A DOPAMINE GRADIENT CONTROLS ACCESS TO DISTRIBUTED WORKING MEMORY IN LARGE-SCALE MONKEY CORTEX

A PREPRINT

Sean Froudist-Walsh[1], Daniel P. Bliss[1], Xingyu Ding[1], Lucija Rapan[2], Meiqi Niu[2], Kenneth Knoblauch[3,4], Karl Zilles[2,+], Henry Kennedy[3,4,5], Nicola Palomero-Gallagher[2,6,*], and Xiao-Jing Wang[1,6,7,*]

[1]Center for Neural Science, New York University, New York, NY 10003, USA

[2]Research Center Jülich, INM-1, Jülich, Germany

[3]INSERM U846, Stem Cell & Brain Research Institute, 69500 Bron, France

[4]Université de Lyon, Université Lyon I, 69003 Lyon, France

[5]Institute of Neuroscience, State Key Laboratory of Neuroscience, Chinese Academy of Sciences (CAS), Key Laboratory of Primate Neurobiology CAS, Shanghai, China

[6]co-senior authors

[7]lead contact

[+]deceased 26 April 2020

[*]correspondence, XJW: xjwang@nyu.edu, NPG: n.palomero-gallagher@fz-juelich.de

July 20, 2021

Summary

Dopamine is required for working memory, but how it modulates large-scale cortex is unknown. Here we report that dopamine receptor density per neuron, measured by autoradiography, displays a macroscopic gradient along the macaque monkey cortical hierarchy. This gradient is incorporated in a connectome-based large-scale cortex model endowed with multiple neuron types. The model captures an inverted-U-shaped dependence of working memory on dopamine, and spatial patterns of persistent activity observed over 90 experimental studies. Moreover, we show that dopamine is crucial for filtering out irrelevant stimuli by enhancing inhibition from dendrite-targeting interneurons. Our model revealed that an activity-silent memory trace can be realized by facilitation of inter-areal connections, and adjusting cortical dopamine induces a switch from an internal memory state to active persistent activity. Taken together, our work represents a cross-level understanding from molecules and cell types to recurrent circuit dynamics underlying a core cognitive function distributed across the primate cortex.

Keywords dopamine · working memory · large-scale brain model · parvalbumin · calbindin · somatostatin · calretinin · VIP · interneurons · activity-silent · short-term synaptic plasticity · persistent activity · distributed working memory

Introduction

Our ability to think through difficult problems without distraction is a hallmark of cognition. When faced with a constant stream of information, we must keep certain information in mind and protect it from distraction. For instance, when at the supermarket looking for your favorite butter, it is important to keep in mind its distinctive golden packaging, and not be distracted by the many other dairy products. This brain function is called working memory. Working memory often engages persistent neural activity that is specific to the information that must be remembered. This mnemonic activity is internally sustained across multiple cortical and subcortical areas in the absence of external stimulation (Funahashi et al. 1989; Fuster and Alexander 1971; Guo et al. 2017; Leavitt et al. 2017; Mendoza-Halliday et al. 2014; Romo et al. 1999; Romo and Salinas 2003; Vergara et al. 2016; Wang 2001; Zhang et al. 2019).

Working memory and the prefrontal cortex are under the influence of monoaminergic modulation (Goldman-Rakic 1995; Robbins and Arnsten 2009). In fact, depletion of dopamine from the prefrontal cortex and complete ablation of the prefrontal cortex cause similar working memory deficits (Brozoski et al. 1979). Dopamine modulates cortical activity through its receptors. D1 receptors are the most densely expressed dopamine receptor type in cortex. Prefrontal neuron activity during working memory depends on precise levels of activation of the dopamine D1 receptors, with both too little and too much D1R stimulation disrupting delay period activity (Vijayraghavan et al. 2007; Wang et al. 2019). However, the density of D1 receptors is known only for relatively small sections of monkey cortex (Goldman-Rakic et al. 1990; Impieri et al. 2019; Lidow et al. 1991; Niu et al. 2020; Richfield et al. 1989). Due to the shortage of areas analysed across studies, it is not clear if variation in D1 receptor densities across cortical areas represents random heterogeneity or a systematic gradient of cortical dopamine modulation.

Dopamine receptors are also differently expressed across different types of inhibitory neuron (Mueller et al. 2019, 2018). Distinct inhibitory cell types primarily focus their inhibition on the dendrites or soma of pyramidal cells, or on other inhibitory neurons (Jiang et al. 2015; Tremblay et al. 2016). Through its differing effects on distinct interneurons, dopamine decreases inhibition to the soma of pyramidal cells, and increases inhibition to the dendrites (Gao et al. 2003). An early theoretical study proposed that inhibition targeted more strongly towards the dendrites, and away from the soma of pyramidal cells could increase the resistance of working memory to distraction (Wang et al. 2004a). The functional significance of dopamine's differential effects on distinct inhibitory neuron types has not yet been investigated.

In this work, we tackled two open questions. First, how does dopamine modulate distributed working memory across a multi-regional large-scale cortical system? Second, in light of an emphasis on cell types in modern cortical physiology, does dopamine contribute to robust working memory against distractors by virtue of differential impacts on different neuron classes? To address these questions, we performed quantitative mapping of dopamine D1 receptor densities across 109 cortical areas using in-vitro autoradiography and constructed a large-scale computational model of macaque cortex that is capable of performing working memory tasks. The model is built using retrograde tract-tracing connectivity data and incorporates gradients of D1 receptors and excitatory synapses. Moreover, to our knowledge this is the first large-scale cortex model endowed with three subtypes of inhibitory neurons. Our results suggest that the firing of dopamine neurons can engage distractor-resistant stimulus-selective sustained activity across multiple brain regions in response to behaviorally-relevant stimuli. Furthermore, we extend, from a local area to a multi-regional cortex, an activity-silent state mechanism that has been proposed for certain forms of short-term memory trace without persistent

activity (Mongillo et al. 2008; Rose et al. 2016; Wolff et al. 2017). We found that this scenario relies principally on short-term facilitation of inter-areal connections, but fails to resist distractors. An enhanced dopamine modulation can convert an internal memory trace to an active persistent activity state needed to filter out distractors. Therefore, our findings contribute to resolving the current debate about the two contrasting scenarios that contribute to working memory (Constantinidis et al. 2018; Lundqvist et al. 2018; Watanabe and Funahashi 2014), and under what conditions which mechanism is implemented (Barbosa et al. 2020; Masse et al. 2019; Trübutschek et al. 2019).

Results

A hierarchical gradient of dopamine D1 receptors per neuron across monkey cortex

We first analyzed D1 and D2 receptor distribution patterns throughout the macaque brain using *in-vitro* receptor autoradiography (Fig. S1). Autoradiography enables the quantification of endogenous receptors in the cell membrane through the use of radioactive ligands (Niu et al. 2020; Palomero-Gallagher and Zilles 2018; Rapan et al. 2021). The highest densities (in fmol/mg protein) of both receptor types were found in the basal ganglia, with the caudate nucleus (D1 298 ± 28; D2 188 ± 30) and putamen (D1 273 ± 40; D2 203 ± 37) presenting considerably higher values than the internal (D1 97 ± 34; D2 22 ± 12) or external (D1 55 ± 16; D2 30 ± 11) subdivisions of the globus pallidus. Raw cortical D1 receptor densities ranged from 49 ± 13 fmol/mg protein in area 4a of the primary motor cortex to 101 ± 35 fmol/mg protein in orbitofrontal area 11l (Fig 1A). The density of the D2 receptor in cortex is so low that it is not detectable with the method used here.

To compare the gradient of D1 receptors to other known gradients of anatomical organization in monkey cortex, we carefully mapped the receptor data (Fig 1A), as well as data on neuronal density (Fig 1B) (Collins et al. 2010) and spine count (Fig 1C) (Elston 2007) onto the Yerkes19 common cortical template, to which anatomical tract tracing data (Fig 1D i) has previously been mapped (Donahue et al. 2016). Here we include retrograde tracing data from 40 regions, quantified using the same protocol as in previous publications with fewer injected regions (Markov et al. 2014b). We estimated the cortical hierarchy using laminar connectivity data (Fig 1D ii, Methods; Markov et al. 2014a), expanding previous descriptions of the cortical hierarchy based on fewer regions (Markov et al. 2014a; Mejias et al. 2016). A one-dimensional hierarchy is probably an oversimplification of the cortical connectivity structure. As we have connectivity data for two distinct sensory modalities, we also calculated a circular embedding of the connectivity data, with radial distance from the edge representing the hierarchical position and angular distance between points representing the inverse of their connectivity strength (Chaudhuri et al. 2015). In this circular representation, separate visual and somatosensory hierarchies can clearly be appreciated, with association regions falling at angles off the main sensory hierarchy axes (Fig 1E).

To facilitate functional interpretation, we divided the D1 receptor density by the neuron density (Collins et al. 2010), to allow estimation of the degree to which dopamine modulates individual neurons across cortex. D1 receptor density per neuron peaked in the parietal and frontal cortex, and was relatively low in early sensory cortex (Fig 1F). There was a strong positive correlation between the D1 receptor density per neuron and the cortical hierarchy (Fig 1G; r = 0.81). Due to spatial autocorrelation between cortical features (i.e. nearby parts of cortex tend to have similar anatomy), it is possible to detect spurious correlations between distinct features of brain anatomy. To account for

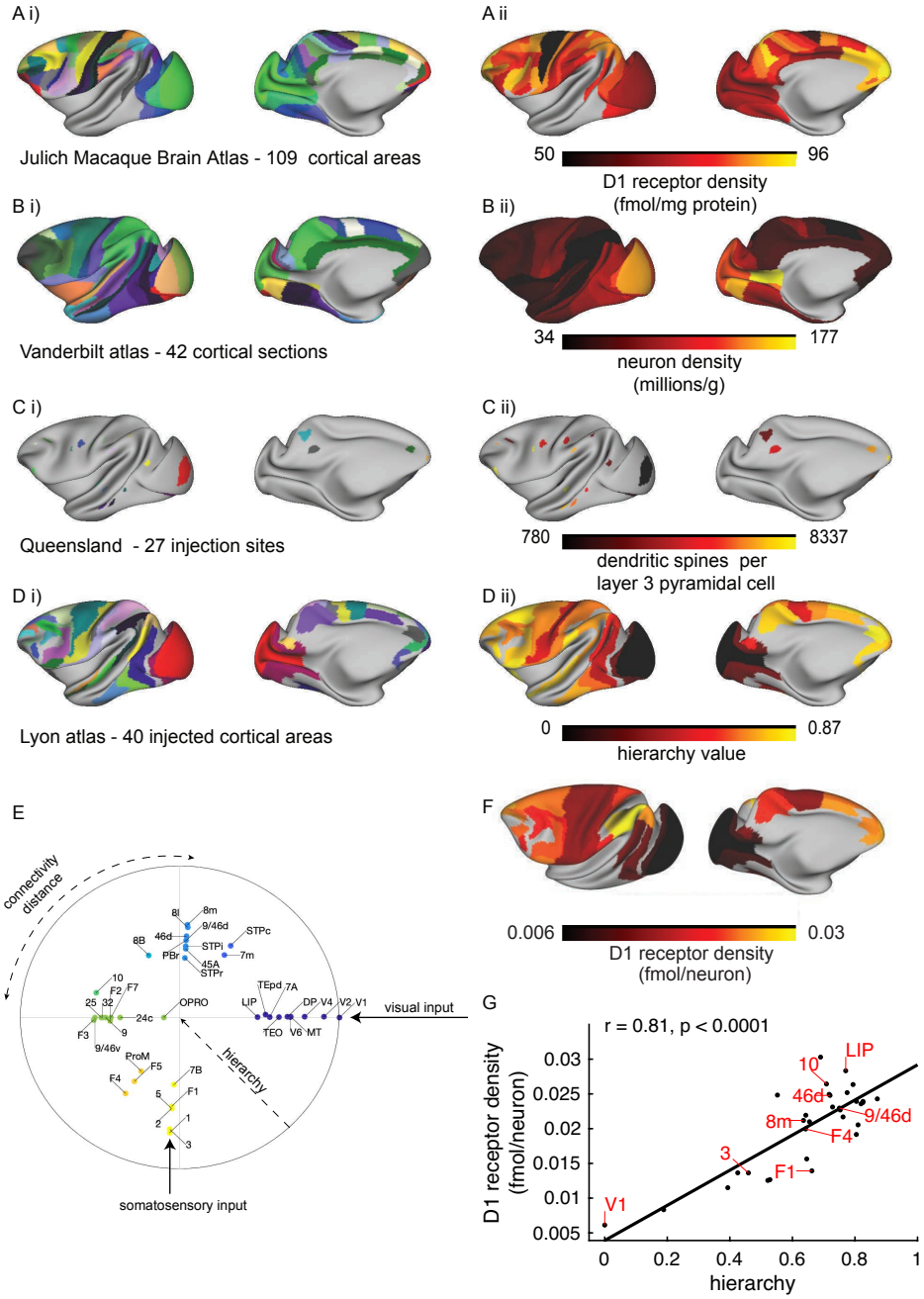

Figure 1: A gradient of dopamine D1 receptors per neuron across monkey cortex. A i) 109 cortical regions of the Julich Macaque Brain Atlas, identified by receptor and cytoarchitecture. A ii) D1-receptor density. Note that the receptor density shown here does not take into account differences in neuron density across areas. B i) Collins et al., (2010) divided the macaque cortex into 42 slabs of tissue, here mapped onto the Yerkes19 surface. B ii) Neuron density across cortex. C i) Injection sites for the studies of dendritic spine density by Elston and colleagues. C ii) Number of dendritic spines on the basal dendrites of layer III pyramidal cells. D i) 40 injected areas in the retrograde tract-tracing database of Kennedy and colleagues (Markov et al. 2014b). D ii) Cortical hierarchy. E) A circular embedding of the cortical hierarchical connectivity structure. Radial distance to the center represents the hierarchical position of the area, with the areas lowest in the hierarchy closest to the edge. Angular distance between areas represents the inverse of connectivity strength (FLN), so that areas that are plotted at similar angles are more strongly connected to each other. Colors represent the angle on the circle. Note that clear visual and somatosensory hierarchies emerge from this circular embedding of the connectivity data (highlighted with arrows). Association areas lie at angles off the main visual and somatosensory hierarchies. F) The density of D1 receptors divided by neuron density. Regions that have not yet been measured shown in gray. G) There was a strong positive correlation between the D1 receptor density per neuron and the cortical hierarchy.

this, we generated 10,000 surrogate maps with similar spatial autocorrelation to the hierarchy map (Burt et al. 2020). None of these surrogate maps were as strongly correlated with the D1 receptor density map as the hierarchy, giving a p-value < 0.0001 for the D1 receptor - hierarchy correlation. There was no significant relationship between D1 receptor expression and whether a cortical area had a granular layer IV (Wilcoxon rank-sum $Z = 0.39$, $p = 0.70$), or to the degree of externopyramidalisation (Kruskall-Wallis Chi-sq = 1.47, $p = 0.48$, Goulas et al. 2018; Sanides 1962, Fig S2). This pattern of receptor expression suggests that dopamine principally modulates areas contributing to higher cognitive processing.

A cortical circuit with three types of inhibitory neurons modulated by dopamine

We built a model of a local cortical circuit which contains pyramidal cells and three types of inhibitory neurons (Fig 2A). The cortical circuit is based on a disinhibitory motif that was originally predicted theoretically (Wang et al. 2004a), with details of the connectivity structure chosen to reflect recent experimental findings (Methods).

In our model, dopamine acted by increasing the synaptic strength of inhibition to the dendrite, and reducing the synaptic strength of inhibition to the cell body of pyramidal cells (Fig 2B) (Gao et al. 2003). In addition, dopamine increased the strength of transmission via NMDA receptors (Seamans et al. 2001). On the other hand, high stimulation of D1 receptors resulted in increased adaptation in excitatory cells (potentially an M-current, via KCNQ potassium channels; Arnsten et al. 2019) , mimicking the net inhibitory effect of high concentrations of D1-agonists.

A large-scale model of macaque cortex incorporating multiple macroscopic gradients

We then built a large-scale model of macaque cortex. We placed the local circuit in each of the 40 cortical areas across macaque cortex (Fig 2A, right). Properties of these local circuits varied across areas in the form of macroscopic gradients (Wang 2020) of long-distance connectivity (set by tracing data), strength of excitation (set according to the spine count) and modulation by D1 receptors (set according to the receptor autoradiography data). We defined the connections between areas using the quantitative retrograde tract-tracing data. In the model, inter-areal connections are excitatory and target the dendrites of pyramidal cells (Petreanu et al. 2009). Inter-areal excitatory connections also target CR/VIP cells to a greater degree than PV or CB/SST cells (Lee et al. 2013; Wall et al. 2016). The frontal eye fields (FEF) have an unusually high density of calretinin (here CR/VIP) cells (Pouget et al. 2009). To account for this, we increased the proportion of inter-areal input to CR/VIP cells in FEF and reduced the strength of input to the PV and CB/SST cells.

An inverted-U relationship between cortical D1 receptor stimulation and distributed working memory activity

We simulated the large-scale cortical model during performance of a working memory task (Fig 2C) with different levels of cortical dopamine availability. In simulations, stimulus-selective activity propagated from visual cortex to temporal, parietal and frontal cortex. Activity in visual cortex was relatively insensitive to dopamine (Fig 2E,F). Dopamine modulation had little-to-no effect on the initial peak of activity in early visual areas, but it did modulate the later peak of activity in these areas (Fig S3), consistent with a specific role of feedback connections in late visual activity (Self et al. 2012). In all cases, there was a strong transient response in visual areas, prior to a rapid return to baseline firing rates. This is similar to the response seen in neurons recorded from area V1 in behaving monkeys (Van Vugt et al. 2018).

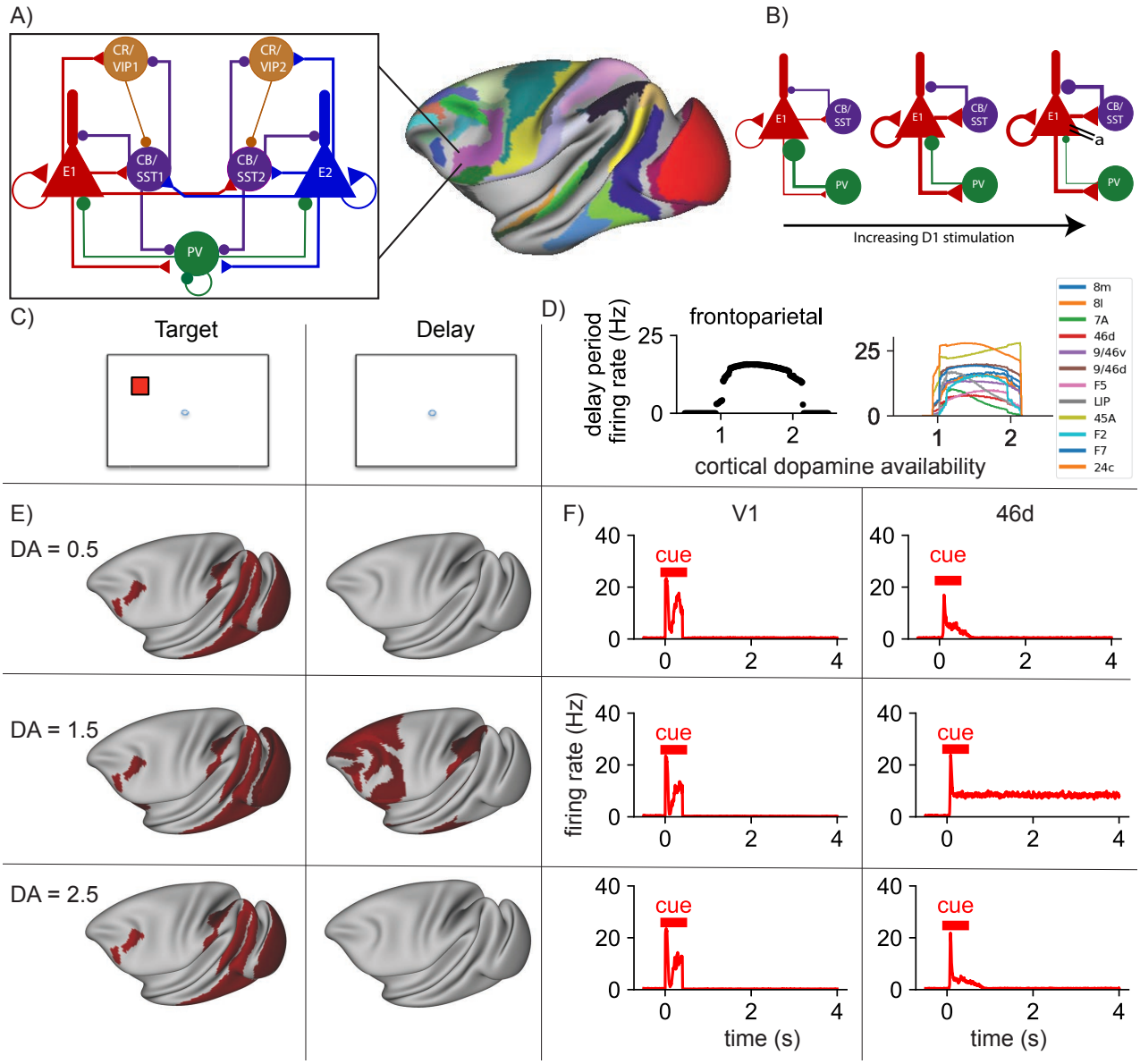

Figure 2: An inverted-U relationship between D1 receptor stimulation and distributed frontoparietal delay-period activity. A, left) Local circuit design. The circuit contains two populations of excitatory cells (red and blue), each selective to a particular spatial location. The cell bodies (triangles) and dendrites (cylinders) are modeled as separate compartments. PV cells (green), CB/SST (purple) and CR/VIP cells (light brown) have characteristic connectivity patterns. A, right) The local circuit is placed at each of 40 cortical locations (various colours). Cortical areas differ in 1) the inter-areal connections, 2) the spine count and 3) the dopamine D1 receptor density. B) Stimulation of D1 receptors affects the cortical circuit via 1) an increase of inhibition targeting the dendrites, with a corresponding decrease in inhibition to the soma of pyramidal cells, 2) an increase in NMDA-dependent excitatory transmission for low-to-medium levels of stimulation and 3) increasing adaptation for high levels of stimulation. C) Structure of the task. The cortical network was presented with a stimulus, which it had to maintain through a delay period. D, left) Mean firing rate in the frontoparietal network at the end of the delay period, for different levels of dopamine release. D, right) Mean delay-period activity of cortical areas as a function of dopamine release. All areas shown display persistent activity in experiments (Leavitt et al. 2017). E) Activity is shown across the cortex at different stages in the working memory task (left to right), with increasing levels of dopamine release (from top to bottom). Red represents activity in the excitatory population sensitive to the target stimulus. Very low or very high levels of dopamine release resulted in reduced propagation of stimulus-related activity to frontal areas and a failure to engage persistent activity. Mid-level dopamine release enables distributed persistent activity. F) Timecourses of activity in selected cortical areas. The horizontal bars indicate the timing of cue (red) input to area V1. DA, cortical dopamine availability.

We observed similar transient activity in somatosensory areas in response to stimulus input to somatosensory cortex (Fig S4), as seen experimentally (Romo and Rossi-Pool 2020). Delay-period activity in a large network of prefrontal, lateral parietal and temporal areas showed an inverted-U relationship with dopamine levels (Fig 2D). A mid-range level of dopamine release engaged a distributed pattern of persistent activity throughout these areas (Fig 2E,F), but too low or too high release led to only a transient response (Fig 2F). A similar pattern of delay period activity was observed following somatosensory input (Fig S4). The inverted-U relationship between D1 receptor stimulation and working memory activity has been shown locally in prefrontal cortex in experimental and computational studies (Brunel and Wang 2001; Vijayraghavan et al. 2007), but has not previously been described throughout the distributed cortical system.

Inter-areal connectivity determines the distributed working memory activity pattern

We next compared the pattern of delay-period activity in the model to delay-period activity observed in over 90 electrophysiology studies (Leavitt et al. 2017). Of the 19 cortical areas in which such activity has been assessed during the delay period in at least three papers, 18 were in agreement between the simulation and experimental results ($\chi^2 = 15.03, p = 0.0001$ Fig 3A). Overall, the experimentally observed persistent activity from numerous studies is reproduced, validating the model. This allows us to inspect the anatomical properties that underlie the distributed activity pattern and gain insight into the brain mechanisms that may produce it.

We repeated model simulations after shuffling the anatomical data. The delay period activity patterns for 30,000 simulations based on the shuffled anatomy were compared to the pattern observed experimentally. Ten thousand simulations were run using shuffled inter-areal connections, shuffled D1 receptor expression and shuffled dendritic spine expression, separately. The overlap between the experimental persistent activity pattern and the model persistent activity pattern was strongly dependent on the pattern of inter-areal connections (p=0.0004), but not on the pattern of D1 receptors (p = 0.71) or dendritic spine count (p = 0.46) (Fig 3B). This analysis suggests that the inter-areal connectivity is important for defining the spatial pattern of delay-period activity, but tells us little about how individual cortical areas differentially contribute to distributed working memory.

