, Y., and Lambe, E. K. (2020). Behavioral pharmacology of the cholinergic system. *Curr. Top. Behav. Neurosci.* 45, 47–69. doi: 10.1007/7854_2020_138

Vijayraghavan, S., Major, A., and Everling, S. (2018). Muscarinic M1 receptor overstimulation disrupts working memory activity for rules in primate prefrontal cortex. *Neuron* 98, 1256–1268.e4. doi: 10.1016/j.neuron.2018.05.027

Vijayraghavan, S., Major, A. J., and Everling, S. (2016). Dopamine D1 and D2 receptors make dissociable contributions to dorsolateral prefrontal cortical regulation of rule-guided oculomotor behavior. *Cell Rep.* 16, 805–816. doi: 10.1016/j.celrep.2016.06.031

Vijayraghavan, S., Major, A. J., and Everling, S. (2017). Neuromodulation of prefrontal cortex in non-human primates by dopaminergic receptors during rule-guided flexible behavior and cognitive control. *Front. Neural Circ.* 11:91. doi: 10.3389/fncir.2017.00091

Voytko, M., Olton, D., Richardson, R., Gorman, L., Tobin, J., and Price, D. (1994). Basal forebrain lesions in monkeys disrupt attention but not learning and memory [published erratum appears in J Neurosci 1995 Mar;15(3): following table of contents]. *J. Neurosci.* 14, 167–186. doi: 10.1523/jneurosci.14-01-00167.1994

Wallace, T. L., and Bertrand, D. (2013). Importance of the nicotinic acetylcholine receptor system in the prefrontal cortex. *Biochem. Pharmacol.* 85, 1713–1720. doi: 10.1016/j.bcp.2013.04.001

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

Wang, M., Vijayraghavan, S., and Goldman-Rakic, P. S. (2004). Selective D2 receptor actions on the functional circuitry of working memory. *Science* 303, 853–856. doi: 10.1126/science.1091162

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

Yang, Y., Paspalas, C. D., Jin, L. E., Picciotto, M. R., Arnsten, A. F. T., and Wang, M. (2013). Nicotinic α7 receptors enhance NMDA cognitive circuits in dorsolateral prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12078–12083. doi: 10.1073/pnas.1307849110

Yeomans, J. (1995). Role of tegmental cholinergic neurons in dopaminergic activation, antimuscarinic psychosis and schizophrenia. *Neuropsychopharmacology* 12, 3–16.

Yi, F., Ball, J., Stoll, K. E., Satpute, V. C., Mitchell, S. M., Pauli, J. L., et al. (2014). Direct excitation of parvalbumin-positive interneurons by M1 muscarinic acetylcholine receptors: roles in cellular excitability, inhibitory transmission and cognition. *J. Physiol.* 592, 3463–3494. doi: 10.1113/jphysiol.2014.275453

Young, M. B., and Thomas, S. A. (2014). M1-muscarinic receptors promote fear memory consolidation via phospholipase C and the M-current. *J. Neurosci.* 34, 1570–1578. doi: 10.1523/JNEUROSCI.1040-13.2014

Zhang, W., Basile, A. S., Gomeza, J., Volpicelli, L. A., Levey, A. I., and Wess, J. (2002). Characterization of central inhibitory muscarinic autoreceptors by the use of muscarinic acetylcholine receptor knock-out mice. *J. Neurosci.* 22, 1709–1717. doi: 10.1523/jneurosci.22-05-01709.2002

Zhou, X., Qi, X.-L. L., Douglas, K., Palaninathan, K., Kang, H. S., Buccafusco, J. J., et al. (2011). Cholinergic modulation of working memory activity in primate prefrontal cortex. *J. Neurophysiol.* 106, 2180–2188. doi: 10.1152/jn.00148.2011

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neural Circuits

ORIGINAL RESEARCH
published: 21 April 2021
doi: 10.3389/fncir.2021.647944

Activity Stabilization in a Population Model of Working Memory by Sinusoidal and Noisy Inputs

Nikita Novikov[1], Denis Zakharov[1]*, Victoria Moiseeva[1] and Boris Gutkin[1,2]**

[1] Centre for Cognition and Decision Making, HSE University, Moscow, Russia, [2] Group for Neural Theory, LNC2 INSERM U960, Département d'Études Cognitives, École Normale Supérieure, PSL Research Université, Paris, France

According to mechanistic theories of working memory (WM), information is retained as stimulus-dependent persistent spiking activity of cortical neural networks. Yet, how this activity is related to changes in the oscillatory profile observed during WM tasks remains a largely open issue. We explore joint effects of input gamma-band oscillations and noise on the dynamics of several firing rate models of WM. The considered models have a metastable active regime, i.e., they demonstrate long-lasting transient post-stimulus firing rate elevation. We start from a single excitatory-inhibitory circuit and demonstrate that either gamma-band or noise input could stabilize the active regime, thus supporting WM retention. We then consider a system of two circuits with excitatory intercoupling. We find that fast coupling allows for better stabilization by common noise compared to independent noise and stronger amplification of this effect by in-phase gamma inputs compared to anti-phase inputs. Finally, we consider a multi-circuit system comprised of two clusters, each containing a group of circuits receiving a common noise input and a group of circuits receiving independent noise. Each cluster is associated with its own local gamma generator, so all its circuits receive gamma-band input in the same phase. We find that gamma-band input differentially stabilizes the activity of the "common-noise" groups compared to the "independent-noise" groups. If the inter-cluster connections are fast, this effect is more pronounced when the gamma-band input is delivered to the clusters in the same phase rather than in the anti-phase. Assuming that the common noise comes from a large-scale distributed WM representation, our results demonstrate that local gamma oscillations can stabilize the activity of the corresponding parts of this representation, with stronger effect for fast long-range connections and synchronized gamma oscillations.

Keywords: gamma oscillations, noise, in-phase oscillations, anti-phase oscillations, AMPA, NMDA, working memory

OPEN ACCESS

Edited by:
Emmanuel Procyk, Institut National de la Santé et de la Recherche Médicale (INSERM), France

Reviewed by:
Axel Hutt, Inria Nancy – Grand-Est Research Centre, France
Alex Roxin, Centre de Recerca Matemàtica, Spain

****Correspondence:***
Nikita Novikov
nikknovikov@gmail.com
Denis Zakharov
dgzakharov@hse.ru; dzakh76@gmail.com
Boris Gutkin
boris.gutkin@ens.fr

Received: *30 December 2020*
Accepted: *19 March 2021*
Published: *21 April 2021*

Citation:
Novikov N, Zakharov D, Moiseeva V and Gutkin B (2021) Activity Stabilization in a Population Model of Working Memory by Sinusoidal and Noisy Inputs. Front. Neural Circuits 15:647944. doi: 10.3389/fncir.2021.647944

INTRODUCTION

The concept of working memory (WM) characterizes the ability of the brain to retain in an active form certain information that is relevant to a current task, but is not perceived at the particular moment by sensory systems (Baddeley, 2003). One of the main mechanisms supposedly underlying WM is self-sustained activity of neural populations (Goldman-Rakic, 1995; Compte, 2006), which

can be turned on or off in a short period of time (about 100 ms) and continue for a time interval of up to tens of seconds (Wang, 2001). Cells whose activity is maintained at an elevated level during the retention of information in WM have been found in various parts of the brain, primarily in the prefrontal cortex (Fuster and Alexander, 1971; Funahashi et al., 1989; Miller et al., 1996; Chafee and Goldman-Rakic, 1998; Constantinidis and Goldman-Rakic, 2002).

The process of retaining information in WM is also associated with changes in the collective rhythmic activity of brain networks (which are neuronal oscillations) in various frequency bands (Sauseng et al., 2009; Siegel et al., 2009; Haegens et al., 2010; Liebe et al., 2012; Kornblith et al., 2016; Lundqvist et al., 2016, 2018; Wimmer et al., 2016). Among these changes, increase of the gamma-band activity is of special interest. Gamma activity usually reflects activation of neural populations and coincides with episodes of firing rate elevation. In WM tasks, gamma activity increases most strongly during presentation of stimuli (Kornblith et al., 2016; Lundqvist et al., 2016; Wimmer et al., 2016), but in the delay period (i.e., in the time period after stimulus termination and before an instruction to make a response) it is still higher than in the baseline (Lutzenberger et al., 2002; Kaiser et al., 2003; Jokisch and Jensen, 2007; Haegens et al., 2010; Palva et al., 2011; Kornblith et al., 2016; Lundqvist et al., 2016; Wimmer et al., 2016). We note that during this delay period, the subject has to withhold the response, yet needs to retain "online" information necessary to generate the appropriate response. The delay-period gamma activity presumably reflects activation of the neural populations that represent the WM content (Roux and Uhlhaas, 2014). This is supported by the findings that the delay-period gamma activity is higher than in the passive observation task (Wimmer et al., 2016), increases with WM load (Howard et al., 2003; van Vugt et al., 2010; Kornblith et al., 2016; Lundqvist et al., 2016) and only in the task-relevant regions (Kaiser et al., 2003; Jokisch and Jensen, 2007) or at those cortical sites that contain neurons selective to the WM content (Kornblith et al., 2016; Lundqvist et al., 2016).

Recently, the delay-period gamma oscillations were more directly linked to activation of WM representations. It was demonstrated that gamma activity is irregular at the single-trial level, and the episodes of increased gamma power ("gamma-bursts") are associated with elevated firing rates and increased amount of information about the WM content that could be decoded from spiking activity (Lundqvist et al., 2016, 2018; Bastos et al., 2018). It is not fully clear, however, whether the gamma power increase plays a functional role in WM retention or is it merely a consequence of transient firing rate increase during spontaneous reactivations of WM representations.

Besides the gamma power increase, an increase in gamma-band coherence between different cortical sites during the delay period was reported (Lutzenberger et al., 2002; Kaiser et al., 2003; Palva et al., 2010; Kornblith et al., 2016). Furthermore, it was shown that transcranial gamma-band electrical stimulation of two distant sites could improve performance in a WM task (Tseng et al., 2016). Interestingly, the improvement was observed only under anti-phase (but not under in-phase) stimulation. This result suggests that gamma-band coherence presumably plays a functional role in WM retention, and is not merely an epiphenomenon.

Nowadays, a number of computational WM models exist. Most of them are based on multistable neural networks. In the simplest case, a system has two stable states, one of which (with low firing rates) corresponds to the background regime, and the other one (with higher firing rates) relates to the active regime, in which an object is retained in WM. Transition from the background to the active state occurs under the action of a short excitatory pulse that mimics the arrival of a to-be-memorized stimulus. As an alternative, there are models, in which the active retention regime is metastable, and the system slowly returns to the background state after a stimulus presentation (Lim and Goldman, 2013). In many WM models, the self-sustained post-stimulus firing rate elevation is provided by reverberation of excitation in the network due to synaptic interactions (Amit and Brunel, 1997; Brunel and Wang, 2001). In addition, there are models in which post-stimulus enhancement of synaptic connections due to short-term plasticity plays a significant role (Mongillo et al., 2008, 2012; Hansel and Mato, 2013).

Many theoretical papers, following Amit and Brunel (1997), described WM models with asynchronous spiking activity. It was shown that if a model contains only fast excitation, then even slight synchronization returns it to the background state (Gutkin et al., 2001; Laing and Chow, 2001; Compte, 2006). However, generation of oscillations (i.e., synchronization) in a WM model is possible in the presence of slow NMDA receptors (Tegnér et al., 2002) or in the case of modular structure of the system (Lundqvist et al., 2010, 2011). Besides investigating the mechanisms of oscillations' appearance in WM models (Tegnér et al., 2002; Roxin and Compte, 2016), several theoretical studies explore possible functional roles of oscillations in WM (Lisman and Idiart, 1995; Ardid et al., 2010; Lundqvist et al., 2010, 2011; Kopell et al., 2011; Chik, 2013; Dipoppa and Gutkin, 2013; Pina et al., 2018; Schmidt et al., 2018; Sherfey et al., 2020). The methodology, however, differs substantially between these studies, and a unified theoretical framework in this field is still lacking.

We follow a paradigm used in Dipoppa and Gutkin (2013); Schmidt et al. (2018), in which a WM system is resonant and receives an external oscillatory input that controls its behavior. Schmidt et al. (2018) showed that gamma-band input could stabilize WM retention even if the active regime is initially metastable (which means that the firing rate increases after stimulus presentation, but then slowly returns to the background level in the absence of input oscillations). In the present study, we make a step further and consider systems with metastable active regime that are comprised of several excitatory-inhibitory circuits coupled by symmetrical excitatory connections and receiving the same stimulus-related signal (such systems could serve as models of a distributed representation of an object in WM). We explore stabilization of the metastable active regime by external gamma-band and white-noise inputs, both of which are assumed to come from neural populations not explicitly included into the model.

We assume that the gamma-band inputs are generated locally (and thus could have different phases for different

circuits), and that the noise inputs originate from distributed networks (that could jointly participate in WM retention, thus producing correlated activity). The model parameters in the focus of our research are the following: (1) in-phase or anti-phase character of the gamma inputs to the circuits, (2) commonality or independence of the noise inputs to the circuits, (3) NMDA:AMPA ratio of inter-circuit connections. By varying these parameters, we aim to understand whether gamma-band oscillations are able to preferentially stabilize the active regime in those circuits that participate in collective WM-related activity (and thus receive a common noise input rather than independent inputs). Another goal of our study is to explore whether synchronization of gamma generators (leading to in-phase gamma inputs to different circuits) affects the ability of the gamma input to stabilize WM retention, depending on whether the inter-circuit connection in the system are fast or slow.

The paper is organized as follows. We start from a single-circuit model and demonstrate that the active regime could be stabilized by gamma-band or white-noise input. Next, we explore joint stabilizing effect of gamma-band and white-noise inputs in a system of two circuits with mutual excitation. We vary the NMDA:AMPA ratio of the inter-circuit connections and parameters of the inputs (including phase difference between the gamma inputs to the circuits and commonality/independence of the noise inputs); for each parameter combination we evaluate the effectiveness of stabilization. Finally, we consider a multi-circuit system comprised of two local clusters, with all the circuits in a cluster receiving the same gamma input. Each cluster, in turn, contains two circuit groups: the circuits from the first groups receive a common noise input and the circuits from the second groups – independent noise inputs. We explored the stabilizing effect of gamma input on each circuit group, depending on whether it is delivered to the clusters in the same phase or in the anti-phase, conditioned on the type (slow/fast) of the inter-cluster connections.

MATERIALS AND METHODS

Model Description

In this article, we consider firing rate models of WM that consist of one, two, or many circuits each containing interacting excitatory and inhibitory neural populations (**Figure 1**). These circuits serve as representations of various parts or features of WM content and are linked by symmetrical excitatory connections. Each circuit receives an external input consisting of several components: (1) tonic (constant) input, (2) stimulus-related input, (3) zero-mean sinusoidal oscillatory input, and (4) white noise. The stimulus-related input is implemented as a rectangular pulse whose amplitude and duration are the same for all circuits in the model. The stimulus input is projected 80% to the excitatory population of a circuit and 20% to the inhibitory population. The oscillatory input impinges only on the excitatory populations of circuits and represents a modulatory signal from the outside of the modeled WM network. White noise mimics the input produced by activity of a larger network into which our model is embedded, but which was not modeled explicitly. We explore different cases, in which the circuits receive in-phase or anti-phase oscillatory inputs, as well as identical or independent noisy inputs.

The state of circuit populations is described by the following dynamical variables: (1) firing rates, (2) mean input currents (AMPA, NMDA, and GABAA), and (3) population variances of the input currents (AMPA and GABAA). We further argue that if the mean values of the AMPA and NMDA currents are of the same order of magnitude, the NMDA current variance has to be much smaller than the AMPA variance, since the NMDA time constant is much larger than the AMPA time constant. Hence, we set the NMDA current variance to zero. Thus, our model of interacting circuits is described by the following equations:

$$\begin{cases} \tau_{ra}\frac{dr_a^p}{dt} &= -r_a^p + F_{ra}(\mu_a^p, \sigma^p_{a,AMPA}, \sigma^p_{a,GABAA}) \\ \mu_a^p &= \sum_S \mu^p_{a,S},\ S = AMPA, NMDA, GABAA \\ \tau_S \cdot \frac{d\mu^p_{a,S}}{dt} &= -\mu^p_{a,S} + \tilde{\mu}^{p,rec}_{a,S} + \tilde{\mu}^{p,cc}_{a,S} + \tilde{\mu}^{p,ext}_{a,S} \\ \frac{\tau_S}{2}\cdot\frac{d\left(\sigma^p_{a,S}\right)^2}{dt} &= -\left(\sigma^p_{a,S}\right)^2 + \left(\tilde{\sigma}^{p,rec}_{a,S}\right)^2 + \left(\tilde{\sigma}^{p,cc}_{a,S}\right)^2 + \left(\tilde{\sigma}^{p,ext}_{a,S}\right)^2 \end{cases}$$

where, the lower index a denotes a population type (e, excitatory; i, inhibitory), the upper index p denotes a circuit number. The variable r_a^p is the firing rate; μ_a^p is the total mean input current, $\mu^p_{a,S}$ is the mean input current via the synapses of the type S (S denotes AMPA, NMDA, or GABAA); $\left(\sigma^p_{a,S}\right)^2$ is a population variance of the input current via the synapses of the type S; τ_{ra} is the time constant that governs the firing rate dynamics, τ_S is the synaptic time constant for the synapses of the type S; $\tilde{\mu}^{p,rec}_{a,S}$, $\left(\tilde{\sigma}^{p,rec}_{a,S}\right)^2$ are the population mean and variance of the recurrent (intra-circuit) inputs; $\tilde{\mu}^{p,cc}_{a,S}$, $\left(\tilde{\sigma}^{p,cc}_{a,S}\right)^2$ are the population mean and variance of the input from other modeled circuits; $\tilde{\mu}^{p,ext}_{a,S}$ and $\left(\tilde{\sigma}^{p,ext}_{a,S}\right)^2$ are the population mean and variance of the external inputs; F_{ra} are the gain (transfer) functions. Note that the time constant for the population variance of each current is twice smaller than the time constant for the population mean of the same current (due to the properties of linear first-order stochastic ODE's; (see Renart et al., 2007) for an example of a model with dynamical mean and variance of the input current). Within-circuit excitatory-to-excitatory connections demonstrate short-term plasticity (Tsodyks and Markram, 1997), see the full description in the **Supplementary Material**.

We also need to define the gain functions for the neural populations. Instead of defining these gain functions F_{re} and F_{ri} *a priori*, we calculate them following an approach similar to the one previously used in Schaffer et al. (2013) and Augustin et al. (2017) that allows to match low-dimensional models with spiking networks, making them more biologically plausible. According to this approach, the gain functions were

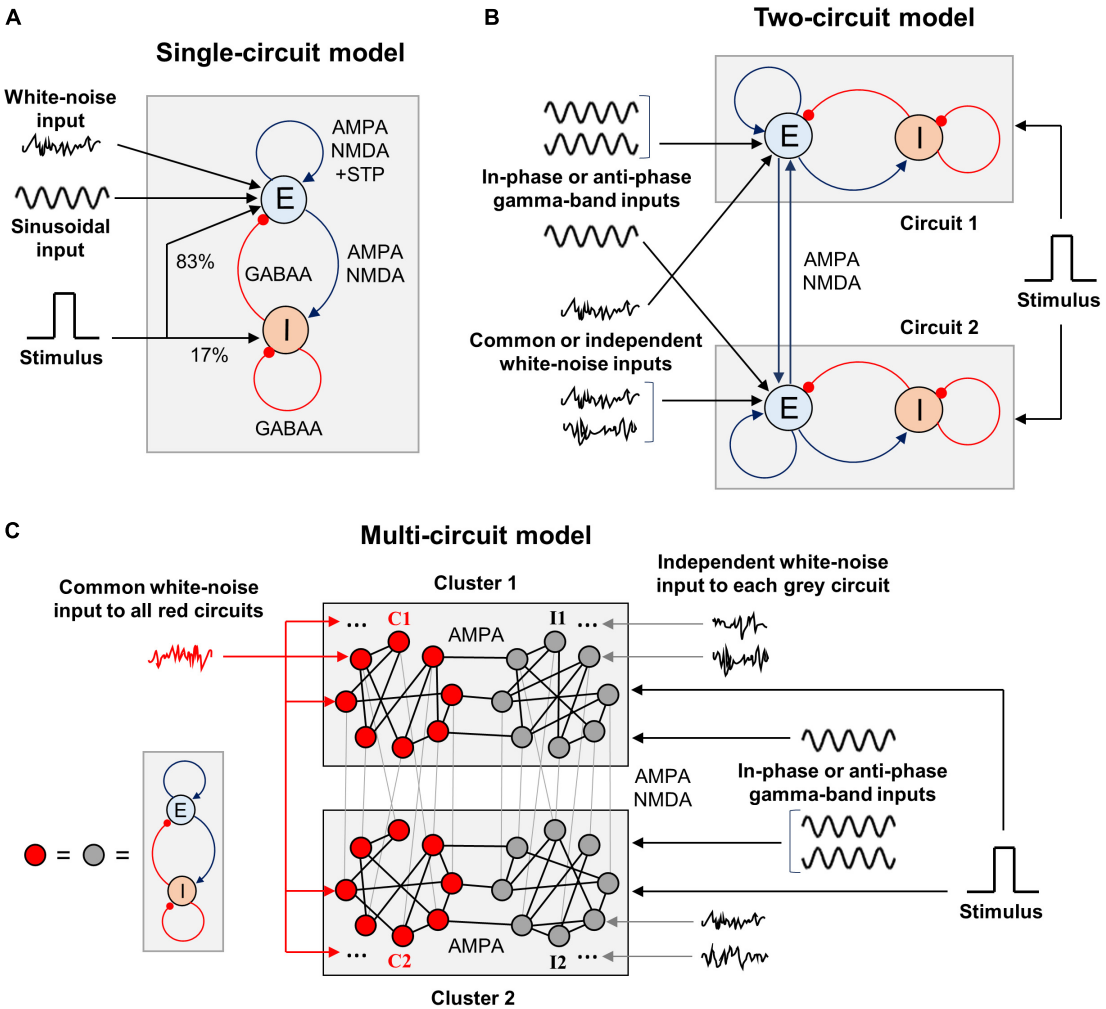


FIGURE 1 | Schematic representation of the WM model types we explore in this study. **(A)** Single-circuit system. E, excitatory population; I, inhibitory population. The excitation is mediated by AMPA and NMDA receptors, inhibition – by GABAA receptors. The excitatory-to-excitatory connections demonstrate short-term plasticity. White-noise and oscillatory inputs are delivered to the excitatory population. A rectangular stimulus-related signal is delivered to the excitatory and inhibitory populations in the proportion of 5:1. Both populations also receive constant inputs (not shown). **(B)** Two-circuit system. The circuits are structurally identical to the one shown in **(A)** and intercoupled via excitatory connections with varying NMDA:AMPA ratio. Gamma-band sinusoidal input is delivered to the excitatory populations of the circuits, either in the same phase or in the anti-phase. White-noise inputs (identical or independently generated) are also delivered to the excitatory populations. The stimulus-related signal is delivered in the same way as in **(A)**, and it is identical for both circuits. **(C)** Multi-circuit model. Each circle represents a circuit, structurally identical to the one shown in **(A)**. The circuits are grouped into two local clusters; each cluster contains two circuit groups. Lines connecting the circles represent symmetrical excitatory inter-circuit connections. Within-cluster inter-circuit connections are mainly AMPA-mediated (NMDA:AMPA ratio is 0.1), inter-cluster connections could be predominantly mediated either by AMPA or NMDA. The circuits from the "red" groups (C1 and C2) receive a common white-noise input; the circuits from the "gray" groups (I1 and I2) receive independent white-noise inputs. All the circuits from each cluster receive gamma-band input in the same phase; the inputs to the clusters could have the same phase or the opposite phases. The same stimulus-related signal is delivered to each circuit.

pre-calculated on a cubic grid (with the coordinates: total mean current, AMPA current std., and GABAA current std.) by numerical simulations of leaky integrate-and-fire (LIF) neurons having typical properties of the regular-spiking (pyramidal) neurons and the fast-spiking interneurons of the cortex. For an arbitrary point, the value of a gain function was obtained by interpolation between the pre-calculated values at the closest grid nodes. To show the shape of the gain functions, in **Figures 2C,D** we present their projections to the plane of mean firing rate and total mean input current. The projections were made for two different combinations of AMPA and GABAA current std. that correspond to the background and active states of the bistable single-circuit system (these states are depicted in **Figure 2A**; see the next section for details). Since our model is tuned to operate in a subthreshold regime [which is typical for cortical neurons; (Compte et al., 2003; Wang, 2010)], the aforementioned plots have an exponential-like, concave shape.

The full system of model equations and its detailed description are given in the **Supplementary Material**.

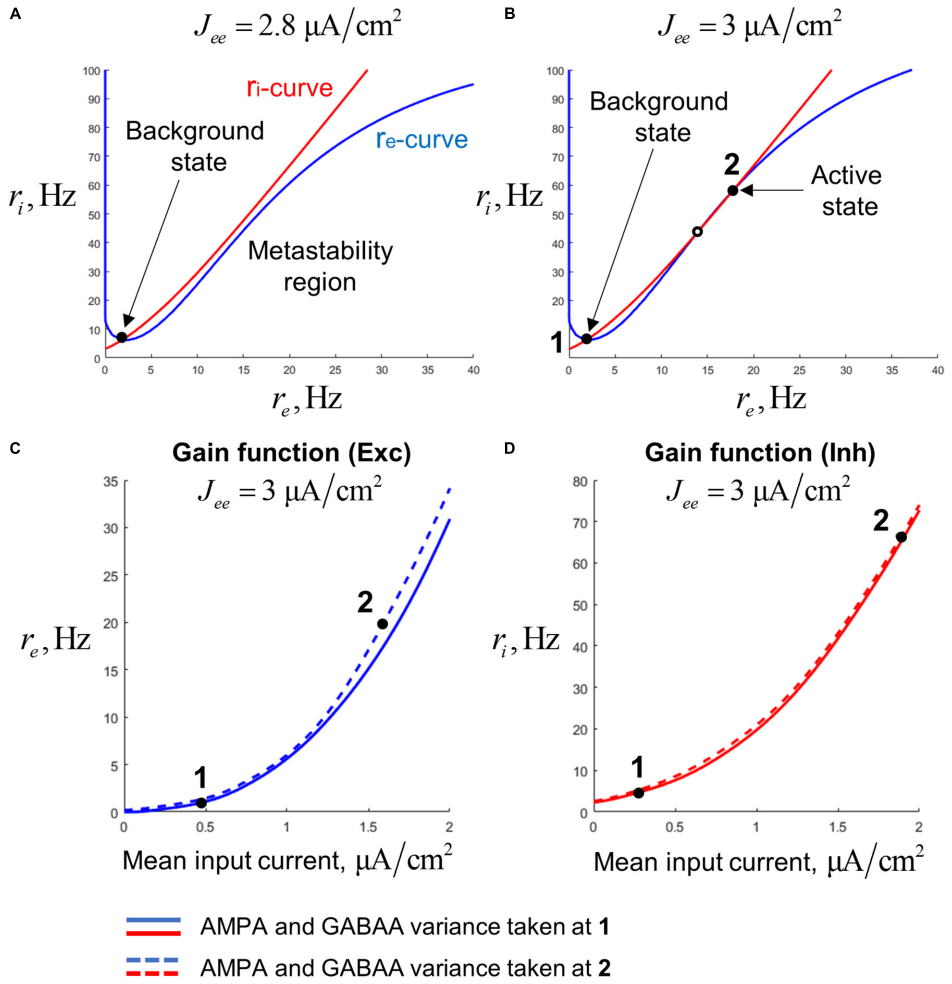


FIGURE 2 | Steady states of a single-circuit model. **(A)** Phase plane for the single-circuit model analyzed in our study. For the points on the blue curve (re-curve), all the derivatives, except of dri/dt are zero; for the point on the red curve (ri-curve), all the derivatives, except of dre/dt are zero. The system has the single (background) steady state and the region of slowly decaying activity (metastability region), where the re- and ri-curves are close to each other. **(B)** Phase plane for a system with increased recurrent excitation. The system has the background steady state (denoted as 1) and the active steady state (denoted as 2). **(C)** Two slices of the gain function of the excitatory population, representing dependence of the excitatory firing rate on the mean input current to the excitatory population. Solid curve – the slice taken at the constant values of the AMPA- and GABAA-current variance, equal to the values calculated at the steady state 1. Dashed curve – the slice taken at the variance values calculated at the steady state 2. **(D)** Same as **(C)**, but for the inhibitory population.

Model Parameters

Single Circuit

We set the parameters of a circuit (**Figure 1A**) in such way that: (1) it has a stable background steady state with a low level of activity, (2) it responds to a stimulus by prolonged activity increase with subsequent return to the vicinity of the background state, and (3) it has gamma-band resonance during the post-stimulus increased activity. We refer to the increased post-stimulus activity as metastable active regime and refer to a system with such regime as metastable system.

To obtain the required behavior of the circuit (metastability and gamma-band resonance), we set the parameters in the following way. We start by pre-selecting parameters that provide bistability in a circuit. In this case, there are three equilibria in the phase space: two stable ones (corresponding to the background and active states) and a saddle (**Figure 2B**). We tuned the parameters in such way that the second (active) equilibrium has a pair of complex-conjugate eigenvalues with a small negative real part and an imaginary part corresponding to the gamma band. An orbit of such a system, when starting near the active steady state, shows slowly decaying gamma-band oscillations returning back to the active state (an example is presented in **Supplementary Figure 1B**). Then, we decreased the weight of the excitatory-to-excitatory synaptic connections, until the upper equilibrium disappears through a fold bifurcation (see the phase plane with the single steady state in **Figure 2A**). In fact, the upper equilibrium leaves a "ghost" near which the dynamics are slow. Thus, the resulting

system is metastable, and the slow dynamics near the "ghost" corresponds to the metastable active regime. This regime inherits gamma-band resonance from the active steady state that existed in the bistable system before the bifurcation. An example orbit of the metastable system showing damped oscillations with subsequent decay to the background is presented in **Supplementary Figure 1A**.

Finally, we explored the ability of white-noise and sinusoidal inputs to stabilize the metastable active regime.

Two Circuits

We also considered a system of two identical interacting circuits (**Figure 1B**). In this case, we symmetrically connect the circuits using excitatory coupling between their excitatory populations. At the same time, to compensate this additional inter-circuit excitation, we decreased the background inputs to the excitatory populations. In this way, we keep the metastable behavior in each circuit and could study the influence of input oscillations and noise.

We varied NMDA:AMPA ratio for the inter-circuit connections and explored the ability of in-phase/anti-phase oscillatory inputs and common/independent white-noise inputs to stabilize the active regime. We used duration of increased post-stimulus activity as the measure of the active regime stability. The post-stimulus activity of a circuit was considered to be terminated when the time-course of its excitatory population firing rate smoothed with the 100 ms time window fell below the level of 3 Hz.

Multiple Circuits

Finally, we considered a multi-circuit system schematically presented in **Figure 1C**. Each circuit is represented by a (red or black) circle. Links between the circles correspond to mutual excitatory connections between the circuits. In the system considered, there were two circuit clusters that mimic two spatially separated local networks. All circuits in each cluster receive input oscillations of the same phase, whereas the circuits in different clusters may receive either in-phase or anti-phase oscillations.

Each cluster contains two groups of circuits. Circuits from the first groups C1, C2 (red-colored circuits in both clusters in **Figure 1C**) receive a common noise input. Circuits from the second groups I1, I2 (gray circuits in **Figure 1C**) receive independent noise inputs.

Each group in our model is a random graph with eight nodes (circuits) having the constant in-degree of three; two groups within the same cluster (C1–I1 and C2–I2) are connected by three randomly assigned links; corresponding groups in different clusters (C1–C2 and I1–I2) are connected by eight randomly assigned links (with one link per each node); non-corresponding groups in different clusters (C1–I2 and C2–I1) are not connected. All links between circuits are bi-directional.

We considered two multi-circuit models: one with the fast inter-cluster connections (90% AMPA and 10% NMDA), and the other one with slow inter-cluster connections (100% NMDA). All other parameters were identical. We simulated both models in several regimes: (1) no oscillatory input, (2) only one cluster receives oscillatory input, (3) the clusters receive in-phase oscillations, (4) the clusters receive anti-phase oscillations. For each regime, we were interested in the average duration of post-stimulus activity of each circuit group (C1, I1, C2, and I2).

RESULTS

Activity of Single-Circuit Model

A single-circuit model consisting of an excitatory and an inhibitory population (**Figure 1A**) could be either bistable or monostable, depending on the strength of the self-excitation. For strong enough self-excitation, the model has three equilibria (**Figure 2B**): two stable steady states and a saddle. The stable equilibria correspond to two types of activity: the background (low-rate) and active (high-rate) regimes, while the saddle manifold separates their basins of attraction. For weaker self-excitation, the equilibria that correspond to the active state and the saddle disappear through the fold bifurcation, and a metastable active regime (with high firing rate) remains as a ghost of the active state (**Figure 2A**). In this case, the circuit could be excited from the background state by a stimulus and demonstrate relatively long transient high-firing-rate activity.

The metastable active regime can be stabilized (i.e., the decay of the post-stimulus activity could be slowed or prevented) either by an oscillatory signal (**Figure 3**), or by a noisy input (**Figure 4**). In the case of oscillatory input, there is a range of its amplitudes, for which the stabilization occurs only if the input frequency falls into the lower gamma band (middle part of **Figure 3A**). For lower amplitudes, the stabilization does not occur (gray region in **Figure 3A**). For higher amplitudes, the frequency range of stabilization expands to the high gamma and beta bands. If the input amplitude is too high (above the dashed red line in **Figure 3A**), the background regime disappears, i.e., the oscillations put the system into the active (high-firing-rate) regime even without stimulus presentation.

We found that noisy input is also able to stabilize the metastable state (**Figure 4**). The purple curve represents the mean firing rate observed in the absence of a stimulus, as a function of the noise standard deviation. The green curve represents the mean post-stimulus firing rate. At low intensity, the noise was unable to stabilize the active regime, while at high intensity, the noise put the system to the active regime even without stimulus presentation. Thus, an intermediate range of noise intensities is appropriate for WM functioning; it is seen in **Figure 4A** as the range in which the green line goes considerably above the purple line. For a value from this range, the persistent activity is initiated by a transient stimulus, and the WM trace is then kept "alive" by the noise input (**Figure 4B**). At the same time, the system stays in the background regime if the stimulus is not presented (**Figure 4C**).

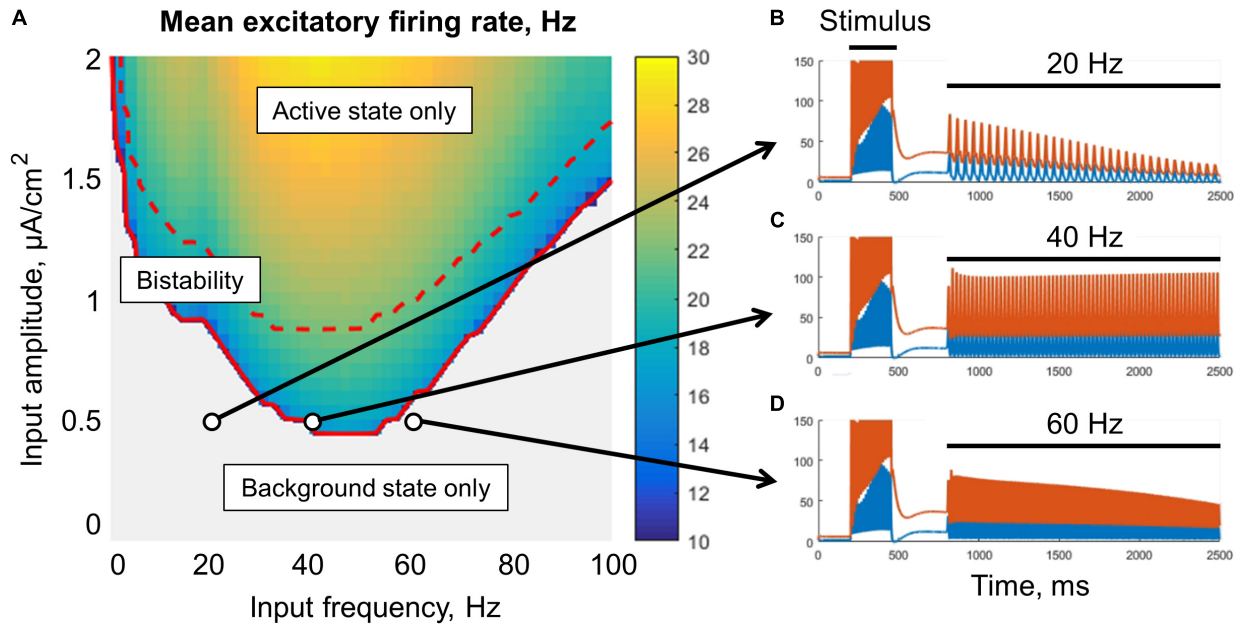


FIGURE 3 | Resonant stabilization of the metastable activity by a sinusoidal signal (without noise) in the single-circuit model. In panel **(A)**, there is a diagram showing the mean excitatory firing rate for different values of the amplitude and frequency of the signal. Simulation was performed for 10 s, and the firing rate was averaged over the last period of the input signal. Above the solid red curve, the active (initially metastable) regime is persistent; above the dashed red curve, there is no background regime. **(B–D)** Examples of the activity traces. Blue, excitatory firing rate; red, inhibitory firing rate. The system was shifted to the metastable active regime by a transient constant input (stimulus), after which an oscillatory input was turned on (20, 40, or 60 Hz zero-mean sinusoid). Note that the 40 Hz periodic input **(C)** provided a persistent active regime.

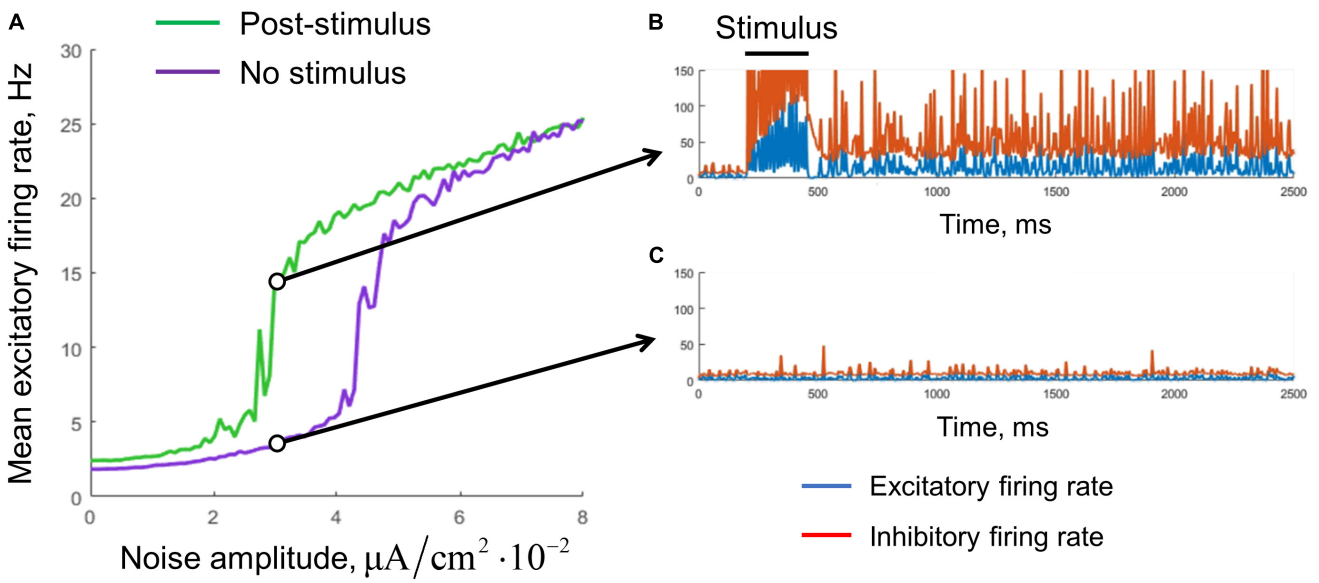


FIGURE 4 | Stabilization of the single-circuit metastable activity by a noisy input. **(A)** Mean excitatory firing rate, averaged over the interval 1–10 s. Green, the result for post-stimulus activity; purple, for background activity (without stimulus presentation). The stimulus presentation interval was 200–450 ms. The noise was delivered during the whole simulation time. Examples of the post-stimulus and the background activity traces ($A_{ex,NOISE} = 3 \times 10^{-2} \mu A/cm^2$) are presented in **(B,C)**, respectively. Blue curves, excitatory firing rate; red curves, inhibitory firing rate.

Activity of Two-Circuit Model

In this section, we explore joint stabilizing effect of input noise and oscillations on the system of two circuits with mutual excitatory connections (**Figure 1B**). We investigate different cases when circuits receive in-phase or anti-phase oscillations, as well as common or independent noise. The inter-circuit connections

contain both fast AMPA and slow NMDA components. We varied the proportion of the inter-circuit current that flows via NMDA receptors (which we denoted as k^{cross}_{NMDA}), while keeping the total inter-circuit connection strength at the constant level.

Dependence of the post-stimulus activity duration (which we use as the measure of the active regime stability) on k^{cross}_{NMDA} value and the amplitude of the input oscillations is presented in **Figure 5**. In the case of fast (AMPA) inter-circuit connections ($k^{cross}_{NMDA} = 0$), the in-phase oscillatory input to both circuits leads to stabilization of the active regime with high efficiency (**Figure 5**, lower part of the left panel). Anti-phase input oscillations in this case need to have a very high amplitude to stabilize the active regime (**Figure 5**, lower part of the right panel). With increasing k^{cross}_{NMDA} (i.e., the portion of the slow NMDA component of the inter-circuit current), effectiveness of the in-phase input deteriorates, while effectiveness of the anti-phase input increases (see the opposite trends in the left and right panels of **Figure 5**). For $k^{cross}_{NMDA} \cong 1$, both types of the input have approximately the same effect (**Figure 5**, upper parts of the panels).

In the presence of noise, our system generates irregular gamma-band quasi-oscillations. They act together with the external gamma-band input in stabilizing the active regime. The joint dependence of the post-stimulus activity duration on the input oscillations' amplitude and on the noise standard deviation is presented in **Figure 6**. In general, the activity duration increases when either the oscillations or the noise become stronger.

In the case of fast inter-circuit connections ($k^{cross}_{NMDA} = 0$, **Figures 6A,B**), common noise has stronger stabilizing effect than independent noise (compare the left and middle panels of **Figures 6A,B**). This is true in the presence of either in-phase or anti-phase oscillations, but in the case of anti-phase oscillations, the stabilizing effect occurs at much higher amplitudes compared to in-phase oscillations (see different horizontal scales in **Figures 6A,B**). The yellow region in the right panels of **Figures 6A,B** contains the combinations of oscillations' amplitude and noise standard deviation, for which there is an evident difference in the post-stimulus activity duration between the common-noise and the independent-noise cases (within our simulation time). Importantly, for intermediate noise intensities, an evident difference is observed only if the oscillations are sufficiently strong (see the horizontal black lines in the right panels of **Figures 6A,B** cross the yellow region near the right parts of the diagrams). In other words, an external gamma-band input with an appropriate amplitude stabilizes the active regime (i.e., WM retention) only when the two circuits receive a common noise input (e.g., when they belong to the same distributed representation), but not when they receive independent noise inputs of the same strength (e.g., when they are parts of different representations).

In the case of slow inter-circuit connections ($k^{cross}_{NMDA} = 1$, **Figures 6C,D**), common noise has weaker stabilizing effect than independent noise (compare the left and middle panels of **Figures 6A,B**). This difference, however, is less pronounced than in the case of fast inter-circuit connections (compare the right panels of **Figures 6A–D**). Furthermore, in-phase and anti-phase gamma-band inputs have almost the same effect (compare **Figures 6C,D**, note that the horizontal scale

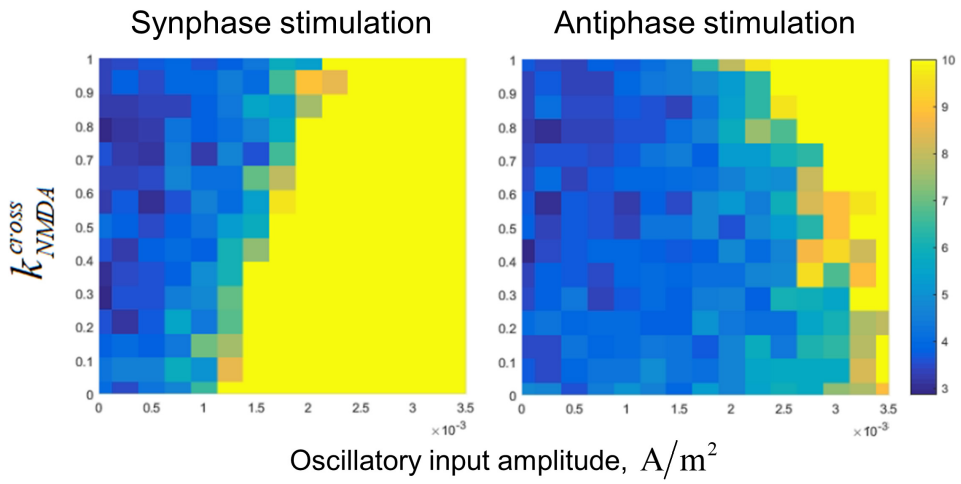


FIGURE 5 | Dependence of post-stimulus activity duration of the two-circuit system on the amplitude of input sinusoidal oscillations and the NMDA-to-total inter-circuit current ratio (k^{cross}_{NMDA}). Left panel: the oscillations are delivered to the circuits in the same phase, right panel: oscillations are delivered in the antiphase. Oscillation frequency: 40 Hz, noise amplitude: 0.014 $\mu A/cm^2$. Note that increasing of k^{cross}_{NMDA} reduces the stabilizing effect of the in-phase input, but increases the effect of the anti-phase input.

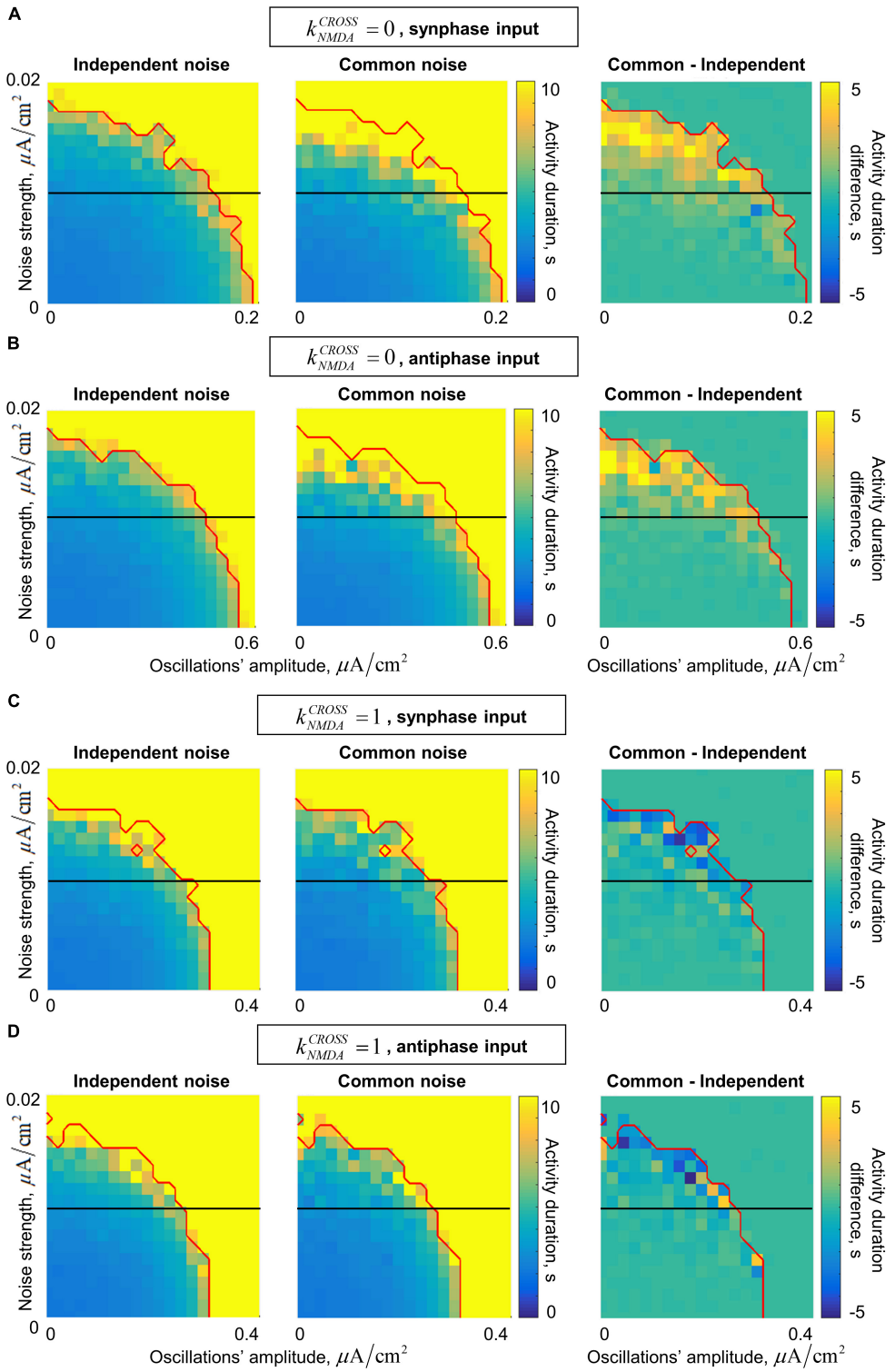


FIGURE 6 | Dependence of the post-stimulus activity duration in the two-circuit system on the amplitude of the input oscillations and the strength of the input noise for the in/anti-phase oscillations and two values of the NMDA connections strength. **(A,B)** Fast inter-circuit connection, **(C,D)** slow inter-circuit connections. **(A,C)** The circuits receive oscillatory signals in the same phase, **(B,D)** the circuits receive oscillatory signals in the opposite phases. Left column – each circuit receives an independent noise input; middle column – the circuits receive a common noise input; right column – the difference between these two cases. Red curve denotes the border of the saturation region: above this curve, the activity duration either for the common or for the independent noise case equals to the simulation time (i.e., above this curve comparison between the noise types does not make sense). Horizontal black lines denote the noise level of 0.01 (used further in the text). When the inter-circuit connections are fast, the activity is more robust (its duration is longer) in the case of common noise (yellow regions in the right panels of **(A,B)**). When the connections are slow, the activity is slightly more robust in the case of independent noise (blue regions in the right panels of **(C,D)**).

is the same). This is in contrast with the $k_{\mathrm{NMDA}}^{\mathrm{cross}} = 0$ case, in which in-phase input had much stronger effect than the anti-phase input. Such difference between $k_{\mathrm{NMDA}}^{\mathrm{cross}} = 0$ and $k_{\mathrm{NMDA}}^{\mathrm{cross}} = 1$ is in agreement with the result presented in **Figure 5** for a constant noise intensity. Let us, again, select a certain intermediate noise intensity (horizontal black lines in **Figures 6C,D**), and start to increase the gamma-band input amplitude. When the amplitude becomes high enough, the gamma-band input begins to stabilize the activity regime. This happens at slightly smaller amplitudes if the noise is independent, but irrespectively of whether the oscillatory input is in-phase or anti-phase.

Oscillatory Control of Multi-Circuit System

In this section, we show how the behavior of a multi-circuit system can be controlled by various types of oscillatory inputs. The results presented here are, in a large part, based on the effects of oscillations on two-circuit models described in the previous section. The multi-circuit system under consideration contains two clusters of circuits. Each cluster contains a group of circuits receiving a common noise input (the same signal for both clusters) and a group of circuits receiving independent noise inputs (we denote the "common-noise" groups of the first and second cluster as C1 and C2, respectively, and the "independent-noise" groups as I1 and I2). We consider two models – with fast and slow inter-cluster connections, respectively. We explore the behavior of the models in four conditions: (1) no oscillatory input ("NONE" condition), (2) only one cluster receives oscillatory input ("1CLUST" condition), (3) the clusters receive in-phase oscillations ("SYNC" condition), (4) the clusters receive anti-phase oscillations ("ANTI" condition). Each simulation was performed 25 times, and the statistics of group-averaged post-stimulus activity duration were collected for each of the four groups (C1, C2, I1, and I2) separately.

Figure 7 summarizes the results on the multi-circuit system behavior under the various oscillatory input conditions. The range of post-stimulus activity durations of the "common-noise" groups (C1 and C2) is represented by vertical red bars; the range of activity durations of the "independent-noise" groups (I1 and I2) – by vertical black bars. The horizontal dash in the middle of a bar marks the corresponding median value. Medians of the black and red bars obtained in the same condition are connected by black lines; a slope of such line demonstrates the difference in activity durations between the "common-noise" and "independent-noise" groups.

Without an oscillatory input ("NONE" condition), the groups receiving common noise input (C1 and C2) stay active for slightly longer time than the groups receiving independent noise inputs (I1 and I2). When the oscillatory input is delivered to the first cluster ("1CLUST" condition), it increases post-stimulus activity duration of both groups that belong to this cluster (C1 and I1). Importantly, the activity duration difference between the "common-noise" group (C1) and the "independent- noise" group (I1) also increases. These effects are almost absent for the second cluster (groups C2 and I2) due to relatively weak inter-cluster interaction.

When the oscillations are delivered to both clusters, either in the same phase or in the antiphase ("SYNC" and "ANTI" conditions, respectively), they produce similar effects to one-cluster oscillatory input, but these effects are stronger and involve both clusters. Thus, post-stimulus activity duration for all the groups (C1, I1, C2, and I2), as well as activity duration difference between the "common-noise" and "independent-noise" groups (C1–I1 and C2–I2) are increased. In **Figure 7**, we can see that the activity duration is higher in "SYNC" and "ANTI" conditions, compared to "NONE" and "1CLUST" conditions for all noise input types (red and black), clusters (1 and 2) and model types (slow and fast inter-cluster connections). It is also seen that the links between the red and black boxes in both clusters (C1–I1 and C2–I2) are steeper in "SYNC"/"ANTI" conditions than in "NONE"/"1CLUST" conditions (for both model types), which reflects increased duration difference between the "common-noise" and "independent-noise" groups.

The models with fast and slow inter-cluster connections differ in their response to in-phase and anti-phase oscillatory inputs. In the model with fast connections, the in-phase input provides strong increase both in the activity duration and in the duration difference between the "common-noise" and "independent-noise" groups, while both these effects are considerably weaker in the case of the anti-phase input. On the contrary, in the model with slow inter-cluster connections, the in-phase and anti-phase oscillatory inputs lead to the effects of roughly the same strength. This difference between the two models is best visible in the results averaged over the clusters (right panels in **Figures 7A,B**). It is seen that the median levels and the slopes of the "red-black" links are higher for "SYNC" condition than for "ANTI" condition in the right panel of **Figure 7A**. At the same time, both the median levels and the slopes for "SYNC" and "ANTI" conditions are close to each other **Figure 7B**.

DISCUSSION

In this study, we considered several WM model variants, comprising of one, two, or many excitatory-inhibitory metastable circuits having mutual excitatory connections and receiving the same stimulus-related signal. First, we explored a single-circuit system and demonstrated that input gamma oscillations or white noise could stabilize (i.e., prolong) post-stimulus metastable activity. Next, we considered a two-circuit model and found that: (1) fast (AMPA) inter-circuit connections provide better stabilization by common-noise input compared to independent-noise inputs, (2) this difference could be further amplified by oscillatory inputs, (3) in-phase oscillatory inputs to the circuits produce better stabilization than anti-phase inputs in the case of fast (AMPA) connections, (4) in-phase and anti-phase inputs produce similar effects in the case of slow (NMDA) connections.

Finally, we developed a more realistic, multi-circuit system able to align with the effects observed in the two-circuit model. The system comprised of two clusters receiving in-phase or anti-phase oscillations and linked by fast (AMPA) or slow

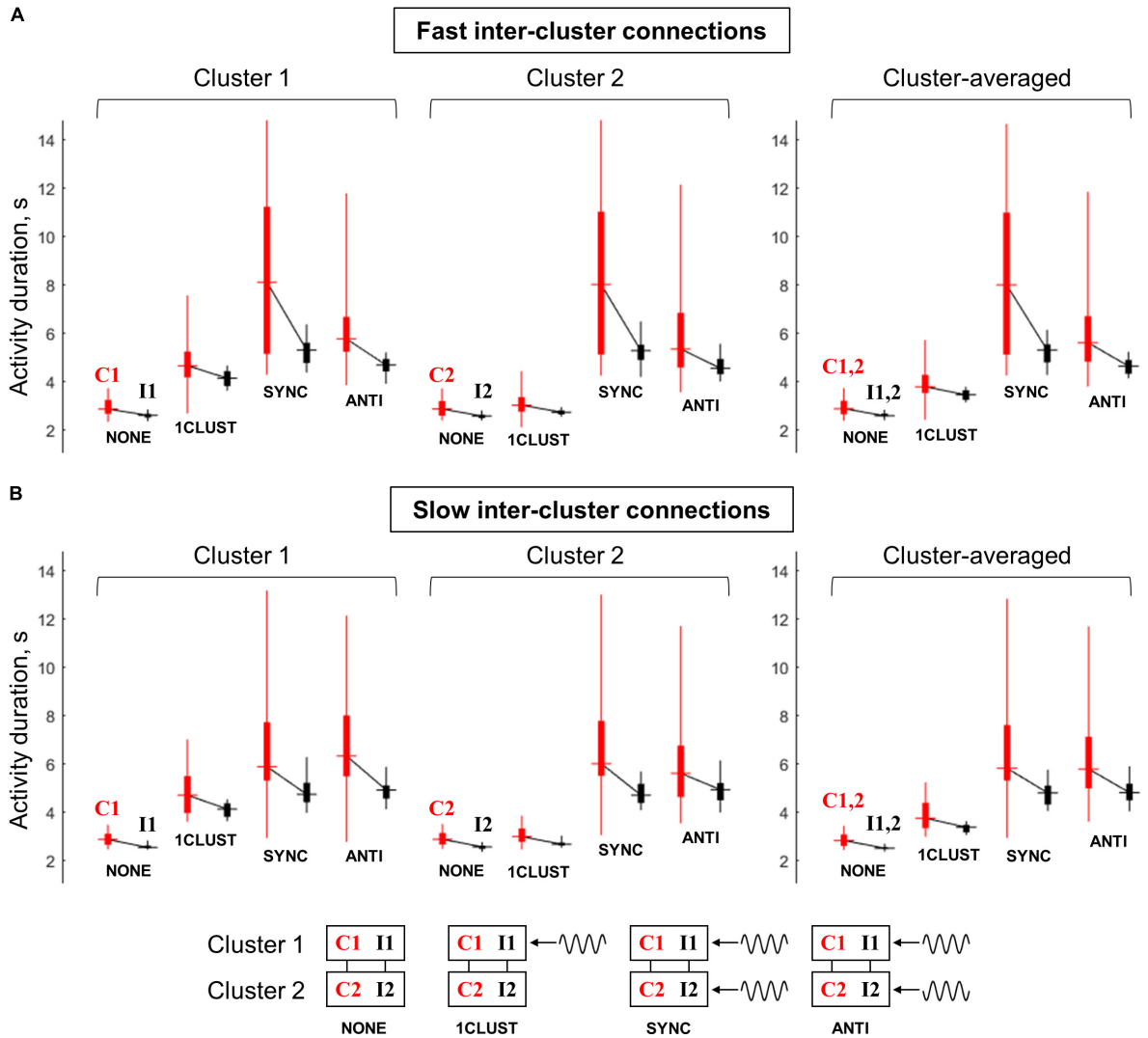


FIGURE 7 | Duration of post-stimulus activity for various circuit groups in the multi-circuit model, depending on the type of oscillatory input. **(A)** The model with fast inter-cluster connections, **(B)** the model with slow inter-cluster connections. The connected pairs of red and black vertical bars correspond to pairs of circuit groups within the same cluster. Red bars correspond to the groups that receive a common noise input (C1 and C2), black bars – to the groups that receive independent noise inputs (I1 and I2). Each bar represents statistics of group-averaged post-stimulus activity durations (obtained in 25 simulation runs): horizontal dash – the median, thick vertical line – two middle quartiles, thin vertical line – range between the minimal and maximal values. Left panel: cluster 1 (groups C1, I1), middle panel: cluster 2 (groups C2 and I2), right panel – average over the two clusters (C1+C2 and I1+I2). Four pairs of boxes in each panel correspond to four types of the oscillatory input: NONE – no oscillations, 1CLUST – oscillations are delivered to one cluster only, SYNC – oscillations are delivered to both clusters in the same phase, ANTI – oscillations are delivered to both clusters in the opposite phases (schematic representations on the input types are presented in the bottom part of the figure). It is seen that oscillatory input increases the activity duration, as well as the difference in activity duration between the "common-noise" and "independent-noise" groups; these effects are stronger when oscillations are delivered to both clusters. The effects of the in-phase input are more prominent than of the anti-phase input for the model with fast inter-cluster connections. For the model with slow inter-cluster connections, the in-phase and anti-phase inputs produce the effects of the same strength.

(NMDA) connections. Each cluster contained a group of circuits that receive a common noise input (representing activity of a distributed WM representation this group is embedded into) and a group of circuits that receive independent inputs (and, thus, do not participate in the distributed WM retention). We found that: (1) oscillatory inputs stabilize activity of all circuits, (2) this stabilization is more pronounced for the "common-noise" groups compared to "independent-noise" groups, (3) in the case of fast inter-cluster connections, both the stabilization and separation between the "common-noise" and "independent-noise" groups are more pronounced for in-phase input compared to anti-phase input, (4) in the case of slow connections, in-phase and anti-phase inputs produce comparable effects.

In summary, we demonstrated that gamma oscillations are able to selectively stabilize activity of the circuits that receive common noise input, thus supporting coherent activity of a

distributed WM representation. If long-range connections are fast, a part of this representation could be further highlighted by synchronizing gamma activity within this part. If these connections are slow, the distributed representation is uniformly stabilized, irrespective of phase differences between gamma oscillations in its local parts.

Single-Circuit System

We started from a single-circuit system and demonstrated that, with an appropriate parameter selection, both the noise and the gamma-band input alone could stabilize WM retention, without affecting the background regime. To explain these effects, we note that our model was tuned to operate in a subthreshold regime, in which the gain functions of the populations (excitatory and inhibitory) were concave. Consequently, even a zero-mean input (noisy or oscillatory) produced additional mean excitation to the populations. In our system, the mean effect of oscillations on the excitatory population outweighed their effect on the inhibitory population, which resulted in oscillation-induced/noise-induced mean firing rate increase and, consequently, to stabilization of the active regime.

A similar effect – stabilization of initially metastable active regime in a WM model by gamma-band input – was previously demonstrated by Schmidt et al. (2018). In this study, the authors considered a purely excitatory system, but took into account the effect of spike-to-spike synchronization, which allowed to achieve resonant behavior without inhibitory population. The system had a resonance in the beta band, but the stabilization occurred under high-gamma input, while beta-band input, on the contrary, switched the system to the background state. A similar switching to the background state by a resonant input was also reported by Dipoppa and Gutkin (2013). In our model, we do not observe this effect, presumably due to the presence of slow NMDA currents ant short-term plasticity, which make the activity more robust to transient episodes of synchronization. As a result, the effect of input oscillations is always excitatory in our model.

We should note that our model is more biologically realistic, compared to Dipoppa and Gutkin (2013) and Schmidt et al. (2018). First, it contains both an excitatory and an inhibitory population. Second, it operates in a subthreshold regime [which is typical for cortical networks during WM retention, (Compte et al., 2003)], while in the aforementioned models, the neurons operated mostly in a suprathreshold regime with regular spiking activity. Third, both these models used highly non-linear pulse-like periodic inputs, while in our case, the oscillatory input was sinusoidal [which is, again, closer to the situation in the neocortex; (see Wang, 2010)].

Two-Circuit System

In the two-circuit system, under a constant noise intensity, we demonstrated that in-phase gamma-band input stabilizes WM retention much more effectively than anti-phase input when inter-circuit interaction is fully mediated by fast AMPA receptors ($k_{NMDA}^{cross} = 0$). With increasing NMDA:AMPA ratio of the inter-circuit connections, effectiveness of the in-phase gamma-band input decreased, while effectiveness of the anti-phase input increased. When the interaction was fully mediated by slow NMDA receptors ($k_{NMDA}^{cross} = 1$), effectiveness of in-phase and anti-phase input was about the same.

The observed effects could be explained by the fact that the eigenmode of the system is in-phase if $k_{NMDA}^{cross} = 0$ and anti-phase if $k_{NMDA}^{cross} = 1$. This fact is illustrated in **Supplementary Figures 2E, 3E**. Without oscillatory input, independent noise inputs lead to in-phase quasi-oscillations if $k_{NMDA}^{cross} = 0$ (left part of **Supplementary Figure 2E**, black curve); however, if $k_{NMDA}^{cross} = 1$, then the noise-induced quasi-oscillations have the average phase difference between the circuits closer to π (left part of **Supplementary Figure 3E**, black curve). The most effective entrainment (and, thus, the most effective stabilization) occurs when the phase difference between the inputs to the circuits matches the eigenmode of the system. In the case of $k_{NMDA}^{cross} = 0$, anti-phase input should change activity of the system from the in-phase to the anti-phase mode to fully entrain it (**Supplementary Figure 2F**), which is accompanied by oscillation-induced desynchronization at certain input amplitudes (**Supplementary Figure 2D**); in the case of $k_{NMDA}^{cross} = 1$, the anti-phase entrainment is much easier (**Supplementary Figures 3D,F**). On the contrary, in-phase oscillations more easily entrain the system if $k_{NMDA}^{cross} = 0$ (**Supplementary Figures 2C,E**), rather than if $k_{NMDA}^{cross} = 1$ (**Supplementary Figures 3C,E**).

We next considered how the duration of post-stimulus activity (i.e., effectiveness of the active regime stabilization) jointly depends on noise intensity and gamma-band input amplitude. For the fast inter-circuit connections ($k_{NMDA}^{cross} = 0$), common noise stabilized the active regime more effectively than independent noise. On the contrary, for the slow inter-circuit connections ($k_{NMDA}^{cross} = 1$), independent noise was slightly more effective. Importantly, for intermediate noise intensities, there was a range of gamma amplitudes, in which the gamma-band input stabilized the active regime with considerably different effectiveness depending on the noise type (common / independent). We used this effect as the basis for selective oscillatory control of activity robustness in the multi-circuit system.

The higher effectiveness of the common noise in the case of $k_{NMDA}^{cross} = 0$ could be, again, explained by the fact that the system's eigenmode in this case is in-phase. Common noise fully projects to this mode, while independent noise projects to it only partially. Thus, entrainment of the in-phase mode by independent noise is weaker, which leads to smaller stabilizing effect. On the contrary, independent noise partially projects to the anti-phase mode (which is the eigenmode of the system with $k_{NMDA}^{cross} = 1$), while common noise is orthogonal to this mode, so independent noise is more effective when $k_{NMDA}^{cross} = 1$.

There is another effect that possibly participates in the link between k_{NMDA}^{cross} and activity robustness. Since slow NMDA receptors provide low-pass filtering of activity, they make the system less resonant. Consequently, increase of k_{NMDA}^{cross} decreases the amplitude of entrained oscillations or noise-induced quasi-oscillations (compare left parts of **Supplementary Figures 2G, 3G**). This effect presumably sums up with the aforementioned effects of eigenmode matching by external inputs.

We note that worsening of the stabilization by in-phase oscillations for higher NMDA:AMPA ratio seemingly contradicts earlier results suggesting the stabilizing role of NMDA receptors in WM retention in the presence of oscillatory activity (Tegnér et al., 2002). This discrepancy could be explained by the fact that the model used in Tegnér et al. (2002) is bistable, but does not contain any slow variables except of the NMDA current. Consequently, the system in the active state is stable in the absence of fluctuations, but fast fluctuations (such as gamma oscillations) make it fall down to the background state if the NMDA:AMPA ratio is too small. On the contrary, our model is metastable, so it spontaneously returns from the active regime to the background state in the absence of fluctuations. At the same time, our system contains slow variables besides the inter-circuit NMDA current – namely, within-circuit NMDA currents and coefficients of within-circuit short-term plasticity. As the result, our system is resistant to fast fluctuations (i.e., it survives fluctuation-induced "gaps" in the activity), even if the within-circuit NMDA:AMPA ratio is zero. Fast fluctuations, instead, increase the level of activity (thus stabilizing it) due to concavity of the gain functions. Notably, the ability of periodic input to stabilize metastable regime even in a model without slow variables was previously demonstrated by Schmidt et al. (2018). We suggest that the stabilizing effect of NMDA current described in Tegnér et al. (2002) could be also present in our model, but it is weak compared to the other effects we described.

Multi-Circuit System

In the analysis of the two-circuit system, we observed two important effects. First, in the case of fast connections and intermediate noise level, input oscillations could strongly stabilize WM retention in a pair of circuits receiving a common noise input, while producing much weaker effect on circuits receiving independent noise inputs. Second, if the connections between two circuits are fast, then the stabilizing effect of oscillations is considerably stronger when oscillations are delivered to the circuits in the same phase rather than in the opposite phases; however, this difference is mitigated if the inter-circuit connections are slow.

In order to demonstrate how these effects could be scaled up and utilized for oscillatory control of WM retention, we developed a multi-circuit system. We assumed that WM retention is based on distributed neural activity, and that local cortical patches could contain circuits involved in this activity, as well as circuits not involved in it (a similar scheme was used in Lundqvist et al. (2011), in which local modules contained parts of various representations, and corresponding parts in different modules were connected by long-range projections). We explicitly included two local patches (referred to as clusters) into our WM system and modeled the inputs from other parts of the cortex as white-noise signals. The circuits in each cluster formed two groups, named C1, I1 in the first cluster, and C2, I2 – in the second one. The groups C1, C2 were considered to participate in the coherent distributed activity related to WM retention, so the circuits from these groups received common noise input. The groups I1, I2 did not participate in this activity, so their circuits received independent noise inputs. We note that, unlike most WM models (e.g., Brunel and Wang, 2001; Lundqvist et al., 2011), participation of a group in the WM representation is not determined at the stimulus presentation stage, since all the circuits in our model receive the same stimulus signal. Instead, a group is considered to be a part of the WM representation if its circuits receive common noise input.

We demonstrated that, as in the two-circuit case, gamma-band input can stabilize WM retention, i.e., increase post-stimulus activity duration. Importantly, this stabilization was more prominent for the "common-noise" groups (C1 and C2), compared to the "independent-noise" groups (I1 and I2). Thus, the gamma-band input also increased "selectivity" of WM retention, predominantly stabilizing those circuits that participate in WM-related distributed coherent activity. This behavior is in agreement with the reported increase in both the firing rates and stimulus selectivity that occurs on top of elevated gamma activity episodes during WM retention (Lundqvist et al., 2016, 2018; Bastos et al., 2018). Such functionality of gamma-input in our model is achieved due to high AMPA-based within-group connectivity and low inter-group connectivity. As a consequence of this, circuit pairs with common noise input and pairs with independent inputs are mostly separated from each other, so the results obtained for the two-circuit system are still applicable. If the inter-group connections are too dense, correlations would spread across the whole system, and the difference between the "common-noise" and "independent-noise" circuits would be diminished.

We suggest that the dense within-group and sparse inter-group connectivity follows naturally from the Hebbian plasticity principles: the circuits that receive the same input (members of C1 and C2) would instantiate links between each other. We assume that the "independent-input" circuits (members of I1 and I2) could also participate in the dominant active representation (and thus receive a common input) in certain cases that we do not model explicitly here: e.g., for different WM content or different task rules. On the contrary, since the groups C1 and I1 (as well as C2 and I2) participate in different representations, their circuits do not usually receive common inputs, which justifies sparse inter-group (C1–I1 and C2–I2) connections.

We also confirmed in the multi-circuit module the influence of NMDA:AMPA ratio on the system's tendency to respond differently depending on the phase between gamma-band oscillatory inputs delivered to its parts. We demonstrated that the stabilizing effect of the input oscillations (i.e., oscillation-induced prolongation of post-stimulus activity) in the system with fast (AMPA-based) inter-cluster connections is stronger when the oscillations are delivered to the clusters in the same phase rather than in the opposite phases. On the contrary, the stabilizing effect of the in-phase and anti-phase inputs to the clusters was the same in the case of slow (NMDA-based) inter-cluster connections. This difference between the slow ant fast inter-cluster connections was observed only when the connectivity between the corresponding groups of different clusters (C1–C2 and I1–I2) was weaker than the within-group connectivity. We suggest that such layout is biologically plausible, due to presumed small-world character of the cortical topology

(Bullmore and Sporns, 2009; Bassett and Bullmore, 2017), with long-range connections being, in general, sparser than short-range connections (Ercsey-Ravasz et al., 2013).

Thus, fast and slow inter-cluster connections provide different functionality. Fast connections allow to selectively increase retention robustness in a part of an active distributed representation by synchronizing it in the gamma-band (e.g., by providing common gamma-band input from a controller cortical or subcortical network). In turn, slow inter-cluster connections provide a basis for robust WM retention by a distributed network with local gamma generators, which do not need to be synchronized. We note that the local character of gamma oscillations is experimentally supported (Donner and Siegel, 2011), and it was also assumed in previous multi-modular models of WM (Lundqvist et al., 2011).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

NN carried out the research. NN, DZ, and BG conceived the research. NN, DZ, VM, and BG discussed the results and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This article is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE University). BG acknowledges support from CNRS, INSERM, ANR-17-EURE-0017, and ANR-10-IDEX-0001-02.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncir.2021.647944/full#supplementary-material

REFERENCES

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

Ardid, S., Wang, X. J., Gomez-Cabrero, D., and Compte, A. (2010). Reconciling coherent oscillation with modulation of irregular spiking activity in selective attention: gamma-range synchronization between sensory and executive cortical areas. *J. Neurosci.* 30, 2856–2870. doi: 10.1523/JNEUROSCI.4222-09.2010

Augustin, M., Ladenbauer, J., Baumann, F., and Obermayer, K. (2017). Low-dimensional spike rate models derived from networks of adaptive integrate-and-fire neurons: comparison and implementation. *PLoS Comput. Biol.* 13:e1005545. doi: 10.1371/journal.pcbi.1005545

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839. doi: 10.1038/nrn1201

Bassett, D. S., and Bullmore, E. T. (2017). Small-world brain networks revisited. *Neuroscientist* 23, 499–516. doi: 10.1177/1073858416667720

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

Brunel, N., and Wang, X. J. (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J. Comput. Neurosci.* 11, 63–85. doi: 10.1023/a:1011204814320

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. doi: 10.1038/nrn2575

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

Chik, D. (2013). Theta-alpha cross-frequency synchronization facilitates working memory control – a modeling study. *Springerplus* 2:14. doi: 10.1186/2193-1801-2-14

Compte, A. (2006). Computational and in vitro studies of persistent activity: edging towards cellular and synaptic mechanisms of working memory. *Neuroscience* 139, 135–151. doi: 10.1016/j.neuroscience.2005.06.011

Compte, A., Constantinidis, C., Tegnér, J., Raghavachari, S., Chafee, M. V., Goldman-Rakic, P. S., et al. (2003). Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *J. Neurophysiol.* 90, 3441–3454. doi: 10.1152/jn.00949.2002

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

Dipoppa, M., and Gutkin, B. S. (2013). Flexible frequency control of cortical oscillations enables computations required for working memory. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12828–12833. doi: 10.1073/pnas.1303270110

Donner, T. H., and Siegel, M. (2011). A framework for local cortical oscillation patterns. *Trends Cogn. Sci.* 15, 191–199. doi: 10.1016/j.tics.2011.03.007

Ercsey-Ravasz, M., Markov, N. T., Lamy, C., Van Essen, D. C., Knoblauch, K., Toroczkai, Z., et al. (2013). A predictive network model of cerebral cortical connectivity based on a distance rule. *Neuron* 80, 184–197. doi: 10.1016/j.neuron.2013.07.036

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

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

Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron* 14, 477–485.

Gutkin, B. S., Laing, C. R., Colby, C. L., Chow, C. C., and Ermentrout, G. B. (2001). Turning on and off with excitation: the role of spike-timing asynchrony and synchrony in sustained neural activity. *J. Comput. Neurosci.* 11, 121–134. doi: 10.1023/A:1012837415096

Haegens, S., Osipova, D., Oostenveld, R., and Jensen, O. (2010). Somatosensory working memory performance in humans depends on both engagement and disengagement of regions in a distributed network. *Hum. Brain Mapp.* 31, 26–35. doi: 10.1002/hbm.20842

Hansel, D., and Mato, G. (2013). Short-term plasticity explains irregular persistent activity in working memory tasks. *J. Neurosci.* 33, 133–149. doi: 10.1523/JNEUROSCI.3455-12.2013

Howard, M. W., Rizzuto, D. S., Caplan, J. B., Madsen, J. R., Lisman, J., Aschenbrenner-Scheibe, R., et al. (2003). Gamma oscillations correlate with working memory load in humans. *Cereb Cortex* 13, 1369–1374. doi: 10.1093/cercor/bhg084

Jokisch, D., and Jensen, O. (2007). Modulation of gamma and alpha activity during a working memory task engaging the dorsal or ventral stream. *J. Neurosci.* 27, 3244–3251. doi: 10.1523/JNEUROSCI.5399-06.2007

Kaiser, J., Ripper, B., Birbaumer, N., and Lutzenberger, W. (2003). Dynamics of gamma-band activity in human magnetoencephalogram during auditory pattern working memory. *Neuroimage* 20, 816–827. doi: 10.1016/S1053-8119(03)00350-1

Kopell, N., Whittington, M. A., and Kramer, M. A. (2011). Neuronal assembly dynamics in the beta1 frequency range permits short-term memory. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3779–3784. doi: 10.1073/pnas.1019676108

Kornblith, S., Buschman, T. J., and Miller, E. K. (2016). Stimulus load and oscillatory activity in higher cortex. *Cereb. Cortex* 26, 3772–3784. doi: 10.1093/cercor/bhv182

Laing, C. R., and Chow, C. C. (2001). Stationary bumps in networks of spiking neurons. *Neural Comput.* 13, 1473–1494. doi: 10.1162/089976601750264974

Liebe, S., Hoerzer, G. M., Logothetis, N. K., and Rainer, G. (2012). Theta coupling between V4 and prefrontal cortex predicts visual short-term memory performance. *Nat. Neurosci.* 15, 456–462. doi: 10.1038/nn.3038

Lim, S., and Goldman, M. S. (2013). Balanced cortical microcircuitry for maintaining information in working memory. *Nat. Neurosci.* 16, 1306–1314. doi: 10.1038/nn.3492

Lisman, J. E., and Idiart, M. A. P. (1995). Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science* 267, 1512–1515. doi: 10.1126/science.7878473

Lundqvist, M., Compte, A., and Lansner, A. (2010). Bistable, irregular firing and population oscillations in a modular attractor memory network. *PLoS Comput. Biol.* 6:e1000803. doi: 10.1371/journal.pcbi.1000803

Lundqvist, M., Herman, P., and Lansner, A. (2011). Theta and gamma power increases and alpha/beta power decreases with memory load in an attractor network model. *J. Cogn. Neurosci.* 23, 3008–3020. doi: 10.1162/jocn_a_00029

Lundqvist, M., Herman, P., Warden, M. R., Brincat, S. L., and Miller, E. K. (2018). Gamma and beta bursts during working memory readout suggest roles in its volitional control. *Nat. Commun.* 9:394. doi: 10.1038/s41467-017-02791-8

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

Lutzenberger, W., Ripper, B., Busse, L., Birbaumer, N., and Kaiser, J. (2002). Dynamics of gamma-band activity during an audiospatial working memory task in humans. *J. Neurosci.* 22, 5630–5638. doi: 10.1523/JNEUROSCI.22-13-05630.2002

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

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

Mongillo, G., Hansel, D., and van Vreeswijk, C. (2012). Bistability and spatio-temporal irregularity in neuronal networks with nonlinear synaptic transmission. *Phys. Rev. Lett.* 108:158101. doi: 10.1103/PhysRevLett.108.158101

Palva, J. M., Monto, S., Kulashekhar, S., and Palva, S. (2010). Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7580–7585. doi: 10.1073/pnas.0913113107

Palva, S., Kulashekhar, S., Hämäläinen, M., and Palva, J. M. (2011). Localization of cortical phase and amplitude dynamics during visual working memory encoding and retention. *J. Neurosci.* 31, 5013–5025. doi: 10.1523/JNEUROSCI.5592-10.2011

Pina, J. E., Bodner, M., and Ermentrout, B. (2018). Oscillations in working memory and neural binding: a mechanism for multiple memories and their interactions. *PLoS Comput. Biol.* 14:e1006517. doi: 10.1371/journal.pcbi.1006517

Renart, A., Moreno-Bote, R., Wang, X. J., and Parga, N. (2007). Mean-driven and fluctuation-driven persistent activity in recurrent networks. *Neural Comput.* 19, 1–46. doi: 10.1162/neco.2007.19.1.1

Roux, F., and Uhlhaas, P. J. (2014). Working memory and neural oscillations: α-γ versus θ-γ codes for distinct WM information? *Trends Cogn. Sci.* 18, 16–25. doi: 10.1016/j.tics.2013.10.010

Roxin, A., and Compte, A. (2016). Oscillations in the bistable regime of neuronal networks. *Phys. Rev. E* 94:012410. doi: 10.1103/PhysRevE.94.012410

Sauseng, P., Klimesch, W., Heise, K. F., Gruber, W. R., Holz, E., Karim, A. A., et al. (2009). Brain oscillatory substrates of visual short-term memory capacity. *Curr. Biol.* 19, 1846–1852. doi: 10.1016/j.cub.2009.08.062

Schaffer, E. S., Ostojic, S., and Abbott, L. F. (2013). A complex-valued firing-rate model that approximates the dynamics of spiking networks. *PLoS Comput. Biol.* 9:e1003301. doi: 10.1371/journal.pcbi.1003301

Schmidt, H., Avitabile, D., Montbrió, E., and Roxin, A. (2018). Network mechanisms underlying the role of oscillations in cognitive tasks. *PLoS Comput. Biol.* 14:e1006430. doi: 10.1371/journal.pcbi.1006430

Sherfey, J., Ardid, S., Miller, E. K., Hasselmo, M. E., and Kopell, N. J. (2020). Prefrontal oscillations modulate the propagation of neuronal activity required for working memory. *Neurobiol. Learn. Mem.* 173:107228. doi: 10.1016/j.nlm.2020.107228

Siegel, M., Warden, M. R., and Miller, E. K. (2009). Phase-dependent neuronal coding of objects in short-term memory. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21341–21346. doi: 10.1073/pnas.0908193106

Tegnér, J., Compte, A., and Wang, X. J. (2002). The dynamical stability of reverberatory neural circuits. *Biol. Cybern.* 87, 471–481. doi: 10.1007/s00422-002-0363-9

Tseng, P., Chang, Y.-T., Chang, C.-F., Liang, W.-K., and Juan, C.-H. (2016). The critical role of phase difference in gamma oscillation within the temporoparietal network for binding visual working memory. *Sci. Rep.* 6:32138. doi: 10.1038/srep32138

Tsodyks, M. V., and Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl. Acad. Sci. U.S.A.* 94, 719–723. doi: 10.1073/pnas.94.2.719

van Vugt, M. K., Schulze-Bonhage, A., Litt, B., Brandt, A., and Kahana, M. J. (2010). Hippocampal gamma oscillations increase with memory load. *J. Neurosci.* 30, 2694–2699. doi: 10.1523/JNEUROSCI.0567-09.2010

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

Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol. Rev.* 90, 1195–1268. doi: 10.1152/physrev.00035.2008

Wimmer, K., Ramon, M., Pasternak, T., and Compte, A. (2016). Transitions between multiband oscillatory patterns characterize memory-guided perceptual decisions in prefrontal circuits. *J. Neurosci.* 36, 489–505. doi: 10.1523/JNEUROSCI.3678-15.2016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neural Circuits

REVIEW
published: 30 June 2021
doi: 10.3389/fncir.2021.679796

Revisiting Persistent Neuronal Activity During Covert Spatial Attention

Julian L. Amengual and Suliann Ben Hamed**

Institut des Sciences Cognitives Marc Jeannerod, CNRS UMR 5229, Université Claude Bernard Lyon I, 67 Boulevard Pinel, Bron, France

Persistent activity has been observed in the prefrontal cortex (PFC), in particular during the delay periods of visual attention tasks. Classical approaches based on the average activity over multiple trials have revealed that such an activity encodes the information about the attentional instruction provided in such tasks. However, single-trial approaches have shown that activity in this area is rather sparse than persistent and highly heterogeneous not only within the trials but also between the different trials. Thus, this observation raised the question of how persistent the actually persistent attention-related prefrontal activity is and how it contributes to spatial attention. In this paper, we review recent evidence of precisely deconstructing the persistence of the neural activity in the PFC in the context of attention orienting. The inclusion of machine-learning methods for decoding the information reveals that attention orienting is a highly dynamic process, possessing intrinsic oscillatory dynamics working at multiple timescales spanning from milliseconds to minutes. Dimensionality reduction methods further show that this persistent activity dynamically incorporates multiple sources of information. This novel framework reflects a high complexity in the neural representation of the attention-related information in the PFC, and how its computational organization predicts behavior.

Keywords: spatial attention, prefrontal cortex, mixed-selectivity, population activity, decoding, neurophysiology, persistent activity, alpha oscillations

OPEN ACCESS

Edited by:
Shintaro Funahashi, Kyoto University, Japan

Reviewed by:
Julio Martinez-Trujillo, Western University, Canada
Liya Ma, Radboud University Nijmegen, Netherlands

****Correspondence:***
Julian L. Amengual
Julian.amengual@isc.cnrs.fr
orcid.org/0000-0002-8613-6325
Suliann Ben Hamed
benhamed@isc.cnrs.fr
orcid.org/0000-0003-1510-7284

Received: *12 March 2021*
Accepted: *03 June 2021*
Published: *30 June 2021*

Citation:
Amengual JL and Ben Hamed S (2021) Revisiting Persistent Neuronal Activity During Covert Spatial Attention. Front. Neural Circuits 15:679796. doi: 10.3389/fncir.2021.679796

INTRODUCTION

Numerous studies report an increase of spiking activity in different brain areas during the performance of visual delayed tasks [see Fuster and Alexander (1971), Goldman-Rakic (1995), Shafi et al. (2007), Barak et al. (2010), Watanabe and Funahashi (2014), Chaudhuri and Fiete (2016), Zylberberg and Strowbridge (2017), Manohar et al. (2019), for a review]. The general structure of the tasks consists of the presentation of an informative visual cue about how the subject should act afterward. After the presentation, there is a delay period in which the subject must keep in mind the information provided by the cue to appropriately respond to the task demands. This information can be spatial (e.g., left vs. right), feature-based (e.g., blue vs. red), or symbolic (e.g., left-pointing arrow vs. right-pointing arrow).

Pioneering electrophysiological studies employing intracortical recordings in non-human primates have identified neurons that not only show activity associated with the sensory stimuli serving as a cue but also show activity in the delay period after the cue

when it is no longer present and the task instructions are being processed (Fuster and Alexander, 1971; Fuster, 1973; Funahashi et al., 1989; Miller et al., 1993). These findings have been extensively corroborated by using different protocols and techniques in humans and non-human primates (Constantinidis et al., 2018 for review). Classically, persistent activity in the prefrontal cortex (PFC) has been considered as a signature of specific cognitive processes such as working memory (Constantinidis et al., 2018). However, working memory interplays with other cognitive functions such as perception or attention. For example, previous studies have succeeded in discriminating perceptual and mnemonic representations of visual features (Mendoza-Halliday and Martinez-Trujillo, 2017a). How the interaction between working memory and attention is theorized depends on whether attention is conceptualized as the processing of a limited source of information (a perception of low-salience visual information) or the selection of information for processing [covert attention: Oberauer (2019)]. In the present review, we will focus on covert attentional processes defined as the *a priori* top-down selection and maintenance of the sensory information for prioritization (e.g., based on its spatial location), in anticipation of its presentation and processing. In this context, covert attention can be considered as an instance of working memory as the information needed to be prioritized, whether feature-based or spatial, is by definition sustained, i.e., held in working memory (Desimone and Duncan, 1995). We will describe the structure and informational content of the observed neuronal activity in the deployment of covert attention, and discuss it in relation to the current views on the dynamic and rhythmic nature of attention [see Gaillard et al. (2020), Gaillard and Ben Hamed (2020) for a review].

Persistent activity during visuospatial attention tasks has been reported both in parietal (Colby et al., 1996; Gottlieb et al., 1998; Ibos et al., 2013) and prefrontal cortices (Moore and Armstrong, 2003; Moore and Fallah, 2004). The tasks involve maintaining a sustained level of information relative to where (spatial attention) or what (feature-based attention) relevant task-related events will need to be processed (Posner and Petersen, 1990). Attention orienting can be driven by bottom-up or stimulus-driven processes, triggered by the salience of the incoming visual stimuli (i.e., their shape or color), and a top-down process that is guided by the relevance of the stimulus (i.e., how much it is useful to the task) defining our internal goals or expectations (Pinto et al., 2013; Katsuki and Constantinidis, 2014). Studies on humans have highlighted the importance of a frontoparietal network in the control of attention, showing the involvement of the parietal cortex and the PFC (Corbetta and Shulman, 2002). In macaque monkeys, the most commonly used model to study the attentional system in non-human primates, a homologous frontoparietal attention network is identified (**Figure 1A**), involving the lateral intraparietal (LIP) area (Gottlieb et al., 1998) and the frontal eye field (FEF; Armstrong et al., 2009; Monosov and Thompson, 2009). The two cortical regions are highly interconnected (Cavada and Goldman-Rakic, 1989; Stanton et al., 1995; Buschman and Miller, 2007; de Schotten et al., 2011; Ibos et al., 2013; Marek, 2018). Reversible inactivation of the two cortical regions results in behavioral impairments both in easy visual search tasks that rely on bottom-up attentional processes (Wardak et al., 2002; Wardak, 2006) and in conjunction with the search tasks that involve top-down attentional processes (Theeuwes, 1993). Although persistent activity has also been described in other regions, including the regions where it is more prevalent compared to the FEF and LIP (Leavitt et al., 2017), we focus on neuronal activity in these two regions in the context of persistent activity during attention orienting.

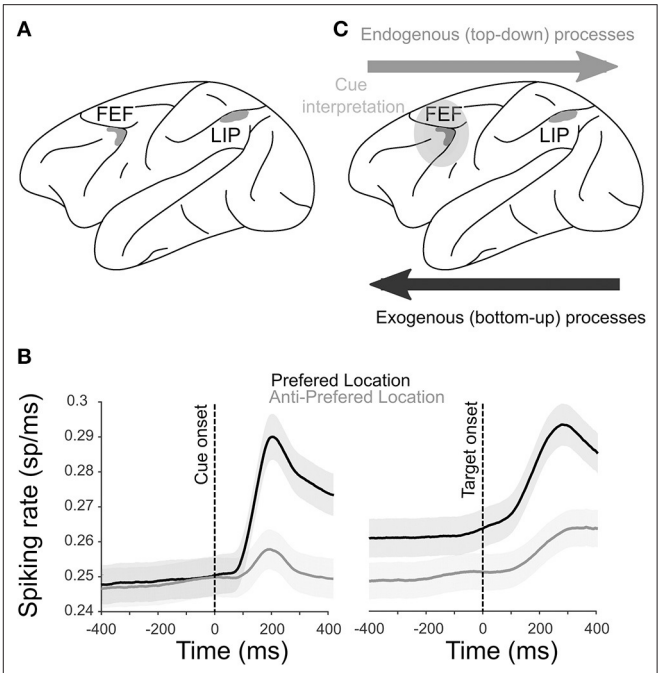


FIGURE 1 | Physiology of the attentional system. **(A)** Anatomical localization of the two core brain regions engaged during spatial attention orienting in the macaque [shaded in gray; see Cohen and Andersen (2002), Ibos et al. (2013), Paneri and Gregoriou (2017), for a review], the FEF and the LIP sulcus. **(B)** The average multiunit activity (MUA; ± s.e.) recorded from the FEF in one monkey when a cue is orienting attention toward the preferred (black) or the anti-preferred (gray) spatial location locked to cue (left) and target (right) onsets. *X*-axis represents the time around cue or target onsets. **(C)** Functional hierarchy in a frontoparietal network during attentional processes. Exogenous processes start in the LIP and project to the FEF. Cue interpretation takes place in the FEF, and the selection of the spatial location and subsequent endogenous processing is projected from the FEF to the LIP.

PERSISTENT NEURONAL ACTIVITY DURING SPATIAL ATTENTION ORIENTING

Persistent activity during sustained attentional processes is classically described in both the FEF and LIP. **Figure 1B** shows the average neuronal responses of a sample of FEF attention-related neurons recorded during the cue-to-target interval of a spatial attention task. A higher activation is observed when the cue is orienting attention toward the preferred spatial position of the neuron (black) compared to when attention

is oriented away from the preferred spatial position (gray). Preferred spatial positions coincide with both enhanced visual cue-related responses as well as enhanced visual target detection responses. Such neuronal response patterns are typical of both FEF and LIP neurons (Ibos et al., 2013). As a result, an important question in the field has been to understand whether parietal and prefrontal attentional responses were functionally identical or not. Simultaneous recordings from both cortical regions allow addressing this question. In the following sections, we will first review this question from the point of view of a single neuron persistent activity and then from the perspective of the neuronal population.

Attention-Related Persistent Responses in Single Neurons

Several studies have addressed the functional interactions between the PFC and the parietal cortex during attentional processes. In easy visual search tasks (e.g., detecting a red square among the green squares), which have been shown to rely on bottom-up attentional processes (Treisman and Gelade, 1980), parietal neurons are activated earlier than prefrontal neurons (Buschman and Miller, 2007). In striking contrast, in conjunction with the visual search tasks (e.g., detecting an orange vertical bar among the red vertical bars and red and orange horizontal bars), which have been shown to rely on top-down attentional processes (Treisman and Gelade, 1980), the reverse is observed (Buschman and Miller, 2007). This suggests that spatial attention or spatial selection mechanisms flow from the parietal cortex to the PFC and the PFC to the parietal cortex when driven by the environment and the subject's internal goals, respectively. However, visual search tasks do not allow researchers to dissociate the neuronal processes related to attention orientation from those related to perceptual cue processing. In order to address this limitation, Ibos et al. (2013) designed a task that allows to temporally dissociate between cue processing, cue interpretation and attention orientation, and target selection. This task had two features (**Supplementary Figure 1**). It was based on a modified version of a rapid serial visual presentation (RSVP) task (Potter, 2018) such that, on each trial, the cue and the target are embedded in two parallel continuous streams (succession) of isoluminent distractors. In such a context, both parietal and prefrontal neurons do not respond to the visual transients between one visual stimulus and the next. Thus, any specific enhancement of neuronal responses to the cue or to the target or in between the cue and the target can be interpreted as an attention orientation signal or a perceptual signal. The second specificity of this task lies in the fact that the attentional orientation cues are highly symbolic. The green cues indicate that the target will appear in the same visual stream as the normal cue while the red cues indicate that the target will appear in the opposite visual stream. In other words, both the left red cues and right green cues oriented attention to the right while both the right red cues and left green cues oriented attention to the left. While both parietal and prefrontal neurons showed an enhanced processing of the cues and targets embedded in the RSVP streams, the cue-related responses had shorter latencies in the parietal cortex than in the PFC, and the target-related responses had shorter latencies in the PFC than in the parietal cortex. Thus, this confirms the idea that spatial selection mechanisms flow from the parietal to the PFC and the PFC to the parietal cortex when driven by the environment and by the subject's internal goals, respectively (**Figure 1C**). In addition, neurons explicitly encoding the instruction for the spatial attention orientation independently of the color and location of a cue were only identified in the PFC and had longer response latencies than the cue-related parietal responses, indicating that the attentional cue interpretation was performed within the PFC (**Figure 1C**). Overall, this thus defined a clear hierarchical functional organization within the parietofrontal network in which the processing of high-saliency stimuli initiates in the LIP; and the active attention orientation control according to the subject's goals takes place in the FEF, thus driving a perception of low-saliency stimuli.

As shown by Ibos et al. (2013) and with relevance to the present review, the prefrontal attention orientation neurons encoded the attention instruction in a sustained manner. This was also the case of a substantial proportion of the cue-related neurons of both cortical regions that responded to one specific category of cues such as cue color or cue position. Thus, these neurons are also expected to contribute to the coding of attention orientation instructions when combined across the population. The fact that FEF neurons explicitly encode the cue instruction suggests functional differences of how both the FEF and LIP represent a spatial orientation signal in the population level and sustain these representations in time. It has been hypothesized that the ability of individual neurons in a recurrent neuronal network to sustain the information over time depends on the correlated fluctuations of activity within the local neuronal microcircuitry (Maimon and Assad, 2009). The recurrent fluctuations of neuronal activity occur over a wide range of timescales depending on the local properties of the brain region (Murray et al., 2014). To measure the timescales of these fluctuations, the time lag autocorrelogram of the spike count of individual neurons is calculated. As this time lag increases, the autocorrelation decays as a function of the fluctuation timescales (Churchland et al., 2011). The timescale of these fluctuations is mathematically characterized by the decay of autocorrelation as a function of the time lag (τ), which corresponds to the fitting of the autocorrelogram with an exponential decay and an offset. The intrinsic timescales differ across the brain areas, showing shorter timescale values in the parietal cortex and longer timescale values in the PFC (Murray et al., 2014). The observation points in the direction of favoring a temporal hierarchical organization between the parietal and the PFC (Murray et al., 2014). One question is whether and how this impacts the functional coding of the neuronal populations as a whole. This is explored in the next section.

Population Activity

Single-neuron responses support an idea of the sustained/persistent neuronal activity (quantified as the

sustained average spiking rate in time when computed across the trials) during delay epochs in cognitive tasks. What is the nature and origin of this "persistence" of persistent activity? Classical models of persistent activity propose that the observed spiking activity of a given neuron during the delay periods of a task is a marker of an active state of the neural ensemble it belongs to, which keeps the neural population available for processing the encoded information (Lundqvist et al., 2018a for a review). However, a closer inspection of the single neuron persistent activity reveals that a spiking activity is better characterized by sparsity than by persistence, suggesting that the information might be held within the local functional network by the changes in synaptic weights rather than in the spiking activity *per se* (Zucker and Regehr, 2002; Lundqvist et al., 2016). It is proposed that this type of information encoding during the delay period might be more long-lasting and more resistant to the disruption by additional inputs compared to a purely persistent spiking activity (Lundqvist et al., 2018b). This observation has been supported by the computational models that predict the sustainability of persistent activity through virtue of recurrent connections between the neurons that have an affinity for shared specific stimulus properties (Compte, 2000; Compte et al., 2003). However, it must be acknowledged that the absence of persistent activity can also be accounted for by analytical and experimental biases. For instance, one must take into account that each cortical neuron receives inputs from several other neurons (up to several thousands), which might cause a high response variability at small timescales but less so at longer timescales. Corroborating this view, Leavitt et al. (2017) show evidence of persistent activity during a working memory task using the temporal scales larger than 400 ms whereas the smaller timescales did not show such an effect. Another possible reason for the absence of persistent activity is the use of single-cell approaches, which, during the mapping of a given cortical area, might miss a specific region in which the persistent activity takes place. In this respect, it is, however, worth noting that sparsity in the spiking activity has been observed in the studies by using dense multielectrode recordings.

Thus, one of the implications of the above framework is that persistent activity is best understood in the level of the neuronal population rather than in the level of individual neurons. This view assumes, among other things, that information coding cannot be unambiguously read out (i.e., decoded) from the spiking rate of a single neuron, but is best characterized by the patterns of activation and connectivity across a population of neurons. That is to say, quoting from Averbeck et al. (2006), "*As in any good democracy, individual neurons count for little; it is population activity that matters.*" Accordingly, there is increasing evidence that individual neuronal response profiles do not fully mirror the dynamics of the functional neuronal population they belong to and that the dynamics of the connection weights between the individual neurons must be taken into account (Barak et al., 2010; Crowe et al., 2010; Stokes et al., 2013).

When using the cued target detection tasks to orient attention, it is often assumed that attention is behaviorally allocated in a stable and sustainable manner in the cued location. Likewise, it is assumed that the neurons that encode this information do so in a stable manner, i.e., with a constant number of spikes in time. However, the evidence points that neither of these assumptions might be correct (see the next section). One way to assess code stability in time is to use cross-temporal decoding approaches (King and Dehaene, 2014; Astrand et al., 2015; Varoquaux et al., 2017). These approaches assume that the attention orientation code is implemented by the neuronal population locked to the cue presentation. Thus, a decoder is trained at identifying whether attention is oriented toward one among multiple spatial locations (e.g., left vs. right) based on the neuronal population activities collected in a given (typically short, ∼100 ms) time interval at a fixed delay from the cue presentation. The decoder is then tested at decoding attention orientation on the novel activities sampled all throughout the cue-to-target interval. If the decoder maintains a high decoding performance at all times (**Figure 2A**), then this indicates a stable code. Alternatively, if the decoder only achieves maximal decoding from the neuronal activities sampled at the same delay from the cue as the training activities, then this indicates a dynamic recurrent coding of attention orientation (**Figure 2B**). This means that the same cascade of neuronal activities unfolds throughout the cue-to-target activity from one trial to the next, reliably encoding spatial attention at each time but with a different neuronal code. Both a stationary regime and a dynamic regime can coexist in a given neuronal population, leading to a mixed cross-temporal decoding map (**Figure 2C**). Using a regularized linear regression classifier as a decoder (Astrand et al., 2014a), training on 70% of the available trials and testing it on the remaining 30% over multiple random draw repetitions (Ben Hamed et al., 2003), Astrand et al. (2015) show that, in the PFC, the neuronal population composed of the attention orientation cells represents the spatial attention orientation in a sustainable manner (**Figure 2A**). The entire task-related FEF neuronal population expresses a mixed cross-temporal decoding map, suggesting the combination of both stationary and dynamic processes (**Figure 2C**). Thus, this indicates that the functional characterization of the individual neuronal responses does not fully account for how the information is encoded in a given area. In contrast, the parietal neuronal population expresses a highly dynamic coding of the spatial attention orientation (**Figure 2B**). In other words, a parietal code for the spatial attention orientation changes from one time to the next.

Additionally, in the PFC, the mixed attention orientation coding coexists with a stationary code for a cue position and a highly dynamic code for a cue color. This indicates that a given functional neuronal population can concurrently code different sources of information in multiple coding regimes, respectively. In other words, the information is multiplexed, and the system is able to simultaneously process (multiplex) the different driving inputs that involve different neuromodulatory sensitivities and synchronization influences (Liu and Hou, 2013; Feng et al., 2014). At this point, it is important to disambiguate multiplexing of information from coding generalization. Whereas, the first refers to the ability of the neural population to simultaneously

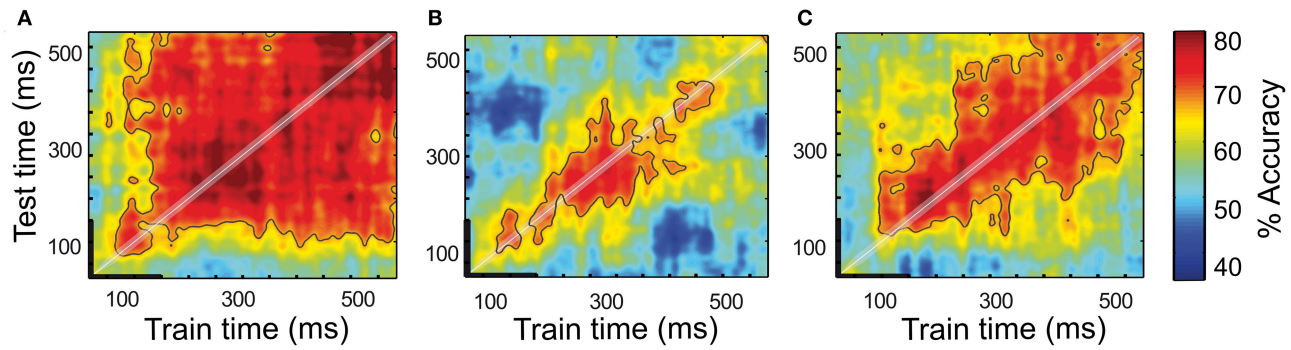


FIGURE 2 | Temporal dynamics of a spatial attention signal. Full cross-temporal classification analysis of the attention-specific FEF population **(A)**, entire LIP population **(B)**, and entire FEF population **(C)**. Classifiers are trained to classify spatial attention from the population activities at every time step within 600 ms following a cue onset and prior to a target onset (*x*-axis, thick black line corresponds to the cue presentation). The performance of each classifier is tested from the population activities obtained from naïve trials at every time step within the same time period relative to the cue. Red colors represent high performance whereas the blue-colored regions represent low performance [chance level at 50%, figure adapted from Astrand et al. (2015)].

encode the different sources of information in multiple coding regimes, the second refers to the ability to decode the different sources of information by using the same code. Prior studies have shown evidence that the PFC neural population codes associated with one specific source of information do not fully generalize (Tremblay et al., 2015b; Mendoza-Halliday and Martinez-Trujillo, 2017b).

In a given cortical region, the specific neuronal coding regime (stationary vs. dynamic) might fully depend on the information to be encoded (e.g., attention orientation would always be encoded in a stationary manner while the color in a dynamic manner), as an intrinsic property of the neuronal population. Alternatively, this could actually be task-dependent (e.g., attention orientation would be encoded in a stationary manner in the cued target detection task, but dynamically in a spontaneous visual exploration task). This remains to be tested. Likewise, how these cortical areas read out these multiple codes and exploit them is a topic of future research. Multiple mechanisms might be at play. For example, similar to the previous working memory studies (Fujisawa et al., 2008; Mongillo et al., 2008), Astrand et al. (2015) propose that the active mechanisms that sustain the attentional information in the neural population might involve short-term plasticity mechanisms. In contrast, the constant inputs to the neuronal population might result in time-dependent response patterns if the synaptic weights that represent the connectivity across the neurons are continuously being changed by the influence of the input pattern of activity (Buonomano and Maass, 2009).

In this section, we have shown that whereas individual neurons in the PFC show a persistent average activity and the underlying neural population encodes the information using different regimes spanning from fully stationary to dynamic and mixed. This calls for a reinterpretation of the persistent activity at a single-neuron level during spatial attention orienting. In the next section, we show that, at the single-trial level, neither the single-neuron responses nor the neuronal population information is persistent.

IS PERSISTENT ATTENTION-RELATED INFORMATION ACTUALLY PERSISTENT?

Single Trial, Spatially, and Temporally Resolved Access to Attention Selection Signals

The use of classification procedures to decode the brain activity associated with specific aspects of human behavior forms the basis of one of the greatest technological achievements in neuroscience for the last two decades, namely brain–computer interfaces (BCIs) (Chapin et al., 1999; Wolpaw et al., 2002). These methods are based on the use of simultaneous neuronal population activities from a given cortical region in order to drive the devices that can help patients with specific dysfunctions or deficits to improve their quality of life. Most BCI technologies are designed to address motor-related dysfunctions such as motor prosthesis or driving external palliative devices such as cursors (Trejo et al., 2006) or robotic arms (Sunny et al., 2016), among others. Little research has been directed to develop the BCI devices that rely on decoding higher-order cognitive processes such as attention (Andersen et al., 2010; Astrand et al., 2014b), due to the fact that such a cognitive content is internally generated by the neuronal signals that are often multiplexed with different types of information, including sensory and motor information. This renders their real-time access very challenging.

Non-human primate studies addressing this question have specifically targeted the cortical regions in which spatial attention has been shown to be sustained, thus favoring the PFC over the parietal cortex. Astrand et al. (2014a) first demonstrated a single-trial left/right attention classification for comparing multiple classifiers. Tremblay et al. (2015a) extended these observations to a four-quadrant classification of attention. Astrand et al. (2016) push this decoding procedure one major step forward, introducing the highly spatially resolved (x, y) tracking of the attentional spotlight [i.e., the actual portion of space being selected (Posner and Petersen, 1990)], at a spatial resolution of the order of 0.1° (see **Supplementary Figure 2** for a description

of the methods). Specifically, a regularized optimal linear estimator is used to associate the recorded bilateral response patterns produced during the correct target detection trials shortly before a target onset with the cued two-dimensional (x, y) spatial location of attention. This decoder is used to predict the (x, y) location of the attentional position inferred from the bilateral prefrontal response patterns recorded in novel trials naïve to the decoder. Decoding is applied at multiple time steps, thus allowing to track the attentional spotlight in time during the cue-to-target period. Using this methodological approach, attention could be decoded everywhere on the workspace during the cue-to-target presentation interval, and not necessarily static at the cued location prior to the target presentation (Astrand et al., 2016, 2020). This corresponds to the distinct neuronal response patterns on successive trials. Therefore, although the average neuronal response patterns might seem to be sustained (**Figure 3A**), on individual trials, spiking probability varies from one trial to the next (**Figure 3B**), corresponding to a different attentional exploration trace from one trial to the next (**Figures 3C,D**).

Importantly, Astrand et al. (2016) show that attention had a higher probability of being closer to the target on the correct target detection trials than on the trials in which the target is missed. Likewise, attention had a higher probability of being closer to a distractor when a false response to the distractor (as opposed to being closer to the target) was performed. This observation was further confirmed on other tasks (Di Bello et al., 2020; Gaillard et al., 2020). In this context, De Sousa et al. (2021) further enhance a correlation between the decoded attention and overt behavior using a novel two-step decoder, the essence of which is to refine the decoder training on only those trials that were initially identified as trials in which attention was oriented close to the target. All in all, the abovementioned studies confirm the association between the decoded readout of spatial attention and the observed task-related behavior of the subject. Despite the clarity of this relationship between the distance of the decoded attention to the real target (or distractor) position and the probability to respond to a target (or distractor), one intriguing question is why trials that are characterized by attention decoded at a similar distance from a target (or distractor) sometimes result in a correct detection (or a false alarm) and other times in a miss (or distractor rejection). Inter-neuronal correlations turn out to be significantly lower on the correct trials than on the miss or false alarm trials, suggesting that error trials might arise when the neuronal population is in a lower informational capacity state characterized by higher noise correlation values (Astrand et al., 2016; Ben Hadj Hassen and Ben Hamed, 2020), which will be explored in Section Noise correlation and neuronal population information capacity. Overall, this work thus demonstrates that, from one trial to the next, although attention is often assumed to be stable at the cued location, it is not quite often. Rather, attention explores space dynamically, shifting from one location to the next every 100 ms or so. Because the attentional dynamics is revealed through the decoding of attention-related information from population neuronal activity, this indicates that the attentional dynamics is subserved by rapid changes in the spiking rates of individual neurons, during the attention orienting delay, as shown in **Figure 3C**. Thus, while stable attention was generally assumed to be subserved by persistent neuronal responses, this section demonstrates that attention is dynamic and is subserved by dynamic neuronal responses and not by persistent neuronal activity.

Attention Explores the Space Rhythmically

Classically, the spotlight theory of attention assumes that attention is only focused at one location of space at a time [Eriksen and St. James, 1986; see the discussion in Posner and Petersen (1990), Gaillard and Ben Hamed (2020)]. This view posits that it is possible to shift the spotlight of attention from one location to another, independent of the eye position and adjustment of its size to the attended location like a zoom lens. Thus, it intrinsically assumes a certain degree of flexibility of attention. Recent behavioral evidence (Venables, 1960; Landau and Fries, 2012; Dugue and VanRullen, 2014; Song et al., 2014) shows that, instead of a smooth and continuous behavior, spatial attention samples the visual environment rhythmically, leading to fluctuating periods of perceptual sensitivity [see VanRullen (2016) for a review]. In other words, these studies suggest that attention and perception might not be attached to a specific location in space (e.g., the cued location), but rather exhibit a temporal rhythmicity between relevant spatial locations [but see Brookshire (2021) for a critical perspective on these observations].

In agreement with these behavioral studies, neurophysiological evidence indicates that the brain activity underlying visual attention, as measured from the local field potentials (LFPs), is rhythmic in the theta band (4–8 Hz) (Lakatos et al., 2008; VanRullen, 2013, 2016; Fiebelkorn et al., 2018; Spyropoulos et al., 2018). For example, Fiebelkorn et al. (2018) show that, in the execution of a cued detection task, monkeys' ability to detect a target fluctuates rhythmically as a function of the time from cue onset, at a rhythm of 4 Hz. Importantly, the likelihood to correctly respond to the target is predicted by the phase of the ongoing oscillations in the prefrontal LFPs in the same frequency band with respect to the cue onset.

Fluctuations in the behavioral attentional performance and in the prefrontal LFP power are in sharp contrast with the notion of a stable prefrontal attentional code following a cue orientation (Astrand et al., 2015). In order to directly address this question, Gaillard et al. (2020) extend the work by Astrand et al. (2016) to a temporally highly resolved decoding of spatial attention (over 50 ms neuronal recording windows, instead of 150 ms). At this temporal resolution, rhythmic fluctuations in the prefrontal attentional information are observed in the 7–12 Hz alpha range. As described in **Figure 3**, these attentional oscillations are not observed at an individual cell level; however, they become apparent when averaging on multiple simultaneously recorded signals (**Figure 3C**). The rhythmic fluctuations in the prefrontal attentional information are decoded as spatio-temporal (x, y) attentional traces, and systematic changes in the location of the decoded attentional spotlight (or attentional saccade) can be seen at a frequency of ∼8 Hz. These traces clearly show that, during the cue-to-target interval, attention explores both the

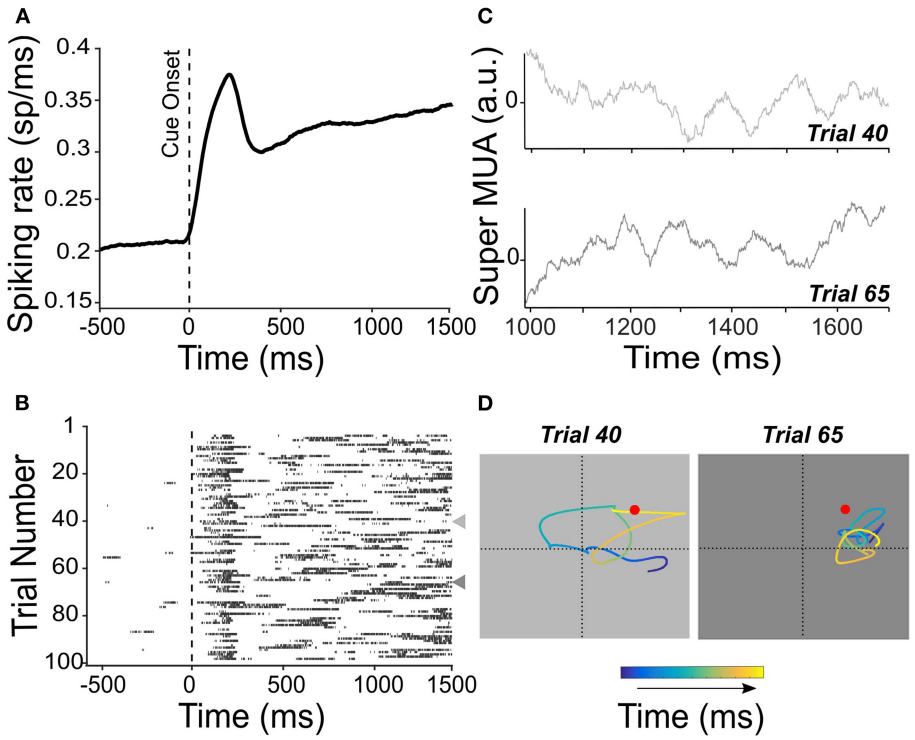


FIGURE 3 | Persistent activity is not persistent from one trial to the next. **(A)** Mean spiking rate activity across 100 trials is recorded from the FEF. Activity is locked to the cue presentation in the preferred location of the neuron. The spiking rate shows an average and sustained increase during the cue-to-target interval. **(B)** Raster plot showing the multiunit activity (MUA) locked to the cue onset for each of the 100 trials used for the average spiking rate shown in **(A)**. Represented individual trial spiking probability sets at a threshold of 65%. Trials 40 and 65 are indicated by a leftward-pointing gray triangle to show a different pattern of temporal activation. These trials are also considered in **(C,D)**. **(C)** The average MUA activity was obtained by averaging the activity recorded from all channels of the same electrode [super MUA, left FEF see Gaillard et al. (2020)] for trials 40 and 65 [indicated in **(B)**]. **(D)** Attentional traces were obtained from high-resolution *x*-, *y*-coordinate decoding of the position of the attentional spotlight from the population neuronal activity, for trials 40 and 65, during the cue-to-target interval. Color code shows the time point along the cue-to-target interval of the decoded position of attention (blue to yellow, 700 ms). Red dot indicates the position of the target stimulus in these trials.

cued locations but also uncued spatial locations. Importantly, this spatial exploration of space by attention also exists prior to attentional cueing, suggesting that the rhythmicity of attention is a default mode. In addition, how the prefrontal attentional trace explores the space varying from one behavioral task to another, indicating that it is under a top-down control. Overall, Gaillard et al. (2020) propose that the rhythmic variations in the attentional exploration subtend an efficient compromise between the exploitation of the prior information and the exploration of the novel information within a given trial.

All these abovementioned studies work under the assumption that the attentional spotlight is unique, a paradigm that has been driving most neurophysiological studies (Moran and Desimone, 1985; Niebur and Koch, 1994; Lee et al., 1999; Martínez et al., 1999; Reynolds et al., 2000; Corchs, 2002). However, this model of attention is limited when one needs to attend to more than one object at a time. In this context, other models of attention have been proposed. One of them is the zoom lens hypothesis, which considers a single attentional spotlight that is able to select the information from multiple locations by adjusting its size (Eriksen and Yeh, 1985; Eriksen and St. James, 1986). Another model proposes that the spotlight can be split, and attention may be simultaneously deployed to multiple spatial regions (Awh and Pashler, 2000; McMains and Somers, 2004; Niebergall et al., 2011; Mayo and Maunsell, 2016). Of utmost interest and relevance, current neurophysiological paradigms do not allow a direct evaluation of these concurrent theoretical models of attention.

All this taken together support the idea that seemingly persistent prefrontal single neuronal and population activity is actually highly dynamic, reflecting complex ongoing endogenous (i.e., covert) processes. These dynamic processes can only be accessed at the single-trial level because they (and their specific associated informational content) vary from one trial to the next. When averaged, these trial-to-trial variations are wiped out.

STATE DEPENDENCE OF PREFRONTAL NEURONAL ACTIVITY

In the previous section, we address the sources of neuronal response variability that correlate with the dynamic nature of attention at the single-cell level and at the population level. In contrast, in this section, we consider the neurophysiological markers that impact the neuronal population information

capacity and overt behavior irrespective of individual neuronal spiking rates and also irrespective of whether neuronal activity is persistent or not. We first discuss the noise correlation across the neuronal population and how it impacts the neuronal population informational capacity. Then, we discuss global fluctuations in the neuronal population attentional information that occur irrespective of ongoing attentional processes but directly impact both attentional neuronal responses and behavior.

Noise Correlation and Neuronal Population Information Capacity

Noise correlations have been shown to critically impact both cortical signal processing and behavioral performance in different domains such as learning and attention (Ben Hadj Hassen and Ben Hamed, 2020). Shared neuronal variability across all recorded neurons is independent of the shared neuronal variability induced by the signal (Ben Hadj Hassen et al., 2019). The accuracy of a population code depends on not only the neuronal correlation arising from a common input (such as sensory information or cognitive control information) but also on the neural correlations that arise from a stimulus-independent activity. Indeed, a noise correlation is shown to interfere with the informational capacity of neuronal populations to represent a given variable and the resilience of this neuronal population to noise interference [see Averbeck et al. (2006), for a review]. For example, Froudarakis et al. (2014) show that the less correlated the firing pattern in V1 neurons, the higher the discriminability of the population code between the different visual stimuli. Likewise, lower noise correlations have also been associated with more efficient memory storage (Olshausen and Field, 2004). However, the relationship between the noise correlation and informational capacity is not straightforward. Indeed, it has been shown that inter-neuronal noise correlations can either improve the overall informational capacity, and hence decoding accuracy or, on the opposite, degrading decoding accuracy mostly depending on how the strength of noise correlations is compared to the strength of signal correlations (Averbeck et al., 2006; Moreno-Bote et al., 2014; Ben Hadj Hassen and Ben Hamed, 2020). How much decoding benefits from the decorrelated neuronal activities thus depends on a variety of experimental and neurophysiological factors (Ben Hadj Hassen and Ben Hamed, 2020).

Astrand et al. (2016) show that the noise correlation in prefrontal neuronal populations is predictive of the overall behavioral performance, which is lower on upcoming correct trials than on upcoming misses or false alarms. They further show that the fluctuations in noise correlations are very slow as noise correlations are globally lower on a given trial either on correct or error trials when the previous trial was a correct trial. In contrast, noise correlations are globally higher in a given correct of error trial when the previous trial was a miss trial. This strongly indicates that more global mechanisms are mediated among other things by a noise correlation, interact with spatial attention processes, and significantly contribute to overt behavioral performance. Importantly, the slow fluctuations in noise correlations are independent of variations in the overall spiking level, confirming that they reflect the state of connectivity of a given neuronal population.

Fluctuations in a noise correlation (and thus in the overall population informational capacity) are also observed at slower timescales than at a trial level. Indeed, Ben Hadj Hassen et al. (2019) show that noise correlations are lower on difficult tasks as compared to easy tasks. This suggests that an active mechanism might contribute to adjusting the neuronal noise correlation to ongoing behavioral demand, thus high-noise correlation states corresponding to a default "relaxed" population state.

In addition, fluctuations in a noise correlation are also characterized at faster timescales. For example, Ben Hadj Hassen et al. (2019) show that prefrontal noise correlations fluctuate within two distinct frequency bands, a high alpha frequency range (10–16 Hz) and a beta frequency band (20–30 Hz). These fluctuations that are independent of fluctuations in neuronal spiking rates are shown to impact behavioral performance and are reproduced in three different behavioral tasks. The authors propose that selective changes of frequency in spike-LFP phase coherence might account for these fluctuations in a noise correlation. Likewise, Womelsdorf et al. (2012) show the fluctuations in V1 noise correlation at an even higher gamma frequency (60–80 Hz), correlating both with the changes in performance and with orientation selectivity as a function of the phase in the gamma cycle.

Overall, the results suggest that noise correlations vary at different timescales, from a very slow to fast, suggesting fluctuations in the overall neuronal population capacity in the same timescale. This is actually confirmed by the observation that the variations in behavioral performance correlate with the variations in noise correlation. The studies cumulatively indicate that the information capacity in a given neuronal population is not only determined by spiking patterns, as described in the previous section but also by inter-neuronal noise correlations, a neurophysiological metric, which is decoupled from the firing rates and still anticorrelated with the attentional information and fluctuates in time in multiple scales. Thus, this further weakens the link between attentional processes and persistent activity.

However, other statistical features of the neural population are also reported to impact the amount of encoded information, such as changes in the network state, neuronal tuning, and global activity modulations (Cohen and Newsome, 2008; Harris and Thiele, 2011; Gutnisky et al., 2017; Verhoef and Maunsell, 2017). In addition, there is no consensus on whether the statistical features of population responses that affect the amount of information encoded in the neural populations also impact behavior (Arandia-Romero et al., 2017; Panzeri et al., 2017). In a very recent paper, Nogueira et al. (2020) have investigated which features of neural population responses most determine the overall amount of encoded information and behavioral performance. Examining neurons

in two different brain areas (the middle temporal area and the lateral PFC), they found that the amount of information encoded in a population and behavioral performance was highly determined by the two statistical features: (1) the length of the vector joining the mean population responses in different experimental conditions [population signal (PS), corresponding to the distance, in lower-dimensional space, between the neuronal response patterns in different conditions] and (2) the inverse population co-variability projected onto the direction of the PS vector [projected prevision (PP), corresponding to the degree of alignment between the low-dimensional representation of the neuronal responses of each experimental condition]. Importantly, keeping the two parameters fixed, the authors did not find a clear relationship between the noise correlation and the amount of encoded information; however, they found a covariation between the latter parameter with PP and PS that could explain the observed effects of noise correlation in the amount of encoded information in the prior studies.

Very Slow Fluctuations in Prefrontal Information Capacity

Until now, we have shown that prefrontal activity during the processing of the attention information is highly dynamic, showing rhythmic fluctuations in the attentional information in the alpha range (∼10 Hz). The fluctuations are associated with a behavioral outcome of the subject, shedding new light on how the attentional system holds the information in a short timescale. However, little is known about the dynamics of the attention information in longer timescales (in the range of minutes and even hours). In this context, previous studies have shown that when attention is actively sustained in time, such as in the context of long-lasting cognitive demands, and the performance seems to decrease (Proctor et al., 1996; Lockley et al., 2004; Bonnefond et al., 2010; Virtanen and Kivimäki, 2018). A recent work by Gaillard et al. (2021) suggests that this might not always be the case. Indeed, they report that behavioral performance in a visual attentional task fluctuates by up to 10% at an ultra-slow rhythm of 4–7 cycles per hour (every 9–15 min), coinciding with phase-locked rhythmic fluctuations in the accuracy of visual and spatial attention information in the PFC. The behavioral and neuronal information fluctuations were not associated with concurrent variations in the spiking rate. However, an enhanced theta (∼6 Hz) and beta (∼24 Hz) oscillatory activity in LFP and an enhanced alpha (∼10 Hz) in LFP coherence were observed during high behavioral performance epochs. Overall, this thus adds a level of complexity to prefrontal activity, in particular during cognitive processing (spatial attention delays), as prefrontal attentional population coding appears to be impacted by long-range distal signals (possibly related to states of vigilance and/or of fatigue and energy depletion), shifting from a high processing efficiency state (associated with enhanced visual and attentional coding accuracies), and a low processing efficiency state (associated with degraded visual and attentional coding accuracies).

PREFRONTAL NEURONAL POPULATION ACTIVITY REFLECTS MULTIPLE PROCESSES

Prefrontal Cortical Population Activity and Mixed Selectivity

We have already described the different population activity regimes that were region-specific but also dependent on the source of encoding information (e.g., position or color of the cue, Section 2.2). Prior studies have demonstrated a specific property of PFC neurons (specifically, the neurons from area 46 in the lateral PFC) called mixed selectivity (Rigotti et al., 2013; Parthasarathy et al., 2017). This property, which has also been reported in the FEF (Brincat et al., 2018; Khanna et al., 2020), allows that the neurons exhibit complex patterns of responses reflecting simultaneously different task-related parameters. Due to the complex functional pattern of activation, single-neuron recording studies on the PFC have found difficulties in relating the parameters to a specific neural activity, since the neurons will encode multiple parameters simultaneously, and the given spiking rate cannot unambiguously be assigned to the specific state of a given function. Approaches based on the average activity from the pre-selected neurons based on the specific criteria across multiple trials have been extensively used as a state-of-the-art in multiple neurophysiological studies (e.g., Ibos et al., 2013). However, these approaches, even though useful in identifying some specific information processes, elude most of the structure of the single-cell responses (Wohrer et al., 2013). This is because complex patterns of behavior might rely on the coordination of different neural mechanisms at a population level rather than on the activity of single neurons. In this context, the analysis of the neural population as a whole allows the extraction of features in the data using dimensionality reduction methods [see Cunningham and Yu (2014), for a review]. One of the methods is the principal component analysis (PCA), which consists of extracting an ordered set of orthogonal directions capturing the greatest variance in data. An important caveat of this method is that the obtained low-dimensional space captures all types of variances—without unmixing the underlying sources. Therefore, mixed selectivity remains in the data after the reduction of dimensionality, preventing from associating a task- or even behavior-specific variance to individual components. This issue has recently been solved by addressing dimensionality reduction methods with explicit information about the variance related to the parameters (Machens, 2010; Mante et al., 2013; Kobak et al., 2016). Specifically, demixed PCA (dPCA; Machens, 2010) is a dimensionality reduction method that aims to decompose the data into features easily interpretable with respect to specific parameters while preserving the original data as much as possible (Kobak et al., 2016).

Unmixing Spatial Attention and States of Inattention (or Attentional Lapses) From Prefrontal Population Activity

One open question in the attention research is to what extent the readout of the attention information fully accounts for

the reported behavior of the subject. Previously, it has been shown that the position of attention with respect to the actual position of the stimulus to be processed accounted for behavior, such that the closer the decoded attentional spotlight to the stimulus (or distractor) prior to the stimulus onset, the more likely the behavioral response to this stimulus (or distractor) (Astrand et al., 2016, 2020). However, these studies show that in trials with a similar distance between the decoded position of attention and the actual cued position, the behavioral outcome could be different, the subject sometimes producing a correct response, and other times producing an error response. This suggests that, on top of the attentional dynamics, different neural states of activity might influence how the system is able to exploit the attention information. In a very recent study, Amengual et al. (2021) isolate, from the PFC population activity, components specifically associated either to the position of the decoded attentional spotlight relative to the expected target position or to the behavioral outcome (hit vs. miss) using dPCA. They consistently find that the components encoded the specific information from each parameter, respectively (attention and reported behavior). Interestingly, they find that the information about the two components partially overlapped (they are not orthogonal), the smaller the overlap the higher the behavioral gain associated with an efficient attention orientation. In other words, the smaller the overlap, the lower the interference at the behavioral level between the spatial attention orientation and the state of inattention encoded by the prefrontal neuronal population.

The results shed new light on the extent to which the system is able to use this information to optimize behavior. It suggests that an accurate performance involving an active engagement in an attentional task depends not only on the active attentional control and readout of the attended information but also on its integration with the activity associated with more general neural states that might correspond to levels of distractibility or impulsivity that allow access to the attended information. In addition, the results call for a functional reconsideration of persistent activity. Indeed, the multiplexing of the multiple states or features in a single population results in an apparent sustained activity. However, the precise informational content of this persistent activity can only be accessed by splitting it into well-defined functional components.

CONCLUSION AND PERSPECTIVES

Electrophysiological studies employed for recording individual cells of the primate PFC have shown clear evidence of persistent spiking activity for visual delay tasks associated with different aspects of cognition. In the present work, we have reviewed the role of the so-called persistent activity in the domain of attention orienting during the delayed visual attention tasks. In this context, classical approaches in the field mostly based on the analysis of single-cell recordings in the FEF- and LIP-averaging neuronal activity across multiple trials have shown that the sustained neuronal spiking activity during the cue-to-target time interval depends on the spatial preference of the cell, being higher when attention is located in the preferred spatial position of the neuron from both areas. However, the "persistence" of the persistent activity has been repeatedly questioned [see Constantinidis et al. (2018), Lundqvist et al. (2018a)].

Accordingly, we have shown clear evidence that, at a single-trial level, the spiking activity of individual neurons is sparse and very heterogeneous across successive trials. In particular, we show that this applies to spatial attention, and that spatial attention is not attached to a specific cued location in space, but rather expresses intrinsic oscillatory dynamics covering the whole visual space in a rhythmic manner at approximately 8 Hz, impacting behavioral performance (Lakatos et al., 2008; Dugue and VanRullen, 2014; VanRullen, 2016; Fiebelkorn et al., 2018; Spyropoulos et al., 2018; Gaillard et al., 2020). In addition, we also show that the neuronal population codes for spatial attention vary at a very slow rhythm of a few cycles per hour. Although the impact of the oscillations on behavioral performance is very strong, their origin is still unknown. The fluctuations in the prefrontal spatial attention codes cannot be tracked on the single neuronal responses, and only become apparent when a larger neuronal population is considered. Lastly, using dimensionality reduction techniques, we consider an additional degree of complexity of delay-related prefrontal activity, identifying specific neuronal sources of variance associated with overt behavioral performance (correct vs. errors) and attention, respectively (Amengual et al., 2021). While most studies on mixed selectivity in the prefrontal neuronal population have focused on task-related information coding, here we consider a condition in which mixed selectivity is associated with a task-independent variable (a state of inattention or attentional lapse) that dynamically and transiently interferes with task-related processes.

Overall, we thus provide a systematic deconstruction of the idea of the persistence of the neuronal activity in the context of attention orienting, and we describe multiple sources of neuronal dynamic processes in the "silent" epochs of cognitive tasks in multiple time scales. An important challenge that remains to be addressed is how this dynamic is organized both at the mesoscopic level of the cortical area and its layers and at the level of the functional network.

AUTHOR CONTRIBUTIONS

SB and JA: conceptualization, figures, writing—original draft, and writing–review and editing. SB: funding acquisition and supervision. All authors contributed to the article and approved the submitted version

FUNDING

This work was supported by ERC Brain3.0 #681978, ANR-11-BSV4-0011, and ANR-14-ASTR-0011-01, LABEX CORTEX funding (ANR-11-LABX-0042) from the Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

ACKNOWLEDGMENTS

We acknowledge Drs. Fardin Afdideh and Corentin Gaillard for their fruitful comments on this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncir.2021.679796/full#supplementary-material

Supplementary Figure 1 | Description of the rapid serial visual presentation (RSVP) task (Ibos et al., 2013). Monkeys have to fixate a central point while a first stream of visual stimuli is presented (stimuli changing every 200 ms). After a few stimuli, a second stream of visual stimuli is presented contralateral to the first stream. A cue is then presented in the first visual stream. The cue can either be green instructing the monkey to maintain attention on this first visual stream (stay cue) because the target will be presented in this stream. Alternatively, the cue can be red, instructing the monkey to shift attention to the second visual stream (shift cue) because the target will be presented in this stream. The cue can thus be red or green (color dimension), presented in the left or in the right visual streams (position dimension), or instruct attention to be oriented to the left or to the right visual streams (attention dimension). Monkeys are rewarded to maintain fixation all throughout the trial and respond to the target presentation with a manual response as fast as possible.

Supplementary Figure 2 | Schema of the construction of a classifier for two-dimensional (2D; *x,y*) readout of the attention information. **(A)** Schema of an electrode with two contacts recording multiunit activity (MUA) activity obtained from two different neurons. Each of the two neurons is tuned to two classes of information (different positions), represented by red and blue colors. **(B)** To quantify the amount of information in the data, a regularized linear regression is applied. *R*: mean neural response in a specific sliding window [shaded gray in **(A)**], W is the synaptic weights representing the contribution of each cell to the final readout and C is a 2D vector describing the two possible classes. A Tikhonov-regularized version is used in order to minimize the cost function to avoid overfitting. A linear decoder estimates a decision boundary of the classifier when discriminating between the population response to class 1 or to class 2 events. **(C)** Each output of the decoder represents an (*x,y*) position relative to the decision boundary, representing the decoded position of the attentional locus in the visual space. **(D)** Decoding procedure applied along the whole cue-to-target interval provides time-resolved decoding of the attentional trajectory associated with each class (position).

REFERENCES

Amengual, J. L., Di Bello, F., Gaillard, C., and Ben Hamed, S. (2021). *Accessing Attention Information in the Frontal Eye Field is Driven by Neural States Associated With Levels of Distractability and Impulsivity*. A NeuroFrance International Virtual Meeting.