Working memory deficits are most severe following lesions to prefrontal areas with high D1 receptor density

We next quantified the degree to which focal lesions to individual areas in the model disrupted persistent activity during the working memory task (without distractors). The effect depended on both the area lesioned, and the level of cortical dopamine (Fig 3C). Lesions to prefrontal and posterior parietal areas caused the greatest reductions in delay-period firing rates (Figure 3D,E). Lesions to frontal areas caused a significantly greater reduction in delay-period firing rates than lesions to parietal areas (Mann Whitney U = 46.0, p = 0.027). We tested the effects of progressively larger lesions to frontal and parietal cortex. To increase the size of the lesions, for each lobe, we first lesioned the area that caused the biggest drop in delay activity when lesioned individually, and then additionally lesioned the area that caused the second biggest drop, and so on (Frontal lesion 1: 46d, lesion 2: 46d+8B, lesion 3: 46d+8B+8m etc.; Parietal lesion 1: LIP, lesion 2: LIP+7m, lesion 3: LIP+7m+7B etc.). On lesioning two frontal regions, mnemonic delay period activity was completely destroyed throughout the cortex, so the network was no longer able to perform the task. In contrast, progressively larger lesions of parietal cortex caused only a gradual decrease in frontoparietal delay activity, and even

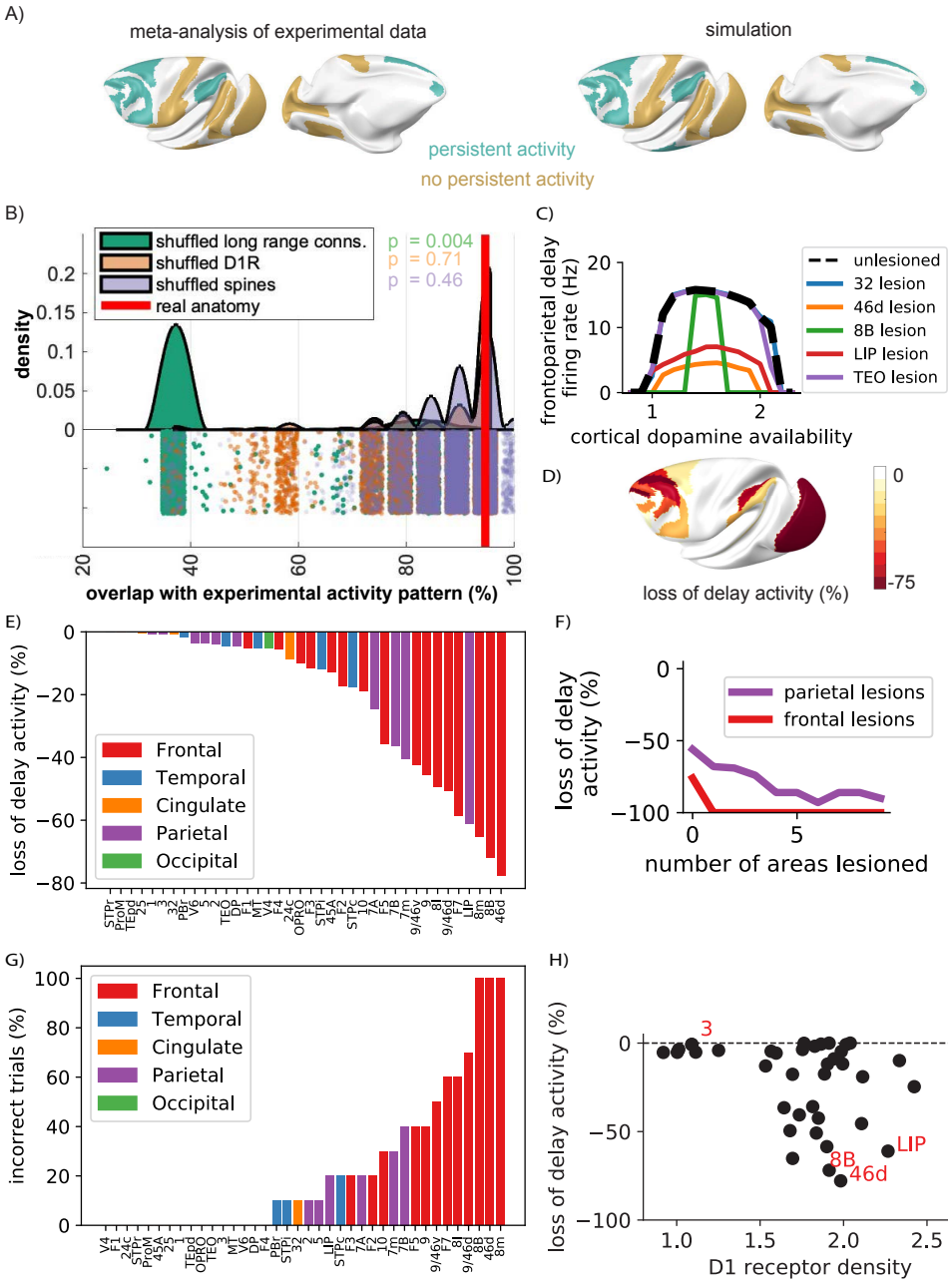

Figure 3: Inter-areal connectivity and D1 receptor density underlie working memory activity and performance. A) There is a strong overlap (18/19 - 95%) between the pattern of persistent activity seen experimentally (Leavitt et al. 2017) and that predicted by the model. B) The results of 10,000 simulations using shuffled inter-areal connections (green), 10,000 simulations using shuffled patterns of D1 receptor expression (orange) and 10,000 simulations using shuffled patterns of dendritic spine counts (purple). The position on the x axis denotes the overlap between the simulated delay activity pattern and the experimental activity pattern identified by Leavitt and colleagues for each simulation based on shuffled anatomical data. The red vertical line denotes the overlap between the simulation based on the real anatomy data and the experimental results. The bottom half of the image shows the results of individual simulations based on shuffled anatomical data. The top half of the image shows the densities. The pattern of inter-areal connections was the most important determinant of the working memory activity pattern. C) Lesions to areas such as 46d and LIP led to reduced delay period firing across for all levels of dopamine release. Following some lesions (such as to area 8B) an optimal level of D1 receptor stimulation could restore close-to-normal working memory activity in the remaining network. D) The level of disruption to distributed working memory activity following lesions to each area, quantified as the total loss of working memory activity in the frontoparietal network summed across all dopamine release levels. E) The percent loss of delay period activity throughout the cortex following a lesion to each area. F) The percent loss of delay period activity following progressively bigger lesions to frontal and parietal areas. G) The percent of failed trials, across all dopamine levels, on a working memory task with a distractor following lesions to each cortical area. H) Lesions to areas with a higher D1 receptor density tended to have a larger impact on working memory activity. D1R, D1 receptor density.

when the entire parietal cortex was removed (10 areas), sufficient residual mnemonic delay-period activity remained to allow the cue stimulus to be decoded (Figure 3F).

We subsequently addressed the ability of the model to maintain cue-specific delay period activity in the presence of distractors following precise lesioning of each cortical area. We analysed trials across all levels of cortical dopamine availability. Lesions to three prefrontal areas (8m, 46d, 8B), but not other areas, caused a complete disruption of distractor-resistant working memory activity in all trials. Lesions to many other areas caused a complete reduction of distractor-resistant working memory activity for some trials (corresponding to a particular dopamine range) but not others. The seven lesions causing the greatest disruption of working memory performance were in frontal cortex (six prefrontal areas, and premotor area F7, Fig 3G). The reduction in performance was significantly greater for lesions to frontal cortical areas than parietal areas (Mann Whitney $U = 48.5$, $p = 0.032$). Our simulations thus suggest that 1) lesions to both prefrontal and posterior parietal cortex can cause a significant disruption of delay period activity, 2) frontal lesions cause a greater effect on behavior than parietal lesions, and 3) smaller lesions, particularly to prefrontal cortex, can significantly disrupt performance on more difficult working memory tasks, such as those with distractors. In contrast, larger lesions are required to disrupt performance on simple working memory tasks.

Lesions to area V1 and V2 led to a complete loss of visual working memory activity (Fig 3D). However, this was due to the fact that a visual stimulus must go through area V1 in order to gain access to the working memory system. We confirmed this by showing that lesions to V1 and V2 had no effect on working memory when somatosensory stimuli were used (with stimulus presented to primary somatosensory area 3). In the somatosensory working memory task, lesions to early somatosensory areas and frontoparietal network areas caused memory deficits (Fig S5). This clearly separates early sensory areas, which are required for signal propagation to the working memory system, from core cross-modal working memory areas in prefrontal and posterior parietal cortex.

D1 receptor density ($F = 4.72$, $p = 0.036$, Fig 3H) was the strongest anatomical predictor of the lesion effects, and adding hierarchy or spine count to the model did not significantly improve the fit. Thus, our model predicts that lesions to areas with a higher D1 receptor density are more likely to disrupt working memory activity. This prediction can be tested experimentally.

Dopamine shifts between activity-silent and persistent activity modes of working memory

Recent experimental and modeling results show that some delay tasks can be solved with little or no persistent activity (Mongillo et al. 2008; Rose et al. 2016; Watanabe and Funahashi 2014; Wolff et al. 2017). This has spurred a debate about whether persistent activity or 'activity-silent' mechanisms underlie working memory (Constantinidis et al. 2018; Lundqvist et al. 2018). Is dopamine modulation throughout the cortex relevant to this debate? We endowed the model with short-term plasticity to assess the possibility of activity-silent working memory in the large-scale network. Short-term plasticity was implemented at all synapses between excitatory cells (using the same parameters as Mongillo et al. 2008), and from excitatory to CB/SST cells. We investigated activity-silent representations by 'pinging' the system, with a neutral stimulus, and reading out the activity generated in response, similar to the experimental protocol in Wolff et al. 2017 (Fig 4A i). For optimal mid levels of dopamine release (Fig 2A ii), the model generated persistent activity that was very similar to the network without short-term plasticity. The strong and distributed activation of frontal and parietal cortex is reminiscent of the ignition response to consciously observed stimuli (Van Vugt et al. 2018).

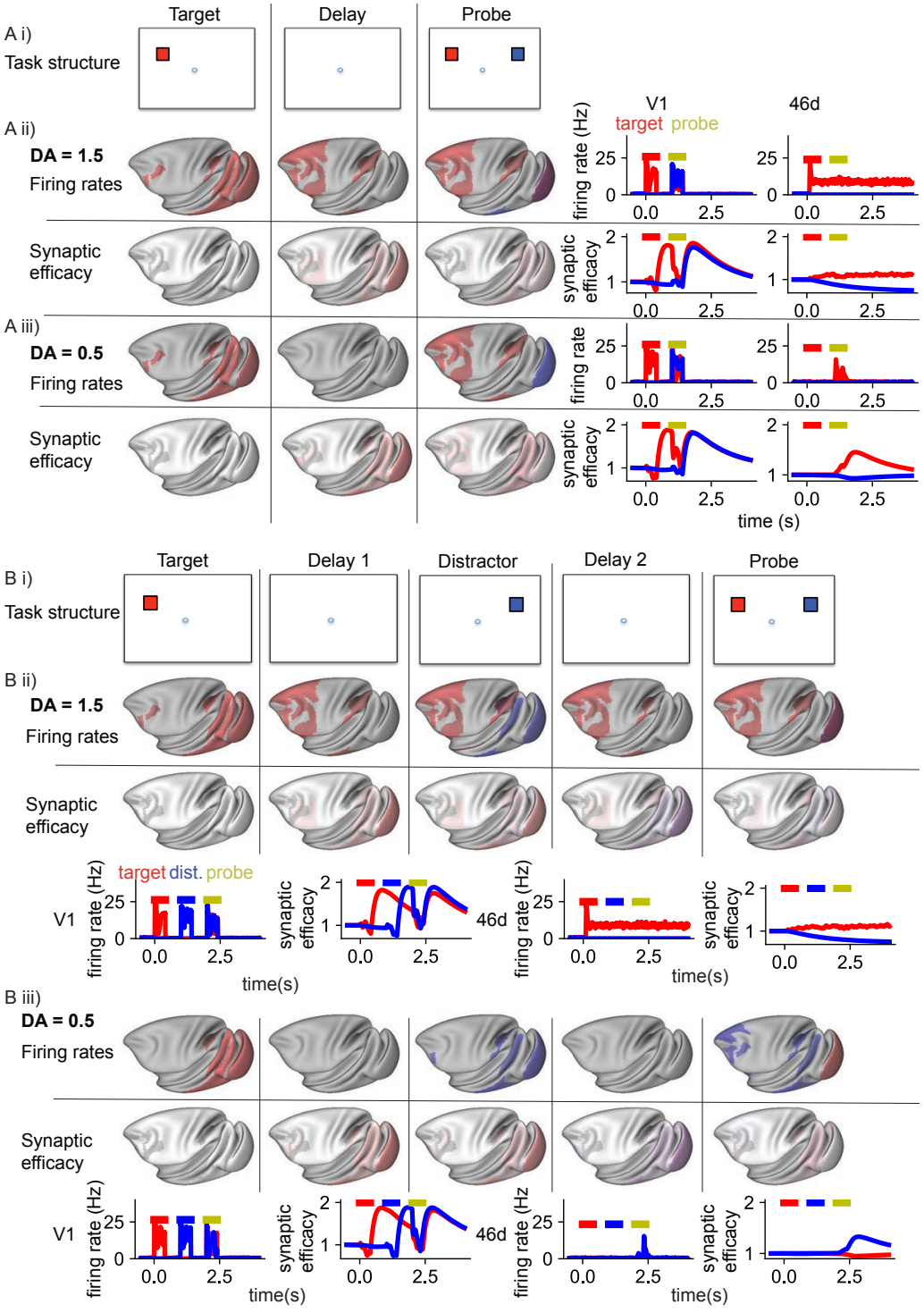

Figure 4: A dopamine-dependent shift between distractible activity-silent and distractor-resistant persistent-activity states. A i) Task structure. A target stimulus was followed by a delay and a probe stimulus. A ii) For mid-level dopamine release, activity relating to the target stimulus propagated from V1 through the hierarchy, and was maintained in persistent activity throughout the frontoparietal network. Top: firing rates on the surface (left) and in selected areas (right). Bottom: synaptic efficacy. A iii) For low-level dopamine release, activity (Top) in response to the stimulus was transient in visual and some frontoparietal areas. There was no persistent activity through the delay period. However, in response to the probe stimulus, activity representing the original target stimulus was regenerated throughout frontoparietal cortex. Bottom: The memory for the stimulus was stored as an increase in synaptic efficacy through the delay period, mostly in connections from sensory areas. B i) Task structure. A target stimulus was followed by a delay period, a distractor, another delay period and a probe stimulus. B ii) For mid-level dopamine release, target-related activity was maintained in persistent activity throughout the frontoparietal network, throughout the delay period through the distractor until the end of the trial. B iii) For low-level dopamine release, frontoparietal activity related to the most recent stimulus (i.e. the distractor) was regenerated during this probe stimulus. DA, cortical dopamine availability.

For both low and high levels of dopamine release there was no persistent activity (Fig 4 A iii). However, when we 'pinged' the system with a neutral stimulus, activity relating to the target cue was transiently generated throughout the frontoparietal network (Fig 4 A iii), suggesting that a memory of the target stimulus was stored internally. During the delay period, the synaptic efficacy increased at connections between neurons coding for the target stimulus. Previous models of activity-silent short-term memory have focused on local synaptic changes in the prefrontal cortex (Mongillo et al. 2008). In our model most of the increase in synaptic efficacy was in synaptic connections from neurons in sensory areas (Fig 4 A iii). When we restricted short-term synaptic plasticity to presynaptic neurons outside the frontoparietal network, pinging the system still resulted in activation of the target-related activity throughout the frontoparietal network (Fig S6). In contrast, when we restricted short-term synaptic plasticity to presynaptic neurons inside the frontoparietal network, pinging the system did not lead to activation of the frontoparietal network (Fig S6). This suggests that synaptic plasticity at connections from (presynaptic) prefrontal cortical neurons is not required for activity-silent memory. When we restricted short-term plasticity to local connections, activity-silent memory recall again failed (Fig S6). This suggests that short-term facilitation in inter-areal feedforward connections from early sensory areas to frontal and parietal cortex is a potential substrate for 'activity-silent' memory in the absence of a strong initial prefrontal response to the stimulus.

Why does the brain have two parallel systems for holding items in short-term memory? To explore this question, we simulated the model using a ping protocol (Wolff et al. 2017) with a distractor. After a behaviorally relevant cue and during the delay period, we introduced a distractor which should be filtered out by the network, followed by a neutral ping stimulus (Fig 4B i). For mid-level dopamine release, persistent activity coding for the target stimulus is engaged, and maintained through the distractor and ping (Fig 4B ii). The distractor is transiently represented in IT and LIP (thus replicating the experimental results in Suzuki and Gottlieb 2013), but does not reach most of the frontoparietal network. In the low and high dopamine cases, during the ping, the activity-silent mechanism regenerates activity related to the last encoded stimulus, namely the distractor, in frontal and parietal cortex (Fig 4B iii). Thus, pinging from the activity-silent state scenario always recalls the latest item but cannot ignore a distractor. Therefore dopamine release may serve to encode salient items in working memory, and protect them from distraction.

Dopamine increases distractor resistance by shifting the subcellular target of inhibition

How does dopamine protect working memory from distraction? To examine this question, we analysed activity within CR/VIP and CB/SST neurons during a working memory task with a distractor (Fig 5A). CB/SST and CR/VIP neurons are in competition, as they mutually inhibit each other. When CB/SST cell firing is higher, the pyramidal cell dendrites are relatively inhibited. Conversely when CR/VIP cell firing is higher, the pyramidal cell dendrites are disinhibited. Each cortical area in the model contains two selective populations of pyramidal, CB/SST and CR/VIP cells. We first analysed trials in which the model successfully ignores the distractor. In the target-selective populations, the CR/VIP neurons fire at a much higher rate than the CB/SST neurons (Fig 5B, C). Thus, the dendrites of the pyramidal cells sensitive to the target stimulus are disinhibited, allowing inter-areal target-related activity to flow between cortical areas. In the distractor sensitive populations, throughout the frontoparietal network, the CB/SST neurons fire at a slightly higher rate than the CR/VIP cells. Thus, activity from other cortical areas is blocked from entering the dendrites of distractor-sensitive pyramidal cells in frontal and parietal cortex.

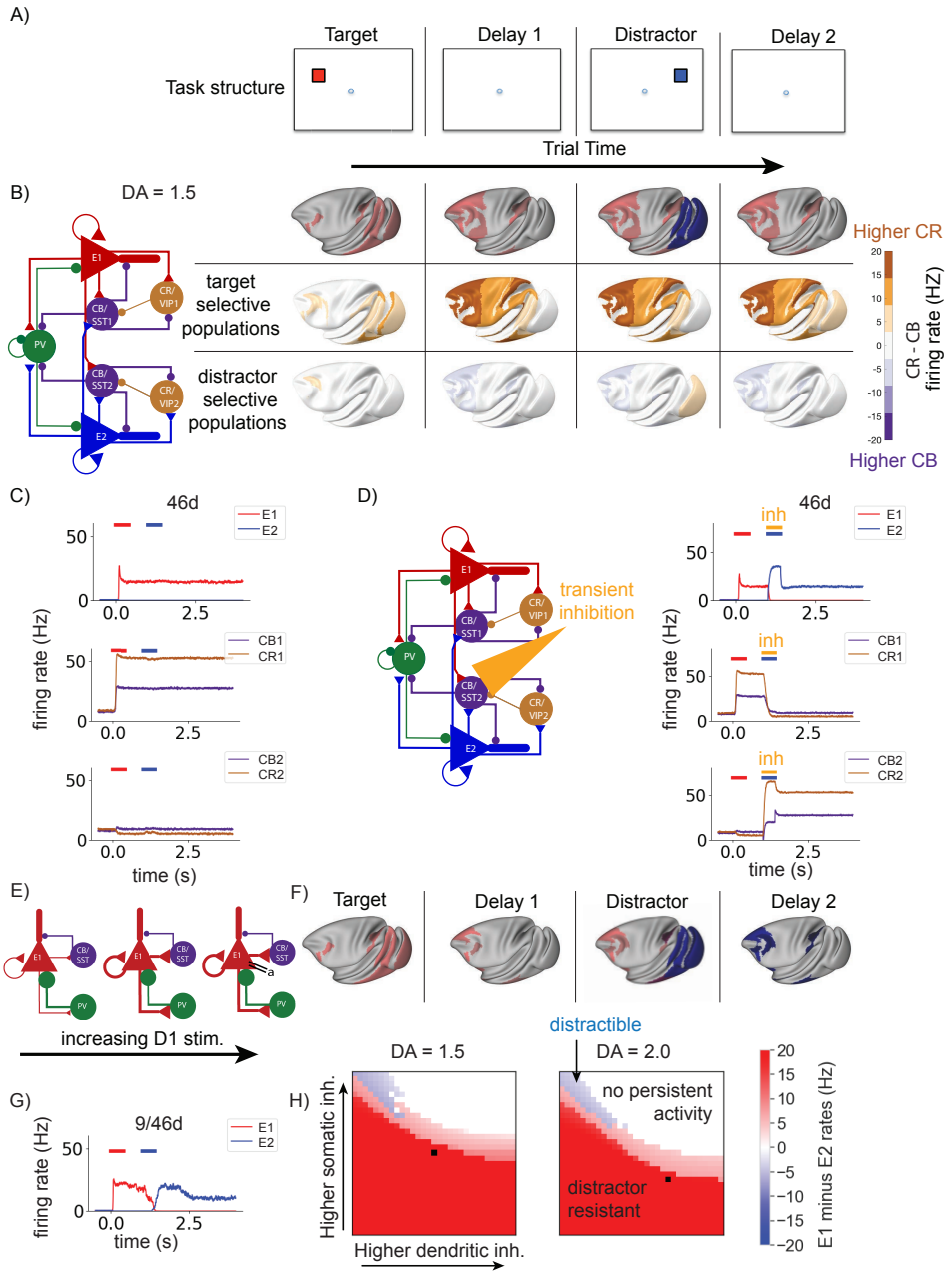

Figure 5: Figure 5. Dopamine increases distractor resistance by shifting the subcellular target of inhibition. A) Task structure. A target stimulus was followed by a delay, a distractor stimulus and another delay period. B) For mid-level dopamine release, persistent target-related activity (red) was present in the frontoparietal network through the delay and the distractor until the end of the trial. Each cortical area contains populations of excitatory, CB/SST and CR/VIP cells that respond to the target stimulus (E1, CB/SST1, CR/VIP1), separate populations sensitive to the distractor stimulus (E2, CB/SST2, CR/VIP2) and PV cells. B and C) Throughout the delay period and distractor stimulus, activity in VIP1 is higher than in CB/SST1, leading to disinhibition of the E1 dendrite. In contrast, activity in CR/VIP2 is slightly lower than in CB/SST2, leading to inhibition of the E2 dendrite. D) We transiently inactivated CB/SST2 populations in the frontoparietal network during the presentation of the distractor stimulus. On trials in which CB/SST2 populations were inhibited, the network became distractible. E) We removed the dopamine modulation of somatic and dendritic inhibition, while maintaining the effects of dopamine on NMDA-dependent excitation and adaptation unchanged. F,G) Without the dopamine-dependent switch towards dendritic inhibition, the network became distractible, with distractor-related activity dominating at the end of the trial. H) Consistently across dopamine levels, higher somatic, and lower dendritic inhibition was associated with distractible working memory (blue). In contrast, lower somatic, and higher somatic inhibition was associated with distractor-resistant working memory (red). High dendritic and high somatic inhibition results in no persistent activity (white). The levels of dendritic and somatic inhibition associated with the standard dopamine modulation used in the rest of the paper marked by a black square. DA, cortical dopamine availability.

To test the importance of this effect, we transiently inhibited the CB/SST2 cells in the frontoparietal network during the presentation of the distractor (CB/SST2, Fig 5D). This transient inhibition of CB/SST2 cells was sufficient to switch the network to a distractible state, with the distractor stimulus held in working memory until the end of the trial (Fig 5D).

As dopamine increases the strength of inhibition to the dendrites, and decreases inhibition to the soma, it is possible that this aspect of dopamine modulation enhances distractor-resistance of the system. We removed this effect of dopamine modulation, while leaving dopamine's effects on NMDA and adaptation currents as before (Fig 5E). We repeated the working memory task in the presence of the distractor with a mid-level of dopamine, which normally results in distractor-resistant working memory. Without the dopamine-dependent shift of inhibition from the soma to the dendrite, the system becomes distractible (Fig 5F, G). We searched the parameter space for the strength of inter-areal and local excitatory-to-excitatory connections, and found that, when a subset of local cortical areas were endowed with sufficient recurrent excitation to generate persistent activity in isolation (e.g., $\mu_{E,E} = 1.25$, $g_{E,E}^{self} = 0.33nA$), high somatic inhibition and low dendritic inhibition was generally associated with distractibility (Fig 5H, Fig S7). Low somatic and high dendritic inhibition was associated with distractor-resistant behaviour (Fig 5H, Fig S7). Therefore the action of dopamine in shifting inhibition from the soma to the dendrite (Gao et al. 2003), via its strong effect on CB/SST cells (Mueller et al. 2019), prevents distractor-related activity from sensory areas disrupting ongoing persistent activity in the frontoparietal network.

Learning to optimally time dopamine release through reinforcement

In real life we experience a constant flow of sensory inputs, and our working memory system must be flexible in determining the timing of relevant versus irrelevant information. Dopamine neurons fire in response to task-relevant stimuli (Schultz et al. 1993), but should not fire in response to task-irrelevant distracting stimuli, regardless of timing. We hypothesised that the correct timing of dopamine release could be learned by simple reward-learning mechanisms.

We added a simplified model of the ventral tegmental area (VTA) with GABAergic and dopaminergic neurons to our large-scale cortical model (Fig 6A) (c.f. Braver and Cohen 2000). Cortical pyramidal cells target both GABAergic and dopaminergic cells in the VTA (Soden et al. 2020; Watabe-Uchida et al. 2012). Dopaminergic cells are also strongly inhibited by local VTA GABAergic cells (Soden et al. 2020). Dopamine in the model is released in cortex in response to VTA dopaminergic neuron firing, and cortical dopamine levels slowly return to baseline following cessation of dopaminergic neuron firing (Muller et al. 2014). In the model, the synaptic strengths of cortical inputs from the selected populations to VTA populations are increased following a reward, and weakened following an incorrect response (Harnett et al. 2009; Soltani and Wang 2006).