Andersen, R. A., Hwang, E. J., and Mulliken, G. H. (2010). Cognitive neural prosthetics. *Annu. Rev. Psychol.* 61, 169–190. doi: 10.1146/annurev.psych.093008.100503

Arandia-Romero, I., Nogueira, R., Mochol, G., and Moreno-Bote, R. (2017). What can neuronal populations tell us about cognition? *Curr. Opin. Neurobiol.* 46, 48–57. doi: 10.1016/j.conb.2017.07.008

Armstrong, K. M., Chang, M. H., and Moore, T. (2009). Selection and maintenance of spatial information by frontal eye field neurons. *J. Neurosci.* 29, 15621–15629. doi: 10.1523/JNEUROSCI.4465-09.2009

Astrand, E., Enel, P., Ibos, G., Dominey, P. F., Baraduc, P., and Ben Hamed, S. (2014a). Comparison of classifiers for decoding sensory and cognitive information from prefrontal neuronal populations. *PLoS ONE* 9:e86314. doi: 10.1371/journal.pone.0086314

Astrand, E., Ibos, G., Duhamel, J.-R., and Ben Hamed, S. (2015). Differential dynamics of spatial attention, position, and color coding within the parietofrontal network. *J. Neurosci.* 35, 3174–3189. doi: 10.1523/JNEUROSCI.2370-14.2015

Astrand, E., Wardak, C., Baraduc, P., and Ben Hamed, S. (2016). Direct two-dimensional access to the spatial location of covert attention in macaque prefrontal cortex. *Curr. Biol.* 26, 1699–1704. doi: 10.1016/j.cub.2016.04.054

Astrand, E., Wardak, C., and Ben Hamed, S. (2014b). Selective visual attention to drive cognitive brainâ€“machine interfaces: from concepts to neurofeedback and rehabilitation applications. *Front. Syst. Neurosci.* 8:144. doi: 10.3389/fnsys.2014.00144

Astrand, E., Wardak, C., and Ben Hamed, S. (2020). Neuronal population correlates of target selection and distractor filtering. *Neuroimage* 209:116517. doi: 10.1016/j.neuroimage.2020.116517

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7, 358–366. doi: 10.1038/nrn1888

Awh, E., and Pashler, H. (2000). Evidence for split attentional foci. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 834–846. doi: 10.1037/0096-1523.26.2.834

Barak, O., Tsodyks, M., and Romo, R. (2010). Neuronal population coding of parametric working memory. *J. Neurosci.* 30, 9424–9430. doi: 10.1523/JNEUROSCI.1875-10.2010

Ben Hadj Hassen, S., and Ben Hamed, S. (2020). Functional and behavioural correlates of shared neuronal noise variability in vision and visual cognition. *Curr. Opin. Physiol.* 16, 85–97. doi: 10.1016/j.cophys.2020.07.015

Ben Hadj Hassen, S., Gaillard, C., Astrand, E., Wardak, C., and Ben Hamed, S. (2019). Interneuronal correlations dynamically adjust to task demands at multiple time-scales. *bioRxiv*. 547802. doi: 10.1101/547802

Ben Hamed, S., Page, W., Duffy, C., and Pouget, A. (2003). MSTd neuronal basis functions for the population encoding of heading direction. *J. Neurophysiol.* 90, 549–558. doi: 10.1152/jn.00639.2002

Bonnefond, A., Doignon-Camus, N., Touzalin-Chretien, P., and Dufour, A. (2010). Vigilance and intrinsic maintenance of alert state: an ERP study. *Behav. Brain Res.* 211, 185–190. doi: 10.1016/j.bbr.2010.03.030

Brincat, S. L., Siegel, M., von Nicolai, C., and Miller, E. K. (2018). Gradual progression from sensory to task-related processing in cerebral cortex. *Proc. Natl. Acad. Sci. U.S.A.* 115, E7202–E7211. doi: 10.1073/pnas.1717075115

Brookshire, G. (2021). Re-evaluating rhythmic attentional switching: Spurious oscillations from shuffling-in-time. *bioRxiv*. 2021.05.07.443101. doi: 10.1101/2021.05.07.443101

Buonomano, D. V., and Maass, W. (2009). State-dependent computations: Spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* 10, 113–125. doi: 10.1038/nrn2558

Buschman, T. J., and Miller, E. K. (2007). Top-Down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science.* 315, 1860–1862. doi: 10.1126/science.1138071

Cavada, C., and Goldman-Rakic, P. S. (1989). Posterior parietal cortex in rhesus monkey: II. Evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. *J. Comparat. Neurol.* 287, 422–445. doi: 10.1002/cne.902870403

Chapin, J. K., Moxon, K. A., Markowitz, R. S., and Nicolelis, M. A. L. (1999). Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nat. Neurosci.* 2, 664–670. doi: 10.1038/10223

Chaudhuri, R., and Fiete, I. (2016). Computational principles of memory. *Nat. Neurosci.* 19, 394–403. doi: 10.1038/nn.4237

Churchland, A. K., Kiani, R., Chaudhuri, R., Wang, X.-J., Pouget, A., Shadlen, M. N. (2011). Variance as a signature of neural computations during decision making. *Neuron* 69, 818–831. doi: 10.1016/j.neuron.2010.12.037

Cohen, M. R., and Newsome, W. T. (2008). Context-Dependent changes in functional circuitry in visual area MT. *Neuron* 60, 162–173. doi: 10.1016/j.neuron.2008.08.007

Cohen, Y. E., and Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nat. Rev. Neurosci.* 3, 553–562. doi: 10.1038/nrn873

Colby, C. L., Duhamel, J. R., and Goldberg, M. E. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *J. Neurophysiol.* 76, 2841–2852. doi: 10.1152/jn.1996.76.5.2841

Compte, A. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* 10, 910–923. doi: 10.1093/cercor/10.9.910

Compte, A., Constantinidis, C., Tegnér, J., Raghavachari, S., Chafee, M. V., Goldman-Rakic, P. S., et al. (2003). Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *J. Neurophysiol.* 90, 3441–3454. doi: 10.1152/jn.00949.2002

Constantinidis, C., Funahashi, S., Lee, D., Murray, J. D., Qi, X.-L., Wang, M., et al. (2018). Persistent spiking activity underlies working memory. *J. Neurosci.* 38, 7020–7028. doi: 10.1523/JNEUROSCI.2486-17.2018

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Corchs, S. (2002). Large-scale neural model for visual attention: integration of experimental single-cell and fMRI data. *Cereb. Cortex* 12, 339–348. doi: 10.1093/cercor/12.4.339

Crowe, D. A., Averbeck, B. B., and Chafee, M. V. (2010). Rapid sequences of population activity patterns dynamically encode task-critical spatial information in parietal cortex. *J. Neurosci.* 30, 11640–11653. doi: 10.1523/JNEUROSCI.0954-10.2010

Cunningham, J. P., and Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* 17, 1500–1509. doi: 10.1038/nn.3776

de Schotten, M. T., Dell'Acqua, F., Forkel, S. J., Simmons, A., Vergani, F., Murphy, D. G. M., et al. (2011). A lateralized brain network for visuospatial attention. *Nat. Neurosci.* 14, 1245–1246. doi: 10.1038/nn.2905

De Sousa, C., Gaillard, C., Di Bello, C., Ben Hadj Hassen, F., and Ben Hamed, S. (2021). Behavioral validation of novel high resolution attention decoding method from multi-units & local field potentials. *Neuroimage* 231:117853. doi: 10.1016/j.neuroimage.2021.117853

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222. doi: 10.1146/annurev.ne.18.030195.001205

Di Bello, F., Ben Hadj Hassen, S., Astrand, E., and Ben Hamed, S. (2020). Selection and suppression of visual information in the macaque prefrontal cortex. *bioRxiv*. 2020.03.25.007922 doi: 10.1101/2020.03.25.007922

Dugue, L., and VanRullen, R. (2014). The dynamics of attentional sampling during visual search revealed by fourier analysis of periodic noise interference. *J. Vis.* 14:11. doi: 10.1167/14.2.11

Eriksen, C. W., and St. James, J. D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Percept. Psychophys.* 40, 225–240. doi: 10.3758/BF03211502

Eriksen, C. W., and Yeh, Y.-Y. (1985). Allocation of attention in the visual field. *J. Exp. Psychol. Hum. Percept. Perform.* 11, 583–597. doi: 10.1037/0096-1523.11.5.583

Feng, S. F., Schwemmer, M., Gershman, S. J., and Cohen, J. D. (2014). Multitasking versus multiplexing: toward a normative account of limitations in the simultaneous execution of control-demanding behaviors. *Cogn. Affect. Behav. Neurosci.* 14, 129–146. doi: 10.3758/s13415-013-0236-9

Fiebelkorn, I. C., Pinsk, M. A., and Kastner, S. (2018). A dynamic interplay within the frontoparietal network underlies rhythmic spatial attention. *Neuron* 99, 842–853.e8. doi: 10.1016/j.neuron.2018.07.038

Froudarakis, E., Berens, P., Ecker, A. S., Cotton, R. J., Sinz, F. H., Yatsenko, D., et al. (2014). Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat. Neurosci.* 17, 851–857. doi: 10.1038/nn.3707

Fujisawa, S., Amarasingham, A., Harrison, M. T., and Buzsáki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat. Neurosci.* 11, 823–833. doi: 10.1038/nn.2134

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

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

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

Gaillard, C., Ben Hadj Hassen, S., Di Bello, F., Bihan-Poudec, Y., VanRullen, R., and Ben Hamed, S. (2020). Prefrontal attentional saccades explore space rhythmically. *Nat. Commun.* 11:925. doi: 10.1038/s41467-020-14649-7

Gaillard, C., and Ben Hamed, S. (2020). The neural bases of spatial attention and perceptual rhythms. *Euro. J. Neurosci.* 53, 1661–1671. doi: 10.1111/ejn.15044

Gaillard, C., De Sousa, C., Amengual, J. L., Loriette, C., Ziane, C., and Ben Hamed, S. (2021). Attentional brain rhythms during prolonged cognitive activity. *BioRxiv*. 2021.05.26.445730. doi: 10.1101/2021.05.26.445730

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

Gottlieb, J. P., Kusunoki, M., and Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature* 391, 481–484. doi: 10.1038/35135

Gutnisky, D. A., Beaman, C., Lew, S. E., and Dragoi, V. (2017). Cortical response states for enhanced sensory discrimination. *Elife* 6:e29226. doi: 10.7554/eLife.29226.021

Harris, K. D., and Thiele, A. (2011). Cortical state and attention. *Nat. Rev. Neurosci.* 12, 509–523. doi: 10.1038/nrn3084

Ibos, G., Duhamel, J.-R., and Ben Hamed, S. (2013). A functional hierarchy within the parietofrontal network in stimulus selection and attention control. *J. Neurosci.* 33, 8359–8369. doi: 10.1523/JNEUROSCI.4058-12.2013

Katsuki, F., and Constantinidis, C. (2014). Bottom-Up and top-down attention: different processes and overlapping neural systems. *Neuroscientist* 20, 509–521. doi: 10.1177/1073858413514136

Khanna, S. B., Scott, J. A., and Smith, M. A. (2020). Dynamic shifts of visual and saccadic signals in prefrontal cortical regions 8Ar and FEF. *J. Neurophysiol.* 124, 1774–1791. doi: 10.1152/jn.00669.2019

King, J.-R., and Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn. Sci.* 18, 203–210. doi: 10.1016/j.tics.2014.01.002

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C. E., Kepecs, A., Mainen, Z. F., et al. (2016). Demixed principal component analysis of neural population data. *Elife* 5:e10989. doi: 10.7554/eLife.10989.022

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113. doi: 10.1126/science.1154735

Landau, A. N., and Fries, P. (2012). Attention samples stimuli rhythmically. *Curr. Biol.* 22, 1000–1004. doi: 10.1016/j.cub.2012.03.054

Leavitt, M. L., Pieper, F., Sachs, A. J., and Martinez-Trujillo, J. C. (2017). Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles. *Proc. Natl. Acad. Sci. U.S.A.* 114, E2494–E2503. doi: 10.1073/pnas.1619949114

Lee, D. K., Itti, L., Koch, C., and Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nat. Neurosci.* 2, 375–381. doi: 10.1038/7286

Liu, T., and Hou, Y. (2013). A hierarchy of attentional priority signals in human frontoparietal cortex. *J. Neurosci.* 33, 16606–16616. doi: 10.1523/JNEUROSCI.1780-13.2013

Lockley, S. W., Cronin, J. W., Evans, E. E., Cade, B. E., Lee, C. J., Landrigan, C. P., et al. (2004). Effect of reducing interns' weekly work hours on sleep and attentional failures. *N. Engl. J. Med.* 351, 1829–1837. doi: 10.1056/NEJMoa041404

Lundqvist, M., Herman, P., and Miller, E. K. (2018a). Working memory: delay activity, yes! persistent activity? Maybe not. *J. Neurosci.* 38, 7013–7019. doi: 10.1523/JNEUROSCI.2485-17.2018

Lundqvist, M., Herman, P., Warden, M. R., Brincat, S. L., and Miller, E. K. (2018b). Gamma and beta bursts during working memory readout suggest roles in its volitional control. *Nat. Commun.* 9:394. doi: 10.1038/s41467-017-02791-8

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

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

Maimon, G., and Assad, J. A. (2009). Beyond poisson: increased spike-time regularity across primate parietal cortex. *Neuron* 62, 426–440. doi: 10.1016/j.neuron.2009.03.021

Manohar, S. G., Zokaei, N., Fallon, S. J., Vogels, T. P., and Husain, M. (2019). Neural mechanisms of attending to items in working memory.

Neurosci. Biobehav. Rev. 101, 1–12. doi: 10.1016/j.neubiorev.2019.03.017

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

Marek, S. (2018). The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. *Dialogues Clin. Neurosci.* 20, 133–140. doi: 10.31887/DCNS.2018.20.2/smarek

Martínez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., et al. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat. Neurosci.* 2, 364–369. doi: 10.1038/7274

Mayo, J. P., and Maunsell, J. H. R. (2016). Graded neuronal modulations related to visual spatial attention. *J. Neurosci.* 36, 5353–5361. doi: 10.1523/JNEUROSCI.0192-16.2016

McMains, S. A., and Somers, D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron* 42, 677–686. doi: 10.1016/S0896-6273(04)00263-6

Mendoza-Halliday, D., and Martinez-Trujillo, J. C. (2017a). Neuronal population coding of perceived and memorized visual features in the lateral prefrontal cortex. *Nat. Commun.* 8:15471. doi: 10.1038/ncomms15471

Mendoza-Halliday, D., and Martinez-Trujillo, J. C. (2017b). Neuronal population coding of perceived and memorized visual features in the lateral prefrontal cortex. *Nat. Commun.* 8:15471.

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

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

Monosov, I. E., and Thompson, K. G. (2009). Frontal eye field activity enhances object identification during covert visual search. *J. Neurophysiol.* 102, 3656–3672. doi: 10.1152/jn.00750.2009

Moore, T., and Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature* 421, 370–373. doi: 10.1038/nature01341

Moore, T., and Fallah, M. (2004). Microstimulation of the frontal eye field and its effects on covert spatial attention. *J. Neurophysiol.* 91, 152–162. doi: 10.1152/jn.00741.2002

Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science* 229, 782–784. doi: 10.1126/science.4023713

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. *Nat. Neurosci.* 17, 1410–1417. doi: 10.1038/nn.3807

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., et al. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* 17, 1661–1663. doi: 10.1038/nn.3862

Niebergall, R., Khayat, P. S., Treue, S., and Martinez-Trujillo, J. C. (2011). Multifocal attention filters targets from distracters within and beyond primate MT neurons' receptive field boundaries. *Neuron* 72, 1067–1079. doi: 10.1016/j.neuron.2011.10.013

Niebur, E., and Koch, C. (1994). A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *J. Comput. Neurosci.* 1, 141–158. doi: 10.1007/BF00962722

Nogueira, R., Peltier, N. E., Anzai, A., DeAngelis, G. C., Martínez-Trujillo, J., and Moreno-Bote, R. (2020). The effects of population tuning and trial-by-trial variability on information encoding and behavior. *J. Neurosci.* 40, 1066–1083. doi: 10.1523/JNEUROSCI.0859-19.2019

Oberauer, K. (2019). Working memory and attention – a conceptual analysis and review. *J. Cogn.* 2:36. doi: 10.5334/joc.58

Olshausen, B., and Field, D. (2004). Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* 14, 481–487. doi: 10.1016/j.conb.2004.07.007

Paneri, S., and Gregoriou, G. G. (2017). Top-Down control of visual attention by the prefrontal cortex. Functional specialization and long-range interactions. *Front. Neurosci.* 11:545. doi: 10.3389/fnins.2017.00545

Panzeri, S., Harvey, C. D., Piasini, E., Latham, P. E., and Fellin, T. (2017). Cracking the neural code for sensory perception by combining statistics, intervention, and behavior. *Neuron* 93, 491–507. doi: 10.1016/j.neuron.2016.12.036

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

Pinto, Y., van der Leij, A. R., Sligte, I. G., Lamme, V. A. F., and Scholte, H. S. (2013). Bottom-up and top-down attention are independent. *J. Vis.* 13, 16. doi: 10.1167/13.3.16

Posner, M. I., and Petersen, S. E. (1990). The attention system of the human brain. *Annu. Rev. Neurosci.* 13, 25–42. doi: 10.1146/annurev.ne.13.030190.000325

Potter, M. C. (2018). "Rapid serial visual presentation (RSVP)," in *New Methods in Reading Comprehension Research, 1st Edn*, eds D. E. Kieras, and M. A. Just (London: Routledge), 91–118. doi: 10.4324/9780429505379-5

Proctor, S. P., White, R., Robins, T., Echeverria, D., and Rocskay, A. (1996). Effect of overtime work on cognitive function in automotive workers. *Scand. J. Work Environ. Health* 22, 124–132. doi: 10.5271/sjweh.120

Reynolds, J. H., Pasternak, T., and Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron* 26, 703–714. doi: 10.1016/S0896-6273(00)81206-4

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

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

Song, K., Meng, M., Chen, L., Zhou, K., and Luo, H. (2014). Behavioral oscillations in attention: rhythmic α pulses mediated through θ band. *J. Neurosci.* 34, 4837–4844. doi: 10.1523/JNEUROSCI.4856-13.2014

Spyropoulos, G., Bosman, C. A., and Fries, P. (2018). A theta rhythm in macaque visual cortex and its attentional modulation. *Proc. Natl. Acad. Sci. U.S.A.* 115, E5614–E5623. doi: 10.1073/pnas.1719433115

Stanton, G. B., Bruce, C. J., and Goldberg, M. E. (1995). Topography of projections to posterior cortical areas from the macaque frontal eye fields. *J. Comp. Neurol.* 353, 291–305. doi: 10.1002/cne.903530210

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

Sunny, T. D., Aparna, T., Neethu, P., Venkateswaran, J., Vishnupriya, V., and Vyas, P. S. (2016). Robotic arm with brain – computer interfacing. *Proc. Technol.* 24, 1089–1096. doi: 10.1016/j.protcy.2016.05.241

Theeuwes, J. (1993). Visual selective attention: a theoretical analysis. *Acta Psychol.* 83, 93–154. doi: 10.1016/0001-6918(93)90042-P

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136. doi: 10.1016/0010-0285(80)90005-5

Trejo, L. J., Rosipal, R., and Matthews, B. (2006). Brain–Computer interfaces for 1-d and 2-d cursor control: designs using volitional control of the EEG spectrum or steady-state visual evoked potentials. *IEEE Trans. Neural Syst. Rehabil. Eng.* 14, 225–229. doi: 10.1109/TNSRE.2006.875578

Tremblay, S., Doucet, G., Pieper, F., Sachs, A., and Martinez-Trujillo, J. (2015a). Single-Trial decoding of visual attention from local field potentials in the primate lateral prefrontal cortex is frequency-dependent. *J. Neurosci.* 35, 9038–9049. doi: 10.1523/JNEUROSCI.1041-15.2015

Tremblay, S., Pieper, F., Sachs, A., and Martinez-Trujillo, J. (2015b). Attentional filtering of visual information by neuronal ensembles in the primate lateral prefrontal cortex. *Neuron* 85, 202–215. doi: 10.1016/j.neuron.2014.11.021

VanRullen, R. (2013). Visual attention: a rhythmic process? *Curr. Biol.* 23, R1110–1112. doi: 10.1016/j.cub.2013.11.006

VanRullen, R. (2016). Perceptual cycles. *Trends Cogn. Sci.* 20, 723–735. doi: 10.1016/j.tics.2016.07.006

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., and Thirion, B. (2017). Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *Neuroimage* 145, 166–179. doi: 10.1016/j.neuroimage.2016.10.038

Venables, P. H. (1960). Periodicity in reaction time. *Br. J. Psychol.* 51, 37–43. doi: 10.1111/j.2044-8295.1960.tb00722.x

Verhoef, B.-E., and Maunsell, J. H. R. (2017). Attention-related changes in correlated neuronal activity arise from normalization mechanisms. *Nat. Neurosci.* 20, 969–977. doi: 10.1038/nn.4572

Virtanen, M., and Kivimäki, M. (2018). Long working hours and risk of cardiovascular disease. *Curr. Cardiol. Rep.* 20:123. doi: 10.1007/s11886-018-1049-9

Wardak, C. (2006). Contribution of the monkey frontal eye field to covert visual attention. *J. Neurosci.* 26, 4228–4235. doi: 10.1523/JNEUROSCI.3336-05.2006

Wardak, C., Olivier, E., and Duhamel, J.-R. (2002). Saccadic target selection deficits after lateral intraparietal area inactivation in monkeys. *J. Neurosci.* 22, 9877–9884. doi: 10.1523/JNEUROSCI.22-22-09877.2002

Watanabe, K., and Funahashi, S. (2014). Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nat. Neurosci.* 17, 601–611. doi: 10.1038/nn.3667

Wohrer, A., Humphries, M. D., and Machens, C. K. (2013). Population-wide distributions of neural activity during perceptual decision-making. *Progr. Neurobiol.* 103, 156–193. doi: 10.1016/j.pneurobio.2012.09.004

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Womelsdorf, T., Lima, B., Vinck, M., Oostenveld, R., Singer, W., Neuenschwander, S., et al. (2012). Orientation selectivity and noise correlation in awake monkey area V1 are modulated by the gamma cycle. *Proc. Natl. Acad. Sci.* 109, 4302–4307. doi: 10.1073/pnas.1114223109

Zucker, R. S., and Regehr, W. G. (2002). Short-Term synaptic plasticity. *Annu. Rev. Physiol.* 64, 355–405. doi: 10.1146/annurev.physiol.64.092501.114547

Zylberberg, J., and Strowbridge, B. W. (2017). Mechanisms of persistent activity in cortical circuits: possible neural substrates for working memory. *Annu. Rev. Neurosci.* 40, 603–627. doi: 10.1146/annurev-neuro-070815-014006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neural Circuits

ORIGINAL RESEARCH
published: 08 July 2021
doi: 10.3389/fncir.2021.648538

Online Learning and Memory of Neural Trajectory Replays for Prefrontal Persistent and Dynamic Representations in the Irregular Asynchronous State

Matthieu X. B. Sarazin[1*†], *Julie Victor*[2†], *David Medernach*[1], *Jérémie Naudé*[3†] *and Bruno Delord*[1*†]

[1] *Institut des Systèmes Intelligents et de Robotique, CNRS, Inserm, Sorbonne Université, Paris, France,* [2] *CEA Paris-Saclay, CNRS, NeuroSpin, Saclay, France,* [3] *Neuroscience Paris Seine - Institut de biologie Paris Seine, CNRS, Inserm, Sorbonne Université, Paris, France*

In the prefrontal cortex (PFC), higher-order cognitive functions and adaptive flexible behaviors rely on continuous dynamical sequences of spiking activity that constitute neural trajectories in the state space of activity. Neural trajectories subserve diverse representations, from explicit mappings in physical spaces to generalized mappings in the task space, and up to complex abstract transformations such as working memory, decision-making and behavioral planning. Computational models have separately assessed learning and replay of neural trajectories, often using unrealistic learning rules or decoupling simulations for learning from replay. Hence, the question remains open of how neural trajectories are learned, memorized and replayed online, with permanently acting biological plasticity rules. The asynchronous irregular regime characterizing cortical dynamics in awake conditions exerts a major source of disorder that may jeopardize plasticity and replay of locally ordered activity. Here, we show that a recurrent model of local PFC circuitry endowed with realistic synaptic spike timing-dependent plasticity and scaling processes can learn, memorize and replay large-size neural trajectories online under asynchronous irregular dynamics, at regular or fast (sped-up) timescale. Presented trajectories are quickly learned (within seconds) as synaptic engrams in the network, and the model is able to chunk overlapping trajectories presented separately. These trajectory engrams last long-term (dozen hours) and trajectory replays can be triggered over an hour. In turn, we show the conditions under which trajectory engrams and replays preserve asynchronous irregular dynamics in the network. Functionally, spiking activity during trajectory replays at regular timescale accounts for both dynamical coding with temporal tuning in individual neurons, persistent activity at the population level, and large levels of variability consistent with observed cognitive-related PFC dynamics. Together, these results offer a consistent theoretical framework accounting for how neural trajectories can be learned, memorized and replayed in PFC networks circuits to subserve flexible dynamic representations and adaptive behaviors.

Keywords: prefrontal cortex, neural trajectory, attractor, persistent and dynamical coding, working memory, learning, replay, asynchronous irregular state

OPEN ACCESS

Edited by:
Shintaro Funahashi, Kyoto University, Japan

Reviewed by:
Shantanu P. Jadhav, Brandeis University, United States
Lukas Ian Schmitt, RIKEN Center for Brain Science (CBS), Japan

***Correspondence:**
Matthieu X. B. Sarazin
matthieu.sarazin@live.fr
Bruno Delord
bruno.delord@sorbonne-universite.fr

[†] These authors have contributed equally to this work

Received: *31 December 2020*
Accepted: *31 May 2021*
Published: *08 July 2021*

Citation:
Sarazin MXB, Victor J, Medernach D, Naudé J and Delord B (2021) Online Learning and Memory of Neural Trajectory Replays for Prefrontal Persistent and Dynamic Representations in the Irregular Asynchronous State. Front. Neural Circuits 15:648538. doi: 10.3389/fncir.2021.648538

INTRODUCTION

As when a few introductory notes recall a melody, in the immense space of known melodies, cerebral networks are able to memorize and replay complex temporal patterns in a flexible way. Such temporal patterns rely on continuous dynamical sequences of spiking activity, i.e., neural trajectories, that occur in recurrent neural networks of the prefrontal cortex (PFC) (Bakhurin et al., 2017; Paton and Buonomano, 2018; Wang et al., 2018). These neural trajectories emerge with learning, relying on dynamical engrams, which distinguish them from classical static engrams underlying Hebbian neuronal assemblies. In turn, these engrams likely arise through activity-dependent synaptic plasticity (Goto et al., 2010; Bittner et al., 2017). Hence, a robust understanding of the interplay between prefrontal dynamics and biological plastic processes is necessary to understand the emergence of functional neural trajectories and engrams. In the PFC of behaving animals, neural trajectories are embedded in an asynchronous and irregular background state activity that is markedly disordered (Destexhe et al., 2003; London et al., 2010). However, how synaptic plasticity builds engrams that are not erased by spontaneous activity and yet are not strong enough to alter irregular PFC dynamics remains an open question.

Neural trajectories correspond to organized spatio-temporal representations that peregrinate within the neural space (Shenoy et al., 2013). They are prominent in prefrontal cortices (Mante et al., 2013), where they subserve higher-order cognitive functions at diverse levels of abstraction (Wutz et al., 2018). In prefrontal areas, at the lowest levels of abstraction, neural trajectories can map the actual animal's position during effective trajectories within explicit spaces during visual perception (Mante et al., 2013) or navigation (Fujisawa et al., 2008; Zielinski et al., 2019). Beyond spatial mapping, neural trajectories can also depict generalized topological locations that are isomorphic to the task space, by multiplexing position, representation of goal locations and choice-related information (Fujisawa et al., 2008; Mashhoori et al., 2018; Yu et al., 2018; Kaefer et al., 2020). Neural trajectories have also been shown to subserve dynamical coding and manipulation of information during delay activities in working memory tasks involving the PFC (Lundqvist et al., 2018). In this context, neural trajectories do not represent explicit trajectories in external spaces, but implicit representations—of ongoing information and cognitive operations—that may prove useful for the task.

Rather than static maintenance of persistent activity in a group of cells, many working-memory representations unfold in the space of neural activity under the form of continuous trajectories, as neurons successively activate in "relay races" sequences of transient activity (Batuev, 1994; Brody et al., 2003; Cromer et al., 2010; Yang et al., 2014; Schmitt et al., 2017; Enel et al., 2020). In the PFC, neural trajectories can form the substrate for dynamic (Sreenivasan et al., 2014) but also, counterintuitively, for stable representations (Druckmann and Chklovskii, 2012). Neural trajectory-mediated dynamical representations can subserve the retrospective working memory of spatial (Batuev, 1994; Yang et al., 2014) or quantitative (Brody et al., 2003) cues, symbolic categories (Cromer et al., 2010), values (Enel et al., 2020), or behavioral rules (Schmitt et al., 2017). They can also serve prospective working memory in computational processes transforming previously encoded information, such as, for e.g., in visuo-motor transformations (Spaak et al., 2017), in the representation of elapsed time (Tiganj et al., 2017) or in the encoding of forthcoming behaviors (Fujisawa et al., 2008; Ito et al., 2015; Nakajima et al., 2019; Passecker et al., 2019). Neural trajectories in the neural space can also appear as sequences of states that involve combinations of active neurons (Batuev, 1994; Abeles et al., 1995; Seidemann et al., 1996; La Camera et al., 2019). Thus, neural trajectories appear in diverse forms and in different functional contexts where they can map actual trajectories in external spaces, remember previously encountered trajectories, or predict forthcoming trajectories during active computational processes requiring dynamical representations.

Neural trajectories in the PFC are adaptive (Euston et al., 2012; Mante et al., 2013): they are learned and memorized, to be "replayed" later. The timescale of the replay depends on the behavioral context. Regular timescale replays operate at the behavioral timescale, lasting seconds (Batuev, 1994; Fujisawa et al., 2008; Cromer et al., 2010; Mante et al., 2013; Yang et al., 2014; Ito et al., 2015; Schmitt et al., 2017; Tiganj et al., 2017; Nakajima et al., 2019; Passecker et al., 2019; Enel et al., 2020). Thus, such replays unfold online as current behavior is executed in interaction with the external world, to subserve retrospective working memory of past information, ongoing dynamical computations, or prospective representation of forthcoming behaviors. Typically, regular replays are triggered by behaviorally–relevant external events (e.g., cues or go signals in working memory tasks, or the current position in navigational tasks). Some replays that may appear as spontaneous can be presumably triggered by internal self-paced decision signals within the PFC (e.g., choices). In all cases, such triggered regular replays rely on internal mechanisms within PFC circuits allowing for the autonomous propagation of proper sequences of activity, once initial neurons of the neural trajectory have been triggered. A major goal of the present study is to decipher how plastic processes allow PFC circuits to learn and replay trajectories, i.e., autonomously generate neural trajectory completion, based on an initial trigger.

Besides, fast timescale replays exist that last a few hundred milliseconds during awake (Jadhav et al., 2016; Mashhoori et al., 2018; Yu et al., 2018; Shin et al., 2019; Kaefer et al., 2020) and sleeping (Euston et al., 2007; Peyrache et al., 2009) states. Beyond their much shorter duration, PFC fast replays are distinct from regular ones, in that they typically operate offline and often co-occur with fast replays in the hippocampal CA1 field (Jadhav et al., 2016). Replay activity in PFC and CA1 presents high degrees of task-dependent spatial and temporal correlations (Jadhav et al., 2016; Yu et al., 2018; Shin et al., 2019), subserving functional coordination combining metric (hippocampus) and task-related (PFC) spatial representations (Pfeiffer and Foster, 2013; Zielinski et al., 2019). These fast replays occur during sharp-wave ripples (SWR) episodes (Jadhav et al., 2016; Yu et al., 2018; Shin et al., 2019), which represent critical events for behavioral learning (Jadhav et al., 2012) and during which animals forge forthcoming decisions (choices,

trajectories, for e.g., Jadhav et al., 2016; Mashhoori et al., 2018; Kaefer et al., 2020), based on the recall of past experiences (actions, trajectories, outcomes, for e.g., Jadhav et al., 2012; Mashhoori et al., 2018). Such coordination across both structures presumably emerges through their reciprocal, direct and indirect, synaptic interactions (Witter and Amaral, 2004). Different studies have pointed out information flow biases from CA1 to PFC (Jadhav et al., 2016) or from PFC to CA1 (Ito et al., 2015) directions, depending on behavioral contexts. However, SWR-related replays in the hippocampus correlate with fast replays in reduced subsets of PFC neurons (Jadhav et al., 2016; Yu et al., 2018) that carry generalized spatial representations but not specific trajectories (Yu et al., 2018). Moreover, fast timescale PFC replays are independent of hippocampal replays during computational processes inherent to the PFC, such as rule switching tasks (Kaefer et al., 2020). Therefore, as for regular replays, we examined how plastic processes allow for the emergence of fast timescale replays autonomously within local recurrent PFC circuits.

Neuronal trajectories consist of robust forms of ordered local activity occurring within a disordered global activity, i.e., the chaotic, asynchronous irregular (AI) state characteristic of the prefrontal cortex in the waking state (Destexhe et al., 2003; London et al., 2010). This coexistence poses a problem at the plasticity level, because the noisy AI regime constitutes a potential source of perturbation for synaptic engrams (Boustani et al., 2012; Litwin-Kumar and Doiron, 2014), whereas strengthened connectivity pathways may exert a synchronizing influence on the network, dramatically altering the chaotic nature of background activity. However, there is currently no biophysically-grounded theoretical framework accounting for the way neural trajectories are learned, memorized and replayed within recurrent cortical networks. In principle, synaptic plasticity, a major substrate of learning, may sculpt oriented connective pathways promoting the propagation of neuronal trajectories, because modifications of synaptic connections are activity-dependent. Specifically, the sequential activation of differentially tuned neurons during successively crossed spatial positions (during navigational trajectories) or representational states (during dynamical cognitive processes) could strengthen connections between neurons, creating oriented pathways (referred to as trajectory engrams hereafter) within recurrent cortical networks. If sufficiently strengthened, engrams could allow the propagation of packets of neuronal activity along them. From an initial stimulation of neurons located at the beginning of the engram, due to the strong connections linking them in the direction of the trajectory, neurons could reactivate sequentially, i.e., perform trajectory replay.

Recurrent neural network models have shown that activity-dependent synaptic plasticity rules can enable the formation of trajectory engrams due to long-term potentiation (LTP) and depression (LTD) together with homeostatic scaling (Liu and Buonomano, 2009; Clopath et al., 2010; Fiete et al., 2010; Klampfl and Maass, 2013). Moreover, trajectory engrams can propagate neuronal trajectories through sequential activation of neurons in recurrent model networks (Liu and Buonomano, 2009; Fiete et al., 2010; Klampfl and Maass, 2013; Laje and Buonomano, 2013; Chenkov et al., 2017). However, the above models of neural trajectories do not elucidate the biological basis of learning and replay in neurophysiological situations encountered by PFC networks for several reasons. First, in these models, trajectory learning is either ignored (hard-written trajectory engram; Chenkov et al., 2017), unrelated to behavior (random formation of arbitrary trajectory; Liu and Buonomano, 2009; Fiete et al., 2010), based on artificial learning rules (Laje and Buonomano, 2013) or on biophysically unrealistic rules in terms of neuronal activity and synaptic plasticity constraints (Liu and Buonomano, 2009; Fiete et al., 2010; Klampfl and Maass, 2013). Moreover, trajectory replay is absent (Clopath et al., 2010) or unable to operate from an initial trigger (Klampfl and Maass, 2013), or the ability to memorize and replay trajectory engrams and replays long-term is not tested (Liu and Buonomano, 2009; Clopath et al., 2010; Fiete et al., 2010; Klampfl and Maass, 2013; Laje and Buonomano, 2013; Chenkov et al., 2017). Finally, none of these models evaluate the capacity for trajectory learning and replay in the realistic context where network activity undergoes AI dynamics, whereas it is characteristic of the awake state in the cortex (Destexhe et al., 2003; London et al., 2010). The interactions between synaptic plasticity and AI dynamics has so far only been assessed for static Hebbian engrams (Morrison et al., 2007; Boustani et al., 2012; Litwin-Kumar and Doiron, 2014) but not for dynamic trajectories.

The disordered activity of AI cortical dynamics represents a potentially important source of disturbance at many stages. Indeed, AI regime activity may spontaneously engage plastic processes (before any trajectory presentation), affecting the synaptic network matrix, and leading to altered network dynamics with divergence toward silence or saturation (Siri et al., 2007). Noisy activity may also interfere with the learning of the trajectory engram, by adding erratic entries of calcium to trajectory presentation-induced calcium, leading to jeopardized downstream decoding of calcium as well as erratic switches between long-term potentiation (LTP) and long-term depression (LTD) of synaptic weights. After learning, the continuous effects of AI regime activity-induced plastic processes (LTD or scaling) might erase the trajectory engram during memorization and jeopardize trajectory replay through the destabilizing influence of activity noise. On the other side of the interaction, trajectory learning through Hebbian synaptic plasticity may potentially, in turn, seriously disrupt AI regime activity (Morrison et al., 2007; Siri et al., 2007). Therefore, it remains uncertain whether realistic biological synaptic plasticity rules are well-suited for proper learning and memorizing of trajectory engrams as well as replay of learned trajectories in PFC physiological conditions.

Here, we assessed how learning, memorization and replay of trajectories can arise from biologically realistic synaptic learning rules in physiological PFC networks displaying disordered AI regime activity. To do so, we built a local recurrent biophysical network model designed to capture replay events like those observed in the PFC. Although designed to fit PFC collective spontaneous and triggered neural dynamics, its intrinsic, synaptic and architectural properties are shared across other cortices, allowing for generalization of the results to other non-PFC cortical areas displaying replays. The model displayed AI

dynamics and was endowed with realistic Hebbian (Hebb, 1949) spike timing-dependent plasticity (STDP) of excitatory synapses (Bi and Poo, 1998). Synaptic modifications operate through calcium-signaling dynamics capturing NMDA-dependent non-linear pre- to post-synaptic associativity (Graupner and Brunel, 2012) and calcium-dependent phosphorylation of synaptic weights with realistic activity-dependent kinase/phosphatase (aKP) dynamics, conferring a rapid, graded and bidirectional induction together with slow maintenance, consistent with learning and memory timescales observed in animal and human (Delord et al., 2007). Moreover, the model incorporates synaptic scaling, which ensures normalization of pre-synaptic weights, as found in the cortex (Turrigiano et al., 1998; Wang and Gao, 2012; Sweatt, 2016). We show, that, in this realistic model, presenting a stimulus trajectory allowed for rapid learning of a trajectory engram as well as long-term memorization of the trajectory engram despite the disturbing influence of the AI regime. In turn, the STDP learning rule and trajectory engram did not affect the spontaneous AI regime despite their influence on all excitatory neurons from the network. Moreover, we show that trajectory replay accounted for essential aspects of information coding in the PFC, including robustness of replays at the timescale of seconds, fast and regular replays, chunking, large inter-trial variability, and the ability to account for the dual dynamical and persistent aspects of working memory representations.

MATERIALS AND METHODS

Model of Biophysical Local Recurrent Neural Network

We built a biophysical model of a prefrontal local recurrent neural network, endowed with detailed biological properties of its neurons and connections. While the model is presented as PFC, its synaptic and neural properties are generally preserved across cortical areas, allowing for generalization of the results to non-PFC cortical areas. The network model contained N neurons that were either excitatory (E) or inhibitory (I) (neurons projecting only glutamate or GABA, respectively; Dale, 1935), with probabilities p_E and $p_I = 1 - p_E$, respectively, and $\frac{p_E}{p_I} = 4$ (Beaulieu et al., 1992). Connectivity was sparse (i.e., only a fraction of all possible connections exists, see $p_{E\to E}$, $p_{E\to I}$, $p_{I\to E}$, $p_{I\to I}$ parameter values; Thomson, 2002) with no autapses (self-connections) and EE connections (from E to E neurons) drawn to insure the over-representation of bidirectional connections in cortical networks (four times more than randomly drawn according to a Bernoulli scheme; Song et al., 2005; Wang et al., 2006). The synaptic weights $w_{(i,j)}$ of existing connections were drawn identically and independently from a log-normal distribution of parameters μ_w and σ_w (Song et al., 2005).

To cope with simulation times required for the massive explorations ran in the parameter space, neurons were modeled as leaky integrate-and-fire (LIF) neurons. The membrane potential of neuron j followed

$$\begin{cases} C\frac{dV_{(j)}}{dt} = -(I_{L(j)} + I_{Syn.Rec(j)} + I_{Syn.FF(j)}) \\ V_{(j)} > \theta \rightarrow V_{(j)} = V_{rest} \end{cases}$$

where neurons spike when the membrane potential reaches the threshold θ, and repolarization to V_{rest} occurred after a refractory period $\Delta\ t_{AP}$.

The leak current followed

$$I_{L(j)} = \overline{g}_L\ \left(V_{(j)} - V_L\right)$$

where $\overline{g}_L$ is the maximal conductance and V_L the equilibrium potential of the leak current.

The recurrent synaptic current on post-synaptic neuron j, from—either excitatory or inhibitory—pre-synaptic neurons (indexed by i), was

$$I_{Syn.Rec(j)} = \sum_i \Big(I_{AMPA(i,j)} + I_{NMDA(i,j)} + I_{GABA_A(i,j)} + I_{GABA_B(i,j)} \Big)$$

The delay for synaptic conduction and transmission, Δt_{syn}, was considered uniform across the network (Brunel and Wang, 2001). Synaptic recurrent currents followed

$$I_{x(i,j)} = \overline{g}_x\ w_{(i,j)}\ p_{x(i)}\ \left(V_{(j)} - V_x\right)$$

where $w_{(i,j)}$ is the synaptic weight, $p_{x(i)}$ the opening probability of channel-receptors and V_x the reversal potential of the current. The NMDA current followed

$$I_{NMDA(i,j)} = \overline{g}_{NMDA}\ w_{(i,j)}\ p_{NMDA(i)}\ x_{NMDA}\left(V_{(j)}\right)\left(V_{(j)} - V_{NMDA}\right)$$

incorporating the magnesium block voltage-dependence modeled (Jahr and Stevens, 1990) as

$$x_{NMDA}\ (V) = \left(1 + \left[Mg^{2+}\right] e^{-0.062\ V}/3.57\right)^{-1}$$

The channel rise times were approximated as instantaneous (Brunel and Wang, 2001) and bounded, with first-order decay

$$\frac{dp_{x(i)}}{dt} = -\frac{p_{x(i)}}{\tau_x} + p_x\ \left(1 - p_{x(i)}\right)\ \delta\left(t - t_{(i)}\right)$$

where δ is the dirac function and $t_{(i)}$ the times of the pre-synaptic action potentials (APs).

Recurrent excitatory and inhibitory currents were balanced in each post-synaptic neuron (Shu et al., 2003; Haider et al., 2006; Xue et al., 2014), according to driving forces and excitation/inhibition weight ratio, through

$$\begin{cases} \overline{g}_{GABA_A} = g_{GABA_A}\ \frac{-(V_{mean} - V_{AMPA})}{(V_{mean} - V_{GABA_A})}\ \frac{\sum_{i\in Exc} w_{(i,j)}}{\sum_{i\in Inh} w_{(i,j)}} \\ \overline{g}_{GABA_B} = g_{GABA_B}\ \frac{-(V_{mean} - V_{AMPA})}{(V_{mean} - V_{GABA_B})}\ \frac{\sum_{i\in Exc} w_{(i,j)}}{\sum_{i\in Inh} w_{(i,j)}} \end{cases}$$

with $V_{mean} = \frac{(\theta + V_{rest})}{2}$ being an approximation of the average membrane potential.

Furthermore, all recurrent maximal conductances were multiplied by g_{Rec}, and by $g_{E\to E}$, $g_{E\to I}$, $g_{I\to E}$ or $g_{I\to I}$ according to the excitatory or inhibitory nature of pre- and post-synaptic populations.

The feed-forward synaptic current $I_{Syn.FF(j)}$ (putatively arising from sub-cortical and cortical inputs) consisted of an AMPA component.

$$I_{Syn.FF(j)} = \overline{g}_{AMPA}\, p_{AMPA.FF}\, \left(V_{(j)} - V_{AMPA}\right)$$

with a constant opening probability $p_{AMPA.FF}$.

Synaptic Spike Timing-Dependent Plasticity (STDP)

We used a biophysical model of spike timing-dependent plasticity of excitatory synapses of the network. This rule operated constantly on the weights of the excitatory synapses during simulations. Synaptic weights evolved according to a first-order dynamic (Shouval et al., 2002; Delord et al., 2007) under the control of intra-synaptic calcium (Graupner and Brunel, 2012) through

$$\dot{w}_{(i,j)}(t) = K_{\max}\frac{Ca\,(t)^{nH}}{{K_{Ca}}^{nH} + Ca\,(t)^{nH}} - P_{\max}\frac{Ca\,(t)^{nH}}{{P_{Ca}}^{nH} + Ca\,(t)^{nH}}\, w_{ij}$$

where the plastic modifications of the synapses, i.e., the phosphorylation and dephosphorylation processes of the synaptic receptor channels, depended on a kinase (e.g., PKC type) and a phosphatase (e.g., calcineurin type) whose allosteric activation was dependent on calcium. Here, $K_{\max}$ represents the maximum reaction rate of the kinase, $P_{\max}$ that of the phosphatase, K_{Ca} and P_{Ca} the calcium half-activation concentration, Ca the synaptic calcium concentration and nH is the Hill's coefficient. The term t-LTP, kinase-related, was independent of synaptic weight ("additive" t-LTP) while t-LTD, phosphatase-related, was weight-proportional ("multiplicative" t-LTD), consistent with the literature (Bi and Poo, 1998; van Rossum et al., 2000). This model of STDP is extremely simple, but a detailed implementation would be prohibitive in an RNN of the order of a thousand neurons. There was no term related to the auto-phosphorylation of CaMKII present in many models to implement a form of molecular memory, because on one hand it is not actually involved in the maintenance of memory of synaptic modifications (Chen et al., 2001), and on the other hand memory is ensured here by the dynamics of kinase and phosphatase at low calcium concentration (Delord et al., 2007).

The time dependence of the APs (Bi and Poo, 1998; He et al., 2015) came from calcium dynamics, according to the model of Graupner and Brunel (2012). In this model, synaptic calcium followed

$$Ca\,(t) = Ca_0 + Ca_{pre}(t) + Ca_{post}\,(t)$$

where the total calcium concentration takes into account pre- and post-synaptic calcium contributions.

Pre-synaptic spiking mediated calcium dynamics followed

$$\dot{Ca}_{pre}(t) = -\frac{Ca_{pre}(t)}{\tau_{Ca}} + \Delta Ca_{pre}\sum_i \delta\left(t - t_{(i)} - D\right)$$

where the first term corresponds to calcium extrusion/buffering with time constant τ_{Ca} and the second term to voltage-dependent calcium channels (VDCC)-mediated calcium entry due to pre-synaptic spiking, with Ca_{pre} the amplitude of calcium entering at each AP of the presynaptic neuron, $t_{(i)}$ the times of the pre-synaptic APs, and D a delay modeling the time required for the activation of AMPA channels, the depolarizing rise of the associated excitatory post-synaptic potential (EPSP) and the subsequent opening of VDCC that induces this calcium entry.

Post-synaptic spiking-mediated calcium dynamics evolved according to

$$\dot{Ca}_{post}(t) = -\frac{Ca_{post}(t)}{\tau_{Ca}} + \Delta Ca_{post}\sum_j \delta\left(t - t_{(j)}\right) + \xi_{PrePost}\sum_j \delta\left(t - t_{(j)}\right) Ca_{pre}(t)$$

and modeled extrusion/buffering (first-term) as well as calcium entries due to post-synaptic, back-propagated spiking from the post-synaptic soma along the dendritic tree to the synapse, opening VDCC (central term) and NMDA channels (right term). $\xi_{PrePost}$ is an interaction coefficient and $t_{(j)}$ corresponds to the AP time of the post-synaptic neuron. NMDA activation is non-linear and depends on the product of a pre- and a post-synaptic term, representing the dependence of NMDA channel openings on the associative conjunction of pre-synaptic glutamate and post-synaptic depolarization, which releases the magnesium blockade of NMDA channels.

Synaptic Scaling

Synaptic weights were subjected to a homeostatic form of synaptic normalization, present in the cortex (Turrigiano et al., 1998; Wang and Gao, 2012; Sweatt, 2016), which was modeled in a simplified, multiplicative and instantaneous form (Zenke et al., 2013), following at each time step

$$w_{(ij)}(t + dt) = w_{(ij)}(t)\frac{\sum_i w_{ij}(t = 0)}{\sum_i w_{ij}(t)}$$

This procedure ensured that the sum of the incoming weights on a post-synaptic neuron remained constant despite the plastic modifications due to STDP.

Estimation of the Time Constant of STDP With Synaptic Scaling

Without synaptic scaling, $\dot{w}_{ij} = \dot{w}_{STDP} = K\,(Ca) - P\,(Ca)\,w$. However, synaptic scaling plays an important role in the slow decay of weights, so to study the time constant of this decay we needed to incorporate the effect of synaptic scaling. Considering n weights of average value μ_w incoming upon a post-synaptic neuron, where a proportion p of weights undergo STDP of value

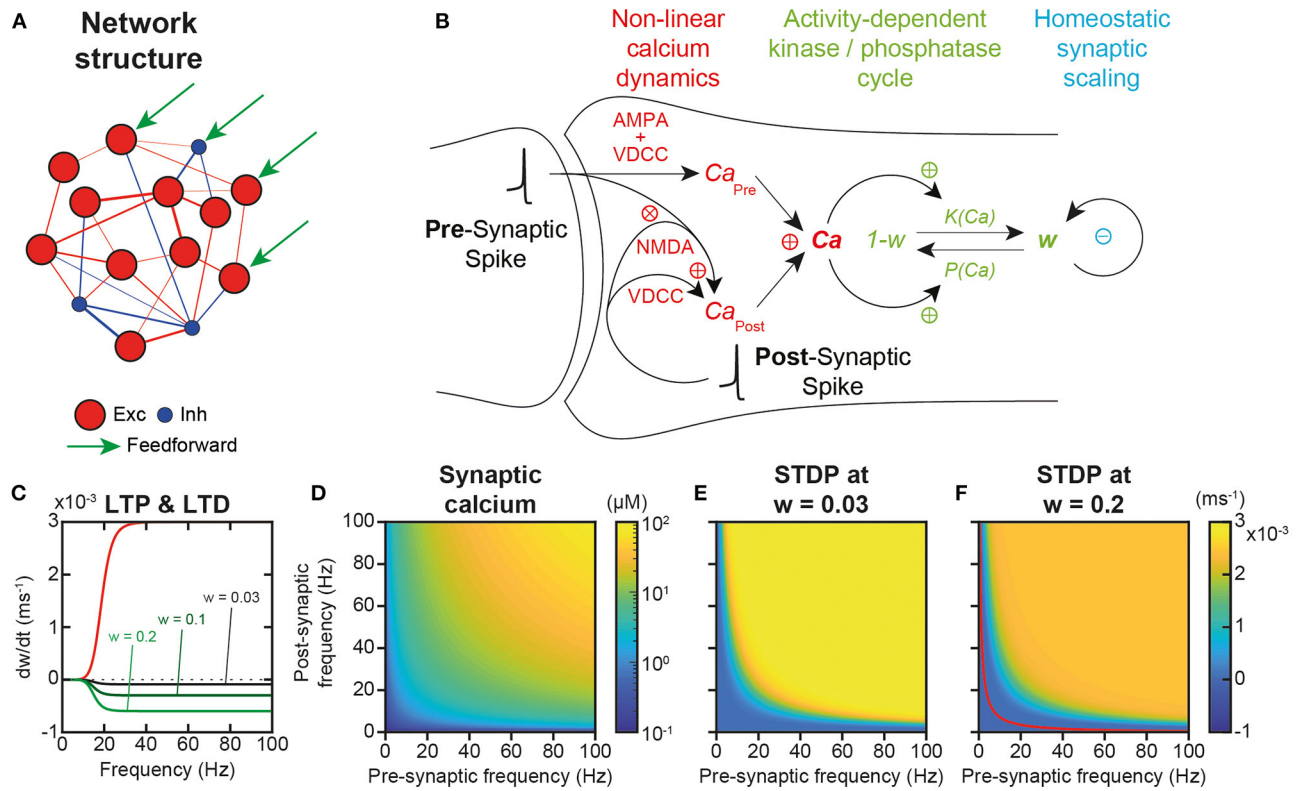


FIGURE 1 | Network structure and plastic properties. **(A)** Scheme of the recurrent network model of the local prefrontal cortex circuit with 484 excitatory (red) and 121 inhibitory (blue) integrate and fire (IAF) neurons. **(B)** Scheme of excitatory synaptic plastic processes. In the post-synaptic compartment, calcium dynamics originates from two distinct sources (Ca_{Pre} and Ca_{Post}), as well as from extrusion/buffering (Graupner and Brunel, 2012). Ca_{Pre} arises from pre-synaptic spiking mediated through α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) depolarization and the subsequent activation of voltage-dependent calcium (VDCC) channels. Ca_{Post} models calcium entries due to post-synaptic spiking back-propagated from the post-synaptic soma to the synapse, opening VDCC and N-methyl-D-aspartate (NMDA) channels. NMDA activation is non-linear and depends on the interaction of pre- and post-synaptic spiking to account for the associative dependence of NMDA channel openings on the conjunction of pre-synaptic glutamate and post-synaptic depolarization that releases magnesium blockade. Plastic modifications operate through calcium-dependent phosphorylation and dephosphorylation of channel AMPA receptors that determine the synaptic weight (aKP model; Delord et al., 2007). Synaptic scaling continuously normalizes weights so as to insure the homeostatic regulation of the sum of incoming (pre-synaptic) weights for each individual neuron (Turrigiano et al., 1998). **(C)** Both long-term spike timing-dependent potentiation (t-LTP) and long-term spike timing-dependent depression (t-LTD) increase non-linearly with pre- and post-synaptic spiking frequency ($\nu_{Pre} = \nu_{Post} = \nu$), due to the allosteric calcium-activation of both enzymes. Kinase-mediated t-LTP is additive, i.e., independent of synaptic weight, while phosphatase-mediated t-LTD is multiplicative, i.e., weight-proportional (Bi and Poo, 1998; van Rossum et al., 2000). **(D)** Because of the associative dependance of NMDA-mediated calcium entry to pre- and post-spiking, synaptic calcium depends multiplicatively on pre- and post-synaptic spiking frequencies. **(E)** In the spontaneous AI regime, plastic modifications are virtually null because STDP plasticity occurs similarly at all synapses, with synaptic scaling compensating STDP (see Results). **(F)**. In synapses connecting neurons in the engram of a learned trajectory, where plasticity has occurred in a subset of synapses, Hebbian t-LTP dominates at large multiplicative pre-/post- frequencies and Hebbian t-LTD at lower frequencies (separated by the red curve for which plasticity is null, see Results).

$\dot{w}_{STDP}$ at time step t followed by scaling, then for a given weight w within the proportion p,

$$w(t+\Delta t) = \left(w(t) + \dot{w}_{STDP}\Delta t\right)\left(\frac{n\mu_w}{n\mu_w + np\dot{w}_{STDP}\Delta t}\right)$$

so that after algebra, one obtains

$$\frac{w(t+\Delta t) - w(t)}{\Delta t} = \left(1 - p\frac{w(t) + \dot{w}_{STDP}\Delta t}{\mu_w + p\dot{w}_{STDP}\Delta t}\right)\dot{w}_{STDP}$$

Passing to the limit $\Delta t \to 0$, one finds:

$$\dot{w} = \left(1 - p\frac{w}{\mu_w}\right)\dot{w}_{STDP}$$

i.e.

$$\dot{w} = \left(1 - p\frac{w}{\mu_w}\right)\left(K(Ca) - P(Ca)w\right)$$

To find an estimate of the time constant of plasticity, linearization aroundμ_w gives

$$\dot{w} \sim \left(P(Ca)\left(2p-1\right) - \frac{K(Ca)p}{\mu_w}\right)w + K(Ca) - pP(Ca)\mu_w$$

so that

$$\tau \sim \frac{\mu_w}{\left|pK\left(Ca\right) - \left(2p-1\right)P\left(Ca\right)\mu_w\right|}$$

Theoretical Dependences Under Asynchronous Irregular Dynamics

The steady-state theoretical concentration of calcium in individual synapses was obtained from fixed-points of Ca_{Pre} and Ca_{Post}, which yielded

$$Ca^*(\nu_{Pre}, \nu_{Post}) \sim Ca_0 + \tau_{Ca}(\Delta Ca_{Pre}\nu_{Pre} + \Delta Ca_{Post}\nu_{Post} + \xi_{PrePost}\Delta Ca_{Pre}\nu_{Pre}\nu_{Post})$$

which was used to determine STDP modification rates

$$\dot{w} = K(Ca^*) - P(Ca^*)w$$

and to determine the time constant for plasticity, in the case of the network asynchronous irregular regime at low frequency, where $p = 1$, i.e.

$$\tau \sim \frac{\mu_w}{\left|K\left(Ca^*\right) - P\left(Ca^*\right)\mu_w\right|}$$

Weights Within and Outside the Engram

Initial excitatory weights (before the 1 h simulation) were convolved with a centered normalized Gaussian function ($\sigma = 5$ *neurons*). Convolved weights with values above 0.1 (times $p_{E\to E} = 0.35$ to take into account inexistent weights) were considered within the engram, the other weights were considered outside the engram. Both weight populations were kept constant and their evolution was studied across time (see **Figures 6**, **7**).

Trajectory Replay Detection

In order to detect coherent propagating activity pulse packets along the synaptic pathway, we convolved spiking activity across time and neurons with centered normalized Gaussian functions ($\sigma = 30$ *ms* and $\sigma \sim 10$ *neurons*). Neurons were considered "active" when at least 40% of the convolved frequencies which include them (>5% of normalized Gaussian function maximum) are above 12.5*Hz*. We considered the emergence of an activity packet when it contained more than 20 neurons.

Spiking Irregularity

To capture spiking irregularity, we quantified the CV (coefficient of variation), CV2 and Lv (time-local variation) of the inter-spike interval (ISI) distribution of the spiking trains of neurons in the network (Compte, 2003; Shinomoto et al., 2005) according to

$$CV = \frac{\sigma_{ISI}}{< ISI >}$$

$$CV_2 =< 2\frac{\left|ISI_{k+1} - ISI_k\right|}{ISI_{k+1} + ISI_k} >_k$$

$$Lv =< 3\frac{\left(ISI_k - ISI_{k+1}\right)^2}{\left(ISI_k + ISI_{k+1}\right)^2} >_k$$

where $CV = CV_2 = Lv = 1$ for a homogeneous Poisson spike train and $= 0$ for a perfectly regular spike train where all *ISI* are equal. CV stands around 1 to 2 *in vivo* (Compte, 2003; Shinomoto et al., 2005), representing the global variability of an entire ISI sequence, but is sensitive to firing rate fluctuations. CV2 and Lv stand around 0.25 to 1.25 and 0 to 2, respectively *in vivo* (Compte, 2003; Shinomoto et al., 2005), evaluating the ISI variability locally in order to be less sensitive to firing rate fluctuations. The CV was calculated on every ISI across neurons, while the CV2 and Lv were calculated for each excitatory neuron and averaged across the whole population.

Spiking Synchrony

Three measures of synchrony were adopted, a synchrony measure *S* (Golomb et al., 2001), pairwise correlation coefficient averaged over all pairs of excitatory neurons $< \rho >$ (Tchumatchenko et al., 2010), and Fano factor *F*. The first two were calculated on the estimated instantaneous neural frequency *f* (Gaussian convolution of spikes, $\sigma = 30ms$), while the last was calculated on the population sum of spike counts *s*, following

$$S = \sqrt{\frac{Var\left(< f >_n\right)}{< Var\left(f_{(n)}\right) >_n}}$$

$$< \rho >= \frac{1}{N(N-1)/2}\sum_i\sum_{j>i}\frac{cov\left(f_{(i)}, f_{(j)}\right)}{\sqrt{Var\left(f_{(i)}\right)Var\left(f_{(j)}\right)}}$$

$$F = \frac{Var\left(\sum_n s_n\right)}{< \sum_n s_n >_t}$$

These measures equal $S = \frac{1}{\sqrt{n_E}} \sim 0.0455$, $< \rho >= 0$ and $F = 1$ for perfectly asynchronous network activity, and $S =< \rho >= 1$ while F increases for perfectly synchronous network activity.

Procedures and Parameters

Models were simulated and explored using custom developed code (MATLAB) and were numerically integrated using the forward Euler method with time-step $\Delta t = 0.5ms$ in network models. Unless indicated in the text, standard parameter values were as following. Concerning the network architecture, $N = 605$ *neurons*, $n_E = 484$ *neurons*, $n_I = 121$ *neurons*, $p_{E\to E} = 0.35$, $p_{E\to I} = 0.2056$, $p_{I\to E} = 0.22$, $p_{I\to I} = 0.25$, $\mu_w = 0.03$, $\sigma_w = 0.02$. Concerning the Integrate-and-Fire neural properties, $C = 1\ \mu F.cm^{-2}$, $\theta = -52\ mV$, $V_{rest} = -67\ mV$, $\Delta t_{AP} = 3\ ms$. Concerning currents, $\bar{g}_L = 0.05\ mS.cm^{-2}$, $V_L = -70\ mV$, $\Delta t_{syn} = 0.5\ ms$, $\bar{g}_{AMPA} = 0.23\ mS.cm^{-2}$, $\bar{g}_{NMDA} = 0.9\ mS.cm^{-2}$, $g_{GABA_A} = 0.3\ mS.cm^{-2}$, $g_{GABA_B} = 0.017\ mS.cm^{-2}$, $V_{AMPA} = V_{NMDA} = 0\ mV$, $V_{GABA_A} = -70\ mV$, $V_{GABA_B} = -90\ mV$, $\left[Mg^{2+}\right] = 1.5\ mM$, $\tau_{AMPA} = 2.5\ ms$, $\tau_{NMDA} = 62\ ms$, $\tau_{GABA_A} = 10\ ms$, $\tau_{GABA_B} = 25\ ms$, $p_{AMPA} = p_{NMDA} = p_{GABA_A} = p_{GABA_B} = 0.1$, $g_{Rec} = 0.65$, $g_{E\to E} = g_{E\to I} = g_{I\to E} = 1$, $g_{I\to I} = 0.7$, $p_{AMPA.FF} \sim 0.0951$. Concerning synaptic properties, $K_{\max} = 3.10^{-3}\ ms^{-1}$, $K_{Ca} = 3\ \mu M$, $P_{\max} = 3.10^{-3}\ ms^{-1}$, $P_{Ca} = 2\ \mu M, nH = 4$, $Ca_0 = 0.1\ \mu M$, $\tau_{Ca} = 100\ ms$, $\Delta Ca_{pre} = 0.02\ \mu M$, $D = 10\ ms$, $\Delta Ca_{post} = 0.02\ \mu M$, $\xi_{PrePost} = 4\ ms^{-1}$.

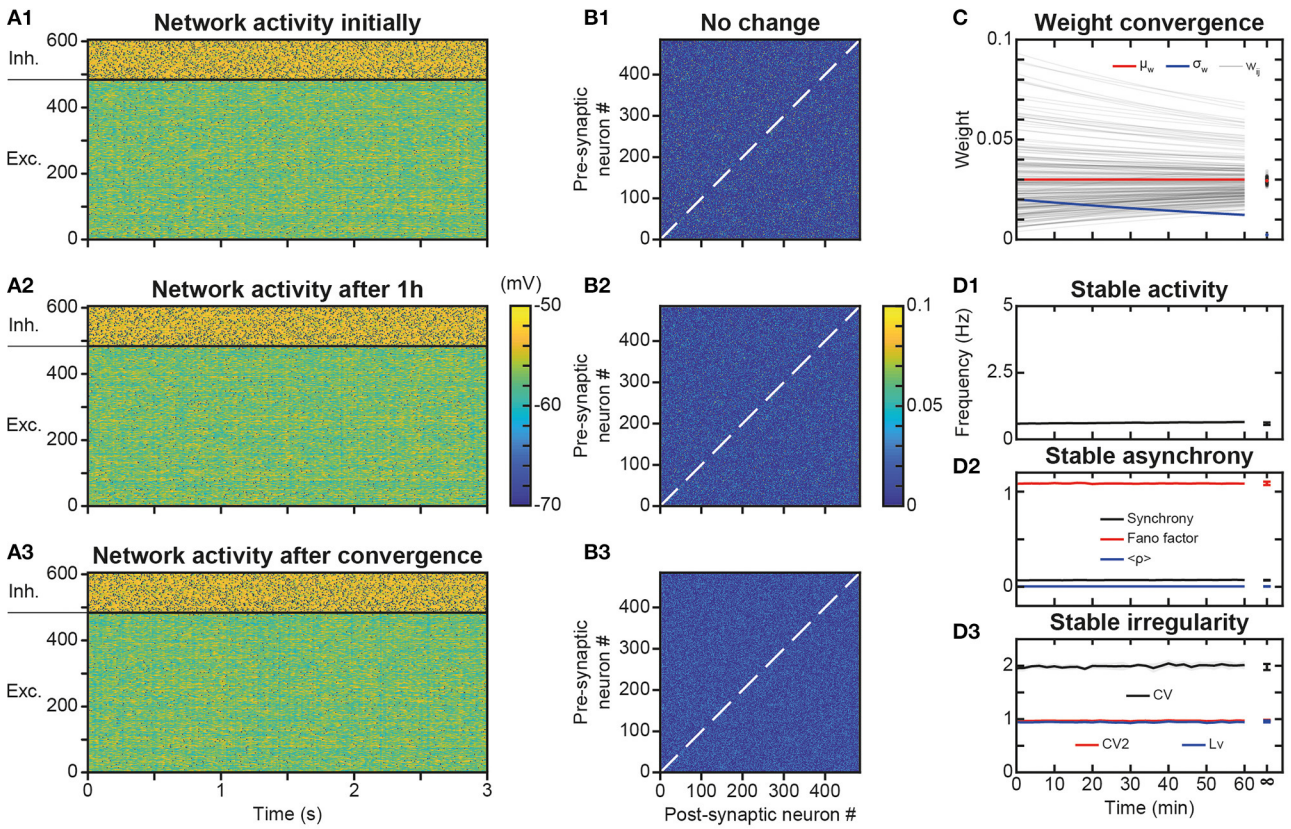


FIGURE 2 | Stability of spontaneous irregular asynchronous (AI) network dynamics under synaptic plasticity. **(A1–A3)** Membrane potential of network neurons during 3 s of spontaneous AI regime in the absence of plasticity **(A1)**, after 1 h of plasticity **(A2)** and after full convergence of synaptic weights due to plasticity **(A3)**. The same initial random connectivity matrix is used for simulations in **(A1–A3)**. Spikes indicated by black dots. Full convergence of the synaptic matrix was obtained by simulating the networks with very fast kinetic constants. **(B1–B3)** Synaptic weights between excitatory neurons of the network at the end of each of simulations presented in **A1–A3**. **(C)** Convergence of synaptic weights toward the mean weight of their post-synaptic neuron as a function of time, due to synaptic scaling normalization (black curves, see Results). Time evolution of the mean (red curve) and standard deviation (blue curve) of synaptic weights. For sake of clarity, only a random selection of synapses is shown. The mean is constant and the standard deviation decreases with time, due to scaling. **(D)**. Average excitatory neural spiking frequency **(D1)** and irregularity **(D3)** and excitatory population synchrony **(D2)** quantifiers, as a function of time, for five different simulations of the network with different realizations of the initial random synaptic matrix. Dots on the right indicate values obtained from network simulations after full convergence of synaptic weights. Shaded areas represent 95% confidence intervals of the mean.

RESULTS

Predicting Fundamental Plastic Properties of PFC Recurrent Networks

To evaluate neural trajectory learning, memorization and replay, we studied a local prefrontal cortex (PFC) recurrent network model, with 484 excitatory and 121 inhibitory integrate and fire (IAF) neurons with topographically tuned feed-forward inputs. Synaptic connections were constrained by cortical connectivity data, following Dale's law, sparseness and log-normal weight distributions, and α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) and N-methyl-D-aspartate (NMDA) excitatory and γ-aminobutyric acid (GABA-A and GABA-B) inhibitory synaptic currents (**Figure 1A;** see *Materials and Methods*). Most synaptic and neural properties, while present in PFC, are generic across cortex, such that the following results can be generalized to non-PFC cortical areas.

Excitatory synapses were plastic, i.e., endowed with realistic calcium dynamics (Graupner and Brunel, 2012) accounting for linear voltage-dependent calcium channels (VDCC)-dependent and non-linear NMDA calcium entries, as well as for linear extrusion and buffering (**Figure 1B**). These calcium dynamics are responsible for the temporal asymmetry of pre- and post-synaptic spike-timing dependent (STDP) plastic modifications (Bi and Poo, 1998; He et al., 2015). Note, however, that with these realistic calcium dynamics, plasticity essentially depends on firing frequency rather than on the precise timing of spikes, because of the frequency and variability of *in vivo*-like spiking (Graupner et al., 2016).

Plastic modifications operated through calcium-dependent kinase-phosphatase kinetics (Delord et al., 2007), which accounts for their fast induction and slower maintenance dynamics (**Figure 1B**). No Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) auto-phosphorylation was present because it

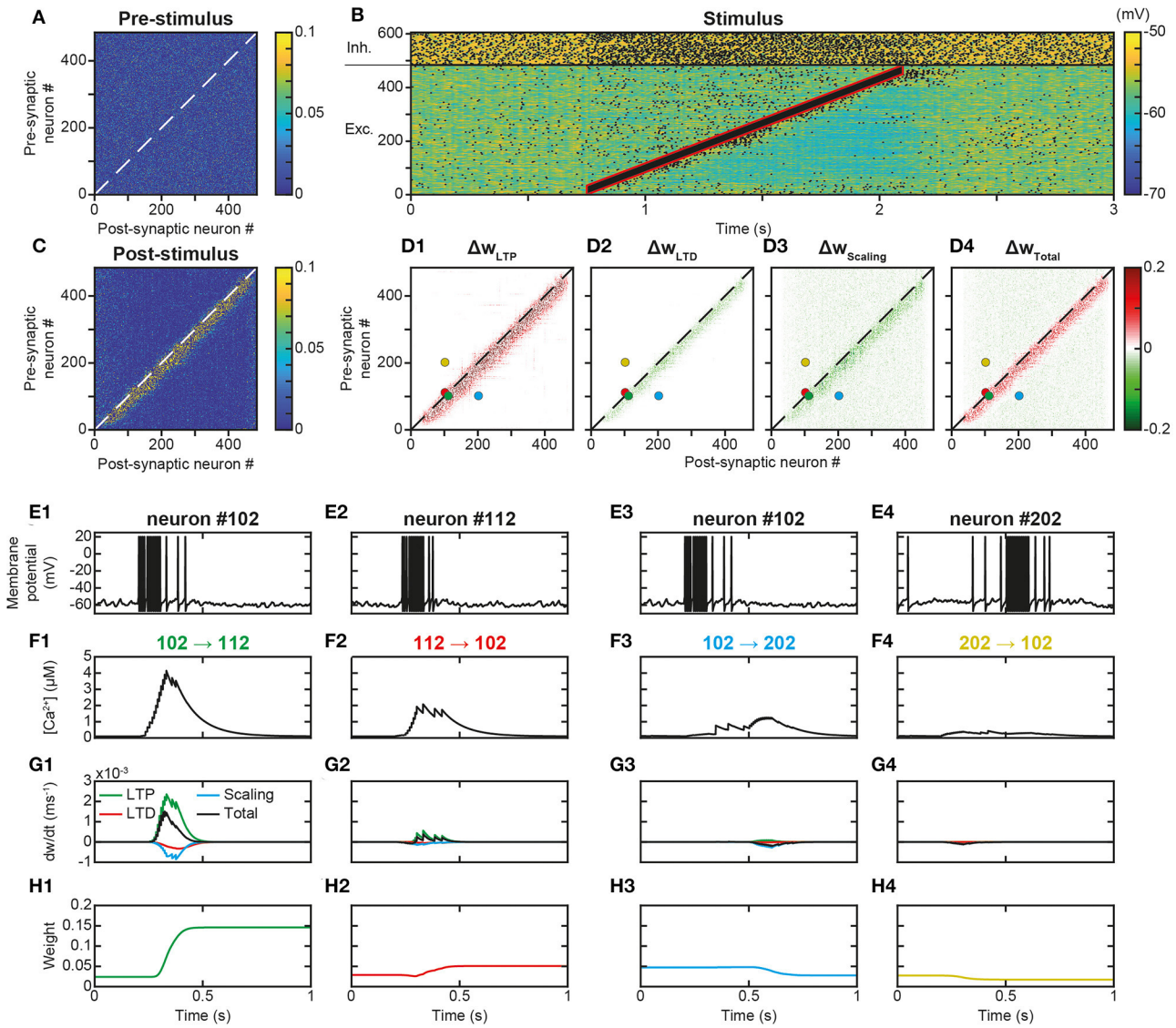


FIGURE 3 | Learning a trajectory stimulus into a trajectory engram. **(A)**. Synaptic matrix between excitatory neurons prior to stimulus presentation. **(B)**. Membrane potential of network neurons in response to the presentation of a trajectory stimulus (stimulus in red) that successively activates all excitatory neurons over a duration of 1,350 ms. Spikes indicated by black dots. **(C)**. Synaptic matrix between excitatory neurons after stimulus presentation. **(D1–D4)**. Weight modifications resulting, after trajectory presentation, from t-LTP **(D1)**, t-LTD **(D2)**, scaling **(D3)**, and their sum **(D4)**. **(E–H)** Membrane potential **(E)**, calcium **(F)**, plastic rates **(G)** and synaptic weight dynamics **(H)** during the passage of the trajectory stimulus in a pair of neurons with nearby topographical tuning #102 **(E1)** and #112 **(E2)** and their reciprocal connections 102→112 **(F1-H1)** and 112→102 **(F2-H2)**, and in a pair of neurons with more distant topographical tuning #102 **(E3)** and #202 **(E4)** and their reciprocal connections 102→202 **(F3-H3)** and 202→102 **(F4-H4)**.

is actually not involved in the maintenance of synaptic modifications (Chen et al., 2001; Lengyel et al., 2004). Rather, the long-term maintenance of plastic modifications emerges from kinase and phosphatase dynamics at low calcium concentrations (see below; Delord et al., 2007). Besides, synapses underwent synaptic scaling (**Figure 1B**), which ensures total weight normalization at the neuron level, as observed in the cortex (Turrigiano et al., 1998; Wang and Gao, 2012; Sweatt, 2016) and, as a consequence, introduces competition between synaptic weights within each neuron (intra-neuronal inter-synaptic competition).

Most importantly, plasticity operated online—i.e., permanently, without offline learning periods—on excitatory synaptic weights, as a function of neuronal activity in the network, whether it corresponds to the spontaneous, asynchronous and irregular (AI) activity of the network, the activity evoked by the feed-forward currents during the input presentation of an example trajectory, or the replay activity

after learning (see below). Both kinase-mediated long-term spike timing-dependent potentiation (t-LTP) and phosphatase-mediated long-term spike timing-dependent depression (t-LTD) increased non-linearly with pre- and post-synaptic spiking frequency, due to the allosteric activation of enzymes by calcium (**Figure 1C**). However, they differed in that kinase-mediated t-LTP was independent of synaptic weight (additive or hard-bounded) while phosphatase-mediated t-LTD was weight-proportional (multiplicative or soft-bounded), consistent with the literature (Bi and Poo, 1998; van Rossum et al., 2000; **Figure 1C**). In the model, the steady-state theoretical concentration of calcium in individual synapses depended multiplicatively upon pre-synaptic and post-synaptic spiking activity (**Figure 1D**), from which one could compute the rate of STDP as a function of pre- and post-synaptic spiking frequency (**Figures 1E,F**) see Materials and Methods). In conditions with weak synaptic weights, such as prior to learning, t-LTP dominated at all frequencies because t-LTD is multiplicative and thus scaled by, here, very low synaptic weights. Thus, STDP effects were always positive and depended multiplicatively on pre- and post-synaptic frequencies (**Figure 1E**). By contrast, when plasticity had previously occurred ($w = 0.2$), such as in the engram of a learned trajectory (see below), t-LTD was stronger due to the stronger weights, and the model predicted Hebbian t-LTP at large multiplicative pre-/post-frequencies and t-LTD at lower frequencies (**Figure 1F**). In the following, we explore the extent to which these predictions are correct in simulations of the whole network model under spontaneous AI dynamics with synaptic scaling, and when assessing learning and memorization upon trajectory presentation.

Stability of Network AI Dynamics Under Synaptic Plasticity

A potential issue of synaptic plasticity in network models remains its sensitivity to spontaneous activity. Hence, before testing the possible role of STDP in trajectory learning and replay, we first studied the effect of STDP on the spontaneous regime, with the aim of verifying that network activity remained stable over the long term and that neurons always discharged in the AI regime. Indeed, Hebbian or post-Hebbian rules of the STDP type, by modifying the matrix of synaptic weights, may lead to saturation of neuronal activity and a collapse of the complexity of the dynamics, from initially AI chaotic activity characteristic of the waking state (Destexhe et al., 2003; London et al., 2010), to activity of the limit-cycle or fixed point type (Siri et al., 2007). We considered here as long term the 1 h time scale, which is the scale classically used experimentally to test the memory of synaptic plasticity modifications (Bi and Poo, 1998). Moreover, a duration of 1 h extends way beyond the classical time scales used in models (Morrison et al., 2007; Boustani et al., 2012; Litwin-Kumar and Doiron, 2014). For this purpose, we have observed the activity (**Figure 2A**) and connectivity (**Figure 2B**) of the network at different time scales, in order to reveal possible modifications in the network behavior.

Simulations showed that the spontaneous activity of the network was identical without plasticity (**Figure 2A1**), after 1 h in the presence of plasticity (**Figure 2A2**) and after full convergence (**Figure 2A3**) of weight matrix dynamics. This observation is consistent with the absence of changes in the connectivity matrix in the presence of STDP, even after 1 h of simulation (**Figures 2B2,B3**), compared to the condition without STDP (**Figure 2B1**). Mechanistically, the low spiking frequency of neurons resulted in moderate average elevations of calcium above its basal concentration in synapses, so that kinase and phosphatase were only very weakly activated. Therefore, weights underwent extremely slow plastic modifications where additive t-LTP (which dominated the multiplicative t-LTD at weak weights) was compensated by synaptic scaling. Due to these effects, weights converged toward the mean initial weight of their post-synaptic neuron (**Figure 2C**) with an apparent time constant of 2 h, close to the theoretical estimation of the time constant of plasticity (see *Materials and Methods* and *Discussion*), which predicts a time constant of 1.95 h during learning at low spiking frequencies and calcium concentrations ($Ca \sim Ca_0$) in the AI regime. These steady-state values were normally distributed, with a constant mean value (due to the synaptic scaling) and a decreasing standard deviation, due to the homogenization of weights within each post-synaptic neuron (**Figure 2C**). Even with this more homogeneous synaptic matrix (**Figure 2B3**), AI dynamics were preserved (**Figure 2A3**). Indeed, excitatory frequency was stable (**Figure 2D1**), as well as markers of synchrony (**Figure 2D2**) and irregularity (**Figure 2D3**). Thus, overall, the activity regime of the network was not altered by the presence of plastic processes. Note that in PFC circuits experiencing dynamically changing feed-forward inputs, convergence of the synaptic matrix may be attenuated or even non-existent.

Learning Trajectory Engrams Under AI Dynamics

Trajectory learning during network activity has already been investigated in the theoretical literature, but either without chaotic dynamics or using biologically unrealistic learning rules (see *Introduction*). To test for the possibility of learning trajectories within physiologically irregular activity, we presented to the network a moving stimulus (**Figure 1A**, feedforward connections) that successively activated all the excitatory neurons over 1,350 ms (**Figure 3B**). Such a stimulation corresponds to a displacement speed of ∼0.3 neurons/ms, where each excitatory neuron was stimulated for ∼100 ms and discharged at ∼100 Hz. This single stimulus presentation triggered neural activity much stronger than the spontaneous activity, sufficient to modify the matrix of synaptic weights. Indeed, whereas the synaptic matrix was initially formed of low random weights (**Figure 3A**), after presentation, the weights of synapses connecting neurons activated by the stimulus at close successive times were increased (**Figure 3C**). This diagonal band of increased weights formed an oriented connectivity path along stimulus-activated neurons and is referred to as the trajectory engram hereafter. Weight modifications inside and outside this trajectory engram resulted from increases due to t-LTP (**Figure 3D1**, Δw_{LTP}) and decreases due to t-LTD (**Figure 3D2**, Δw_{LTD}). Moreover, the homeostatic

process of synaptic scaling, which ensures the constancy of the sum of the incoming weights of the cortical neurons, decreased the total incoming synaptic weights on post-synaptic neurons, in order to compensate for weight modifications due to STDP (**Figure 3D3**, $\Delta w_{Scaling}$). In fine, STDP and scaling led together to an increase in engram weights and a slight decrease in off-engram weights (**Figure 3D4**, Δw_{Total}; also observe the darker area in **Figure 3C**, compared to **Figure 3A**).

The observation, on a local scale, of the details of the processes at work for the synapses linking the neurons of the engram allowed for a better understanding of these network effects. For illustration, neurons #102 and #112, with close spatial topographical tuning, discharged one following the other with partial overlap during the stimulus (**Figure 3E**). At the level of the synapse between neurons #102 and #112 (102→112), whose orientation was that of the trajectory, the arrival of pre-synaptic action potentials (APs) was followed by that of postsynaptic APs (pre #102 then post #112 neuron, **Figures 3E1,E2**), which triggered a massive input of calcium via the VDCC channels and the NMDA receptor channels (**Figure 3F1**). Conversely, in the synapse 112→102, for which the sequence of arrival of the APs was reversed (pre #112 then post #102 neuron), NMDA channels did not open (see above), such that the calcium input resulted only from the VDCC channels and was thus moderate (**Figure 3F2**). These calcium elevations activated the kinases and phosphatases, which, respectively, phosphorylated and dephosphorylated AMPA channels, increasing (t-LTP) and decreasing (t-LTD) synaptic weights (only phosphorylated AMPA channels are functional and ensure synaptic transmission). These kinase and phosphatase activations were important for synapse 102→112 (**Figure 3G1**), but less so for the synapse 112→102 (**Figure 3G2**). For both synapses (**Figures 3G1,G2**), the phosphatase was more strongly activated (lower half-activation; Delord et al., 2007), but the resulting t-LTD modification rate was low, because it is multiplicative, i.e., it scales with synaptic weight, which was low. Conversely, the rate of modification due to t-LTP was higher because it is additive and depends only on kinase activation (van Rossum et al., 2000). These STDP effects, cumulated with those of scaling, resulted in a positive speed (increase in weight), which was strong for synapse 102→112 (**Figure 3G1**) and very weak for synapse 112→102 (**Figure 3G2**). Together, these plastic processes increased the weight of the synapse oriented in the same direction as the stimulus (**Figure 3H1**) leaving the weight of the synapse of opposite orientation almost unchanged (**Figure 3H2**).

For neurons whose receptive fields were more spatially distant, activation by the stimulus occurred at more temporally distant times (for example, neurons #102 and #202, **Figures 3E3,E4**). In this case, regardless of the sequence of arrival of the APs in both neurons, their succession was too distant in time to open NMDA channels, so that incoming calcium came only from the VDCC channels and was therefore low (**Figures 3F3,F4**). Consequently, kinase and phosphatase were weakly activated, resulting in virtually null STDP velocity (**Figures 3G3,G4**). Synaptic scaling (**Figures 3G3,G4**), induced by the increase of weights in the engram (**Figures 3H1,H2**), ultimately decreased synaptic weights (**Figures 3H3,H4**). As such, there was no learning of any trajectory between distant neurons, contrary to what happened between closer neurons.

Trajectory Replays From Learned Trajectory Engrams

In behaving animals, learnt trajectories are replayed later in appropriate behavioral conditions. In the model, we assessed whether trajectories could be replayed, the dynamics of trajectory replays and the way they affect the network connectivity compared to before they occur (**Figure 4A**). Trajectory replay was defined as the reactivation of neurons of the entire trajectory engram, after temporarily stimulating only initial neurons at the beginning of the engram. To assess trajectory replay in the network, we applied a stimulus of 100 ms to the first 50 neurons of the engram, 500 ms after trajectory learning was completed (**Figure 4B**). We found that the network was able to replay the trajectory entirely after learning (**Figure 4B1**). Fundamentally, the replay emerged because neurons were linked by strong synapses so that preceding neurons activated subsequent neurons in the engram, forming an oriented propagating wave (**Figure 4B2**).

Because it activated neurons at several tens of Hz, the replay could have brought into play plastic processes at the synapses forming the engram, and, in doing so, either reinforce or diminish their weights, possibly disturbing or even destroying the engram. To evaluate these possibilities, we observed the variation of synaptic weights before and after the replay. We found that after replay, the engram was still present (**Figure 4C**) and its structure identical to that before replay (**Figure 4A**). However, when dissecting the effects at work, we found that the engram had slightly thickened during the trajectory replay, due to the combined effect of t-LTP (**Figure 4D1** Δw_{LTP}), t-LTD (**Figure 4D2** Δw_{LTD}) and scaling (**Figure 4D3** $\Delta w_{Scaling}$). Weights above and below the engram increased, whereas weights slightly decreased within the engram (**Figure 4D4**, Δw_{Total}, red fringes).

Up to this point, the neural trajectory was presented as a whole. However, whole trajectories are generally not accessible directly to the PFC. Rather, PFC circuits generally encounter elementary trajectory fragments at separate points in time to produce prospective planning of future behaviors (Ito et al., 2015; Mashhoori et al., 2018; Kaefer et al., 2020), as well as learn transitions between them and chunk fragments together as whole trajectories independently of their presentation order (ordinal knowledge) (Ostlund et al., 2009; Dehaene et al., 2015). We trained the network with four fragments of the whole trajectory, noted A-D, that overlapped at their extremities and which were presented sequentially every 2 s, so as to learn separately different parts of the trajectory (**Figure 4E**). We found that, once fragments were presented in forward order (ABCD), stimulating neurons at the beginning of the A fragment induced propagation of activity that recapitulated the whole trajectory, by subsequently recalling ABCD fragments in the forward order

(**Figure 4E1**). Therefore, the network was able to learn trajectory fragments themselves and the transitions between fragments so as to chunk them into a whole trajectory. Moreover, we found that chunking was possible even when fragments had been learned in reverse order (DCBA; **Figure 4E2**). Hence, the network was able to replay a chunked trajectory based on the presentation of overlapping stimuli, independently of their order of presentation.

Functional Diversity of Trajectory Replays

Neural activity during the replay was less focused than the stimulus trajectory (**Figure 4B**), i.e., it involved more (∼90 vs. 35) neurons, spiking at a lower (∼65 vs. 100 Hz) discharge frequency. The replay also unfolded at a faster speed, lasting ∼750 ms—for a stimulus of 1,350 ms—so that it exhibited a temporal compression factor (tCF) of ∼1.8, which is situated between fast and regular timescale replays observed in animals. Regular timescale replays operate at the timescale of behaviors they were learnt from, i.e., a few seconds (in navigation or working memory tasks, e.g.), hence typically displaying tCF∼1. By contrast, fast timescale replays last several hundred ms in the awake PFC (200–1,500 ms; Jadhav et al., 2016; Mashhoori et al., 2018; Kaefer et al., 2020), yielding several-fold compression factors (tCF∼2–15). We assessed whether varying biophysical parameters of the network could account for durations and tCF ranges characterizing regular and fast replays. As regular and fast timescale replays frequently alternate within trials in behavioral tasks, we discarded trivial replay speed control that can be readily obtained by scaling structural parameters that vary at extremely slow timescales (e.g., number of neurons in the trajectory, synaptic delay, etc., not shown). Rather, we focused on synaptic and intrinsic neuronal properties likely to be rapidly regulated by ongoing neuromodulation in the PFC, as attentional demands or reward outcomes vary at the trial timescale. Among passive and synaptic neuronal parameters tested, the NMDA conductance decay time constant (τ_{NMDA}) emerged as a critical factor controlling the duration and tCF of replays. Hence, the same network, taught with the same trajectory and stimulated with the same initiation stimulus, could generate a large range of replay timescales spanning from regular (duration 1,680 ms, tCF = 0.8; **Figure 5A1**) to fast (duration 375 ms, CF∼3.6; **Figure 5A2**) replays, when the decay time constant of NMDA, τ_{NMDA}, was varied. Consistently, dopaminergic neuromodulation, the major determinant of reward signaling, rapidly slows the decaying dynamics of NMDA currents in PFC circuits (Chen et al., 2004; Onn and Wang, 2005; Onn et al., 2006). Such neuromodulatory effects, as well as others forms of neuromodulation of NMDA dynamics (Lutzu and Castillo, 2021) may control the duration and compression factor of trajectory replays, as well as the relative rate of occurrence of regular vs. fast timescale replays. Inspecting neuronal activity during replays in terms of firing frequency, we found that in single replays individual neurons displayed a sequence of overlapping transient bumps of activity of a few hundred milliseconds (**Figure 5B1**) resembling "relay race" of PFC individual activities during regular replays in working memory tasks (Batuev, 1994; Brody et al., 2003; Cromer et al., 2010; Yang et al., 2014; Schmitt et al., 2017). By contrast, the averaged frequency over the population of excitatory neurons displayed a persistent decaying activity pattern that lasted at the second time scale (**Figure 5B2**) and mimicked population-level working memory maintenance in the PFC (Murray et al., 2017; Cavanagh et al., 2018; Enel et al., 2020). This dichotomy recalls that found in the PFC, whereby individual neurons encode information at short timescale while the population holds stabilized persistent representations on longer timescales (Meyers et al., 2008; Murray et al., 2017; Cavanagh et al., 2018). Moreover, we found that inter-trial variability for each neuron was important, due to disordered network AI dynamics, and that it increased as activity traveled later in the trajectory in individual neurons (**Figure 5B3**) and at the population level (**Figure 5B4**), as found experimentally (Compte, 2003; Shafi et al., 2007; Tiganj et al., 2017).

Globally, the model thus not only indicated that it was possible to learn trajectories online by creating synaptic engrams, thanks to the STDP-type plasticity rule. It also showed that learned trajectories were functional as a memory process, in the sense that their replay was possible and globally preserved the synaptic structure of the learned engram. Finally, the model accounted for the large functional diversity of replays observed in behaving animals, both with regard to the timescale (fast vs. regular) they exhibit, as well as to the type of coding (dynamical vs. stable) they may subserve in navigational or working memory tasks.

Stability of Network AI Dynamics in the Presence of Trajectory Engrams

After evaluating the stability of the learned trajectory in the presence of AI network activity, we asked the symmetrical question, i.e., whether the engram of a previously learned trajectory could alter the irregular features of spontaneous network dynamics. Indeed, the altered synaptic structure (which implies large weights in all neurons of the recurrent network) may induce correlated activations of neurons (e.g., partial replays) resulting in runaway activity-plasticity interactions and drifts in network activity and synaptic structure. We monitored network connectivity (**Figure 6A**) and activity dynamics (**Figures 6B1–B3**) for 1 h to assess the stability of the spontaneous AI regime in the presence of the engram. We observed that following learning of the engram, synaptic weights outside the engram (i.e., responsible for the AI dynamics) increased exponentially toward their new steady-state in a very slow manner (**Figure 6A**) with an apparent time constant of 1.91 h, consistent with the theoretical estimation of 1.95 h (see above). This increase resulted from the decrease of within-engram large synaptic weights via synaptic scaling (**Figure 6E1**, see above). Despite this slow and moderate structural reorganization, AI dynamics were preserved with stable frequency (**Figure 6B1**), synchrony (**Figure 6B2**), and irregularity (**Figure 6B3**). Thus, overall, both the synaptic structure outside the engram as well as the spontaneous AI regime remained stable in the presence of the engram.

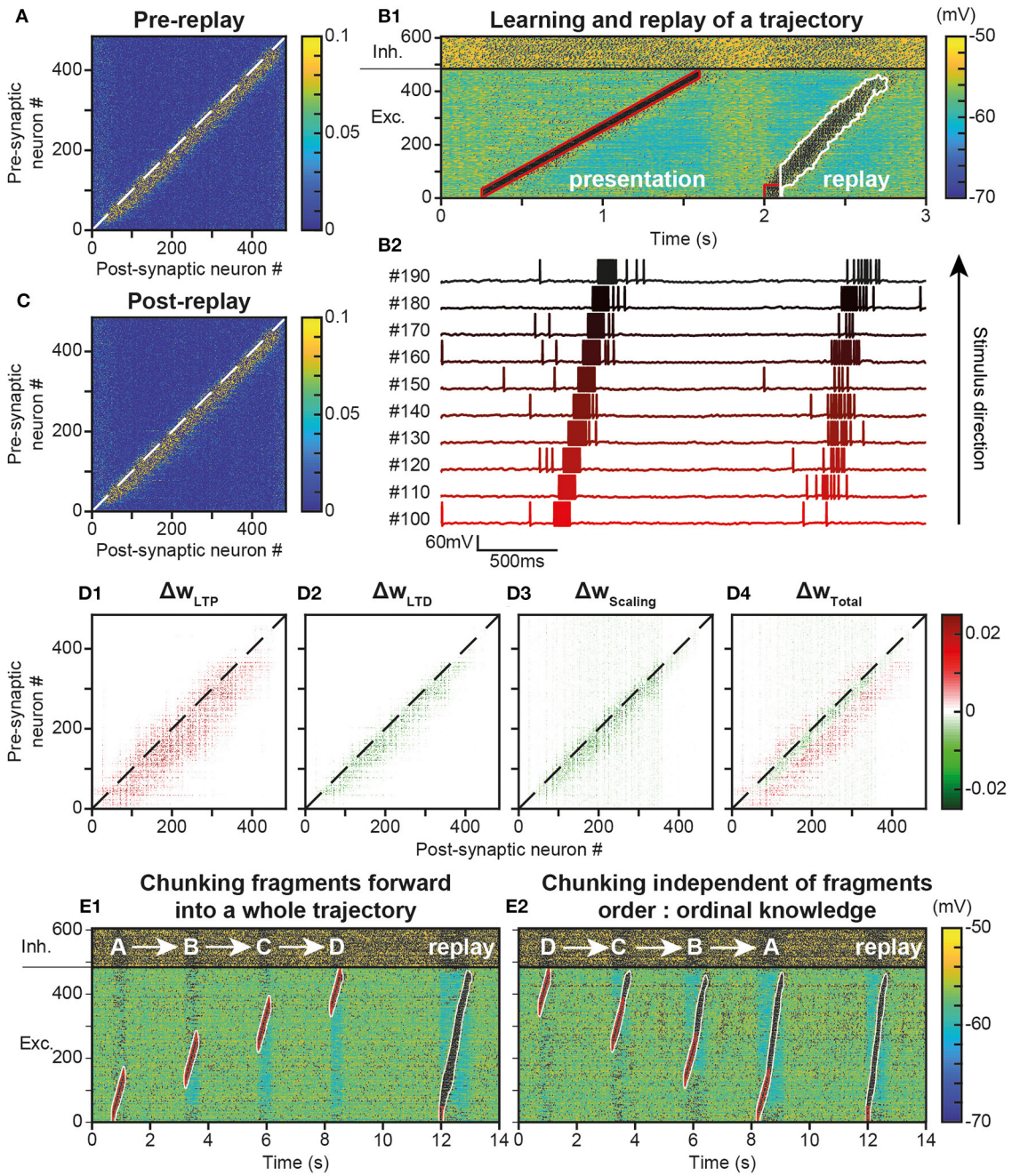


FIGURE 4 | Replay of learned trajectories. **(A)**. Synaptic matrix between excitatory neurons after stimulus presentation but prior to trajectory replay. **(B)**. Membrane potential of network neurons (**B1**, spikes indicated by black dots) in response to the trajectory stimulus, followed by a transient trajectory replay triggered by stimulating the start of the trajectory (neurons #1–50, stimulus in red). Membrane potential of a selected subset of neurons along the trajectory (**B2**, arbitrary colors). **(C)**. Synaptic matrix between excitatory neurons after stimulus and replay. **(D)**. Weight modifications resulting, after compared to before trajectory replay, from t-LTP **(D1)**, t-LTD **(D2)**, scaling **(D3)**, and their sum **(D4)**. **(E)** Recapitulation of the whole trajectory after separately learning four individual trajectory fragments (ABCD) in the forward order (**E1**; chunking) or backward order (**E2**; ordinal knowledge). Each fragment corresponds to 180 neurons. Fragments overlap over 65 neurons.

Memory of Trajectory Engrams in the Presence of Network AI Dynamics

We then studied whether the spontaneous AI activity could disrupt the engram of the learned trajectory and the possibility for trajectory replay. Indeed, the trajectory engram may be gradually erased, due to AI activity at low frequency favoring t-LTD, or even amplified, due to the activity in the trajectory engram caused by plasticity (resulting in further plasticity

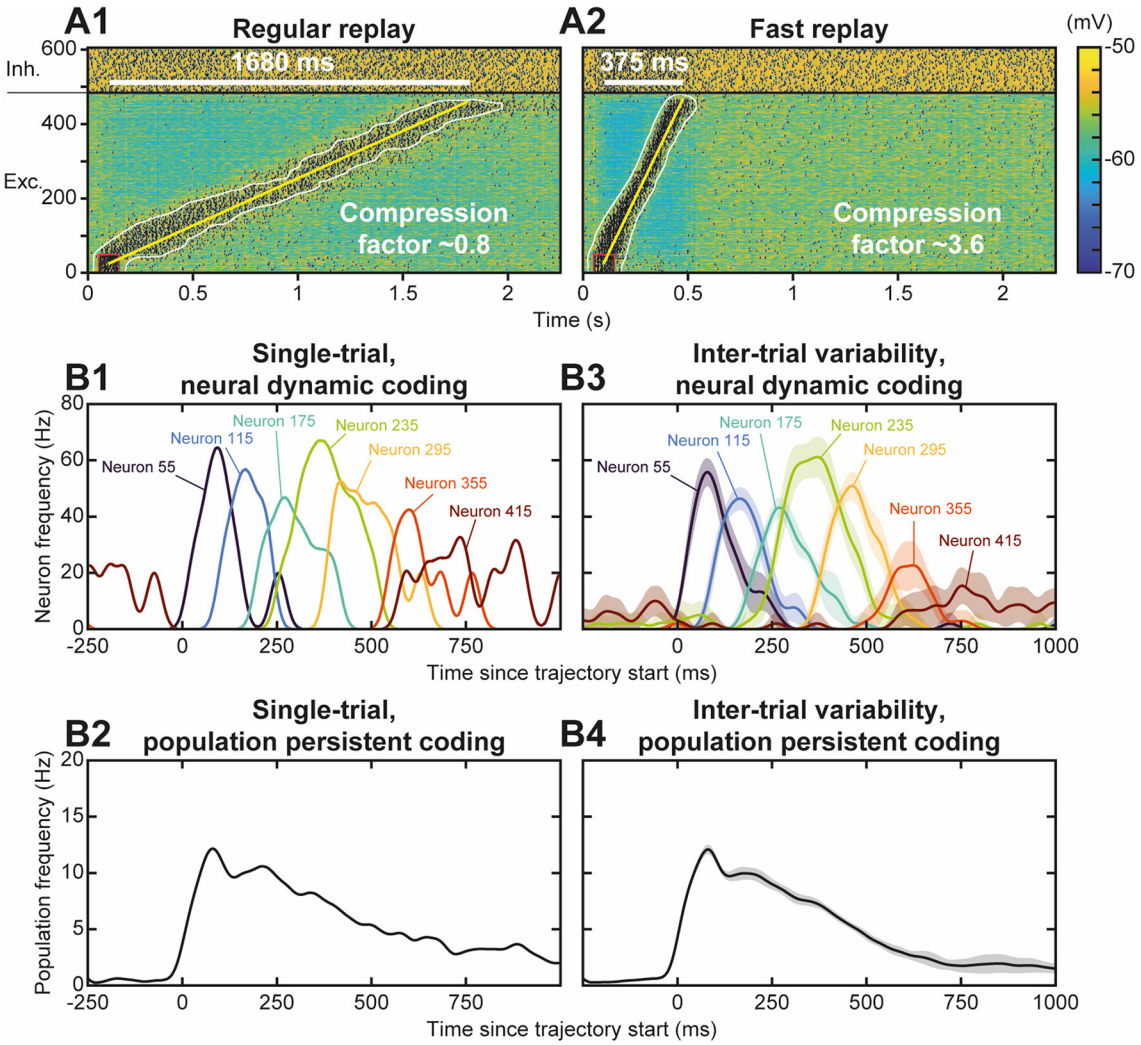


FIGURE 5 | Functional diversity of trajectory replays. **(A)** Trajectory replay duration (upper left white bar) and compression factor (tCF; lower right) depend on the NMDA conductance decay time constant (τ_{NMDA}, range 30–150 ms). NMDA maximal conductance was scaled (range 0.475–1.8) so as to insure similar levels of firing frequency drive during trajectory replays. Regular **(A1)** and fast **(A2)** timescale replay are due to slower and faster NMDA dynamics. **(B)**. Single-trial **(B1, B2)** and inter-trial variability **(B3,B4)** of firing frequency of individual neurons **(B1,B3)** and of the population **(B2,B4)** for 10 different simulations similar to the replay shown **Figure 4B**. Lines represent mean values, shaded regions represent 95% confidence intervals of the mean.

runaway). To do so, we assessed the timescale of potential drifts in engram connectivity and activity following learning, and of the network ability to replay the engram. Intuitively, engram erasure, runaway or stability probably depended on network dynamics after learning: spontaneous AI regime, spontaneous replays, or other forms of activity.

To address these questions, we simulated the network for 1 h after trajectory learning and recorded "snapshots" of the continuous evolution of the synaptic matrix every minute. Using these successive recorded matrices as initial conditions for independent simulations of replays, we were able to quantify network ability for trajectory replay, at different times of the evolution of the network. We found that while trajectory replay occurred in full after 1 s, activating all neurons of the trajectory (**Figure 6C1**), it was slightly attenuated after 1 min (last neurons spiking at lower frequency; **Figure 6C2**) and failed after 1 h (**Figure 6C3**). Observing the synaptic matrix at these three moments allowed us to understand the origin of this degradation in replay ability. Indeed, whereas after 1 min (**Figure 6D2**), the synaptic weights of the engram changed only a little compared to 1 s (**Figure 6D1**), the engram was narrowed and weights attenuated after 1 h (**Figure 6D3**). Such degradation of the engram was probably the cause of the failure to replay the trajectory 1 h after learning.

To more precisely monitor degradation of the trajectory engram and replay, we measured averaged engram weights as well as replay frequency and duration across time. We found that the engram weights declined exponentially with a fitted time constant of 1.91 h (**Figure 6E1**), very close to that predicted by the theory (1.95 h). The measures of trajectory replay decreased

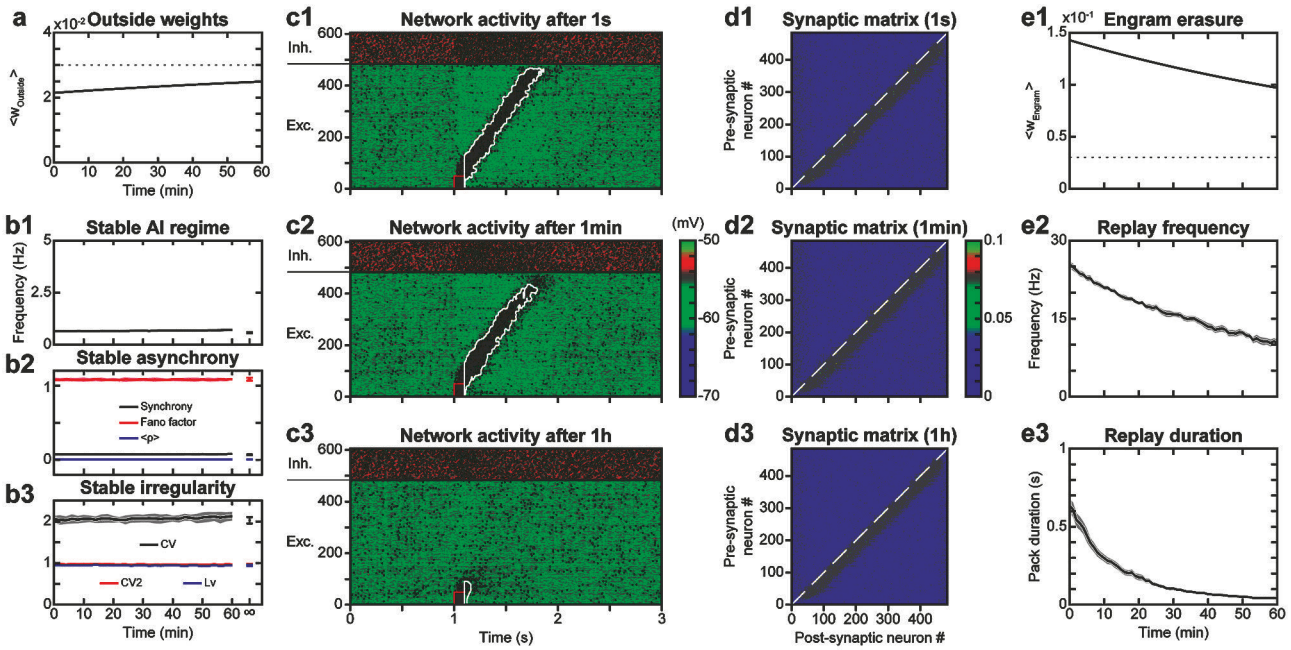


FIGURE 6 | Stability of the spontaneous AI regime in the presence of the engram. **(A)**. Average synaptic weights outside the engram after learning for 1 h. Shaded areas represent 95% confidence intervals of the mean for 5 network simulations. **(B)**. Networks dynamics after learning for 1 h: frequency **(B1)**, synchrony **(B2)**, and irregularity **(B3)** of excitatory neurons. Shaded areas as in **(A)**. **(C)**. Membrane potential of neurons in the neural network for 3 s following a replay stimulation of the 50 first neurons at 1 s **(C1)**, 1 min **(C2)** or 1 h **(C3)** after trajectory learning. **(D)** Synaptic matrices between excitatory neurons of the network, at the end of the simulations presented in **(C)**. **(E)** Network engram synaptic weight average **(E1)** as well as frequency **(E2)** and duration **(E3)** of trajectory replays during 1 h after trajectory learning. Shaded areas as in **(A)**.

faster than the engram weights, with time constants of ∼54 min for mean frequency during the replay (**Figure 6E2**) and ∼13 min for replay duration (**Figure 6E3**). Specifically, replay of the full trajectory lasted 4 min. The degradation of trajectory replay was mainly due to progressive replay failure in the neurons located later in the trajectory engram. The faster decrease in trajectory activity, compared to the average engram weights, was probably a consequence of a cooperative mechanism of propagation in the engram: the non-linearity in NMDA current activation, requiring synergistic activation of pre- and post-synaptic neurons in the engram, rendered the propagation of activity non-linearly sensitive to decreases in engram weights.

Repeated Trajectory Replays Can Destabilize Trajectory Engrams and Replays

We have observed that a single replay of the trajectory only marginally modified the engram (**Figure 4C** vs. **Figure 4A**). However, we assessed whether replay repetitions could strengthen the engram significantly further. Such strengthening through repetition could compensate for the engram erasure due to spontaneous activity after the learning (**Figure 6E1**) and its functional consequence, the relatively rapid loss of replay capacity (**Figures 6E2,E3**). Intuitively, the partial increase in weight at the border of the trajectory engram after one replay (**Figure 4D4** Δw_{Total}, red fringes) could, after repeated replays, be strong enough to counteract the decrease observed outside replays during memorization (**Figure 6D3**, light blue fringe). To test this possibility, we repeated the replay stimulus every 3 s for 30 s after the presentation of the initial trajectory stimulus (**Figure 7A**). We observed, from the very first seconds, and even before we could test the effect of the protocol at larger timescales, that these successive stimuli, initially triggering correct trajectory replays, rapidly led to hyperactivity involving most of the neurons in the network (**Figure 7A1**). Such paroxysmal activity typically appeared via avalanche dynamics activating neurons at the end of the trajectory (a fraction of the network, therefore), which propagated to the whole network at increasingly higher discharge frequencies (up to tens of Hz). Moreover, this activity had an oscillatory component, visible on the time course of the frequency of the excitatory and inhibitory neurons (**Figure 7A2**). This paroxysmal activity partially erased the engram of the learned trajectory via synaptic scaling (**Figure 7B**), making it impossible to replay the trajectory following this seizure (see last stimulus, **Figure 7A1**), consistent with similar effects found in empirical observation during epileptic seizures (Hu et al., 2005; Meador, 2007; Truccolo et al., 2011).

Slow Learning Stabilizes Trajectory Engram and Replays

As the repetition of replay learning led to over-activation of the trajectory with plasticity speed parameters sufficiently fast for a single stimulus presentation to be learned and replayed, we investigated how slower STDP kinetic coefficients could prevent

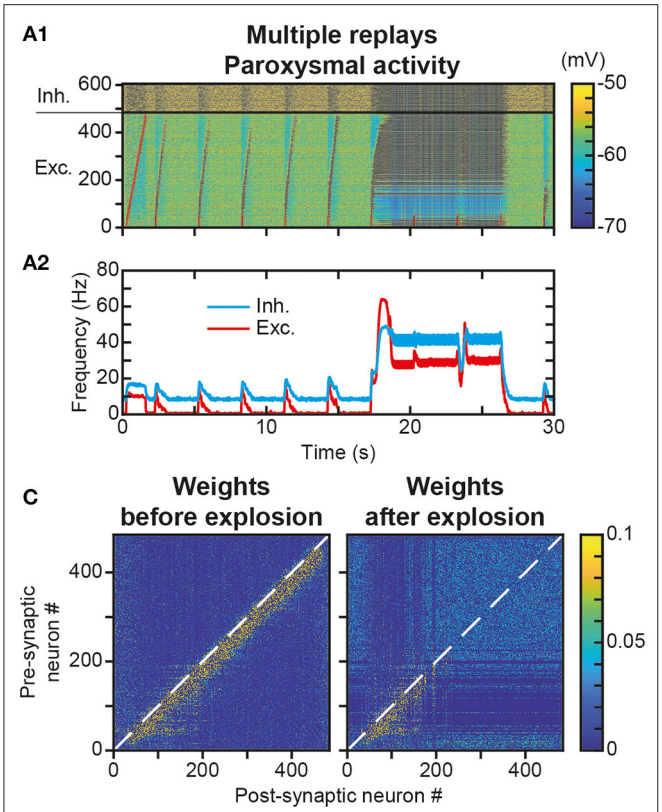


FIGURE 7 | Unstable engram and network dynamics after repeated trajectory replays. **(A)**. Membrane potential of neurons in the neural network **(A1)** for 30 s during which a replay stimulus is performed on the first 50 neurons every 3 s. Mean activity of excitatory (red) and inhibitory (blue) neurons **(A2)**. **(B)**. Matrix of synaptic weights between excitatory neurons before (left) and after (right) paroxysmal network activity.

paroxysmal activity during stimulus presentations and replays. For this, we used smaller values of K_{max} and P_{max}, i.e., here, divided by a factor of 6. With these values, 4 presentations of the trajectory stimulus were necessary for increasing the engram weights enough to sustain trajectory replays (**Figure 8A**). After such a learning protocol, the replay of the full trajectory was possible even beyond 1 h after learning (**Figure 8B**), whereas replay ability lasted only a few minutes with previous parameters (**Figures 6E2,E3**). This increase in replay memory timescale is consistent with that of the engram time constant, which was 11.5 h (**Figure 8C**), of the order of its theoretical estimation ∼11.7 h, i.e., it was increased by a factor 6 compared to that obtained with previous parameters (1.91 and 1.95 h, respectively **Figure 6E1**). Remarkably, the memory of trajectory replay was increased by a factor >20 (trajectory completely replayed at >1.4 h vs. 4 min with previous parameters), so that, relatively to the timescale of the trajectory engram, the timescale for trajectory replay was further increased by a factor 3.5. Indeed, the presentation of several stimuli recruited a thicker-tailed weight distribution, with higher probability of large weights (blue curve above the red one in ∼0.05–0.125; **Figure 8D**) but lowered probabilities of highly-weighted synapses (blue curve with negligible probabilities above 0.15; **Figure 8D**), because successive trajectory stimuli simultaneously evoked progressively stronger trajectory replays, recruiting more neurons at lower frequencies (**Figure 8A**), therefore imprinting larger engrams. Thus, slower plasticity kinetics required a larger number of successive presentations to learn the trajectory, but ensured a more robust engram involving more synapses, resulting in a better resilience to forgetting, i.e., a better quality of learning.

Finally, we assessed whether slow plasticity with multiple stimulus presentations also preserved network dynamics. AI dynamics were preserved with stable frequency (**Figure 8E1**), synchrony (**Figure 8E2**), and irregularity (**Figure 8E3**). We then repeated the replay stimulus every 3 s for 30 s after the presentation of the initial trajectory stimulus, a protocol which led to paroxysmal activity when considering fast plasticity. With slower kinetics, multiple replay stimuli triggered correct trajectory replays for the whole duration of the simulation (**Figure 8F**). We then asked whether a threshold of plasticity speed exists above which paroxysmal activity is triggered, or, conversely, the risk of paroxysmal activity linearly scales with the ability to learn fast. To do so, we parametrically explored simulations with plasticity rate divided by a slowdown factor in the range 1–10. The minimal number of stimulus presentations required to form a strong enough engram (i.e., allowing a replay) increased slowly with slower plasticity kinetics (**Figure 8G**, red). In parallel, the increase in the maximal number of replays before turning network dynamics into paroxysmal activity was much larger (**Figure 8G**, black), so that slowing plasticity kinetics increased the physiological range allowing learning while preserving network dynamics from paroxysmal activity. Hence, plasticity slow enough to preserve healthy dynamics may constitute a key constraint on the ability to learn rapidly. Furthermore, if the product of plasticity speed with the number of stimulus presentation was constant, it would indicate a linear summation of plastic effects arising from each presentation. By contrast, the number of stimulus presentations necessary for replay was lower than the factor of plasticity slowdown (5 stimuli for 10x plasticity slowdown instead of 10 stimuli, **Figure 8G**). This is due to successive stimulations overlapping with replays (i.e., stimulus presentations after the first one induce replays, **Figure 8A**), suggesting progressive facilitation of learning at slow plasticity speeds.

DISCUSSION

Here, we show that it is possible to learn neural trajectories (dynamical representations) using a spike timing-dependent plasticity (STDP) learning rule in local PFC circuits displaying spontaneous activity in the asynchronous irregular (AI) regime. We used a physiological model of plasticity (Delord et al., 2007; Graupner and Brunel, 2012; He et al., 2015) continuously occurring online, i.e., without decoupling simulations of learning and activity. Presentation of a dynamic stimulus, the trajectory, resulted in the writing of a synaptic engram of the trajectory on a rapid timescale (seconds), as well as its long-term storage at

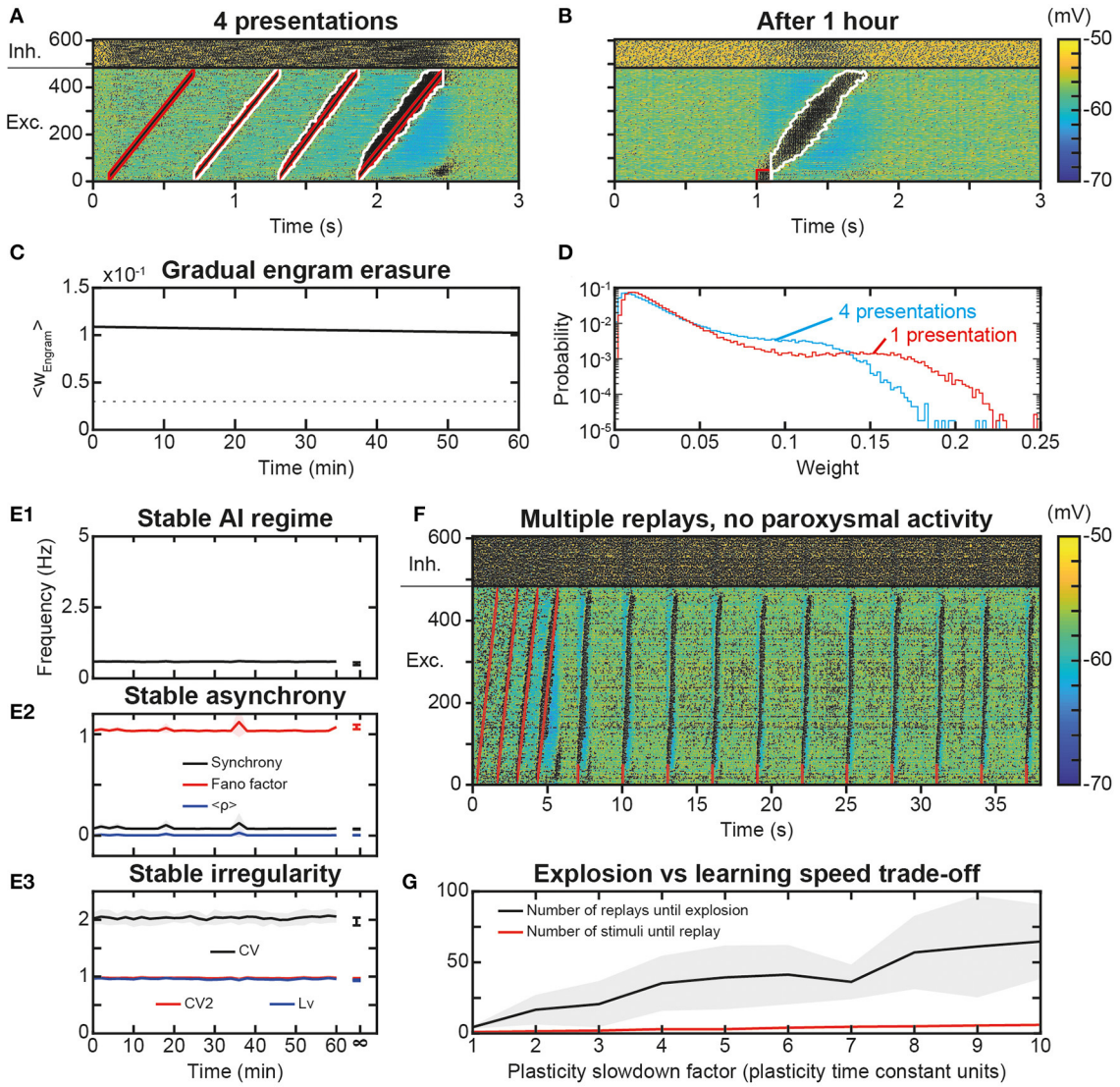


FIGURE 8 | Slower learning stabilization of the engram and network dynamics. **(A)**. Membrane potential of the neural network in response to the presentation of 4 trajectory stimuli in the presence of slower STDP learning kinetics. **(B)**. Membrane voltage of the neural network for 3 s following a replay stimulation on the first 50 neurons at 1 s, 1 h after trajectory learning. **(C)**. Average weight of all engram synapses after learning for 1 h. **(D)**. Probability distribution of the synaptic weights of the excitatory synapses after 4 presentations of the trajectory stimulus during slow learning (blue), and after one presentation of the trajectory stimulus during learning with faster (standard) parameters (red). **(E)** Networks dynamics after learning with slow plasticity: frequency **(E1)**, synchrony **(E2)**, and irregularity **(E3)** of excitatory neurons. Shaded areas represent 95% confidence intervals of the mean for five network simulations. **(F)** Membrane potential of the neural network for 38 s during which a replay stimulus is performed on 50 neurons every 3 s for 10 total repetitions (as in **Figure 7A**) after 4 trajectory stimuli in the presence of slower STDP learning kinetics (as in **Figure 8A**). **(G)** Minimal number of stimulus presentations required to learn a replay (red) and maximal number of replays before paroxysmal activity (black), as a function of the plasticity slowdown factor expressed in units of plasticity standard time constant (i.e., by which slowdown factor plasticity rates are divided). The number of replays until explosion is evaluated with the same weight matrix (learned at standard plasticity speed or x1 slowdown) across different plasticity speeds, for better comparison of the effect of plasticity speeds on replay. Shaded areas represent 95% confidence intervals of the mean for 10 network simulations.

the timescale of the order of several hours. The network replayed the trajectory upon stimulation of a subset of the engram at the timescale of the order of dozens of minutes. These results indicate that disordered AI activity does not necessarily jeopardize the encoding and replay of neural trajectories. Conversely, the weak but continuous plastic processes that noisy AI produces did not erase the synaptic engram of neural trajectories, at least before several hours. In turn, the learning of a trajectory engram within network synapses was not found to alter the AI characteristics of PFC activity. From a functional perspective, we show that trajectory activity accounted for both types of dynamics subserving working-memory in the PFC, i.e., persistent activity (Constantinidis et al., 2018) and dynamical coding (Lundqvist et al., 2018), and help understanding how they can be reconciled (Murray et al., 2017; Cavanagh et al., 2018; Enel et al., 2020). Together, these results offer a consistent theoretical

framework accounting for how dynamical representations can be learned, memorized and replayed in PFC circuits in physiological conditions.

This model was built to reproduce functional phenomenology of the PFC (learning, replays at different timescales, dynamic or persistent coding, see below), based on biophysical constraints from the experimental literature at the molecular, cellular and network levels, rather than by artificial training. If overall architectural properties of the model are observed in the PFC, such properties are also compatible with other non-prefrontal cortices with trajectory replays, lending strength to the genericity of the current study's results. For example, the excitatory/inhibitory network balance, observed in the PFC (Shu et al., 2003; Haider et al., 2006), is also observed and essential to computations across non-PFC structures (Isaacson and Scanziani, 2011). Similarly, the over-representation of bidirectional connections in the PFC (Wang et al., 2006) is a general property in cortical networks (Song et al., 2005). While the PFC has been less subject to the investigation of synaptic scaling compared to other structures, its presence across many non-PFC cortical structures (for e.g., sensory cortices, hippocampus, motor cortex) and crucial role for synaptic learning stabilization (Keck et al., 2017) makes it a plausible mechanism in PFC. Certain lines of evidence suggest its presence in PFC (Wang and Gao, 2012; Sweatt, 2016), although further confirmation is needed.

In the model, external feedforward inputs are constant, as in previous models of characteristic PFC activity (for e.g., Brunel, 2000). Therefore, the variability of neuronal discharge observed in the network entirely arises from internal dynamics among recurrent connections, as the network is in the asynchronous irregular regime (Destexhe et al., 2003; London et al., 2010). It would be interesting to study versions of the model with feedforward inputs variability, as occurring in real PFC circuits. However, this option was out of scope as we focused on the internal interactions between the spontaneous AI regime, learning processes affecting the synaptic matrix and trajectory replays. As another potential extension to our study, one could explore the influence of rhythmic inputs from the hippocampus (theta rhythms, Siapas et al., 2005; Benchenane et al., 2011) or from the olfactory pathways (delta rhythms, Moberly et al., 2018), which are known to be important for behaviorally-relevant neural activity and memory replays.

Molecular Plasticity and Memory in the PFC

In the PFC, e-STDP necessitates more than the pre-post synaptic pairings used in spike-timing protocols, as long-term potentiation (t-LTP) emerges in the presence of dopaminergic or cholinergic tonic neuromodulation, or when inhibitory synaptic transmission is decreased (Couey et al., 2007; Xu and Yao, 2010; Ruan et al., 2014). Moreover, Hebbian STDP (i.e., t-LTP for pre-then-post and t-LTD for post-then-pre spiking) is observed when followed by phasic noradrenergic, dopaminergic or serotoninergic neuromodulation (He et al., 2015). Hence, we assumed that t-LTP and t-LTD co-exist, and STDP is thus Hebbian, in the PFC of behaving animals, where both phasic and tonic neuromodulation are encountered during behaviorally relevant learning (Dembrow and Johnston, 2014). The present study did not incorporate noradrenergic, serotoninergic and dopaminergic transformation of eligibility traces into effective plastic modifications found at PFC excitatory synapses (He et al., 2015), a possible substrate of context- and reward-modulated learning in PFC circuits (Ellwood et al., 2017). The present work also did not consider alternative biophysical processes that may participate to sculpt dynamical and flexible neural representations in the PFC (Buonomano and Maass, 2009; Stokes, 2015). For instance, short-term synaptic plasticity (Mongillo et al., 2008) may affect network dynamics through slow hidden (e.g., biochemical) variables. Such a silent-based coding of past activity could possibly account for the near-complete disappearance of activity observed sometimes during working memory (Stokes, 2015) and its interaction with activity-based working-memory in the PFC (Barbosa et al., 2020) remains to be elucidated. Similarly, inward current-mediated bistability such as with persistent sodium, or calcium-activated non-specific currents (Delord et al., 1997; Rodriguez et al., 2018), can produce cellular forms of memory that may take part in dynamic representations in the PFC, either through retrospective memory of past information or in prospective computations of forthcoming decisions and actions. Finally, the present study did not consider anti-homeostatic forms of intrinsic plasticity (i.e., the plasticity of intrinsic properties) which may represent an essential mean to learn and regulate dynamic representations (Zhang and Linden, 2003).

Stable Spontaneous AI Dynamics in the PFC in the Presence of Plasticity and Learning

Hebbian forms of plasticity (Abbott and Nelson, 2000), such as the STDP of excitatory synapses (Markram et al., 2012) modeled here, increase weights between neurons that are frequently co-activated. Stronger synapses potentiated by STDP, in turn, statistically increase the frequency of future co-activations. These rules thus constitute positive feedback loops (anti-homeostatic) between activity and connectivity. As a consequence, synaptic runaway (Keck et al., 2017; Zenke et al., 2017) produces network instability toward saturated or quiescent activity and connectivity. In recurrent network models, synaptic plasticity typically decreases the dynamics complexity toward regular activity such as limit-cycle or quasi-periodic attractors (Morrison et al., 2007; Siri et al., 2007; Litwin-Kumar and Doiron, 2014) that resembles neural dynamics encountered during sleep or paroxysmal crises. However, activity in the PFC and other cortices during wakefulness is characterized by asynchronous irregular spiking at low frequency (Ecker et al., 2010; Renart et al., 2010), due to the balance between strong excitatory and inhibitory synaptic currents (Destexhe et al., 2003). AI spiking is compatible with critical or even chaotic dynamics (Beggs and Plenz, 2003; Hahn et al., 2010; London et al., 2010), which may benefit temporally complex computations (Bertschinger and Natschläger, 2004) believed to be performed by the PFC (Compte, 2003).

Many studies show that e-STDP rules are deleterious to AI dynamics such that compensating homeostatic mechanisms are

required to control neuronal activity, for e.g., a metaplastic e-STDP rule with sliding-threshold (Boustani et al., 2012), synaptic scaling (which keeps the sum of pre-synaptic excitatory weights constant, Zenke et al., 2013), STDP of inhibitory synapses (i-STDP; ensuring excitation-inhibition balance, Vogels et al., 2011) or intrinsic plasticity of ionic conductances (regulating action potential threshold, Naudé et al., 2013). In the present detailed biophysical model, we found that a combination of e-STDP where all pre-/post- pairings were taken into account (all-to-all STDP), together with synaptic scaling, preserves AI dynamics. All-to-all e-STDP without scaling can also preserve AI dynamics, but at the price of unstable fluctuating synaptic weights (Morrison et al., 2007), while weight distributions were stable here. Moreover, the present study shows that network stability held not only with random recurrent connections, but also in the presence of an engram involving a significant fraction of strong, potentiated synapses in all excitatory neurons. In the absence of synaptic scaling, learning static patterns into synaptic engrams with e-STDP disrupts AI dynamics toward pathological high-frequency oscillations (Morrison et al., 2007; Litwin-Kumar and Doiron, 2014), or with i-STDP leads to AI activity with unrealistic high firing frequency states and sharp state transitions (Litwin-Kumar and Doiron, 2014), at odds with PFC dynamics in awake animals (Compte, 2003). A metaplastic form of e-STDP conserves AI dynamics on a short-timescale (one second) but AI stability remains unchecked at longer timescales (Boustani et al., 2012). This is only the case with static stimulus, as learning receptive fields using dynamical stimulus leads to a catastrophic decrease in the complexity of the AI regime (Boustani et al., 2012). Altogether, our study thus suggests that synaptic scaling represents a more efficient form of homeostatic compensation (rather than metaplastic e-STDP, or i-STDP) for learning trajectory engrams without the deleterious effects of STDP disrupting AI dynamics. We used here an instantaneous synaptic scaling, because our model, like most models, requires synaptic scaling at faster or equal timescales than synaptic plasticity for stable learning, far from the experimentally observed homeostatic or metaplastic timescales of hours to weeks (Zenke et al., 2017). This constraint suggests the existence of as yet unidentified rapid compensatory processes, potential candidates being heterosynaptic plasticity (Fiete et al., 2010), intrinsic plasticity (Zhang and Linden, 2003; Naudé et al., 2013), input normalization by feed-forward inhibition (Pouille et al., 2009; Keck et al., 2012), and the implication of astrocytes (Papouin et al., 2017). Additionally, at slower timescales, sleep-dependent consolidation mechanisms may provide global compensatory synaptic down-scaling offline (Tononi and Cirelli, 2003).

Learning Dynamical Representations in the PFC Under AI Dynamics

Phenomenological e-STDP models fail to learn engrams in noisy AI states because of their sensitivity to spontaneous activity. The absence of STDP weight-dependence forbids learning and induces the direct loss of engrams (Boustani et al., 2012), while without synaptic scaling, learning fails with catastrophic consequences in terms of network dynamics (see above; Morrison et al., 2007). A weight-dependent e-STDP rule endowed with homeostatic metaplasticity (instead of synaptic scaling, as here) allowed learning the engram of a presented stimulus while preserving AI dynamics, although it unrealistically left neurons of the engram in a state of permanent activity (Boustani et al., 2012). Likewise, i-STDP enables learning of engrams, but with unrealistic AI activity (see above; Litwin-Kumar and Doiron, 2014). Here, we find that the combination of a weight-dependent Hebbian e-STDP rule and synaptic scaling allows for the learning of engrams in local PFC recurrent networks under conditions of AI dynamics, as found in behaving mammals.

Phenomenological STDP models based on neighboring spike-doublet or spike-triplet schemes often produce side effects (either sensitivity to noisy activity, or runaway plasticity) due to the temporal bounds of the pre- and post-couplings they consider (Boustani et al., 2012). The present STDP model describes continuous post-synaptic biophysical dynamics that account for all pre-/post-pairings (all-to-all STDP) and is thus more realistic than phenomenological STDP models. Here, the temporal asymmetry of the spike-timing dependence of the e-STDP rule arises from a detailed description of calcium dynamics. Calcium arises from two different sources of calcium that originate from the influence of AMPA, NMDA and VDCC channel activations (see *Materials and Methods*; Graupner and Brunel, 2012), which accounts for the relative influence of pre-synaptic evoked excitatory post-synaptic potentials and of backpropagating post-synaptic activity. However, this rule remains simple compared to models describing more complete signaling scenarios (Manninen et al., 2010), allowing simulation at the network scale.

In feed-forward networks endowed with this STDP rule, and for conditions of spiking frequency and irregularity similar to AI activity, plastic modifications essentially depend on firing frequency rather than on the precise timing of spikes, because equivalent probabilities of encountering pre-then-post and post-then-pre spike pairs in conditions of stationary spiking essentially blurs net spike-timing effects (Graupner et al., 2016). Moreover, t-LTP dominates t-LTD, because t-LTD is multiplicative (Bi and Poo, 1998; van Rossum et al., 2000), i.e., scaled by weak weight values (Graupner et al., 2016). Consistent with these observations, in the present PFC recurrent network model, plasticity was essentially frequency-dependent under conditions of stationary spiking, and t-LTP dominated t-LTD under spontaneous AI dynamics, being principally compensated by synaptic scaling. However, during trajectory presentation or trajectory replay, i.e., when pre-post spiking was enforced to be temporally asymmetric, t-LTD nevertheless contributed to compensate t-LTP and determined overall resulting modifications on the same order than scaling.

The previous studies that have addressed the possibility of engram learning in recurrent networks with AI dynamics focused on static stimuli (Morrison et al., 2007; Boustani et al., 2012; Litwin-Kumar and Doiron, 2014). By contrast, our study demonstrates engram learning and activity replay of dynamical stimuli, such as the sequences or trajectories of activity that

occur during cortical AI dynamics in behaving animals (Kaefer et al., 2020). Standard static Hebbian assemblies, which learn static stimuli through strong bidirectional connections between neurons of the assembly and replay the static activity through pattern completion, induce avalanche-like convergent dynamics toward a static attractor, which are too low-dimensional to account for physiological data. Remarkably, the present study demonstrates the possibility for engrams of dynamic stimuli in the disordered AI state, despite the fact that they relied on mono-directional strengthening of synaptic connections, which favors propagation of activity, but does not allow for the convergent effect of static patterns and the positive feedback inherent to it.

Long-Term Memory of Dynamical Representations in the PFC Under AI Dynamics

The present study underlines the importance of slow plasticity kinetics together with repeated presentations for learning dynamic representations in PFC networks. Faster kinetics allowed one-shot learning of trajectory engrams, but extensive training could then induce paroxysmal activity during the trajectory replays that partly erased the engram, which was ultimately detrimental to the learning and replay process. This synchronous increase in neuronal activity in the model is reminiscent of epileptic seizures (Truccolo et al., 2011), which have been found to cancel out the plasticity effects of synaptic weights (Hu et al., 2005), and affect memory (Meador, 2007), as we found here. By contrast, slower kinetics resulted in more stable engrams, while highlighting the importance of repeated presentations of the dynamic stimulus, similarly to observations with static patterns (Boustani et al., 2012). Parametric exploration of plasticity kinetics showed a tradeoff between the number of stimulus repetitions required to form an engram and the risk of paroxysmal activity. However, slowing down plasticity decreased the risk of over-activation while preserving the ability to learn fast (even though not through one-shot learning). Consistent with our results, learning occurs gradually in the PFC, and at a slower pace than in the hippocampus and basal ganglia (Pasupathy and Miller, 2005; Buschman and Miller, 2014). The tradeoff between fast learning and paroxysmal risk may constitute a constraint for the PFC, with the preservation of asynchronous irregular dynamics preventing one-shot learning based on synaptic plasticity alone. One-shot learning, which occurs in well-trained animals, may thus require additional mechanisms for structural learning (Gallistel and Matzel, 2013).

Fast learning together with stable memory is considered in many synaptic plasticity models to rely on auto-phosphorylation of the calmodulin-dependent protein kinase II (CaMKII). CaMKII auto-phosphorylation is appealing because it constitutes a positive-feedback loop (inducing fast plasticity) underlying bistable dynamics (providing infinite memory of a single potentiated synaptic state). However, we did not consider CaMKII in the present model, because CamKII is not necessary to the maintenance of synaptic modifications (Chen et al., 2001; Lengyel et al., 2004). Moreover, activity-dependent synaptic modifications are not systematically bistable (i.e., they can be graded; Montgomery and Madison, 2002; Tanaka et al., 2008; Enoki et al., 2009) and they can fade with time scales from seconds to minutes (Hempel et al., 2000).

Here, the stability of molecular memory originated from extremely slow synaptic weight dynamics, resulting in slow exponential forgetting of the engram. Slow weight dynamics arose from activity-dependent kinase and phosphatase (aKP, Delord et al., 2007), which are weakly activated at near-basal calcium concentrations associated with low spiking frequency during AI dynamics. Such aKP signaling processes are ubiquitous (e.g., PKA, PKC, calcineurin) and confer an activity-dependent control over the rate of plasticity and memory (Delord et al., 2007), which is essential for flexible learning in the PFC (Fusi et al., 2005). Alternatively, when implemented with low copy molecule numbers at individual synapses, bistable models faced with noise also exhibit exponential forgetting of memory when averaged over synapses and trials (Fusi et al., 2005). Here, the memory of the trajectory engram admitted an effective time constant of the order of 2 h in network simulations, consistent with its theoretical prediction (see *Materials and Methods*), but longer memories could be expected for lower values of $P_{\max}$ and $K_{\max}$, the maximum phosphatase and kinase activations. However, the time constant for plasticity would also increase, slowing learning too, while its current value is compatible with induction times of synaptic plasticity (Malenka et al., 1992). Alternatively, a higher calcium phosphatase half-activation (P_{Ca}), which is physiologically possible (Delord et al., 2007), would allow for a longer memory timescale while preserving rapid learning (at large calcium, the time constant of plasticity is independent of P_{Ca}). Hence, specifying biophysical models with precise kinetic parameters is essential because they have huge consequences on the stability of network dynamics, learning and the time scale of memory (Zenke et al., 2013). Specifically, homeostatic scaling appeared important here as for learning, since its absence was reported to forbid the memory of static patterns in recurrent network models because of catastrophic forgetting due to fluctuating synaptic weights (Morrison et al., 2007).