We tested the model on a variant of the target-distractor-ping task introduced earlier (Fig 4B i; 6B). For the first 30 trials, the first stimulus (Cue 1, red) was rewarded (rule 1). For the following 30 trials the second stimulus (Cue 2, blue) was rewarded (rule 2). For the final 30 trials, we switched back to rule 1 (Fig 6B). By the seventh trial of the first block distractor-resistant persistent activity emerges, and the first cue is correctly remembered. This behaviour persisted until the next block. Following a few trials of the second block, dopamine release in response to the first stimulus was reduced, and neural populations throughout the cortex only transiently represented the first (now irrelevant) stimulus. However, dopamine response to the second stimulus increased, so that persistent activity was engaged following the second stimulus. Following the second rule switch, the system again switched back to engaging persistent activity in

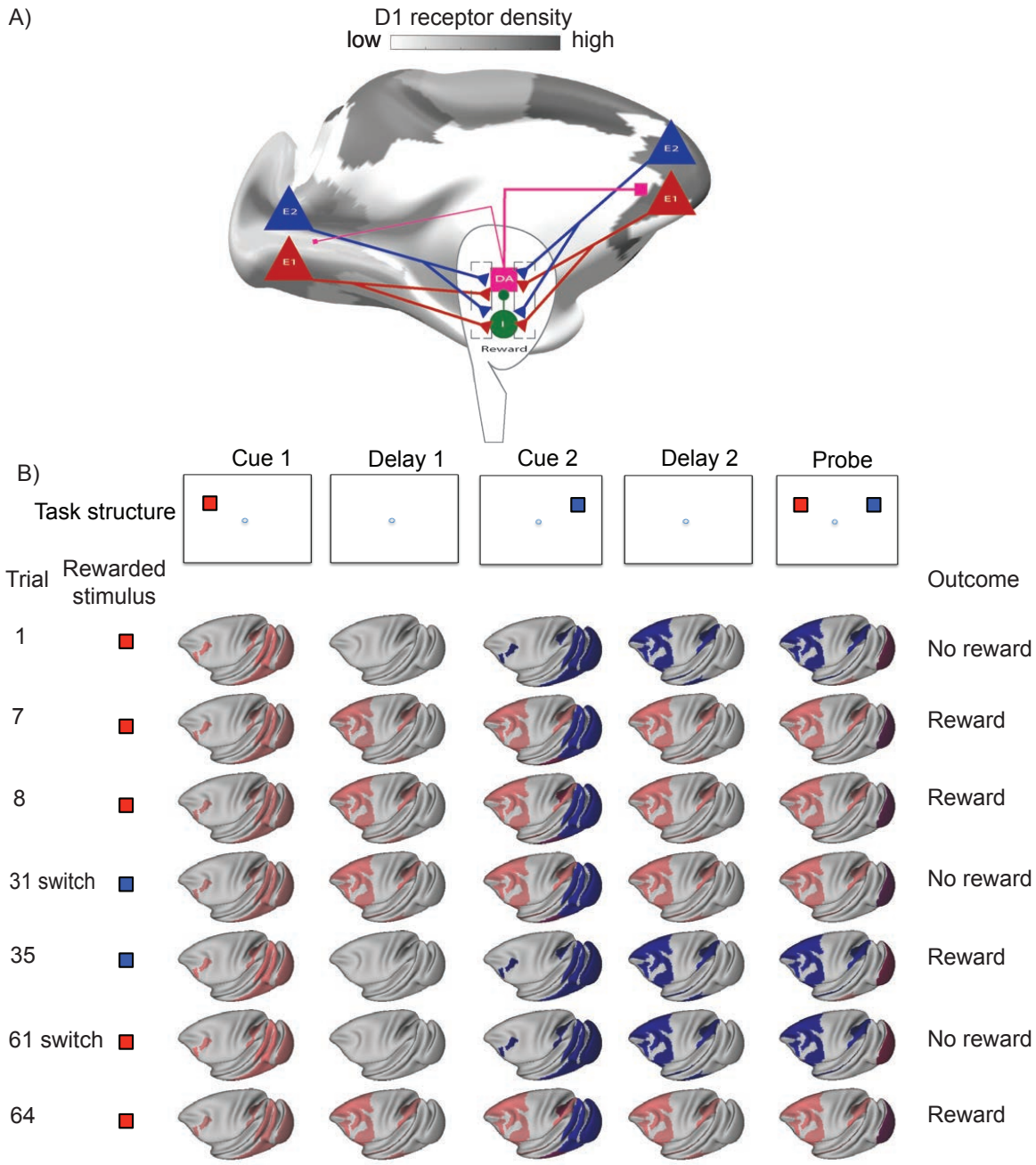

Figure 6: Figure 6. Reward-dependent learning of dopamine release appropriately engages persistent activity mechanisms to enable reversal learning. A) We designed a simplified VTA model and connected this bidirectionally to the large-scale cortical model. The VTA contained dopaminergic and GABAergic neuron populations. Dopamine was released dynamically depending on dopaminergic neuron activity. The strength of cortical inputs to VTA dopaminergic and GABAergic cells was updated at the end of each trial on the basis of trial outcome and choice. B) We simulated a task with two cues (red and blue) followed by a probe stimulus. The rewarded stimulus changed every 30 trials. Following each switch, after a few trials the network learns to store the appropriate stimulus in distributed persistent activity. This depends on high dopamine release in response to the rewarded stimulus and low release in response to the unrewarded stimulus.

response to the first cue. Additionally, the number of trials to engage appropriate persistent activity gradually decreased with each switch. We further tested the model on a version of the task in which the relevant red cue could be shown either first or second within a block, before the blue cue became relevant in the second block. The model was also able to learn this task, although it took more trials (10-15) to learn the switch (for the first few blocks). Thus, by means of simple reward-learning mechanisms, the optimal timing of dopamine release can be learned, allowing flexible engagement of distributed persistent activity in working memory.

Discussion

We uncovered a macroscopic gradient of dopamine D1 receptor density along the cortical hierarchy. By building a novel large-scale anatomically-constrained model of monkey cortex, we showed how dopamine can engage robust distributed persistent activity mechanisms across connected higher cortical areas, and protect memories of behaviourally relevant-stimuli from distraction. This work leads to new predictions that would not have been possible with local circuit models. For example, the model shows that dopamine's enhancement of inhibition from CB/SST-expressing cells to the dendrites of pyramidal cells blocks distracting sensory information from entering the fronto-parietal working memory network. Second, when an initial stimulus fails to robustly activate prefrontal cortex, we found that the memory of the original stimulus can be recalled through an activity-silent synaptic mechanism in inter-areal connections from sensory to fronto-parietal cortex. Lastly, our model predicts that dopamine can switch between activity-silent and distributed persistent activity mechanisms, and the timing of dopamine release could be learned through reinforcement. This suggests that distributed persistent activity may be engaged for behaviorally-relevant stimuli that need protection from distractors.

A gradient of D1 receptors along the cortical hierarchy

In order to create a high-resolution, and high-fidelity map of cortical dopamine receptor architecture, we used quantitative *in-vitro* receptor autoradiography. PET and SPECT scans provide the advantages of *in-vivo* measurements, such as information on individual and group differences, but are limited in spatial resolution and signal-to-noise ratio (Abi-Dargham et al. 2002; Froudist-Walsh et al. 2017a; Roffman et al. 2016; Slifstein et al. 2015) and are often unreliable for cortical measurements (Egerton et al. 2010; Farde et al. 1988). Gene expression methods have certain advantages, especially RNA sequencing which can provide cell-specific data. However, mRNA expression is not always closely related to, or even positively correlated with the receptor density at the synapse (Arnatkeviciute et al. 2019; Beliveau et al. 2017). Receptor density at the synapse is the functionally important quantity, and is directly measured here. The map of D1 receptor density here greatly expands previous descriptions of D1 receptor densities (Goldman-Rakic et al. 1990; Impieri et al. 2019; Lidow et al. 1991; Niu et al. 2020; Richfield et al. 1989). We show that D1 receptor density increases along the cortical hierarchy, peaking in prefrontal and posterior parietal cortex. A previous study of 12 cortical areas suggested a posterior-anterior gradient of D1 receptor expression (Lidow et al. 1991). Here we expand that work to include 109 regions, take into account variation in neuron density across the cortex, and show that the D1 receptor gradient more closely follows the cortical hierarchy than a strict posterior-anterior gradient. The distinction is clear with higher levels of D1 receptor density per neuron in areas of posterior parietal cortex than somatosensory and primary motor cortex. Future work is required to test the degree to which gradients of gene expression accurately capture the

receptor gradient (Beliveau et al. 2017; Hurd et al. 2001). The gradient of dopamine D1 receptors is similar to gradients of other anatomical and functional properties described across the cortex, many of which increase or decrease along the hierarchy (Burt et al. 2018; Fulcher et al. 2019; Goulas et al. 2018; Margulies et al. 2016; Sanides 1962; Shafiei et al. 2020; Wang 2020). We observed some interesting patterns of D1R density per neuron (Fig. 1F), such as a gradual caudo-rostral increase within the prefrontal cortex, which resembles previously reported gradients of plasticity, laminar connectivity and abstraction (Badre and D'esposito 2009; Riley et al. 2018; Vezoli et al. 2021). Due to the small number of animals, and relatively similar D1R expression levels in several areas of frontal and parietal cortex, comparison of D1R density between pairs of areas is difficult. As originally shown in Markov et al. (2014) the hierarchy itself is steep for early sensory areas, and becomes shallower for higher association areas (Markov et al. 2014a). Therefore, the exact positions of areas like LIP or 10 are not as robustly distinguishable as those of V1, V2 and V4. Nonetheless, we expect the general pattern of an increase in D1R density per neuron along the cortical hierarchy to hold. While the D1R labeling per neuron, as well as synaptic excitation and inhibition, display a smooth gradient, quantitative variations of circuit properties can give rise to a non-smooth pattern of persistent activity along the cortical hierarchy, through a phenomenon akin to bifurcations described by the theory of nonlinear dynamical systems (Mejias and Wang 2020; Wang 2020). Such a sudden transition was observed in a monkey experiment where elevated persistent activity associated with working memory was absent in MT but significantly present one synapse away in the nearby MST (Mendoza-Halliday et al. 2014). Simultaneous recording from many parcellated areas using new tools such as Neuropixels (Jun et al. 2017) from behaving animals could test our model prediction systematically in future experiments. This increasing gradient of dopamine receptors along the cortical hierarchy is a major anatomical basis by which dopamine can modulate higher cognitive processing.

An inverted-U relationship between dopamine and distributed working memory activity

Previous experimental and modelling studies have shown an inverted-U relationship between D1 receptor stimulation and persistent activity in the prefrontal cortex in monkeys performing working memory tasks (Brunel and Wang 2001; Vijayraghavan et al. 2007; Wang et al. 2019). Dopamine activity in the VTA is relatively low during the delay period, but still has an inverted-U shape relationship with short-term memory peformance in the rat (Choi et al. 2020). In our model this may be interpreted as the VTA continuing to provide low-level dopamine to the cortex in order to maintain cortical dopamine levels within the appropriate bounds for distributed persistent activity. We found dense D1 and D2 receptor labelling in the striatum. However, we focused our working memory modeling on the cortex and VTA. Notably, optogenetic manipulation of substantia nigra pars compacta dopamine neurons (which principally target the striatum) does not have specific short-term memory effects (Choi et al. 2020). This suggests that cortical, rather than striatal, dopamine release is likely more important to short-term memory. By constructing a novel-large scale model based on the D1 receptor map and tract-tracing data, we found that the inverted-U relationship between D1 receptor stimulation and persistent activity held across frontal and parietal cortex during working memory. The working memory activity pattern was strikingly similar to that seen experimentally, according to a meta-analysis of 90 electophysiology studies of delay period activity in monkey cortex (Leavitt et al. 2017). Analysing the model showed that the pattern of inter-areal connections was the strongest determinant of the pattern of working memory activity.

Noudoost and Moore found that injecting a D1-antagonist into FEF, led to an increase in firing rates in V4 (Noudoost and Moore 2011). Similarly, in our model, when cortical dopamine levels are close to optimal range for working

memory (i.e. the peak of the inverted U), then reducing D1 receptor stimulation via an antagonist would lead to an increase in V4 activity during the second peak of the response to visual stimulation (Figure S3). Note however that our model focused on distributed working memory in a large-scale cortical system, and was not built to uncover mechanisms of attention or decision making. Recent electrophysiology and modeling studies of non-human primate attention have suggested that the dominant net effect of attention on neural activity in sensory cortex is inhibition (Huang et al. 2019; Yoo et al. 2021). This may be consistent with a subtle enhancement of firing for neurons whose receptive field is in the focus of attention, combined with a greater inhibition of neurons with nearby receptive fields. We showed that somatosensory and visuospatial working memory tasks engage largely overlapping higher cortical areas during the delay period. It is likely that at a neural level, these networks may be only partially overlapping. To simulate this mixed inhibitory and excitatory effects of attention, and to identify the degree to which different types of working memory engage the same neurons, future models will require more neural populations per area, perhaps with structured connectivity, such as a ring (Ardid et al. 2007). Local circuit modelling has previously shown that a circuit designed for working memory turned is suitable for decision-making (Wang 2002). Our model may also be suitable to consider decision processes distributed across cortical areas.

Prefrontal and parietal contributions to distributed working memory

Some studies have aimed to dissociate the contribution of prefrontal and parietal cortex to working memory via temporary inactivations. For example, Chaffee and Goldman-Rakic examined the effects of reversibly cooling prefrontal or parietal cortex on activity in the other area and behavior during a visuospatial working memory task without a distractor (Chafee and Goldman-Rakic 2000). Cooling affected the frontal eye fields (area 8) and nearby prefrontal cortex including the principal sulcus (areas 46, 9). Cooling of the parietal cortex included LIP, as well as parts of area DP, 7A and 5. Cooling parietal cortex led to a substantial reduction in prefrontal firing rates, but with only a minor effect on performance. Cooling prefrontal cortex led to a substantial reduction in parietal firing rates, and a large increase in behavioral errors (Chafee and Goldman-Rakic 2000). This is consistent with our simulation results that both prefrontal and parietal inactivation can have a robust effect on reducing mnemonic delay activity, but prefrontal inactivation has much larger effects on performance (Figures 3E,F).

Suzuki and Gottlieb inactivated areas LIP and dlPFC using the GABA-A receptor agonist muscimol and assessed performance on a similar visuospatial working memory task, with and without distractor stimuli (Suzuki and Gottlieb 2013). In these experiments neither LIP nor dlPFC inactivation caused errors in trials without distractors (Suzuki and Gottlieb 2013). However, inactivation of dlPFC, but not LIP, led to a dramatic increase in errors on trials with distractors (Suzuki and Gottlieb 2013). This is consistent with our simulation results showing that precise lesions to dlPFC affect behavior on challenging working memory trials with distractor stimuli, but larger lesions are required to disrupt performance in simple working memory trials without distractors, and lesions to LIP have only subtle effects on performance. The effects of lesions on model performance are consistent with recent reports that there is a distinction between areas that are active during normal behavior, and those that are essential for a computation (Pinto et al. 2019; Zatka-Haas et al. 2021), and that performance on challenging tasks is more sensitive to a variety of cortical lesions (Pinto et al. 2019).

Lesions to areas with a high D1 receptor density disrupt working memory

Working memory activity was most disrupted by lesions to areas with a high D1 receptor density, a prediction that can be tested experimentally. Human patients with traumatic brain injury often have working memory deficits (Dunning et al. 2016). Pharmacological treatments of these deficits, including with dopaminergic drugs, have met mixed success (Froudist-Walsh et al. 2017b). Our model simulations suggest that D1 agonists or antagonists could be effective at restoring normal working memory functioning following lesions to particular cortical areas, but the correct treatment may depend on the baseline cortical dopamine levels of the patient. Dopaminergic drugs have also been suggested as treatments for schizophrenia patients with working memory deficits (Yang and Chen 2005). In patients with schizophrenia, PV and SST gene expression is reduced across multiple areas of the cortical working memory network (Tsubomoto et al. 2019). Disruption to these inhibitory neurons is likely to contribute to working memory deficits. Future adaptations of our model could allow simulation of working memory deficits and motivate potential treatments for patients based on their particular anatomy, gene expression and patterns of cortical dopamine release or receptor density (Abi-Dargham et al. 2002; Slifstein et al. 2015).

A dopamine switch between the activity-silent state and persistent activity

For very low or high levels of D1 receptor stimulation, it was possible to maintain stimulus information in the absence of persistent activity, via synaptic mechanisms. This pattern of successful memory recall without frontoparietal delay-period activity is reminiscent of a passive short-term memory trace which is thought to rely on 'activity-silent' synaptic mechanisms (Rose et al. 2016; Trübutschek et al. 2017; Wolff et al. 2017) and which could occur without ignition of fronto-parietal cortex (Trübutschek et al. 2017, 2019). Previous models with short-term synaptic plasticity have focused on local activity in the prefrontal cortex (Mongillo et al. 2008) and thus implicitly imply that the initial stimulus must significantly engage prefrontal neural activity and store the memory trace via short-term plasticity in local prefrontal connections. However, some stimuli may be remembered without a strong initial prefrontal response. We found that short-term synaptic plasticity in inter-areal connections from sensory to frontoparietal areas was most important to maintaining the silent memory trace. In particular, this a potential mechanism for activity-silent short-term memory in the absence of a strong initial prefrontal response to the stimulus. It has recently been proposed that nonspecific excitatory or inhibitory currents could account for switches between active and silent states (Barbosa et al. 2020). Our model suggests that dopamine could in fact account for the switch from silent to active state. Indeed, due to the inverted-U relationship between dopamine and persistent firing, a dopamine response to the reward at the end of a trial could also terminate persistent activity. Another recent proposal suggests that activity-silent short term memory could be undertaken via hippocampal-prefrontal episodic memory mechanisms, perhaps in combination with short-term synaptic changes in the cortex (Beukers et al. 2021). Future studies should aim to disentangle the contributions of rapid synaptic changes within prefrontal cortex (Mongillo et al. 2008), at inter-areal connections from sensory areas (this paper), or in the hippocampus (Beukers et al. 2021) to activity-silent short-term memory in the primate. We found that in the activity-silent state, the most recently encoded stimulus was always encoded most strongly. This prediction should hold as the number of distractors is increased. The activity-silent system may still be able to recall earlier stimuli for a limited time if another input biases the network towards the activity pattern used during encoding of the earlier stimulus. This could be achieved by presenting a cue that partially overlaps with the original remembered stimulus,

which can trigger pattern completion, and recall of the memory (Manohar et al. 2019). Stimuli stored in persistent activity (and thus dependent on mid-level dopamine release) were more robust against distraction, consistent with drug studies in humans (Fallon et al. 2017, 2016). Thus, dopamine release may engage distributed persistent activity in order to protect memories of important stimuli from distraction.

Dopamine increases distractor resistance by shifting the subcellular target of inhibition

The resilience of the active working memory state in the model depended on CB/SST cells blocking distracting inputs from sensory areas to the dendrites of pyramidal cells in frontal and parietal cortex. Previous modelling work on local cortical circuits has suggested that greater dendritic and less somatic inhibition could increase distractor-resistance (Wang et al. 2004a), and that selective disinhibition of the dendrite (through CR/VIP cells) could allow specific information to be passed through the network (Yang et al. 2016). In our large-scale model, CR/VIP cells selectively disinhibited the dendrites of cells selective to the target stimulus, allowing target-related activity to flow through the cortical network. D1 receptors in monkey cortex are more strongly expressed on CB/SST neurons than other interneuron types (Mueller et al. 2019). In agreement with these anatomical findings, application of dopamine to a frontal cortex slice increases inhibition to the dendrite, and decreases inhibition to the soma of pyramidal cells (Gao et al. 2003). We found that, as long as local cortical areas (or potentially cortico-subcortical loops) are capable of maintaining persistent activity, then shifting the balance of inhibition from the soma to the dendrite can allow for maintenance of an active representation of a stimulus in persistent activity, while shielding it from distracting input from sensory areas. Note that the ability of cortical areas to maintain persistent activity itself depends on dopaminergic enhancement of NMDA-dependent excitation. In mice, inhibition of SST neurons in mPFC during the sample period of a spatial working memory task impairs performance and increases representation of irrelevant information in prefrontal activity (Abbas et al. 2018). Consistent with our model, this suggests that somatostatin neurons gate the entry of information into working memory, and inhibition of somatostatin neurons in frontoparietal areas allows distracting information to enter.

Learning to engage distributed persistent activity through reinforcement

Distractor-resistance in response to all stimuli could render the working memory system inflexible, and unresponsive to new, potentially important inputs. Previous studies have shown that lesioning the prefrontal cortex impairs the ability to switch attention between stimuli across trials (Rossi et al. 2007). Our model predicts that the prefrontal cortex is more crucial to persistent activity than activity-silent short-term memory, which can rely on short-term synaptic changes outside the prefrontal cortex. We show that by using a simple reward-based learning mechanism, a cortical-VTA model (c.f. Braver and Cohen 2000; Frank 2005), can successfully perform a task with reversals between the memory cue and distractor stimuli across trials. In our model, the timing of dopamine release in the cortex can be learned in order to engage distributed persistent activity throughout the frontoparietal network only in response to reward-predicting cues. Dopamine neurons burst about 130-150ms after reward-predicting stimuli, coinciding with a rise in activity in frontal cortical neurons (Lafuente and Romo 2012). Due to the slow dynamics of cortical dopamine (Muller et al. 2014), a transient increase in dopamine release in response to the target stimulus (Choi et al. 2020; Schultz et al. 1993) is sufficient to maintain distributed persistent activity for several seconds. This mechanism may thus be reserved for behaviorally important stimuli that must be protected from distraction, even when the behaviorally-relevant stimuli

change from trial to trial. In contrast, irrelevant, or less salient stimuli result in lower dopamine release, and may be remembered via silent mechanisms, or forgotten. We investigated model performance on a reversal learning task with identical, repeated trials within a block. In natural life, no two situations are exactly the same. It is likely that the brain generalizes across similar situations in order to enable reinforcement learning to be used in practice. This ability to generalize may itself arise from dopamine-dependent plasticity in prefrontal cortex (Wang et al. 2018). The classical reward-prediction-error hypothesis treats dopamine as a global scalar reward prediction error signal that is spatiotemporally uniform (Schultz 1998). Here we aim to highlight one form of spatial heterogeneity, and suggest that broad dopamine release will affect each cortical area according to the D1 receptor density per neuron. Recent work suggests that there is temporal heterogeneity in dopamine release, which is released in waves in the mouse striatum (Hamid et al. 2021). Whether such dopamine waves also occur in cortex, or indeed in primates, remains to be seen. Even if dopamine is released in waves across the cortex, its impact on cortical areas will be dependent on the D1 receptor gradient presented here.

Roles of other neuromodulatory and subcortical systems

In addition to dopamine, other neuromodulators, such as acetylcholine (Croxson et al. 2011; Sun et al. 2017; Yang et al. 2013) and noradrenaline (Arnsten et al. 2012) affect prefrontal delay period firing and performance on visuospatial working memory tasks. Cholinergic mechanisms may complement dopaminergic mechanisms. For example, nicotinic alpha-7 receptors depolarize pyramidal cells, which enables NMDA-receptors to be engaged via the removal of the magnesium block (Yang et al. 2013). This may compensate for a reduction in presynaptic glutamate release in response to D1 stimulation, and enable dopamine's permissive effects on NMDA transmission (Seamans et al. 2001). Muscarinic M1 receptor activation closes KCNQ channels, which contribute to the hyperpolarizing effect of high levels of D1 stimulation (Arnsten et al. 2012; Galvin et al. 2020). Thus M1 stimulation may enable persistent activity over a larger range of dopamine release. The effects of noradrenaline on working memory circuits depend on the targeted adrenergic receptors. Moderate release of noradrenaline engages adrenergic α_{2A} receptors, which may counteract the hyperpolarizing effects of HCN channels (Arnsten 2000; Arnsten et al. 2012; Li and Mei 1994; Robbins and Arnsten 2009) and keep the D1 effects in check by decreasing calcium-cAMP signalling. Greater noradrenergic levels engage α_1 and β_1 receptors, which promote calcium-cAMP signaling and at high levels provide negative feedback via KCNQ and HCN channels (Arnsten et al. 2020). Studies linking neuromodulators to working memory have focused on dorsolateral prefrontal cortex. Much less is known about the influence of these and other neuromodulators on the distributed network activity that underlies working memory outside of prefrontal cortex. Future work should focus on the interaction of distinct neuromodulators, and how the release of different combinations of neuromodulators may affect distributed activity patterns and behavior, taking into account the different distributions of these receptors across the cortex (Froudist-Walsh et al. 2021). Subcortical structures, such as the thalamus may play a significant role in working memory (Fuster and Alexander 1971; Guo et al. 2017; Jaramillo et al. 2019; Watanabe and Funahashi 2012). However, future experiments and computational modelling studies should aim to disentangle the contribution of the thalamus to sensory working memory and motor preparation (Guo et al. 2017; Watanabe and Funahashi 2012); and clarify the degree to which such mechanisms are shared across species. Once appropriate weighted and directed connectivity data becomes available, future large-scale cortical models should also integrate further structures, such as

the thalamus (Jaramillo et al. 2019), basal ganglia (Wei and Wang 2016), claustrum and cerebellum to identify their contributions to working memory.

Conclusion

We found experimentally a macroscopic gradient of dopamine D1 receptor density along the cortical hierarchy. By building a novel connectome-based biophysical model of monkey cortex, endowed with multiple types of inhibitiory cells, we show how dopamine can engage robust distributed persistent activity mechanisms across connected higher cortical areas, and protect memories of salient stimuli from distraction. As distributed persistent activity is necessary for the internal manipulation of information in working memory (Masse et al. 2019; Takeda and Funahashi 2004; Trübutschek et al. 2019), dopamine release in the cortex may be a key step towards higher cognition and thought.

Acknowledgements

This project was funded by (NIH/BMBF) CRCNS grant (nos. R01MH122024 and 01GQ1902) to NPG and XJW; NIH grant R01MH062349, ONR grant N00014 and James Simons foundation grant 543057SPI to XJW; the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreements 785907 (Human Brain Project SGA2) and 945539 (Human Brain Project SGA3) to KZ and NPG; ANR grant DUAL_STREAMS ANR-19-CE37-0025 to KK; Grants LABEX CORTEX ANR-11-LABX-0042, Université de Lyon ANR-11-IDEX-0007; A2P2MC ANR-17-NENC-0004; CORTICITY ANR-17-HBPR-0003; REGION AUVERGNE-RHON-ALPES SCUSI 1700933701; and Chinese Academy of Sciences President's International Fellowship Initiative Grant No 2018VBA0011 to HK. The authors would like to dedicate this work to Prof. Karl Zilles, who passed away during the finalisation of this manuscript. He was an inspiring mentor, colleague and friend.

Inclusion and Diversity

While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

Author contributions

Conceptualization - SFW, XJW. Methodology - SFW, DB, XD, KK, HK, KZ, NPG, XJW. Software - SFW. Validation - SFW, NPG, DB, XD, KZ, XJW. Formal Analysis - SFW. Investigation - NPG, LJR, MQ, HK, KZ. Resources - NPG, KK, HK, KZ, XJW. Writing - original draft preparation - SFW. Writing - review and editing - all authors. Visualization - SFW. Supervision - NPG, KZ, XJW. Funding acquisition - SFW, NPG, KK, HK, KZ, XJW.

Declaration of Interests

The authors declare no competing interests.