The timescale of trajectory replay scaled with that of the engram. This is because replay requires a sufficiently preserved engram to emerge from synaptic interactions between neurons. However, the lifetime of trajectory replay was an order of magnitude smaller than that of the trajectory engram, because replay requires neuronal interactions that are non-linear and therefore sensitive to decreases in synaptic weights. Interestingly, the long-term degradation of trajectory replay was due to incomplete replay at the end of the trajectories learned, in a manner consistent with the primacy effect of medium-term learned sequences (Greene et al., 2000). Besides, the memory of trajectory replay did not only rely on biophysical parameters but also on the learning protocol. Indeed, slower learning with repetitions increased the quality of engram by better anchoring the learned trajectory, through a larger number of synapses. Slow plasticity of a large number of synapses from a recurrent network, through repetition, may thus underlie the robustness of PFC-dependent memories (Buschman and

Miller, 2014). In addition to extensive training, the maintenance of trajectory engrams over longer timescales may be reached by regular replays, as observed in PFC-dependent active executive processes such as trajectory reactivations (Stokes, 2015), spontaneous replays (Kaefer et al., 2020), rehearsal and refreshing (Raye et al., 2007), or consolidations (Dudai, 2012). At the molecular scale, the possibility of synaptic tagging could be incorporated in the model (Clopath et al., 2008) in order to stabilize the engram and account for longer memory timescales.

Humans or animals generally learn complex navigational paths such as sensory, motor or behavioral sequences in a progressive manner. Thus, PFC circuits are often challenged with the necessity to process several parts of whole neural trajectories that are discovered as sequences of elementary parts encountered at separate points in time. Moreover, prospective processes in the PFC require recombining elementary neural trajectories into new trajectory representations serving the planning of future actions, choices or navigational paths, for e.g., during rule switching and behavioral adaptation (Ito et al., 2015; Mashhoori et al., 2018; Kaefer et al., 2020). Besides, sequences of non-spatial items have been shown to be processed in a spatial frame in primates (Jensen et al., 2013), likely involving neural trajectories. We found that STDP-based trajectory learning and replay in the network was able to learn trajectory fragments, transitions between fragments, and to chunk them into a whole trajectory, as found in the PFC (Ostlund et al., 2009; Dehaene et al., 2015). Moreover, the network displayed the ability to reconstitute a whole trajectory (i.e., a macroscopic sequence) based on trajectory fragments (i.e., overlapping microscopic sequences), independently of their order of presentation, i.e., to acquire ordinal knowledge about sequences of trajectory fragments (Jensen et al., 2013; Dehaene et al., 2015). However, STDP-based trajectory learning in our PFC network model was unable to learn higher-order representations of algebraic patterns or more complex nested structures (Dehaene et al., 2015), or to categorize sequences into specific classes (Shima et al., 2007). Assessing such possibilities using more elaborated, reward-dependent, forms of STDP learning rules might deserve future explorations.

Multiple Functional Relevance of STDP-Based Neural Trajectories in the PFC

We found in our model that the same network, taught with the same stimulus, could generate a large range of replay duration and compression factors, including those characterizing regular (Batuev, 1994; Fujisawa et al., 2008; Cromer et al., 2010; Mante et al., 2013; Yang et al., 2014; Ito et al., 2015; Markowitz et al., 2015; Schmitt et al., 2017; Tiganj et al., 2017; Nakajima et al., 2019; Passecker et al., 2019; Enel et al., 2020) and fast (Jadhav et al., 2016; Tiganj et al., 2017; Mashhoori et al., 2018; Yu et al., 2018; Shin et al., 2019; Kaefer et al., 2020) timescale replays in behaving animal. We found that the time constant of NMDA decay dynamics was essential in controlling the duration and compression factor of trajectory replays. In PFC circuits, dopamine slows decaying dynamics of NMDA-mediated EPSPs through D1-receptors (Chen et al., 2004; Onn et al., 2006) in an almost instantaneous manner (Onn and Wang, 2005). In addition to dopaminergic regulation, other forms of neuromodulation affect NMDA dynamics (Lutzu and Castillo, 2021). Our results suggest that rapid and bidirectional regulation of biophysical parameters in PFC networks by ongoing neuromodulation—as attentional demands and reward outcomes vary at the trial timescale—may control replay duration, compression factors, and the relative rate of regular vs. fast timescale replays.

Besides, individual neuronal activity displayed lower firing frequency during replay compared to the activity induced by the stimulus, consistent with sparse coding of representations after learning. Firing rates of individual neurons during stimuli or delays in working memory tasks, as well as in navigation tasks, vary considerably across species and behavioral contexts, spanning two orders of magnitude from ∼1 to ∼100 Hz (Fuster and Alexander, 1971; Batuev, 1994; Romo et al., 1999; Baeg et al., 2003; Yang et al., 2014; Markowitz et al., 2015; Tiganj et al., 2017). Frequencies of dozens Hz are common in individual PFC neurons (Funahashi et al., 1989; Romo et al., 1999; Brody et al., 2003; Fujii and Graybiel, 2003; Shinomoto et al., 2003; Jun et al., 2010; Tiganj et al., 2017; Enel et al., 2020). In the present model, frequencies of individual neurons were actually ∼100 Hz during stimuli and presentations, and 20–60 Hz during replays (**Figures 5B1,B3**). Thus, although larger than those observed during stimuli, individual frequencies were globally of the order of magnitude of those empirically observed. Mean frequencies in our network ranged below 10 Hz (**Figures 5B1,B3**), (**7A2**), in accord with experimental literature (Funahashi et al., 1989; Romo et al., 1999; Brody et al., 2003; Fujii and Graybiel, 2003; Shinomoto et al., 2003; Jun et al., 2010; Tiganj et al., 2017; Enel et al., 2020).

In the PFC, representations for executive functions and cognition can present less explicit dynamic coding schemes than regular timescale neural trajectories presented here. For instance, working memory can display intricate patterns of complex (heterogeneous but non-random) dynamic activities that can hardly be disentangled into simpler well-separate transient patterns of activity (Jun et al., 2010). However, during working memory tasks, PFC persistent delay activity is selective and maintains online content-specific representations. Working memory does often, but not systematically, require underlying persistent activities, often in a stable activity state (Goldman-Rakic, 1995; Compte et al., 2000; Durstewitz et al., 2000; Wang, 2001; Constantinidis et al., 2018). It can also rely on dynamical sequences of activities disappearing and reappearing, depending on instantaneous computational task-relevant requirements (Sreenivasan et al., 2014; Stokes, 2015; Lundqvist et al., 2018). The coexistence of stable population coding together with heterogeneous neural dynamics has been observed in the PFC during working memory tasks (Murray et al., 2017).

Here, trajectory replays offer a possible unified framework that can participate to reconcile opposite views regarding the nature of information persistent vs. dynamic coding in the PFC (Constantinidis et al., 2018; Lundqvist et al., 2018). Indeed, we find that while individual neurons displayed transient (hundreds

of milliseconds) overlapping bumps of activity, implementing a "relay race" form of explicit dynamic coding (Batuev, 1994; Brody et al., 2003; Cromer et al., 2010; Yang et al., 2014; Schmitt et al., 2017), their population activity persisted at the second timescale, ensuring the maintenance of the representation across time (Murray et al., 2017; Cavanagh et al., 2018; Enel et al., 2020). Depending on the functional context, neural trajectories learned here could be interpreted as the actual explicit representation of a trajectory unfolding online, granted that the decoding downstream neural structure can resolve individual activities of the network. Alternatively, if the downstream decoding neural structure only globally decodes the population average of network dynamics, activity would then be interpreted as an integrated and stable persistent representation of the trajectory as a whole (i.e., as a symbolic entity). This dichotomy is congruent with that found in the PFC, whereby individual neurons encode information at short timescales while the population as a whole persistently maintains information at longer time scales (Meyers et al., 2008). In this scheme, working memory representations would rely on individual neurons collectively stabilizing a dynamic population-level process (Murray et al., 2017; Cavanagh et al., 2018; Enel et al., 2020).

Interestingly, we found that the population activity of trajectory replays accounted for the decreasing pattern of activity that can be observed in the PFC (Cavanagh et al., 2018; Enel et al., 2020). Trajectory replays also displayed strong variability, as observed in the PFC during delay activities (Compte, 2003; Shafi et al., 2007). While within-trial variability across neurons essentially came from the fact that neurons spiked at distinct periods along the trajectory, inter-trial variability for each neuron originated from the noisy AI dynamics. Inter-trial variability accumulated over time for neurons situated later in the trajectory, henceforth the temporal tuning of neurons widened with their position in the sequence (Tiganj et al., 2017).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

MS, JV, JN, and BD developed the model. MS, JV, DM, and BD contributed to numerical simulations and their analysis. MS, JV, DM, JN, and BD wrote the article. All authors contributed to the article and approved the submitted version.

REFERENCES

Abbott, L. F., and Nelson, S. B. (2000). Synaptic plasticity: taming the beast. *Nat. Neurosci.* 3, 1178–1183. doi: 10.1038/81453

Abeles, M., Bergman, H., Gat, I., Meilijson, I., Seidemann, E., Tishby, N., et al. (1995). Cortical activity flips among quasi-stationary states. *Proc. Natl. Acad. Sci. U.S.A.* 92, 8616–8620. doi: 10.1073/pnas.92.19.8616

Baeg, E. H., Kim, Y. B., Huh, K., Mook-Jung, I., Kim, H. T., and Jung, M. W. (2003). Dynamics of population code for working memory in PFC. *Neuron* 40, 177–188. doi: 10.1016/S0896-6273(03)00597-X

Bakhurin, K. I., Goudar, V., Shobe, J. L., Claar, L. D., Buonomano, D. V., and Masmanidis, S. C. (2017). Differential encoding of time by prefrontal and striatal network dynamics. *J. Neurosci.* 37, 854–870. doi: 10.1523/JNEUROSCI.1789-16.2016

Barbosa, J., Stein, H., Martinez, R. L., Galan-Gadea, A., Li, S., Dalmau, 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, 1016–1024. doi: 10.1038/s41593-020-0644-4

Batuev, A. S. (1994). Two neuronal systems involved in short-term spatial memory in monkeys. *Acta Neurobiol. Exp.* 54, 335–344.

Beaulieu, C., Kisvarday, Z., Somogyi, P., Cynader, M., and Cowey, A. (1992). Quantitative distribution of gaba-immunopositive and -immunonegative neurons and synapses in the monkey striate cortex (area 17). *Cerebral Cortex* 2, 295–309. doi: 10.1093/cercor/2.4.295

Beggs, J. M., and Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *J. Neurosci.* 23, 11167–11177. doi: 10.1523/JNEUROSCI.23-35-11167.2003

Benchenane, K., Tiesinga, P. H., and Battaglia, F. P. (2011). Oscillations in the prefrontal cortex: a gateway to memory and attention. *Curr. Opin. Neurobiol.* 21, 475–485. doi: 10.1016/j.conb.2011.01.004

Bertschinger, N., and Natschläger, T. (2004). Real-time computation at the edge of chaos in recurrent neural networks. *Neural Comput.* 16, 1413–1436. doi: 10.1162/089976604323057443

Bi, G. Q., and Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* 18, 10464–10472. doi: 10.1523/JNEUROSCI.18-24-10464.1998

Bittner, K. C., Milstein, A. D., Grienberger, C., Romani, S., and Magee, J. C. (2017). Behavioral time scale synaptic plasticity underlies CA1 place fields. *Science* 357, 1033–1036. doi: 10.1126/science.aan3846

Boustani, S. E., Yger, P., Frégnac, Y., and Destexhe, A. (2012). Stable learning in stochastic network states. *J. Neurosci.* 32, 194–214. doi: 10.1523/JNEUROSCI.2496-11.2012

Brody, C. D., Hernández, A., Zainos, A., and Romo, R. (2003). Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cereb Cortex* 13, 1196–1207. doi: 10.1093/cercor/bhg100

Brunel, N. (2000). Dynamics of Sparsely Connected Networks of Excitatory and Inhibitory Spiking Neurons. *J Comput Neurosci* 8, 183–208. doi: 10.1023/A:1008925309027

Brunel, N., and Wang, X. J. (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J. Comput. Neurosci.* 11, 63–85. doi: 10.1023/A:1011204814320

Buonomano, D. V., and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* 10, 113–125. doi: 10.1038/nrn2558

Buschman, T. J., and Miller, E. K. (2014). Goal-direction and top-down control. *Philos. Transac. R. Soc. B Biol. Sci.* 369:20130471. doi: 10.1098/rstb.2013.0471

Cavanagh, S. E., Towers, J. P., Wallis, J. D., Hunt, L. T., and Kennerley, S. W. (2018). Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex. *Nat. Commun.* 9:3498. doi: 10.1038/s41467-018-05873-3

Chen, G., Greengard, P., and Yan, Z. (2004). Potentiation of NMDA receptor currents by dopamine D1 receptors in prefrontal cortex. *PNAS* 101, 2596–2600. doi: 10.1073/pnas.0308618100

Chen, H.-X., Otmakhov, N., Strack, S., Colbran, R. J., and Lisman, J. E. (2001). Is persistent activity of calcium/calmodulin-dependent kinase

required for the maintenance of LTP? *J. Neurophysiol.* 85, 1368–1376. doi: 10.1152/jn.2001.85.4.1368

Chenkov, N., Sprekeler, H., and Kempter, R. (2017). Memory replay in balanced recurrent networks. *PLoS Comput. Biol.* 13:e1005359. doi: 10.1371/journal.pcbi.1005359

Clopath, C., Büsing, L., Vasilaki, E., and Gerstner, W. (2010). Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nat. Neurosci.* 13, 344–352. doi: 10.1038/nn.2479

Clopath, C., Ziegler, L., Vasilaki, E., Büsing, L., and Gerstner, W. (2008). Tag-trigger-consolidation: a model of early and late long-term-potentiation and depression. *PLoS Comput. Biol.* 4:e1000248. doi: 10.1371/journal.pcbi.1000248

Compte, A. (2003). Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *J. Neurophysiol.* 90, 3441–3454. doi: 10.1152/jn.00949.2002

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

Constantinidis, C., Funahashi, S., Lee, D., Murray, J. D., Qi, X.-L., Wang, M., et al. (2018). Persistent spiking activity underlies working memory. *J. Neurosci.* 38, 7020–7028. doi: 10.1523/JNEUROSCI.2486-17.2018

Couey, J. J., Meredith, R. M., Spijker, S., Poorthuis, R. B., Smit, A. B., Brussaard, A. B., et al. (2007). Distributed network actions by nicotine increase the threshold for spike-timing-dependent plasticity in prefrontal cortex. *Neuron* 54, 73–87. doi: 10.1016/j.neuron.2007.03.006

Cromer, J. A., Roy, J. E., and Miller, E. K. (2010). Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron* 66, 796–807. doi: 10.1016/j.neuron.2010.05.005

Dale, H. (1935). Pharmacology and nerve-endings (Walter Ernest Dixon memorial lecture): (Section of Therapeutics and Pharmacology). *Proc. R. Soc. Med.* 28, 319–332. doi: 10.1177/003591573502800330

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* 88, 2–19. doi: 10.1016/j.neuron.2015.09.019

Delord, B., Berry, H., Guigon, E., and Genet, S. (2007). A new principle for information storage in an enzymatic pathway model. *PLoS Comput. Biol.* 3:e124. doi: 10.1371/journal.pcbi.0030124

Delord, B., Klaassen, A. J., Burnod, Y., Costalat, R., and Guigon, E. (1997). Bistable behaviour in a neocortical neurone model. *Neuroreport* 8, 1019–1023. doi: 10.1097/00001756-199703030-00040

Dembrow, N., and Johnston, D. (2014). Subcircuit-specific neuromodulation in the prefrontal cortex. *Front. Neural Circuits* 8:54. doi: 10.3389/fncir.2014.00054

Destexhe, A., Rudolph, M., and Paré, D. (2003). The high-conductance state of neocortical neurons *in vivo*. *Nat. Rev. Neurosci.* 4, 739–751. doi: 10.1038/nrn1198

Druckmann, S., and Chklovskii, D. B. (2012). Neuronal circuits underlying persistent representations despite time varying activity. *Curr. Biol.* 22, 2095–2103. doi: 10.1016/j.cub.2012.08.058

Dudai, Y. (2012). The restless engram: consolidations never end. *Annu. Rev. Neurosci.* 35, 227–247. doi: 10.1146/annurev-neuro-062111-150500

Durstewitz, D., Seamans, J. K., and Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nat. Neurosci.* 3, 1184–1191. doi: 10.1038/81460

Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., Logothetis, N. K., and Tolias, A. S. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science* 327, 584–587. doi: 10.1126/science.1179867

Ellwood, I. T., Patel, T., Wadia, V., Lee, A. T., Liptak, A. T., Bender, K. J., et al. (2017). Tonic or phasic stimulation of dopaminergic projections to prefrontal cortex causes mice to maintain or deviate from previously learned behavioral strategies. *J. Neurosci.* 37, 8315–8329. doi: 10.1523/JNEUROSCI.1221-17.2017

Enel, P., Wallis, J. D., and Rich, E. L. (2020). Stable and dynamic representations of value in the prefrontal cortex. *Elife* 9:e54313. doi: 10.7554/eLife.54313.sa2

Enoki, R., Hu, Y., Hamilton, D., and Fine, A. (2009). Expression of long-term plasticity at individual synapses in hippocampus is graded, bidirectional, and mainly presynaptic: optical quantal analysis. *Neuron* 62, 242–253. doi: 10.1016/j.neuron.2009.02.026

Euston, D. R., Gruber, A. J., and McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron* 76, 1057–1070. doi: 10.1016/j.neuron.2012.12.002

Euston, D. R., Tatsuno, M., and McNaughton, B. L. (2007). Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science* 318, 1147–1150. doi: 10.1126/science.1148979

Fiete, I. R., Senn, W., Wang, C. Z. H., and Hahnloser, R. H. R. (2010). Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron* 65, 563–576. doi: 10.1016/j.neuron.2010.02.003

Fujii, N., and Graybiel, A. M. (2003). Representation of action sequence boundaries by macaque prefrontal cortical neurons. *Science* 301, 1246–1249. doi: 10.1126/science.1086872

Fujisawa, S., Amarasingham, A., Harrison, M. T., and Buzsáki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat. Neurosci.* 11, 823–833. doi: 10.1038/nn.2134

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

Fusi, S., Drew, P. J., and Abbott, L. F. (2005). Cascade models of synaptically stored memories. *Neuron* 45, 599–611. doi: 10.1016/j.neuron.2005.02.001

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

Gallistel, C. R., and Matzel, L. D. (2013). The neuroscience of learning: beyond the Hebbian synapse. *Annu. Rev. Psychol.* 64, 169–200. doi: 10.1146/annurev-psych-113011-143807

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

Golomb, D., Hansel, D., and Mato, G. (2001). "Mechanisms of synchrony of neural activity in large networks," in Handbook of Biological Physics, *Volume 4: Neuro-Informatics and Neural Modelling*, eds F. Moss, and S. Gielen (Amsterdam: Elsevier Science), 887–968.

Goto, Y., Yang, C. R., and Otani, S. (2010). Functional and dysfunctional synaptic plasticity in prefrontal cortex: roles in psychiatric disorders. *Biol. Psychiatry* 67, 199–207. doi: 10.1016/j.biopsych.2009.08.026

Graupner, M., and Brunel, N. (2012). Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *PNAS* 109, 3991–3996. doi: 10.1073/pnas.1109359109

Graupner, M., Wallisch, P., and Ostojic, S. (2016). Natural firing patterns imply low sensitivity of synaptic plasticity to spike timing compared with firing rate. *J. Neurosci.* 36, 11238–11258. doi: 10.1523/JNEUROSCI.0104-16.2016

Greene, A. J., Prepscius, C., and Levy, W. B. (2000). Primacy versus recency in a quantitative model: activity is the critical distinction. *Learn. Mem.* 7, 48–57. doi: 10.1101/lm.7.1.48

Hahn, G., Petermann, T., Havenith, M. N., Yu, S., Singer, W., Plenz, D., et al. (2010). Neuronal avalanches in spontaneous activity *in vivo*. *J. Neurophysiol.* 104, 3312–3322. doi: 10.1152/jn.00953.2009

Haider, B., Duque, A., Hasenstaub, A. R., and McCormick, D. A. (2006). Neocortical network activity *in vivo* is generated through a dynamic balance of excitation and inhibition. *J. Neurosci.* 26, 4535–4545. doi: 10.1523/JNEUROSCI.5297-05.2006

He, K., Huertas, M., Hong, S. Z., Tie, X., Hell, J. W., Shouval, H., et al. (2015). Distinct eligibility traces for LTP and LTD in cortical synapses. *Neuron* 88, 528–538. doi: 10.1016/j.neuron.2015.09.037

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsycholocigal Theory*. New York, NY: John Wiley & Sons, Inc.

Hempel, C. M., Hartman, K. H., Wang, X.-J., Turrigiano, G. G., and Nelson, S. B. (2000). Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex. *J. Neurophysiol.* 83, 3031–3041. doi: 10.1152/jn.2000.83.5.3031

Hu, B., Sergei, K., Lei, Z., and Armin, S. (2005). Reversal of hippocampal LTP by spontaneous seizure-like activity: role of group i mGlur and cell depolarization. *J. Neurophysiol.* 93, 316–336. doi: 10.1152/jn.00172.2004

Isaacson, J. S., and Scanziani, M. (2011). How inhibition shapes cortical activity. *Neuron* 72, 231–243. doi: 10.1016/j.neuron.2011.09.027

Ito, H. T., Zhang, S.-J., Witter, M. P., Moser, E. I., and Moser, M.-B. (2015). A prefrontal–thalamo–hippocampal circuit for goal-directed spatial navigation. *Nature* 522, 50–55. doi: 10.1038/nature14396

Jadhav, S. P., Kemere, C., German, P. W., and Frank, L. M. (2012). Awake hippocampal sharp-wave ripples support spatial memory. *Science* 336, 1454–1458. doi: 10.1126/science.1217230

Jadhav, S. P., Rothschild, G., Roumis, D. K., and Frank, L. M. (2016). Coordinated excitation and inhibition of prefrontal ensembles during awake hippocampal sharp-wave ripple events. *Neuron* 90, 113–127. doi: 10.1016/j.neuron.2016.02.010

Jahr, C. E., and Stevens, C. F. (1990). Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics. *J. Neurosci.* 10, 3178–3182. doi: 10.1523/JNEUROSCI.10-09-03178.1990

Jensen, G., Altschul, D., Danly, E., and Terrace, H. (2013). Transfer of a serial representation between two distinct tasks by rhesus macaques. *PLoS ONE* 8:e70285. doi: 10.1371/journal.pone.0070285

Jun, J. K., Miller, P., Hernández, A., Zainos, A., Lemus, L., Brody, C. D., et al. (2010). Heterogenous population coding of a short-term memory and decision task. *J. Neurosci.* 30, 916–929. doi: 10.1523/JNEUROSCI.2062-09.2010

Kaefer, K., Nardin, M., Blahna, K., and Csicsvari, J. (2020). Replay of behavioral sequences in the medial prefrontal cortex during rule switching. *Neuron* 106, 154–165.e6. doi: 10.1016/j.neuron.2020.01.015

Keck, C., Savin, C., and Lücke, J. (2012). Feedforward inhibition and synaptic scaling – two sides of the same coin? *PLoS Comput. Biol.* 8:e1002432. doi: 10.1371/journal.pcbi.1002432

Keck, T., Toyoizumi, T., Chen, L., Doiron, B., Feldman, D. E., Fox, K., et al. (2017). Integrating Hebbian and homeostatic plasticity: the current state of the field and future research directions. *Philos. Transac. R. Soc. B Biol. Sci.* 372:20160158. doi: 10.1098/rstb.2016.0158

Klampfl, S., and Maass, W. (2013). Emergence of dynamic memory traces in cortical microcircuit models through STDP. *J. Neurosci.* 33, 11515–11529. doi: 10.1523/JNEUROSCI.5044-12.2013

La Camera, G., Fontanini, A., and Mazzucato, L. (2019). Cortical computations via metastable activity. *Curr. Opin. Neurobiol.* 58, 37–45. doi: 10.1016/j.conb.2019.06.007

Laje, R., and Buonomano, D. V. (2013). Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nat. Neurosci.* 16, 925–933. doi: 10.1038/nn.3405

Lengyel, I., Voss, K., Cammarota, M., Bradshaw, K., Brent, V., Murphy, K. P. S. J., et al. (2004). Autonomous activity of CaMKII is only transiently increased following the induction of long-term potentiation in the rat hippocampus. *Eur. J. Neurosci.* 20, 3063–3072. doi: 10.1111/j.1460-9568.2004.03748.x

Litwin-Kumar, A., and Doiron, B. (2014). Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nat. Commun.* 5:5319. doi: 10.1038/ncomms6319

Liu, J. K., and Buonomano, D. V. (2009). Embedding multiple trajectories in simulated recurrent neural networks in a self-organizing manner. *J. Neurosci.* 29, 13172–13181. doi: 10.1523/JNEUROSCI.2358-09.2009

London, M., Roth, A., Beeren, L., Häusser, M., and Latham, P. E. (2010). Sensitivity to perturbations *in vivo* implies high noise and suggests rate coding in cortex. *Nature* 466, 123–127. doi: 10.1038/nature09086

Lundqvist, M., Herman, P., and Miller, E. K. (2018). Working Memory: delay activity, yes! Persistent activity? Maybe not. *J. Neurosci.* 38, 7013–7019. doi: 10.1523/JNEUROSCI.2485-17.2018

Lutzu, S., and Castillo, P. E. (2021). Modulation of NMDA receptors by g-protein-coupled receptors: role in synaptic transmission, plasticity and beyond. *Neuroscience* 456, 27–42. doi: 10.1016/j.neuroscience.2020.02.019

Malenka, R. C., Lancaster, B., and Zucker, R. S. (1992). Temporal limits on the rise in postsynaptic calcium required for the induction of long-term potentiation. *Neuron* 9, 121–128. doi: 10.1016/0896-6273(92)90227-5

Manninen, T., Hituri, K., Hellgren Kotaleski, J., Blackwell, K. T., and Linne, M.-L. (2010). Postsynaptic signal transduction models for long-term potentiation and depression. *Front. Comput. Neurosci.* 4:152. doi: 10.3389/fncom.2010.00152

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

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

Markram, H., Gerstner, W., and Sjöström, P. J. (2012). Spike-timing-dependent plasticity: a comprehensive overview. *Front. Synaptic Neurosci.* 4:2. doi: 10.3389/fnsyn.2012.00002

Mashhoori, A., Hashemnia, S., McNaughton, B. L., Euston, D. R., and Gruber, A. J. (2018). Rat anterior cingulate cortex recalls features of remote reward locations after disfavoured reinforcements. *Elife* 7:e29793. doi: 10.7554/eLife.29793

Meador, K. J. (2007). The basic science of memory as it applies to epilepsy. *Epilepsia* 48, 23–25. doi: 10.1111/j.1528-1167.2007.01396.x

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., and Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J. Neurophysiol.* 100, 1407–1419. doi: 10.1152/jn.90248.2008

Moberly, A. H., Schreck, M., Bhattarai, J. P., Zweifel, L. S., Luo, W., and Ma, M. (2018). Olfactory inputs modulate respiration-related rhythmic activity in the prefrontal cortex and freezing behavior. *Nat. Commun.* 9:1528. doi: 10.1038/s41467-018-03988-1

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

Montgomery, J. M., and Madison, D. V. (2002). State-dependent heterogeneity in synaptic depression between pyramidal cell pairs. *Neuron* 33, 765–777. doi: 10.1016/S0896-6273(02)00606-2

Morrison, A., Aertsen, A., and Diesmann, M. (2007). Spike-timing-dependent plasticity in balanced random networks. *Neural Comput.* 19, 1437–1467. doi: 10.1162/neco.2007.19.6.1437

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

Nakajima, M., Schmitt, L. I., and Halassa, M. M. (2019). Prefrontal cortex regulates sensory filtering through a basal ganglia-to-thalamus pathway. *Neuron* 103, 445–458.e10. doi: 10.1016/j.neuron.2019.05.026

Naudé, J., Cessac, B., Berry, H., and Delord, B. (2013). Effects of cellular homeostatic intrinsic plasticity on dynamical and computational properties of biological recurrent neural networks. *J. Neurosci.* 33, 15032–15043. doi: 10.1523/JNEUROSCI.0870-13.2013

Onn, S.-P., and Wang, X.-B. (2005). Differential modulation of anterior cingulate cortical activity by afferents from ventral tegmental area and mediodorsal thalamus. *Eur. J. Neurosci.* 21, 2975–2992. doi: 10.1111/j.1460-9568.2005.04122.x

Onn, S.-P., Wang, X.-B., Lin, M., and Grace, A. A. (2006). Dopamine D1 and D4 receptor subtypes differentially modulate recurrent excitatory synapses in prefrontal cortical pyramidal neurons. *Neuropsychopharmacology* 31, 318–338. doi: 10.1038/sj.npp.1300829

Ostlund, S. B., Winterbauer, N. E., and Balleine, B. W. (2009). Evidence of action sequence chunking in goal-directed instrumental conditioning and its dependence on the dorsomedial prefrontal cortex. *J. Neurosci.* 29, 8280–8287. doi: 10.1523/JNEUROSCI.1176-09.2009

Papouin, T., Dunphy, J., Tolman, M., Foley, J. C., and Haydon, P. G. (2017). Astrocytic control of synaptic function. *Philos. Transac. R. Soc. B Biol. Sci.* 372:20160154. doi: 10.1098/rstb.2016.0154

Passecker, J., Mikus, N., Malagon-Vina, H., Anner, P., Dimidschstein, J., Fishell, G., et al. (2019). Activity of prefrontal neurons predict future choices during gambling. *Neuron* 101, 152–164.e7. doi: 10.1016/j.neuron.2018.10.050

Pasupathy, A., and Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433, 873–876. doi: 10.1038/nature03287

Paton, J. J., and Buonomano, D. V. (2018). The neural basis of timing: distributed mechanisms for diverse functions. *Neuron* 98, 687–705. doi: 10.1016/j.neuron.2018.03.045

Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I., and Battaglia, F. P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nat. Neurosci.* 12, 919–926. doi: 10.1038/nn.2337

Pfeiffer, B. E., and Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* 497, 74–79. doi: 10.1038/nature12112

Pouille, F., Marin-Burgin, A., Adesnik, H., Atallah, B. V., and Scanziani, M. (2009). Input normalization by global feedforward inhibition expands cortical dynamic range. *Nat. Neurosci.* 12, 1577–1585. doi: 10.1038/nn.2441

Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., and Johnson, M. R. (2007). Refreshing: a minimal executive function. *Cortex* 43, 135–145. doi: 10.1016/S0010-9452(08)70451-9

Renart, A., Rocha, J., de la, Bartho, P., Hollender, L., Parga, N., Reyes, A., et al. (2010). The asynchronous state in cortical circuits. *Science* 327, 587–590. doi: 10.1126/science.1179850

Rodriguez, G., Sarazin, M., Clemente, A., Holden, S., Paz, J. T., and Delord, B. (2018). Conditional bistability, a generic cellular mnemonic mechanism for robust and flexible working memory computations. *J. Neurosci.* 38, 5209–5219. doi: 10.1523/JNEUROSCI.1992-17.2017

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

Ruan, H., Saur, T., and Yao, W.-D. (2014). Dopamine-enabled anti-Hebbian timing-dependent plasticity in prefrontal circuitry. *Front. Neural Circuits* 8:38. doi: 10.3389/fncir.2014.00038

Schmitt, L. I., Wimmer, R. D., Nakajima, M., Happ, M., Mofakham, S., and Halassa, M. M. (2017). Thalamic amplification of cortical connectivity sustains attentional control. *Nature* 545, 219–223. doi: 10.1038/nature22073

Seidemann, E., Meilijson, I., Abeles, M., Bergman, H., and Vaadia, E. (1996). Simultaneously recorded single units in the frontal cortex go through sequences of discrete and stable states in monkeys performing a delayed localization task. *J. Neurosci.* 16, 752–768. doi: 10.1523/JNEUROSCI.16-02-00752.1996

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

Shenoy, K. V., Sahani, M., and Churchland, M. M. (2013). Cortical control of arm movements: a dynamical systems perspective. *Annu. Rev. Neurosci.* 36, 337–359. doi: 10.1146/annurev-neuro-062111-150509

Shima, K., Isoda, M., Mushiake, H., and Tanji, J. (2007). Categorization of behavioural sequences in the prefrontal cortex. *Nature* 445, 315–318. doi: 10.1038/nature05470

Shin, J. D., Tang, W., and Jadhav, S. P. (2019). Dynamics of awake hippocampal-prefrontal replay for spatial learning and memory-guided decision making. *Neuron* 104, 1110–1125.e7. doi: 10.1016/j.neuron.2019.09.012

Shinomoto, S., Miyazaki, Y., Tamura, H., and Fujita, I. (2005). Regional and laminar differences in *in vivo* firing patterns of primate cortical neurons. *J. Neurophysiol.* 94, 567–575. doi: 10.1152/jn.00896.2004

Shinomoto, S., Shima, K., and Tanji, J. (2003). Differences in spiking patterns among cortical neurons. *Neural Comput.* 15, 2823–2842. doi: 10.1162/089976603322518759

Shouval, H. Z., Bear, M. F., and Cooper, L. N. (2002). A unified model of NMDA receptor-dependent bidirectional synaptic plasticity. *PNAS* 99, 10831–10836. doi: 10.1073/pnas.152343099

Shu, Y., Hasenstaub, A., and McCormick, D. A. (2003). Turning on and off recurrent balanced cortical activity. *Nature* 423, 288–293. doi: 10.1038/nature01616

Siapas, A. G., Lubenov, E. V., and Wilson, M. A. (2005). Prefrontal phase locking to hippocampal theta oscillations. *Neuron* 46, 141–151. doi: 10.1016/j.neuron.2005.02.028

Siri, B. B., Quoy, M., Delord, B., Cessac, B., and Berry, H. (2007). Effects of Hebbian learning on the dynamics and structure of random networks with inhibitory and excitatory neurons. *J. Physiol. Paris* 101, 136–148. doi: 10.1016/j.jphysparis.2007.10.003

Song, S., Sjöström, P. J., Reigl, M., Nelson, S., and Chklovskii, D. B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* 3:e68. doi: 10.1371/journal.pbio.0030068

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

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

Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* 19, 394–405. doi: 10.1016/j.tics.2015.05.004

Sweatt, J. D. (2016). Dynamic DNA methylation controls glutamate receptor trafficking and synaptic scaling. *J. Neurochem.* 137, 312–330. doi: 10.1111/jnc.13564

Tanaka, J., Horiike, Y., Matsuzaki, M., Miyazaki, T., Ellis-Davies, G. C. R., and Kasai, H. (2008). Protein synthesis and neurotrophin-dependent structural plasticity of single dendritic spines. *Science* 319, 1683–1687. doi: 10.1126/science.1152864

Tchumatchenko, T., Geisel, T., Volgushev, M., and Wolf, F. (2010). Signatures of synchrony in pairwise count correlations. *Front. Comput. Neurosci.* 4:1. doi: 10.3389/neuro.10.001.2010

Thomson, A. M. (2002). Synaptic connections and small circuits involving excitatory and inhibitory neurons in layers 2-5 of adult rat and cat neocortex: triple intracellular recordings and biocytin labelling *in vitro*. *Cerebral Cortex* 12, 936–953. doi: 10.1093/cercor/12.9.936

Tiganj, Z., Jung, M. W., Kim, J., and Howard, M. W. (2017). Sequential firing codes for time in rodent medial prefrontal cortex. *Cerebral Cortex* 27, 5663–5671. doi: 10.1093/cercor/bhw336

Tononi, G., and Cirelli, C. (2003). Sleep and synaptic homeostasis: a hypothesis. *Brain Res. Bull.* 62, 143–150. doi: 10.1016/j.brainresbull.2003.09.004

Truccolo, W., Donoghue, J. A., Hochberg, L. R., Eskandar, E. N., Madsen, J. R., Anderson, W. S., et al. (2011). Single-neuron dynamics in human focal epilepsy. *Nat. Neurosci.* 14, 635–641. doi: 10.1038/nn.2782

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., and Nelson, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature* 391, 892–896. doi: 10.1038/36103

van Rossum, M. C., Bi, G. Q., and Turrigiano, G. G. (2000). Stable Hebbian learning from spike timing-dependent plasticity. *J. Neurosci.* 20, 8812–8821. doi: 10.1523/JNEUROSCI.20-23-08812.2000

Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C., and Gerstner, W. (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science* 334, 1569–1573. doi: 10.1126/science.1211095

Wang, H.-X., and Gao, W.-J. (2012). Prolonged exposure to NMDAR antagonist induces cell-type specific changes of glutamatergic receptors in rat prefrontal cortex. *Neuropharmacology* 62, 1808–1822. doi: 10.1016/j.neuropharm.2011.11.024

Wang, J., Narain, D., Hosseini, E. A., and Jazayeri, M. (2018). Flexible timing by temporal scaling of cortical responses. *Nat. Neurosci.* 21, 102–110. doi: 10.1038/s41593-017-0028-6

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

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

Witter, M. P., and Amaral, D. G. (2004). "CHAPTER 21 - Hippocampal Formation," in *The Rat Nervous System (Third Edition)*, ed G. Paxinos (Burlington, NJ: Academic Press), 635–704. doi: 10.1016/B978-012547638-6/50022-5

Wutz, A., Loonis, R., Roy, J. E., Donoghue, J. A., and Miller, E. K. (2018). Different levels of category abstraction by different dynamics in different prefrontal areas. *Neuron* 97, 716–726.e8. doi: 10.1016/j.neuron.2018.01.009

Xu, T.-X., and Yao, W.-D. (2010). D1 and D2 dopamine receptors in separate circuits cooperate to drive associative long-term potentiation in the prefrontal cortex. *PNAS* 107, 16366–16371. doi: 10.1073/pnas.1004108107

Xue, M., Atallah, B. V., and Scanziani, M. (2014). Equalizing excitation–inhibition ratios across visual cortical neurons. *Nature* 511, 596–600. doi: 10.1038/nature13321

Yang, S.-T., Shi, Y., Wang, Q., Peng, J.-Y., and Li, B.-M. (2014). Neuronal representation of working memory in the medial prefrontal cortex of rats. *Mol. Brain* 7:61. doi: 10.1186/s13041-014-0061-2

Yu, J. Y., Liu, D. F., Loback, A., Grossrubatscher, I., and Frank, L. M. (2018). Specific hippocampal representations are linked to generalized cortical representations in memory. *Nat. Commun.* 9:2209. doi: 10.1038/s41467-018-04498-w

Zenke, F., Gerstner, W., and Ganguli, S. (2017). The temporal paradox of Hebbian learning and homeostatic plasticity. *Curr. Opin. Neurobiol.* 43, 166–176. doi: 10.1016/j.conb.2017.03.015

Zenke, F., Hennequin, G., and Gerstner, W. (2013). Synaptic plasticity in neural networks needs homeostasis with a fast rate detector. *PLoS Comput. Biol.* 9:e1003330. doi: 10.1371/journal.pcbi.1003330

Zhang, W., and Linden, D. J. (2003). The other side of the engram: experience-driven changes in neuronal intrinsic excitability. *Nat. Rev. Neurosci.* 4:885. doi: 10.1038/nrn1248

Zielinski, M. C., Shin, J. D., and Jadhav, S. P. (2019). Coherent coding of spatial position mediated by theta oscillations in the hippocampus and prefrontal cortex. *J. Neurosci.* 39, 4550–4565. doi: 10.1523/JNEUROSCI.0106-19.2019

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neural Circuits

REVIEW
published: 21 July 2021
doi: 10.3389/fncir.2021.696060

Persistent Activity During Working Memory From Front to Back

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

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

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

Keywords: working memory, saccades, PFC, FEF, Lip, fMRI, decoding, visual cortex

OPEN ACCESS

Edited by:
Shintaro Funahashi,
Kyoto University, Japan

Reviewed by:
Athanasia Papoutsi
Foundation of Research and Technology-Hellas (FORTH), Greece
Susan Courtney,
Johns Hopkins University, United States

***Correspondence:**
Clayton E. Curtis
clayton.curtis@nyu.edu

Received: *16 April 2021*
Accepted: *28 June 2021*
Published: *21 July 2021*

Citation:
Curtis CE and Sprague TC (2021) Persistent Activity During Working Memory From Front to Back. Front. Neural Circuits 15:696060. doi: 10.3389/fncir.2021.696060

INTRODUCTION

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

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

NEURAL ACTIVITY PERSISTS IN THE PREFRONTAL CORTEX

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

Following these pioneering studies, the experimental techniques matured and over the next 30 years our knowledge about the relationships between persistent activity and WM accumulated. For example, persistent activity in PFC neurons is not limited to spatial WM. PFC neurons that show preferences for both simple (e.g., color) and complex (e.g., face) objects exhibit activity that persists while monkeys maintain these objects in WM (Quintana et al., 1988; Miller et al., 1996; Ó Scalaidhe et al., 1999; Fuster et al., 2000; Panichello and Buschman, 2021). Assuming that stimulus selective persistent activity is the mechanism by which WM representations are stored, PFC neurons appear to store any type of stimulus feature, including the frequency of tactile flutter (Romo et al., 1999), the direction of dot motion (Zaksas and Pasternak, 2006; Mendoza-Halliday et al., 2014), sound location (Fuster et al., 2000; Kikuchi-Yorioka and Sawaguchi, 2000), and audiovisual macaque vocalizations (Hwang and Romanski, 2015). Moreover, they encode memory-guided prospective motor plans (Funahashi et al., 1993b; Takeda and Funahashi, 2002; Markowitz et al., 2015) and the prospective sensory features of a delayed paired associate (Rainer et al., 1999; Fuster et al., 2000). Finally, persistent activity appears to even encode complex task rules and contexts (Asaad et al., 2000; Wallis et al., 2001), abstract categories (Freedman et al., 2001), and selective conjunctions of objects and locations (Rao et al., 1997; Rainer et al., 1998) that cannot be explained by simpler stimulus or location specific representations.

CANONICAL PFC MICROCIRCUIT MODEL OF WM

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

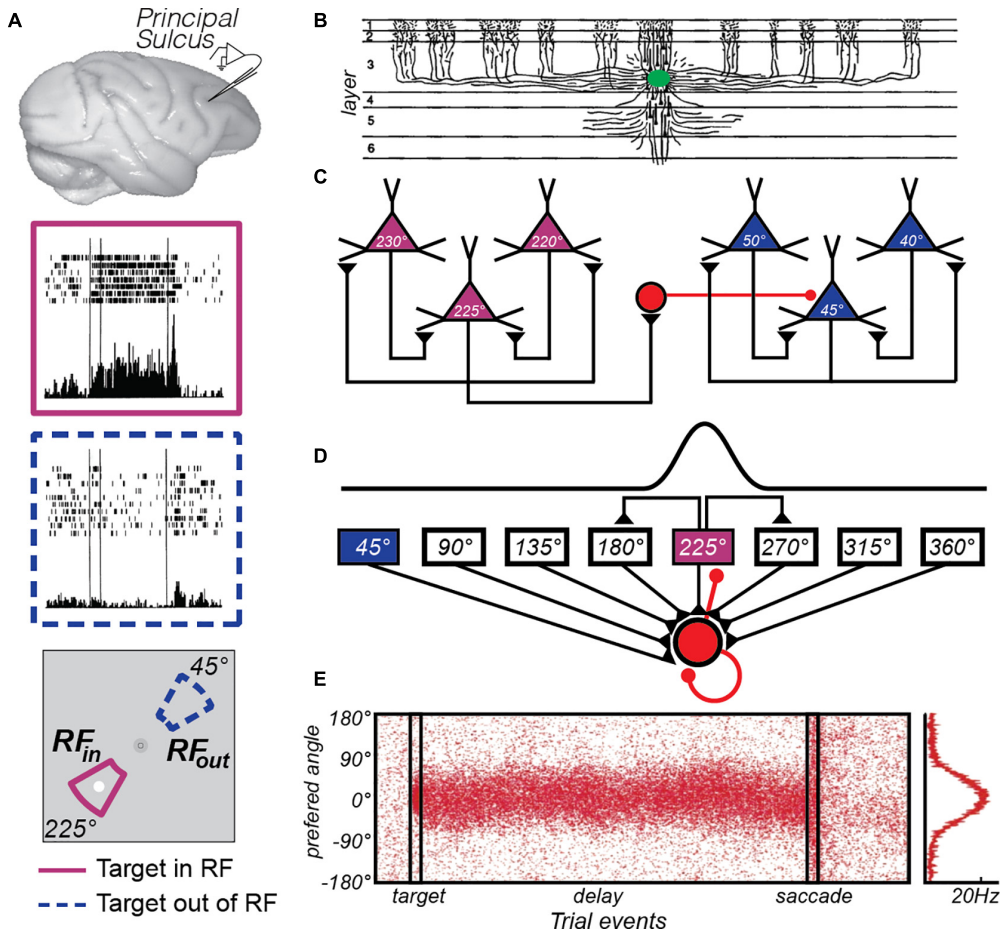


FIGURE 1 | The canonical PFC microcircuit model of WM. **(A)** Neural activity recorded from the principal sulcus in the macaque dorsolateral PFC. Activity persists during the delay period of memory-guided saccade tasks. The two insets depict a PFC neuron's response when the memory target appears in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

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

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

TRANSLATING THE PRIMATE PFC MODEL OF HUMAN WM

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

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

Measuring neural activity with fMRI while humans perform spatial WM tasks, including memory-guided saccade WM tasks like those used to initially study the macaque PFC (Funahashi et al., 1989), we find that fMRI is perfectly sensitive to WM representations of single items (Curtis et al., 2004; Curtis and D'Esposito, 2006; Schluppeck et al., 2006; Srimal and Curtis, 2008; Tark and Curtis, 2009; Jerde et al., 2012; Sprague et al., 2014; Saber et al., 2015; Rahmati et al., 2020; Hallenbeck et al., in press). However, in none of the above cited studies did we find evidence that neural activity persists in the human dorsolateral PFC during simple spatial WM tasks. On the other hand, in each one of those studies we found evidence that activity persists, in a variety of meaningful ways, in the superior spur of the precentral sulcus (PCS) in the frontal cortex and/or in the posterior part of the intraparietal sulcus (IPS) (**Figure 2A**).

On the face of it, these results conflict between the two species. In the monkey, neurons in dorsolateral PFC show persistent activity and lesions cause WM impairments. However, in humans neural activity only persists in the PCS, not in more anterior parts of the PFC in areas homologous to the macaque principal sulcus. We generated two hypotheses to explain these conflicting results based on the impact that lesions to the PFC and PCS had on WM (**Figure 1A**). If lesions to the dorsolateral PFC that spare the PCS cause WM impairments, like they do in monkeys, this would indicate that fMRI may not be sensitive enough to measure persistent activity in that part of the brain (hypothesis 1). On the other hand, if lesions to the PCS, rather than PFC, cause WM impairments, this would indicate that the human dorsolateral PFC is not necessary for WM like it is in the monkey (hypothesis 2). In support of hypothesis 2, the accuracy of memory-guided saccades was unimpacted by dorsolateral PFC resections as long as they spared the PCS (**Figure 2B**). PCS lesions increased the magnitude of memory errors largely when the target was in the contralesional hemifield (Mackey et al., 2016b, 2017; Mackey and Curtis, 2017). In order to rule out other factors, like reorganization or compensation in the lesion patients, we repeated the study using transcranial magnetic stimulation (TMS) applied to the superior PCS and the intermediate frontal sulcus in the PFC during the memory delay in a healthy cohort of participants (Mackey et al., 2016b, 2017; Mackey and Curtis, 2017). The TMS results replicated the patient study; TMS to the sPCS, but not dorsolateral PFC, caused an increase in memory-guided saccade errors (**Figure 2C**). These results are consistent with previous studies that have investigated the impact of dorsolateral PFC and/or PCS damage on both spatial and non-spatial forms of WM (D'Esposito and Postle, 1999; Ploner et al., 1999; Postle et al., 2003).

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

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

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

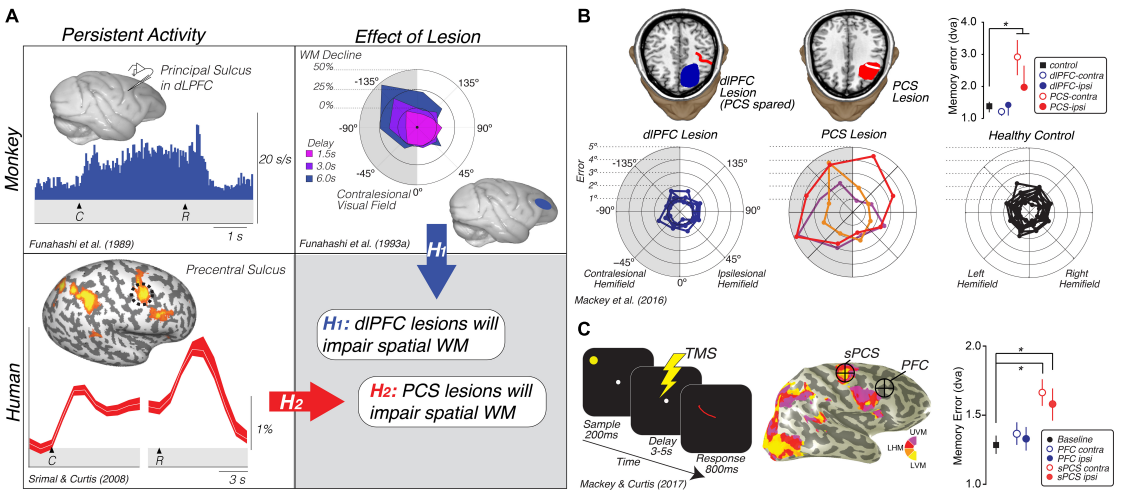


FIGURE 2 | Translating the canonical model of WM to humans. **(A)** Rationale and hypotheses of studies of lesion (Mackey et al., 2016b) and TMS perturbation (Mackey and Curtis, 2017) of human PFC. Neural activity persists in the monkey dlPFC during the retention interval of memory-guided saccade tasks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

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

NEURAL ACTIVITY PERSISTS BEYOND PFC

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

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

In humans, measuring delay period activity with fMRI BOLD supports these non-human primate reports in just how widely distributed persistent activity appears to be during WM. There have been a number of reviews recently of human neuroimaging studies of WM (e.g., Sreenivasan et al., 2014; D'Esposito and Postle, 2015; Christophel et al., 2017; Sreenivasan and D'Esposito, 2019) and thus we will instead focus on instructive examples of persistent activity measured in humans with fMRI. Moreover, we focus on spatial WM because of its widespread use in both species and the ease with which neural encoding properties can be measured in both species. In many of the monkey electrophysiological studies reviewed above, an important first step involved characterizing each neuron's RF or its preferred stimulus feature. This then allowed researchers to compare memory responses between stimuli placed within and outside of each neuron's RF (or compare between preferred and non-preferred stimuli). Utilizing the same logic, advances in population receptive field (pRF) mapping (Dumoulin and Wandell, 2008; Wandell and Winawer, 2015; Mackey et al., 2017) allow researchers to compare BOLD estimates of persistent activity between trials in which the memoranda fall within and outside of a voxel's pRF. In **Figure 3**, the time courses of BOLD activity during a memory-guided saccade task are shown for ten visual field maps (Rahmati et al., 2020; Hallenbeck et al., in press). Each visual field map contains either an upright or inverted representation of the contralateral visual field. Within each map, the location and size of the pRF of each voxel in these maps can be estimated using non-linear optimization techniques. Then, averaging BOLD signal over trials can be performed in a principled way according to the match between voxels' pRF positions and the locations of the memorized targets. Overall, activity persists during the delay period in almost all of these maps. Moreover, the amplitude of BOLD activity is generally greater among voxels with pRFs matching the target, compared to voxels with pRFs 180° away from the target (on the opposite side of fixation).

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

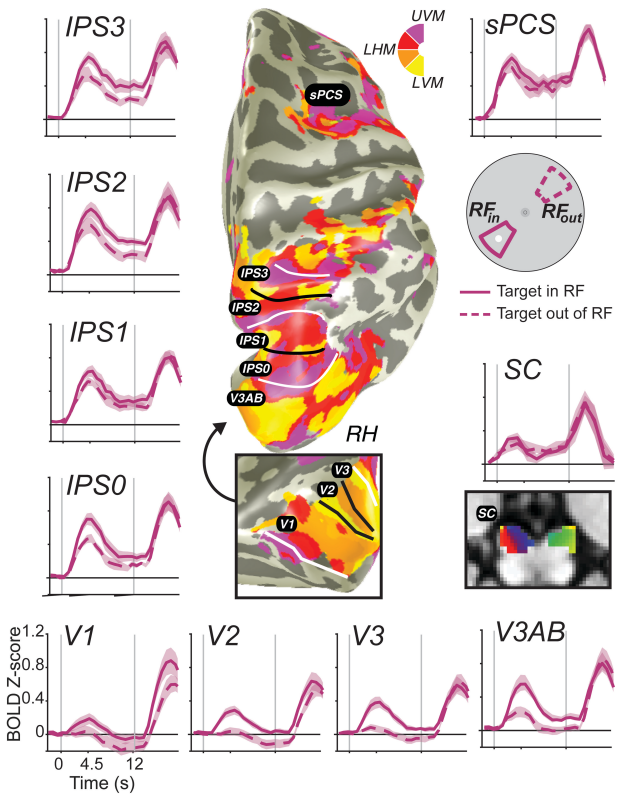


FIGURE 3 | Stimulus-selective persistent activity across the human brain measured with fMRI. BOLD activity persists during the retention interval of memory-guided saccade tasks in many cortical and subcortical brain areas (Hallenbeck et al., in press). Each region-of-interest was defined using modified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

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

DECODING WM CONTENTS FROM POPULATION-LEVEL ACTIVATION PATTERNS

As reviewed above in the section on translational studies, attempts to identify persistent activity in human dorsolateral PFC during WM tasks that require simple maintenance have largely been unsuccessful. However, these studies typically leveraged mass univariate analysis approaches which average responses across all trials in the experiment and over neighboring voxels by way of smoothing. As a result, these analyses effectively focus on the *similarities* in fMRI activation across all unique remembered stimuli. That is–to isolate activation related to the maintenance of information over the delay period, trials corresponding to all possible WM contents are combined. Such averaging necessarily masks important differences in activation associated with specific types of stimuli–for example, the particular location, orientation, or color held in WM.

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

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

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

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

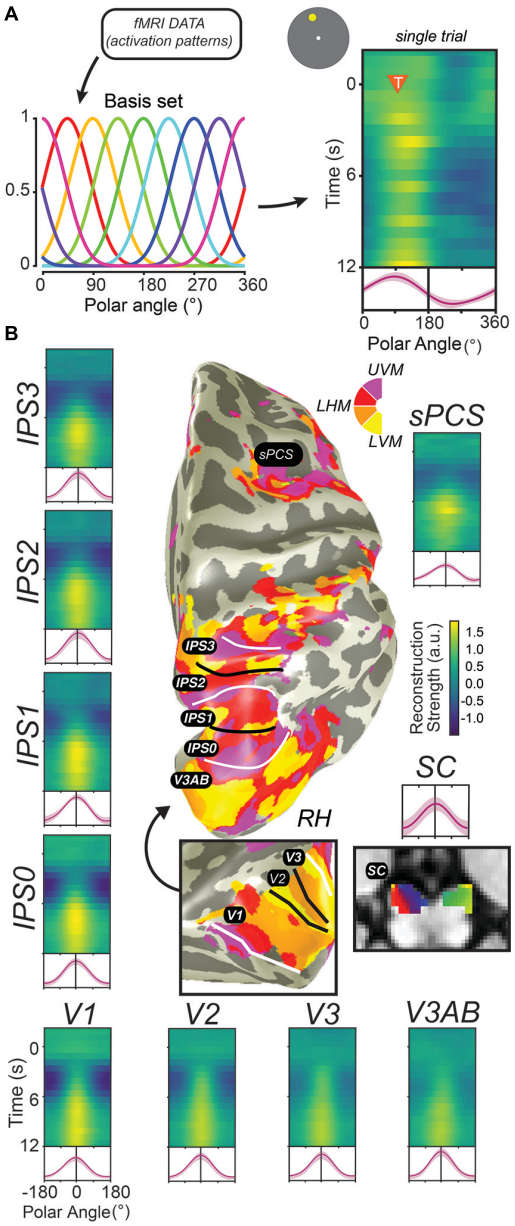


FIGURE 4 | Decoding methods reveal stimulus-selective persistent activity across cortical and subcortical regions in the human brain. **(A)** Recent studies have employed "IEMs," which model the activation of each voxel as a weighted combination of neural information channels (Brouwer and Heeger, 2009;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

FIGURE 4 | Continued
along with the average channel response profile over the final 1.5 s of the delay period (red line). Data adapted from Hallenbeck et al. (in press) (**A,B**; all ROIs except SC) and Rahmati et al. (2020) (**B**; SC). This Figure depicts the same data shown in **Figure 3** analyzed in a different way.

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

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

When such an analysis is applied to data acquired during a memory-guided saccade task analogous to that used in macaques, greater activation is measured in voxels with spatial RFs near the remembered location as compared to those voxels with spatial RFs farther away (**Figure 3**; Saber et al., 2015; Hallenbeck et al., in press). Moreover, this holds for features like orientation: recent studies which have instead attempted to "localize" voxels preferring one or another orientation and directly compare activation between these subpopulations support this notion: voxels labeled with the orientation remembered on a trial show elevated activation compared with those labeled with the non-remembered orientation (Lawrence et al., 2018). These results track with those observed in the classical studies of macaque DLPFC which show elevated neural firing for neurons which prefer the remembered location as compared to those with more distal preferences (Funahashi et al., 1989).

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

FURTHER CHALLENGES TO THE CANONICAL PFC MODEL OF WM

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

Is Persistent Activity in PFC an Artifact of Averaging?

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

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

Dynamic Codes for WM Content

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

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

While there certainly does exist ample evidence that dynamic responses at the single-unit level can be observed, and that they can in some cases support a stable population-level neural code, it is critical to note that these studies do not negate the existence nor importance of other stable coding mechanisms, some of which are observed in the same studies. For example, it has been shown that neurons with dynamic responses and those with stable responses coexist in PFC, and their response dynamics can be well-predicted by their intrinsic "time constant" [their autocorrelation function measured from inter-trial intervals; (Wasmuht et al., 2018)]. That is–nearby neurons in the same brain region can either show evidence for dynamic coding or stable coding. In another study, macaques performed an oculomotor delayed response task with an intervening irrelevant distractor stimulus. Activity patterns measured from LPFC "morphed" following the distractor, but patterns measured from the FEF of the same animals did not show evidence for such dynamic morphing (Parthasarathy et al., 2017). These results show that even when dynamic codes are observed, stable subspaces (consistent with a fixed readout rule) can account for a large amount of the response dynamics, and moreover, that stable coding is simultaneously observed in other neurons and/or brain regions.

Mixed Selectivity in PFC

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

"Activity-Silent" WM Representations

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

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

Distractors Impact WM Representations in Sensory Regions

Cognitive theories about the nature of WM representations have long been informed by behavioral studies of the distracting effects of material presented during WM retention intervals (Baddeley, 1986). Intervening information is more disruptive when its features match the contents of WM. For instance, intervening phonological but not visual information impairs one's ability to maintain visually presented strings of letters, suggesting an important role of articulatory processes for items that are verbalizable (Logie et al., 1990). Similarly, intervening visuospatial processing and oculomotion selectively impacts spatial WM (Postle et al., 2006). In general, many conclusions about the formats of WM representations depend on the logic that the effectiveness of distraction depends on how well the representational formats of the distractor and memoranda are matched. Neuroscientific studies have also relied on a similar logic, assuming that the competition or interaction between the neural representations of the memoranda and the intervening distractor disrupts memory. As reviewed above, early electrophysiology and neuroimaging studies focused primarily on the importance of persistent activity in the PFC. Until recently, the potential importance of posterior cortical areas in WM had been largely neglected. Indeed, neural activity in monkey inferotemporal cortex is less robust during memory delays and the selectivity of activity appears to be disrupted by intervening distractors, while PFC representations appear resistant to distraction (Miller et al., 1996). Similarly, memoranda-specific delay period activity of neurons in the monkey PFC resists the effect of distractors, especially when compared to neurons in posterior parietal cortex (di Pellegrino and Wise, 1993; Constantinidis and Steinmetz, 1996; Suzuki and Gottlieb, 2013). Inferences stemming from the underlying logic of these distractor studies imply that the PFC, rather than posterior cortical areas, is critical for WM storage.

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

CONCLUDING REMARKS: PERSISTENT ACTIVITY PERSISTS

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

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

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

Moving forward, there are a number of questions that we instead need to address. Most relevant is: what, then, *is* the role of the dorsolateral PFC? Note that persistent activity is simply an observation and is not synonymous with WM maintenance (Curtis and Lee, 2010). Perhaps persistent activity in PFC reflects not the storage of WM features, but rather some mechanism related to the control of WM representations stored elsewhere, maybe by their own persistent activity (Miller and Cohen, 2001; Curtis and D'Esposito, 2003; Emrich et al., 2013; D'Esposito and Postle, 2015; Postle and Yu, 2020). Also, why do we see evidence of persistent activity, even for a simple single item WM task, in so many cortical and subcortical brain areas (Christophel et al., 2017; Leavitt et al., 2017)? Redundancy is good to a point, but future research should try to figure out which of these numerous areas are necessary, what types of features they might be representing, and if they might be encoding different representational formats of WM. For example, disrupting persistent activity with intervening distraction (Lorenc et al., 2018; Rademaker et al., 2019; e.g., Hallenbeck et al., in press) or TMS (e.g., Mackey and Curtis, 2017; Rademaker et al., 2017) may be able to disentangle the relative roles of different cortical regions. However, such efforts are tricky, as a distractor may not affect a top-down control signal, especially when passively viewed. Experiments parametrically manipulating task demands in concert with visual distraction may help further clarify the relative role different brain regions play in WM tasks. An especially promising avenue for future exploration is comparing decoded feature values from single trials of fMRI activation to behavioral errors on those same trials (Ester et al., 2016; Hallenbeck et al., in press; Li et al., in press).

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

AUTHOR CONTRIBUTIONS

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

FUNDING

This work was supported by CEC: R01 EY-016407 and R01 EY-027925 and TCS: F32-EY028438; Sloan Research Fellowship.

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Lorenc, E. S., and Sreenivasan, K. K. (2021). Reframing the debate: the distributed systems view of working memory. *Vis. Cogn.* 29, 1–9. doi: 10.1080/13506285.2021.1899091

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

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

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

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

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

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

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

Mackey, W. E., Devinsky, O., Doyle, W. K., Meager, M. R., and Curtis, C. E. (2016b). Human dorsolateral prefrontal cortex is not necessary for spatial working memory. *J. Neurosci.* 36, 2847–2856. doi: 10.1523/jneurosci.3618-15.2016

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

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

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

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

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457. doi: 10.1037/0033-295x.102.3.419

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

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

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Postle, B. R., Idzikowski, C., Sala, S. D., Logie, R. H., and Baddeley, A. D. (2006). The selective disruption of spatial working memory by eye movements. *Q. J. Exp. Psychol.* 59, 100–120. doi: 10.1080/17470210500151410

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neural Circuits

BRIEF RESEARCH REPORT
published: 20 September 2021
doi: 10.3389/fncir.2021.716965

Across-Area Synchronization Supports Feature Integration in a Biophysical Network Model of Working Memory

*Joao Barbosa[1,2], Vahan Babushkin[3], Ainsley Temudo[3], Kartik K. Sreenivasan[3] and Albert Compte[1]**

[1] Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain, [2] Laboratoire de Neurosciences Cognitives et Computationnelles, INSERM U960, Ecole Normale Supérieure – PSL Research University, Paris, France, [3] Division of Science and Mathematics, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Working memory function is severely limited. One key limitation that constrains the ability to maintain multiple items in working memory simultaneously is so-called swap errors. These errors occur when an inaccurate response is in fact accurate relative to a non-target stimulus, reflecting the failure to maintain the appropriate association or "binding" between the features that define one object (e.g., color and location). The mechanisms underlying feature binding in working memory remain unknown. Here, we tested the hypothesis that features are bound in memory through synchrony across feature-specific neural assemblies. We built a biophysical neural network model composed of two one-dimensional attractor networks – one for color and one for location – simulating feature storage in different cortical areas. Within each area, gamma oscillations were induced during bump attractor activity through the interplay of fast recurrent excitation and slower feedback inhibition. As a result, different memorized items were held at different phases of the network's oscillation. These two areas were then reciprocally connected via weak cortico-cortical excitation, accomplishing binding between color and location through the synchronization of pairs of bumps across the two areas. Encoding and decoding of color-location associations was accomplished through rate coding, overcoming a long-standing limitation of binding through synchrony. In some simulations, swap errors arose: "color bumps" abruptly changed their phase relationship with "location bumps." This model, which leverages the explanatory power of similar attractor models, specifies a plausible mechanism for feature binding and makes specific predictions about swap errors that are testable at behavioral and neurophysiological levels.

Keywords: working memory, binding, oscillations, multi-area, attractor network

OPEN ACCESS

Edited by:
Shintaro Funahashi, Kyoto University, Japan

Reviewed by:
Leonid L. Rubchinsky, Indiana University, Purdue University Indianapolis, United States
Shoji Tanaka, Sophia University, Japan
Louis Kang, RIKEN Center for Brain Science (CBS), Japan

****Correspondence:***
Albert Compte
acompte@clinic.cat

Received: *29 May 2021*
Accepted: *11 August 2021*
Published: *20 September 2021*

Citation:
Barbosa J, Babushkin V, Temudo A, Sreenivasan KK and Compte A (2021) Across-Area Synchronization Supports Feature Integration in a Biophysical Network Model of Working Memory. Front. Neural Circuits 15:716965. doi: 10.3389/fncir.2021.716965

INTRODUCTION

Working memory, our ability to hold information in mind for short time periods, is a hallmark of cognition but is severely limited on several fronts (Ma et al., 2014). Some of its limitations, such as its capacity, precision, or specific quantitative biases have been successfully accounted for by a family of biophysically-constrained models, mostly on the basis of a ring attractor network that maintains

memoranda through sustained reverberatory neural activity (activity bumps) (Compte et al., 2000; Edin et al., 2009; Wei et al., 2012; Wimmer et al., 2014; Almeida et al., 2015; Papadimitriou et al., 2015; Nassar et al., 2018; Bouchacourt and Buschman, 2019; Qi et al., 2019; Barbosa et al., 2020). A feature of working memory that constrains the simultaneous storage of several items is the presence of swap errors (Schneegans and Bays, 2019). These errors occur when an inaccurate response to the target item is in fact accurate relative to a non-target item, reflecting the failure to maintain the appropriate association or "binding" between the separate features that define each item (e.g., color and location). The neural mechanisms supporting feature binding remain unclear, with different computational models implementing two alternative hypotheses (Raffone and Wolters, 2001; Swan and Wyble, 2014; Matthey et al., 2015; Schneegans et al., 2016; Pina et al., 2018; Schneegans and Bays, 2019).

The first type of models are based on selective synchronization (Raffone and Wolters, 2001; Pina et al., 2018). In these models, different neuronal populations selective to each feature that define an object are bound together through synchronized oscillatory activity. This would answer the longstanding question of how independently encoded features could be flexibly encoded as a single concept (Singer, 1999). Thanks to this flexibility, at least conceptually, these models do not suffer from combinatorial explosion as an increasing number of feature combinations are considered. There are, however, important questions about the biological plausibility of this hypothesis. Crucially, such a framework would need a temporal *encoder* that tags bound features by a "temporal code" and a temporal *decoder* that is able to distinguish which features are associated by detecting ensembles oscillating in precise synchrony. Both the encoder and decoder would thus depend on undefined biological mechanisms for spike coincidence detection (Shadlen and Movshon, 1999), which would struggle with the known high variability of neural spiking in sustained activity (Compte et al., 2003; Shafi et al., 2007). However, there is ample evidence for oscillatory dynamics during working memory. For instance, oscillatory activity in the gamma band (roughly defined between 30 and 100 Hz) increases during the mnemonic periods, both locally (Pesaran et al., 2002; Wimmer et al., 2016) and across sites (Lutzenberger et al., 2002; Kaiser et al., 2003; Palva et al., 2011; Kornblith et al., 2016), and further increases with memory load (Howard et al., 2003; van Vugt et al., 2010; Kornblith et al., 2016; Lundqvist et al., 2016). Importantly, gamma-band activity seems to play a functional role, as working memory binding performance is increased when transcranial stimulation at gamma frequency (40 Hz) is applied at two different sites (left temporal and parietal), but only when in anti-phase (Tseng et al., 2016) in line with monkey electrophysiology showing that different items are stored in different oscillatory phases (Siegel et al., 2009) and the more general framework of phase-coding in working memory (Fell and Axmacher, 2011).

Another class of models achieve feature binding through "conjunction neurons" – neurons that are selective to all features being bound. Since neurons with mixed selectivity are ubiquitous in the brain (Rigotti et al., 2013; Fusi et al., 2016), these models seem more biologically plausible than those relying on unrealistically precise spike synchronization. Nevertheless, they suffer from some important limitations. First, the number of possible combinations explode quickly with an increasing number of features (Matthey et al., 2015; Schneegans et al., 2016; Schneegans and Bays, 2017, 2019; but see Swan and Wyble, 2014). Second, these models do not have independent storage systems for each feature that define an object, to which there is converging evidence (Olson and Jiang, 2002; Wheeler and Treisman, 2002; Xu, 2002; Delvenne and Bruyer, 2004; Bays et al., 2011b; Fougnie and Alvarez, 2011; Parra et al., 2011). See Ma et al. (2014) and Schneegans and Bays (2019) for recent reviews on the experimental evidence that should constrain multi-item working memory models, in particular those aiming to explain feature binding.

Here, we propose a hybrid model that overcomes several limitations from both types of models. We connected two ring attractor networks – one ring representing and memorizing colors and another ring storing locations – via weak excitation. This is an explicit implementation of the independent storage of individual features, where each feature might be represented in different cortical areas (e.g., color in inferior temporal cortex and location in posterior parietal cortex). Within each area, oscillatory mnemonic activity occured naturally through the interplay between fast recurrent excitation and slower inhibitory feedback. Feature binding was accomplished through the selective synchronization of pairs of bumps across the two networks. Furthermore, encoding and decoding of specific color-location associations was accomplished through rate coding. Our hybrid model of rate/temporal coding shares the rich explanatory power of classical ring-attractor models of working memory (Edin et al., 2009; Wei et al., 2012; Wimmer et al., 2014; Almeida et al., 2015; Papadimitriou et al., 2015; Nassar et al., 2018; Bouchacourt and Buschman, 2019; Qi et al., 2019; Barbosa et al., 2020) and derives new predictions that can be tested on multiple levels.

MATERIALS AND METHODS

Neural Network Model

We extended a previously proposed computational model (Compte et al., 2000). In particular, we connected two one-dimensional ring networks via weak, cortico-cortical excitatory synapses governed by AMPAR-dynamics. Each network consists of 2,048 excitatory and 512 inhibitory leaky integrate-and-fire neurons fully connected through AMPAR-, NMDAR-, and $GABA_AR$-mediated synaptic transmission as in Compte et al. (2000). Moreover, excitatory and inhibitory neurons were spatially distributed on a ring so that nearby neurons encoded nearby spatial locations. All connections were all-to-all and spatially tuned, so that nearby neurons with similar preferred directions had stronger than average connections, while distant neurons had weaker connections. Inhibitory-to-inhibitory and across-network connectivity was untuned. Intrinsic parameters for both cell types and all the connectivity parameters were taken from Compte et al. (2000), except the following for networks

holding up to two stimuli or capacity-2 networks (notation consistent with Compte et al., 2000):

$$G_{EE,\,AMPA} = 0.09\text{ nS},\ G_{EI,\,AMPA} = 0.256\text{ nS},$$
$$G_{EE,\,NMDA} = 0.24\text{ nS},\ G_{EI,\,NMDA} = 0.11\text{ nS},$$
$$G_{II,\,GABA} = 2\text{ nS},\ G_{IE,\,GABA} = 3\text{ nS},$$
$$g_{ext,\,I} = 2.74\text{ nS},\ g_{ext,\,E} = 3.5\text{ nS},$$
$$J_{EE}^{+} = 10,\ \sigma_{EE} = 9,\ J_{EI}^{+} = J_{IE}^{+} = 2.4,\ \sigma_{EI} = \sigma_{IE} = 18.$$

For networks holding up to three stimuli (capacity-3 networks),

$$G_{EE,\,AMPA} = 0.126\text{ nS},\ G_{EI,\,AMPA} = 0.256\text{ nS},$$
$$G_{EE,\,NMDA} = 0.2\text{ nS},\ G_{EI,\,NMDA} = 0.11\text{ nS},$$
$$G_{II,\,GABA} = 2\text{ nS},\ G_{IE,\,GABA} = 3\text{ nS},$$
$$g_{ext,\,I} = 2.8\text{ nS},\ g_{ext,\,E} = 3.58\text{ nS},$$
$$J_{EE}^{+} = 11,\ \sigma_{EE} = 9,\ J_{EI}^{+} = J_{IE}^{+} = 2.6,\ \sigma_{EI} = \sigma_{IE} = 30.$$

Connectivity across networks was determined by the following conductances (for unconnected simulations, these conductances were set to zero):

$$G_{EE,\,AMPA,\,across} = 0.45\text{ nS},\ G_{EI,\,AMPA,\,across} = 0.18\text{ nS},$$
$$G_{EE,\,NMDA,\,across} = G_{EI,\,NMDA,\,across} = 0\text{ nS}.$$

These parameters were adjusted to have within-network oscillations, which was accomplished by increasing the ratio between fast and slow excitation, supported, respectively, by AMPAR and NMDAR channels, as previously shown (Compte et al., 2000). The main dynamics described in this study were robust to a broad range of parameter values (**Figures 1–4**).

Cross-Correlations

For the cross-correlation analyses, we computed spike counts in bins of 5 ms, collapsing all neurons around the stimulus presentation location (here called a *bump*, ±340 neurons). Moreover, we computed within- and across-network correlations by, respectively, considering neurons in bumps from the same or different circuits. For the cross-frequency correlation plots (e.g., **Figure 2B**), we further computed the power spectrum of the resulting cross-correlation functions, averaged across all possible (only within- or only across-) pairs of bumps.

Conversion of Spikes Into Local Field Potentials

For the conversion of simulated spike trains into local field potentials, we convolved the aggregated spike times (t_s) of all the neurons engaged in a bump (or in the network, depending on the analysis) with an alpha-function synaptic kernel:

$$LFP(t) = \sum_{t_s} \Theta(t - t_s)\frac{t - t_s}{\tau}\exp\left(-\frac{t - t_s}{\tau}\right)$$

with $\Theta(t)$ being the Heaviside theta function, and $\tau = 5$ ms.

Phase-Preservation Index

To measure how an oscillating activity bump kept its oscillatory phase over multiple trials ($k = 1,\ldots,N$) of our simulation, we first converted spike times into local-field potentials (see above). Through wavelet analysis, we determined the phase $\phi^k(f_0, t)$ of the LFP at $f_0 = 30$ Hz (the approximate frequency of oscillations in the network) at all time points t of the simulation, and then we used the phase-preservation index (PPI), a method originally developed by Mazaheri and Jensen (2006) for EEG data.

The PPI is defined by taking a reference time point (in our case t_{ref} = stimulus offset), and then computing the average consistency of the phases at the specific frequency of interest $f_0$with the rest of the time points:

$$PPI(f_0, t) = \frac{1}{N}\left|\sum_{k=1}^{N} e^{i\phi^k(f_0,t_{ref}) - i\phi^k(f_0,t)}\right|$$

Phase-preservation index values thus vary between 0 and 1, with 1 indicating perfect phase consistency.

Extracting Behavioral Output With a Mixture of Gaussians

The final behavioral output, for simplicity, was extracted by fitting a mixture of two gaussians to the late-delay average activity of the color network. We then selected the central value (color) of the gaussian component with larger amplitude, or stronger mixture component. We fit the mixture of gaussians using the Python function *sklearn.mixture.GMM*. This algorithmic read-out could be replaced by a biologically plausible downstream network connected to the color circuit, and tuned to be in a winner-take-all regime – i.e., only able to maintain one bump at a time.

RESULTS

Working Memory Load Modulates Oscillation Power and Frequency

We built a computational network model of a local neocortical circuit, with excitatory and inhibitory spiking neurons (*leaky integrate-and-fire* neuron model) connected reciprocally via excitatory AMPAR-mediated and NMDAR-mediated synapses and inhibitory $GABA_AR$-mediated synapses (see "Materials and Methods"). The ring-attractor network model was adjusted to support bump attractor dynamics with up to three simultaneous bumps (Edin et al., 2009), and further adjustment of the relative weights of AMPAR- and NMDAR-mediated currents was performed to set active reverberant neurons in the oscillatory regime (Wang, 1999). Using this computational model we started by investigating the dynamics that originated within each network.

In our model, multiple bumps showed anti-correlated oscillatory activity (**Figure 1**). As we stored more bumps in the network, lateral inhibition originating from simultaneous memories established anti-phase oscillatory dynamics during the memory period. These oscillatory dynamics were irregular, as illustrated in quickly dampened correlation functions (**Figure 1A**,

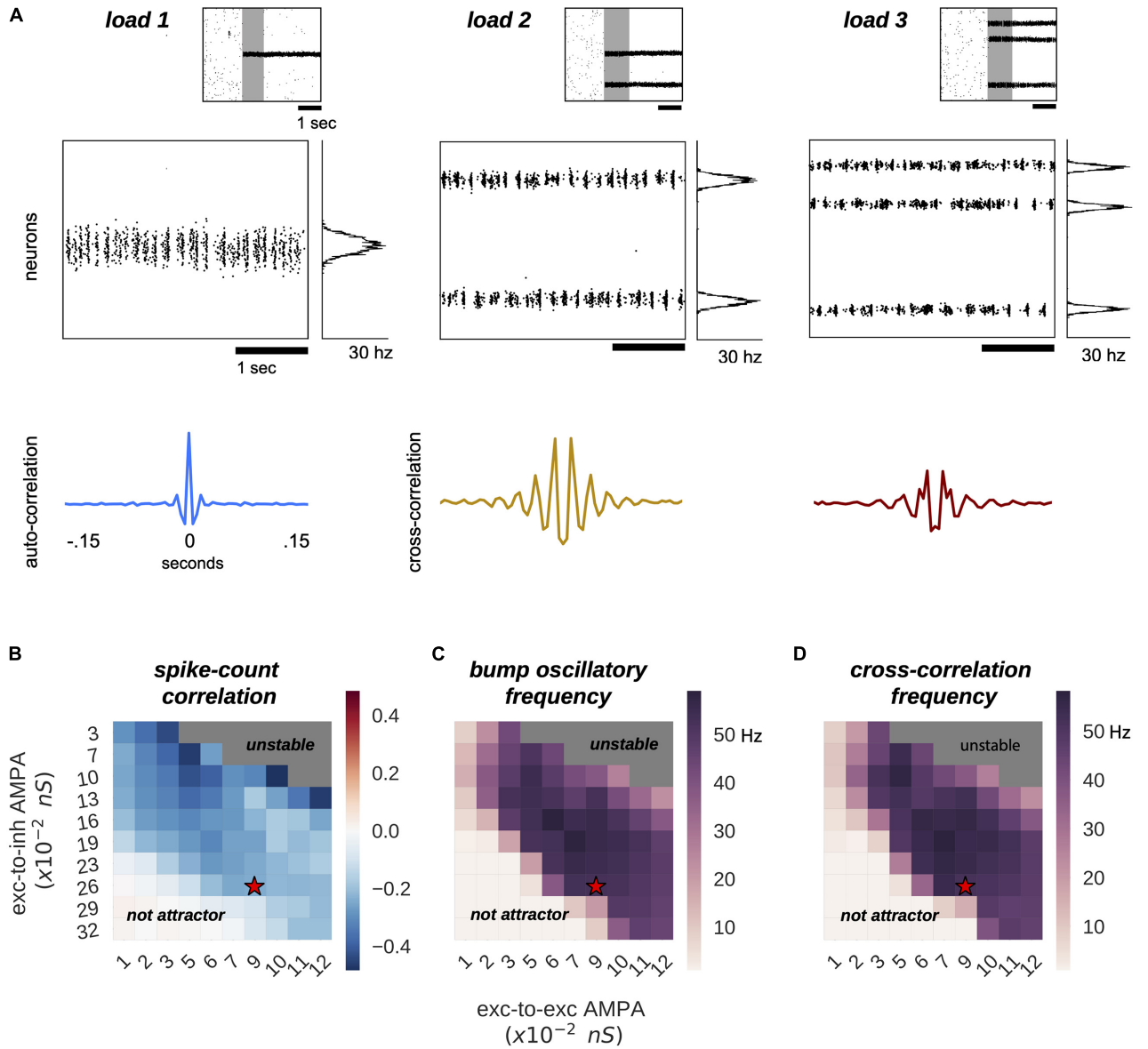


FIGURE 1 | *Multiple bumps are spontaneously anti-correlated.* **(A)** Raster plots of three sample simulations of load 1–3 (*top*) and delay-period zooms (*middle*) show clear bump oscillatory activity, confirmed by correlation functions (*bottom*). Notably, irregular activity coexists with oscillatory dynamics. **(B–D)** Anti-correlated oscillatory dynamics as excitation is manipulated in the network (AMPAR conductance for excitatory to inhibitory connections, Y-axis, and for excitatory to excitatory connections, X-axis) in simulations with load 2 for the capacity-2 network used in **Figures 3**, **4** (to facilitate comparisons). **(B)** Anti-phase dynamics as measured by zero-lag cross-correlation between bumps. **(C)** Dominant frequency of the auto-correlation function computed independently for each bump (computed with Fourier analysis). **(D)** Dominant frequency of the cross-correlation between the two bumps. Red stars mark the parameter values of model simulations used throughout the study. Plots in panels **(B–D)** summarize the dynamics of ∼10,000 simulations (total) of 100 different networks.

bottom). Moreover, we found that the anti-phase behavior was robust in a wide range of values for AMPAR conductances (**Figure 1B**), consistently in the gamma range of frequencies (**Figures 1C,D**). Having seen these anti-phase dynamics between simultaneous bumps, we sought to contrast two opposite scenarios as we increased the number of stored memories (*memory load*). Under one alternative, bumps may oscillate at a fixed frequency irrespectively of load, so that the global network oscillation (adding up the activity of fixed-frequency out-of-phase bumps) would have a frequency that should increase linearly with memory load (scenario 1, dashed line **Figure 2C**). Alternatively, the network global oscillation could have a fixed frequency for different loads, and simultaneous bumps would take turns to fire in the available active periods. This would lead to halving each bump's oscillation frequency as we double the memory load (scenario 2, dashed line in **Figure 2D**). We tested our model simulations to identify if our biophysical model adhered to one of these scenarios. To

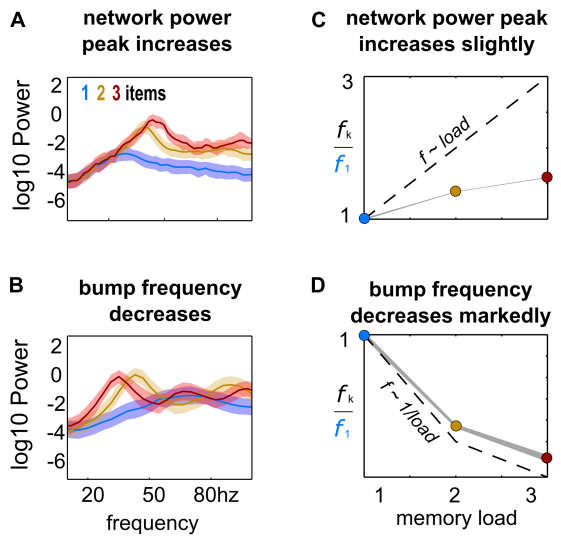


FIGURE 2 | *Load-modulation of network and bump oscillatory dynamics.* Power spectrum of LFPs computed from simulations with increasing load (1–3), using the activity of the whole network **(A)** or of each bump's activity **(B)**. **(C,D)** Peak-frequency f_k computed from simulations with increasing load k, normalized to frequency f_1 from simulations with a single bump and computed from LFPs of the whole network activity **(C)** and from LFPs of each individual bump's activity **(D)**.

this end, we ran multiple simulations with three different loads (presenting 1–3 separate bumps during the encoding cue period) and we computed power spectra from either the aggregate activity of the whole network (network power) or from separate populations centered around each presented target (bump power). We then extracted the frequency of the peak network and bump power to study their dependency with load. We found signatures of both scenarios (**Figures 2A,B**). As we increased the memory load, the overall network activity oscillated at slightly increasing frequencies (**Figures 2A,C**). In contrast, each bump, corresponding to different memories, oscillated at markedly slower frequencies as load increased (**Figures 2B,D**). We quantified which were the dominant dynamics by plotting both the network's and each bump's oscillating frequency against memory load. For better comparison, we normalized the frequency associated with different loads to the one of load 1. Moreover, we compared the effect of memory load against scenario 1 and 2 (dashed lines in **Figures 2C,D**). Qualitatively, we found that our network dynamics was more consistent with the latter.

We therefore conclude that our biophysical network maintains a relatively constant global oscillation as more items are loaded into memory, and individual memory oscillations instead start skipping cycles to sustain out-of-phase dynamics with other memories. Thus, the interplay between recurrent (fast) excitation and (slower) feedback inhibition acting locally is the basis of the oscillatory bump behavior. Moreover, we now show that anti-phase dynamics of simultaneous bumps occurs due to bump competition, accomplished by lateral inhibition. This competition increases with memory load, leading to longer periods of silence during the delay-activity of each bump. These dynamics generalize previous findings in simplified rate models (Pina et al., 2018), and extend them to biologically realistic ring attractor networks.

Uniform Coupling Achieves Feature Binding

The binding between color and location is accomplished through the spontaneous synchronization of pairs of bumps across two networks connected via weak cortico-cortical excitation (**Figure 3**). In particular, we connected two ring-attractors in the regime described above with all-to-all, untuned excitatory connectivity. This connectivity was weak and it was mediated exclusively by AMPARs (**Figure 3A**), acting on all excitatory and inhibitory neurons. Interestingly, anti-phase dynamics within each network (as described above) was maintained robustly for a wide range of connectivity strength values (**Figures 3E,F**). Across networks, each bump's activity was in phase with one bump in the other network (**Figures 3B,C**, black) but out of phase with the other (**Figures 3B,C**, red). On the majority of the simulations, this selective synchronization was maintained through the whole delay period (see **Figures 3C,D** for an example simulation). This set of dynamics is an interesting possible mechanism that binds and maintains the information of each presented stimulus. To this end, however, there are several aspects to resolve in relation to the encoding and decoding of this bound information.

On the one hand, synchronization selection was noise-induced in our simulations, resulting in across-networks associations between random pairs of bumps for different simulations. To control this association at the time of stimulus encoding, we stimulated strongly (7.5 times the intensity of sensory stimuli) and simultaneously one bump in each network for a brief period of 50 ms (**Figures 3B**, **4A**, green period), forcing these two bumps (one in each network) to engage in correlated activity during the delay period. Nevertheless, this phase-locked dynamics could be broken by noisy fluctuations, leading to possible misbinding of memorized features and swap trials (**Figures 4A,B**).

On the other hand, our model raised the question of how this binding of information could reasonably be decoded without resorting to complex mechanisms for spike coincidence detection. In our task, the "behavioral" output consisted in answering which "color" was initially associated with a particular "location," and this was accomplished by evaluating which bump of the color network maintained in-phase synchronization with the bump of the probed location at the end of the delay. We found that this did not require complex coincidence detection, but could instead be simulated in a rate formalism as follows. For each trial, we probed one *location* by stimulating weakly ($\frac{1}{4}$ of stimulus intensity) corresponding neurons in the *location network* at the end of the delay. This simulated the visual presentation of a location probe at the end of the delay. This increased the firing rate of the corresponding location bump, and we found that it also resulted in an increase of activity of the associated, in-phase *color bump*. Finally, we

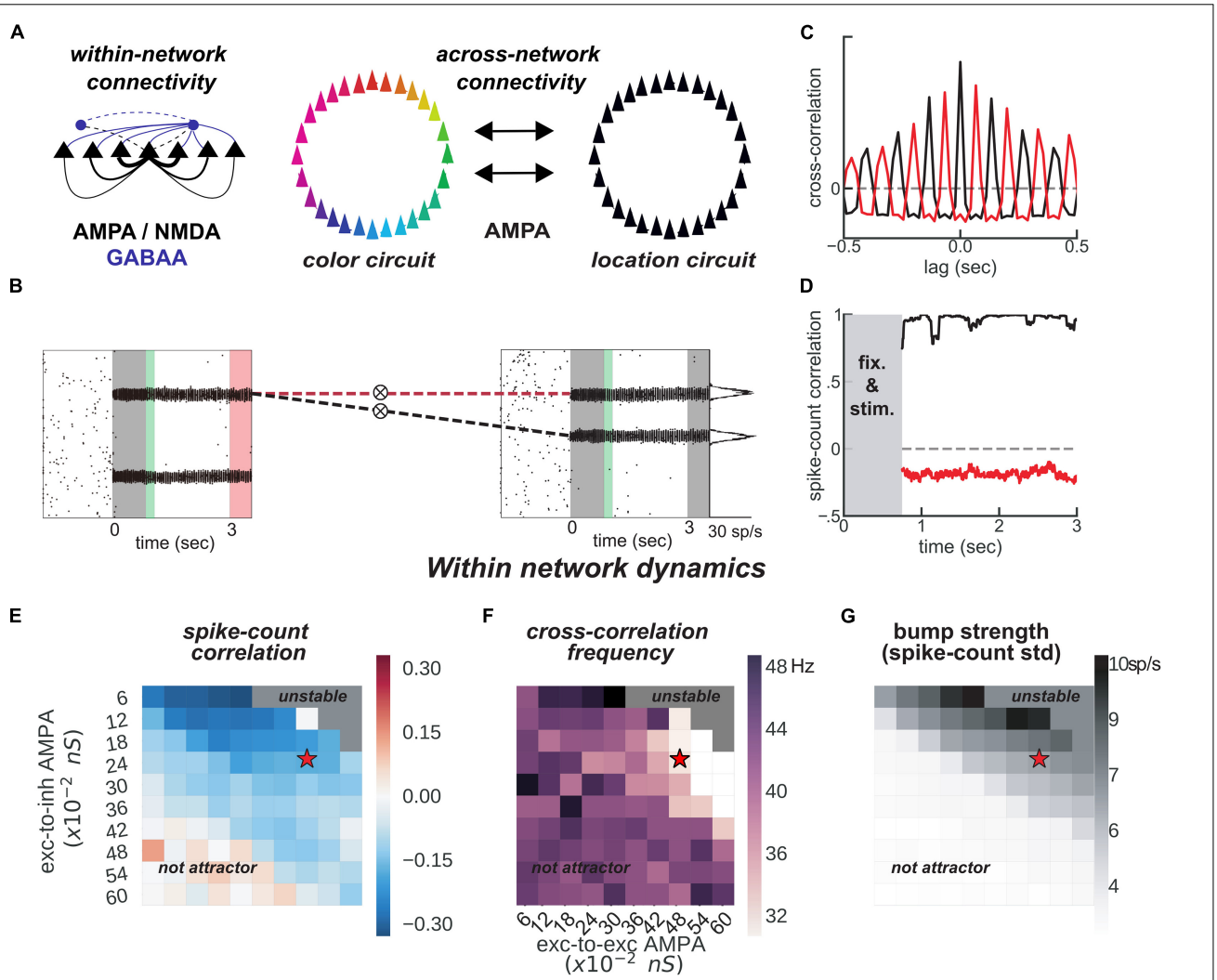


FIGURE 3 | *Feature-binding through weak, uniform coupling of 2 ring attractors.* **(A)** Left, schematics of within-network excitatory (black) and inhibitory (blue) connectivity to excitatory (solid) and inhibitory (dashed) neurons. Neurons with similar selectivity were strongly connected as illustrated by the line width. Right, schematics of the 2-network architecture, consisting of 2 ring-attractors with all-to-all, uniform connectivity. Each ring is able to store memories from one feature space (e.g., color or location) as activity bumps (**Figure 1**). **(B)** One example simulation for the two networks. The pink-shaded area marks the period in which we read out the activity of the entire color network, while injecting current at one specific location in the location network (right gray-shaded area in the location rastergram, see main text for details about encoding/decoding). **(C)** Cross-correlation computed between 2 pairs of bumps across networks [as marked with dashed red and black lines in panel **(B)**]. Across networks, oscillating bumps synchronize in phase (black, positive zero-lag cross-correlation) or out of phase (red, negative zero-lag cross-correlation). **(D)** Spike count correlation (in count bins of 5 ms and correlation windows of 100 ms) of both associations through the memory delay is stable for this simulation. **(E,F)** Similar to **Figures 1B,D**, but manipulating AMPAR conductances across networks. **(E)** Robustness of anti-phase dynamics within each network as measured by spike count correlation between bumps (**Figure 1B**). **(F)** Dominant frequency of cross-correlation between the two bumps within each network (**Figure 1D**). **(G)** Bump strength measured as standard deviation of spike-counts across model neurons at the end of the delay. **(E–G)** summarize the dynamics of 22,000 simulations (total) of 100 × 2 networks. Stars indicate parameters and dynamical regime of network simulations shown in panels **(B–D)**.

extracted the behavioral output by fitting a mixture of gaussians ("Materials and Methods") applied to the mean firing rate activity across the *color network* during the location-probing period (0.5 s). **Figure 4B** shows color readouts from 1,000 of such simulated trials. Applying our encoding/decoding method to our simulations, resulted in 30% of trials wrongly associated with the non-target color (swap trials, **Figure 4B**). We then separated *swap* trials from *on-target* trials and computed the spike-count correlation in windows of 5 ms through the whole trial period (**Figure 4D**), and confirmed that on-target trials were in fact characterized by stable phase-locked activity, while the correlation between bumps in swap trials progressively approached the opposite dynamics (in-phase/anti-phase for the bound/unbound items, **Figure 4D**). Importantly, networks maintained synchronized in-phase dynamics for bound features robustly over a broad range of inter-network connectivity parameter values (**Figures 4E,F**). Additionally, we identified three sources of swap errors in our simulations, classified as

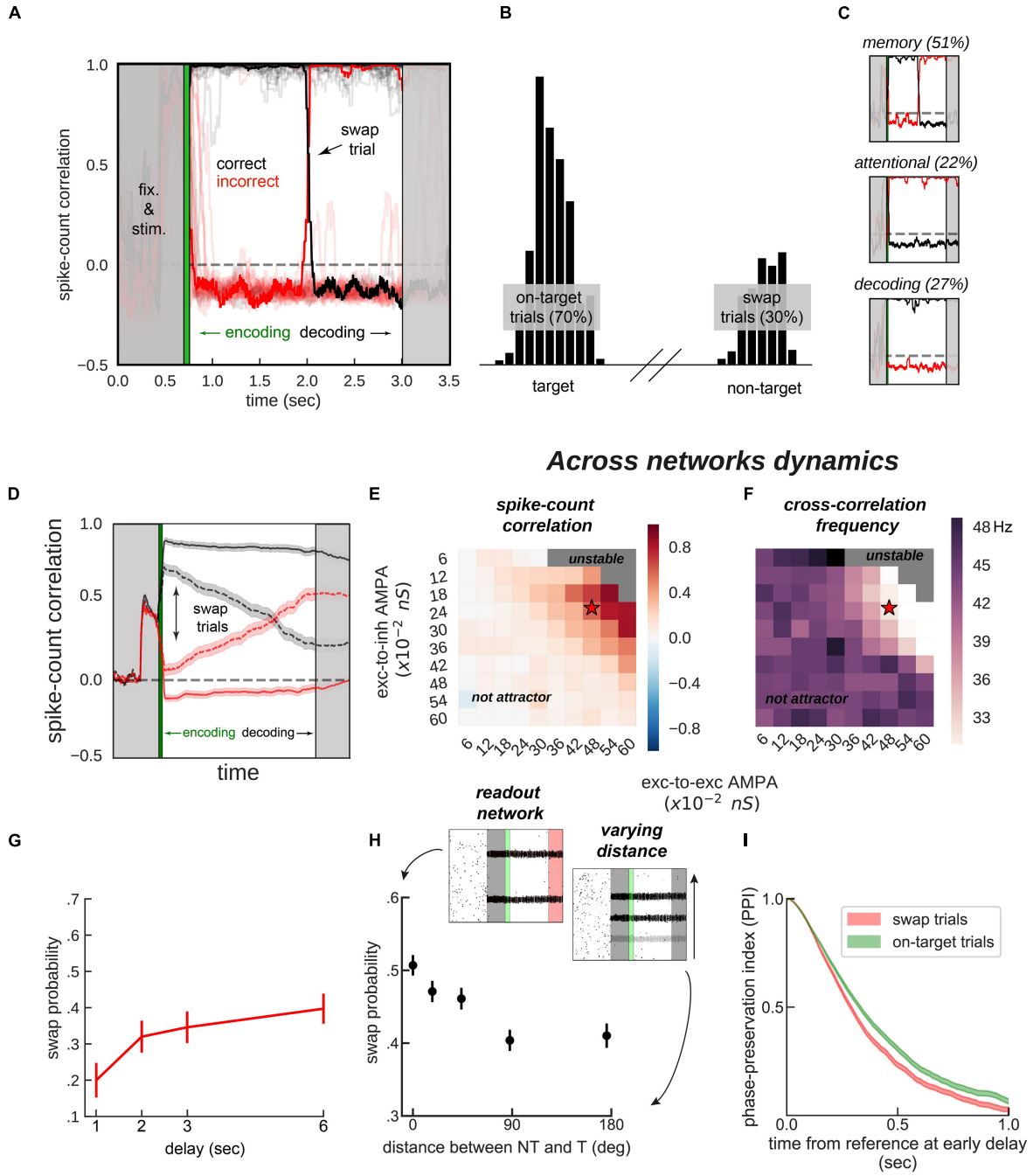


FIGURE 4 | *Encoding and decoding without temporal precision.* **(A)** Spike-count correlation (in count bins of 5 ms and correlation windows of 100 ms) during the delay for 20 sample simulations. During the encoding period (green), immediately after the stimulus presentation, we bound two bumps, one from each network, by simultaneously stimulating them strongly. This ensured those two bumps were correlated through the trial more often than chance (black lines in the figures), and the other cross-network association synchronized mostly out-of-phase (red lines). On some trials (only one in a), noisy fluctuations reversed these correlations suddenly (swap trials). During the decoding period (light gray, on the right) we simulated the probe period of a working memory task, by stimulating the cued location (0.5 s) of one network, while decoding mean firing rates from the color network. **(B)** Color readout histogram in 1,000 simulations. Bumps bound during encoding (target, centered at neuron 520 out of 2,048) were more likely to be read-out than unbound bumps (non-target, centered at 1,480). **(C)** Three types of swaps: *memory swaps* (top), *attentional swaps* (middle), or *decoding swaps* (bottom). **(D)** Same as panel **(A)**, averaging across all trials separately for swap and on-target trials, as defined by the decoder, shown in panel **(B)**. **(E,F)** summary of the dynamics of 22,000 simulations (total) of 100 connected (×2) networks as a function of inter-network connectivity. **(E)** Binding stability measured as the average spike-count correlation between initially bound bump pairs during the delay (black, in figures). **(F)** Dominant frequency of the cross-correlation between bound bump pairs. Red stars mark the parameter values of the model used for sample simulations. **(G)** Swap errors increase with delay duration and **(H)** simulations (3 s delay) where target (T) and non-target (NT) bumps are stored close-by (varying

(Continued)

FIGURE 4 | (Continued)
distance) increase swap error probability, relative to when they are further apart. **(I)** Swap-error trials (red, n = 3,000), compared with on-target trials (green, n = 3,000) in the model are associated with a lower phase consistency of oscillatory activity in the delay period, as measured with phase-preservation index (PPI, "Materials and Methods") using early delay as the reference time point. Error-bars are bootstrapped standard errors (n = 500).

memory swaps if the correct association based on in-phase bump synchronization changed abruptly during the delay (51% of the swap trials), *attentional swaps* if the wrong association was encoded during the encoding period (22%) or *decoding swaps* if the correct association was encoded and maintained during the memory period, but the decoding failed (27%). See **Figure 4C** for example simulations.

Together, our biologically-constrained simulations demonstrate that feature-binding can be robustly accomplished through selective synchronization. Crucially, encoding/decoding location-color associations was done without a *temporally precise code*, a long-standing limitation in the *binding by synchrony* framework (Shadlen and Movshon, 1999). Moreover, we identified three sources of swap errors. Based on these computational findings, we investigated model predictions that could be compared with existing data or could generate hypotheses for new experimental studies.

Swap Errors Increase With Delay and Item Competition

In our model, swap errors are induced by noisy fluctuations. This results in two behavioral predictions, congruent with previous findings (Emrich and Ferber, 2012; Pertzov et al., 2017; Schneegans and Bays, 2017). First, longer memory delays should increase the probability of a noisy fluctuation that is sufficiently large to induce a swap (**Figure 4G**). Second, **Figure 4H** shows how swap errors decrease with target to non-target distances. For very close locations, feedback inhibition is strongest, leading to strong competition between nearby bumps, explaining an increase of swap errors for such distances. This is similar to previous studies (Wei et al., 2012; Almeida et al., 2015; Nassar et al., 2018), in which simultaneous bumps interfere (repulsively and through their phase relationship, which is in this case less stable through the delay). Experimentally, these two regimes correspond to different scenarios. In the first case, one color is forgotten, while in the second scenario, there is an actual *swap* error. This prediction could be tested experimentally by probing the subject's memory on all items, instead of just one (Adam et al., 2017).

In sum, our model is able to describe a previously found dependence of swap errors with delay duration and with target to non-target distance, and it offers mechanistic explanations for such dependencies.

Neural Prediction: Swap Trials Show Less Phase Preservation Through the Delay

Finally, abrupt changes in the phase relationship between oscillating bumps is the central mechanism of swap errors in our model (**Figures 4A,B**). Therefore, it is worth deriving a testable neurophysiological prediction from this mechanism. Additionally, because these changes in phase relationships are abrupt, they require experiments using techniques with high temporal resolution such as MEG or EEG. Intuitively, swap errors in our model simulations are characterized by inconsistent phase relationships between brain signals when comparing the beginning and the end of the delay period. We therefore considered applying an analysis that has been proposed to test phase consistency in EEG/MEG: the phase-preservation index (PPI, Mazaheri and Jensen, 2006). We first derived LFP signals from our network's spiking activity ("Materials and Methods"). We then calculated the phase-preservation index (PPI, see Mazaheri and Jensen, 2006 and "Materials and Methods") at the end of the delay, relative to the beginning of the delay, and separately for on-target and swap trials defined "behaviorally" (**Figure 4B**). As we expected based on our model simulations (**Figure 4**), this analysis applied to our simulated data showed that trials containing swap errors had a lower PPI, compared to on-target trials (**Figure 4I**). This prediction can be tested with MEG/EEG data recorded from humans performing this task, based on an analysis of behavioral responses able to discriminate swap and correct error trials (Bays et al., 2009).

DISCUSSION

Aiming to account for swap-errors, a prominent source of multi-item working memory interference (Schneegans and Bays, 2019), we extended the ring-attractor model (Compte et al., 2000). Our biologically-constrained model offers a plausible mechanism for feature-binding. Briefly, the encoding and decoding of associations is accomplished through rate-coding, while their maintenance is accomplished through selective synchronization of oscillatory mnemonic activity. Oscillatory dynamics emerges naturally from bump competition, which increases with memory load and is in line with previous EEG experiments in humans (Roux et al., 2012) and LFP recordings from monkey PFC (Lundqvist et al., 2018). Finally, our model reveals different origins of swap errors (Mitchell et al., 2018; Pratte, 2019), how they depend on delay duration and inter-item distances (Emrich and Ferber, 2012; Pertzov et al., 2017; Schneegans and Bays, 2017), and predicts that phase-locked oscillatory activity during the memory periods should reflect swap errors.

Other Multi-Area Models for Working Memory

Our multi-area model adds to a large body of computational work (Ardid et al., 2007, 2010; Edin et al., 2009; Engel and Wang, 2011; Murray et al., 2017; Bouchacourt and Buschman, 2019; Mejias and Wang, 2019; Froudist-Walsh et al., 2020; Min et al., 2020; Novikov et al., 2021) attempting to account for

the distributed nature of working memory (Christophel et al., 2017). While several of these models have implemented across-area interactions through oscillatory dynamics (Ardid et al., 2010; Novikov et al., 2021), they did not attribute a clear mechanistic role to inter-area synchronization dynamics. This is in contrast to our model, where feature-binding in working memory is accomplished through selective synchronization of oscillatory activity in different brain areas.

Comparison With Previous Binding Models

Previously proposed models by Pina et al. (2018) and Raffone and Wolters (2001) as well as our model are explicit implementations of the synchronization mechanism for feature binding in working memory. While similar in the approach, there are important differences. As argued by Schneegans and Bays (2019), a major difficulty with previous synchronization models was that they were unable to show their capacity of reproducing the rich phenomenology of working memory behavior that other models can explain. Our model, on the basis of its architecture with ring attractor models of spiking neural networks, overcomes the limitation of earlier discrete population models (Raffone and Wolters, 2001; Pina et al., 2018) and keeps all the demonstrated explanatory power that is characteristic of these attractor models, such as explaining several behavioral working memory biases in humans (Almeida et al., 2015; Kiyonaga et al., 2017; Barbosa and Compte, 2018; Kilpatrick, 2018; Nassar et al., 2018; Stein et al., 2020) and monkeys (Papadimitriou et al., 2015; Barbosa et al., 2020); as well as explaining key neurophysiological dynamics during working memory maintenance periods (see Barbosa, 2017 for a short review) in humans (Edin et al., 2009; Kamiński et al., 2017) and monkeys (Wimmer et al., 2014; Sajad et al., 2016). Our model also goes beyond previous synchronization models in that (1) by virtue of its 2-ring architecture, it explicitly implements the storage of different features in independent systems or brain areas, as shown experimentally (Schneegans and Bays, 2019), and that (2) it provides a plausible rate-based readout mechanism of working memory associations without resorting to complex synchrony detection processes, a major difficulty for this sort of models (Shadlen and Movshon, 1999). Indeed, we show that our proposed mechanisms is robust to the noise inherent in spiking networks, which together with the need of precise spike coincidence detectors were major concerns of the binding through synchronized activity hypothesis in general (Shadlen and Movshon, 1999) and previous implementations in particular (Raffone and Wolters, 2001; Pina et al., 2018).

Thus, our model now brings back synchronization-based feature binding in working memory as a plausible alternative to recent conjunction binding proposals, such as the *binding pool* (Swan and Wyble, 2014) and the *conjunctive coding* model (Matthey et al., 2015; Schneegans and Bays, 2017). These models implement binding mechanisms that are fundamentally different from ours. In these models, binding of separated features is accomplished through conjunction neurons, which are neurons selective to mixtures of those features. While there is evidence for such neurons in the cortex (Rigotti et al., 2013; Fusi et al., 2016), their role in feature-binding is not clear, given the consistent evidence for separate feature storage underlying working memory binding (Olson and Jiang, 2002; Wheeler and Treisman, 2002; Xu, 2002; Delvenne and Bruyer, 2004; Bays et al., 2011b; Fougnie and Alvarez, 2011; Parra et al., 2011). Importantly, such a mechanism scales exponentially with the number of feature combinations, thus seemingly inconsistent with our ability to flexibly bind never seen combinations (Schneegans and Bays, 2019). However, it is to be noted that some conjunction models have mitigated this scaling problem through the construction of random conjunctions in an interposed network (Swan and Wyble, 2014; Bouchacourt and Buschman, 2019).

Encoding With Rate Code

In our hybrid model, only the maintenance of associations is accomplished through correlated oscillatory activity or, in other words, relies on a *temporal code*. Instead, encoding and decoding of associations is achieved through a *rate code*. Encoding and decoding is accomplished by delivering flat pulses (i.e., without the need to be temporally precise) to both the to-be-bound features exclusively (*encoding*) or just to one of them (*decoding*).

Encoding the association between two different features through a pulse delivered simultaneously to each corresponding bump resembles the sequential encoding hypothesis in working memory (Wolfe, 1994; Bays et al., 2011a). Moreover, there is evidence that a mechanism combining sequential and parallel encoding is implemented in the brain when solving multi-item working memory tasks (Bays et al., 2011a). Our model implements such a combination. First, information about independent features arrives simultaneously to memory-encoding areas from upstream sensory areas. Then, the correct associations are sequentially encoded by brief excitatory pulses, possibly as a result of overt selective attention to each stimulus sequentially (Schoenfeld et al., 2014). Speaking to this, humans take longer to encode combined features than they take to encode the same amount of independent features (Schneegans and Bays, 2019).

Decoding With Rate Code

Works modelling multi-item working memory though the storage of several bumps in a network (Wei et al., 2012; Krishnan et al., 2018; Nassar et al., 2018) including our own (Edin et al., 2009; Almeida et al., 2015) often used approaches that are biologically implausible to extract the location of one bump, while ignoring other simultaneously maintained bumps. Our approach, however, matches closely the "cueing" period of a multi-item working memory task, which consists of stimulating the "cued" locations while reading out from the whole color network population. Moreover, our encoding/decoding mechanism proposes that swap errors can be of different origins (attention, memory, or decoding; **Figure 4C**). Indeed, experimental designs that require subjects to rate their confidence on a trial-by-trial basis show that swap errors occur both in high- and low-confidence trials, suggesting different origins (Mitchell et al., 2018; Pratte, 2019).

Future Work: Toward Biological Plausibility of Binding Through Dynamics

We found anti-phase dynamics within each network and phase-locking across networks, the central mechanisms for feature-binding in our model, to occur naturally in a broad range of parameters, indicating that the mechanisms proposed here do not require fine-tuning. Because our model is to some degree biologically constrained, it is a proof of concept that working memory binding through synchronized activity is *at least* possible to occur in the brain. In fact, we simulated noisy integrate-and-fire neurons, supporting that the central mechanism implemented in our model has some degree of robustness to noise.

Our model is, however, limited in several ways that could be addressed in future studies. First, we did not simulate trials demanding binding of load 3 or higher. We expect that the main challenges associated with that improvement will be the encoding of more associations. We also did not explore conditions with asymmetric number of bumps (e.g., two colors/locations at/with one location/color), as this would lead to different experimental paradigms. Second, we did not investigate how feature-binding is impacted by incoming distractors. Previous work has shown that oscillatory activity on different bands can play a role in filtering distractors (Dipoppa and Gutkin, 2013). Future work combining these models is necessary. Third, as a proof of concept, we only simulated two connected networks, while humans can encode and decode the association of many more features (Schneegans and Bays, 2019). Relatedly, our two-dimensional network architecture should be taken as a proof of concept, rather than being a literal anatomical representation of a specific brain structure. Finally, the oscillatory regime in which our model is operating, in which neurons are strongly synchronized with the population rhythm (**Figure 4C**), however, derived from biologically constrained neuronal models, is arguably not biological itself. While there is abundant evidence that neuronal populations show strong oscillatory dynamics in working memory (e.g., Pesaran et al., 2002), single neuron dynamics approaches a Poisson process (Softky and Koch, 1993; Compte et al., 2003) therefore not oscillatory at this scale (but see Lundqvist et al., 2016). Early theoretical work (Brunel and Hakim, 1999; Brunel, 2000; Brunel and Wang, 2003) has demonstrated that such oscillatory dynamics at the population level can coexist with noisy, unsynchronized neurons when randomly connected. Future work that connects randomly connected networks that store multiple stable bump-attractors (Hansel and Mato, 2013), but operating in anti-correlated oscillatory activity such as in our simulations could be an appropriate avenue for the future work attempting to overcome these limitations.

DATA AVAILABILITY STATEMENT

The multi-area model as well as the code for measuring the phase-preservation index is available at https://github.com/comptelab/binding.

AUTHOR CONTRIBUTIONS

JB carried out the research. JB and AC conceived the research and wrote the manuscript. KKS, VB, and AT conceived the electrophysiological prediction. All authors edited and approved the final version of the manuscript.

FUNDING

This work was funded by the Spanish Ministry of Science, Innovation and Universities and European Regional Development Fund (Refs: BFU2015-65315-R and RTI2018-094190-B-I00); by the Institute Carlos III, Spain (Grants PIE 16/00014 and AC20/00071); by the Cellex Foundation; by the Generalitat de Catalunya (AGAUR 2014SGR1265 and 2017SGR01565); and by the CERCA Programme/Generalitat de Catalunya. JB was supported by the Spanish Ministry of Economy and Competitiveness (FPI program) and by the Bial Foundation (Ref: 356/18). This work was developed at the building Centro Esther Koplowitz, Barcelona.

REFERENCES

Adam, K. C. S., Vogel, E. K., and Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cogn. Psychol.* 97, 79–97. doi: 10.1016/j.cogpsych.2017.07.001

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

Ardid, S., Wang, X.-J., and Compte, A. (2007). An integrated microcircuit model of attentional processing in the neocortex. *J. Neurosci.* 27, 8486–8495. doi: 10.1523/JNEUROSCI.1145-07.2007

Ardid, S., Wang, X.-J., Gomez-Cabrero, D., and Compte, A. (2010). Reconciling coherent oscillation with modulation of irregular spiking activity in selective attention: gamma-range synchronization between sensory and executive cortical areas. *J. Neurosci.* 30, 2856–2870. doi: 10.1523/JNEUROSCI.4222-09.2010

Barbosa, J. (2017). Working memories are maintained in a stable code. *J. Neurosci.* 37, 8309–8311. doi: 10.1523/JNEUROSCI.1547-17.2017

Barbosa, J., and Compte, A. (2018). Build-up of serial dependence in color working memory. *BioRxiv* [Preprint] doi: 10.1101/503185

Barbosa, J., Stein, H., Martinez, R. L., Galan-Gadea, A., Li, S., Dalmau, 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, 1016–1024. doi: 10.1038/s41593-020-0644-4

Bays, P. M., Catalao, R. F. G., and Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *J. Vis.* 9, 7.1–11. doi: 10.1167/9.10.7

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., and Husain, M. (2011a). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *J. Vis.* 11, 6.1–15. doi: 10.1167/11.10.6

Bays, P. M., Wu, E. Y., and Husain, M. (2011b). Storage and binding of object features in visual working memory. *Neuropsychologia* 49, 1622–1631. doi: 10.1016/j.neuropsychologia.2010.12.023

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

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.* 8, 183–208. doi: 10.1023/a:1008925309027

Brunel, N., and Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput.* 11, 1621–1671. doi: 10.1162/089976699300016179

Brunel, N., and Wang, X.-J. (2003). What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance. *J. Neurophysiol.* 90, 415–430. doi: 10.1152/jn.01095.2002

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., and Haynes, J.-D. (2017). The distributed nature of working memory. *Trends Cogn. Sci. (Regul. Ed.)* 21, 111–124. doi: 10.1016/j.tics.2016.12.007

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

Compte, A., Constantinidis, C., Tegner, J., Raghavachari, S., Chafee, M. V., Goldman-Rakic, P. S., et al. (2003). Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *J. Neurophysiol.* 90, 3441–3454. doi: 10.1152/jn.00949.2002

Delvenne, J., and Bruyer, R. (2004). Does visual short-term memory store bound features? *Vis. Cogn.* 11, 1–27. doi: 10.1080/13506280344000167

Dipoppa, M., and Gutkin, B. S. (2013). Flexible frequency control of cortical oscillations enables computations required for working memory. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12828–12833. doi: 10.1073/pnas.1303270110

Edin, F., Klingberg, T., Johansson, P., McNab, F., Tegnér, J., and Compte, A. (2009). Mechanism for top-down control of working memory capacity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6802–6807. doi: 10.1073/pnas.0901894106

Emrich, S. M., and Ferber, S. (2012). Competition increases binding errors in visual working memory. *J. Vis.* 12:12. doi: 10.1167/12.4.12

Engel, T. A., and Wang, X.-J. (2011). Same or different? A neural circuit mechanism of similarity-based pattern match decision making. *J. Neurosci.* 31, 6982–6996. doi: 10.1523/JNEUROSCI.6150-10.2011

Fell, J., and Axmacher, N. (2011). The role of phase synchronization in memory processes. *Nat. Rev. Neurosci.* 12, 105–118. doi: 10.1038/nrn2979

Fougnie, D., and Alvarez, G. A. (2011). Object features fail independently in visual working memory: evidence for a probabilistic feature-store model. *J. Vis.* 11:3. doi: 10.1167/11.12.3

Froudist-Walsh, S., Bliss, D. P., Ding, X., Jankovic-Rapan, L., Niu, M., Knoblauch, K., et al. (2020). A dopamine gradient controls access to distributed working memory in monkey cortex. *BioRxiv* [Preprint] doi: 10.1101/2020.09.07.286500

Fusi, S., Miller, E. K., and Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* 37, 66–74. doi: 10.1016/j.conb.2016.01.010

Hansel, D., and Mato, G. (2013). Short-term plasticity explains irregular persistent activity in working memory tasks. *J. Neurosci.* 33, 133–149. doi: 10.1523/JNEUROSCI.3455-12.2013

Howard, M. W., Rizzuto, D. S., Caplan, J. B., Madsen, J. R., Lisman, J., Aschenbrenner-Scheibe, R., et al. (2003). Gamma oscillations correlate with working memory load in humans. *Cereb. Cortex* 13, 1369–1374. doi: 10.1093/cercor/bhg084

Kaiser, J., Ripper, B., Birbaumer, N., and Lutzenberger, W. (2003). Dynamics of gamma-band activity in human magnetoencephalogram during auditory pattern working memory. *Neuroimage* 20, 816–827. doi: 10.1016/S1053-8119(03)00350-1

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

Kilpatrick, Z. P. (2018). Synaptic mechanisms of interference in working memory. *Sci. Rep.* 8:7879. doi: 10.1038/s41598-018-25958-9

Kiyonaga, A., Scimeca, J. M., Bliss, D. P., and Whitney, D. (2017). Serial dependence across perception, attention, and memory. *Trends Cogn. Sci. (Regul. Ed.)* 21, 493–497. doi: 10.1016/j.tics.2017.04.011

Kornblith, S., Buschman, T. J., and Miller, E. K. (2016). Stimulus load and oscillatory activity in higher cortex. *Cereb. Cortex* 26, 3772–3784. doi: 10.1093/cercor/bhv182

Krishnan, N., Poll, D. B., and Kilpatrick, Z. P. (2018). Synaptic efficacy shapes resource limitations in working memory. *J. Comput. Neurosci.* 44, 273–295. doi: 10.1007/s10827-018-0679-7

Lundqvist, M., Herman, P., Warden, M. R., Brincat, S. L., and Miller, E. K. (2018). Gamma and beta bursts during working memory readout suggest roles in its volitional control. *Nat. Commun.* 9:394. doi: 10.1038/s41467-017-02791-8

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

Lutzenberger, W., Ripper, B., Busse, L., Birbaumer, N., and Kaiser, J. (2002). Dynamics of gamma-band activity during an audiospatial working memory task in humans. *J. Neurosci.* 22, 5630–5638.

Ma, W. J., Husain, M., and Bays, P. M. (2014). Changing concepts of working memory. *Nat. Neurosci.* 17, 347–356. doi: 10.1038/nn.3655

Matthey, L., Bays, P. M., and Dayan, P. (2015). A probabilistic palimpsest model of visual short-term memory. *PLoS Comput. Biol.* 11:e1004003. doi: 10.1371/journal.pcbi.1004003

Mazaheri, A., and Jensen, O. (2006). Posterior alpha activity is not phase-reset by visual stimuli. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2948–2952. doi: 10.1073/pnas.0505785103

Mejias, J. F., and Wang, X.-J. (2019). Mechanisms of distributed working memory in a large-scale model of the macaque neocortex. *BioRxiv* [Preprint]. doi: 10.1101/760231

Min, B., Bliss, D. P., Sarma, A., Freedman, D. J., and Wang, X.-J. (2020). A neural circuit mechanism of categorical perception: top-down signaling in the primate cortex. *BioRxiv* [Preprint]. doi: 10.1101/2020.06.15.151506

Mitchell, D. J., Cusack, R., and Cam-Can. (2018). Visual short-term memory through the lifespan: Preserved benefits of context and metacognition. *Psychol. Aging* 33, 841–854. doi: 10.1037/pag0000265

Murray, J. D., Jaramillo, J., and Wang, X.-J. (2017). Working Memory and Decision-Making in a Frontoparietal Circuit Model. *J. Neurosci.* 37, 12167–12186. doi: 10.1523/JNEUROSCI.0343-17.2017

Nassar, M. R., Helmers, J. C., and Frank, M. J. (2018). Chunking as a rational strategy for lossy data compression in visual working memory. *Psychol. Rev.* 125, 486–511. doi: 10.1037/rev0000101

Novikov, N., Zakharov, D., Moiseeva, V., and Gutkin, B. (2021). Activity stabilization in a population model of working memory by sinusoidal and noisy inputs. *Front. Neural Circuits* 15:647944. doi: 10.3389/fncir.2021.647944

Olson, I. R., and Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Percept. Psychophys.* 64, 1055–1067. doi: 10.3758/bf03194756

Palva, S., Kulashekhar, S., Hämäläinen, M., and Palva, J. M. (2011). Localization of cortical phase and amplitude dynamics during visual working memory encoding and retention. *J. Neurosci.* 31, 5013–5025. doi: 10.1523/JNEUROSCI.5592-10.2011

Papadimitriou, C., Ferdoash, A., and Snyder, L. H. (2015). Ghosts in the machine: memory interference from the previous trial. *J. Neurophysiol.* 113, 567–577. doi: 10.1152/jn.00402.2014

Parra, M. A., Cubelli, R., and Della Sala, S. (2011). Lack of color integration in visual short-term memory binding. *Mem. Cognit.* 39, 1187–1197. doi: 10.3758/s13421-011-0107-y

Pertzov, Y., Manohar, S., and Husain, M. (2017). Rapid forgetting results from competition over time between items in visual working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 43, 528–536. doi: 10.1037/xlm0000328

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

Pina, J. E., Bodner, M., and Ermentrout, B. (2018). Oscillations in working memory and neural binding: A mechanism for multiple memories and their interactions. *PLoS Comput. Biol.* 14:e1006517. doi: 10.1371/journal.pcbi.1006517

Pratte, M. S. (2019). Swap errors in spatial working memory are guesses. *Psychon. Bull. Rev.* 26, 958–966. doi: 10.3758/s13423-018-1524-8

Qi, X.-L., Liu, R., Vazdarjanova, A. I., Blake, D. T., and Constantinidis, C. (2019). Nucleus basalis stimulation stabilizes attractor networks and enhances task representation in prefrontal cortex. *BioRxiv* [Preprint] doi: 10.1101/674465

Raffone, A., and Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *J. Cogn. Neurosci.* 13, 766–785. doi: 10.1162/08989290152541430

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

Roux, F., Wibral, M., Mohr, H. M., Singer, W., and Uhlhaas, P. J. (2012). Gamma-band activity in human prefrontal cortex codes for the number of relevant items maintained in working memory. *J. Neurosci.* 32, 12411–12420. doi: 10.1523/JNEUROSCI.0421-12.2012

Sajad, A., Sadeh, M., Yan, X., Wang, H., and Crawford, J. D. (2016). Transition from target to gaze coding in primate frontal eye field during memory delay and memory-motor transformation. *Eneuro* 3, e0040–16.2016 1–20. doi: 10.1523/ENEURO.0040-16.2016

Schneegans, S., and Bays, P. M. (2017). Neural architecture for feature binding in visual working memory. *J. Neurosci.* 37, 3913–3925. doi: 10.1523/JNEUROSCI.3493-16.2017

Schneegans, S., and Bays, P. M. (2019). New perspectives on binding in visual working memory. *Br. J. Psychol.* 110, 207–244. doi: 10.1111/bjop.12345

Schneegans, S., Spencer, J. P., and Schöner, G. (2016). "Integrating "what" and "where": Visual working memory for objects in a scene. Oxford series in developmental cognitive neuroscience," in *In Dynamic thinking: A primer on dynamic field theory*, eds G. Schöner, J. P. Spencer, and DFT Research Group (Oxford: Oxford University Press), 197–226.

Schoenfeld, M. A., Hopf, J.-M., Merkel, C., Heinze, H.-J., and Hillyard, S. A. (2014). Object-based attention involves the sequential activation of feature-specific cortical modules. *Nat. Neurosci.* 17, 619–624. doi: 10.1038/nn.3656

Shadlen, M. N., and Movshon, J. A. (1999). Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron* 24, 67–77, 111-125. doi: 10.1016/s0896-6273(00)80822-3

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

Siegel, M., Warden, M. R., and Miller, E. K. (2009). Phase-dependent neuronal coding of objects in short-term memory. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21341–21346. doi: 10.1073/pnas.0908193106

Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24, 49–65, 111–125. doi: 10.1016/s0896-6273(00)80821-1

Softky, W. R., and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci.* 13, 334–350. doi: 10.1523/JNEUROSCI.13-01-00334.1993

Stein, H., Barbosa, J., Rosa-Justicia, M., Prades, L., Morató, A., Galan-Gadea, A., et al. (2020). Reduced serial dependence suggests deficits in synaptic potentiation in anti-NMDAR encephalitis and schizophrenia. *Nat. Commun.* 11:4250. doi: 10.1038/s41467-020-18033-3

Swan, G., and Wyble, B. (2014). The binding pool: a model of shared neural resources for distinct items in visual working memory. *Atten. Percept. Psychophys.* 76, 2136–2157. doi: 10.3758/s13414-014-0633-3

Tseng, P., Chang, Y.-T., Chang, C.-F., Liang, W.-K., and Juan, C.-H. (2016). The critical role of phase difference in gamma oscillation within the temporoparietal network for binding visual working memory. *Sci. Rep.* 6:32138. doi: 10.1038/srep32138

van Vugt, M. K., Schulze-Bonhage, A., Litt, B., Brandt, A., and Kahana, M. J. (2010). Hippocampal gamma oscillations increase with memory load. *J. Neurosci.* 30, 2694–2699. doi: 10.1523/JNEUROSCI.0567-09.2010

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

Wei, Z., Wang, X.-J., and Wang, D.-H. (2012). From distributed resources to limited slots in multiple-item working memory: a spiking network model with normalization. *J. Neurosci.* 32, 11228–11240. doi: 10.1523/JNEUROSCI.0735-12.2012

Wheeler, M. E., and Treisman, A. M. (2002). Binding in short-term visual memory. *J. Exp. Psychol. Gen.* 131, 48–64. doi: 10.1037//0096-3445.131.1.48

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

Wimmer, K., Ramon, M., Pasternak, T., and Compte, A. (2016). Transitions between multiband oscillatory patterns characterize memory-guided perceptual decisions in prefrontal circuits. *J. Neurosci.* 36, 489–505. doi: 10.1523/JNEUROSCI.3678-15.2016

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychon. Bull. Rev.* 1, 202–238. doi: 10.3758/BF03200774

Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 458–468. doi: 10.1037//0096-1523.28.2.458

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neural Circuits

REVIEW
published: 26 November 2021
doi: 10.3389/fncir.2021.764177

Neural Substrates of Visual Perception and Working Memory: Two Sides of the Same Coin or Two Different Coins?

Megan Roussy[1], *Diego Mendoza-Halliday*[2] *and Julio C. Martinez-Trujillo*[1]*

[1] *Department of Physiology and Pharmacology, Schulich School of Medicine & Dentistry, Robarts Research Institute, University of Western Ontario, London, ON, Canada,* [2] *McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, United States*

Visual perception occurs when a set of physical signals emanating from the environment enter the visual system and the brain interprets such signals as a percept. Visual working memory occurs when the brain produces and maintains a mental representation of a percept while the physical signals corresponding to that percept are not available. Early studies in humans and non-human primates demonstrated that lesions of the prefrontal cortex impair performance during visual working memory tasks but not during perceptual tasks. These studies attributed a fundamental role in working memory and a lesser role in visual perception to the prefrontal cortex. Indeed, single cell recording studies have found that neurons in the lateral prefrontal cortex of macaques encode working memory representations via persistent firing, validating the results of lesion studies. However, other studies have reported that neurons in some areas of the parietal and temporal lobe—classically associated with visual perception—similarly encode working memory representations via persistent firing. This prompted a line of enquiry about the role of the prefrontal and other associative cortices in working memory and perception. Here, we review evidence from single neuron studies in macaque monkeys examining working memory representations across different areas of the visual hierarchy and link them to studies examining the role of the same areas in visual perception. We conclude that neurons in early visual areas of both ventral (V1-V2-V4) and dorsal (V1-V3-MT) visual pathways of macaques mainly encode perceptual signals. On the other hand, areas downstream from V4 and MT contain subpopulations of neurons that encode both perceptual and/or working memory signals. Differences in cortical architecture (neuronal types, layer composition, and synaptic density and distribution) may be linked to the differential encoding of perceptual and working memory signals between early visual areas and higher association areas.

Keywords: visual perception, working memory, persistent activity, prefrontal cortex, visual system

OPEN ACCESS

Edited by:
Shintaro Funahashi,
Kyoto University, Japan

Reviewed by:
Guido Marco Cicchini,
National Research Council (CNR), Italy
Paolo Medini,
Umeå University, Sweden

***Correspondence:**
Julio C. Martinez-Trujillo
julio.martinez@robarts.ca

Received: *25 August 2021*
Accepted: *25 October 2021*
Published: *26 November 2021*

Citation:
Roussy M, Mendoza-Halliday D and Martinez-Trujillo JC (2021) Neural Substrates of Visual Perception and Working Memory: Two Sides of the Same Coin or Two Different Coins? Front. Neural Circuits 15:764177. doi: 10.3389/fncir.2021.764177

1. INTRODUCTION

1.1. Are Working Memory and Perception Two Distinct Cognitive Functions?

Visual perception is defined as the ability to interpret the surrounding environment from electromagnetic signals entering the retinas. Visual perception occurs when neurons across different areas of the visual system are activated by retinal inputs and the brain produces "a percept" or interpretation of the physical reality (e.g., seeing a red shirt) (Chalupa and Werner, 2003). Visual working memory is the ability to remember and manipulate, for short periods of time, an interpretation of the physical reality when the corresponding physical signals are no longer entering the retinas (Baddeley, 2010) (e.g., the mental image or memory of the same red shirt). Perhaps the best operational distinction between visual perception and working memory is that the former is linked to the flow of visual inputs, while the latter is not. The distinction between perceptual and mnemonic states seems intuitive. Indeed, a typical human subject can distinguish when they "see" an image of a red shirt (perceptual) and when they "remember" an image of a red shirt (mnemonic). Thus, for typical individuals, the mental states corresponding to visual perception and working memory are different and distinguishable.

It is important to clarify that perception is not always a lawful reflection of the physical properties of stimuli. Phenomena such as perceptual illusions have taught us that perception is a creative process, and under particular circumstances of ambiguity, we could "misinterpret" the physical environment or even interpret the same environment in multiple ways (Todorović, 2020). However, we would argue that in general, perception reflects the physical reality in a predictable manner. Therefore, in the current review, we refer to perception as a predictable and stable process and exclude cases of perceptual illusions or variations (Foster, 2011). We focus on the distinction between the physical presence of an object (visual perception) and the mental image of the same object when unavailable to the senses (mnemonic representation).

A somewhat related review of this topic largely based on findings from human experiments using non- invasive signal measurement techniques has been recently published (Dijkstra et al., 2021). In this current review, we primarily refer to data collected in experiments using invasive techniques in non-human primates such as lesion studies and electrophysiological recordings. We make the reasonable axiomatic assumption that anthropoid non-human primates with a developed visual system and brain areas that have human homologs (Petrides, 2005) use perception and working memory as part of their cognitive repertoire (Beran et al., 2016).

The distinction between working memory and perceptual functions can be traced to lesions studies conducted more than a century ago in humans and animals. They reported that damage to certain brain areas can produce selective deficits of working memory while sparing visual perception (reviewed in the next section). However, more recent studies have reported co-existence of signal correlates of visual perception and working memory across brain areas and have questioned the segregation of the neural substrates for these two functions in the brain (reviewed in the section "Dissociating Visual Working Memory and Perception: Electrophysiological Studies of Single Neurons Across Brain Areas"). Influential in this latter view, have been findings of functional imaging and EEG/MEG studies in human subjects (Dijkstra et al., 2021).

On a cautionary note, we have found that the diversity of techniques used to record brain signals in humans and non-human primates and that of paradigms (tasks) used to explore working memory and perception makes it difficult to examine the relationship between the neural correlates of the two functions across species. This is in part because different techniques used in humans and non-human primates explore different spatial and temporal scales of brain activity and record different types of signals. It is therefore difficult to reconcile the results of studies in different species. In this review, we have taken a focused approach to examine reports mainly from studies in non-human primates using different methodologies to study working memory coding along areas of the visual processing pathways and its relationship to visual perception. We also assume that over the short temporal scales of perception and working memory, action potentials are the central elements of information coding and transmission between neurons and neuronal networks over distances that extend beyond synapses. Therefore, we concentrate on studies that have directly recorded action potentials from neurons or neuronal populations during behavioral tasks that involve visual perception and working memory.

2. MAIN

2.1. Dissociating Visual Working Memory and Perception: Lesion Studies

The idea that perceptual and mnemonic representations are separable in the brain originated by investigations into patients with localized cortical damage. Although they did not directly measure working memory, early case studies describe independent impairments in top-down driven representations (visual imagery) or perception. Charcot and Bernard first described a patient in 1883 that could identify objects but was neither able to form mental representations of these objects nor envision them from memory (Charcot and Bernard, 1883). The opposite deficit has also been described in which patients are unable to perceive objects yet can describe them in detail based on clear mental representations. A well-known case of this, described in patient C.K, was presented by Behrmann and colleagues in the early 1990's. C.K was unable to identify either simple or complex items but was able to produce clear and detailed drawings of those same items (Behrmann et al., 1994).

Early lesion studies in non-human primates supported the dissociation between working memory and perception. Jacobsen (1936) conducted a series of lesion experiments in the prefrontal cortex (PFC) of different species of non-human primates [*Macaca mulatta* (rhesus macaque), *Cercocebus torquatus* (mangabey), and *Papio papio* (baboon)] and noticed that the lesions produced selective performance deficits in delayed response tasks, where animals had to remember the

locations or features of objects for a short period of time. Importantly, the animals could perform other perceptual tasks without major difficulty (Jacobsen, 1936). These results suggested that lesions of the PFC affect mainly working memory while sparing perception. In another study, Chow, Blum and Blum conducted lesion experiments of the posterior association areas of the parieto-occipital temporal region and the prefrontal areas close to the frontal pole in macaque monkeys (Chow et al., 1951). They found that posterior lesions did not substantially affect performance in a delayed response task. On the other hand, prefrontal lesions did affect the animals' performance without substantially affecting other discrimination abilities. They concluded that the PFC plays a selective role in the delayed aspects of the task.

In 1952, Harlow and colleagues reported two distinct deficits associated with lesions of the posterior cortices and anterior (prefrontal) cortices in macaque monkeys. The animals with posterior lesions had stronger deficits in discrimination tasks, whereas animals with anterior prefrontal lesions had stronger deficits in delayed response tasks (Harlow et al., 1952). Curiously, lesions to the posterior parietal cortex have little effect on the performance of delayed response tasks. In the case of complete and bilateral posterior parietal cortex lesions, visuospatial information may possibly arrive to PFC through alternate connections (i.e., anterior/posterior cingulate cortex) or through connections to preoccipital regions (i.e., dorsomedial area DP), via the occipitofrontal fascicle (Selemon and Goldman-Rakic, 1988; Yeterian and Pandya, 2010; Yeterian et al., 2012; Arnsten, 2013).

In 1952, Pribram and coworkers described that lesion of the PFC in baboons (*Papio papio*) also produced performance deficits in delayed response tasks. Dorsolateral lesioned animals had greater alterations in all tasks compared to ventromedial lesioned animals (Pribram et al., 1952). In 1969, a study by Butters and Pandya (1969) reported a more specific finding concerning the role of the PFC in working memory tasks. They compared the performance of lesioned and control rhesus macaques in delayed alternation tasks. Lesions included bilateral inferior parietal cortex lesions and three types of prefrontal lesions around the principal sulcus. Animals with lesions of the anterior and posterior thirds of the principal sulcus as well as periarcuate and parietal lesions could re-learn the delay alternation task but animals with lesions of the central part of the arcuate sulcus could not re-learn the task and showed permanent deficits. A later study by Warren and Divac (1972) demonstrated that the effect of principal sulcus lesions extends to delayed response tasks.

Importantly, decades earlier, Malmo (1942) and Orbach and Fischer (1959) reported the importance of the PFC in maintaining working memory representations in the presence of irrelevant incoming visual signals. Without PFC, stored mental representations can be disrupted by incoming sensory signals. These studies highlighted the importance of PFC to guard mental representations from distracters.