Tables and Legends - NOTE TABLES 1-4 SHOULD APPEAR IN STAR METHODS

$$
\begin{array}{cc}
 & \text{from} \\
G_E & \begin{array}{cc} E1^{soma} & E2^{soma} \end{array} \\
\begin{array}{c} E1^{soma} \\ E2^{soma} \\ PV \\ to\ CB1 \\ CB2 \\ CR1 \\ CR2 \end{array} & \begin{pmatrix} g_{E,E}^{self} & 0 \\ 0 & g_{E,E}^{self} \\ g_{PV,E} & g_{PV,E} \\ g_{CB,E}^{self} & g_{CB,E}^{cross} \\ g_{CB,E}^{cross} & g_{CB,E}^{self} \\ g_{CR,E} & 0 \\ 0 & g_{CR,E} \end{pmatrix}
\end{array}
\qquad
\begin{array}{cc}
 & \text{from} \\
G_{I,[k]}^{dend} & \begin{array}{cc} CB1 & CB2 \end{array} \\
\begin{array}{c} E1^{dend} \\ E2^{dend} \end{array} & \begin{pmatrix} g_{E^{dend},CB,[k]}^{DA} & 0 \\ 0 & g_{E^{dend},CB,[k]}^{DA} \end{pmatrix}
\end{array}
$$

Table 1. Left: Local excitatory output connections target excitatory and inhibitory populations. Right: CB/SST interneurons target the dendrites of pyramidal cells.

$$
\begin{array}{cc}
 & \text{from} \\
G_{I,[k]}^{soma} & \begin{array}{ccccc} PV & CB1 & CB2 & CR1 & CR2 \end{array} \\
\begin{array}{c} E1^{soma} \\ E2^{soma} \\ PV \\ to\ CB1 \\ CB2 \\ CR1 \\ CR2 \end{array} & \begin{pmatrix} g_{E_{soma},PV,[k]}^{DA} & 0 & 0 & 0 & 0 \\ g_{E_{soma},PV,[k]}^{DA} & 0 & 0 & 0 & 0 \\ g_{PV,PV} & g_{PV,CB} & g_{PV,CB} & 0 & 0 \\ 0 & 0 & 0 & g_{CB,CR} & 0 \\ 0 & 0 & 0 & 0 & g_{CB,CR} \\ 0 & g_{CR,CB} & 0 & 0 & 0 \\ 0 & 0 & g_{CR,CB} & 0 & 0 \end{pmatrix}
\end{array}
$$

Table 2. PV cells inhibit the cell body of pyramidal cells, but are themselves inhibited by other PV cells and CB/SST cells. CB/SST cells and CR/VIP cells mutually inhibit each other.

$$
\begin{array}{cc}
 & \text{from} \\
J^{E,E} & \begin{array}{cc} E1_{soma} & E2_{soma} \end{array} \\
\begin{array}{c} E1_{soma} \\ E2_{soma} \\ to \\ E1_{dend} \\ E2_{dend} \end{array} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ g_{E,E}^{LR,self} & g_{E,E}^{LR,cross} \\ g_{E,E}^{LR,cross} & g_{E,E}^{LR,self} \end{pmatrix}
\end{array}
\qquad
\begin{array}{cc}
 & \text{from} \\
J^{I,E} & \begin{array}{cc} E1_{soma} & E2_{soma} \end{array} \\
\begin{array}{c} PV \\ SST1 \\ to\ SST2 \\ VIP1 \\ VIP2 \end{array} & \begin{pmatrix} g_{PV,E}^{LR} & g_{PV,E}^{LR} \\ g_{SST,E}^{LR} & 0 \\ 0 & g_{SST,E}^{LR} \\ g_{VIP,E} & 0 \\ 0 & g_{VIP,E} \end{pmatrix}
\end{array}
$$

Table 3. Long-range targets onto excitatory (left) and inhibitory (right) cells

Table 4. Parameters for Numerical Simulations

Parameter	Description	Value
$g^{self}_{E,E}$, $g_{PV,E}$, $g^{self}_{SST,E}$, $g^{cross}_{SST,E}$, $g_{VIP,E}$	Excitatory synaptic strengths	0.18nA, 0.174nA, 0.0435nA, 0.0435nA, 0.058nA
$g^{min}_{E_{soma},PV}$, $g^{max}_{E_{soma},PV}$, $g_{PV,PV}$	PV synaptic strengths	-0.001nA, -0.4nA, -0.18nA
$g^{min}_{E_{dend},SST}$, $g^{max}_{E_{dend},SST}$, $g_{PV,SST}$, $g_{VIP,SST}$	CB/SST synaptic strengths	-0.09nA, -0.11nA, -0.17nA, -0.1nA
$g_{SST,VIP}$	CR/VIP synaptic strengths	-0.05nA
τ^{NMDA}, τ^{AMPA}	E synaptic time constants	60ms,2ms
τ^{GABA}, $\tau^{GABA,dend}$	I synaptic time constants	5ms, 10ms
τ^{a}	adaptation time constant	100ms
γ_{NMDA}, γ_{AMPA}, γ_I	synaptic rise constants	1.282, 5, 2
κ_{PV}, κ_{other}	$\frac{NMDA}{NMDA+AMPA}$ fraction	0.8, 0.9
z^{min}	Min spine val	0.45
σ_{noise}	std. dev. of noise	0.005nA
$I^{bg}_{E_{soma}}$, $I^{bg}_{i\epsilon Inh}$, $I^{bg}_{E_{dend}}$	Background inputs	0.31nA, 0.30nA, 0.03nA
c_{1-6}	Dendrite parameters	0.12nA, 0.13624nA, 7, 0nA, 0.00964nA, 0.02nA
g^{a}_{PV}, g^{a}_{other}	Adaptation strength	0nA, -0.004nA
a, b, d	f-I curve (E)	0.135 Hz/nA, 54Hz, 0.308s
$c_{SST,VIP}$, $r^0_{SST,VIP}$	f-I curve (SST, VIP)	132Hz/nA, 33Hz
c_{PV}, r^0_{PV}	f-I curve (PV cells)	330Hz/nA, 95Hz
b_1	rescale FLN	0.3
$g^{LR,self}_{E,E}$, $g^{LR,cross}_{E,E}$	Long-range E targets	0.9, 0.1
$g^{LR}_{PV,E}$, $g^{LR}_{SST,E}$, $g^{LR}_{VIP,E}$	Long-range I targets	0.31, 0.22, 0.47
$g^{LR,FEF}_{PV,E}$, $g^{LR,FEF}_{SST,E}$, $g^{LR,FEF}_{VIP,E}$	Long-range I targets FEF	0.2, 0.1, 0.7
b^o, c^o	D1 occupancy	2, 1
b^N, c^N, α	DA-NMDA modulation	0.35, 10, 0.6
b^M, c^M, g^m_E, g^m_I	DA-M current	0.85, 14, -0.5, 0
$\mu^{E,E}$, $\mu^{I,E}$	Long-range connectivity	1.45, 2.24
I^{stim}	target/distractor stimulus	0.1nA (main figures), 0.2nA (Figs S2 and S3)

STAR Methods

KEY RESOURCES TABLE		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Large-scale dynamical model simulation and analysis software	This paper	https://github.com/seanfw/dopamine-dist-wm
Data		
D1R/neuron data	This paper	https://balsa.wustl.edu/
Connectivity data	Markov et al. 2014b	https://core-nets.org/
Spine count data	Elston 2007	DOI:10.1016/B0-12-370878-8/00164-6
Cortical representation of all anatomical data	This paper	https://balsa.wustl.edu/

Materials availability

This study did not generate new unique reagents.

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Xiao-Jing Wang (xjwang@nyu.edu).

METHOD DETAILS

Overview of anatomical data

In this study, we combine post-mortem anatomical data on receptor densities, white matter connectivity, neuron densities and dendritic spine counts. Each of these four anatomical measures was originally quantified using different parcellations of cortex. Large sections of the temporal lobe are not yet quantified for either the receptor autoradiography

data, or the tract-tracing connectivity data. Collection of this data is underway and will be made available in future studies. With the exception of the receptor densities in the posterior parietal cortex (Impieri et al. 2019; Niu et al. 2020, 2021), all D1 receptor densities are reported for the first time in this study. The connectivity data for ten of the 40 cortical areas is used here for the first time, but will be described in more detail in an upcoming publication from the Kennedy lab. This enabled us to expand the calculation of the cortical hierarchy to 40 regions.

A note on notation

Subscripts in square brackets, such as $[k]$ are used to denote cortical areas themselves. Subscripts not in brackets, such as i are used to denote populations of neurons within a cortical area. Superscripts are used to provide further clarifying information. We use the convention that targets are listed before sources, so that $g_{i,j}$ would denote the strength of a connection from neural population j to neural population i. Parameter values are listed in Table 4.

Quantification of receptor density across cortex - in-vitro autoradiography

In order to create a high-resolution, and high-fidelity map of cortical dopamine receptor architecture, we used quantitative *in-vitro* receptor autoradiography (Palomero-Gallagher and Zilles 2018). Previous dopamine receptor autoradiography has focused on relatively small sections of cortex (Goldman-Rakic et al. 1990; Impieri et al. 2019; Lidow et al. 1991; Niu et al. 2020; Richfield et al. 1989). To create a more comprehensive map of the cortical dopamine receptors, we measured D1 receptor density across 109 cortical areas, and D1 and D2 receptors in the basal ganglia.

We analysed the brains of three adult male *Macaca fascicularis* specimens (between 6 and 8 years old; body weight between 5.2 and 6.6 kg) obtained from Covance, Münster, where they were used as control animals for pharmaceutical studies performed in compliance with legal requirements.

All experimental protocols were in accordance with the guidelines of the European laws for the care and use of animals for scientific purposes. Animals were sacrificed by means of an intravenous lethal dose of sodium pentobarbital. Brains were removed immediately from the skull, and brain stem and cerebellum were dissected off in close proximity to the cerebral peduncles. Hemispheres were separated and then cut into a rostral and a caudal block by a cut in the coronal plane of sectioning between the central and arcuate sulci. These blocks were frozen in isopentane at -40C to -50C, and then stored in airtight plastic bags at -70C. Each block was serially sectioned in the coronal plane (section thickness 20 μm) using a cryostat microtome (CM 3050, Leica, Germany). Sections were thaw-mounted on gelatine-coated slides, freeze-dried overnight and processed for visualization of D1 or D2 receptors, cell bodies (Merker 1983) or myelin (Gallyas 1979). Quantitative *in-vitro* receptor autoradiography was applied to label dopaminergic D1 and D2 receptors according to previously published protocols (Palomero-Gallagher and Zilles 2018; Zilles et al. 2002) encompassing a preincubation, a main incubation and a final rinsing step. For visualization of the D1 receptor, sections were first rehydrated and endogenous substances removed during a 20 minute preincubation at room temperature in a 50 mM Tris-HCl buffer (pH 7.4) containing 120 mM NaCl, 5 mM KCl, 2 mM $CaCl_2$ and 1 mM $MgCl_2$. During the main incubation, sections were incubated with either 0.5 nM [^{3}H]SCH 23390 alone (to determine total binding), or with 0.5 nM [^{3}H]SCH 23390 and 1 mM of the displacer mianserin (to determine the proportion of displaceable, non-specific binding) for 90 minutes at room temperature in the same buffer as used for the preincubation. Finally, the rinsing procedure consisted of two 20 minutes washing steps in cold buffer followed by a short dip in distilled water. For

visualization of the D2 receptor, sections were preincubated 50 mM Tris-HCl buffer (pH 7.4) containing 150 mM NaCl and 1% ascorbate. In the main incubation, sections were incubated with either 0.3 nM [^{3}H]raclopride alone, or with 0.3 nM [^{3}H]raclopride and 1 μM of the displacer 1 μM butaclamol for 45 minutes at room temperature in the same buffer as used for the preincubation. Rinsing consisted of six 1 minute washing steps in cold buffer followed by a short dip in distilled water. Specific binding is the difference between total and non-specific binding. Since the ligands and binding protocols used resulted in a displaceable binding, which was less than 5% of the total binding, total binding is considered to be equivalent of specific binding. Sections were dried in a cold stream of air, exposed together with plastic scales of known radioactivity against tritium-sensitive films (Hyperfilm, Amersham) for six (for the D1 receptor) or eight (for the D2 receptor) weeks, and ensuing autoradiographs processed by densitometry with a video-based image analysing technique (Palomero-Gallagher and Zilles 2018; Zilles et al. 2002). Autoradiographs were digitized using a CCD-camera, and stored as 8-bit grey value images with a spatial resolution of 2080x1542 pixels. Grey values (g) in the co-exposed scales as well as experimental conditions were used to create a regression curve with which grey values in each pixel of an autoradiograph were transformed into binding site densities (Bmax) in fmol/mg protein by means of the formula

$$B_{max} = \frac{gR}{EBW^b s^a} \cdot \frac{K^D + L}{L} \tag{1}$$

where R is the radioactivity concentration (cpm) in a scale, E the efficiency of the scintillation counter used to determine the amount of radioactivity in the incubation buffer, B the number of decays per unit of time and radioactivity, W_b the protein weight of a standard, S_a the specific activity of the ligand, K_D the dissociation constant of the ligand, and L the free concentration of the ligand during incubation. For visualization purposes solely, autoradiographs were subsequently pseudo-colour coded by linear contrast enhancement and assignment of equally spaced density ranges to a spectral arrangement of eleven colours.

Cortical areas were identified by cytoarchitectonic analysis and receptor densities measured at comparable sites in the adjacent sections processed for receptor visualization. The mean receptor density for each area over a series of 3–5 sections per animal and receptor was determined by density profiles extracted vertical to the cortical surface using Matlab-based in house software (Palomero-Gallagher and Zilles 2018).

Retrograde tract-tracing

The inter-areal connectivity data in this paper is part of an ongoing effort to map the cortical connectome of the macaque using retrograde tract-tracing (Markov et al. 2013, 2014a,b). For each target area, a retrograde tracer was injected into the cortex. The tracer was taken up in the axon terminals in this area, and retrogradely transported to the cell bodies of neurons that projected to the target. These cell bodies could be throughout the brain. Each of these cell bodies in cortex was counted as a labelled neuron (LN). The amount of labelled neurons was counted in all cortical areas except for the injected target area. The cortical areas that send axons to the target area are called source areas. As there are uncontrollable differences in tracer volume and uptake between injections, we estimated the strength of connections as follows. For a given injection, the total number of cell bodies in the cortex outside of the injected (target) area was counted. The number of labeled neurons was within a source cortical area was then divided by the number of labeled neurons in the whole cortex (excluding the target area), to give a fraction of labeled neurons (FLN). The FLN was

averaged across all injections in a given target area. For this calculation, we include all areas in the entire cortical hemisphere ($n^{areas} = 91$).

$$FLN_{[k,l]} = \frac{LN_{[k,l]}}{\sum_{l=1}^{n^{areas}} LN_{[k,l]}} \tag{2}$$

In addition, for each connection we defined the supragranular labeled neurons (SLN) as the fraction of neurons in the source area whose cell bodies were in the superficial (aka supragranular) layers.

$$SLN_{[k,l]} = \frac{LN_{[k,l]}^{supra}}{LN_{[k,l]}^{supra} + LN_{[k,l]}^{infra}} \tag{3}$$

The subiculum (SUB) and piriform cortex (PIR) have a qualitatively different laminar structure to the neocortical areas, and thus supra- and infra-laminar connections (and thus the SLN) from these areas are undefined. We thus removed all connections from these areas from the following calculations ($n^{areas,SLN} = 89$). These connectivity data will be available on the core-nets website.

Estimation of the cortical hierarchy

Following (Markov et al. 2014a), we estimate the hierarchical position h of each area using the SLN values of its connections. Feedforward connections tend to originate in the supragranular layers, while feedback connections tend to originate in the deep layers of the source area (Barone et al. 2000; Felleman and Van Essen 1991). Moreover, if a target area occupies a much higher hierarchical position than the source area, a greater proportion of the neurons emerge from the supragranular layers of the source area than if the two areas are closer in the hierarchy (Barone et al. 2000). Likewise for the feedback connections, a greater hierarchical distance between the areas implies that the higher area sends a greater proportion of it projections from the infragranular layers. This implies that the fraction of neurons coming from the supragranular layers in a given connection gives an estimate of the relative hierarchical position of two connected areas (Barone et al. 2000; Markov et al. 2014a). Here, following (Markov et al. 2014a), we estimate a set of hierarchical levels (one per area) that best predicts the SLN values for all connections in the dataset.

The model to estimate the hierarchy has the form

$$g(E(SLN)) = X\beta \tag{4}$$

where g is a function that links the SLN of the connection between areas to the hierarchical distance between them. β is a column vector of length $n^{areas,SLN}$, containing the hierarchy values to be estimated. X is an incidence matrix of shape $n^{conns} \times n^{areas,SLN}$, where n^{conns} (= 2619) is the number of observed (non-zero) connections between cortical areas in the remaining dataset. Each row in X represents a connection, and each column represents a cortical area. All entries in each row equal 0 except for the column corresponding to the source area, which has a value of -1, and the target (recipient) area, which has a value of 1 (Strang 1993).

The hierarchical values can be estimated with maximum likelihood regression. However, the model is singular (the rows sum to zero). In order to make the model identifiable, we therefore removed one column from X. We chose to remove the column corresponding to area V1, which is therefore forced to have a hierarchical value of 0. However, the

choice of column is unimportant, as it is possible to estimate negative hierarchical values (in the case that other areas are lower than V1 in the hierarchy).

We used the beta-binomial model. The binomial parameter p corresponds to the proportion of successes. This is thought to be a random variable following a Beta distribution. The beta-binomial distribution depends on two parameters, the mean (μ, here the SLN), and the dispersion (ϕ). The beta-binomial model can account for the overdispersion of the neural count data. Note that the SLN of each measured connection is input into the model, without averaging across repeated injections.

The likelihood is written as

$$f(\mu, \phi; q, n) = \binom{n}{q} \frac{B\left(\mu(\frac{1-\phi}{\phi}) + q, (1-\mu)(\frac{1-\phi}{\phi}) + n - q\right)}{B\left(\mu(\frac{1-\phi}{\phi}), (1-\mu)(\frac{1-\phi}{\phi})\right)} \tag{5}$$

where q is the number of neurons projecting from the supragranular layers, n is the number of neurons projecting from all layers, and B is the beta function defined as

$$B(x, y) = \int_0^1 p^{x-1}(1-p)^{y-1} dp \tag{6}$$

with $x, y > 0$. We fit the model using $\mu = \Phi(X\beta)$, where Φ is the cumulative Gaussian, as it maps the real numbers to the (0,1) range. $\Phi^{-1} = g$ in equation 4 is the probit link function. The hierarchy is estimated by minimising the log-likelihood. For more details see (Markov et al. 2014a).

We then rescaled the hierarchy so that the maximum hierarchial value within the 40 region complete subgraph (containing all injected areas) equaled 1:

$$h_{[k]} = \frac{\beta_{[k]}}{max(\beta^{subgraph})} \tag{7}$$

for all cortical areas k in the complete 40-area subgraph.

For the circular embedding of the connectivity data, we estimate angles $\theta_{i,j}$ between areas A_i and A_j so that a smaller angular distance between areas corresponds to a higher connectivity strength (Chaudhuri et al. 2015). The dissimilarity $d(A_i, A_j)$ is defined as

$$d(A_i, A_j) = \begin{cases} -log_{10}(FLN(A_i, A_j)) & for\, FLN(A_i, A_j) \geq 0 \\ -log_{10}(FLN_{min}) & for\, FLN(A_i, A_j) = 0 \end{cases}$$

where $FLN_{min} = 10^{-7}$, a value smaller than any FLN in the dataset.

The angles θ_i are assigned to each area such that

$$d(A_i, A_j) \approx min(|\theta_i - \theta_j|, 2\pi - |\theta_i - \theta_j|)$$

The estimated angles θ_i are constrained to lie within the range $[0, 1]$ and then mapped onto $[0, 2\pi]$.

The radial distance from the center of the circle is $r_i = \sqrt{1 - h_i}$, where h_i is the hierarchical value of the area, as defined above.

Integration of anatomical datasets

All anatomical data was mapped to the appropriate parcellations on the Yerkes19 surface. For the present study, we mapped all data to the 40 area Lyon subgraph (Markov et al. 2014b), as the areas in this parcellation were generally larger than those in the Julich Macaque Brain Atlas (Impieri et al. 2019; Niu et al. 2020; Rapan et al. 2021, this paper) and the Queensland (spine count) injection sites (Elston 2007), and closer to standard areal descriptions than the Vanderbilt (neuronal density) (Collins et al. 2010) sections.

The receptor densities were quantified in 109 cortical regions defined by cyto- and receptor-architecture. The method for the delineation of cortical region borders is described in (Impieri et al. 2019; Niu et al. 2020; Rapan et al. 2021). Using the same method, anatomists (NPG, MN, LR) identified cortical areas on the basis of the receptor and cyto-architecture. See Figure 1 and associated data for the definition of the areas. Anatomists carefully drew (NPG, MN, LR) and independently revised (NPG, MN, LR, SFW) defined borders on the Yerkes19 cortical surface (Donahue et al. 2016) to enable comparison with other data types. The D1 receptor data was mapped to the Lyon atlas as follows. For each area in the Lyon atlas, we searched for overlaps with areas in the Julich Macaque Brain Atlas. If more than 50% of the vertices within the area were also in the Julich Macaque Brain Atlas, the D1 receptor density for the area was calculated. All vertices within each Julich area were assigned the mean value for that area. We averaged the D1 receptor density across all vertices that lay within both the Lyon area and the Julich Macaque Brain Atlas, thus performing a weighted average of the D1 receptor densities according to the degree of spatial overlap. Thirty-two of the 40 Lyon areas were assigned D1 receptor density in this way, with the remaining eight areas not overlapping sufficiently with the Julich Macaque Brain Atlas. Due to the strong positive correlation between the D1 receptor/neuron density and the hierarchy (Fig 1), for the simulations we inferred values for the remaining eight regions using linear regression with hierarchy as the independent variable and D1 receptor/neuron density as the dependent variable.

The *in-vitro* autoradiography data accurately quantifies the density of receptors across cortex. However, it is important to bear in mind that the density of neurons also varies across the cortex. Collins and colleagues measured the density of neurons across the entire macaque cortex using the isotropic fractionator (a.k.a. brain soup) method (Collins et al. 2010). In the original paper, the cortex was divided into 42 regions and displayed on a flatmap, with anatomical landmarks labeled (Fig 2 and S1 of that paper). The borders of these regions were drawn on the Yerkes19 surface by SFW with reference to the original paper (Collins et al. 2010), several anatomical papers from the same group (Beck and Kaas 1999; Cerkevich et al. 2014; Kaas 2004), the Julich Macaque (109 areas) and the Lyon (Markov-132) atlases (Donahue et al. 2016; Markov et al. 2014b), and were independently assessed by anatomists (LR, MN, NPG). The neural density data covered the entire cortex. As such, we assigned neural density to each area in the Lyon atlas, weighted by the spatial overlap with the original regions in the Vanderbilt atlas. D1 receptor density was divided by the neuron density to give the D1 receptor/neuron density in each area. The neuron density was in units of neurons per gram. To estimate the receptor density in fmol per neuron, we used the previously reported figure that 8% of brain tissue is protein (McIlwain and Bachelard 1972). This amounts to multiplying by a constant, and does not affect the correlations or the effect of the dopamine gradient in the model.

The Lyon atlas used to define the interareal connectivity data (Markov et al. 2014b) is already available on the Yerkes19 surface (Donahue et al. 2016). The complete subgraph of injected areas including bidirectional connectivity has been expanded from 29 areas in Donahue et al. 2016 to the 40 areas used in this paper.

For the spine count data, outlines of the 27 injection sites were drawn on the Yerkes19 surface by SFW with reference to the original papers (most of which had substantial anatomical description and hand-drawn maps), as well as anatomical papers cited within the original papers (Cavada and Goldman-Rakic 1989; Preuss and Goldman-Rakic 1991; Seltzer and Pandya 1978) and the Lyon and Julich Macaque Brain Atlases. Direct comparison with the hand-drawn maps was possible for areas V1, V2, MT, LIPv, 7a, V4, TEO, STP, IT, Ant. Cing., Post. Cing, TEpd, 12vl, A1, 3b, 4, 5, 6, 7b, 9, 13, 46, 7m (Elston 2007). Areas 10, 11 and 12 were described with reference to (Preuss and Goldman-Rakic 1991). The injection in area TEa used the maps in Seltzer and Pandya 1978 for area definition. We used these maps to approximate the injection location. Area STP was identified with the corresponding region STPp in the atlas of Felleman and Van Essen (Felleman and Van Essen 1991). Area FEF was identified as lying on the anterior bank of the medial aspect of the arcuate sulcus, as described by Elston (Elston 2007). All identified injection sites on the cortical surface were independently verified by MN, LR and NPG. Spine count data was expressed according to injection sites, rather than entire cortical areas. As such, we found the number of vertices from each injection site overlapping with each area in the Lyon atlas. For each Lyon area, the spine count was an average of the spine counts for all the injection sites overlapping with the area, weighted by the number of vertices of each injection site contained within the area. In this way we estimated the spine counts on pyramidal cells in 24 of the 40 regions in the Lyon atlas. Based on the strong positive correlation between spine count and cortical hierarchy ($r = 0.61$, $p = 0.001$), and following previous work (Chaudhuri et al. 2015; Mejias and Wang 2020), we inferred the spine count for the remaining regions based on the hierarchy using linear regression.

Neuroanatomists (NPG, LR, MN) classified each of the 109 cortical areas for which D1 receptor data is available as being either granular, or agranular, and according to the ratio of cell body size between layers III and V.

Delineations of the areal borders for each atlas, and the anatomical data in the Yerkes19 space are available on the BALSA database.

Overview of dynamical models

We first describe the connectivity structure of our local circuit model, and how dopamine modulates the efficacy of these connections. We then describe a large-scale dynamical model, in which the local circuit is used as a building block, and placed in each of 40 cortical areas. We describe the various steps to building the large-scale model, including how to connect the cortical areas, apply heterogeneity of excitation and the gradient of dopamine. Lastly, we describe how we simulated working memory tasks, lesions and transient inhibition in this model.

Description of the local cortical circuit

We describe a local cortical circuit containing populations of four distinct types of neurons. This is conceptually related to previous computational models of working memory involving multiple types of interneurons (Tanaka 1999; Wang et al. 2004a), and uses a mean field reduction of a spiking model (Brunel and Wang 2001; Wong and Wang 2006). PV, CB/SST and CR/VIP cells differed in the threshold and slope of their input-output function (f-I curve) (Bacci et al.

2003), local (Adesnik et al. 2012; Jiang et al. 2015; Muñoz et al. 2017; Pfeffer et al. 2013; Tremblay et al. 2016) and long-range connectivity (Lee et al. 2013; Wall et al. 2016), adaptation rates (Kawaguchi 1993; Mendonça et al. 2016; Schuman et al. 2019), and NMDA/AMPA ratio (Lu et al. 2007).

The connectivity structure and strengths of the local circuit, are based on a synthesis of anatomical and physiological studies, and are captured in the local connectivity matrix G (Tables 1 and 2) (Jiang et al. 2015; Kalisman et al. 2005; Lee et al. 2013; Ma et al. 2012; Markram et al. 1997; Pfeffer et al. 2013; Silberberg and Markram 2007; Walker et al. 2016). Note that connection probability and synaptic strength between neural types are generally positive correlated (Jiang et al. 2015). This simplifies the process of identifying the relative strengths of connections between neural populations in the circuit.

We grouped the pyramidal neurons into two separate populations. Each of these populations is selective to a particular visual feature (such as a region of visual space). Pyramidal cells excite all cell types in the circuit, with different strengths. We model two compartments in the pyramidal cells. One compartment represents the soma and proximal dendrites, and the other the distal dendrites. The dendrite is modelled as a simplified nonlinear function, adapted from Yang et al. 2016. Pyramidal cells target the soma and proximal dendrites of other pyramidal cells in the same cortical area (Kalisman et al. 2005; Markram et al. 1997; Petreanu et al. 2009). Each type of inhibitory neuron has a unique pattern of connectivity. The first inhibitory cell type targets the perisomatic area of the pyramidal cells. These cells express parvalbumin (PV) and are fast spiking (Jiang et al. 2015; Kawaguchi 1993, 1995). They are basket cells with axons that branch across wide distances, which allows them to inhibit pyramidal cells in neighboring populations (Helmstaedter et al. 2009; Kawaguchi 1995). They also inhibit other PV neurons (Jiang et al. 2015; Pfeffer et al. 2013). Compared to other inhibitory neurons, PV neurons receive a smaller proportion of excitatory inputs via NMDA receptors (Lu et al. 2007; Wang and Gao 2009). The second type of inhibitory neuron targets the distal dendrites of excitatory cells. In non-human primates, dendrite-targeting cells express calbindin (DeFelipe et al. 1989). The best characterised dendrite-targeting cell type in rodents is the Martinotti cell, which expresses somatostatin (CB/SST) (Wang et al. 2004b). These cells target all other cell types, while avoiding other Martinotti cells (Jiang et al. 2015). They also receive a strong lateral projection from pyramidal cells in neighboring columns (Adesnik et al. 2012) and receive most of their excitation via NMDA receptors (Lu et al. 2007). The third type of interneuron expresses calretinin and vasoactive intestinal peptide (CR/VIP) (Tremblay et al. 2016) and targets CB/SST inhibitory neurons (Lee et al. 2013). Although gene expression of PV, SST and VIP have been used to successfully distinguish non-overlapping classes of interneurons in primates (Hodge et al. 2019; Krienen et al. 2020), in primates SST antibodies often label relatively few cells (Hendry et al. 1984; Mueller et al. 2019, 2018). SST is often, but not always co-expressed with CB (González-Albo et al. 2001; Lake et al. 2016). CB and SST expressing cells show a similar pattern of expression across cortical layers and areas in the macaque (Dienel et al. 2020). CR is expressed in most VIP neurons in primate cortex (Gabbott and Bacon 1997; Lake et al. 2016), and both VIP and CR show a similar expression across layers and cortical areas in the macaque (Dienel et al. 2020). However, the investigation of cross-species interneuron type similarities and differences is ongoing and not resolved (Hodge et al. 2019; Kooijmans et al. 2020; Krienen et al. 2020). In our model, the three interneuron types should be more appropriately interpreted according to their synaptic targets, rather than other cellular markers.

See Table 4 for all parameter values.

Dopamine modulation

The density of dopamine D1 receptors per neuron was rescaled, so that the area with minimum density ρ_{min}^{raw} was set to zero, and the area with maximum density ρ_{max}^{raw} was set to one, with all other areas lying in between.

$$\rho_{[k]} = \frac{\rho_{[k]}^{raw} - \rho_{min}^{raw}}{\rho_{max}^{raw} - \rho_{min}^{raw}}$$

for all cortical areas k.

Network behavior was investigated for differing amounts of cortical dopamine availability (λ^{DA}). The specific value of λ^{DA} used for each simulation is shown in the figures and main text. Note that for Figure 6, λ^{DA} is calculated dynamically throughout each trial. Cortical dopamine availability is related to the fraction of occupied D1 receptors λ^{occ} through a sigmoid function. The fraction of occupied D1 receptors thus lies between 0 and 1, as expected.

$$\lambda^{occ} = \frac{e^{b^o(\lambda^{DA} - c^o)}}{1 + e^{b^o(\lambda^{DA} - c^o)}} \tag{8}$$

Dopamine increases the proportion of inhibition onto the dendrites of pyramidal cells (Gao et al. 2003). Therefore, we simulated the effect of dopamine on dendritic inhibition as follows. The total amount of dendritic inhibition increases (from a minimum to a maximum strength) as the total amount of occupied receptors increases. The total amount of occupied receptors is equal to the receptor density multiplied by the fraction of occupied receptors.

$$g_{E_{dend},SST,[k]}^{DA} = g_{E_{dend},SST}^{min} + \lambda^{occ}\rho_{[k]}(g_{E_{dend},SST}^{max} - g_{E_{dend},SST}^{min}) \tag{9}$$

Dopamine decreases the proportion of inhibition onto the soma of pyramidal cells (Gao et al. 2003). Therefore, we simulated the effect of dopamine on somatic inhibition as follows. The total amount of somatic inhibition decreases (from a maximum to a minimum strength) as the total amount of occupied receptors increases.

$$g_{E_{soma},PV,[k]}^{DA} = g_{E_{soma},PV}^{max} + \lambda^{occ}\rho_{[k]}(g_{E_{soma},PV}^{min} - g_{E_{soma},PV}^{max}) \tag{10}$$

Dopamine also increases the strength of excitatory synaptic transmission via NMDA receptors (Seamans et al. 2001). We modeled this with a sigmoid function, so that dopamine primarily increases NMDA conductances at low and medium dopamine concentrations, before reaching a plateau (Brunel and Wang 2001).

$$\nu_{[k]} = \frac{e^{b^\nu(\lambda^{occ}\rho_{[k]} - c^\nu)}}{1 + e^{b^\nu(\lambda^{occ}\rho_{[k]} - c^\nu)}} \tag{11}$$

Here b^ν sets the slope of the sigmoid function, c^ν sets the midpoint.

The effects of dopamine on NMDA transmission is then defined as

$$\nu_{[k]}^{DA} = 1 + \alpha\nu_{[k]} \tag{12}$$

where α controls the strength of dopamine modulation on NMDA transmission.

High levels of D1 agonism lead to a reduction in pyramidal cell firing, particularly during the delay period of working memory tasks. D1 receptor stimulation may lead to inhibition of ongoing activity by engaging an intracellular pathway involving cyclic AMP, protein kinase A and either HCN or KCNQ channels (Arnsten et al. 2019; Gamo et al. 2015; Vijayraghavan et al. 2007). The mechanisms by which HCN channels may hyperpolarise the cell are still under debate (George et al. 2009; Pereira 2014). We simulated an increase in adaptation for very high levels of D1 receptor stimulation with a sigmoid function, so that adaptation increases at high dopamine concentrations, before reaching a plateau.

$$\mu^M_{[k]} = \frac{e^{b^M(\lambda^{occ}\rho_{[k]} - c^M)}}{1 + e^{b^M(\lambda^{occ}\rho_{[k]} - c^M)}} \tag{13}$$

Description of dynamical variables

The neural populations interact via synapses that contain NMDA, AMPA and GABA receptors. Each receptor has its own dynamics, governed by the following equations.

The synaptic variables are updated as follows (Wang 1999; Wong and Wang 2006; Yang et al. 2016)

$$\frac{ds^{NMDA}}{dt} = -\frac{s^{NMDA}}{\tau^{NMDA}} + (1 - s^{NMDA})\gamma_{NMDA} r_E \tag{14}$$

$$\frac{ds^{AMPA}}{dt} = -\frac{s^{AMPA}}{\tau^{AMPA}} + \gamma_{AMPA} r_E \tag{15}$$

$$\frac{ds^{GABA}}{dt} = -\frac{s^{GABA}}{\tau^{GABA}} + \gamma_I r_I \tag{16}$$

$$\frac{ds^{GABA,dend}}{dt} = -\frac{s^{GABA,dend}}{\tau^{GABA,dend}} + \gamma_I r_I \tag{17}$$

where s is the synaptic drive onto a particular receptor type, τ is the time constant of decay of that receptor and γ_{NMDA}, γ_{AMPA} and γ_I are constants. r_E and r_I are the firing rates of the presynaptic excitatory and inhibitory cells targeting the NMDA, AMPA and GABA receptors, calculated below. Note that the inhibition onto the dendrite is slower than inhibition elsewhere ($\tau^{GABA,dend} > \tau^{GABA}$) (Ali and Thomson 2008). Hence we calculate dynamics of dendritic and somatic inhibition separately.

Adaptation acts to reduce the firing rate when the rate is high and has been frequently modeled in the following simple form (Engel and Wang 2011; Hansel and Sompolinsky 1998; Laing and Chow 2002; Shpiro et al. 2007; Theodoni et al. 2011), derived from a spiking model (Liu and Wang 2001; Theodoni et al. 2011)

$$\frac{da}{dt} = -\frac{a}{\tau^a} + r \tag{18}$$

where a is the adaptation variable, τ^A is the adaptation time constant, and r is the firing rate of the neural population.

NMDA/AMPA ratio

The fraction of excitatory postsynaptic current that is dependent on NMDA vs AMPA receptors differs by cell type (e.g. with relatively more current via the NMDA receptors in CB/SST vs PV cells) (Lu et al. 2007). Thus, we allowed the strength of excitatory transmission via NMDA and AMPA receptors to vary by cell type, described in the NMDA fraction, κ (Table 4).

Modulation of excitatory connections by dendritic spines

Approximately 90% of excitatory synapses on neocortical pyramidal cells are on dendritic spines (Nimchinsky et al. 2002). On this basis, we modulate the strength of excitatory connections according to the dendritic spine count.

$$\zeta_{[k]} = \frac{\zeta^{raw}_{[k]} - \zeta^{raw}_{min}}{\zeta^{raw}_{max} - \zeta^{raw}_{min}}$$

for all cortical areas $[k]$.

$$z_{[k]} = z^{min} + \zeta_{[k]}(1 - z^{min}) \tag{19}$$

where z^{min} sets the lower bound for the modulation of excitatory connections by the spine count, ζ.

Description of local currents

The local NMDA current is calculated as follows

$$I^{NMDA,local}_{i,[k]} = z_{[k]}\kappa_i \nu^{DA}_{[k]} \sum_{j\epsilon\{E_1,E_2\}} g^E_{i,j} s^{NMDA}_j \tag{20}$$

where the local excitatory connections via the NMDA receptors are scaled by the NMDA receptor fraction κ_i, the dendritic spine count $z_{[k]}$ and the D1 receptor stimulation $\nu^{DA}_{[k]}$ for all populations of neurons i and cortical areas k.

Similarly local excitatory connections via the AMPA receptors are scaled by the AMPA receptor fraction $1 - \kappa_i$ and the dendritic spine count $z_{[k]}$.

$$I^{AMPA,local}_{i,[k]} = z_{[k]}(1 - \kappa_i) \sum_{j\epsilon\{E_1,E_2\}} g^E_{i,j} s^{AMPA}_j \tag{21}$$

Local inhibitory connections are not explicitly modulated by the dendritic spine count (as spines are the locations of synapses between excitatory cortical neurons). Note however, that the connectivity structure g_{GABA} is modulated by the dopamine receptor density and occupancy (See Tables 1, 2 and 4).

$$I^{GABA}_i = \sum_{j\epsilon\{Inh\}} g^{GABA}_{i,j} s^{GABA}_j \tag{22}$$

where Inh is the set of inhibitory neuron populations.

The currents onto the dendrites are calculated separately, in order to calculate the nonlinear transformation of the current in the dendrite. They depend on the noise and background currents, so are described below.

Description of noise and background currents

Noise is modeled as an Ornstein-Uhlenbeck process, separately for each population.

$$\tau^{AMPA}\frac{dI^{noise}(t)}{dt} = -I^{noise}(t) + \eta(t)\sqrt{\tau^{AMPA}\sigma^2_{noise}} \tag{23}$$

where σ_{noise} is the standard deviation of the noise and η is Gaussian white noise with zero mean and unit variance.

A constant background current I^{bg} was also added to each population (Table 4). This represents input from brain areas that are not explicitly modeled.

Description of the adaptation current

We include adaptation in excitatory cells (Kawaguchi 1993), CB/SST (Kawaguchi 1993, 1995) and CR/VIP cells (Mendonça et al. 2016; Schuman et al. 2019), but not PV cells (Kawaguchi 1993, 1995). This is reflected in their differing adaptation strengths g^{adapt}_{PV} and g^{adapt}, where $g^{adapt}_{PV} = 0$.

The adaptation current is

$$I^{adapt}_{i,[k]} = (g^a_i + g^m_i \mu^M_{[k]})a_{i,[k]} \tag{24}$$

for all local populations i and cortical areas k.

Note that g^a_i represents the non-dopamine dependent adaptation, while $g^m_i \mu^M_{[k]}$ controls the dopamine-dependent adaptation, which depends on both dopamine release and receptor density (equation 13).

Large-scale connectivity structure

Each of the cortical areas is connected using connectivity strengths derived from the retrograde tract-tracing data. Parts of this dataset of been included in previous publications (Markov et al. 2013, 2014a,b). The long-range connectivity matrices are built from the FLN matrix. However, as noted in (Mejias et al. 2016), the FLN matrix spans 5 orders of magnitude. The relationship between anatomical and physiological connectivity strengths is not clear, but if we were to use the raw FLN values in the large-scale model, many of the weaker connections would become irrelevant. To deal with this, we rescale the FLN matrix in order to increase the influence of smaller connections while maintaining the topological structure (Mejias et al. 2016; Mejias and Wang 2020).

$$w_{[k,l]} = \frac{FLN^{b_1}_{[k,l]}}{\sum_{l=1}^{n^{sub}} FLN^{b_1}_{[k,l]}} \tag{25}$$

Here we restrict calculations to the injected cortical areas i, j, which allows us to simulate the complete bidirectional connectivity structure within the subgraph ($n^{sub} = 40$). We use the same parameter values as in (Mejias et al. 2016; Mejias and Wang 2020) (Table 4) to construct our interareal connectivity matrix W .

As noted previously, feedforward projections tend to originate in the supragranular layers, while feedback connections originate in the deep layers. Feedforward and feedback connections also likely have different cellular targets. Therefore it is useful to separate the long-distance feedforward and feedback connections.

$$w^{supra}_{[k,l]} = SLN_{[k,l]} w_{[k,l]} \tag{26}$$

$$w^{infra}_{[k,l]} = (1 - SLN_{[k,l]}) w_{[k,l]} \tag{27}$$

Interareal population interactions

The majority of interareal connections contain a mixture of axons projecting from deep and superficial layers. Long distance connections onto excitatory cells primarily target the distal dendrites (Petreanu et al. 2009) (Table 3). Therefore, in the model we assume that long-distance connections target the dendrites of excitatory cells. CR/VIP cells receive the strongest long-distance inputs of all inhibitory cells, while CB/SST receives the weakest (Lee et al. 2013; Wall et al. 2016) (Table 3, Table 4). This suggests that long-range connections effectively disinhibit the dendrite in the target area by exciting CR/VIP interneurons, while concurrently exciting the dendrite, to maximize the probability of information passing from the source area into the target area. Following Mejias and Wang 2020 we assume that feedback connections target inhibitory cells more strongly than feedforward connections.

Excitatory cells in different cortical areas with the same receptive fields are more likely to be functionally connected (Zandvakili and Kohn 2015). This is reflected in our model as follows. In the source area, there are two excitatory populations, 1 and 2, each sensitive to a particular feature of a visual stimulus (such as a location in the visual field). Likewise in the target area, there are two populations 1 and 2, sensitive to the same visual features. We assume that 90% of the output of population 1 in the source area goes to population 1 in the target area, and the remaining 10% to population 2. The converse is true for population 2 in the source area (it targets 10% population 1, 90% population 2; Table 3, Table 4).

Disinhibitory circuit in the frontal eye fields

The frontal eye fields (areas 8m and 8l in the model), have a very high percentage of calretinin neurons, and relatively fewer parvalbumin and calbindin neurons (Pouget et al. 2009). To account for this in the model, we relatively increased the long-range inputs to CR/VIP cells in areas 8m and 8l, as detailed in Table 4. These changes are critical for persistent activity in areas 8l and 8m, but otherwise do not greatly affect the behavior of the model. Without this change, the overlap between the simulated delay activity pattern and the experimental delay activity pattern (as in Figure 3A) is still extremely high (17/19 areas correct, chi-square = 12.31 p = 0.0004), and the activity pattern depends on both the long-range connectivity (p = 0.001), and D1 receptor distribution (p = 0.008), but not the spine count (p = 0.19), and lesions to areas 8l and 8m have a smaller effect on distributed persistent activity. All other results are unchanged. We also increased the relative strength of local CR/VIP connections and reduced the relative strength of local PV connections in FEF, but found that this had no effect on model behaviour, so the simulations in the paper are presented without the local changes in FEF.

Calculation of long-range currents

Long-range interactions are applied as follows:

$$I_{i[k]}^{NMDA,E,E} = z_{[k]}\mu^{E,E}\nu_{[k]}^{DA}\kappa_i \sum_{l=1}^{n^{sub}} w_{[k,l]}^{supra} \sum_{j\epsilon\{E_1,E_2\}} g_{i,j}^{E,E} S_{j[l]}^{NMDA} \tag{28}$$

where $z_{[k]}$ is the dendritic spine count for area k (as defined above), $\mu^{E,E}$ is the long-range connectivity strength onto excitatory cells (See Table 4), $\nu_{[k]}^{DA}$ is the degree of dopamine modulation of NMDA currents for area k, κ_i is the NMDA/AMPA fraction for population i, $w_{[k,l]}$ is the connection strength from area l to area k, $g_{i,j}^{E,E}$ sets the long-range strength from population j to population i (Tables 3 and 4) and $S_{j[l]}^{NMDA}$ is the synaptic NMDA drive from population j in source area l.

Similarly,

$$I_{i[k]}^{NMDA,I,E} = z_{[k]}\mu^{I,E}\nu_{[k]}^{DA}\kappa_i \sum_{l=1}^{n^{sub}} w_{[k,l]}^{infra} \sum_{j\epsilon\{E_1,E_2\}} g_{i,j}^{I,E} S_{j[l]}^{NMDA} \tag{29}$$

The total long-range current via the NMDA receptors, is simply the concatenation of the two above terms $I^{NMDA,E,E}$ and $I^{NMDA,I,E}$.

$$I^{NMDA,LR} = (I^{NMDA,E,E}, I^{NMDA,I,E}) \tag{30}$$

The long-range AMPA current is calculated similarly,

$$I_{i[k]}^{AMPA,E,E} = z_{[k]}\mu^{E,E}(1-\kappa_i) \sum_{l=1}^{n^{sub}} w_{[k,l]}^{supra} \sum_{j\epsilon\{E_1,E_2\}} g_{i,j}^{E,E} S_{j[l]}^{AMPA} \tag{31}$$

$$I_{i[k]}^{AMPA,I,E} = z_{[k]}\mu^{I,E}(1-\kappa_i) \sum_{l=1}^{n^{sub}} w_{[k,l]}^{infra} \sum_{j\epsilon\{E_1,E_2\}} g_{i,j}^{I,E} S_{j[l]}^{AMPA} \tag{32}$$

$$I^{AMPA,LR} = (I^{AMPA,E,E}, I^{AMPA,I,E}) \tag{33}$$

Description of dendritic currents

The inhibitory current onto the dendrite comes from CB/SST cells and is modulated by dopamine (Table 1, equation 9)

$$I_i^{dend,inh} = \sum_{j\epsilon\{SST1,SST2\}} g_{i,j}^{GABA,dend} s_j^{GABA} \tag{34}$$

The distal dendrites receive long-range input (from neurons in other areas), noise and background input. In addition, if the area receives a stimulus directly, then the external stimulus also targets the dendrites. Note that most local connections target the area around the soma (Markram et al. 1997; Petreanu et al. 2009). This is reflected in the model by having local connections exclusively target the soma compartment of pyramidal cells.

$$I_{i,[k]}^{dend,exc} = I_{i,[k]}^{NMDA,LR} + I_{i,[k]}^{AMPA,LR} + I_{i,[k]}^{stim} + I_{i,[k]}^{noise} + I_{i}^{background} \tag{35}$$

The dendritic nonlinearity is adapted from (Yang et al. 2016) and modeled as follows:

$$I^{soma,dend} = f_I(I^{dend,exc}, I^{dend,inh}) = c_1.\left[\tanh\left(\frac{I^{dend,exc} + c_3 I^{dend,inh} + c_4}{c_5 e^{-I^{dend,inh}/c6}}\right)\right] + c_2 \tag{36}$$

where $I^{soma,dend}$ is the total current passed from the dendrite to the soma, $I^{dend,exc}$ and $I^{dend,inh}$ are the total excitatory and inhibitory current onto the dendrite, respectively. $c1$ to $c6$ control the gain, shift, inversion point and shape of the nonlinear function. These parameters are set to ensure that strong inhibition to the dendrite effectively blocks dendritic activity, but has little effect on somatic firing if the soma is directly stimulated (See Table 4) (Marlin and Carter 2014).

Application of external stimuli for tasks

In all simulations, the first stimulus is applied for 400ms. The second stimulus (Figures 4-6) is applied 600ms after the removal of the target stimulus for another 400ms. The two stimuli are of equal strength and duration, although the results are robust to a range of stimulus strengths (See Table 4 for parameter values). For Figures 2-6 in the main text, a stimulus was applied to the dendrite of excitatory population 1 in area V1. For Figures 3-6 a second stimulus was applied to the dendrite of excitatory population 2 of area V1. For Supplementary Figures 4 and 5, the stimuli were applied to area 3 of primary somatosensory cortex instead. In all equations, the target and distractor stimuli are designated by the term I^{stim}.

Total current in large-scale model

The total current equals the sum of all long-range, local and external inputs, and intrinsic currents.

$$I_{total} = I^{NMDA,LR} + I^{AMPA,LR} + I^{NMDA,local} + I^{AMPA,local} + I^{GABA,local} + I^{soma,dend} + I^{adapt} + I^{noise} + I^{bg} + I^{stim} \tag{37}$$

Description of f-I curves

The f-I (current to frequency) curve of the excitatory population is

$$f(I_E^{total}) = \frac{aI_E^{total} - b}{1 - e^{-d(aI_E^{total} - b)}} \tag{38}$$

where r_E is the firing rate of an populations of excitatory cells, I_E^{total} is the total input to the population, a is a gain factor, d determines the shape of $f(I_E^{total})$, such that if d is large, $f(I_E^{total})$ acts like a threshold-linear function, with threshold b (Abbott and Chance 2005).

The f-I curves for the inhibitory neuron populations are modeled using a threshold-linear function

$$f(I_i^{total}) = \left\{ \begin{array}{ll} c_i I_i^{total} + r_i^0 & \text{for } I_i^{total} \geq -r_i^0/c_i \\ 0, & \text{otherwise} \end{array} \right\} \tag{39}$$

where r_i is the firing rate of a population of inhibitory cells, I_i^{total} is the total input to the population.

The threshold r_i^0 and slope c_i depend on the cell type i (Bacci et al. 2003). See Table 4 for parameter values.

The firing rates are updated as follows

$$\tau^{AMPA}\frac{dR}{dt} = -R + f(I^{total}) \tag{40}$$

for all cell types.

Short-term synaptic plasticity

For Figure 4, we added short-term plasticity to synapses from excitatory cells to excitatory cells (Hempel et al. 2000; Wang et al. 2004b) and CB/SST cells (Lee et al. 2013; Silberberg and Markram 2007) as follows (Mongillo et al. 2008).

$$\frac{ds^{NMDA}}{dt} = -\frac{s^{NMDA}}{\tau^{NMDA}} + xu(1 - s^{NMDA})\gamma_{NMDA}\gamma_{xu}r_E \tag{41}$$

$$\frac{ds^{AMPA}}{dt} = -\frac{s^{AMPA}}{\tau^{AMPA}} + xu\gamma_{AMPA}\gamma_{xu}r_E \tag{42}$$

$$\frac{du}{dt} = \frac{U - u}{\tau^u} + U(1 - u)r_E \tag{43}$$

$$\frac{dx}{dt} = \frac{1 - x}{\tau^x} - uxr_E \tag{44}$$

with $U = 0.2, \tau^u = 1.5s, \tau^x = 0.2s$, as in Mongillo et al. 2008. We also added a term $\gamma_{xu} = 2.5$ to account for the fact that the product xu is usually less then 1, and to keep firing rates similar to those in other simulations.

Simulated transient inhibition of SST2 populations

In Figure 5, we simulate the effects of transient inhibition to the SST2 populations in cortical areas in the frontoparietal network. The frontoparietal network is defined according to the results of Leavitt et al. 2017, as in Figure 3. To do this, we apply an external inhibitory stimulus of 0.1nA to these populations for the duration of the distractor stimulus.

Dynamics and connectivity within VTA

For Figure 6, we investigate whether the dynamics of dopamine release can be learned in order to selectively maintain the desired working memory content. Previous cortico-basal ganglia models have tackled similar problems (Braver and Cohen 2000; Frank 2005). Note both dopaminergic and GABAergic cells in the VTA receive excitatory input from the cortex, while the majority of inhibition to dopaminergic cells comes from local VTA GABAergic cells (Soden et al. 2020).

The total current input to the dopamine cells in VTA is

$$I_{DA}^{total} = I_{DA}^{bg} + \sum_{k=1}^{n_{areas}}\sum_{j=1}^{2} c_{Ej}^{vta,ctx} g_{DA,E_j}^{vta,ctx} S_{NMDA,E_j}^{k} + \sum_{k=1}^{n_{areas}}\sum_{j=1}^{2} c_{Ej}^{vta,ctx} g_{DA,E_j}^{vta,ctx} S_{AMPA,E_j}^{k} + g_{DA,I}^{vta} S_{GABA}^{vta}$$

where $g^{vta,ctx}_{DA,E_j}$ sets the maximum strength of cortical-VTA connections. $c^{vta,ctx}_{Ej}$ is the fraction of synapses in an up state (Soltani and Wang 2006), and is updated via reinforcement learning (see below). Initial values are $c^{vta,ctx}_1 = 0.7$, $c^{vta,ctx}_2 = 1$. $g^{vta,ctx}_{DA,E_j} = 0.047nA$ and $g^{vta}_{DA,I} = -0.55nA$.

The input to VTA inhibitory cells is

$$I^{total}_{vta,I} = I^{bg}_{vta,I} + \sum_{k=1}^{n_{areas}} \sum_{j=1}^{2} c^{vta,ctx}_{Ej} g^{vta,ctx}_{I,E_j} S^k_{NMDA,E_j} + \sum_{k=1}^{n_{areas}} \sum_{j=1}^{2} c^{vta,ctx}_{Ej} g^{vta,ctx}_{I,E_j} S^k_{AMPA,E_j}$$

where $g^{vta,ctx}_{I,E_j} = 0.02nA$

Synaptic inputs to the VTA inhibitory are driven by facilitating synapses (Soden et al. 2020), as in equations 41-44, but with $x = 0.87$ held constant and $\tau^u = 200ms$

The firing rates of the dopamine cells r_{DA} as in equations 38 and 40. The firing rates of GABAergic cells are updated as in equations 39-40.

Cortical dopamine availability

Dopamine neurons fire bursts in response to stimuli that predict reward in working memory tasks (Schultz et al. 1993). Following release in the cortex, dopamine levels remain elevated for seconds (Muller et al. 2014). This is approximately the period of one trial in our simulations. Therefore, for the majority of simulations we approximated this by setting dopamine to a constant value for each trial.

For Figure 6 the cortical model is the same as in previous figures, with the exception that dopamine availability in the cortex λ^{DA} changes dynamically and depends on the firing rates in the dopamine neurons.

$$\frac{d\lambda^{DA}}{dt} = -\frac{\lambda^{DA}}{\tau^{DA}} + \gamma_{DA} r_{DA}$$

where $\tau^{DA} = 3s$ and $\gamma_{DA} = 0.05$. In addition, we removed the effect of dopamine on adaptation currents to simplify the learning process.

Reward-based learning

The fraction of cortex to VTA synapses in the up state is updated according to the outcome of the previous trial, using the simplified learning rule of Soltani and Wang 2006

$$c_{Ej}(T+1) = c_{Ej}(T) + \alpha[1 - c_{Ej}(T)]$$

if target j is selected and rewarded and

$$c_{Ej}(T+1) = c_{Ej}(T) - \alpha[c_{Ej}(T)]$$

if target j is selected and not rewarded. T is the current trial and $\alpha = 0.2$ is the learning rate.

QUANTIFICATION AND STATISTICAL ANALYSIS

Correlation between D1 receptor density and other anatomical features

Many aspects of brain anatomy are spatially autocorrelated, with nearby brain areas displaying similar anatomy. This spatial autocorrelation is not accounted for in conventional statistical tests, which often assume independence of data points. Failing to account for the spatial autocorrelation can lead to spurious correlations between brain maps. To overcome this problem, we generated random surrogate brain maps, with a spatial autocorrelation that closely matched the hierarchy map (Burt et al. 2020). This is done by first randomly permuting the values in the hierarchy map, and then smoothing and rescaling the permuted map to recover the lost spatial autocorrelation. The smoothing is perfomed by a local kernel-weighted sum of values of the k nearest neighbour regions, where k is chosen to best match the autocorrelation of the original hierarchy map (Burt et al. 2020). Each of the randomly generated surrogate maps is then correlated with the D1 receptor map. The spatially-corrected p-value is then the fraction of surrogate maps that show a stronger Pearson correlation (negative or positive) with the D1 receptor map than the hierarchy map.

To compare the D1R density between granular and agranular cortical areas, we used a non-parametric Wilcoxon rank-sum test. To compare D1R density between areas with internopyramidisation, externopyramidisation and equal layer III and layer V pyramid sizes, we used a non-parametric Kruskall-Wallis test.

Comparing the simulated and experimental patterns of delay activity

In Figure 3A and 3B we compare the activity pattern of the model to the experimental pattern, and investigate its dependence on anatomical features. The experimental electrophysiology data was taken from a mega-analysis by Leavitt and colleagues of over 90 electrophysiology studies of delay period activity during working memory tasks (Leavitt et al. 2017). We first divided the cortex into persistent activity and non-persistent activity areas for both the experimental data and simulation (Supplementary Table 1). Areas were classified in the persistent activity group if at least 3 more studies observed persistent delay period activity than a lack of such activity. We excluded areas that have been assessed in less than three studies. Of the areas that have been studied in at least three studies, we classify an area as having persistent activity, if more than 50% of studies have found persistent activity. However, the conclusions are not dependent on this threshold, or the minimum number of studies (Supplementary Table 2). Areas in the simulation were classified as having persistent activity if, for the last 500ms of the trial, they had mean firing rates of at least 5Hz greater than the pre-stimulus baseline firing rates.

To shuffle anatomical connections, we shuffled connections within rows of the FLN matrix, so that the distribution of connections and connection strengths to each area remained constant, with the identity of the connections changing. The same reordering was applied to the SLN matrix. D1 receptor densities and spine counts were shuffled separately. Results were visualised using a custom version of a Raincloud Plot (Allen et al. 2019) to enable concurrent visualisation of the distribution and individual simulation results. The p-value is calculated as the fraction of simulations based on shuffled anatomical data that produce a delay activity pattern that overlaps with the experimental data as well as (or better than) the original simulation.

Lesioning of cortical areas

In Figure 3C-F, we simulate the effects of a lesion to individual cortical areas. We do this by removing all input and output connections of the lesioned area in the connectivity matrices $W^{E,E}$ and $W^{I,E}$. For the statistical analysis of the relationship between anatomical features and visual effects, we removed areas V1 and V2 from the analysis. This was due to the fact that these areas were crucial to the propagation of the visual stimulus, but not working memory *per se* (Fig 3 and Fig S3). We performed a stepwise-linear regression approach.

DATA AND SOFTWARE AVAILABILITY

Software was written in the Python (https://www.python.org/) and MATLAB (https://www.mathworks.com/products/matlab.html) programming languages. Network simulations for the cortical network and cortical-VTA networks are available at the GitHub repository: https:// github.com/seanfw/dopamine-dist-wm. The cortical representation of all anatomical data used in this study is available via the BALSA database: (https://balsa.wustl.edu/).

References

Abbas, Atheir I., Marina J. M. Sundiang, Britt Henoch, Mitchell P. Morton, Scott S. Bolkan, Alan J. Park, Alexander Z. Harris, Christoph Kellendonk, and Joshua A. Gordon (2018). "Somatostatin interneurons facilitate hippocampal-prefrontal synchrony and prefrontal spatial encoding". *Neuron* 100.4, 926–939.e3.

Abbott, Larry F. and Frances S. Chance (2005). "Drivers and modulators from push-pull and balanced synaptic input". In: *Progress in Brain Research*. Vol. 149. Cortical Function: a View from the Thalamus. Elsevier, pp. 147–155.

Abi-Dargham, Anissa, Osama Mawlawi, Ilise Lombardo, Roberto Gil, Diana Martinez, Yiyun Huang, Dah-Ren Hwang, John Keilp, Lisa Kochan, Ronald Van Heertum, Jack M. Gorman, and Marc Laruelle (2002). "Prefrontal Dopamine D1 Receptors and Working Memory in Schizophrenia". *Journal of Neuroscience* 22.9, pp. 3708–3719.

Adesnik, Hillel, William Bruns, Hiroki Taniguchi, Z. Josh Huang, and Massimo Scanziani (2012). "A neural circuit for spatial summation in visual cortex". *Nature* 490.7419, p. 226.

Ali, Afia B. and Alex M. Thomson (2008). "Synaptic α5 subunit–containing GABAA receptors mediate IPSPs elicited by dendrite-preferring cells in rat neocortex". *Cerebral Cortex* 18.6, pp. 1260–1271.

Allen, Micah, Davide Poggiali, Kirstie Whitaker, Tom Rhys Marshall, and Rogier A. Kievit (2019). "Raincloud plots: a multi-platform tool for robust data visualization". *Wellcome open research* 4.

Ardid, Salva, Xiao-Jing Wang, and Albert Compte (2007). "An Integrated Microcircuit Model of Attentional Processing in the Neocortex". *Journal of Neuroscience* 27.32, pp. 8486–8495.

Arnatkeviciute, Aurina, Ben D. Fulcher, and Alex Fornito (2019). "A practical guide to linking brain-wide gene expression and neuroimaging data". *NeuroImage* 189, pp. 353–367.

Arnsten, Amy F. T. (2000). "Through the Looking Glass: Differential Noradenergic Modulation of Prefrontal Cortical Function". *Neural Plasticity* 7.1-2, pp. 133–146.

Arnsten, Amy F. T., Dibyadeep Datta, and Min Wang (2020). "The genie in the bottle-magnified calcium signaling in dorsolateral prefrontal cortex". *Molecular Psychiatry*, pp. 1–17.

Arnsten, Amy F. T., Lu E. Jin, Nao J. Gamo, Brian Ramos, Constantinos D. Paspalas, Yury M. Morozov, Anna Kata, Nigel S. Bamford, Mark F. Yeckel, Leonard K. Kaczmarek, and Lynda El-Hassar (2019). "Role of KCNQ potassium channels in stress-induced deficit of working memory". *Neurobiology of Stress* 11, p. 100187.

Arnsten, Amy F. T., Min J. Wang, and Constantinos D. Paspalas (2012). "Neuromodulation of Thought: Flexibilities and Vulnerabilities in Prefrontal Cortical Network Synapses". *Neuron* 76.1, pp. 223–239.

Bacci, Alberto, Uwe Rudolph, John R. Huguenard, and David A. Prince (2003). "Major differences in inhibitory synaptic transmission onto two neocortical interneuron subclasses". *Journal of Neuroscience* 23.29, pp. 9664–9674.

Badre, David and Mark D'esposito (2009). "Is the rostro-caudal axis of the frontal lobe hierarchical?" *Nature Reviews Neuroscience* 10.9, pp. 659–669.

Barbosa, Joao, Heike Stein, Rebecca L. Martinez, Adrià Galan-Gadea, Sihai Li, Josep Dalmau, Kirsten C. S. Adam, Josep Valls-Solé, Christos Constantinidis, and Albert Compte (2020). "Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory". *Nature Neuroscience*, pp. 1–9.

Barone, Pascal, Alexandre Batardiere, Kenneth Knoblauch, and Henry Kennedy (2000). "Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule". *Journal of Neuroscience* 20.9, pp. 3263–3281.

Beck, Pamela D. and Jon H. Kaas (1999). "Cortical connections of the dorsomedial visual area in Old World macaque monkeys". *Journal of Comparative Neurology* 406.4, pp. 487–502.

Beliveau, Vincent, Melanie Ganz, Ling Feng, Brice Ozenne, Liselotte Højgaard, Patrick M. Fisher, Claus Svarer, Douglas N. Greve, and Gitte M. Knudsen (2017). "A High-Resolution In Vivo Atlas of the Human Brain's Serotonin System". *Journal of Neuroscience* 37.1, pp. 120–128.

Beukers, Andre O., Timothy J. Buschman, Jonathan D. Cohen, and Kenneth A. Norman (2021). "Is Activity Silent Working Memory Simply Episodic Memory?" *Trends in Cognitive Sciences* 25.4, pp. 284–293.

Braver, Todd S and Jonathan D Cohen (2000). "On the control of control: the role of dopamine in regulating prefrontal function and working memory". In: *Control of Cognitive Processes*. Attention and Performance XVIII, pp. 713–737.

Brozoski, T. J., R. M. Brown, H. E. Rosvold, and P. S. Goldman (1979). "Cognitive deficit caused by regional depletion of dopamine in prefrontal cortex of rhesus monkey". *Science* 205.4409, pp. 929–932.

Brunel, Nicolas and Xiao-Jing Wang (2001). "Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition". *Journal of Computational Neuroscience* 11.1, pp. 63–85.

Burt, Joshua B., Murat Demirtaş, William J. Eckner, Natasha M. Navejar, Jie Lisa Ji, William J. Martin, Alberto Bernacchia, Alan Anticevic, and John D. Murray (2018). "Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography". *Nature Neuroscience* 21.9, p. 1251.

Burt, Joshua B., Markus Helmer, Maxwell Shinn, Alan Anticevic, and John D. Murray (2020). "Generative modeling of brain maps with spatial autocorrelation". *NeuroImage* 220, p. 117038.

Cavada, Carmen and Patricia S. Goldman-Rakic (1989). "Posterior parietal cortex in rhesus monkey: I. Parcellation of areas based on distinctive limbic and sensory corticocortical connections". *Journal of Comparative Neurology* 287.4, pp. 393–421.

Cerkevich, Christina M., Hui-Xin Qi, and Jon H. Kaas (2014). "Corticocortical projections to representations of the teeth, tongue, and face in somatosensory area 3b of macaques". *Journal of Comparative Neurology* 522.3, pp. 546–572.

Chafee, Matthew V. and Patricia S. Goldman-Rakic (2000). "Inactivation of Parietal and Prefrontal Cortex Reveals Interdependence of Neural Activity During Memory-Guided Saccades". *Journal of Neurophysiology* 83.3, pp. 1550–1566.

Chaudhuri, Rishidev, Kenneth Knoblauch, Marie-Alice Gariel, Henry Kennedy, and Xiao-Jing Wang (2015). "A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex". *Neuron* 88.2, pp. 419–431.

Choi, Jung Yoon, Hee Jae Jang, Sharon Ornelas, Weston T. Fleming, Daniel Fürth, Jennifer Au, Akhil Bandi, Esteban A. Engel, and Ilana B. Witten (2020). "A Comparison of Dopaminergic and Cholinergic Populations Reveals Unique Contributions of VTA Dopamine Neurons to Short-Term Memory". *Cell Reports* 33.11.

Collins, Christine E., David C. Airey, Nicole A. Young, Duncan B. Leitch, and Jon H. Kaas (2010). "Neuron densities vary across and within cortical areas in primates". *Proceedings of the National Academy of Sciences* 107.36, pp. 15927–15932.

Constantinidis, Christos, Shintaro Funahashi, Daeyeol Lee, John D. Murray, Xue-Lian Qi, Min Wang, and Amy F. T. Arnsten (2018). "Persistent Spiking Activity Underlies Working Memory". *Journal of Neuroscience* 38.32, pp. 7020–7028.

Croxson, Paula L., Diana A. Kyriazis, and Mark G. Baxter (2011). "Cholinergic modulation of a specific memory function of prefrontal cortex". *Nature Neuroscience* 14.12, pp. 1510–1512.

DeFelipe, Javier, Stewart H. C. Hendry, and Edward G. Jones (1989). "Synapses of double bouquet cells in monkey cerebral cortex visualized by calbindin immunoreactivity". *Brain Research* 503.1, pp. 49–54.

Dienel, Samuel J, Andrew J Ciesielski, Holly H Bazmi, Elizabeth A Profozich, Kenneth N Fish, and David A Lewis (2020). "Distinct Laminar and Cellular Patterns of GABA Neuron Transcript Expression in Monkey Prefrontal and Visual Cortices". *Cerebral Cortex* bhaa341.

Donahue, C. J., S. N. Sotiropoulos, S. Jbabdi, M. Hernandez-Fernandez, T. E. Behrens, T. B. Dyrby, T. Coalson, H. Kennedy, K. Knoblauch, D. C. Van Essen, and M. F. Glasser (2016). "Using diffusion tractography to predict cortical connection strength and distance: a quantitative comparison with tracers in the monkey". *J. Neurosci.* 36, pp. 6758–6770.

Dunning, Darren L., Briony Westgate, and Anna-Lynne R. Adlam (2016). "A meta-analysis of working memory impairments in survivors of moderate-to-severe traumatic brain injury." *Neuropsychology* 30.7, pp. 811–819.

Egerton, Alice, Arsime Demjaha, Philip McGuire, Mitul A. Mehta, and Oliver D. Howes (2010). "The test–retest reliability of 18F-DOPA PET in assessing striatal and extrastriatal presynaptic dopaminergic function". *NeuroImage* 50.2, pp. 524–531.

Elston, Guy N. (2007). "Specialization of the neocortical pyramidal cell during primate evolution". In: *Evolution of Nervous Systems*. Elsevier, pp. 191–242.

Engel, Tatiana A. and Xiao-Jing Wang (2011). "Same or different? A neural circuit mechanism of similarity-based pattern match decision making". *Journal of Neuroscience* 31.19, pp. 6982–6996.

Fallon, Sean James, Rozemarijn Margaretha Mattiesing, Kinan Muhammed, Sanjay Manohar, and Masud Husain (2017). "Fractionating the neurocognitive mechanisms underlying working memory: Independent effects of dopamine and Parkinson's disease". *Cerebral Cortex* 27.12, pp. 5727–5738.

Fallon, Sean James, Marieke E. van der Schaaf, Niels ter Huurne, and Roshan Cools (2016). "The neurocognitive cost of enhancing cognition with methylphenidate: improved distractor resistance but impaired updating". *Journal of Cognitive Neuroscience* 29.4, pp. 652–663.

Farde, Lars, Stefan Pauli, Håkan Hall, Lars Eriksson, Christer Halldin, Thomas Högberg, Lars Nilsson, Irene Sjögren, and Sharon Stone-Elander (1988). "Stereoselective binding of 11C-raclopride in living human brain — a search for extrastriatal central D2-dopamine receptors by PET". *Psychopharmacology* 94.4, pp. 471–478.

Felleman, Daniel J. and David C. Van Essen (1991). "Distributed hierarchical processing in the primate cerebral cortex." *Cerebral cortex* 1.1, pp. 1–47.

Frank, Michael J. (2005). "Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism". *Journal of Cognitive Neuroscience* 17.1, pp. 51–72.

Froudist-Walsh, Sean, Michael AP Bloomfield, Mattia Veronese, Jasmin Kroll, Vyacheslav R. Karolis, Sameer Jauhar, Ilaria Bonoldi, Philip K. McGuire, Shitij Kapur, Robin M. Murray, Chiara Nosarti, and Oliver Howes (2017a). "The effect of perinatal brain injury on dopaminergic function and hippocampal volume in adult life". *eLife* 6, e29088.

Froudist-Walsh, Sean, Diana López-Barroso, María José Torres-Prioris, Paula Croxson, and Marcelo L. Berthier (2017b). "Plasticity in the working memory system: life span changes and response to injury". *The Neuroscientist*, p. 1073858417717210.

Froudist-Walsh, Sean, Ting Xu, Meiqi Niu, Lucija Rapan, Karl Zilles, Daniel S. Margulies, Xiao-Jing Wang, and Nicola Palomero-Gallagher (2021). "Gradients of receptor expression in the macaque cortex". *bioRxiv*, p. 2021.02.22.432173.

Fulcher, Ben D., John D. Murray, Valerio Zerbi, and Xiao-Jing Wang (2019). "Multimodal gradients across mouse cortex". *Proceedings of the National Academy of Sciences* 116.10, pp. 4689–4695.

Funahashi, Shintaro, Charles J. Bruce, and Patricia S. Goldman-Rakic (1989). "Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex". *Journal of Neurophysiology* 61.2, pp. 331–349.

Fuster, Joaquin M. and Garrett E. Alexander (1971). "Neuron activity related to short-term memory". *Science* 173.3997, pp. 652–654.

Gabbott, P. L. A and S. J Bacon (1997). "Vasoactive intestinal polypeptide containing neurones in monkey medial prefrontal cortex (mPFC): colocalisation with calretinin". *Brain Research* 744.1, pp. 179–184.

Gallyas, Ferenc (1979). "Silver staining of myelin by means of physical development". *Neurological Research* 1.2, pp. 203–209.

Galvin, Veronica C., Sheng Tao Yang, Constantinos D. Paspalas, Yang Yang, Lu E. Jin, Dibyadeep Datta, Yury M. Morozov, Taber C. Lightbourne, Adam S. Lowet, Pasko Rakic, Amy F. T. Arnsten, and Min Wang (2020). "Muscarinic M1 Receptors Modulate Working Memory Performance and Activity via KCNQ Potassium Channels in the Primate Prefrontal Cortex". *Neuron* 106.4, 649–661.e4.

Gamo, Nao J., Gyorgy Lur, Michael J. Higley, Min Wang, Constantinos D. Paspalas, Susheel Vijayraghavan, Yang Yang, Brian P. Ramos, Kathy Peng, Anna Kata, Lindsay Boven, Faith Lin, Lisette Roman, Daeyeol Lee, and Amy F. T. Arnsten (2015). "Stress impairs prefrontal cortical function via d1 dopamine receptor interactions with hyperpolarization-activated cyclic nucleotide-gated channels". *Biological Psychiatry*. Mechanisms of Stress, Memory, and Posttraumatic Stress Disorder 78.12, pp. 860–870.

Gao, Wen-Jun, Yun Wang, and Patricia S. Goldman-Rakic (2003). "Dopamine modulation of perisomatic and peridendritic inhibition in prefrontal cortex". *Journal of Neuroscience* 23.5, pp. 1622–1630.

George, Meena S., L. F. Abbott, and Steven A. Siegelbaum (2009). "HCN hyperpolarization-activated cation channels inhibit EPSPs by interactions with M-type K + channels". *Nature Neuroscience* 12.5, pp. 577–584.

Goldman-Rakic, P. S (1995). "Cellular basis of working memory". *Neuron* 14.3, pp. 477–485.

Goldman-Rakic, Patricia S., M. S. Lidow, and Dorothy. W. Gallager (1990). "Overlap of dopaminergic, adrenergic, and serotoninergic receptors and complementarity of their subtypes in primate prefrontal cortex". *Journal of Neuroscience* 10.7, pp. 2125–2138.

González-Albo, María C., Guy N. Elston, and Javier DeFelipe (2001). "The Human Temporal Cortex: Characterization of Neurons Expressing Nitric Oxide Synthase, Neuropeptides and Calcium-binding Proteins, and their Glutamate Receptor Subunit Profiles". *Cerebral Cortex* 11.12, pp. 1170–1181.

Goulas, Alexandros, Karl Zilles, and Claus C. Hilgetag (2018). "Cortical Gradients and Laminar Projections in Mammals". *Trends in Neurosciences* 41.11, pp. 775–788.

Guo, Zengcai V., Hidehiko K. Inagaki, Kayvon Daie, Shaul Druckmann, Charles R. Gerfen, and Karel Svoboda (2017). "Maintenance of persistent activity in a frontal thalamocortical loop". *Nature* 545.7653, pp. 181–186.

Hamid, Arif A., Michael J. Frank, and Christopher I. Moore (2021). "Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment". *Cell* 184.10, 2733–2749.e16.

Hansel, David and Haim Sompolinsky (1998). "Modeling Feature Selectivity in Local Cortical Circuits". In: *Methods in Neuronal Modeling: From Synapse to Networks*, pp. 499–567.

Harnett, Mark T., Brian E. Bernier, Kee-Chan Ahn, and Hitoshi Morikawa (2009). "Burst-Timing-Dependent Plasticity of NMDA Receptor-Mediated Transmission in Midbrain Dopamine Neurons". *Neuron* 62.6, pp. 826–838.

Helmstaedter, Moritz, Bert Sakmann, and Dirk Feldmeyer (2009). "Neuronal correlates of local, lateral, and translaminar inhibition with reference to cortical columns". *Cerebral Cortex* 19.4, pp. 926–937.

Hempel, Chris M., Kenichi H. Hartman, X.-J. Wang, Gina G. Turrigiano, and Sacha B. Nelson (2000). "Multiple Forms of Short-Term Plasticity at Excitatory Synapses in Rat Medial Prefrontal Cortex". *Journal of Neurophysiology* 83.5, pp. 3031–3041.

Hendry, S. H., E. G. Jones, and P. C. Emson (1984). "Morphology, distribution, and synaptic relations of somatostatin- and neuropeptide Y-immunoreactive neurons in rat and monkey neocortex". *Journal of Neuroscience* 4.10, pp. 2497–2517.

Hodge, Rebecca D. et al. (2019). "Conserved cell types with divergent features in human versus mouse cortex". *Nature* 573.7772, pp. 61–68.

Huang, Chengcheng, Douglas A. Ruff, Ryan Pyle, Robert Rosenbaum, Marlene R. Cohen, and Brent Doiron (2019). "Circuit Models of Low-Dimensional Shared Variability in Cortical Networks". *Neuron* 101.2, 337–348.e4.

Hurd, Yasmin L., Michio Suzuki, and Göran C. Sedvall (2001). "D1 and D2 dopamine receptor mRNA expression in whole hemisphere sections of the human brain". *Journal of Chemical Neuroanatomy* 22.1, pp. 127–137.

Impieri, Daniele, Karl Zilles, Meiqi Niu, Lucija Rapan, Nicole Schubert, Claudio Galletti, and Nicola Palomero-Gallagher (2019). "Receptor density pattern confirms and enhances the anatomic-functional features of the macaque superior parietal lobule areas". *Brain Structure and Function* 224.8, pp. 2733–2756.

Jaramillo, Jorge, Jorge F. Mejias, and Xiao-Jing Wang (2019). "Engagement of Pulvino-cortical Feedforward and Feedback Pathways in Cognitive Computations". *Neuron* 101.2, 321–336.e9.

Jiang, Xiaolong, Shan Shen, Cathryn R. Cadwell, Philipp Berens, Fabian Sinz, Alexander S. Ecker, Saumil Patel, and Andreas S. Tolias (2015). "Principles of connectivity among morphologically defined cell types in adult neocortex". *Science* 350.6264, aac9462.

Jun, James J. et al. (2017). "Fully integrated silicon probes for high-density recording of neural activity". *Nature* 551.7679, pp. 232–236.

Kaas, Jon H. (2004). "Somatosensory system". *The Human Nervous System*, pp. 1059–1092.

Kalisman, Nir, Gilad Silberberg, and Henry Markram (2005). "The neocortical microcircuit as a tabula rasa". *Proceedings of the National Academy of Sciences* 102.3, pp. 880–885.

Kawaguchi, Yasuo (1993). "Groupings of nonpyramidal and pyramidal cells with specific physiological and morphological characteristics in rat frontal cortex". *Journal of Neurophysiology* 69.2, pp. 416–431.

– (1995). "Physiological subgroups of nonpyramidal cells with specific morphological characteristics in layer II/III of rat frontal cortex". *Journal of Neuroscience* 15.4, pp. 2638–2655.

Kooijmans, Roxana N, Wesley Sierhuis, Matthew W Self, and Pieter R Roelfsema (2020). "A quantitative comparison of inhibitory interneuron size and distribution between mouse and macaque v1, using calcium-binding proteins". *Cerebral Cortex Communications* 1.tgaa068.

Krienen, Fenna M. et al. (2020). "Innovations present in the primate interneuron repertoire". *Nature*, pp. 1–8.

Lafuente, V. de and R. Romo (2012). "Dopaminergic activity coincides with stimulus detection by the frontal lobe". *Neuroscience* 218, pp. 181–184.

Laing, Carlo R. and Carson C. Chow (2002). "A Spiking Neuron Model for Binocular Rivalry". *Journal of Computational Neuroscience* 12.1, pp. 39–53.

Lake, Blue B. et al. (2016). "Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain". *Science* 352.6293, pp. 1586–1590.

Leavitt, Matthew L., Diego Mendoza-Halliday, and Julio C. Martinez-Trujillo (2017). "Sustained Activity Encoding Working Memories: Not Fully Distributed". *Trends in Neurosciences* 40.6, pp. 328–346.

Lee, Soohyun, Illya Kruglikov, Z. Josh Huang, Gord Fishell, and Bernardo Rudy (2013). "A disinhibitory circuit mediates motor integration in the somatosensory cortex". *Nature Neuroscience* 16.11, pp. 1662–1670.

Li, Bao-Ming and Zhen-Tong Mei (1994). "Delayed-response deficit induced by local injection of the α2-adrenergic antagonist yohimbine into the dorsolateral prefrontal cortex in young adult monkeys". *Behavioral and Neural Biology* 62.2, pp. 134–139.

Lidow, Michael S., Patricia S. Goldman-Rakic, D. W. Gallager, and Pasko Rakic (1991). "Distribution of dopaminergic receptors in the primate cerebral cortex: Quantitative autoradiographic analysis using [3H]raclopride, [3H]spiperone and [3H]SCH23390". *Neuroscience* 40.3, pp. 657–671.

Liu, Ying-Hui and Xiao-Jing Wang (2001). "Spike-Frequency Adaptation of a Generalized Leaky Integrate-and-Fire Model Neuron". *Journal of Computational Neuroscience* 10.1, pp. 25–45.

Lu, Jiang-teng, Cheng-yu Li, Jian-Ping Zhao, Mu-ming Poo, and Xiao-hui Zhang (2007). "Spike-timing-dependent plasticity of neocortical excitatory synapses on inhibitory interneurons depends on target cell type". *Journal of Neuroscience* 27.36, pp. 9711–9720.

Lundqvist, Mikael, Pawel Herman, and Earl K. Miller (2018). "Working memory: delay activity, yes! Persistent activity? Maybe not". *Journal of Neuroscience* 38.32, pp. 7013–7019.

Ma, Yunyong, Hang Hu, and Ariel Agmon (2012). "Short-term plasticity of unitary inhibitory-to-inhibitory synapses depends on the presynaptic interneuron subtype". *Journal of Neuroscience* 32.3, pp. 983–988.

Manohar, Sanjay G., Nahid Zokaei, Sean J. Fallon, Tim P. Vogels, and Masud Husain (2019). "Neural mechanisms of attending to items in working memory". *Neuroscience & Biobehavioral Reviews* 101, pp. 1–12.

Margulies, Daniel S., Satrajit S. Ghosh, Alexandros Goulas, Marcel Falkiewicz, Julia M. Huntenburg, Georg Langs, Gleb Bezgin, Simon B. Eickhoff, F. Xavier Castellanos, Michael Petrides, Elizabeth Jefferies, and Jonathan Smallwood (2016). "Situating the default-mode network along a principal gradient of macroscale cortical organization". *Proceedings of the National Academy of Sciences* 113.44, pp. 12574–12579.

Markov, Nikola T., Mária Ercsey-Ravasz, David C. Van Essen, Kenneth Knoblauch, Zoltán Toroczkai, and Henry Kennedy (2013). "Cortical high-density counterstream architectures". *Science* 342.6158, p. 1238406.

Markov, Nikola T., Julien Vezoli, Pascal Chameau, Arnaud Falchier, René Quilodran, Cyril Huissoud, Camille Lamy, Pierre Misery, Pascale Giroud, Shimon Ullman, Pascal Barone, Colette Dehay, Kenneth Knoblauch, and Henry Kennedy (2014a). "Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex". *Journal of Comparative Neurology* 522.1, pp. 225–259.

Markov, Nikola T. et al. (2014b). "A weighted and directed interareal connectivity matrix for macaque cerebral cortex". *Cerebral Cortex* 24.1, pp. 17–36.

Markram, Henry, Joachim Lübke, Michael Frotscher, Arnd Roth, and Bert Sakmann (1997). "Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex." *The Journal of Physiology* 500.2, pp. 409–440.

Marlin, Joseph J. and Adam G. Carter (2014). "Gaba-a receptor inhibition of local calcium signaling in spines and dendrites". *Journal of Neuroscience* 34.48, pp. 15898–15911.

Masse, Nicolas Y., Guangyu R. Yang, H. Francis Song, Xiao-Jing Wang, and David J. Freedman (2019). "Circuit mechanisms for the maintenance and manipulation of information in working memory". *Nature Neuroscience* 22.7, pp. 1159–1167.

McIlwain, Henry and Herman Stanton Bachelard (1972). "Biochemistry and the central nervous system".

Mejias, Jorge F., John D. Murray, Henry Kennedy, and Xiao-Jing Wang (2016). "Feedforward and feedback frequency-dependent interactions in a large-scale laminar network of the primate cortex". *Science Advances* 2.11, e1601335.

Mejias, Jorge F. and Xiao-Jing Wang (2020). "Mechanisms of distributed working memory in a large-scale model of macaque neocortex". *bioRxiv*, p. 760231.

Mendonça, Philipe RF, Mariana Vargas-Caballero, Ferenc Erdélyi, Gábor Szabó, Ole Paulsen, and Hugh PC Robinson (2016). "Stochastic and deterministic dynamics of intrinsically irregular firing in cortical inhibitory interneurons". *eLife* 5. Ed. by Frances K Skinner, e16475.

Mendoza-Halliday, Diego, Santiago Torres, and Julio C. Martinez-Trujillo (2014). "Sharp emergence of feature-selective sustained activity along the dorsal visual pathway". *Nature Neuroscience* 17.9, pp. 1255–1262.

Merker, Björn (1983). "Silver staining of cell bodies by means of physical development". *Journal of Neuroscience Methods* 9.3, pp. 235–241.

Mongillo, Gianluigi, Omri Barak, and Misha Tsodyks (2008). "Synaptic Theory of Working Memory". *Science* 319.5869, pp. 1543–1546.

Mueller, Adrienne, Rebecca M. Krock, Steven Shepard, and Tirin Moore (2019). "Dopamine receptor expression among local and visual cortex-projecting frontal eye field neurons". *Cerebral Cortex*, bhz–78.

Mueller, Adrienne, Steven B. Shepard, and Tirin Moore (2018). "Differential Expression of Dopamine D5 Receptors across Neuronal Subtypes in Macaque Frontal Eye Field". *Frontiers in Neural Circuits* 12.

Muller, Arnaud, Victory Joseph, Paul A. Slesinger, and David Kleinfeld (2014). "Cell-based reporters reveal in vivo dynamics of dopamine and norepinephrine release in murine cortex". *Nature methods* 11.12, pp. 1245–1252.

Muñoz, William, Robin Tremblay, Daniel Levenstein, and Bernardo Rudy (2017). "Layer-specific modulation of neocortical dendritic inhibition during active wakefulness". *Science* 355.6328, pp. 954–959.

Nimchinsky, Esther A., Bernardo L. Sabatini, and Karel Svoboda (2002). "Structure and function of dendritic spines". *Annual Review of Physiology* 64.1, pp. 313–353.

Niu, Meiqi, Daniele Impieri, Lucija Rapan, Thomas Funck, Nicola Palomero-Gallagher, and Karl Zilles (2020). "Receptor-driven, multimodal mapping of cortical areas in the macaque monkey intraparietal sulcus". *eLife* 9. Ed. by Timothy E Behrens, e55979.

Niu, Meiqi, Lucija Rapan, Thomas Funck, Seán Froudist-Walsh, Ling Zhao, Karl Zilles, and Nicola Palomero-Gallagher (2021). "Organization of the macaque monkey inferior parietal lobule based on multimodal receptor architectonics". *NeuroImage* 231, p. 117843.

Noudoost, Behrad and Tirin Moore (2011). "Control of visual cortical signals by prefrontal dopamine". *Nature* 474.7351, pp. 372–375.

Palomero-Gallagher, Nicola and Karl Zilles (2018). "Chapter 24 - Cyto- and receptor architectonic mapping of the human brain". In: *Handbook of Clinical Neurology*. Ed. by Ingeborg Huitinga and Maree J. Webster. Vol. 150. Brain Banking. Elsevier, pp. 355–387.

Pereira, Jacinto José Fonseca (2014). "Computational modeling of prefrontal cortex circuits: from neurons to networks". PhD Thesis. Universidade NOVA de Lisboa (Portugal).

Petreanu, Leopoldo, Tianyi Mao, Scott M. Sternson, and Karel Svoboda (2009). "The subcellular organization of neocortical excitatory connections". *Nature* 457.7233, pp. 1142–1145.

Pfeffer, Carsten K., Mingshan Xue, Miao He, Z. Josh Huang, and Massimo Scanziani (2013). "Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons". *Nature Neuroscience* 16.8, pp. 1068–1076.

Pinto, Lucas, Kanaka Rajan, Brian DePasquale, Stephan Y. Thiberge, David W. Tank, and Carlos D. Brody (2019). "Task-Dependent Changes in the Large-Scale Dynamics and Necessity of Cortical Regions". *Neuron* 104.4, 810–824.e9.

Pouget, Pierre, Iwona Stepniewska, Erin A. Crowder, Melanie W. Leslie, Erik E. Emeric, Matthew J. Nelson, and Jeffrey D. Schall (2009). "Visual and motor connectivity and the distribution of calcium-binding proteins in macaque frontal eye field: implications for saccade target selection". *Frontiers in Neuroanatomy* 3.

Preuss, Todd M. and Patricia S. Goldman-Rakic (1991). "Myelo- and cytoarchitecture of the granular frontal cortex and surrounding regions in the strepsirhine primate Galago and the anthropoid primate Macaca". *Journal of Comparative Neurology* 310.4, pp. 429–474.

Rapan, Lucija, Sean Froudist-Walsh, Meiqi Niu, Ting Xu, Thomas Funck, Karl Zilles, and Nicola Palomero-Gallagher (2021). "Multimodal 3D atlas of the macaque monkey motor and premotor cortex". *NeuroImage*, p. 117574.

Richfield, Eric K., Anne B. Young, and John B. Penney (1989). "Comparative distributions of dopamine D-1 and D-2 receptors in the cerebral cortex of rats, cats, and monkeys". *Journal of Comparative Neurology* 286.4, pp. 409–426.

Riley, Mitchell R., Xue-Lian Qi, Xin Zhou, and Christos Constantinidis (2018). "Anterior-posterior gradient of plasticity in primate prefrontal cortex". *Nature Communications* 9.1, p. 3790.

Robbins, T.W. and A.F.T. Arnsten (2009). "The Neuropsychopharmacology of Fronto-Executive Function: Monoaminergic Modulation". *Annual Review of Neuroscience* 32.1, pp. 267–287.

Roffman, Joshua L., Alexandra S. Tanner, Hamdi Eryilmaz, Anais Rodriguez-Thompson, Noah J. Silverstein, New Fei Ho, Adam Z. Nitenson, Daniel B. Chonde, Douglas N. Greve, and Anissa Abi-Dargham (2016). "Dopamine D 1 signaling organizes network dynamics underlying working memory". *Science advances* 2.6, e1501672.

Romo, Ranulfo, Carlos D. Brody, Adrián Hernández, and Luis Lemus (1999). "Neuronal correlates of parametric working memory in the prefrontal cortex". *Nature* 399.6735, pp. 470–473.

Romo, Ranulfo and Román Rossi-Pool (2020). "Turning Touch into Perception". *Neuron* 105.1, pp. 16–33.

Romo, Ranulfo and Emilio Salinas (2003). "Flutter Discrimination: neural codes, perception, memory and decision making". *Nature Reviews Neuroscience* 4.3, pp. 203–218.

Rose, Nathan S., Joshua J. LaRocque, Adam C. Riggall, Olivia Gosseries, Michael J. Starrett, Emma E. Meyering, and Bradley R. Postle (2016). "Reactivation of latent working memories with transcranial magnetic stimulation". *Science* 354.6316, pp. 1136–1139.

Rossi, Andrew F., Narcisse P. Bichot, Robert Desimone, and Leslie G. Ungerleider (2007). "Top–Down Attentional Deficits in Macaques with Lesions of Lateral Prefrontal Cortex". *Journal of Neuroscience* 27.42, pp. 11306–11314.

Sanides, F. (1962). *Die Architektonik des Menschlichen Stirnhirns: Zugleich eine Darstellung der Prinzipien Seiner Gestaltung als Spiegel der Stammesgeschichtlichen Differenzierung der Grosshirnrinde*. Monographien aus dem Gesamtgebiete der Neurologie und Psychiatrie. Berlin Heidelberg: Springer-Verlag.

Schultz, Wolfram (1998). "Predictive Reward Signal of Dopamine Neurons". *Journal of Neurophysiology* 80.1, pp. 1–27.

Schultz, Wolfram, Paul Apicella, and Tomas Ljungberg (1993). "Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task". *Journal of Neuroscience* 13.3, pp. 900–913.

Schuman, Benjamin, Robert P. Machold, Yoshiko Hashikawa, János Fuzik, Gord J. Fishell, and Bernardo Rudy (2019). "Four unique interneuron populations reside in neocortical layer 1". *Journal of Neuroscience* 39.1, pp. 125–139.

Seamans, Jeremy K., Daniel Durstewitz, Brian R. Christie, Charles F. Stevens, and Terrence J. Sejnowski (2001). "Dopamine D1/D5 receptor modulation of excitatory synaptic inputs to layer V prefrontal cortex neurons". *Proceedings of the National Academy of Sciences* 98.1, pp. 301–306.

Self, Matthew W., Roxana N. Kooijmans, Hans Supèr, Victor A. Lamme, and Pieter R. Roelfsema (2012). "Different glutamate receptors convey feedforward and recurrent processing in macaque V1". *Proceedings of the National Academy of Sciences* 109.27, pp. 11031–11036.

Seltzer, Benjamin and Deepak N. Pandya (1978). "Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey". *Brain research* 149.1, pp. 1–24.

Shafiei, Golia, Ross D Markello, Reinder Vos de Wael, Boris C Bernhardt, Ben D Fulcher, and Bratislav Misic (2020). "Topographic gradients of intrinsic dynamics across neocortex". *eLife* 9. Ed. by Lucina Q Uddin, Chris I Baker, Lucina Q Uddin, and Maxwell Bertolero, e62116.

Shpiro, Asya, Rodica Curtu, John Rinzel, and Nava Rubin (2007). "Dynamical Characteristics Common to Neuronal Competition Models". *Journal of Neurophysiology* 97.1, pp. 462–473.

Silberberg, Gilad and Henry Markram (2007). "Disynaptic inhibition between neocortical pyramidal cells mediated by martinotti cells". *Neuron* 53.5, pp. 735–746.

Slifstein, Mark, Elsmarieke van de Giessen, Jared Van Snellenberg, Judy L. Thompson, Rajesh Narendran, Roberto Gil, Elizabeth Hackett, Ragy Girgis, Najate Ojeil, Holly Moore, Deepak D'Souza, Robert T. Malison, Yiyun Huang, Keunpoong Lim, Nabeel Nabulsi, Richard E. Carson, Jeffrey A. Lieberman, and Anissa Abi-Dargham (2015). "Deficits in prefrontal cortical and extrastriatal dopamine release in schizophrenia: a positron emission tomographic functional magnetic resonance imaging study". *JAMA Psychiatry* 72.4, pp. 316–324.

Soden, Marta E., Amanda S. Chung, Beatriz Cuevas, Jesse M. Resnick, Rajeshwar Awatramani, and Larry S. Zweifel (2020). "Anatomic resolution of neurotransmitter-specific projections to the VTA reveals diversity of GABAergic inputs". *Nature Neuroscience*, pp. 1–13.

Soltani, Alireza and Xiao-Jing Wang (2006). "A Biophysically Based Neural Model of Matching Law Behavior: Melioration by Stochastic Synapses". *Journal of Neuroscience* 26.14, pp. 3731–3744.

Strang, Gilbert (1993). *Introduction to Linear Algebra*. Vol. 3. Wellesley-Cambridge Press Wellesley, MA.

Sun, Yongan, Yang Yang, Veronica C. Galvin, Shengtao Yang, Amy F. Arnsten, and Min Wang (2017). "Nicotinic $\alpha 4\beta 2$ Cholinergic Receptor Influences on Dorsolateral Prefrontal Cortical Neuronal Firing during a Working Memory Task". *Journal of Neuroscience* 37.21, pp. 5366–5377.

Suzuki, Mototaka and Jacqueline Gottlieb (2013). "Distinct neural mechanisms of distractor suppression in the frontal and parietal lobe". *Nature Neuroscience* 16.1, pp. 98–104.

Takeda, Kazuyoshi and Shintaro Funahashi (2004). "Population Vector Analysis of Primate Prefrontal Activity during Spatial Working Memory". *Cerebral Cortex* 14.12, pp. 1328–1339.

Tanaka, Shoji (1999). "Architecture and dynamics of the primate prefrontal cortical circuit for spatial working memory". *Neural Networks* 12.7, pp. 1007–1020.

Theodoni, Panagiota, Theofanis I. Panagiotaropoulos, Vishal Kapoor, Nikos K. Logothetis, and Gustavo Deco (2011). "Cortical Microcircuit Dynamics Mediating Binocular Rivalry: The Role of Adaptation in Inhibition". *Frontiers in Human Neuroscience* 5.

Tremblay, Robin, Soohyun Lee, and Bernardo Rudy (2016). "Gabaergic interneurons in the neocortex: from cellular properties to circuits". *Neuron* 91.2, pp. 260–292.

Trübutschek, Darinka, Sébastien Marti, Andrés Ojeda, Jean-Rémi King, Yuanyuan Mi, Misha Tsodyks, and Stanislas Dehaene (2017). "A theory of working memory without consciousness or sustained activity". *eLife* 6. Ed. by Tatiana Pasternak, e23871.

Trübutschek, Darinka, Sébastien Marti, Henrik Ueberschär, and Stanislas Dehaene (2019). "Probing the limits of activity-silent non-conscious working memory". *Proceedings of the National Academy of Sciences* 116.28, pp. 14358–14367.

Tsubomoto, Makoto, Rika Kawabata, Xiaonan Zhu, Yoshio Minabe, Kehui Chen, David A Lewis, and Takanori Hashimoto (2019). "Expression of Transcripts Selective for GABA Neuron Subpopulations across the Cortical Visuospatial Working Memory Network in the Healthy State and Schizophrenia". *Cerebral Cortex* 29.8, pp. 3540–3550.

Van Vugt, B., Bruno Dagnino, Devavrat Vartak, Houman Safaai, Stefano Panzeri, Stanislas Dehaene, and Pieter R. Roelfsema (2018). "The threshold for conscious report: Signal loss and response bias in visual and frontal cortex". *Science* 360.6388, pp. 537–542.

Vergara, José, Natsuko Rivera, Román Rossi-Pool, and Ranulfo Romo (2016). "A Neural Parametric Code for Storing Information of More than One Sensory Modality in Working Memory". *Neuron* 89.1, pp. 54–62.

Vezoli, Julien, Loïc Magrou, Rainer Goebel, Xiao-Jing Wang, Kenneth Knoblauch, Martin Vinck, and Henry Kennedy (2021). "Cortical hierarchy, dual counterstream architecture and the importance of top-down generative networks". *NeuroImage* 225, p. 117479.

Vijayraghavan, Susheel, Min Wang, Shari G. Birnbaum, Graham V. Williams, and Amy F. T. Arnsten (2007). "Inverted-U dopamine D1 receptor actions on prefrontal neurons engaged in working memory". *Nature Neuroscience* 10.3, pp. 376–384.

Walker, Florian, Martin Möck, Michael Feyerabend, Julien Guy, Robin J. Wagener, Dirk Schubert, Jochen F. Staiger, and Mirko Witte (2016). "Parvalbumin- and vasoactive intestinal polypeptide-expressing neocortical interneurons impose differential inhibition on Martinotti cells". *Nature Communications* 7, p. 13664.

Wall, Nicholas R., Mauricio De La Parra, Jordan M. Sorokin, Hiroki Taniguchi, Z. Josh Huang, and Edward M. Callaway (2016). "Brain-wide maps of synaptic input to cortical interneurons". *Journal of Neuroscience* 36.14, pp. 4000–4009.

Wang, Huai-Xing and Wen-Jun Gao (2009). "Cell Type-Specific Development of NMDA Receptors in the Interneurons of Rat Prefrontal Cortex". *Neuropsychopharmacology* 34.8, pp. 2028–2040.

Wang, Jane X., Zeb Kurth-Nelson, Dharshan Kumaran, Dhruva Tirumala, Hubert Soyer, Joel Z. Leibo, Demis Hassabis, and Matthew Botvinick (2018). "Prefrontal cortex as a meta-reinforcement learning system". *Nature Neuroscience* 21.6, pp. 860–868.

Wang, Min, Dibyadeep Datta, John Enwright, Veronica Galvin, Sheng-Tao Yang, Constantinos Paspalas, Rouba Kozak, David L. Gray, David A. Lewis, and Amy F. T. Arnsten (2019). "A novel dopamine D1 receptor agonist excites delay-dependent working memory-related neuronal firing in primate dorsolateral prefrontal cortex". *Neuropharmacology* 150, pp. 46–58.

Wang, X.-J., J. Tegnér, C. Constantinidis, and P. S. Goldman-Rakic (2004a). "Division of labor among distinct subtypes of inhibitory neurons in a cortical microcircuit of working memory". *Proceedings of the National Academy of Sciences of the United States of America* 101.5, pp. 1368–1373.

Wang, Xiao-Jing (1999). "Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory". *Journal of Neuroscience* 19.21, pp. 9587–9603.

– (2001). "Synaptic reverberation underlying mnemonic persistent activity". *Trends in Neurosciences* 24.8, pp. 455–463.

– (2002). "Probabilistic Decision Making by Slow Reverberation in Cortical Circuits". *Neuron* 36.5, pp. 955–968.

– (2020). "Macroscopic gradients of synaptic excitation and inhibition in the neocortex". *Nature Reviews Neuroscience* 21.3, pp. 169–178.

Wang, Yun, Maria Toledo-Rodriguez, Anirudh Gupta, Caizhi Wu, Gilad Silberberg, Junyi Luo, and Henry Markram (2004b). "Anatomical, physiological and molecular properties of Martinotti cells in the somatosensory cortex of the juvenile rat". *The Journal of Physiology* 561.1, pp. 65–90.

Watabe-Uchida, Mitsuko, Lisa Zhu, Sachie K. Ogawa, Archana Vamanrao, and Naoshige Uchida (2012). "Whole-Brain Mapping of Direct Inputs to Midbrain Dopamine Neurons". *Neuron* 74.5, pp. 858–873.

Watanabe, Kei and Shintaro Funahashi (2014). "Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex". *Nature Neuroscience* 17.4, pp. 601–611.

Watanabe, Yumiko and Shintaro Funahashi (2012). "Thalamic mediodorsal nucleus and working memory". *Neuroscience & Biobehavioral Reviews* 36.1, pp. 134–142.

Wei, Wei and Xiao-Jing Wang (2016). "Inhibitory Control in the Cortico-Basal Ganglia-Thalamocortical Loop: Complex Regulation and Interplay with Memory and Decision Processes". *Neuron* 92.5, pp. 1093–1105.

Wolff, Michael J., Janina Jochim, Elkan G. Akyürek, and Mark G. Stokes (2017). "Dynamic hidden states underlying working-memory-guided behavior". *Nature Neuroscience* 20.6, pp. 864–871.

Wong, Kong-Fatt and Xiao-Jing Wang (2006). "A recurrent network mechanism of time integration in perceptual decisions". *Journal of Neuroscience* 26.4, pp. 1314–1328.

Yang, Charles R. and Long Chen (2005). "Targeting Prefrontal Cortical Dopamine D1 and N-Methyl-D-Aspartate Receptor Interactions in Schizophrenia Treatment". *The Neuroscientist* 11.5, pp. 452–470.

Yang, Guangyu Robert, John D. Murray, and Xiao-Jing Wang (2016). "A dendritic disinhibitory circuit mechanism for pathway-specific gating". *Nature Communications* 7.1, p. 12815.

Yang, Yang, Constantinos D. Paspalas, Lu E. Jin, Marina R. Picciotto, Amy F. T. Arnsten, and Min Wang (2013). "Nicotinic α7 receptors enhance NMDA cognitive circuits in dorsolateral prefrontal cortex". *Proceedings of the National Academy of Sciences* 110.29, pp. 12078–12083.

Yoo, Sang-Ah, Julio C. Martinez-Trujillo, Stefan Treue, John K. Tsotsos, and Mazyar Fallah (2021). "Feature-based attention induces surround suppression during the perception of visual motion". *bioRxiv*, p. 2021.02.17.431646.

Zandvakili, Amin and Adam Kohn (2015). "Coordinated neuronal activity enhances corticocortical communication". *Neuron* 87.4, pp. 827–839.

Zatka-Haas, Peter, Nicholas A. Steinmetz, Matteo Carandini, and Kenneth D. Harris (2021). "Sensory coding and causal impact of mouse cortex in a visual decision". *bioRxiv*, p. 501627.

Zhang, Xiaoxing, Wenjun Yan, Wenliang Wang, Hongmei Fan, Ruiqing Hou, Yulei Chen, Zhaoqin Chen, Chaofan Ge, Shumin Duan, Albert Compte, and Chengyu T Li (2019). "Active information maintenance in working memory by a sensory cortex". *eLife* 8. Ed. by Catherine Dulac, Upinder Singh Bhalla, and Kartik Sreenivasan, e43191.

Zilles, Karl, Axel Schleicher, Nicola Palomero-Gallagher, and Katrin Amunts (2002). "Quantitative analysis of cyto-and receptor architecture of the human brain". In: *Brain Mapping: The Methods (Second Edition)*. Elsevier, pp. 573–602.

Supplementary Tables

Supplementary Table 1. Experimental evidence for delay activity (from Leavitt et al., 2017)

Cortical area	Positive findings	Negative findings
V1	2	3
V2	1	0
V4	2	3
1	0	3
3	0	3
MT	0	5
V6	0	0
DP	0	0
TEO	0	0
8m	18	0
F4	1	0
5	1	0
2	1	2
8l	15	0
STPc	1	0
7A	6	1
10	0	1
F3	2	0
TEpd	2	2
46d	24	0
9/46v	31	2
PBr	0	0
9/46d	32	2
F5	3	0
7m	0	0
25	0	0
LIP	7	0
32	0	0
STPi	1	0
9	1	0
45A	6	0
8B	2	0
7b	0	0
F2	3	0
F7	6	0
ProM	0	0
STPr	1	0
24c	6	0
OPRO	0	0

Supplementary Table 1. Experimental evidence for delay activity (from Leavitt et al., 2017), Related to STAR Methods. Areas considered 'persistent activity areas' in Figure 3 are shown in a red background. Areas considered 'non-persistent activity areas' are shown in a blue background. Areas with insufficient studies are shown with a white background. Note that the overlap between the model and the experimental data is high regardless of the minimum number of studies chosen, and the threshold to decide which are 'persistent activity areas' - see Supplementary Table 2.

Supplementary Table 2. Overlap of simulated activity with experimental data (%)					
Thresholds	1 study	2 studies	3 studies	4 studies	5 studies
50%	95	95	93	100	100
60%	95	95	93	100	100
70%	95	95	93	100	100
80%	95	95	93	100	100
90%	90	89	86	92	90

Supplementary Table 2. Overlap of simulated activity with experimental data (%), Related to STAR Methods. The overlap between the simulated delay activity pattern (from the model based on real anatomy) and the experimental pattern was high, regardless of the criteria used. Two thresholds were used. First, a threshold based on the number of studies was used to determine which brain areas to compare between the experimental data and simulations. This 'number of studies' threshold is shown in the columns. Once the areas for comparison have been chosen, we must decide which areas have significant evidence for persistent activity. This is based on the 'percent threshold', shown here in the rows. For a percent threshold of 50%, any included areas for which more than 50% of studies have found persistent activity is counted as a 'persistent activity area'. The experimental persistent activity areas are then compared to the areas showing persistent activity in the model simulation. In Figure 3 in the main text, we use a number of studies threshold of 2 studies and 50% percent threshold.

Supplementary Figures

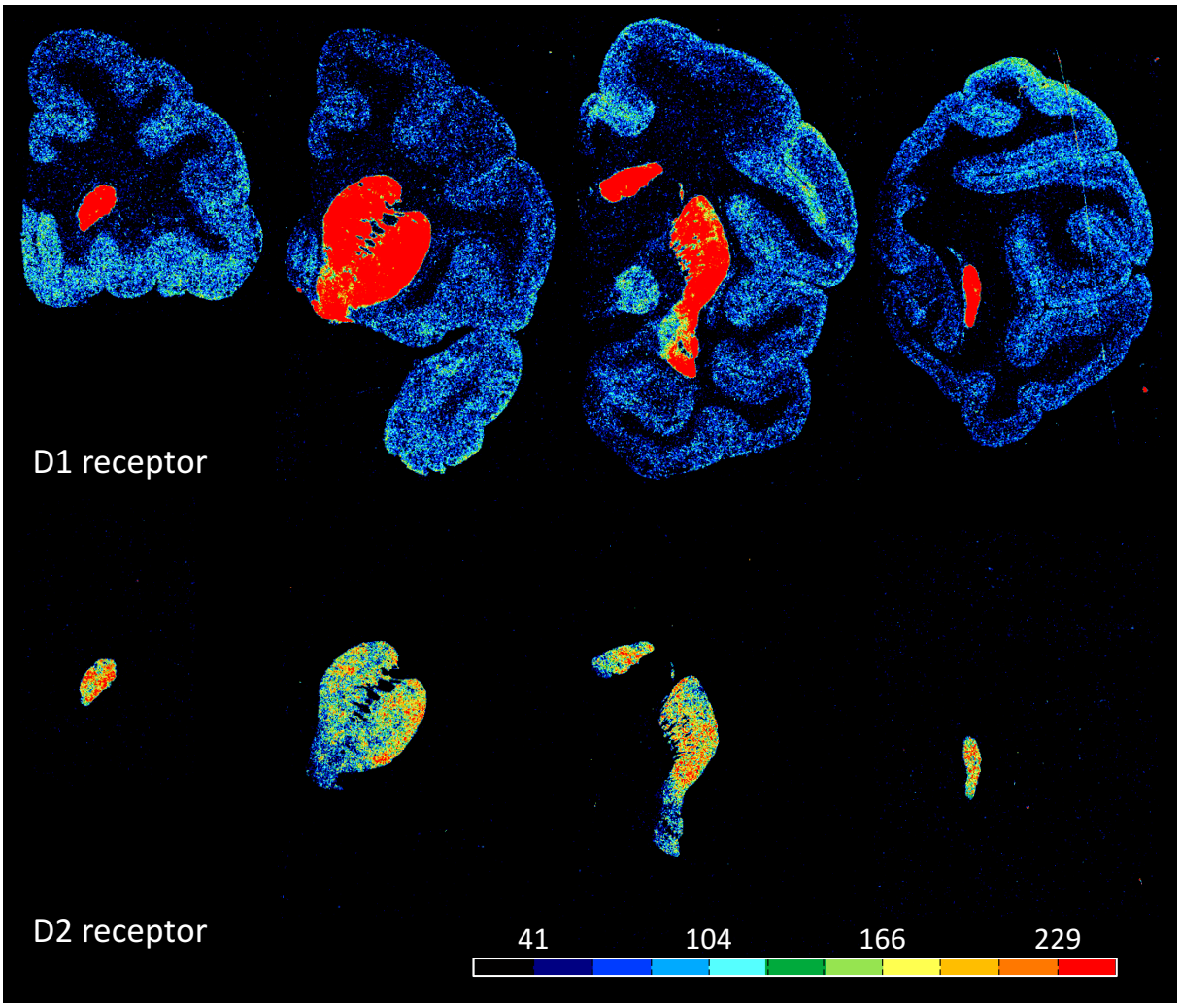

Figure S1: Example coronal sections through the macaque brain and processed for visualization of dopamine D1 and D2 receptors by means of quantitative in-vitro receptor autoradiography. Related to Figure 1. Note, that D2 receptor density in cortex is so low, that it is not detectable by means of the here applied method. Scale bar codes for receptor densities in fmol/mg protein.

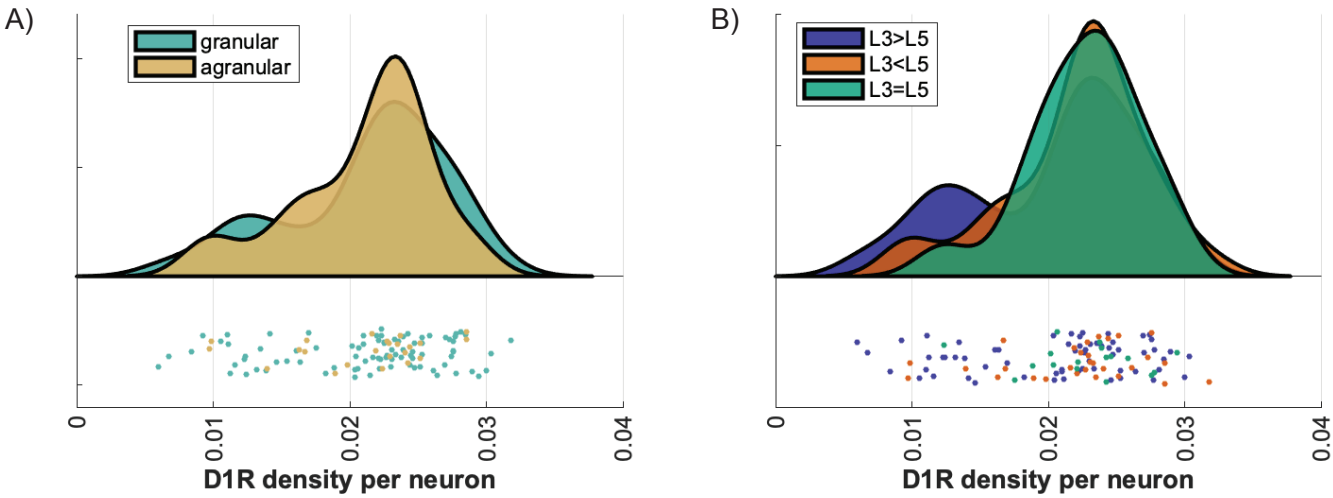

Figure S2: Anatomical distribution of D1 receptors. Related to Figure 1. A) The D1 receptor density per neuron did not significantly differ between granular (green) and agranular (yellow) areas. B) The D1 receptor density per neuron did not significantly differ between areas with larger pyramids in layer III (blue), those with larger pyramids in layer V (orange) or those with roughly equal sized pyramids in layers III and V (green)

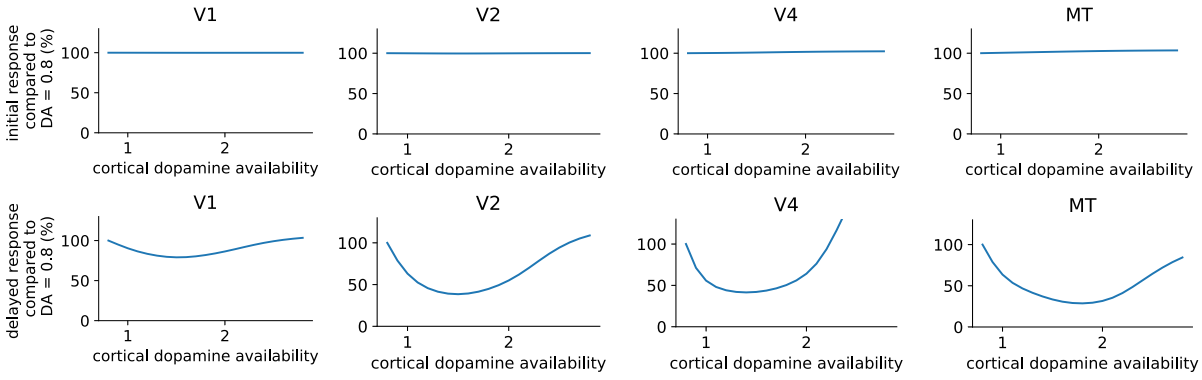

Figure S3: DA affects delayed, but not initial responses in early sensory areas. Related to Figure 2. Top row: The height of the initial peak response to the stimulus is essentially unaffected by dopamine levels in the cortex. Bottom row: The height of the delayed sensory response is affected by cortical dopamine levels, likely via feedback connections from higher cortical areas.

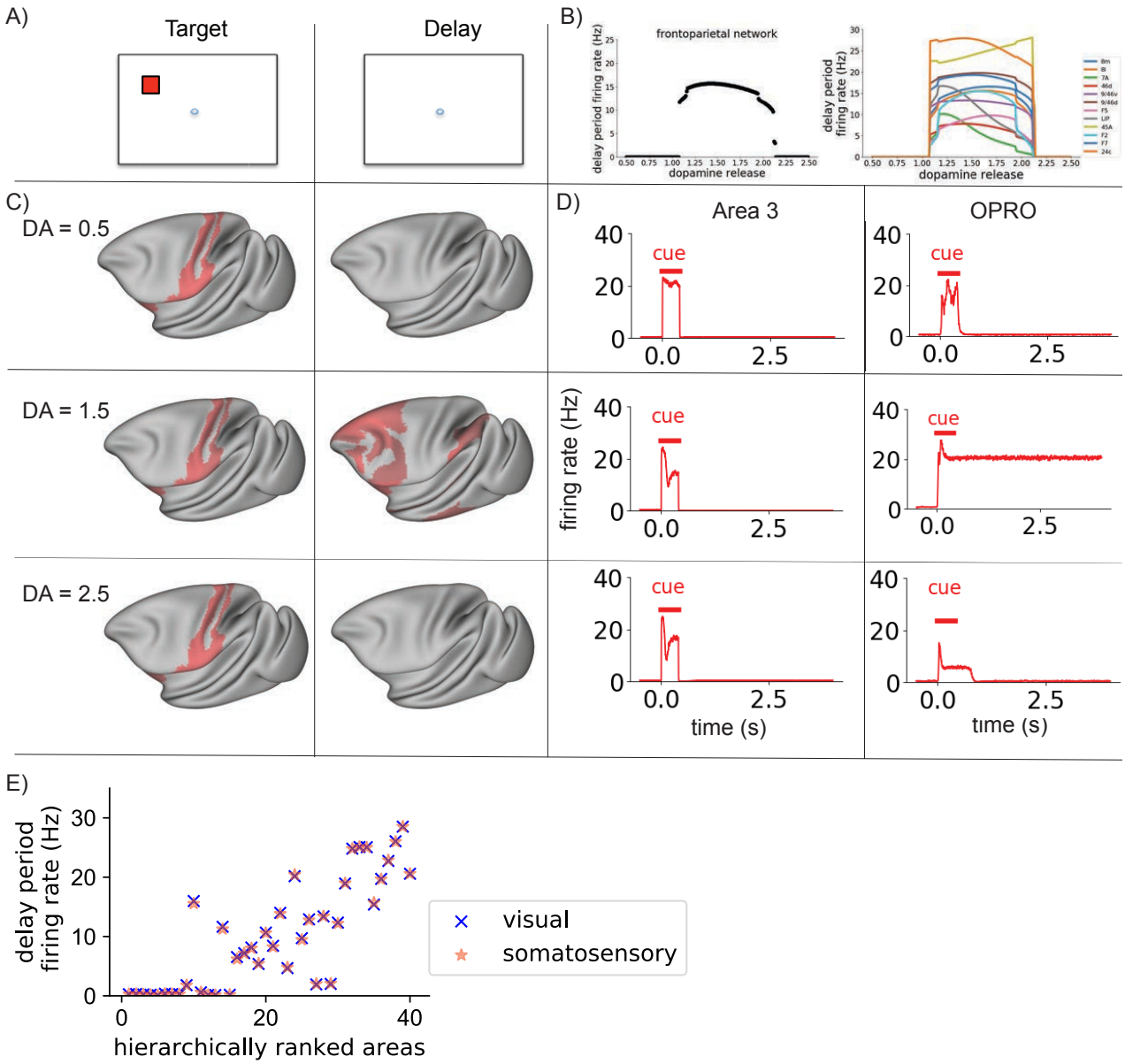

Figure S4: Dopamine release enables distributed somatosensory working memory. Related to Figure 2. A) Structure of the task. The cortical network was presented with a stimulus, which it had to maintain through a delay period. The tactile stimulus is presented to primary somatosensory cortex (area 3). B, left) Mean firing rate in the frontoparietal network at the end of the delay period, for different levels of dopamine release. There is an inverted-U relationship between dopamine release and delay period activity across the frontoparietal network, as for visual working memory. B, right) Mean delay-period activity of cortical areas as a function of dopamine release. All areas shown display persistent activity in experiments (Leavitt et al. 2017). C) Activity is shown across the cortex at different stages in the working memory task (left to right), with increasing levels of dopamine release (from top to bottom). Red represents activity in the excitatory population sensitive to the location of the target stimulus. Very low or very high levels of dopamine release resulted in reduced propagation of stimulus-related activity to frontal areas and a failure to engage persistent activity. Mid-level dopamine release enables distributed persistent activity. D) Timecourses of activity in selected cortical areas. The horizontal bars indicate the timing of cue (red) input to area 3. Activity in early somatosensory areas such as area 3 peaks in response to the stimulus, but quickly decays away after stimulus removal for all levels of dopamine release. In contrast, there is dopamine-dependent persistent activity in area OPRO. E) The pattern of activity at the end of the delay period is highly overlapping following visual and somatosensory working memory tasks. DA, cortical dopamine availability.

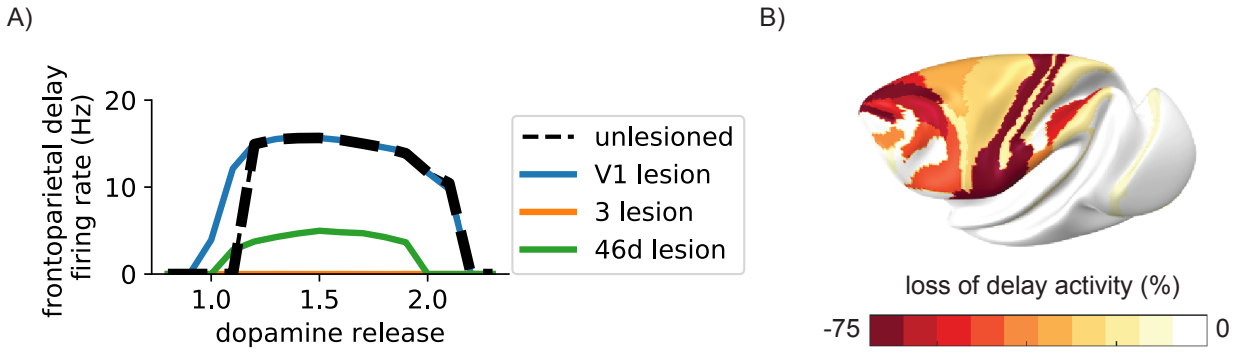

Figure S5: Lesions to visual areas do not disrupt somatosensory working memory. Related to Figure 3. A) Lesions to areas such as 46d and LIP led to reduced delay period firing across for all levels of dopamine release. Lesions to ares 3 and 2 of somatosensory cortex disrupted the ability to perform the somatosensory working memory task. In contrast, lesions to visual areas such as V1 did not significantly affect somatosensory working memory. B) Map showing the severity of lesions to cortical areas on somatosensory working memory. More severe effects are shown in deeper red.

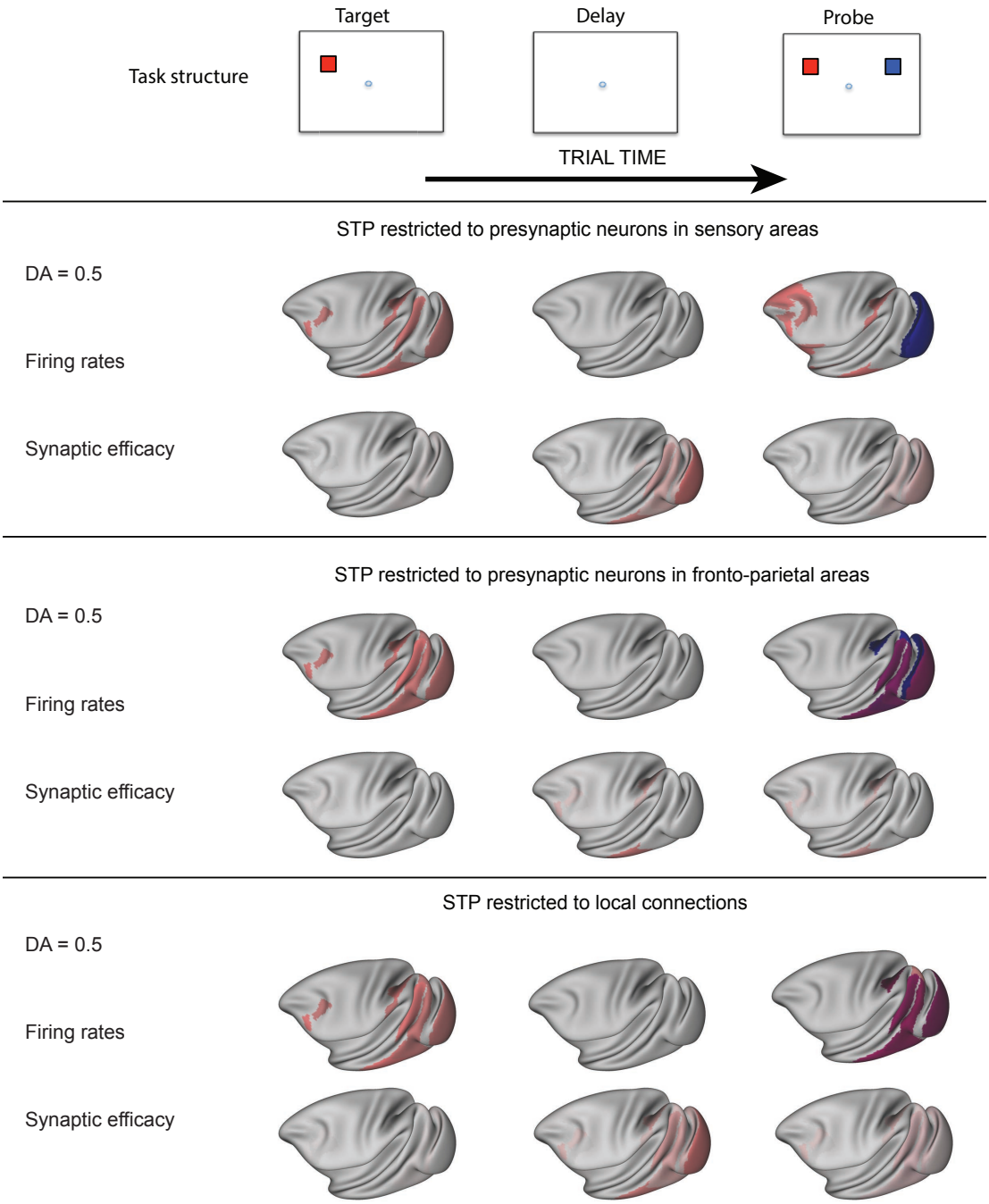

Figure S6: Activity-silent working memory without short-term plasticity in local prefrontal synapses. Related to Figure 4. Top row. The structure of the 'ping' short-term memory task. Second row. Reactivation of latent working memory representations was possible upon pinging the system, with short-term plasticity restricted to connections from neurons in sensory areas. Third row. Reactivation of latent working memory representations was not possible upon pinging the system, when short-term plasticity was restricted to connections from neurons in frontoparietal cortex. Bottom row. Reactivation of latent working memory representations was not possible upon pinging the system, when short-term plasticity was restricted to local connections between neurons in the same area.

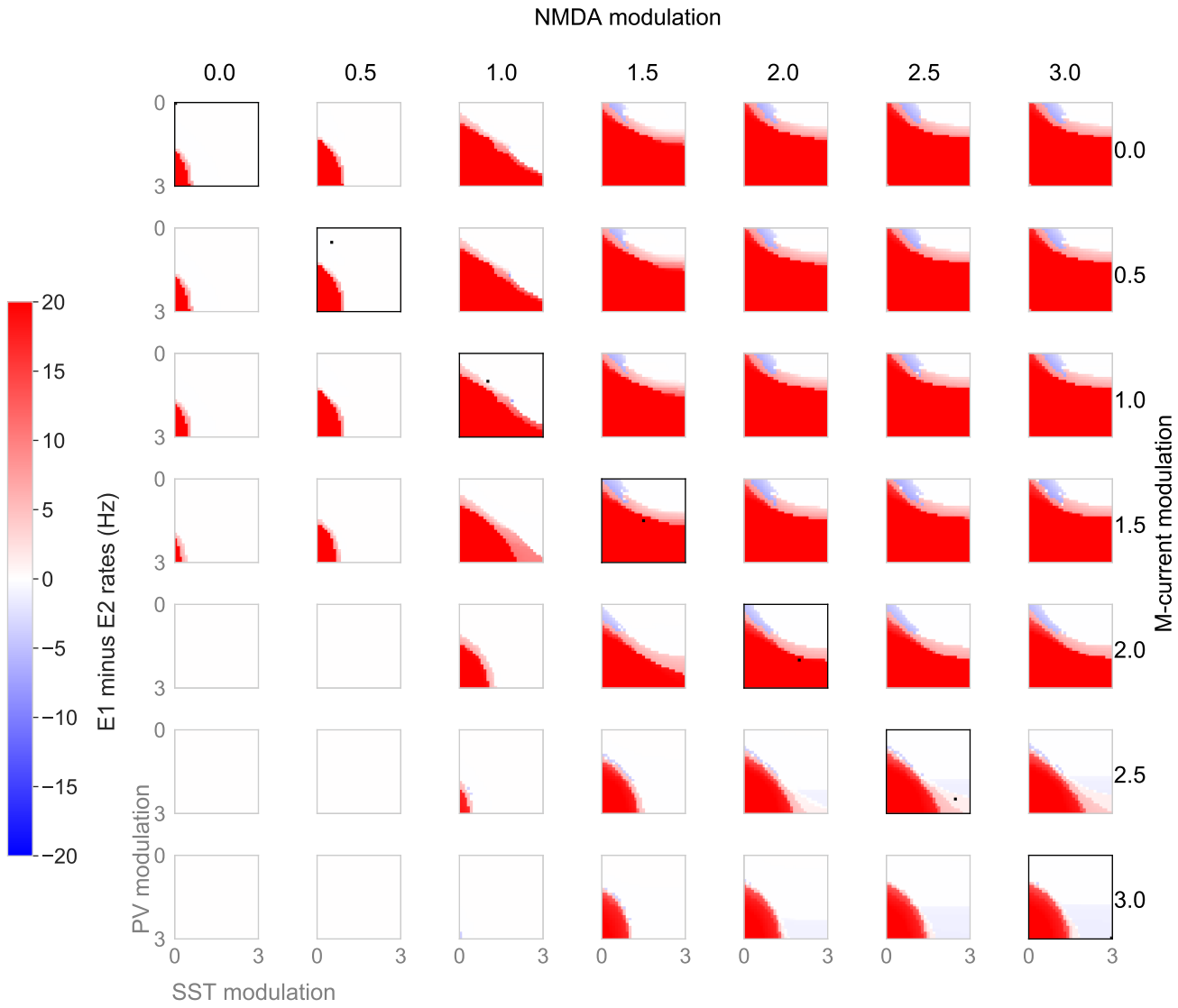

Figure S7: Distractor-resistance depends on the high dendritic inhibition. Related to Figure 5. We identified the model behaviour for different dopamine levels, across different levels of dendritic and somatic inhibition. Consistently across dopamine levels, higher somatic, and lower dendritic inhibition was associated with distractible working memory (blue). In contrast, lower somatic, and higher somatic inhibition was associated with distractor-resistant working memory (red). High dendritic and high somatic inhibition results in no persistant activity (white). The levels of dendritic and somatic inhibition associated with the standard dopamine modulation used in the rest of the paper marked by a black square. Note that high PV modulation by dopamine results in lower PV inhibition of the soma.