In 1960, Miles and Blomquist (1960) reported that lesions of the PFC in squirrel monkeys (*Samiri sciureus*), a new world primate, produced a similar syndrome as the one observed in the old world species. The syndrome consisted of hyperactivity, deficits in delayed response tasks, and no adverse effects on the ability to solve discrimination tasks when the stimulus was present. This study extends the observed effects of prefrontal lesions to new world monkeys, with a relatively less expanded PFC than their old world relatives (Passingham and Wise, 2012).

More recently, in the second half of the twentieth century, spatially refined lesion and pharmaceutical inactivation studies in the PFC of macaque monkeys further demonstrated perturbation of visuospatial working memory representations and sparing of perceptual representations (Sawaguchi and Goldman-Rakic, 1991; Funahashi et al., 1993; Iba and Sawaguchi, 2003). This work introduced the concept of mnemonic scotoma, a deficit in remembering a certain spatial location during a delayed response task induced by inactivating small regions in the lateral prefrontal cortex (LPFC) (Funahashi et al., 1993). However, animals with mnemonic scotomas are able to make saccades to the region of the mnemonic scotoma when the target object is visually available. The latter not only confirmed the results of previous studies, but also emphasized a major role of the PFC in visual working memory and a lesser role in visual perception. Thus, from lesions studies, one may conclude the PFC is needed for maintaining information in working memory, but it is not essential for visual perception (i.e., when visual information remains available). **Table 1** shows a summary of studies that explore the effects of lesions in perceptual and working memory tasks in non-human primates. **Figure 1** provides a graphical summary of this information.

2.2. Dissociating Visual Working Memory and Perception Along the Visual Pathways

Departing from the accumulated evidence in early lesion studies in non-human primates (reviewed above) and the development of single cell recording techniques in behaving animals (Hubel, 1957), Fuster and Alexander (1971) recorded the responses of neurons in the LPFC and mediodorsal nucleus of the thalamus in macaque monkeys during delayed response tasks. They discovered cells in the LPFC that represent remembered locations and features of visual stimuli via persistent firing: an increase in firing rate above baseline tuned for the location of the items held in working memory. One important feature of persistent firing is that it occurs in the absence of sensory inputs, when the cue or sample stimulus disappears from the visual field—the so-called delay period of working memory tasks. An amount of controversy has been accumulating around the concept of persistent firing. For example, whether it is sustained during the entire delay period by single neurons or populations, or it has a temporal structure (e.g., oscillations in certain frequency bands) (Sreenivasan et al., 2014; Lundqvist et al., 2016, 2018; Constantinidis et al., 2018). In the original report Fuster and Alexander (1971) do not make considerations about the temporal structure of persistent firing in individual trials but used trial averages. Although clarifying the temporal structure of persistent firing is important to reveal the mechanisms of working memory coding, this review will not expand on this topic. We will consider persistent firing as increases in firing rate that encode

TABLE 1 | Lesion studies.

References	Species	Main finding
Bianchi (1895)	*Papio cynocephalus*	Lesions of the frontal cortex resulted in attentional but not perceptual deficits. Concludes that the frontal lobes serve to fuse incoming sensory signals and motor output forming associative representations.
Jacobsen et al. (1935)	*Pan troglodytes*	Bilateral lesions of the prefrontal cortex diminished performance on a delayed response task.
Jacobsen (1936)	*Macaca mulatta* *Cerocebus torquatus* *Papio papio*	Bilateral lesions of the prefrontal cortex diminished performance on a delayed response task.
Jacobsen and Nissen (1937)	*Macaca mulatta*	Bilateral lesions of the prefrontal cortex diminished performance on a delayed alternation task.
Malmo (1942)	*Macaca mulatta* *Cerocebus torquatus*	Bilateral prefrontal lesions made animals more susceptible to extraneous stimuli occurring during the delay interval of a delayed response task.
Finan (1942)	*Cerocebus torquatus*	Bilateral prefrontal lesions decrease performance of a delayed response task. Pre-rewarded food increased performance.
Spaet and Harlow (1943)	*Macaca mulatta*	Bilateral prefrontal lesions created greater deficits in delayed reaction problems (non-spatial delayed reaction, spatial delayed reaction) than in stimulus-object discrimination problems.
Campbell and Harlow (1945)	*Macaca mulatta*	Bilateral lesions of the frontal cortex related in reduced performance on a spatial delayed response task. Performance differed based on recovery time from surgery.
Pribram (1950)	*Papio porcarius*	Bilateral lesion of the prefrontal cortex anterior to FEF decreased performance on a delayed response task. Insulin administration, cooling and fasting increased performance likely through increased reward value of the stimulus (food).
Chow et al. (1951)	*Macaca mulatta*	Animals with bilateral lesions of the prefrontal cortex showed similar performance deficits on a delayed reaction test as animals with prefrontal lesions and additional damage to parietal and temporal regions. Sedative drugs did not improve performance.
Harlow et al. (1952)	*Macaca mulatta*	Anterior and posterior lesions produce predominantly delay response and discrimination deficits respectively.
Pribram et al. (1952)	*Papio papio*	Dorsolateral lesions reduced performance on delayed response-type problems but showed little effect on visual-discrimination task performance. Two of the four animals with ventromedial lesions showed no change in task performance.
Blum (1952)	*Macaca mulatta*	Lesions to the ventrolateral and dorsal region produced smaller deficits in a visual and auditory delay reaction tasks while lesions in the midlateral region (region anterior to the arcuate sulcus) produced large deficits.
Mishkin and Pribram (1955)	Macaque (unknown)	Lesions to the anterolateral frontal cortex resulted in poor performance on a series of delayed alternation problems.
Mishkin and Pribram (1956)	*Macaca mulatta*	Animals with bilateral anterolateral prefrontal lesions were tested on a series of delayed response tasks. Lesions resulted in deficits in the performance of traditional delayed response tasks, but performance increased when traditional cues are replaced by non-positional cues.
Orbach (1956)	*Macaca mulatta*	Bilateral prefrontal lesions resulted in deficits in a delayed response task within hours after surgery. This deficit was present 14 days after surgery though there was a slight recovery in performance.
Rosvold and Delgado (1956)	*Macaca mulatta*	Stimulation in the region of the head of the caudate nucleus impaired alternation without affecting visual discrimination, as did tissue destruction in the same site.
Mishkin (1957)	*Macaca mulatta*	Lesions of the midlateral region of the prefrontal cortex (anterior to arcuate sulcus) produced a deficit in a delayed alternation task that was as severe as total anterior frontal lesions.
Orbach and Fischer (1959)	*Macaca mulatta*	Bilateral lesions of the frontal granular cortex reduced performance on a delayed response task. Performance in animals with lesions was further reduced with added light interruption. Retraining on the task after surgery did improve performance.
Miles and Blomquist (1960)	*Saimiri sciureus*	Bilateral frontal lesions result in reduced delayed response performance but show no change in discrimination learning.
Gross and Weiskrantz (1962)	*Macaca mulatta*	Lesions surrounding the principal sulcus resulted in greater impairment on delayed response tasks whereas frontal lesions excluding tissue surrounding the principal sulcus resulted in greater impairment on auditory-discrimination tasks. Lesions in either area did not affect performance of a visual-discrimination task.

(Continued)

TABLE 1 | Continued

References	Species	Main finding
Tucker and Kling (1967)	*Macaca mulatta* *Macaca speciosa*	Bilateral lesions of the dorsolateral frontal granular cortex at either the 35th postnatal day or 3 years of age showed similar deficits in a delayed alternation task but performance on a delayed response task was better in animals with earlier lesions.
Butters and Pandya (1969)	*Macaca mulatta*	Bilateral lesions were performed in the anterior, middle, or posterior thirds of the principal sulcus, of the periarcuate prefrontal region, or of the inferior parietal lobule. Lesions within the middle third of the principal sulcus produced deficits on a delayed alternation task whereas lesions in other regions had little effect.
Fuster and Alexander (1970)	*Macaca mulatta*	Performance of a delayed response task was impaired by bilateral cooling of the dorsolateral prefrontal cortex.
Goldman and Rosvold (1970)	*Macaca mulatta*	Lesions around the principal sulcus impaired performance on the spatial task with delay and lesions around the arcuate impaired performance on the spatial task without delay.
Goldman et al. (1971)	*Macaca mulatta*	Lesions to the dorsolateral prefrontal cortex and to regions along the principal sulcus resulted in deficits in both a spatial discrimination task and spatial delayed response task.
Stamm and Weber-Levine (1971)	*Macaca mulatta*	Total bilateral lesions of the dorsolateral prefrontal cortex and lesions of the banks and floor of the principal sulcus produced the greatest deficits on a delayed alternation task while lesions to the surrounding dorsolateral cortical strips produced smaller deficits.
Butters et al. (1971)	*Macaca mulatta*	Lesions were made in the superior and/or inferior banks of the middle third of principal sulcus. Lesions which involved both banks led to greater deficits in a spatial delayed alternation and place reversal task than lesions to either bank alone.
Warren and Divac (1972)	*Macaca mulatta*	Lesions of the middle third of principal sulcus decrease performance of a delayed response and delayed alternation task.
Fuster and Bauer (1974)	*Macaca mulatta*	Cooling of the prefrontal cortex reduced performance of a delayed matching-to-sample task with bilateral cooling having a greater effect than unilateral cooling. Cooling of the parietal cortex did not produce a deficit.
Oscar-Berman et al. (1975)	*Macaca mulatta*	Lesions to the dorsolateral prefrontal cortex produced greater deficits in a delayed response task than lesions to the ventrolateral orbito-frontal cortex but had a smaller impact on visual and auditory discrimination tasks.
Passingham (1975)	*Macaca mulatta*	Dorsal prefrontal lesions decreased performance of a spatial delayed alternation task but had little impact on a delayed matching task for colors. Ventral prefrontal lesions impaired performance on the delayed matching task for colors.
Bauer and Fuster (1976)	*Macaca mulatta*	Delayed matching and delayed response deficit from cooling dorsolateral prefrontal cortex in monkeys.
Mishkin and Manning (1978)	*Macaca mulatta*	Lesions surrounding the principal sulcus resulted in deficits on delayed spatial memory tasks but had little effect on three non-spatial tasks such as delayed object matching, and delayed color matching.
Brozoski et al. (1979)	*Macaca mulatta*	Depletion of prefrontal dopamine leads to deficits on delayed alternation but not visual pattern discrimination.
Sawaguchi and Goldman-Rakic (1991)	*Macaca mulatta*	Local injections of selective D1 receptor antagonists into the prefrontal cortex reduced performance of an oculomotor delayed response task but had no effect on performance of a visually guided saccade task.
Funahashi et al. (1993)	*Macaca mulatta*	Unilateral lesions of the dorsolateral prefrontal cortex produced the greatest deficits in an oculomotor delayed response task for contralateral targets. Deficits were not seen for a visually guided saccade task suggesting the existence of mnemonic scotomas.
Petrides (1995)	*Macaca nemestrina*	Lesions of the mid-dorsal part of the lateral produced deficits in non-spatial self-ordered and externally ordered working memory tasks. The number of remembered items influenced performance. Deficits were not seen after lesions of the posterior dorsolateral frontal cortex (surrounds the arcuate sulcus).
Petrides (2000)	*Macaca nemestrina*	Increasing the number of stimuli to be remembered during a visual working memory task impaired performance after mid-dorsolateral lesions but not after anterior inferotemporal lesions whereas the opposite was true after extending the duration of the delay period. Full lesion of the mid-dorsolateral region created greater deficits than lesions on area 9 alone.

(Continued)

TABLE 1 | Continued

References	Species	Main finding
Sawaguchi and Iba (2001)	*Macaca mulatta*	Local injection of muscimol into the dorsolateral prefrontal cortex produced deficits in an oculomotor delayed response task to specific and typically contralateral target locations. No deficits we identified for a visually guided saccade task.
Croxson et al. (2011)	*Macaca mulatta*	Selective lesions of cholinergic input to prefrontal cortex severely impaired on a spatial working memory task while leaving unimpaired decision-making and episodic memory.
Upright et al. (2018)	*Macaca mulatta*	Reversible chemogenetic inhibition of only 3% of prefrontal neurons is sufficient for impairing performance on a spatial delayed response task.

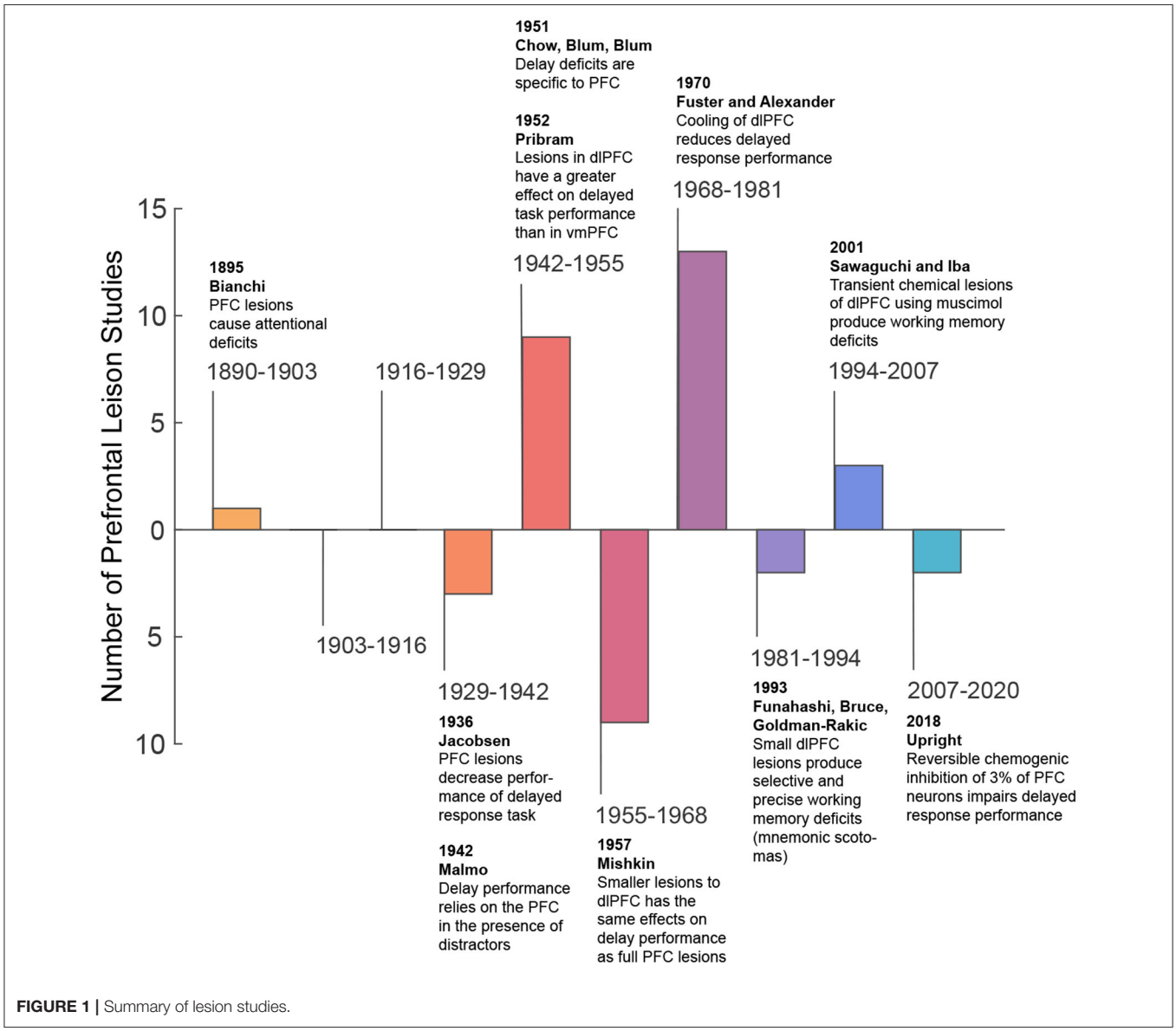


FIGURE 1 | Summary of lesion studies.

the contents of working memory. The temporal structure of such changes may be variable in individual neurons and across tasks.

It must be noted that rodent models are commonly used to study short-term memory and delay activity has been reported in areas associated with rodent cognition, in particular the

medial prefrontal cortex (Park et al., 2019; Ozdemir et al., 2020). Although experiments using rodent models have enriched our understanding of short-term memory mechanisms, the rodent visual system diverges from that of primates: rodents lack a granular prefrontal cortex making the comparison with primate brain regions problematic (Uylings et al., 2003; Passingham and Wise, 2012). Interareal connectivity between rat medial prefrontal cortex also diverges from primate lPFC in which it was shown to be more similar to primate premotor regions (Schaeffer et al., 2020), further complicating direct comparisons. The topic of similarities and differences between short term or working memory mechanisms in rodents (mice and rats) and primates necessitates an extensive discussion. Our review will therefore focus on experiments in primates.

We must also indicate here that we are not distinguishing different aspects of working memory in this review. What some believe makes working memory distinct is that it implies manipulation of information and not simply maintenance in its original form (Baddeley, 2010) (e.g., a mental rotation of an object or a reference frame transformation from retina-centered to space centered). However, physiological studies in non-human primates have not classically made that distinction, and refer to working memory in its maintenance aspect (Goldman-Rakic, 1995). We will continue this tradition here and acknowledge that work needs to be done to clarify this issue.

The initial results of Fuster and Alexander in the LPFC were confirmed by other studies (Kubota and Niki, 1971), thus supporting the hypothesis that the neural substrates of working memory is allocated to the LPFC in primates (areas 46/9, around the principal sulcus). Importantly, the existence of persistent firing pointed toward a different mechanism for working memory coding compared to the mechanisms of permanent synaptic storage for long-term memory (Eccles, 1986). The fundamental idea is that the memory is maintained as long as persistent firing is maintained; therefore, it dissipates when neurons stop firing. This matches the behavioral observations of working memory as a mechanism susceptible to temporal decay (Baddeley, 2010). It also agrees with the fact that most representations held in working memory are not transferred into long-term memory. Such a continuous transfer would be wasteful in many situations since many items held in working memory are "temporally useful" and therefore not needed to be kept in long-term memory (e.g., the location of a car in a parking lot after driving out of the parking lot).

Fuster and Alexander also reported in their seminal work that a number of neurons in the LPFC were activated during the cue period of the delayed response task, whereas others were active only when the cue stimulus disappears. They suggested that the activity during the cue period may be related to attention since many neurons did not show selectivity for the position of the cue (Fuster and Alexander, 1971). Importantly, the fact that a group of neurons show activity exclusively during the delay period (mnemonic cells) suggests that, at the level of individual neurons, the neural correlates of working memory can be dissociated from those of visual perception (**Figure 2A**).

After this initial report, several studies have reported that persistent firing representing the contents of working memory can also be found in association areas of the frontal, parietal (Andersen et al., 1985), and temporal lobes (Mikami and Kubota, 1980; Fuster and Jervey, 1981); for a review see (Leavitt et al., 2017a). These findings sparked the debate on what the role of association areas outside the LPFC in WM coding is (Riley and Constantinidis, 2015). This question remains mostly unanswered but something that is common to studies in the PFC and posterior association cortices is the existence of neurons that represent information during different task periods. Thus, no matter where persistent firing has been reported, neurons showing selectivity for a visual cue are not necessarily the same as neurons showing persistent firing when a representation of a cue is held in working memory. The latter could be interpreted as evidence in favor of the hypothesis that the substrates for perception and working memory are at least partially segregated within areas such as the LPFC.

One study has reported that the proportion of neurons encoding information during the cue and delay period of a delayed match-to-sample task changes as one moves along the hierarchy of visual processing from area MT (neurons almost exclusively encode during the sample period) to MST (neurons predominantly encode information during the sample period but a proportion of cells also encode information during the delay period) to LPFC (a similar proportion of neurons encode information during the sample and delayed period) (see **Figures 2B–D**) (Mendoza-Halliday et al., 2014). Bisley et al. (2001) reported that microstimulation of area MT during the encoding stage of a working memory task for motion direction biased the neural response to direction but stimulation during the delay period did not. The latter supports the hypothesis that although sensory areas are recruited during visual processing and perception, which is require for encoding information during working memory tasks, they may play a lesser role in maintaining working memory representations. These results match the pattern revealed by lesion studies with neurons in the posterior early sensory and association areas encoding predominantly perceptual information and neurons in the PFC encoding mnemonic signals (**Figures 2B–D**). One may also conclude that a population of neurons in areas such as LPFC seem to encode information about the cue during all task periods (Mendoza-Halliday and Martinez-Trujillo, 2017).

Although we, as most researchers, discuss independent properties of various brain regions, it is important to expand beyond the local-circuit model and recognize the impact that cortical—cortical connections have in generating persistent activity. In 1998, Chafee and Goldman-Rakic made the observation that patterns of neuronal activity in the dorsolateral prefrontal cortex and parietal area LIP/7a were remarkably similar including their spatial tuning and ability to generate persistent activity (Chafee and Goldman-Rakic, 1998). They later demonstrated, using cortical cooling, that WM memory related activity in both regions were dependent on shared reciprocal activity (Chafee and Goldman-Rakic, 2000). Synchronized activity between PFC and PPC underlying working memory has since been substantiated (Salazar et al., 2012). The prefrontal and parietal cortices thus represent two regions in which persistent activity is frequently observed but the role of their

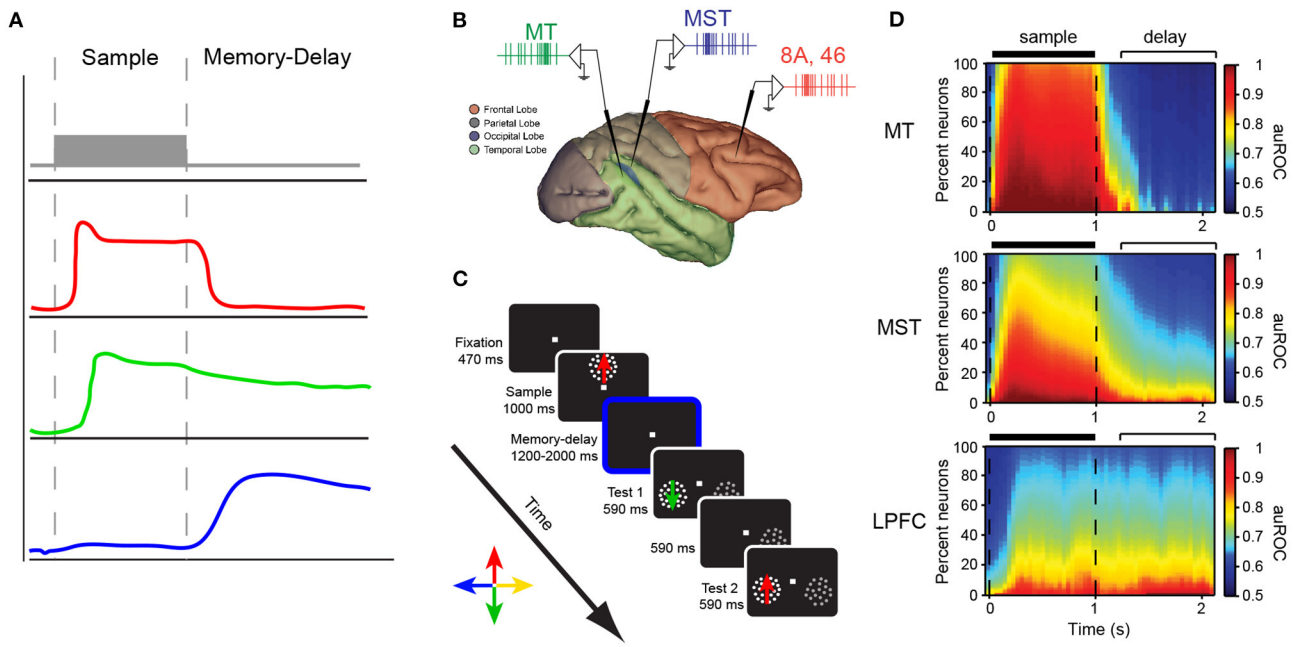


FIGURE 2 | (A) Different response profiles of neurons in the LPFC of macaque monkeys during the sample and delay periods of a delayed response task. **(B)** Recording locations in the study of Mendoza-Halliday et al. (2014). MT (green), MST (blue), LPFC areas 8A/46 (red) during a match-to-sample task for motion direction **(C)**. **(D)** The proportion of neurons showing encoding of motion direction during the sample and delay period of the task in the three areas. The color scale represents the strength of direction selectivity quantified by the area under the receiver operating characteristic curve (auROC).

reciprocal connections is still debated (Christophel et al., 2017; Constantinidis et al., 2018).

To explore the function of these prefrontal- parietal connections, Murray et al. (2017) developed a computational model of two bidirectionally connected modules that biophysically represented local networks of PFC and PPC. This model shows that PPC functions in a weak attractor state and transiently encodes the stimulus and propagates this sensory signal to PFC. Although both maintain the WM representation after stimulus offset, the attractor state is stronger in PFC module, allowing for robustness against distractors. Feedback projections from PFC can additionally switch PPC neurons back to encoding target stimuli after distractor presentation. Therefore, in this model, persistent activity was supported by both local and long-range network connections.

Synchronized activity was also identified between area MT and LPFC through observations of phase- coherent local field potential oscillations during a motion direction match to sample task. This observation suggests that persistent activity in LPFC modulates synaptic activity in MT, again showing a top-down mechanism by which memory signals in LPFC influence stimulus processing (Mendoza-Halliday et al., 2014).

Regarding the neural correlates of visual perception, there is a large body of literature starting as early as when single cell recording techniques became popular (Hubel, 1957). Early studies of Hubel and Wiesel demonstrated that neurons in the monkey primary visual cortex (V1) encode the features of sensory stimuli shown inside their receptive field (RF) (Hubel and Wiesel, 1968). Later studies discovered similar selectivity in other brain areas of both the dorsal and ventral visual pathways (Mikami et al., 1986). The selectivity for features and their conjunction becomes more complex in areas downstream from V1 (e.g., linear motion in MT and complex optic flow motion in MST, or color and orientation selectivity in V4 and face selectivity in IT) (Felleman and Van Essen, 1991). However, most of these studies focused on the specific role of brain areas in conscious visual perception rather than in the distinction between perception and mnemonic processes. For example, lesions of area V1 leaves subjects cortically blind; however, lesioned individuals may show some residual vision or blindsight, likely suggesting that some perception can happen without V1 (reviewed in Leopold, 2012). Nevertheless, many agree that visual perception is deeply impaired after V1 lesions, suggesting that V1 is a bottleneck for visual signals entering higher level areas of the visual pathways (Leopold, 2012).

Remarkably, selective deficits in motion perception without affecting contrast thresholds can be observed after lesions of area MT (Newsome and Pare, 1988). Area MT contains a high proportion of direction selective neurons that receive inputs from direction selective neurons in area V1 (Born and Bradley, 2005). These observations suggests that V1 is not sufficient for motion perception but necessitates area MT. This hypothesis has been supported by reports of electrical microstimulation in area MT neurons, biasing motion perception (Salzman and Newsome,

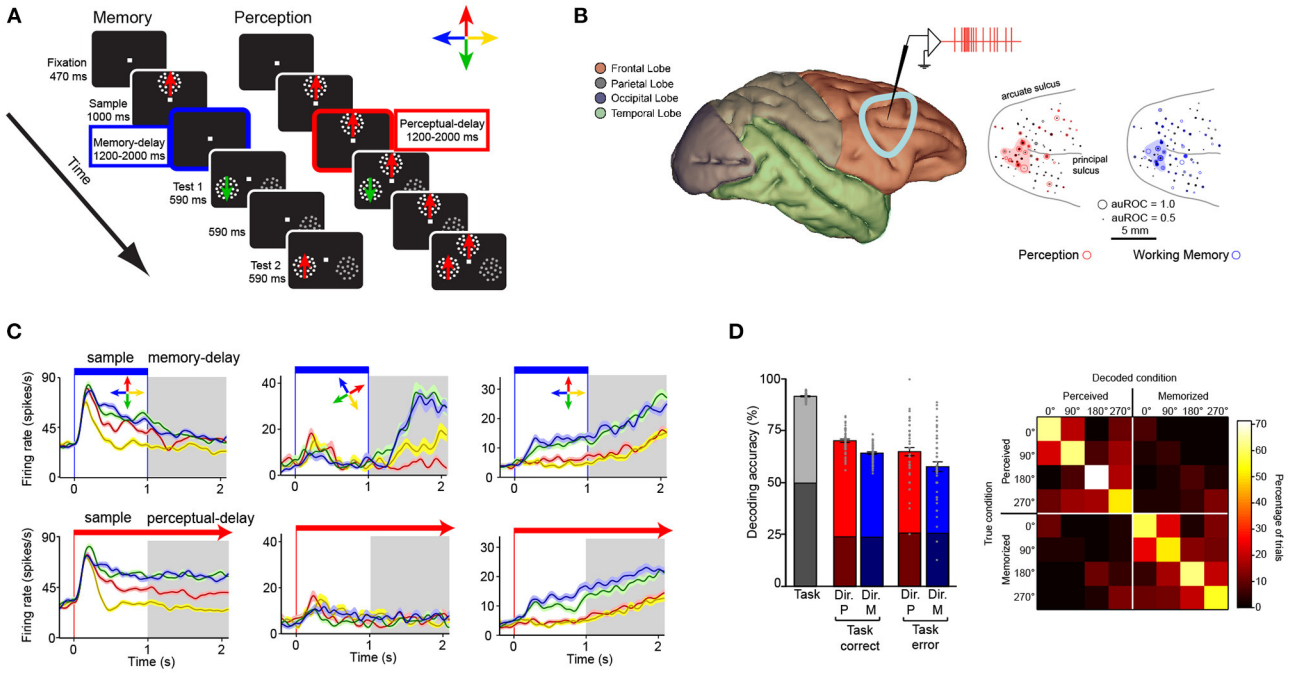


FIGURE 3 | Encoding of perceptual and working memory representations by subpopulations of neurons within the LPFC. **(A)** Memory and perception tasks. Panels illustrate stimuli and monitor setup. Animals fixate a dot at the center of a computer screen and press a button. Then a sample Random Dot Pattern (RDP) appears moving in one of four directions. In the memory task (left) the sample disappears after 1,000 ms. A delay period of 1,200 to 2,000 ms then occurs in which only the fixation point is on the screen. At the end of the delay period two patterns, a test RDP moving in either the same or different direction as the sample, and a distracter RDP that contains dots moving in random directions are presented. The animal releases a button if the test matches the direction of the sample or waits until the test disappears, and a second test RDP is presented after a 590-ms delay period. During the perceptual (right) task the sample RDP does not disappear but stays on during the duration of the trial. **(B)** Recording locations in the LPFC. The dots indicate the location of units with selectivity during the memory (blue) and perceptual (red) tasks. **(C)** Firing rate (y axis) over time (x axis) for three example neurons (left, center, and right columns) during the working memory task (top row) and perceptual task (bottom row). The task periods are indicated on top. **(D)** left bar graph: Accuracy of a linear classifier to decode, from the population of recorded neurons, the task (working memory vs. perceptual tasks, gray bar), the direction of the stimulus in each task in trials with correct and incorrect task decoding (red and blue bars). Right panel: confusion matrix for the classification of perceived and memorized direction corresponding to the correct trials.

1994). On the ventral pathways, damage to areas of the temporal lobe, such as the fusiform face area, leads to prosopagnosia: a selective deficit in face perception (Barton, 2003). Cells selective for faces have been extensively reported in the macaque inferiortemporal cortex (Perrett et al., 1984; Freiwald and Tsao, 2012). One influential study used visual rivalry, a phenomenon in which two different images are presented separately to each eye, the subject experiences alternating percepts of each image and periods of fusion of the two images. Single neuron activity is reported to more accurately reflect the percept downstream from area V1 (Leopold and Logothetis, 1996). The latter suggests that although V1 activity is essential to perception, the phenomenology that triggers perceptual awareness may occur or at least be triggered in downstream areas such as MT or MST, where neurons selective for the perceived features exist.

A central question to this review is whether the neural substrates that support visual perception and those that support working memory are the same or different. From the previous sections we may conclude that: (1) there is a set of areas in which neurons represent visual attributes such as motion (Duffy and Wurtz, 1991) and complex shapes (Rolls, 1984) during both perception and working memory tasks (Miller et al., 1991; Mendoza-Halliday et al., 2014), (2) there is a set of areas where neurons encode perceptual but not mnemonic representations of visual attributes, mainly early areas in the hierarchy of visual processing (i.e., V1 to MT in the dorsal pathway, and V1 to V4 in the ventral pathway), and (3) the relative proportion of neurons showing selectivity for perceptual and mnemonic visual attributes changes along the hierarchy of visual processing (i.e., the proportion of cells encoding mnemonic relative to perceptual representations is lower in MST than in LPFC), and (4) there are different subpopulations of neurons encoding perceptual and mnemonic representations in association areas, as well as a subpopulation of neurons that encode both types of representations.

2.3. Coding of Perceptual and Working Memory Representations by Subpopulations of Neurons Within Brain Areas

The exclusive role of the PFC and association cortices in working memory coding has recently been put into question (Pasternak and Greenlee, 2005; Christophel et al., 2017; Scimeca et al., 2018).

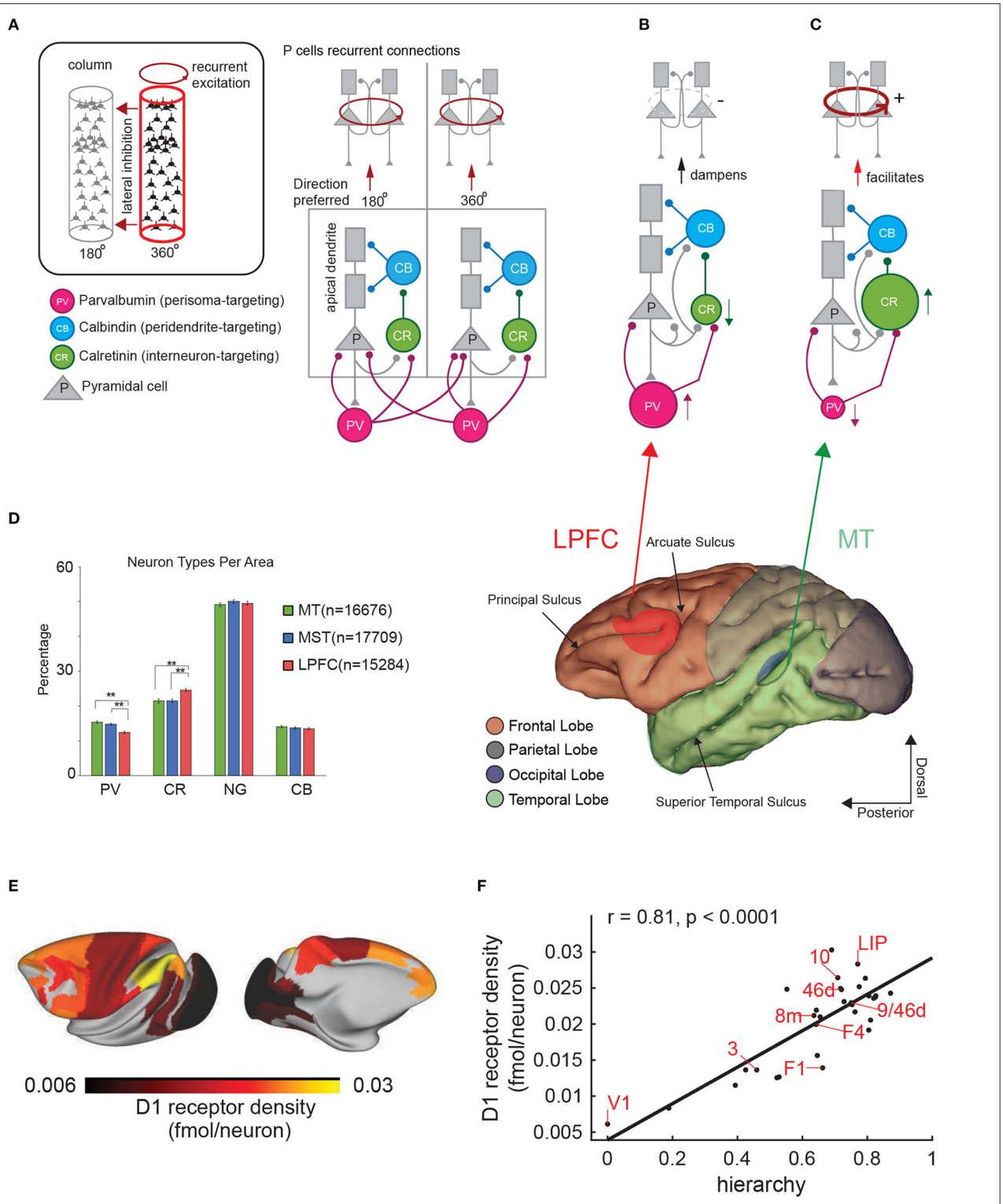


FIGURE 4 | Cortical architectures for perception and working memory. **(A)** Diagram showing the structure of two nearby cortical columns and the four main cell types (see inset). Observe pyramidal cells have at least two distinct compartments, the apical (distal) dendrites (gray rectangles) and the cell body. **(B,C)** different architectures based on the proportion of CR and PV interneurons and the ability to produce persistent firing. Lower panel shows a side view of the macaque brain and the different lobes in different colors. **(D)** Percentages of the 4 main cell types in areas MT, MST, and the LPFC (from Torres-Gomez et al., 2020). Distribution of Dopamine D1 receptors in the macaque brain. The color scale indicates the receptor density. **(F)** Correlation between position of a brain area in the hierarchy of visual processing and D1 receptor density. Each data point represents a brain area. The correlation coefficient and associated *p*-value are indicated courtesy of Froudist-Walsh et al. (2020). **(E)** D1 receptor density across the macaque cortex.

Some studies have proposed that neurons in sensory areas such as V1 and V4 encode working memory representations (Pasternak and Greenlee, 2005; Tong and Pratte, 2012). One argument in favor of this idea is that single neurons and neuronal populations in early sensory areas contain precise maps of visual attributes (Hubel and Wiesel, 1968; Albright, 1984; Born and Bradley, 2005). Thus, these populations must be recruited for perceiving such attributes accurately (Ester et al., 2013). However, encoding of visual attributes by single neurons and populations does not exclusively occur in early sensory areas such as V1, MT, and V4 but also occurs in downstream association areas where the neural correlates of working memory have been isolated. One example is coding of linear motion direction, which has been found not only in MT, but also in MST and LPFC (Bisley et al., 2004; Zaksas and Pasternak, 2006; Mendoza-Halliday et al., 2014; Mendoza-Halliday and Martinez-Trujillo, 2017), as well as in areas such as the Lateral Intraparietal (LIP) area (Freedman and Assad, 2006). Another example is encoding of color which has been reported not only in area V4, but also in the LPFC (Schwedhelm et al., 2020). Something to point out is that feature-selective neurons in the LPFC do not exhibit the retinotopic or feature-topic organization observed in early sensory areas (see **Figure 3B**; Mendoza-Halliday and Martinez-Trujillo, 2017). Thus, human studies using functional imaging techniques or EEG/MEG, that pool activity over cubic millimeters of cortical tissue, may underestimate selectivity for individual features or locations.

One important detail we have already mentioned is that feature selectivity in association areas does not only occur during delayed response tasks, but also during perceptual tasks when a stimulus remains visible (Mendoza-Halliday and Martinez-Trujillo, 2017). Interestingly, single unit responses to the same visual attribute become more correlated with behavioral outcomes as one advances downstream from V1 in the hierarchy of visual processing, for example from MST to LPFC (Freedman et al., 2001; Freedman and Assad, 2006; Mendoza-Halliday et al., 2014). Thus, association areas are equipped with "copies" of perceptual representations likely inherited from upstream areas, as well as with mnemonic representations that may emerge as a result of local processing. Unlike in visual areas, such "copies" are sensitive to the statistics of the environment and can form categories within a single feature dimension (Freedman et al., 2001).

Indeed, association areas in the frontal lobe such as the LPFC (around the posterior third of the principal sulcus) contain neurons that encode motion direction during a delayed match-to-sample task as well as neurons that encode memory representations of the same motion direction (Mendoza-Halliday and Martinez-Trujillo, 2017) (**Figures 3A,B**). A study found that about 1/3 of the neurons encoded perceptual representations of motion direction but not mnemonic representations, another 1/3 encoded mnemonic representations but not perceptual representations, and another 1/3 encoded a mix of both perceptual and mnemonic representations (Mendoza-Halliday and Martinez-Trujillo, 2017). Importantly, mnemonic cells are selective for motion direction only during the delay period and not during the visual presentation of the same motion direction (**Figure 3C** middle panel). Perceptual cells show the opposite pattern. Perceptual and mnemonic cells show a concentration within the posterior end of the principal sulcus and were also found to be spread within area 9/46 but without any apparent clustering by the type of representation (perceptual or mnemonic) or the feature they encode (**Figure 3B**). The latter deviates from observations in early sensory areas such as MT where neurons are topographically organized according to their RF location and motion direction they encode (Born and Bradley, 2005). As mentioned before, exploring the fine granulated functional architecture of the LPFC using BOLD signal measurement or EEG/MEG with spatial resolution of millimeters may cause an under estimation of feature selectivity or selectivity for perceptual and mnemonic representations.

The segregation of the different populations (perceptual and mnemonic) within LPFC allows a linear decoder to use single neuron activity to estimate whether a direction of motion is held in working memory or is visually presented (perception-memory decoder) as well as which direction is perceived or memorized (direction decoder) (**Figure 3D**). This indicates that perceptual and mnemonic signals as well as the features they encode can be discriminated, with reasonable accuracy, from the activity of neurons within the LPFC circuitry.

The existence of subpopulations of perceptual and mnemonic neurons within the LPFC circuitry may be considered as evidence in favor of separate substrates for perception and working memory "concentrated" within a single brain area microcircuit. One potential functional relevance of such a concentration is that a "read-out" of the population activity in the LPFC can provide a substrate for rapidly "identifying" the nature of the representation—perceptual or mnemonic—as well as its content. In the language of dynamical systems, the different activity profiles during the perceptual and mnemonic states could serve as attractors for corresponding cognitive states respectively (Wimmer et al., 2014). Interestingly, in patients with schizophrenia that lose the ability to differentiate between perceptual and mental representations (e.g., during hallucinations and delusions), abnormal patterns of activity are commonly reported in areas such as the LPFC (Callicott et al., 2000). Working memory deficits are also common in patients with schizophrenia and abnormal LPFC activity is consistently reported (Glahn et al., 2005; Forbes et al., 2009). In favor of this hypothesis, we have recently reported that systemic administration of ketamine, a drug often used to model symptoms of schizophrenia, modulates the activity landscape in the LPFC of macaques. In this experiment, ketamine drastically reduced performance during a working memory task by destroying the tuning of prefrontal neuron delay activity for remembered locations but had no effect on a perceptual control version of the same task (Roussy et al., 2021).

Another possible functional relevance to the coexistence of perceptual and mnemonic signals in the LPFC, is that information transfer from perceptual to mnemonic neurons can happen locally through short range connections within the area microcircuit, without the need for transfer through long range connections (e.g., perceptual neurons in MST transferring information about the cue to mnemonic cells in LPFC). For

example, during delayed matched-to-sample tasks, a read-out from sensory areas can be "loaded" into the perceptual cells and transferred "internally" to mnemonic cells that will "maintain" the representation via persistent firing. The role of perceptual and mnemonic cells in the generation of feedback signals that influence processing in early sensory areas is not clear. One study has documented synchrony between spikes in LPFC and local field potentials (LFPs) in MT during the delay period of a memory task (Mendoza-Halliday et al., 2014). Other studies have documented that microstimulation of areas such as the Frontal Eye Fields (FEF), posterior to LPFC, produces a modulation of responses in area V4 (Moore and Armstrong, 2003). Thus, it is possible that perceptual and mnemonic cells in LPFC play a critical role in modulating the activity of neurons in early visual areas during tasks that require attention either to sensory (perceptual) or mnemonic representations.

Finally, a concentration of neurons holding different representations of space, objects, and their attributes within a relatively small brain volume may facilitate the implementation of other cognitive operations such as attention. The predominant hypothesis of how attention is implemented is through competition via inhibitory interactions between neurons encoding representations of targets and distracters (Reynolds et al., 1999). Studies have reported evidence that the strength of such competition increases in association areas downstream from V1 (Buffalo et al., 2010; Lennert and Martinez-Trujillo, 2013). The strength of the competition also increases when targets and distracters become closer in space (Treue and Martínez Trujillo, 1999). Interestingly, association areas in the PFC possess spatial representations of the entire visual field, which may allow implementing competition between neurons representing targets and distracters in opposite hemifields via short range inhibitory connections within a local circuitry (Lennert and Martinez-Trujillo, 2013; Duong et al., 2019). Such operations could be more difficult to implement through short range projections between neurons in areas such as V1 or MT, where neurons represent stimuli in the opposite hemifield (Born and Bradley, 2005). Additionally, for the particular case of V1, with a large surface area, short range connections may be insufficient to implement operations when targets and distracters are far apart but still within the same hemifield. The latter may suggest the reduction in surface area from early visual areas relative to areas downstream facilitates interactions between neurons encoding different representations via short range connections.

2.4. Cortical Architectures for Perceptual and Mnemonic Coding

The primate cerebral cortex is not homogenous. Cortical architecture varies between early sensory and association areas in terms of thickness of cortical layers (Yang et al., 2018), neuronal densities (Collins et al., 2010), and proportion of different interneuron types (Torres-Gomez et al., 2020). The latter has been related to the ability of some local microcircuits to generate persistent firing in the absence of sensory stimulation (Leavitt et al., 2017a; Torres-Gomez et al., 2020). Indeed, the neural basis of persistent firing has been linked to the existence of recurrent connections between pyramidal cells within a local area circuitry (Goldman-Rakic, 1995). Empirical evidence shows more numerous excitatory synapses between pyramidal cells as well as differences in the distribution of long time constant NMDA receptors relative to short time constant AMPA receptors in the LPFC compared to the early visual cortex (Wang, 1999; Gonza'lez-Burgos et al., 2000; Zaitsev et al., 2012; Yang et al., 2018). These differences in excitatory synapse numbers and glutamate receptor types may explain the larger integration times found in association and executive areas of the visual processing hierarchy relative to sensory areas (Murray et al., 2014) and the ability of the former set of areas to encode working memory representations.

More recently, a larger proportion of interneurons that disinhibit pyramidal cells (e.g., calretinin positive (CR) cells) relative to interneurons that directly inhibit pyramidal cell firing (e.g., parvalbumin (PV) positive cells) have been reported in the LPFC compared to early visual areas like MT (Torres-Gomez et al., 2020). Wang has elaborated on a model that incorporates different cell types within the LPFC circuitry such as the calretinin positive (CR, sometimes identified as functionally similar to vasointestinal peptide (VIP)-expressing neurons in mice) and the calbindin positive neuron (CB, sometimes identified as functionally similar to somatostatin (SST)-expressing neurons in mice) (Wang et al., 2004; Wang, 2009). CR cells receive inputs from pyramidal cells and inhibit CB cells. The CB cells inhibit inputs into the dendrites of pyramidal cells (**Figure 4A**). Thus, an increase in the number or activation strength of CR neurons or their synapses onto CB cells would have a positive impact on the activation of the pyramidal cells (**Figure 4B**). A decrease in CR numbers or synaptic strength on their targets may have the opposite effect (**Figure 4C**). On the other hand, for PV neurons, an increase in their proportion or relative synaptic strength would increase the inhibition of pyramidal cells. A high ratio of CR to PV cells in LPFC relative to sensory areas may favor the emergence of persistent firing encoding working memory via facilitation of recurrent excitatory dynamics amongst pyramidal cells (Torres-Gomez et al., 2020) (**Figure 4D**). A low ratio of CR to PV cells (e.g., a relatively high proportion of PV cells or synaptic strength onto their target pyramidal cells) may cause strong inhibition of pyramidal cell firing and dampening of recurrent excitatory dynamics (perceptual encoding).

Supporting the idea that cortical architectures differ in their interneuron type proportions, a recent study has compared transcriptomic profiles of different neuronal types [PV, SST, VIP, and LAMP5 (Lysosome associated membrane protein 5 expressing interneurons)] in areas V1 and PFC of different species of primates (common marmosets, rhesus macaques, and humans). SST and PV originate from the Medial Ganglionic Eminence (MGE), while the VIP and LAMP5 originate from the Caudal Ganglionic Eminence (CGE). Neurons originating from the MGE tend to be more numerous in the deep layers while those originating in the CGE tend to be more numerous in the superficial layers. The study found that whereas PV and SST cells are more abundant in area V1, VIP, and LAMP5

are relatively more abundant in PFC. These differences may be due to the expansion of superficial (supragranular) cortical layers in primate association cortices, better documented in LPFC (Arnsten et al., 2012). Interestingly, such differences in the proportion of interneuron types were not found in the mouse (Krienen et al., 2020; but see Kim et al., 2017). This suggests that gradients of interneuron types may have become pronounced in primate neocortex, which is compatible with studies reporting a larger proportion of interneurons in primates relative to rodents (Džaja et al., 2014), as well as a larger proportion of CR cells in LPFC relative to sensory areas (Torres-Gomez et al., 2020).

One issue that remains unclear is why areas such as MST, where neurons show persistent firing during working memory tasks, do not show the same increase in the ratio of CR to PV neurons observed in the LPFC. There may be two possible explanations for this result. First, that persistent firing in areas such as MST is not intrinsic to the area circuitry and needs strong feedback signals from LPFC. Second, it is possible that the differences in CR interneurons proportion described in previous studies (Torres-Gomez et al., 2020) is not directly related to the ability to produce persistent firing, but to the ability of a local area circuitry to make persistent firing encoding working memory representations less disrupted by incoming distracting sensory signals (e.g., sensory signals unrelated to the representation held in working memory but co-occurring during the period of memory maintenance). In favor of the latter explanation, inactivation of the LPFC, where CR interneurons are abundant, increases distracter interference during working memory tasks and activity in LPFC is less disrupted by incoming distracting signals than in areas such as LIP (Suzuki and Gottlieb, 2013).

Wang and Yang have proposed a model circuit motif composed of the same cell types referred to earlier (pyramidal, CB, CR, and PV). Here the dendrite targeting CB neurons can regulate the flow of signals into dendritic trees. These neurons are controlled by CR interneurons. An increase in these cell type proportions and their control by cognitive signals encoding the behavioral relevance of stimuli in the environment in areas like LPFC where the filtering of distracter signals is particularly strong (Lennert and Martinez-Trujillo, 2011), may allow flexible "gating" of inputs into a pyramidal cell network. An increase in the proportion of SST neurons, a putative functional homolog of CB neurons in primates, has been reported in association areas of the mouse neocortex (Kim et al., 2017). One issue that remains unclear is how the gating of sensory inputs from upstream areas interplay with the gating of recurrent excitatory inputs from neighboring cells within the area. Further exploration will clarify apparent contradictions between the aforementioned hypotheses regarding cell type gradients.

Another difference between early sensory and association cortices concerns the distribution of receptors for neuromodulators that have been classically involved in working memory functions (Brozoski et al., 1979). Froudist-Walsh and coworkers have recently shown that receptors for neuromodulators that regulates working memory function such as the dopamine D1 receptor (D1R) (Williams and Goldman-Rakic, 1995) has an unequal distribution in the macaque cerebral cortex (Froudist-Walsh et al., 2020) (**Figures 4E,F**). D1 dopamine receptors action have been associated with the ability to filter distracter stimuli (Jacob et al., 2016). The concentration of D1 receptors increases along the hierarchy of visual processing reaching their maximal concentration in the parietal and prefrontal cortices. Froudist-Walsh and coworkers elaborated on a computational model in which release of dopamine favors persistent firing and resilience to distracters in association areas via its action on D1 receptors. Insufficient or excessive dopamine release on the other hand, makes persistent firing less robust to distracter interference (Froudist-Walsh et al., 2020). One relevant detail is that the model makes the prediction that dopamine increases the synaptic strength of the inhibition to the apical dendrites of pyramidal cells. Because recurrent excitatory connections between pyramidal cells target the soma and proximal dendrites, which are NMDA dependent and facilitated by D1R, and inhibitory connections from calbindin-expressing interneurons target the apical dendrites and are also facilitated by D1R, the next effect for dopamine release is to facilitate persistent firing via recurrent excitation. Additional details of this model can be found in Froudist-Walsh et al. (2020).

Despite an accumulating body of evidence in favor of different cortical architectures that support perception and working memory, several issues remain unexplained. For example, studies have reported that a noticeable proportion of neurons in areas of the LPFC encode perceptual but not mnemonic representations (see **Figure 3**; Mendoza-Halliday and Martinez-Trujillo, 2017). Here one may conceive the possibility that the LPFC microcircuitry is heterogeneous in composition and may contain features of both perceptual and mnemonic microcircuits. One may speculate that perceptual neurons inherit and "echo" the responses and selectivity from perceptual neurons in upstream visual areas (e.g., MT) via feed-forward inputs, processing these signals within circuits that do not include mnemonic neurons. During working memory, perceptual LPFC neurons then transfer such signals to mnemonic neurons, which are in turn capable of maintaining them via local recurrent excitatory networks that do not necessarily include perceptual neurons.

Is it possible the LPFC is a mosaic of perceptual and mnemonic cortical architectures that differ in basic features such as proportion of interneuron types, or the number of synapses that enable recurrent connections? If that were the case, one may conceive evolution of the neocortex produced such a hybrid architecture for a "purpose": compressing information about the nature of a representation (perceptual or mnemonic) within a brain area. One possibility is that such architecture originates, at least partially, during migrations of interneurons from MGE (e.g., PV) and CGE (e.g., CR/VIP) that produce cortical columns of different composition in areas such as LPFC. It may also be shaped by patterns of inputs and activity during development. As we have proposed earlier, this hybrid architecture may facilitate computations and information transfer within local microcircuits in an efficient manner. On the other hand, it may also make the brain more vulnerable to disorders of perception/imagination when such a circuit undergoes certain deviations from typical development in early

life, as can be seen in schizophrenia. This idea, however, needs to be tested experimentally.

An interesting question related to the possible existence of a hybrid architecture of perceptual or mnemonic "blocks" in the LPFC, is what the resolution of such blocks would be. Some studies have pointed out the existence of a non-retinotopic topography for mnemonic representations of visual space in the macaque monkey LPFC (Leavitt et al., 2017b). Another study found that neurons with the strongest selectivity for perceived and memorized motion directions were concentrated within a small subregion of LPFC near the posterior end of the principal sulcus (Mendoza-Halliday and Martinez-Trujillo, 2017). Moreover, a previous study has described a pattern of stripe-like areas in the LPFC that connects to the ipsilateral parietal cortex and the contralateral LPFC respectively (Goldman-Rakic and Schwartz, 1982). Could such a pattern be related to subregions of LPFC with perceptual and mnemonic architectures such as the ones illustrated in **Figures 4B,C**? One possibility is that neurons in perceptual blocks receive projections from the parietal cortex, while neurons in the mnemonic blocks receive projections from perceptual blocks within the same hemisphere and contralateral blocks in the opposite hemisphere. The latter may allow manipulation of spatial information in working memory (e.g., interhemispheric transfer of information; Brincat et al., 2021). However, this proposal remains speculative and future studies must clarify this issue. With the advent of modern techniques for high-yield electrophysiological recordings and 2-photon imaging of neuronal activity using calcium indicators (Yang and Yuste, 2017), it may be possible to test some of the hypotheses mentioned or proposed here.

2.5. The Case for Overlapping Substrates of Visual Working Memory and Perception

With the advent of modern functional imaging, it has been possible to measure Blood Oxygenation Level Dependent (BOLD) signals in humans performing perceptual and working memory tasks. One common finding is that it is possible to decode the contents of working memory from BOLD signals in early visual areas (V1-V4) (Tong and Pratte, 2012). Yet, electrophysiological studies in monkeys find little evidence of persistent firing of action potential by single neurons (see Leavitt et al., 2017a for a review). These functional imaging findings have been the motivation of a popular hypothesis that proposes early sensory areas are recruited, and may be necessary, for the maintenance of working memory representations (Postle, 2006; Ester et al., 2013; Scimeca et al., 2018). This hypothesis is known as the "sensory recruitment" hypothesis, and has been a matter of debate amongst neuroscientists investigating the topic (Scimeca et al., 2018). At first glance, the sensory recruitment hypothesis does not fully match the results of electrophysiological and lesion studies in non-human primates we have reviewed above. Below, we consider a few explanations for this mismatch.

Boynton (2011) outlines several hypotheses to understand the identified discrepancies between single neuron electrophysiology and fMRI findings. The first outlines that the BOLD signal more closely represents local field potential activity rather than spiking activity. It is possible that sensory areas are not recruited during working memory maintenance and the results of fMRI studies reflect feedback signals from higher-order association areas into early sensory areas. Such signals would increase synaptic activity and oxygen consumption in early visual cortex in a retinotopic or feature-topic fashion, which is sufficient to produce BOLD signals that provide information about remembered locations/features, but insufficient to significantly evoke action potentials from single neurons. In favor of this hypothesis, at least one study in monkeys has reported the direction of a stimulus held in working memory can be decoded from LFP signals recorded in area MT but cannot be decoded from spiking activity of neurons within the area (Mendoza-Halliday et al., 2014). Indeed, previous studies have shown that in certain experimental conditions, it is possible to dissociate between the inputs into a cell and the spiking outputs: BOLD signals are better correlated with LFP signals (as a measure of synaptic inputs) than with spikes (Logothetis and Wandell, 2004). The feedback signals into early visual cortex would help implement top-down attention, facilitating or prioritizing the processing of incoming stimuli that match the features or locations held in working memory (Mendoza et al., 2011). Such effects are commonly found in visual search paradigms (Bichot et al., 2019) and have been interpreted as top-down modulation of neuronal activity in early visual areas by attentional templates (working memory signals) originating in executive control areas of the parietal and PFC.

One issue that also needs clarification is why classification accuracy during working memory tasks is poorer using BOLD signals recorded in parietal areas and the LPFC compared to early visual cortex (e.g., V1, V4, MT) (Bettencourt and Xu, 2016; Ester et al., 2016). One possible explanation is that the retinotopy of visual space is weaker in high-order association cortices, leading to reduced decoding performance for working memory using BOLD signals (Xu, 2017). Here, one may consider that decoding methods used in fMRI rely on the selectivity of voxels for remembered features or locations. Such voxels are usually isotropic and distributed in a way that map BOLD signals in the cortex homogeneously. Although a voxel in areas like V1 and MT may include neurons with similar selectivities (Born and Bradley, 2005), this is not the case in late association areas such as the LPFC, where retinotopic and feature-topic maps are not homogenous (Leavitt et al., 2017b) (see **Figure 3**).

Boynton (2011) also suggest that discrepancies are caused by differences in experimental design including the use of different species. The same research group is unlikely to study both macaques and humans and use both fMRI and single neuron recording techniques. Differences in experimental approach and design and interpretation of results could certainly contribute to the observed discrepancies. Another possible explanation is that humans differ from other primates such as macaque monkeys in the way in which working memory networks encode information in the brain. The recruitment of early sensory areas could be a feature of the human cortex that is not present in macaques and other species of monkeys. This hypothesis is difficult to test. We did not find any study in humans recording neuronal

activity in early visual areas during working memory tasks. Methods such as fMRI, EEG, and MEG do not have sufficient spatial resolution to measure spikes in single neurons. They are most sensitive to transient changes in sensory inputs or behavioral states. Recordings of single neurons from areas such as V1 in human subjects during working memory tasks would clarify the issue. However, these experiments are difficult due to ethical constraints, and are exclusively performed in patients with clinically-implanted electrodes for epilepsy mapping, almost all of which do not target early visual areas. Although we cannot fully discard this hypothesis, it would assume that humans have undergone a major step in the evolution of working memory mechanisms and cortical architectures. Beside the expansion of the PFC and the more pronounced folding of the brain surface in humans, there is no evidence in favor of fundamental changes in circuitry between macaques and humans (Passingham and Wise, 2012). Future studies in humans may clarify this issue.

3. CONCLUSION

We conclude that the neural substrates of working memory and perception are segregated in the non-human primate neocortex. Neurons and neuronal populations in early visual areas mainly encode perceptual signals. In areas downstream, there are populations of neurons that encode both perceptual and working memory signals, with the relative proportion of neurons encoding the latter increasing from early association areas to the PFC. In the LPFC, the activity of neuronal populations can provide a neural substrate for the distinction between perceptual and mnemonic states via population activity profiles that can be translated into attractor landscapes. Changes in the architecture of microcircuits across the hierarchy of visual processing in terms of pyramidal cell morphology and connectivity, proportion of different interneuron types, and distribution of receptors (i.e., NMDA, AMPA, and dopaminergic) also reflect the changes in electrophysiological signals supporting perception and working memory. This suggests a parallel degree of heterogeneity between anatomy and physiology. Finally, the results from non-human primate studies do not match the proposition of a sensory recruitment hypothesis for working memory. The latter could be due to the heterogeneity of signal measurements and their interpretation across studies in humans and non-human primates, or to evolutionary changes in the mechanisms by which humans encode perceptual and working memory signals.

AUTHOR CONTRIBUTIONS

MR and JM-T contributed to the topic development, manuscript writing, and figure and table development. DM-H contributed to manuscript writing and figure development. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by NEURONEX-NSF/CIHR/DFG consortium, CIHR project grant to JM-T, CFI, NSERC Discovery grant to JM-T, and Autism Research Chair Government of Ontario to JM-T.

ACKNOWLEDGMENTS

We would like to thank Dr. Xiao-Jing Wang and Dr. Sean Froudist-Walsh for their valuable feedback on our manuscript.

REFERENCES

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *J. Neurophysiol.* 52, 1106–1130. doi: 10.1152/jn.1984.52.6.1106

Andersen, R. A., Essick, G. K., and Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science* 230, 456–458. doi: 10.1126/science.4048942

Arnsten, A. F. T. (2013). The neurobiology of thought: the groundbreaking discoveries of Patricia Goldman-Rakic 1937–2003. *Cereb. Cortex* 23, 2269–2281. doi: 10.1093/cercor/bht195

Arnsten, A. F. T., Wang, M. J., and Paspalas, C. D. (2012). Neuromodulation of thought: flexibilities and vulnerabilities in prefrontal cortical network synapses. *Neuron* 76, 223–239. doi: 10.1016/j.neuron.2012.08.038

Baddeley, A. (2010). Working memory. *Curr. Biol.* 20, R136–R140. doi: 10.1016/j.cub.2009.12.014

Barton, J. J. (2003). Disorders of face perception and recognition. *Neurol. Clin.* 21, 521–548. doi: 10.1016/s0733-8619(02)00106-8

Bauer, R. H., and Fuster, J. M. (1976). Delayed-matching and delayed-response deficit from cooling dorsolateral prefrontal cortex in monkeys. *J. Comp. Physiol. Psychol.* 90, 293–302. doi: 10.1037/h0087996

Behrmann, M., Moscovitch, M., and Winocur, G. (1994). Intact visual imagery and impaired visual perception in a patient with visual agnosia. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 1068–1087. doi: 10.1037//0096-1523.20.5.1068

Beran, M. J., Menzel, C. R., Parrish, A. E., Perdue, B. M., Sayers, K., Smith, J. D., et al. (2016). Primate cognition: attention, episodic memory, prospective memory, self-control, and metacognition as examples of cognitive control in nonhuman primates. *Wiley Interdiscipl. Rev. Cogn. Sci.* 7, 294–316. doi: 10.1002/wcs.1397

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

Bianchi, L. (1895). The functions of the frontal lobes. *Brain* 18, 497–522. doi: 10.1093/brain/18.4.497

Bichot, N. P., Xu, R., Ghadooshahy, A., Williams, M. L., and Desimone, R. (2019). The role of prefrontal cortex in the control of feature attention in area V4. *Nat. Commun.* 10, 1–12. doi: 10.1038/s41467-019-13761-7

Bisley, J. W., Zaksas, D., Droll, J. A., and Pasternak, T. (2004). Activity of neurons in cortical area MT during a memory for motion task. *J. Neurophysiol.* 9, 286–300. doi: 10.1152/jn.00870.2003

Bisley, J. W., Zaksas, D., and Pasternak, T. (2001). Microstimulation of cortical area MT affects performance on a visual working memory task. *J. Neurophysiol.* 85, 187–196. doi: 10.1152/jn.2001.85.1.187

Blum, R. A. (1952). Effects of subtotal lesions of frontal granular cortex on delayed reaction in monkeys. *AMA Arch. Neurol. Psychiatry* 67, 375–386. doi: 10.1001/archneurpsyc.1952.02320150108012

Born, R. T., and Bradley, D. C. (2005). Structure and function of visual area MT. *Annu. Rev. Neurosci.* 28, 157–189. doi: 10.1146/annurev.neuro.26.041002.131052

Boynton, G. (2011). Spikes, BOLD, attention, and awareness: a comparison of electrophysiological and fMRI signals in V1. *J. Vis.* 11:12. doi: 10.1167/11.5.12

Brincat, S. L., Donoghue, J. A., Mahnke, M. K., Kornblith, S., Lundqvist, M., and Miller, E. K. (2021). Interhemispheric transfer of working memories. *Neuron* 109, 1055–1066.e4. doi: 10.1016/j.neuron.2021.01.016

Brozoski, T. J., Brown, R. M., Rosvold, H. E., and Goldman, P. S. (1979). Cognitive deficit caused by regional depletion of dopamine in prefrontal cortex of rhesus monkey. *Science* 205, 929–932. doi: 10.1126/science.112679

Buffalo, E. A., Fries, P., Landman, R., Liang, H., and Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *Proc. Natl. Acad. Sci. U.S.A.* 107, 361–365. doi: 10.1073/pnas.0907658106

Butters, N., and Pandya, D. (1969). Retention of delayed-alternation: effect of selective lesions of sulcus principalis. *Science* 165, 1271–1273. doi: 10.1126/science.165.3899.1271

Butters, N., Pandya, D., Sanders, K., and Dye, P. (1971). Behavioral deficits in monkeys after selective lesions within the middle third of sulcus principalis. *J. Comp. Physiol. Psychol.* 76, 8–14. doi: 10.1037/h0031037

Callicott, J. H., Bertolino, A., Mattay, V. S., Langheim, F. J., Duyn, J., Coppola, R., et al. (2000). Physiological dysfunction of the dorsolateral prefrontal cortex in schizophrenia revisited. *Cereb. Cortex* 10, 1078–1092. doi: 10.1093/cercor/10.11.1078

Campbell, R. J., and Harlow, H. F. (1945). Problem solution by monkeys following bilateral removal of the prefrontal areas. V. Spatial delayed reactions. *J. Exp. Psychol.* 35, 110–126. doi: 10.1037/h0062913

Chafee, M. V., and Goldman-Rakic, P. S. (1998). Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J. Neurophysiol.* 79, 2919–2940.

Chafee, M. V., and Goldman-Rakic, P. S. (2000). Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. *J. Neurophysiol.* 83, 1550–1566. doi: 10.1152/jn.2000.83.3.1550

Chalupa, L. M., and Werner, J. S. (2003). *The Visual Neurosciences*. Cambridge, MA: MIT Press.

Charcot, M., and Bernard, D. (1883). Un cas de suppression brusque et isole'e de la vision mentaledes signes et des objets (formes et couleurs). *Prog. Med.* 11,568–571.

Chow, K. L., Blum, J. S., and Blum, R. A. (1951). Effects of combined destruction of frontal and posterior associative areas in monkeys. *J. Neurophysiol.* 14, 59–71. doi: 10.1152/jn.1951.14.1.59

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., and Haynes, J.-D. (2017). The distributed nature of working memory. *Trends Cogn. Sci.* 21, 111–124. doi: 10.1016/j.tics.2016.12.007

Collins, C. E., Airey, D. C., Young, N. A., Leitch, D. B., and Kaas, J. H. (2010). Neuron densities vary across and within cortical areas in primates. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15927–15932. doi: 10.1073/pnas.1010356107

Constantinidis, C., Funahashi, S., Lee, D., Murray, J. D., Qi, X.-L., Wang, M., et al. (2018). Persistent spiking activity underlies working memory. *J. Neurosci.* 38, 7020–7028. doi: 10.1523/JNEUROSCI.2486-17.2018

Croxson, P. L., Kyriazis, D. A., and Baxter, M. G. (2011). Cholinergic modulation of a specific memory function of prefrontal cortex. *Nat. Neurosci.* 14, 1510–1512. doi: 10.1038/nn.2971

Dijkstra, N., Kok, P., and Fleming, S. M. (2021). Perceptual reality monitoring: neural mechanisms dissociating imagination from reality. *PsyArXiv*. doi: 10.31234/osf.io/zngeq

Duffy, C. J., and Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *J. Neurophysiol.* 65, 1329–1345. doi: 10.1152/jn.1991.65.6.1329

Duong, L., Leavitt, M., Pieper, F., Sachs, A., and Martinez-Trujillo, J. (2019). A normalization circuit underlying coding of spatial attention in primate lateral prefrontal cortex. *eNeuro* 6:ENEURO.0301–18.2019. doi: 10.1523/ENEURO.0301-18.2019

Džaja, D., Hladnik, A., Bičanič, I., Bakovič, M., and Petanjek, Z. (2014). Neocortical calretinin neurons in primates: increase in proportion and microcircuitry structure. *Front. Neuroanat.* 8:103. doi: 10.3389/fnana.2014.00103

Eccles, J. C. (1986). Mechanisms of long-term memory. *J. Physiol.* 81, 312–317.

Ester, E. F., Anderson, D. E., Serences, J. T., and Awh, E. (2013). A neural measure of precision in visual working memory. *J. Cogn. Neurosci.* 25, 754–761. doi: 10.1162/jocna00357

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

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1-a

Finan, J. (1942). Delayed response with pre-delay reinforcement in monkeys after the removal of the frontal lobes. *Am. J. Psychol.* 55, 202–214. doi: 10.2307/1417079

Forbes, N. F., Carrick, L. A., McIntosh, A. M., and Lawrie, S. M. (2009). Working memory in schizophrenia: a meta-analysis. *Psychol. Med.* 39, 889–905. doi: 10.1017/S0033291708004558

Foster, D. H. (2011). Color constancy. *Vision Res.* 51, 674–700. doi: 10.1016/j.visres.2010.09.006

Freedman, D. J., and Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature* 443, 85–88. doi: 10.1038/nature05078

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

Freiwald, W., and Tsao, D. (2012). *Taking Apart the Neural Machinery of Face Processing*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199559053.013.0036

Froudist-Walsh, S., Bliss, D. P., Ding, X., Jankovic-Rapan, L., Niu, M., Knoblauch, K., et al. (2020). A dopamine gradient controls access to distributed working memory in monkey cortex. *bioRxiv* 2020.09.07.286500. doi: 10.1101/2020.09.07.286500

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

Fuster, J. M., and Alexander, G. E. (1970). Delayed response deficit by cryogenic depression of frontal cortex. *Brain Res.* 20, 85–90. doi: 10.1016/0006-8993(70)90156-3

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

Fuster, J. M., and Bauer, R. H. (1974). Visual short-term memory deficit from hypothermia of frontal cortex. *Brain Res.* 81, 393–400. doi: 10.1016/0006-8993(74)90838-5

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

Glahn, D. C., Ragland, J. D., Abramoff, A., Barrett, J., Laird, A. R., Bearden, C. E., et al. (2005). Beyond hypofrontality: a quantitative meta-analysis of functional neuroimaging studies of working memory in schizophrenia. *Hum. Brain Mapp.* 25, 60–69. doi: 10.1002/hbm.20138

Goldman, P. S., and Rosvold, H. E. (1970). Localization of function within the dorsolateral prefrontal cortex of the rhesus monkey. *Exp. Neurol.* 27, 291–304. doi: 10.1016/0014-4886(70)90222-0

Goldman, P. S., Rosvold, H. E., Vest, B., and Galkin, T. W. (1971). Analysis of the delayed-alternation deficit produced by dorsolateral prefrontal lesions in the rhesus monkey. *J. Comp. Physiol. Psychol.* 77, 212–220. doi: 10.1037/h0031649

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

Goldman-Rakic, P. S., and Schwartz, M. L. (1982). Interdigitation of contralateral and ipsilateral columnar projections to frontal association cortex in primates. *Science* 216, 755–757. doi: 10.1126/science.6177037

Gonza'lez-Burgos, G., Barrionuevo, G., and Lewis, D. A. (2000). Horizontal synaptic connections in monkey prefrontal cortex: an *in vitro* electrophysiological study. *Cereb. Cortex* 10, 82–92. doi: 10.1093/cercor/10.1.82

Gross, C., and Weiskrantz, L. (1962). Evidence for dissociation of impairment on auditory discrimination and delayed response following lateral frontal lesions in monkeys. *Exp. Neurol.* 5, 453–476. doi: 10.1016/0014-4886(62)90057-2

Harlow, H. F., Davis, R. T., Settlage, P. H., and Meyer, D. R. (1952). Analysis of frontal and posterior association syndromes in brain-damaged monkeys. *J. Comp. Physiol. Psychol.* 45, 419–429. doi: 10.1037/h0056634

Hubel, D. H. (1957). Tungsten microelectrode for recording from single units. *Science* 125, 549–550. doi: 10.1126/science.125.3247.549

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243. doi: 10.1113/jphysiol.1968.sp008455

Iba, M., and Sawaguchi, T. (2003). Involvement of the dorsolateral prefrontal cortex of monkeys in visuospatial target selection. *J. Neurophysiol.* 89, 587–599. doi: 10.1152/jn.00148.2002

Jacob, S. N., Stalter, M., and Nieder, A. (2016). Cell-type-specific modulation of targets and distractors by dopamine D1 receptors in primate prefrontal cortex. *Nat. Commun.* 7, 13218–13211. doi: 10.1038/ncomms13218

Jacobsen, C., Wolfe, J., and Jackson, T. A. (1935). An experimental analysis of the functions of the frontal association areas in primates. *J. Nerv. Ment. Dis.* 82, 1–14. doi: 10.1097/00005053-193507000-00001

Jacobsen, C. F. (1936). Studies of cerebral function in primates. I. The functions of the frontal association areas in monkeys. *Comp. Psychol. Monogr.* 13, 1–60.

Jacobsen, C. F., and Nissen, H. W. (1937). Studies of cerebral function in primates. IV. The effects of frontal lobe lesions on the delayed alternation habit in monkeys. *J. Comp. Psychol.* 23, 101–112. doi: 10.1037/h0056632

Kim, Y., Yang, G. R., Pradhan, K., Venkataraju, K. U., Bota, M., Garc'ia Del Molino, L. C., et al. (2017). Brain-wide maps reveal stereotyped cell-type-based cortical architecture and subcortical sexual dimorphism. *Cell* 171, 456–469. doi: 10.1016/j.cell.2017.09.020

Krienen, F. M., Goldman, M., Zhang, Q., del Rosario, R. C. H., Florio, M., Machold, R., et al. (2020). Innovations present in the primate interneuron repertoire. *Nature* 586, 262–269. doi: 10.1038/s41586- 020-2781-z

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

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

Leavitt, M. L., Pieper, F., Sachs, A. J., and Martinez-Trujillo, J. C. (2017b). A quadrantic bias in prefrontal representation of visual-mnemonic space. *Cereb. Cortex* 52, 1–17. doi: 10.1093/cercor/bhx142

Lennert, T., and Martinez-Trujillo, J. (2011). Strength of response suppression to distracter stimuli determines attentional-filtering performance in primate prefrontal neurons. *Neuron* 70, 141–152. doi: 10.1016/j.neuron.2011.02.041

Lennert, T., and Martinez-Trujillo, J. C. (2013). Prefrontal neurons of opposite spatial preference display distinct target selection dynamics. *J. Neurosci.* 33, 9520–9529. doi: 10.1523/JNEUROSCI.5156-12.2013

Leopold, D. A. (2012). Primary visual cortex: awareness and blindsight. *Annu. Rev. Neurosci.* 35, 91–109. doi: 10.1146/annurev-neuro-062111-150356

Leopold, D. A., and Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379, 549–553. doi: 10.1038/379549a0

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

Lundqvist, M., Herman, P., and Miller, E. K. (2018). Working memory: delay activity, yes! Persistent activity? Maybe not. *J. Neurosci.* 38, 7013–7019. doi: 10.1523/JNEUROSCI.2485- 17.2018

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

Malmo, R. B. (1942). Interference factors in delayed response in monkeys after removal of frontal lobes. *J. Neurophysiol.* 5, 295–308. doi: 10.1152/jn.1942.5.4.295

Mendoza, D., Schneiderman, M., Kaul, C., and Martinez-Trujillo, J. (2011). Combined effects of feature-based working memory and feature-based attention on the perception of visual motion direction. *J. Vis.* 11:11. doi: 10.1167/11.1.11

Mendoza-Halliday, D., and Martinez-Trujillo, J. C. (2017). Neuronal population coding of perceived and memorized visual features in the lateral prefrontal cortex. *Nat. Commun.* 8:15471. doi: 10.1038/ncomms15471

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

Mikami, A., and Kubota, K. (1980). Inferotemporal neuron activities and color discrimination with delay. *Brain Res.* 182, 65–78. doi: 10.1016/0006-8993(80)90830-6

Mikami, A., Newsome, W. T., and Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. I. Mechanisms of direction and speed selectivity in extrastriate area MT. *J. Neurophysiol.* 55, 1308–1327. doi: 10.1152/jn.1986.55.6.1308

Miles, R. C., and Blomquist, A. J. (1960). Frontal lesions and behavioral deficits in monkey. *J. Neurophysiol.* 23, 471–484. doi: 10.1152/jn.1960.23.5.471

Miller, E. K., Li, L., and Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science* 254, 1377–1379. doi: 10.1126/science.1962197

Mishkin, M. (1957). Effects of small frontal lesions on delayed alternation in monkeys. *J. Neurophysiol.* 20, 615–622. doi: 10.1152/jn.1957.20.6.615

Mishkin, M., and Manning, F. J. (1978). Non-spatial memory after selective prefrontal lesions in monkeys. *Brain Res.* 143, 313–323. doi: 10.1016/0006-8993(78)90571-1

Mishkin, M., and Pribram, K. H. (1955). Analysis of the effects of frontal lesions in monkeys: I. Variations of delayed alternations. *J. Comp. Physiol. Psychol.* 48, 492–495. doi: 10.1037/h0040318

Mishkin, M., and Pribram, K. H. (1956). Analysis of the effects of frontal lesions in monkey. II. Variations of delayed response. *J. Comp. Physiol. Psychol.* 49, 36–40. doi: 10.1037/h0040592

Moore, T., and Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature* 421, 370–373. doi: 10.1038/nature01341

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., et al. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* 17, 1661–1663. doi: 10.1038/nn.3862

Murray, J. D., Jaramillo, J., and Wang, X.-J. (2017). Working memory and decision-making in a frontoparietal circuit model. *J. Neurosci.* 37, 12167–12186. doi: 10.1523/JNEUROSCI.0343-17.2017

Newsome, W. T., and Pare', E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.* 8, 2201–2211. doi: 10.1523/JNEUROSCI.08-06-02201.1988

Orbach, J. (1956). Immediate and chronic disturbances on the delayed response following transections of the frontal granular cortex in the monkey. *J. Comp. Physiol. Psychol.* 49, 46–51. doi: 10.1037/h0048954

Orbach, J., and Fischer, G. J. (1959). Bilateral resections of frontal granular cortex: factors influencing delayed response and discrimination performance in monkeys. *JAMA Neurol.* 1, 78–86. doi: 10.1001/archneur.1959.03840010080010

Oscar-Berman, M., Zurif, E. B., and Blumstein, S. (1975). Effects of unilateral brain damage on the processing of speech sounds. *Brain Lang.* 2, 345–355. doi: 10.1016/S0093-934X(75)80075-7

Ozdemir, A. T., Lagler, M., Lagoun, S., Malagon-Vina, H., Laszto'czi, B., and Klausberger, T. (2020). Unexpected rule-changes in a working memory task shape the firing of histologically identified delay-tuned neurons in the prefrontal cortex. *Cell Rep.* 30, 1613–1626.e4. doi: 10.1016/j.celrep.2019.12.102

Park, J. C., Bae, J. W., Kim, J., and Jung, M. W. (2019). Dynamically changing neuronal activity supporting working memory for predictable and unpredictable durations. *Sci. Rep.* 9:15512. doi: 10.1038/s41598-019-52017-8

Passingham, R. (1975). Delayed matching after selective prefrontal lesions in monkeys (*Macaca mulatta*). *Brain Res.* 92, 89–102. doi: 10.1016/0006-8993(75)90529-6

Passingham, R. E., and Wise, S. P. (2012). *The Neurobiology of the Prefrontal Cortex*. Oxford: Oxford University Press.

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

Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., et al. (1984). Neurones responsive to faces in the temporal cortex: studies of functional organization, sensitivity to identity and relation to perception. *Hum. Neurobiol.* 3, 197–208.

Petrides, M. (1995). Impairments on nonspatial self-ordered and externally ordered working memory tasks after lesions of the mid-dorsal part of the lateral frontal cortex in the monkey. *J. Neurosci.* 15, 359–375. doi: 10.1523/JNEUROSCI.15-01-00359.1995

Petrides, M. (2000). Dissociable roles of mid-dorsolateral prefrontal and anterior inferotemporal cortex in visual working memory. *J. Neurosci.* 20, 7496–7503. doi: 10.1523/JNEUROSCI.20-19-07496.2000

Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 360, 781–795. doi: 10.1098/rstb.2005.1631

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

Pribram, K. H. (1950). Some physical and pharmacological factors affecting delayed response performance of baboons following frontal lobotomy. *J. Neurophysiol.* 13, 373–382. doi: 10.1152/jn.1950.13.5.373

Pribram, K. H., Mishkin, M., Rosvold, H. E., and Kaplan, S. J. (1952). Effects on delayed-response performance of lesions of dorsolateral and ventromedial frontal cortex of baboons. *J. Comp. Physiol. Psychol.* 45, 565–575. doi: 10.1037/h0061240

Reynolds, J. H., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* 19, 1736–1753. doi: 10.1523/JNEUROSCI.19- 05-01736.1999

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

Rolls, E. T. (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Hum. Neurobiol.* 3, 209–222. doi: 10.1016/0166-4328(85)90062-2

Rosvold, H. E., and Delgado, J. M. R. (1956). The effect on delayed-alternation test performance of stimulating or destroying electrically structures within the frontal lobes of the monkey's brain. *J. Comp. Physiol. Psychol.* 49, 365–372. doi: 10.1037/h0087991

Roussy, M., Luna, R., Duong, L., Corrigan, B., Gulli, R. A., Nogueira, R., et al. (2021). Ketamine disrupts naturalistic coding of working memory in primate lateral prefrontal cortex networks. *Mol. Psychiatry* 1–16. doi: 10.1038/s41380-021-01082-5

Salazar, R. F., Dotson, N. M., Bressler, S. L., and Gray, C. M. (2012). Content-specific fronto-parietal synchronization during visual working memory. *Science* 338, 1097–1100. doi: 10.1126/science.1224000

Salzman, C. D., and Newsome, W. T. (1994). Neural mechanisms for forming a perceptual decision. *Science* 264, 231–237. doi: 10.1126/science.8146653

Sawaguchi, T., and Goldman-Rakic, P. S. (1991). D1 dopamine receptors in prefrontal cortex: involvement in working memory. *Science* 251, 947–950. doi: 10.1126/science.1825731

Sawaguchi, T., and Iba, M. (2001). Prefrontal cortical representation of visuospatial working memory in monkeys examined by local inactivation with muscimol. *J. Neurophysiol.* 86, 2041–2053. doi: 10.1152/jn.2001.86.4.2041

Schaeffer, D. J., Hori, Y., Gilbert, K. M., Joseph, S. G., Menon, R. S., and Everling, S. (2020). Divergence of rodent and primate medial frontal cortex functional connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 117, 21681–21689. doi: 10.1073/pnas.2003181117

Schwedhelm, P., Baldauf, D., and Treue, S. (2020). The lateral prefrontal cortex of primates encodes stimulus colors and their behavioral relevance during a match-to-sample task. *Sci. Rep.* 10:4216. doi: 10.1038/s41598-020-61171-3

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

Selemon, L. D., and Goldman-Rakic, P. S. (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, 4049–4068.

Spaet, T., and Harlow, H. F. (1943). Problem solution by monkeys following bilateral removal of the prefrontal areas. II. Delayed reaction problems involving use of the matching-from-sample method. *J. Exp. Psychol.* 32, 424–434. doi: 10.1037/h0058008

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

Stamm, J. S., and Weber-Levine, M. L. (1971). Delayed alternation impairments following selective prefrontal cortical ablations in monkeys. *Exp. Neurol.* 33, 263–278. doi: 10.1016/0014- 4886(71)90020-3

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

Todorović, D. (2020). What are visual illusions? *Perception* 49, 1128–1199. doi: 10.1177/0301006620962279

Tong, F., and Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annu. Rev. Psychol.* 63, 483–509. doi: 10.1146/annurev-psych-120710-100412

Torres-Gomez, S., Blonde, J. D., Mendoza-Halliday, D., Kuebler, E., Everest, M., Wang, X.-J., 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* 9:557. doi: 10.1093/cercor/bhaa056

Treue, S., and Martínez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399, 575–579. doi: 10.1038/21176

Tucker, T., and Kling, A. (1967). Differential effects of early and late lesions of frontal granular cortex in the monkey. *Brain Res.* 5, 377–389. doi: 10.1016/0006-8993(67)90045-5

Upright, N. A., Brookshire, S. W., Schnebelen, W., Damatac, C. G., Hof, P. R., Browning, P. G. F., et al. (2018). Behavioral effect of chemogenetic inhibition is directly related to receptor transduction levels in rhesus monkeys. *J. Neurosci.* 38, 7969–75. doi: 10.1523/JNEUROSCI.1422-18.2018

Uylings, H. B., Groenewegen, H. J., and Kolb, B. (2003). Do rats have a prefrontal cortex? *Behav. Brain Res.* 146, 3–17. doi: 10.1016/j.bbr.2003.09.028

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

Wang, X.-J. (2009). "A microcircuit model of prefrontal functions: ying and yang of reverberatory neurodynamics in cognition," in *The Frontal Lobes,* eds J. Risberg and J. Grafman (Cambridge: Cambridge University Press), 92–127. doi: 10.1017/CBO9780511545917.006

Wang, X.-J., Tegne'r, J., Constantinidis, C., and Goldman-Rakic, P. S. (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, 1368–1373. doi: 10.1073/pnas.0305337101

Warren, J. M., and Divac, I. (1972). Delayed response performance by rhesus monkeys with midprincipalis lesions. *Psychon. Sci.* 28, 146–148. doi: 10.3758/BF03328689

Williams, G. V., and Goldman-Rakic, P. S. (1995). Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature* 376, 572–575. doi: 10.1038/376572a0

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

Xu, Y. (2017). Reevaluating the sensory account of visual working memory storage. *Trends Cogn. Sci.* 21, 794–815. doi: 10.1016/j.tics.2017.06.013

Yang, S.-T., Wang, M., Paspalas, C. D., Crimins, J. L., Altman, M. T., Mazer, J. A., et al. (2018). Core differences in synaptic signaling between primary visual and dorsolateral prefrontal cortex. *Cereb. Cortex* 28, 1458–1471. doi: 10.1093/cercor/bhx357

Yang, W., and Yuste, R. (2017). *In vivo* imaging of neural activity. *Nat. Methods* 14, 349–359. doi: 10.1038/nmeth.4230

Yeterian, E., and Pandya, D. (2010). Fiber pathways and cortical connections of preoccipital areas in rhesus monkeys. *J. Comp. Neurol.* 518, 3725–3375. doi: 10.1002/cne.22420

Yeterian, E. H., Pandya, D. N., Tomaiuolo, F., and Petrides, M. (2012). The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex* 48, 58–81. doi: 10.1016/j.cortex.2011.03.004

Zaitsev, A. V., Povysheva, N. V., Gonzalez-Burgos, G., and Lewis, D. A. (2012). Electrophysiological classes of layer 2/3 pyramidal cells in monkey prefrontal cortex. *J. Neurophysiol.* 108, 595–609. doi: 10.1152/jn.00859.2011

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

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neural Circuits

REVIEW
published: 18 January 2022
doi: 10.3389/fncir.2021.790691

Cognitive Networks (*Cognits*) Process and Maintain Working Memory

*Joaquín M. Fuster**

University of California, Los Angeles, Los Angeles, CA, United States

Ever since it was discovered in the monkey's prefrontal cortex, persistent neuronal activity during the delay period of delay tasks has been considered a phenomenon of working memory. Operationally, this interpretation is correct, because during that delay those tasks require the memorization of a sensory cue, commonly visual. What is incorrect is the assumption that the persistent activity during the delay is caused exclusively by the retention of the sensory cue. In this brief review, the author takes the position that the neural substrate of working memory is an array of long-term memory networks, that is, of cognitive networks (*cognits*), updated and orderly activated for the attainment of a behavioral goal. In the case of a behavioral task, that activated array of *cognits* has been previously formed in long-term memory (throughout this text, the expression "long-term memory" refers to all experiences acquired after birth, including habits and so-called procedural memory, such as the learning of a behavioral task). The learning of a task is the forming of synaptic associations between neural representations of three cognitive components of the task: perceptual, motor, and reward-related. Thereafter, when needed, the composite *cognit* of the task is activated in an orderly fashion to serve working memory in the *perception-action cycle*. To make his points on a complex issue, which has been the focus of his work, and to delineate a frontier for future research, the author refers to several of his own publications and previously published reviews.

Keywords: cognits, phyletic memory, long-term-memory, perception-action cycle, delay tasks, neuroplasticity

OPEN ACCESS

Edited by:
Kenichi Ohki,
The University of Tokyo, Japan

Reviewed by:
Masamichi Sakagami,
Tamagawa University, Japan
Guilhem Ibos,
UMR7289 Institut de Neurosciences de la Timone (INT), France

***Correspondence:**
Joaquín M. Fuster
joaquinF@ucla.edu

Received: *07 October 2021*
Accepted: *13 December 2021*
Published: *18 January 2022*

Citation:
Fuster JM (2022) Cognitive Networks (Cognits) Process and Maintain Working Memory. Front. Neural Circuits 15:790691. doi: 10.3389/fncir.2021.790691

INTRODUCTION

Hughlings Jackson (1958) noted that the very same neural elements that *represent* a movement in the motor cortex are in charge of its *execution*. A similar statement can be made on sensation in the sensory cortex with regard to sensory representation and perception. Here I extend that principle to the entirety of the nervous system, from genetic "representations" (*phyletic memory*), like the anatomical structure of primary motor, sensory, and reward systems, to the representation of personal memories in the cortex of association. Memory is recalled or put to work by activation of the neural structure that represents it[1].

[1]The analogy with immune systems is remarkable. "Memory T-cells" are characterized by their long immunological memory.

According to this view, a learned delayed-response task, with all its component operations, including working memory, is represented and executed by a vast network of cortical memory that represents sensory stimuli, motor responses, and reward (or approval). Whenever the task requires the mediation of a cross-temporal contingency, as in the delay period, persistent activity links representations of temporally separate task components to bridge time across the contingency between them. If we replace the word *representation* by the word *memory*, in this as in other conditions of the organism, we may reach the conclusion that in the brain there are no systems of memory but there is the memory of systems, and working memory is the temporary activation of perceptual, executive and reward systems' memory toward a goal.

The interpretation of a cortical cell's persistent activity as a phenomenon of working memory is entirely in accord with Baddeley's (1983) basic definition of working memory: the temporary retention of information *for* a behavioral choice or the solution of a problem. Unfortunately, this future aspect of working memory, that is, its "teleonomic" aspect (Monod, 1971), is generally ignored in discussions of persistent activity in the prefrontal cortex.

One clear neural manifestation of the "teleonomic" nature of working memory is the evidence of prefrontal cells whose persistent delay activity is attuned to the animal's approaching motor response to the cue (Niki, 1974; Quintana and Fuster, 1999). Further, in prefrontal area 8 or its proximity, where a visual directional cue is integrated across a delay with a directional eye movement, persistent delay activity reveals their cross-temporal sensory-motor integration in working memory (Funahashi et al., 1989).

The presence of motion-related neurons in the prefrontal cortex is in harmony with the general notion that this cortex is involved in orderly goal-directed behavioral actions. However, the organization of all such actions requires inputs from sensory areas of the posterior cortex engaged with them in the perception-action cycle. Hence, some prefrontal cells exhibit persistent delay activity that discriminates two stimuli of different modalities—e.g., visual and auditory—if they are associated with each other (behaviorally induced "synesthesia") across the delay in the performance of a cross-modal delayed matching task (Fuster et al., 2000). Furthermore, in the expectation of a good behavioral outcome or reward, persistent activity can be observed in the orbito-medial areas of the prefrontal cortex (Moorman and Aston-Jones, 2014), which are intimately connected with limbic structures, notably the amygdala and the hypothalamus.

In contrast to the frontal cortex, vigorous sensory-discriminant delay activity can be observed in the temporal (Fuster and Jervey, 1982; Miller et al., 1993) and parietal (Zhou and Fuster, 1996) cortex. Thus we may draw the general conclusion, as others have done (Christophel et al., 2017), that cells in frontal cortices receive multiple sensory and drive-related inputs from posterior sensory cortices and limbic structures for the performance and monitoring of a working-memory task. During the delay, these multiple inputs of diverse origin, which are part of working memory and dispersed in time and cortical space, may average across trials to a semblance of persistent activity. That appearance, however, hides considerable variability from trial to trial (Shafi et al., 2007), as would be expected, though it is not yet proven, from the asynchronous convergence on the prefrontal cortex of task-related inputs from multiple sources, cortical and subcortical. Probably, that prefrontal activity during the delay is driven alternatively by several inputs from the memory of the task, including the cue, the impending motor response, and the expected reward. These inputs from multiple cognitive sub-networks (component *cognits*) upon prefrontal cell populations can be computationally considered and dealt with as *multiple attractors* (Roussy et al., 2021; Wang, 2021).

That cortical inputs are important for the maintenance of working memory and performance of a delay task is evident because the cooling of posterior cortical areas leads to a reversible deficit in the performance of delay tasks and concomitant deficits of activation of frontal cells (Quintana et al., 1989); conversely, the cooling of lateral prefrontal cortex leads to working-memory deficits and disturbance of cell activity in posterior association cortex. Both these phenomena would be manifestations of impaired frontal "cognitive control" (Miller and Cohen, 2001).

The purpose of this review is to emphasize the intimate dependence of working memory from long-term memory and to defend the hypothesis of a common anatomical substrate for both. A related purpose, largely dependent on the validity of that hypothesis, is to defend the corollary that working memory and the persistent neuronal activity that serves it are highly distributed cortical functions of the perception-action cycle.

DISTRIBUTED MEMORY

After the discovery of persistent delay activity as a neuronal manifestation of working memory (Fuster, 1973; Niki, 1974), there was a large number of single-unit studies conducted on primates during the performance of delay tasks. The principal anatomical targets of those studies were the associative areas of the frontal, parietal and temporal cortices. To this reviewer, several general facts became gradually apparent with regard to the task-related cell activity—persistent or not—during those tasks, especially in the light of observations in the human after cortical damage:

(a) In all cortical regions explored with microelectrodes, a large contingent of cells does not alter their discharge in relation to any of the delay-task components. But those cells that do, usually exhibit considerable variability from trial to trial, consistent with temporal variability in the synaptic associative inputs and outputs related to the task. That is also consistent with the expected fluctuations in the perceptual and executive attention (Amengual and Ben Hamed, 2021)

devoted by the animal to components of the delay-task habit, which in the trained animal can be safely assumed to be part of long-term memory.

(b) Cue-related activity (sensory) is most prominent in areas of the posterior association cortex, whereas choice-related (motor) activity is most prominent in the prefrontal cortex (especially lateral aspects of it). However, in prefrontal areas heavily involved in sensory-motor integration, such as in haptics (Romo et al., 1999) or in oculomotor behavior (Funahashi et al., 1989), remarkable parametric relationships have been observed between neuronal discharge and stimulus and/or motor response, always within the context of the task previously learned, thus of long-term memory.

(c) The cortical regions from which delay-task activity can be recorded have been implicated by lesion studies in the perceptual or motor memory of the task that the activity is correlated with. For instance, the inferior temporal cortex, from which persistent discriminating cells have been recorded during the delay of visual working memory (Fuster and Jervey, 1982; Miller et al., 1993), has been shown by lesion studies to be a focus of long-term memory of visual discriminations. The same can be said for the posterior parietal cortex with regard to spatial working and long-term memory. Lesions of the prefrontal cortex impair the performance of all delay tasks, as well as of other tasks that, like them, require temporal order of actions and/or the mediation of cross-temporal contingencies (Fuster, 2001). All these tasks are in the long-term memory of the trained animal.

(d) It is in the human brain where, thanks to clinical lesion studies, the most direct relations have been shown between cortical damage and memory deficit (Fuster, 1995, 2009, 2015). Thus, posterior lesions result primarily in deficits of perceptual memory (*e.g.*, agnosias, semantic aphasias, and episodic amnesias), whereas frontal lesions result primarily in deficits of executive memory and functions (*e.g.*, executive neglect, motor aphasia, and problems with executive memory, attention, and planning). In the monkey, lesions of homologs of some of the areas involved in human amnesias and other deficits lead to comparable deficits of perceptual and executive memory, including of course working memory.

The aggregate of these facts provides strong evidence for the following conclusions:

1. All the experimental phenomena of working memory, including persistent delay activity, are phenomena of the processing of the testing task, and therefore of the temporary and orderly activation of the associated components of the long-term memory of the task.
2. Persistent delay activity is an expression of the brain's necessity to transfer information across time between two or more of those components if they are mutually contingent on one another (perceptual cue, motor choice, and reward).
3. Working memory and long-term memory share the same neural substrate and mnemonic content; working memory is a portion of the long-term memory activated from its resting state and updated in order to mediate cross-temporal contingencies, and thus to conduct the subject to the goal of a task or the approval of the experimenter, or both.

CORTICAL ORGANIZATION OF MEMORY

The facts above support the general principle that working memory consists of an updated cognitive network of long-term memory selectively and orderly activated to attain a goal. Persistent activity is the prime manifestation of it when the attainment of that goal requires the reconciliation of cross-temporal contingencies between associated items of the activated network. It follows that the analysis of the cortical organization of long-term memory should help us understand the neural infrastructure of working memory and its functional dynamics. Here we need a note of caution: the debate about the neural base of memory of any kind or state is often muddled by the assumption of consciousness, ignoring the fact that memory can be unconsciously active and operative.

Because of limits in spatial and temporal resolution, current methods can only provide us with approximate estimates of the cortical regions harboring the highest densities of the most active neural elements—cells and fibers—engaged in the representation of memory, whether this is sensory, motor, emotional, or associative. Those methods, however, are clearly insufficient to define the fine grain of memory and the distribution of specific memories, in other words, what used to be called the "engrams" or "memory traces". The modern connectome reveals the connective complexity of the cortical substrate of those memories, but cannot tell us about their content any more than a roadmap can tell us about the resources or the economy of a nation. The problem is aggravated by the graded, analog, and probabilistic nature of transactions in the neural cognitive domain. A new paradigm is needed, such as the *cognit* paradigm below, to account for the microstructure and dynamics of cortical memory and cognition.

Nonetheless, as an introduction to the *cognit*, it is useful to consider the general organization of cortical memory at a mesoscopic level, as revealed by the evidence severely summarized above.

Our brain comes to the world with three inherited systems to adapt to it: sensation, motion, and emotion. The *anatomical structure* of these three systems, which in life are going to interact intimately with one another, is a form of memory that we all share and that in the course of evolution our species has acquired to deal with the physical and social environments. I call that neural structure of those three systems *phyletic memory* or "memory of the species" because it represents in the form of neural matter, genetically transmitted, the means by which the species in the "night of times" of evolution has acquired ("learned") to adapt to the environment for subsistence and procreation. Phyletic memory includes the sensory and motor systems and the limbic system, with their peripheral, subcortical, and cortical components. The organism "recalls" and "rehearses" phyletic memory with every sensation, every act, and every emotion.

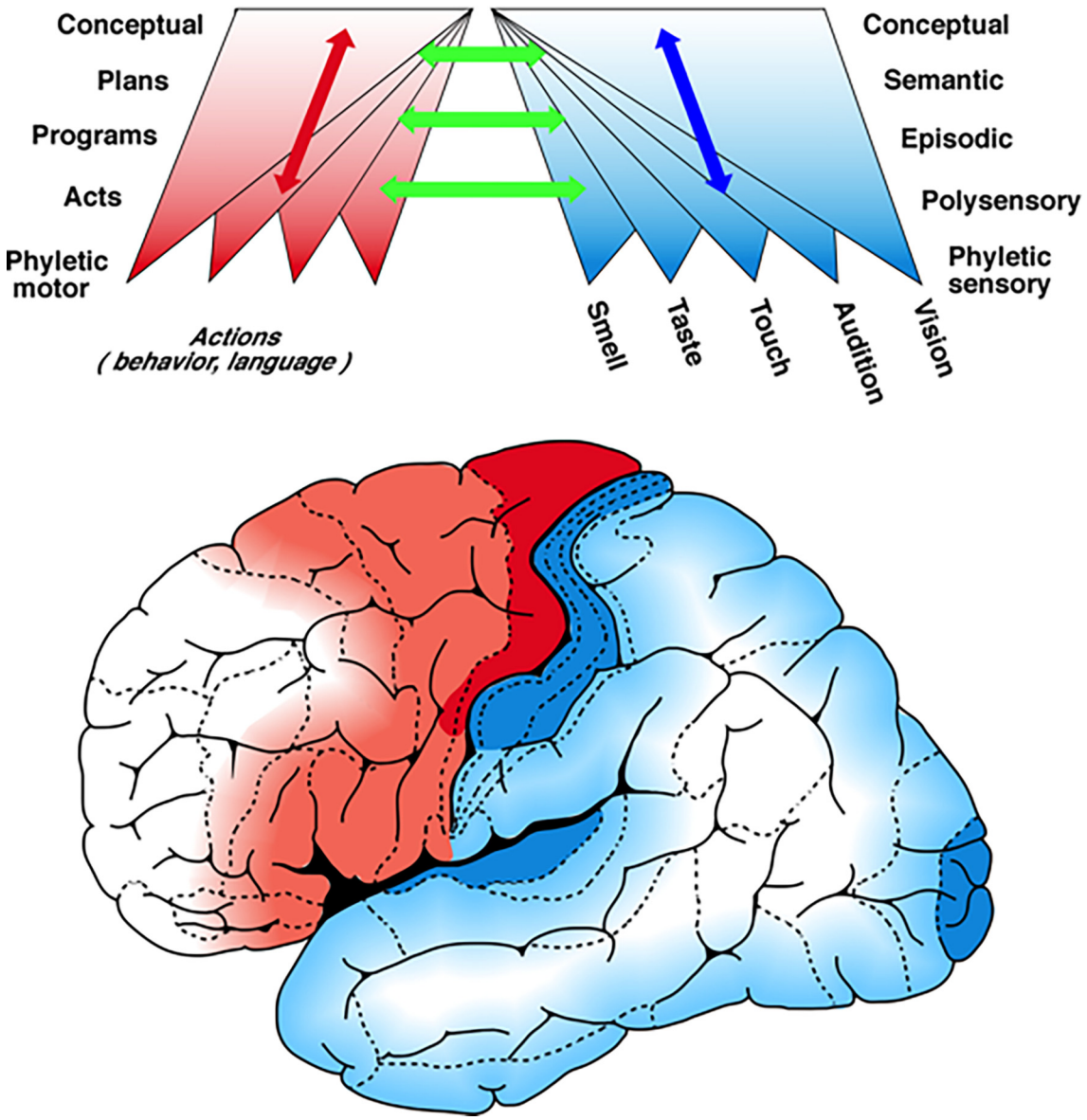


FIGURE 1 | A schema of the hierarchies of cortical memory. The figure portrays in a highly schematic manner a mesoscopic view of the distribution of memory networks. Two hierarchies of memory are shown spread-out tangentially to the cortical surface of the left hemisphere: one in posterior cortex, perceptual, for memories acquired through the senses (*blue to white)* and the other in the frontal cortex, executive, for memories acquired through action (*red to white)*. The two color gradients mark the ascending hierarchies of memory formation and deposition, from the most concrete, sensory- and motor-related, in the lower levels, to the most abstract and complex in the higher levels of both hierarchies. Up each hierarchy, as memory networks accrue and find their hierarchical niche, they connect with preexisting ones at every level by reasons of similarity and common features. As the upper figure implies, the networks of memory and knowledge (semantic memory) are formed by convergence as well as divergence of associative connections. The two bi-directional arrows, one blue and the other red indicate not only the ascending and descending connections within each hierarchy but also the heterarchical connectivity in some memories and their networks. The green arrows symbolize the connections between hierarchies that play such a critical role in the perception-action cycle. From Fuster (2015).

It is on, and from, the basic grounds of phyletic memory—that is primary sensory and motor cortices, and limbic structures—that all individual memories and knowledge will grow into association cortex to form the long-term memory and habits of the individual organism. Once formed, those memories and habits will be available to be activated *ad hoc* in working memory.

In a comprehensive review of primate single-unit studies of working memory in the visual system of primates, Roussy et al. (2021) arrive at conclusions that remarkably support the ideas expressed above about the hierarchical organization of areal memory networks. The authors conclude as I predict from my studies, that beginning with the striate cortex ("phyletic memory"), successively higher areas in a hierarchy that reaches the prefrontal cortex engage progressively less in "perception" and more in executive memory. I would object to the use of the word "perception" instead of *sensation* (perception is already individual memory), but the principle is valid: by ascending the visual hierarchy, vision becomes more memory, perceptual memory that is, and therefore more specific to the individual. This is an argument for the increase and expansion of idiosyncratic connectivity up the hierarchy (symbolized by the upward diverging cones in **Figure 1**). It is also an argument for a common substrate for working and long-term memory.

THE COGNIT

The idiosyncrasy of personal memory can only be understood by the combinatorial power of some 20 billion cortical neurons and their connections. To understand the microstructure of memory, its widespread roots and branches throughout the cerebral cortex, as well as its dynamics in retrieval and working memory, we must construct a new paradigm based on two fundamental principles of neurobiology that apply to all levels of neural cognition, from phyletic to semantic memory:

(a) Every sensation and every movement defines itself, and acquires neural function and meaning, in relation to other stimuli or movements that have been apprehended or learned together with it, whether in evolution or in the life of the individual organism. At the evolutionary level, the "elementary sensation" (Mach, 1885) does not exist (Hayek, 1952), because every sensory feature, however simple, has some spatial or temporal dimension and continuity into itself. The neurocognitive code is basically a *relational code* and all memory is associative, even at the level of the neuronal columns or groups of neurons that represent minimal sensory or motor features. Excitation and inhibition—e.g., in the retina or in antagonistic muscles—provide strength and contrast to each other: this is true, for example, in the flexors and extensors of the leg, whether in its innate defensive leg withdrawal or in normal walking. Context and background provide essential associations to define the memory of a stimulus or a movement.

(b) Even at rest, the connectivity of a cortical memory network, which links neuronal columns or groups together, is never static. Synaptic weights change with general metabolism, circadian rhythm, developmental stage, and age. The synaptic connectivity changes increase markedly with reactivation of the network in retrieval, new learning, or new experience. More generally, abrupt synaptic changes occur by engagement of the network in any kind of sustained cognitive operation, such as working memory or consolidation. All these changes take place in large and specific memory networks that join widely separated cell groups in the cerebral cortex.

Most empirical or computational models of cortical memory ignore those two principles and, in addition, the growing evidence that memory networks serve not only memory operations but also the other cognitive functions: attention, perception, language, and intelligence (Fuster, 2003). Memory, of any kind, is the essential neural substrate those functions operate on and with. Our attention, both serial and parallel, is guided by memory. In perception, we project memory on the environment, "we not only remember what we see but see what we remember" (Helmholtz, 1860). Language is essentially based on semantic memory. Intelligence makes use of all of the above plus executive memory.

It is the evidence that memory networks are the basic neural units *of all* cognitive functions that led me to the *cognit paradigm* and to rename those networks *cognits*. The new paradigm is founded on a new conceptual methodology to approach the cognitive brain, the knowing and remembering brain. Its principal new feature is a Copernican shift of the basic cognitive unit from the neuron or cortical area to the widely distributed network of cortical neurons, where association, connection, and relationship define structure and mechanisms at the microscopic level within widely distributed networks.

The cortical cognitive networks that I propose are considerably different from those in the available literature. Regardless of their empirical or theoretical base, those published networks ordinarily link together anatomically or physiologically defined areas of the cerebral cortex. Instead, my postulated networks link neuronal groups within and between multiple cortical areas, some of those groups widely separated. Here are some of its distinguishing features:

1. A *cognit* consists, in and of, a net of cortical nerve cells and the fibers and synapses that unite them. That structure contains in itself an item of memory or knowledge acquired by life experience. Cognits are exquisitely idiosyncratic, specific for each individual, differing in location, extension, and synaptic strength, depending on such factors as age, experience and training, or education.
2. The anatomical outlines *of a cognit* are diffuse and highly irregular, as it blends at its margins onto other associated cognits with weak or unstable connections. Depending on their synapse and fiber complexity, *cognits* vary considerably in size and cortical coverage. Because they share cell groups and connections representing common associated features, cognits interconnect and overlap profusely with one another.
3. A *cognit* develops out of phyletic memory—primary sensory or motor cortex—and into the associative cortex in accord with Hebbian principles (Hebb, 1949), by associations of spatial and temporal coincidence between new sensory and/or motor—proprioceptive—stimuli. In addition, those stimuli can activate, and establish connections with, pre-existing and related *cognits*, to form with them

more complex cognits, thus expanding prior memory and knowledge. This will occur under inputs and influences from the hippocampus (archicortex)—by still unclear mechanisms—and the amygdala, the latter contributing emotional connotations to the new or updated cognits.

As they develop in the course of life, the new and expanded *cognits* will occupy progressively larger and hierarchically higher areas of the associative cortex, while retaining connections with lower, nested *cognits*. Because of the unlimited possibilities of connection (combinatorial power) between cortical neurons, and because of the graded (not "all-or-none") strength of their synaptic interconnections, the higher generated *cognits* are profusely distributed over cortex, overlapping and interconnecting with one another and with the lower ones nested within them.

By virtue of the practically infinite possibilities of interconnection between cortical cell groups or modules to form a *cognit*, and the interactions between cognits, the size, location, and synaptic stability of a given cognit varies greatly over time. Plasticity under personal experience, attrition with time and age, and anatomical overlap is the norm for all *cognits* and what gives them individuality. Furthermore, because of interactions and overlaps, any cell or group of cells practically anywhere in the association cortex can be part of *many cognits*, thus many memories or items of knowledge.

The new paradigm does not supplant more conventional memory networks linking cortical areas, but it complements them with greatly magnifying "optics". Under their view, the *cognit* is individualized, much more extensive and intricate than areal networks, and it serves not only memory but also the other cognitive functions as well; hence the word *cognit*.

Recently, Fulvi Mari (2021) has published a computational model of memory retrieval in a modular associative network with an architecture extraordinarily similar to that here postulated for the *cognit*. The model suggests storage and retrieval mechanisms across different levels of a memory hierarchy of networks.

Our views of memory leave little room for the traditional classes of memory (episodic, declarative, implicit, etc.) and even less room for their anatomical location. Nonetheless, as indicated in previous sections, there is now sufficient evidence from humans and monkeys to roughly trace the cortical paths of formation of the various *cognits* and the approximate anatomical location of their foci (nodes) of heaviest associations.

In recent years, physiological animal studies have confirmed the upward trend toward higher categories of *cognits* in the perceptual hierarchy that has long been recognized by clinical studies in the human brain. That trend culminates in the prefrontal cortex, where the highest-order sensorimotor *cognits* and integrations take place (Brincat et al., 2018; Reinert et al., 2021).

Connection fibers ascending the two hierarchies, perceptual and executive, from area to area, are reciprocated every step of the way by fibers running in the opposite direction (**Figure 1**). Some fibers descend directly (through the lateral longitudinal fasciculus) from the prefrontal cortex to the cortex of association

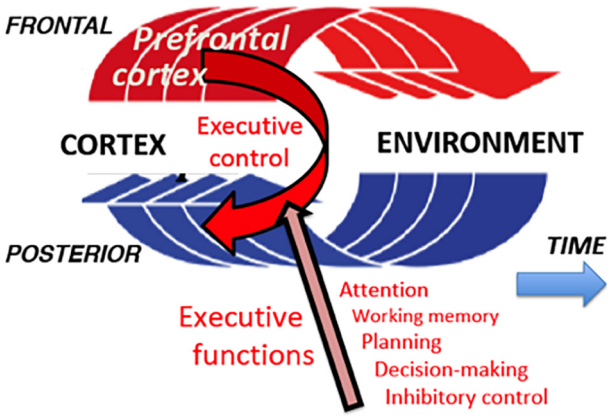


FIGURE 2 | Perception-action cycle. In a sequence of goal-directed actions, each action causes a change in the environment, which generates sensory impulses; these impulses are analyzed in the posterior association cortex (in perceptual long-term memory), and the result of this analysis informs the frontal cortex (executive long-term memory) for the next action. And so on and so forth, cycle after cycle, until the behavioral goal is reached. At every turn of a cycle, the prefrontal cortex matches percept and action to the long-term memory of both in the present context, and exerts updated executive control of them through its executive functions.

beyond sensory cortices (*e.g.*, the inferior temporal cortex). These fibers evidently engage in what has been called "cognitive control" (Miller and Cohen, 2001; Goodwin et al., 2012). Cognitive control is exerted over working memory networks and generates in them persistent activity when there are discontinuities in the perception-action cycle.

WORKING MEMORY IN THE PERCEPTION-ACTION CYCLE

The perception-action cycle is the ultimate evolutionary development into the cerebral cortex of the innate systems and mechanisms of the organism to adapt to changes in the internal and external milieus (Uexküll, 1926). The internal milieu is stabilized by the autonomic nervous system and neuroendocrine systems (homeostasis). For adaptation to the external environment, the organism is born with an array of reflex arcs in the spinal cord and mesencephalon that serve it to satisfy immediately vital needs and may be considered part of phyletic memory. At the level of the cerebral cortex, the cortical regions for adapting cognitive behavior to the physical and social environments constitute the highest substrate of the perception-action cycle. In the aggregate, this substrate forms a highly plastic and versatile system of adaptation. It is a biocybernetic system with feed-forward and feedback that governs cognitive interactions of the organism with the exterior, including such high cognitive functions as is conversational language (**Figure 2**).

In order to understand the physiological functions of the cycle, especially the role of the prefrontal cortex in it, and persistent activity in its neural circuitry, it is useful to consider certain general assumptions that derive from the human brain (Fuster, 2015):

1. The cortex of the frontal lobe is essential for the temporal organization of orderly behavior toward a goal, especially if the behavior and the goal are novel (Luria, 1966).
2. To that effect, the cortex of the frontal lobe necessitates subcortical inputs of drive and motivation, as well as the basal ganglia as outputs to accion, together with the pyramidal system.
3. The cortex of the frontal lobe, especially the lateral and medial prefrontal cortex, has a predictive, anticipatory property that allows the organism to become future-oriented, error-predictive, and "pre-adaptive."
4. The prefrontal cortex, especially its orbital region, is important for inhibitory control of distractions by interfering stimuli, impulses, and memories.

None of these global functions of the human frontal cortex is strictly specific for any frontal region in particular, and there is considerable individual variability in the dominance of any of them in any particular region. There is, however, a group of functions best identified in the prefrontal cortex of the nonhuman primate, which is somewhat topologically related to those of the human and that serves the perception-action cycle in the temporal organization of behavior. These functions (listed on the lower right of **Figure 2**) are the so-called executive functions of the prefrontal cortex, grouped under the heading of *executive control*.

Note that the first in the list is attention, a cognitive function—*not necessarily conscious*—which supports all other executive functions and consists in selectively allocating to them the limited neural resources available. The second executive function is working memory, so dependent on attention that Baddeley (1993) was inclined to consider it attention to an internal representation. The third function, planning, is attention directed to future actions, including attention to the preparation of actions in the short term. Decision-making is selective executive attention by definition. Finally, inhibitory control is also an aspect of attention, by definition, that is, the exclusionary form of attention: it is the inhibition of any source of interference, internal or external that might impede the perception-action cycle to attain its goal.

In the temporal course of the perception-action cycle toward that goal, the focus and content of attention and the role of the prefrontal executive functions over posterior—perceptual—cortical regions shifts, within the present context, from one item in long-term memory to another—updated to the present. That long-term memory can be, for example, the performance of a delay task. Naturally, in the case of a trial of such a task, the items that will attract the most attention in a given trial will be the sensory cue and the motor response, both of which will be novel *for that trial*.

The perception-action cycle can be set into motion in any of its compartments, internal or external. Examples of cycle starters would be an internal plan with a long-term objective, an emotional encounter, a biological urge, a sensory experience, or a combination of any of them. The cycle circulates through cortical memory, perceptual and executive, and through the environment. Cycle after cycle, with changing input and output, though with a consistent goal, the perception-action cycle epitomizes what could be characterized as the adaptive dynamic infrastructure of the cortex.

Some parts of the cycle that are constant and repetitive, such as the habitual actions in every trial of a delay task, circulate through the cortex and, in addition, through subcortical reflex arcs, including the basal ganglia (Daw et al., 2005). The task itself is represented by a high cortical *cognit* and its parts by nested subordinate *cognits* at lower hierarchical levels. All are sequentially recruited and activated under the cognitive control of the prefrontal cortex, which ensures order and guidance to the sequence. But the sequence of active *cognits* is essentially self-generated and self-organized by association. Thus the cognits were initially formed by association, and now by association are sequentially activated in the perception-action cycle. Accordingly, the perception of the cue at the beginning of a delay trial is an act of *recall*, which by association will evoke successive cycles in every trial to attain its reward.

In the enforced delay of any delay task, the short-term memory of the cue and the prospective memory of the response will dominate, the first in sensory association cortex and the second in the prefrontal cortex, both probably maintained and mutually reinforced by reverberating activity between the active cognits of the two cortical regions. Those activated *cognits* will be the two main sources of persistent activity in those cortical regions.

Figure 3 is the result of a *graphic meta-analysis* of a large number of functional neuroimaging studies of human subjects performing visual delay tasks[2], thus in the perception-action cycle. During the delay, when persistent activity—in averages or single trials—is most likely to occur, activation is seen simultaneously in visual association cortex and lateral prefrontal cortex. As the delay progresses, the prefrontal activation grows and advances toward the motor cortex, anticipating the choice-response. The joint activation of the prefrontal and infero-temporal components of the network representing the task, with their loop of persistent activity, serves as a bridge of working memory at the top of the perception-action cycle.

Ascribing to the prefrontal cortex the "seat" of executive control with its five executive functions (**Figure 2**) is supported by a massive amount of data. However, this fact may lead to the mischaracterization of the prefrontal cortex as the "central executive" or the "center of will". It is neither, even though it mediates executive functions, free choice, and creativity. Indeed, to give to our prefrontal cortex the role of the autonomous origin of all our decisions and actions leads inevitably to an infinite regress that should be avoided ("What agency controls the prefrontal cortex? What other agency controls that one?"...and so on *ad infinitum*). The only reasonable solution to the quandary is to place the prefrontal cortex in the perception-action cycle, where the action can originate anywhere, including the cerebral cortex, prefrontal or other.

[2]Because of the limitations of neuroimaging, especially in temporal resolution, as well as differences in the time scales of the studies analyzed (some of them meta-analyses of others), the time course of activations has been estimated from unit studies in the monkey. Because of the large number of studies analyzed, the problem of reverse inference (Poldrack, 2006) is presumably avoided.

Thus the prefrontal cortex does not escape Jackson's principle: the same neural structure harboring the memory of an action is in charge of its execution. So, we have to pose ourselves two questions: what kind of memories does the prefrontal cortex hold for the long-term? What kinds of cognitive networks does it hold that, under certain circumstances, can become executive networks by entering the perception-action cycle? The answer to both questions seems to be that the prefrontal cortex contains memory networks (*cognits)* of plans of action or a series of goal-directed actions, ready to be activated in working memory. Here we are adding an essential parameter of prefrontal memories and *cognits*: time (Fuster, 2001). Associations of timing and order encoded in those networks, pace and time the executive functions that lead those activated and operational networks to their objective. What's more, the prefrontal cortex can create within itself new *cognits*, new memories out of old ones, thus predicting, imagining and creating future actions (Ingvar, 1985; Addis et al., 2007; Fuster and Bressler, 2015), in addition to bridging cross-temporal contingencies with working memory and with persistent activity of its cognitive neuronal networks or *cognits* of the cerebral cortex at large.

Although it can evoke and create a new action, the prefrontal cortex cannot execute it without the intimate cooperation of the other cortical and subcortical participants in the perception-action cycle. A suitable analogy for that cortex would be that of both composer of the music and director of the orchestra.

DISCUSSION

The *cognit* paradigm would have to be rejected if it were shown by reliable methods that a personal memory, a percept, or a sequence of organized action were localized in its entirety in a discrete portion of the cortex. Such evidence would negate the essentially distributed character of a *cognit* and working memory, in addition to the critical phenomenon of perceptual constancy ("a rose is a rose, is a rose," regardless of size, color, aroma, or position in my visual field).

The new paradigm stems in part from the failure of all forms of cerebral *localizationism* of memory. Also, it is an attempt to substantiate by neuroscientific methodology four classic theories of cognition: associationism, Gestalt psychology, connectionism, and the Cajal-Hebb synaptic theory. Despite their shortcomings, all four theories share important properties with the *cognit*, and therefore, of the substrate of working memory I postulate.

Associationism, the psychological doctrine introduced in ancient times by Aristotle and widely advocated by British empiricists (17th–18th centuries), reduces all mental life to associations between mental states, ideas, sensations, reflexes, etc.., dividing the mind into components but ignoring its unity and the functional relations between those components. One exceptionally useful concept of associationism is that of the association between sensation and memory in perception ("we remember what we see, and see what we remember") With regard to memory, associationism fails to recognize its hierarchical organization, and of course, there is no place in it for heterarchical associations.

Associationism and connectionism clearly accommodate the concepts of synapsis and neural network that serve cognition in the brain. Any neurophysiological analysis of cognition based on them, however, must deal with neural signals that are for the most part analog and probabilistic, like firing frequency and field potentials. These are not the most convenient signals for models and machines in the field of artificial intelligence. Nonetheless, connectionist neural-network algorithms applied to language can discover certain categories of grammatical rules based on similarities, much as the *cognit* paradigm can uncover hierarchical categories of language (McClelland and Rumelhart, 1986).

Gestalt psychology (Koffka, 1935) provides the most proximate property to semantic memory, and with it to high-category *cognits*: an object is defined by the relations between its parts, not by the parts themselves, and certainly not by the sum of those parts. However, as in the case of connectionism, the meaning of a Gestalt ("structure") lies in both the relation and the related elements. Given the practically infinite combinatorial power of over 20 billion neurons, the potential variability of human memories is immense, like that of human experience. There are, however, constraints to that variability dictated by anatomy and physiology. One is the innate connectivity of the individual brain. Another is the strength of synaptic connections. These constraints exist at all levels of the cognitive hierarchy, but are presumably more stringent at their higher levels, where semantic knowledge and global action are constituted by convergent affluences from lower, nested, and more concrete *cognits*. Hence, by assumed foci of synaptic strength and fiber convergence, it seems legitimate to grossly delineate the relative position of the various categories of knowledge and memory on the cortical surface (**Figure 1**).

All three psychological theories mentioned in support of the *cognit* paradigm have the most reasonable neurobiological foundation in the synaptic principles of memory formation first proposed by (Cajal, 1894) and (Hebb, 1949). These principles culminate with the idea of the "neuronal assembly," which is the theoretical precursor of the *"cognit"*, though the latter applies to all cognitive functions, not just memory. Further, with regard to memory, Hebb's concepts are based on circuitry mostly circumscribed to the visual and the parastriate cortex, whereas the *cognit* extends to association cortex of all sensory and motor systems. Both conceptions, Cajal's and Hebb's fail to explain the role of the hippocampus, decisive but still poorly understood till now, on the formation of new neocortical memory.

In recent years, the *cognit* paradigm has found some support in the latest investigations of cortical connectionism with the most advanced techniques available. Among the latest initiatives based on those techniques is the *connectome*, the international research program to expose the entire connectivity of the human brain. This effort has led to exquisite maps of cortical connectivity, spectacular for its richness, but so far it has not helped us much to reveal cortical neuroplasticity, one of the objectives of the program. Functional resting-state magnetic resonance imaging (fMRI) is another promising method (Taren et al., 2011).

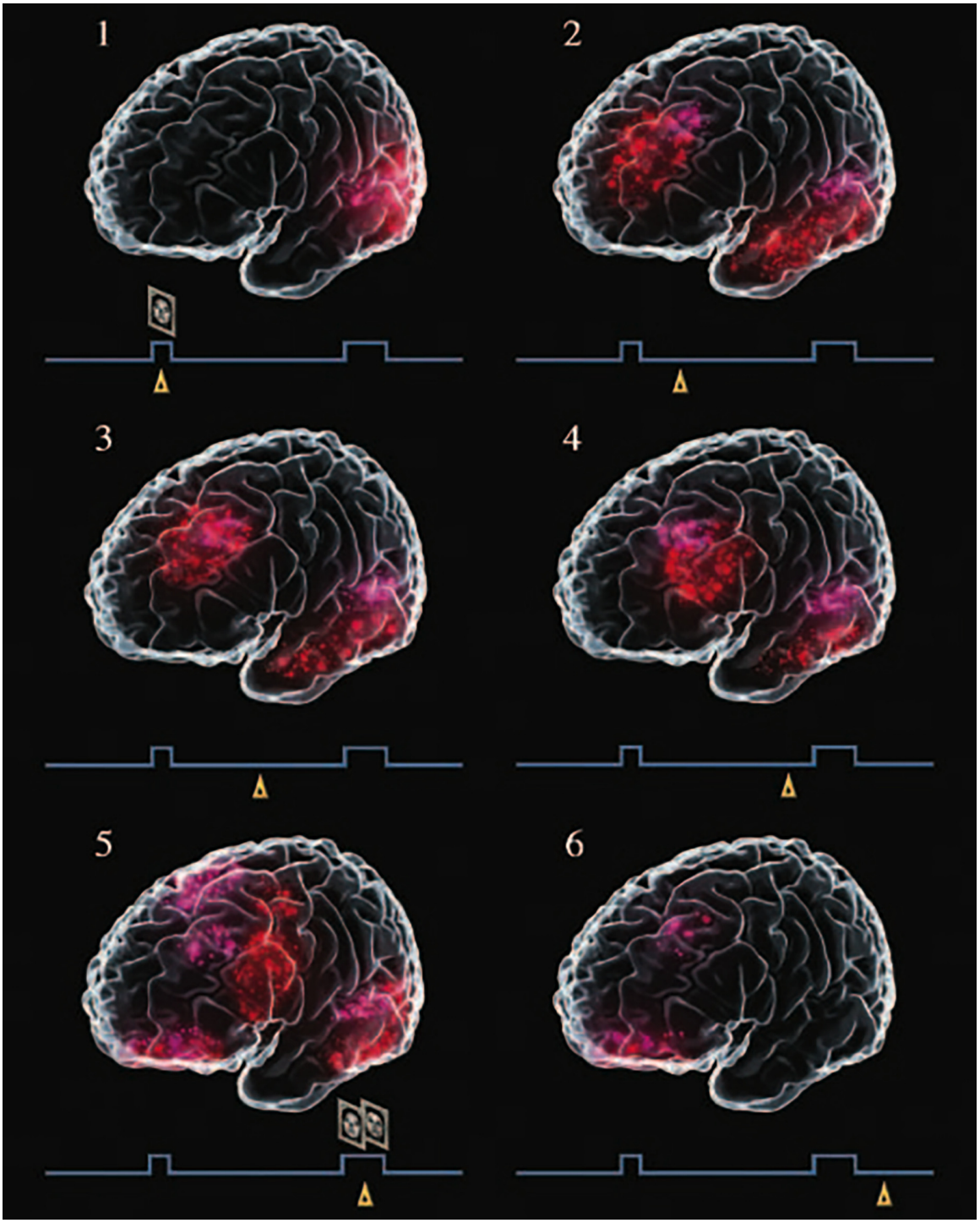


FIGURE 3 | Graphic meta-analysis of cortical activations in the course of a delayed matching-to-sample task, paradigmatic of visual delay tasks in a large number of functional neuroimaging studies. A trial begins with the presentation of a face (sample), which the subject must remember for a delay of 10–15 s, at the end of which the subject is presented with two faces and must choose the sample. Little triangles mark the approximate relative timing of the records in the course of a trial. Note the activation of the visual cortex at the sample (excerpt 1), and of the inferior temporal and prefrontal cortices during the delay (excerpts 2–6). From Fuster (2015).

Future research should be devoted to obtaining better spatial and temporal resolution than we now have of cerebral processes in active memory. The dependency of working memory from long-term memory could be supported by utilizing—in working memory tasks—stimuli (cues) that activated different levels of the memory hierarchy. The critical question would be if the same

cortical hierarchy of memory would be evinced by simple recall as by working memory of sensory stimuli of differing levels. This would further confirm my parsimonious proposal of an identical neural substrate for both conditions of activated memory.

Another issue for future research is the predictive and prospective executive functions of the prefrontal cortex, such as planning, executive attention, and working memory, in the acquisition of memory and knowledge. Training children in those functions is the key to the success of *active learning*, the educational method that capitalizes on the initiative, creativity, and cooperativeness of the child. This method is at the foundation of the most successful modern systems of elementary education, such as the Finnish system.

CONCLUSION

The presence in the primate brain of a system for long-term memory and another for working memory is at odds with all the pertinent empirical evidence. Instead, a massive body of experimental and clinical evidence indicates that working memory consists of the temporary activation of an updated cortical network of long-term memory for the attainment of an objective. That accords with the general principle of this review: under appropriate circumstances, any memory of the organism, from the biological to the most abstract, can become operational in behavior, reasoning, and in the spoken or written language. Working memory is operational memory by definition and the epitome of that principle. Its most elementary substrate is a cortical network of long-term memory, here called *cognit*, formed between neurons by associations according to Hebbian principles. A *cognit* is specific for a given individual; in working memory, it is updated for present context The dynamics of working memory can best be examined and understood in the perception-action cycle, the biocybernetic loop that engages the organism with its environment in goal-directed behavior. Working memory bridges with persistent activity in widely distributed cortical networks any temporal break or discontinuity that may occur in the cycle before reaching its goal.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Addis, D. R., Wong, A. T., and Schacter, D. L. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia* 45, 1363–1377. doi: 10.1016/j.neuropsychologia.2006.10.016

Amengual, J. L., and Ben Hamed, S. (2021). Revisiting persistent neuronal activity during covert spatial attention. *Front. Neural Circuits* 15:679796. doi: 10.3389/fncir.2021.679796

Baddeley, A. (1983). Working memory. *Phil. Trans. R. Soc. Lond. B.* 302, 311–324.

Baddeley, A. D. (1993). "Working memory or working attention? In attention: selection, awareness and control," in *A Tribute to Donald Broadbent*, eds A. D. Baddeley, and L. Weiskrantz (Oxford: Clarendon Press), 152–170

Brincat, S. L., Siegel, M., von Nicolai, C., and Miler, E. K. (2018). Gradual progression from sensory to task-related processing in cerebral cortex. *Proc. Natl. Acad. Sci. U S A* 115, E7202–E7211. doi: 10.1073/pnas.1717075115

Cajal, S. R. (1894). The croonian lecture: la fine structure des centres nerveux. *Proc. R. Soc. Lond.* 55, 444–468. doi: 10.1098/rspl.1894.0063

Christophel, T.B., Klink, P. C., Spitzer, B., Roelfsema, P. R., and Haynes, J.-D. (2017). The distributed nature of working memory. *Trends Cogn. Sci.* 21, 111–124. doi: 10.1016/j.tics.2016.12.007

Daw, N. D., Yael Niv, D., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi: 10.1038/nn1560

Fulvi Mari, C. (2021). Memory retrieval dynamics and storage capacity of a modular network model of association cortex with feature decomposition. *Biosystems* 211:104570. doi: 10.1016/j.biosystems.2021.104570

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

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

Fuster, J. M. (1995). *Memory in the Cerebral Cortex.* Cambridge, MA: MIT Press.

Fuster, J. M. (2001). The prefrontal cortex - an update: time is of the essence. *Neuron* 30, 319–333. doi: 10.1016/s0896-6273(01)00285-9

Fuster, J. M. (2003). *Cortex and Mind: Unifying Cognition.* New York: Oxford University Press.

Fuster, J. M. (2009). Cortex and memory: emergence of a new paradigm. *J. Cogn. Neurosci.* 21, 2047–2072. doi: 10.1162/jocn.2009.21280

Fuster, J. M. (2015). *The Prefrontal Cortex.* 5th Edn. Amsterdam: Elsevier Academic Press. doi: 10.1016/C2012-0-06164-9

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

Fuster, J. M., and Bressler, S. L. (2015). Past makes future: role of pFC in prediction. *J. Cogn. Neurosci.* 27, 639–654. doi: 10.1162/jocn_a_00746

Fuster, J. M., and Jervey, J. P. (1982). Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. *J. Neurosci.* 2, 361–375. doi: 10.1523/JNEUROSCI.02-03-00361.1982

Goodwin, S. J., Blackman, R. K., Sakellaridi, S., and Chafee, M. V. (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, 3499–3515. doi: 10.1523/JNEUROSCI.3585-11.2012

Hayek, F. A. (1952). *The Sensory Order.* Chicago, IL: University of Chicago Press.

Hebb, D. O. (1949). *The Organization of Behavior: a Neuropsychological Theory.* New York: Wiley.

Helmholtz, H. v. (1860). *Treatise on Physiological Optics.* New York: Dover.

Ingvar, D. H. (1985). "Memory of the future": an essay on the temporal organization of conscious awareness. *Hum. Neurobiol.* 4, 127–136.

Jackson, J. H. (1958). *Selected Writings.* New York: Basic Books.

Koffka, K. (1935). *Principles of Gestalt Psychology.* New York: Harcourt Brace and Company.

Luria, A. R. (1966). *Higher Cortical Functions in Man.* Boston: Springer.

Mach, E. (1885). *Die Analyse der Empfindungen.* Jena: G. Fisher.

McClelland, H. A., and Rumelhart, D. E. (1986). *Parallel Distributed Processing.* Cambridge, MA: MIT Press.

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167

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

Monod, J. (1971). *Chance and Necessity.* New York: Knopf.

Moorman, D., and Aston-Jones, G. (2014). Orbitofrontal cortical neurons encode expectation-driven initiation of reward-seeking. *J. Neurosci.* 34, 10234–10246. doi: 10.1523/JNEUROSCI.3216-13.2014

Niki, H. (1974). Differential activity of prefrontal units during right and left delayed response trials. *Brain Res.* 70, 346–349. doi: 10.1016/0006-8993(74)90324-2

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data?. *Trends Cogn. Sci.* 10, 59–63. doi: 10.1016/j.tics.2005.12.004

Quintana, J., and Fuster, J. M. (1999). From perception to action: temporal integrative functions of prefrontal and parietal neurons. *Cereb. Cortex* 9, 213–221. doi: 10.1093/cercor/9.3.213

Quintana, J., Fuster, J. M., and Yajeya, J. (1989). Effects of cooling parietal cortex on prefrontal units in delay tasks. *Brain Res.* 503, 100–110. doi: 10.1016/0006-8993(89)91709-5

Reinert, S., Hubener, M., Bonhoeffer, T., and Goldstein, P. M. (2021). Mouse prefrontal cortex represents learned rules for categorization. *Nature* 593, 411–417. doi: 10.1038/s41586-021-03452-z

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

Roussy, M., Mendoza-Halliday, D., and Martínez-Trujillo, J. C. (2021). Neural substrates of visual percepcion and Working memory: two sides of the same coin or two different coins? *Front. Neural Circuits* 15:764177. doi: 10.3389/fncir.2021.764177

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

Taren, A. A., Venkatraman, V., and Huettel, S. A. (2011). A parallel functional topography between medial and lateral prefrontal cortex: evidence and implications for cognitive control. *J. Neurosci.* 31, 5026–5031. doi: 10.1523/JNEUROSCI.5762-10.2011

Uexküll, J. V. (1926). *Theoretical Biology*. New York: Harcourt, Brace & Co.

Wang, X.-J. (2021). 50 Years of persistent activity: quo vadis? *Trends Neurosci.* 44, 888–902. doi: 10.1016/j.tins.2021.09.001

Zhou, Y., and Fuster, J. M. (1996). Mnemonic neuronal activity in somatosensory cortex. *Proc. Natl. Acad. Sci. U S A* 93, 10533–10537. doi: 10.1073/pnas.93.19.10533

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership