Emergence of Non-Linear Mixed Selectivity in Prefrontal Cortex after Training

Wenhao Dang*, Russell J. Jaffe*, Xue-Lian Qi, Christos Constantinidis

Department of Neurobiology & Anatomy, Wake Forest School of Medicine,

Winston-Salem, NC 27157, USA

* These authors contributed equally to this work

Abbreviated Title: Mixed Selectivity in PFC

Address correspondence to:
Christos Constantinidis, PhD.
Department of Neurobiology and Anatomy,
Wake Forest School of Medicine
Medical Center Blvd, Winston Salem, NC, 27157
E-mail: cconstan@wakehealth.edu

Number of pages: 38
Number of figures: 8
Number of words for Abstract: 217
Number of words for Significance: 102
Number of words for Introduction: 503
Number of words for Discussion: 1279

The authors report no conflicts of interest during the research conducted.

ACKNOWLEDGMENTS

Research reported in this paper was supported by the National Eye Institute of the National Institutes of Health under award number R01 EY017077. We wish to acknowledge Kathini Palaninathan, Austin Lodish, for technical help; Junda Zhu, and Sihai Li for helpful comments on the manuscript.

ABSTRACT

Neurons in the prefrontal cortex are typically activated by multiple factors when performing a cognitive task, and by different tasks altogether. The selectivity of single neurons for the same stimulus dimension often changes depending on context or task performed, a phenomenon known as nonlinear mixed selectivity. It has been hypothesized that neurons with such mixed selectivity offer a computational advantage for performing cognitive tasks due to high-dimensional neural representations. In this study, we sought to determine how nonlinear mixed selectivity is affected by training to perform a cognitive task by examining the neural responses of monkeys before and after they were trained to perform visual working memory tasks. We also compared nonlinear mixed selectivity in different sub-regions of the prefrontal cortex that play different roles in these tasks. Our findings indicate that a small population of prefrontal neurons exhibit nonlinear mixed selectivity even prior to any training to perform cognitive tasks. Learning to perform working memory tasks induces a modest increase in the proportion of neurons with both linear and non-linear mixed selectivity. However, we saw little evidence that nonlinear mixed selectivity is predictive of task performance. Our results provide insights on the representation of stimulus and task information in neuronal populations.

SIGNIFICANCE STATEMENT

Working memory depends on the ability of neurons to represent stimuli in their pattern of discharges when they are no longer present. How neurons represent simultaneously different types of information remains a complex computational problem. It has been hypothesized that nonlinear mixed selectivity emerges as a result of training to perform tasks that require maintenance of stimuli and task parameters in memory. We tested experimentally this hypothesis by examining neuronal responses at different areas of the prefrontal cortex, before and after training to perform cognitive tasks. We reveal the regions of the prefrontal cortex that are most responsible for different types of selectivity, as well as how these types of selectivity vary as a result of training, the context of information represented in working memory tasks, and their modulating factors. These insights are critical to formulating a practical understanding of working memory, and by extension, of memory-related disorders dependent on neural selectivity.

INTRODUCTION

Working memory (WM) is commonly defined as the ability to encode, maintain, and manipulate information in the conscious mind over a period of seconds without the presence of any sensory inputs. Although this is a core component of complex cognitive abilities such as planning and reasoning, the true importance of WM ultimately depends on whether it can maintain task relevant information and manipulate information in task relevant manner (Baddeley, 2012). To do achieve the necessary adaptability in WM, the brain needs to be able to encode multiple variables, including both external sensory inputs and internal task requirements. The mechanisms that underlie and organize this encoding across time and population is one of the most important questions in current WM research.

When humans or animals are required to maintain objects in their WM, neurons in a network of brain areas exhibit selective and sustained increases or decreases in their activity in order to represent the remembered objects through these unique patterns of activity (Constantinidis and Procyk, 2004). The prefrontal cortex (PFC) plays a leading role in this network, and by extension, in the use of WM (Riley and Constantinidis, 2016). For example, when the PFC is damaged or degraded, whether through trauma, illness, or experimental lesions, performance in WM tasks seems to decrease dramatically (Curtis and D'Esposito, 2004; Morris and Baddeley, 1988; Rossi et al., 2007).

PFC neurons often encode more than one variables, and the exact variables encoded are task dependent (Asaad et al., 2000; Machens et al., 2010; Mansouri et al., 2006; Qi et al., 2015; Warden and Miller, 2010). More interestingly, a proportion of

neurons exhibit nonlinear mixed selectivity (NMS) for different variables, which means their response to the combination of variables cannot be predicted by the linear summation of their responses to single variables (Johnston et al., 2020; Parthasarathy et al., 2017; Rigotti et al., 2013). Theoretical studies have shown that NMS is useful for linear readouts of flexible, arbitrary combinations of variables (Buonomano and Maass, 2009; Fusi et al., 2016; Rigotti et al., 2010), and may also control the trade-off between discrimination and generalization (Barak et al., 2013; Johnston et al., 2020).

Despite the proposed importance of NMS on theoretical grounds, some experimental studies have failed to detect neurons with NMS (Cavanagh et al., 2018). It is therefore possible that NMS is a property of only a limited set of prefrontal subdivisions or that NMS emerges exclusively after training to perform specific types of cognitive tasks. Moreover, the implications of NMS on other aspects of encoding, such as code stability, have also not yet been investigated. We were therefore motivated in the current study, to analyze and compare neural data recorded when rhesus macaque monkeys were performing different visual working memory tasks. We report that NMS can be changed by training experience and task context and that NMS cells differ in coding dynamics comparing to linear cells. However, the average amount of linear decodable information is similar for single neurons in both populations.

METHODS

Animals: Six male rhesus monkeys (*Macaca mulatta*), age 5–9 years old, weighing 5–12 kg were used in these experiments. None of the animals had any prior experimentation experience at the onset of the experiments. Monkeys were either single-housed or pair-housed in communal rooms with sensory interactions with other monkeys. All experimental procedures followed guidelines by the U.S. Public Health Service Policy on Humane Care and Use of Laboratory Animals and the National Research Council's Guide for the Care and Use of Laboratory Animals and were reviewed and approved by the Wake Forest University Institutional Animal Care and Use Committee.

Experimental setup: Monkeys sat with their head fixed in a primate chair while viewing a monitor positioned 68 cm away from their eyes with dim ambient illumination. Animals were required to fixate on a 0.2° white square appearing in the center of the screen. During each trial, animals maintained fixation on the square while visual stimuli were presented either at a peripheral location or over the fovea in order to receive a liquid reward. Any break of fixation immediately terminated the trial and no reward was given. Eye position was monitored throughout the trial using a non-invasive, infrared eye position scanning system (model RK-716; ISCAN, Burlington, MA). The system achieved a $< 0.3°$ resolution around the center of vision. Eye position was sampled at 240 Hz, digitized and recorded. Visual stimuli display, monitoring of eye position, and the synchronization of stimuli with neurophysiological data were performed with in-house

131 software55 implemented on the MATLAB environment (Mathworks, Natick, MA), and
132 utilizing the psycho-physics toolbox.
133
134 Pre-training presentation: Following a brief period of fixation training and acclimation to
135 the stimuli, monkeys were required to fixate on a center position while stimuli were
136 displayed on the screen. The monkeys were rewarded for maintaining fixation during the
137 trial with a liquid reward (fruit juice). The stimuli shown were white 2° square stimulus
138 presented in one of nine possible locations arranged in a 3 × 3 grid of 10° distance
139 between adjacent stimuli. The same stimuli were shown following training in a working
140 memory tasking during the "post- training" phase. A fixation interval of 1 s where only
141 the fixation point was displayed was followed by 500 ms of stimulus presentation,
142 followed by a 1.5 s "delay" interval where, again, only the fixation point was displayed.
143 A second stimulus was subsequently shown, either identical in location to the initial
144 stimulus, or diametrically opposite the first stimulus. This second stimulus display was
145 followed by another "delay" period of 1.5 s. The location and identity of stimuli in these
146 experiments was of no behavioral relevance to the monkeys during the "pre- training"
147 phase. In a few sessions, a variable delay period was used. Neurons recorded in these
148 sessions appear in most analyses, though they are excluded from analyses that assume a
149 fixed delay period.
150
151 Working memory task: Four of the six monkeys were trained to complete spatial working
152 memory tasks. The task used in most experiments required the monkeys to remember the
153 spatial location of the first stimulus shown, observe a second stimulus, and report

whether the second stimulus was shown in the same location as the first stimulus or if it was in the diametrically opposite location via saccading to one of two target stimuli. For two of the monkeys, a match would mean a saccade to the green square stimulus while a nonmatch would mean a saccade to a blue square stimulus. Targets for the remaining monkey were an "H" and a diamond shape for match condition/nonmatch condition, respectively. Each target stimulus appeared at locations orthogonal to the cue/sample stimuli while the target feature locations were varied randomly from trial-to-trial. One of the four monkeys was trained in a different spatial task, a variant of the delayed response task. Its structure was identical to the first five epochs of the match/ nonmatch task except that the second stimulus always appeared in the same location as the first stimulus. After the end of the second delay period, the animal then had to saccade to the location where the stimulus was located. Neurons recorded from this animal are excluded from the match/nonmatch analysis.

Surgery and neurophysiology: A 20 mm diameter craniotomy was performed over the PFC and a recording cylinder was implanted over the site. The location of the cylinder was visualized with anatomical MRI imaging and stereotaxic coordinates post-surgery. For two of the four monkeys in the post-training phase (sub- jects MA and EL), the recording cylinder was moved after an initial round of recordings so that an additional surface of the prefrontal cortex could be sampled.

Anatomical localization. Each monkey underwent a magnetic resonance imaging scan prior to neurophysiological recordings. Electrode penetrations were mapped onto the

177 cortical surface. We identified 6 lateral prefrontal regions: a posterior- dorsal region
178 including area 8 A, a mid-dorsal region including area 8B and area 9/ 46, an anterior-
179 dorsal region including area 9 and area 46, a posterior-ventral region including area 45,
180 an anterior-ventral region including area 47/12, and a frontopolar region including area
181 10. However, the frontopolar region was not sampled sufficiently for this analysis.
182 In addition to comparisons between areas segmented in this fashion, other analyses were
183 performed taking into account the position of each neuron along the AP axis. This was
184 defined as the line connecting the genu of the arcuate sulcus to the frontal pole, for the
185 purposes of this analysis. The recording coordinates of each neuron were projected onto
186 this line and position was expressed as a proportion of the length of this line.
187
188 Neuronal recordings: Neural recordings were carried out in areas 8, 9, 9/46, 45, 46, and
189 47/12 of the PFC prior to training and following training in a spatial working memory
190 task. Subsets of the data presented here were previously used to determine the properties
191 of neurons in the dorsal and ventral prefrontal cortex pooled together, and properties of
192 neurons prior to training. Newly acquired data were added here, to determine differences
193 before and after training in posterior-dorsal, mid-dorsal, anterior-dorsal, posterior-ventral,
194 and anterior-ventral prefrontal subdivisions. Extracellular recordings were performed
195 with multiple microelectrodes. These were either glass- or epoxylite-coated tungsten
196 electrodes with a 250 μm diameter and 1–4 MΩ impedance at 1 kHz (Alpha-Omega
197 Engineering, Nazareth, Israel). Arrays of up to 8-microelectrodes spaced 0.2–1.5 mm
198 apart were advanced into the cortex with a Microdrive system (EPS drive, Alpha- Omega
199 Engineering) through the dura into the prefrontal cortex. The signal from each electrode

was amplified and band-pass filtered between 500 Hz and 8 kHz while being recorded with a modular data acquisition system (APM system, FHC, Bowdoin, ME). Waveforms that exceeded a user-defined threshold were sampled at 25 μs resolution, digitized, and stored for off-line analysis. Neurons were sampled in an unbiased fashion, collecting data from all units isolated from our electrodes, with no regard to response properties of a neuron being isolated. Recorded spike waveforms were sorted into separate units using an automated cluster analysis relying on the KlustaKwik algorithm, which applied principal component analysis of the waveforms. To ensure the stability of firing rate in the recordings analyzed, we identified recordings in which a significant effect of trial sequence was evident on the baseline firing rate (ANOVA, $p < 0.05$), e.g., due to a neuron dis- appearing or appearing during a run, as we were collecting data from multiple electrodes. Data from these sessions were truncated so that analysis was only performed on a range of trials with stable firing rate. Less than 10% of neuronal records were corrected in this way.

Identical data collection procedures, recording equipment, and spike sorting algorithms were used before and after training, to ensure that any changes reported between stages were not due to these factors. To also ensure that changes in neuronal firing properties were not the result of systematic differences in the inherent properties of neurons sampled, we compared the Signal-to-Noise Ratio (SNR) of neuronal recordings before and after training10. For each unit, we defined SNR as the ratio of the peak-to-trough height of its mean action potential waveform, divided by the standard deviation of the noise. The latter was computed from the baseline of each waveform, derived from the first 10 data points (corresponding to 0.25 ms) of each sample. SNR provides an overall

measure of unit isolation quality10. We used this measure to identify neurons with excellent isolation, defined based on SNR > 5.

Data analysis: Data analysis was implemented with the MATLAB computational environment (Mathworks, Natick, MA), with statistic tests implemented through Originlab (OriginLab Corporation, Northampton, MA) and StatsDirect (StatsDirect Ltd. England). PSTHs were calculated by moving window average method with a Gaussian window with 200 ms standard deviation, shaded area indicating two times standard error cross trials. For all tasks, only cells with at least 12 correct trials for each cue-sample location/shape pairs were included in the analysis. To classify neurons of the spatial task into different categories of selectivity, we performed two-way ANOVAs on the spike count between either the sample location x matching status in for the trial, or between sample location x task epoch (first or second stimuli presentation). Neurons with classic selectivity (CS) exhibited a main effect of only one factor without significant interaction term. Neurons with linear mixed selectivity (LMS) exhibited main effects of both factors without significant interaction term. Neurons with nonlinear mixed selectivity (NMS) exhibited a significant interactions term. Finally, neurons with no selectivity (NS) exhibited no significant term for both the main effects and the interaction term. Similarly, the two factors for feature task ANOVA analysis were sample shape x matching status for the trial.

PFC areas with more than 50 cells in both pre- and post-training time points were included in the subdivision mixed selectivity analysis. Thus, for the feature task, only data from mid-dorsal, posterior-dorsal and posterior-ventral PFC were analyzed, while

the spatial task analyzed data from the mid-dorsal, posterior-dorsal, posterior-ventral, anterior-dorsal and anterior-ventral PFC.

For comparing mixed selectivity in feature and conjunction tasks, conjunction trials in which cue and sample stimuli both showed in the same location as in the feature task were picked as the dataset for conjunction task, then in the feature task same number of trials using the same shapes were picked as corresponding feature task dataset. Moreover, comparing mixed selectivity in success and error trials, we only utilized neurons that had at least 4 usable error trials across at least two cue-sample locations pairs, in order to avoid anomalous data. The same number of trials from the corresponding cue-sample locations pairs were then randomly chosen in the success trials as the dataset for the correct behavioral response condition. This randomized success trial selection process was repeated for 50 times, in order to allow us to compensate for our relatively limited selection of available trials.

For decoding analysis, spiking responses from 1 second before cue onset to 5 seconds after cue onset were first binned using a 400 wide window and 100 ms steps to create a spike count vector with a length of 57 elements. A pseudo-population was then constructed using the spike count vectors from all the available neurons of all the available animals, thus resulting in a dataset with 96 trials, as if they were recorded simultaneously. A linear SVM decoding algorithm was implemented using fitcecoc function in MATLAB to decode stimuli location, shape, or matching status of trials. 10-fold cross validation was used to estimate the decoder performance, 10 random samplings were implemented to calculate 95% confidence interval. For location and feature task, the decoding baseline for sensor information was 12.5%, since there are 8 different choices.

For conjunction dataset decoding analysis, neurons were treated as if the same shape and location pairs were used for every neuron. For the conjunction task, the decoding chance level is 50% for location and feature.

In the pre- vs post-training decoding analysis (Fig. 6), linear (CS and LMS) and nonlinear (NMS) neurons are first defined by their pre and post training responses in the sample or sample delay period. Each classified population was then applied to decode sensory information (location and shape) and matching status. All informative neurons (CS, LMS, and NMS neurons) in both the sample and sample delay periods were used for the cross temporal decoding analysis (Fig. 7). The SVM decoder was trained on the conjunction dataset and got 57 linear decision boundaries for each time points. The same dataset was then classified by every decision boundary in the vector to produce a 57x57 matrix—a process that was repeated for 10 times in order to plot the mean. For passive-active cross task decoding analysis (Fig. 4C), only neurons that showed nonlinear mixed selectivity during sample period of both tasks were used. The SVM decoder was trained on the passive dataset, then tested on the active spatial dataset—a process that was repeated 10 times to produce a 95 percent interval.

Data availability: All relevant data and code will be available from the corresponding author on reasonable request. Matlab decoder code for figure 6 and 7 are also available at https://github.com/dwhzlh87/mixed-selectivity.git

RESULTS

Extracellular neurophysiological recordings were collected from the lateral prefrontal cortex (LPFC) of six monkeys before and after they were trained to perform a match/nonmatch task (Meyer et al., 2011; Riley et al., 2018). The task required them to view two stimuli appearing in sequence, with delay periods intervening between them and to make a judgement on whether the second stimulus was identical to the first or not (Fig. 1). The two stimuli could differ in terms of their location (spatial task), shape (feature task), or both (conjunction task). If the second stimulus matched the first, the monkey had to choose a green choice target, at a subsequent interval, or a blue target otherwise. A total of 1617 cells from six monkeys and 1495 cells from five monkeys were recorded while the animals were performing passive spatial and passive feature tasks respectively pre-training (Fig. 1A) and 1104 cells from three and 1116 cells from two of the same six animals were collected from post-training time points performing spatial and feature match-to-sample task respectively (Fig. 1B-C). Additionally, 975 cells from two of the same six animals were collected while the monkey performing the conjunction task post-training (Fig. 1D). Besides, we also collected neural data from 247 neurons for the passive spatial task (Fig. 1A) from two monkeys after they were trained for the active spatial task.

Types of selectivity in individual neuronal responses

In our tasks, the exact same stimulus has a different context depending on the task interval during which it is presented and the sequence of stimuli in the trial. We first

considered how selectivity for stimulus location in the spatial working memory task (Fig. 1B) varies when the same stimulus appears as a match (it is preceded by a cue at the same location) or a nonmatch (it is preceded by a cue at a different location). We thus examined firing rate during the second stimulus (sample) presentation as a function of the location the stimulus appeared (eight locations arranged on a 3x3 grid with 10 degrees distance between stimuli, excluding the center location) and on whether this stimulus was a match or a nonmatch. We used a 2-way ANOVA with factors stimulus location and match/nonmatch status to classify neurons into four categories: classical selective (CS) were the neurons that exhibited a significant main effect of stimulus location, but not match/nonmatch status i.e. they were selective for the location of stimuli and did not respond differently when the stimulus appeared as a match or nonmatch. Linear mixed selective (LMS) were the neurons with a significant main effect of location and a significant main effect of match/nonmatch status but no significant interaction, i.e. neurons whose selectivity for stimuli remained the same for the match and nonmatch conditions, but the overall level of response was higher when a stimulus appeared either as a match or nonmatch. Non-linear mixed selective (NMS) neurons were the neurons with significant main effect of location and a significant interaction i.e. neurons whose spatial selectivity differed for the match and nonmatch conditions. Finally, non-selective (NS) were the neurons with no location selectivity. We repeated the identical analysis based on firing rates in the delay period that followed the second stimulus presentation (sample delay).

A second type of NMS was identified in terms of selectivity for a stimulus when it appeared as the first stimulus in the sequence (cue) or the second (sample). For this

analysis, we only examined sample stimuli that matched the cue. In this case too, we set up a 2-way ANOVA model, and identified CS, LMS, NMS, and NS neurons now in terms of how the represented the exact same stimulus when it appeared as a cue and as a sample (match) stimulus.

Effects of training on NMS

Training in the spatial working memory task increased the proportion of NMS cells in both the sample and sample-delay period (sample period: pre-training proportion=6.2%, post-training proportion=12.3%, two-sample proportion test, $z=5.31$, $p=1.13\times10^{-7}$; sample delay period: pre-training proportion= 2.8%, post-training proportion=6.2%, two-sample proportion test, $z=4.62$, $p=4.86\times10^{-5}$). However this increase was not exclusive to NMS cells. The proportion of CS cells also increased in the sample-delay period (sample period: pre-training proportion=17.1%, post-training proportion=15.0%, two-sample proportion test, $z=1.47$, $p=0.142$; sample-delay period: pre-training proportion= 10.58%, post-training proportion=14.8%, two-sample proportion test, $z=3.19$, $p=0.0014$). Nor was the increase in NMS cells evident for all types of training. When we looked at the proportion change for feature task between post- and pre-training, we only found an increase of proportion for CS cells (sample period: pre-training proportion= 12.0%, post-training proportion=15.7%, two-sample proportion test, $z=2.65$, $p=0.0081$; sample-delay period: pre-training proportion= 9.0%, post-training proportion=22.6%, two-sample proportion test, $z=9.37$, $p=0$). No significant increase in the proportion of NMS cells was observed (sample period: pre-training proportion=5.8%, post-training proportion=6.7%, two-sample proportion test, $z=1.01$, $p=0.314$; sample-delay period: pre-training

proportion= 4.2%, post-training proportion=4.6%, two-sample proportion test, z=0.522, p=0.602) (Fig. 2B,C).

Generally the same results held when we used task epoch (cue vs. match) as the second independent variable for the two-way ANOVA (Fig. 2D,E). For the spatial task, training increased proportion of both NMS (sample period: pre-training proportion= 4.1%, post-training proportion=7.3%, two-sample proportion test, $z=3.36$, $p=7.89\times10^{-4}$; sample-delay period: pre-training proportion= 3.0%, post-training proportion=9.5%, two-sample proportion test, $z=6.69$, $p=2.30\times10^{-11}$) and CS (sample period: pre-training proportion= 27.4%, post-training proportion=33.2%, two-sample proportion test, z=3.20, p=0.0014; sample-delay period: pre-training proportion= 30.0%, post-training proportion=37.9%, two-sample proportion test, $z=4.25$, $p=2.15\times10^{-5}$), but for the feature task, only the proportion of CS cells changed (sample period: pre-training proportion= 22.5%, post-training proportion=34.1%, two-sample proportion test, z=6.49, p=0.0014; sample-delay period: pre-training proportion= 25%, post-training proportion=36.4%, two-sample proportion test, $z=4.25$, $p=8.39\times10^{-11}$. NMS cells sample period: pre-training proportion= 9.8%, post-training proportion=9.4%, two-sample proportion test, z=0.364, p=0.716; NMS cells sample-delay period: pre-training proportion= 7.4%, post-training proportion=7.6%, two-sample proportion test, z=0.184, p=0.854).

Regional localization of NMS

It is possible that neurons with NMS are localized in some sub-regions of the prefrontal cortex. To figure out what portion of the LPFC contributed to the observed mixed selectivity changes, we subdivided the lateral PFC intro regions (Fig. 3A), and analyzed

neurophysiological data from five of these regions. Only subregions with more than 50 cells in both pre- and post-training conditions were included in the comparison. We looked at the mixed selectivity defined by location/shape and matching status in the sample period. It was found that the mid-dorsal subdivision underwent the most change in the proportion of NMS cells for the spatial task after training (Fig. 3B). For the feature task, the most apparent change in NMS happened in the posterior-dorsal region, while a small increase in classic and linear mix selectivity could be found in all three subdivisions tested (Fig. 3C).

NMS in task context

The comparison of the naïve and trained conditions allowed us to test the overall incidence of NMS in different populations of prefrontal neurons, sampled randomly before and after training, which lasted over several months. If nonlinear mixed selectivity is critical for the representation of task-relevant information one might expect that it may also dynamically change in the same neurons, when animals are performing the task vs. they are passively viewing stimuli. Our dataset included a condition where this comparison was possible: passive presentation of stimuli to monkeys after they had been trained to perform the tasks. We thus performed a two-way ANOVA on collected neurons from which we recorded responses in both passive and active spatial tasks.

Although we found there was an increase in the proportion of cells that coded matching status during the sample period, as well as more cells coding sensory information in the cue-delay when the animal need to report the matching decision, the increase in the proportion of NMS cells is not significant (passive proportion= 9.3%,

active proportion=11.7%, exact matched pair proportion test, F=1.385, p=0.362) (Fig. 4 A). Interestingly, a large proportion of cells changed their selectivity category across tasks, especially for CS cells. Also it was found that the degree of nonlinear mixed selectivity in NMS cells does not seem to be predictive of whether the cell will fall in the same selectivity category in both tasks (Fig. 4B).

Other than the stability on the cell ensemble level, we also wondered if the encoding code is stable in cells that are informative in both tasks. To do that, we trained SVM classifier in the passive task and tested in the active spatial task, if the code is stable cross tasks, the decoding performance will be close. The cross-task decoding analysis using the same cells showed that the encoding of the sensory information was more stable across tasks, comparing to that of matching information (Fig. 4C). This indicates at least, some cells dynamically change their coding variable, or start to code multiple variables when the task at hand require to do so, while most cells keep the same code for the same requirement cross tasks.

We were able to address the effect of task context on NMS observed in the same cells, in a second data set. We compared responses to identical stimuli when they appeared in the context of the conjunction task, which required maintenance in memory of both the spatial location and feature of stimuli, and when they appeared in the context of the feature task, when only the feature needed to be remembered (all stimuli always appeared at the same location). Responses from the same cells could then be compared in two conditions, and the same number of trials and same stimuli location pairs were chosen in two conditions to make a fair comparison. We did not find any significant differences for neither CS cells (feature task proportion= 11.9%, conjunction task

426 proportion=9.9%, exact matched pair proportion test ,F=1.229, p=0.197) nor NMS cells
427 (feature task proportion= 4.1%, conjunction task proportion=4.4%, exact matched pair
428 proportion test ,F=1.031, p=0.901) in the sample period. Similarly for the sample-delay
429 period, we still did not found significant proportion change for neither CS cells (feature
430 task proportion= 10.8%, conjunction task proportion=11.6%, exact matched pair
431 proportion test ,F=1.069, p=0.681) nor NMS cells (feature task proportion= 4.6%,
432 conjunction task proportion=5.9%, exact matched pair proportion test ,F=1.278, p=0.266)
433 (Fig. 5A). Like the case of passive-active comparison, we found an unstable mapping
434 between tasks in the selectivity categories, for both CS and NMS cells (Fig. 5B), but this
435 may just reflect the fact that most of the informative cells we found are just by chance
436 (p=0.05 was used as threshold for detecting significant terms). In contrast to passive-
437 active spatial dataset, in the current dataset the degree of nonlinear mixed selectivity
438 presented for NMS cells in one task could predict whether the cell would exhibit NMS in
439 the other task (Fig. 5C).
440

441 **Information encoding by NMS neurons**

442 To quantify the amount of task variable information contained in linear (CS and LMS)
443 and nonlinear mixed (NMS) cells, we used a linear Support Vector Machine (SVM)
444 decoder to decode sensory information (location and shape) and match/or nonmatch
445 status information. We performed this comparison on firing rates recorded in the sample
446 and sample-delay period. For each comparison, we picked randomly equal numbers of
447 linear and nonlinear cells in the respective task epoch. Pre- and post-training time points
448 were also plotted side by side to be compared (Fig. 6). As expected, decoding

449 performance for matching information increased above chance level only after the sample
450 had been presented, in the sample and sample-delay period. Some non-chance
451 performance could be detected before training, but performance rose well above chance
452 after training, for both CS and NMS cells. Compared to the shape information, the
453 location information was already well-represented before training.

454 In general, linear and non-linear cells contained near equal amount of linear
455 decodable information about external sensory information as well as task-related
456 variables, though CS appeared slightly more important for shape feature encoding in the
457 sample period, while NMS appeared more important in the spatial task in the sample-
458 delay period. Cross temporal decoding approach was utilized to probe the encoding
459 dynamics for CS and NMS cells in the conjunction task (Fig. 7), the analysis showed that
460 NMS cells are more dynamic compare to CS cells in the sample-delay period for
461 location, indicated by the worse performance off the diagonal in the cross-temporal
462 decoding performance matrix. We also found that even the location matching alone was
463 not enough to solve the conjunction task, location matching information was still
464 maintained in the sample and sample-delay period. That was not true for feature matching
465 information, despite the fact that a subset of neurons could reach near-perfect decoding
466 performance in feature matching task (Fig. 6).

467

468 **NMS in correct and error trials**

469 The presence of decodable information in the prefrontal cortex does not necessarily imply
470 the presence of information in the conscious mind. We thus compared mixed selectivity
471 in correct and error trials. Limited to a small number of error trials, we only analyzed the

trials in which the sample stimuli was a match to the cue. We used location and task epoch as two independent variables for this ANOVA analysis. The number of trials and the number of cue locations included was also matched for each cell. We found that the proportion of cells selective to stimuli locations are higher in error trials (sample period: correct trials proportion= 8.8%, error trials proportion=17.0%, two-sample t test, $t(98)=45.74$, $p=6.38\times10^{-68}$; sample-delay period: correct trials proportion= 9.6%, error trials proportion=15.4%, two-sample t test, $t(98)=0.364$, $p=2.87\times10^{-53}$), indicating that individual neurons are more broadly responding to different locations. To our surprise, we did not find a change in nonlinear mixed selectivity (sample period: correct trials proportion= 4.7%, error trials proportion=4.9%, two-sample t test, $t(98)=1.13$, $p=0.262$; sample-delay period: correct trials proportion= 2.8%, error trials proportion=2.3%, two-sample t test, $t(98)=3.82$, $p=2.33\times10^{-4}$), but that may be due to lack of trials to detect interaction. In both correct and error trials the proportion of interaction term is very low, despite the fact that in the full spatial matching dataset the proportion of interaction term is above chance (Fig. 2).

DISCUSSION

Selectivity for different types of information is critical in representing the multitude of information that can be maintained in WM. However, recent research suggests that the role of neural selectivity may extend far beyond merely serving as a medium for representation in WM, and the increased dimensionality in NMS has been highlighted as a potential means of increasing the efficiency of WM task performance (Johnston et al., 2020; Rigotti et al., 2013; Shaoyu et al., 2016). Specifically, Rigotti et al. correlated dimensional collapse with failed task performance, noting that the increased dimensionality of NMS could lead to greater efficiency in information storage, as well as greater flexibility in adapting to execute new tasks (Rigotti et al., 2013). Moreover, all task relevant information could be decoded from NMS neurons alone, despite their relative scarcity, with decoder accuracy actually increasing as the task became more complex (Rigotti et al., 2013). This implies that NMS may play a role in the successful performance of complex tasks.

A causal relationship between success and dimensionality—and by extension, NMS—was not supported by our results, as we did not observe any significant changes in NMS between error and success trials. Nevertheless, training still resulted in a slight increase in NMS, just as we have previously observed in CS, with NMS increasing in spatial tasks, just as CS increased in feature tasks. These insights seem to imply that NMS plays a key role in spatial WM, and further investigation is therefore encouraged to further elaborate upon the role of this phenomenon.

511 **Effects of training on neural responses**

512 Working memory is considerably plastic and at least some aspects of it, such as mental
513 processing speed and the ability to multitask, can be improved with training (Bherer et
514 al., 2008; Dux et al., 2009; Jaeggi et al., 2008; Klingberg et al., 2005; Klingberg et al.,
515 2002). Working memory training has been proven particularly beneficial for clinical
516 populations, e.g. in traumatic brain injury, attention deficit hyperactivity disorder
517 (ADHD), and schizophrenia (Klingberg et al., 2002; Subramaniam et al., 2012;
518 Westerberg et al., 2007). However, the verdict of whether working memory training
519 confers tangible benefits on normal adults and whether these benefits transfer to
520 untrained domains, remains a matter of heated debate. (Constantinidis and Klingberg,
521 2016; Cortese et al., 2015; Fukuda et al., 2010; Owen et al., 2010; Peijnenborgh et al.,
522 2015; Schwaighofer et al., 2015).

523 This malleability of cognitive performance is thought to be mediated by the
524 underlying plasticity in neural responses, most importantly within the prefrontal cortex
525 (Constantinidis and Klingberg, 2016). In a series of prior studies, we have investigated
526 changes in prefrontal responsiveness and selectivity (Meyer et al., 2011; Meyers et al.,
527 2012; Qi et al., 2011; Riley et al., 2018), as well as other aspects of neuronal discharges
528 such as trial-to-trial variability and correlation between neurons (Qi and Constantinidis,
529 2012a, b). In the present analysis, guided by experimental and theoretical predictions
530 (Rigotti et al., 2013), we examined another potential source of enhanced ability to
531 represent working memory information after training, Nonlinear Mixed Selectivity.

532 In agreement with our hypothesis, we found that training increased the proportion
533 of neurons that exhibit NMS. However, we encountered neurons with NMS even in

534 animals that were naïve to any cognitive training. It is now well established that the
535 human and primate prefrontal cortex represent stimuli in memory if when not prompted
536 to do so (Foster et al., 2017) or even prior to training in working memory tasks (Meyer et
537 al., 2007). Our finding of NMS neurons in naïve monkeys provides another exemplar of
538 that principle, and future experiments should consider investigation NMS as a possible
539 mechanism through which this principle is implemented. Additionally, we determined
540 that the training-induced increase in neuronal selectivity was generalized across
541 categories of neurons, including neurons with Classical Selectivity which were much
542 more abundant in the trained than the naïve prefrontal cortex. Finally, NMS increased
543 only for some types of task information and not for others. The increase in feature
544 selectivity after training was driven almost exclusively by CS cells. This results suggest
545 that NMS may play a role in the simultaneous representation of multiple types of
546 information in memory, however this is not universal across tasks, in agreement with
547 some prior studies, which have failed to uncover substantial NMS in the tasks they
548 employed (Cavanagh et al., 2018).
549

550 **Task Complexity and Difficulty**

551 A possible factor that determines the emergence of NMS is task complexity. In principle,
552 NMS may be only necessary in highly complex tasks, that require subjects to maintain in
553 memory and combine multiple types of information, particularly if the primary role of
554 NMS is to simplify the involved neural circuits, thus achieving greater efficiency, as
555 originally suggested (Rigotti et al., 2013). Our dataset relied on three tasks which differed
556 in complexity (and overall difficulty) and thus allowed us to test the idea. The spatial and

feature tasks each required maintenance of a single stimulus property in memory (location or shape). The conjunction task required both.

Surprising however, we did not observe a higher incidence of NMS in the conjunction task when compared to the feature task. Moreover, we observed a much lower incidence of NMS in the feature than the spatial task although the latter was no more complex, and the monkeys achieved higher overall performance. This implies that NMS may ultimately be a spatially centered phenomenon.

Regional Specialization

Different types of information are represented across the dorso-ventral and anterior-posterior axes of the prefrontal cortex (Constantinidis and Qi, 2018), and it was therefore important to examine the regional distribution of NMS neurons within the prefrontal cortex. We saw that NMS cells were most strongly demonstrated in the mid dorsal area for the spatial task and the posterior dorsal area for the feature task. This pattern was generally consistent with the known distribution of neuronal selectivity for stimuli in the prefrontal cortex (Riley et al., 2018).

Information Content and Task Performance

The most critical hypotheses of the Mixed Selectivity theory are that NMS represent information more efficiently and that it is critical for performance. We relied on a linear SVM decoder to determine specifically what information could be represented by NMS cells, compared to CS cells. Similar quantity of information could be decoded from

(equal-sized) populations of CS and NMS neurons. We only saw a minor preference for spatial information in NMS cells, and for feature information in CS cells.

Similarly, when we compared the NMS levels of successful and failed task trials, we were surprised to find that there was no significant difference. This means that although equal information may be stored in NMS neurons, the information is not necessarily accessible to the conscious mind. An important caveat for this conclusion is that we based our analysis on the feature task for which sufficient number of neurons and error trials per neuron were available (the spatial task was too easy for the animals, and not enough neurons were available in the conjunction task). Since very little NMS was present in the feature task in correct trials, a floor effect may have prevented a further decline from becoming apparent. Nonetheless, our result reinforces the idea that NMS is not necessary in all tasks, without which performance fails. An interesting observation in this analysis was that spatial location was elevated in error trials. The result may imply that task success also depends on the task relevance of the represented information. Ultimately, by comparing and evaluating the conditions in which significant quantities of neurons exhibit NMS, we may decipher the true role of NMS in working memory and beyond.

FIGURES

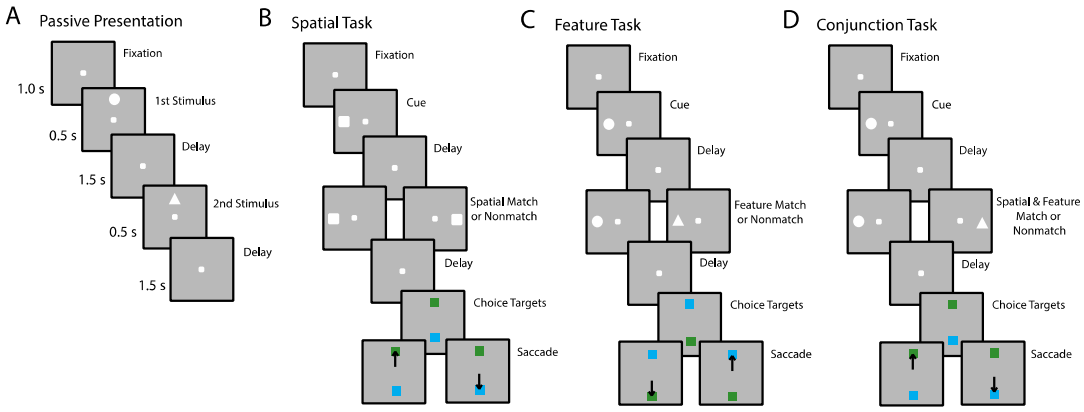
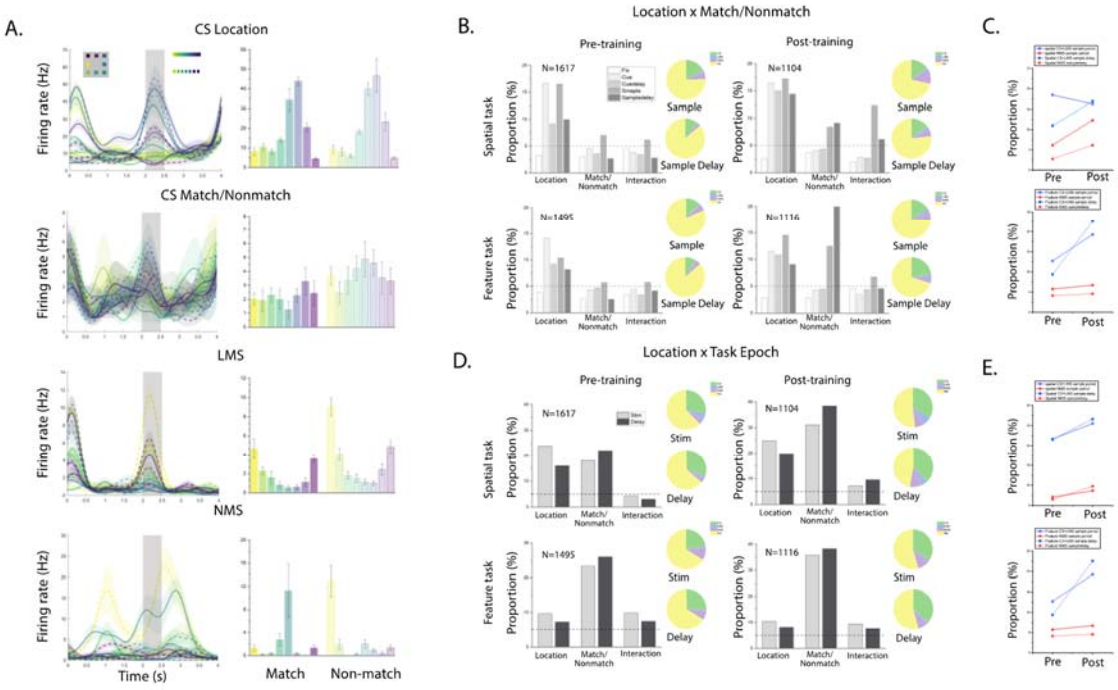


Figure.1 Task structure for (A) Passive presentation; (B) Spatial location match-to-sample task; (C) Shape feature match-to-sample task; (C) Location-Shape conjunction task.

603

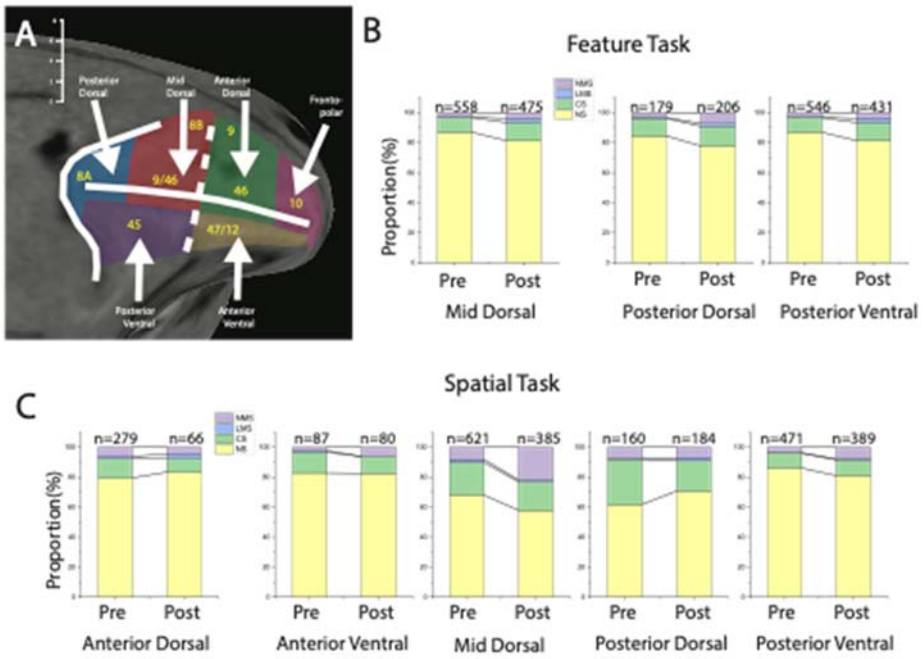


604

605 Figure.2 Training increased mixed selectivity preferentially in the spatial task. (A)
606 Example neurons with classic selectivity (CS), linear mixed selectivity (LMS), and
607 nonlinear mixed selectivity (NMS). (B) and (C) Training increased non-linear mixed
608 selectivity, especially for the spatial task, revealed by two-way ANOVA of stimuli
609 location and match/nonmatch. (D) and (E) Training increased non-linear mixed
610 selectivity, especially for the spatial task, revealed by two-way ANOVA of stimuli
611 location and task epoch.

612

613



614

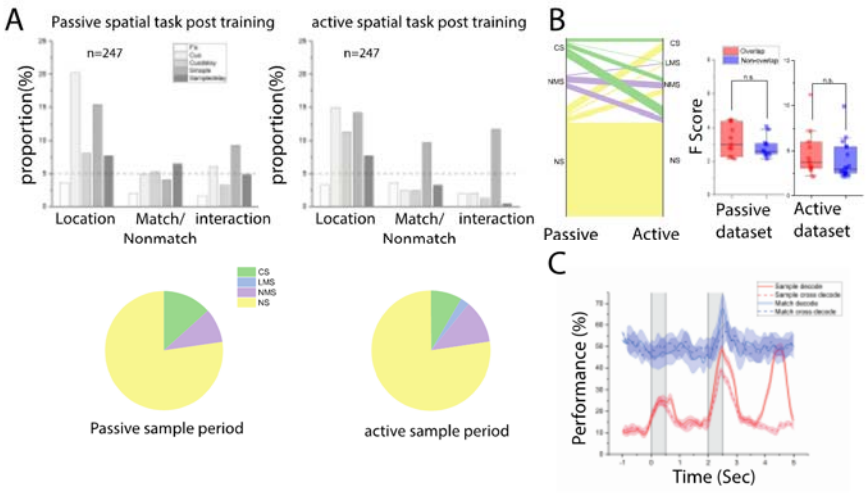
615 Figure.3 Cell selectivity changes by brain regions. (A) recording location subdivision in

616 LPFC. (B) Proportion changes for different selectivity categories in the feature task. (C)

617 Proportion changes for different categories in the spatial task.

618

619

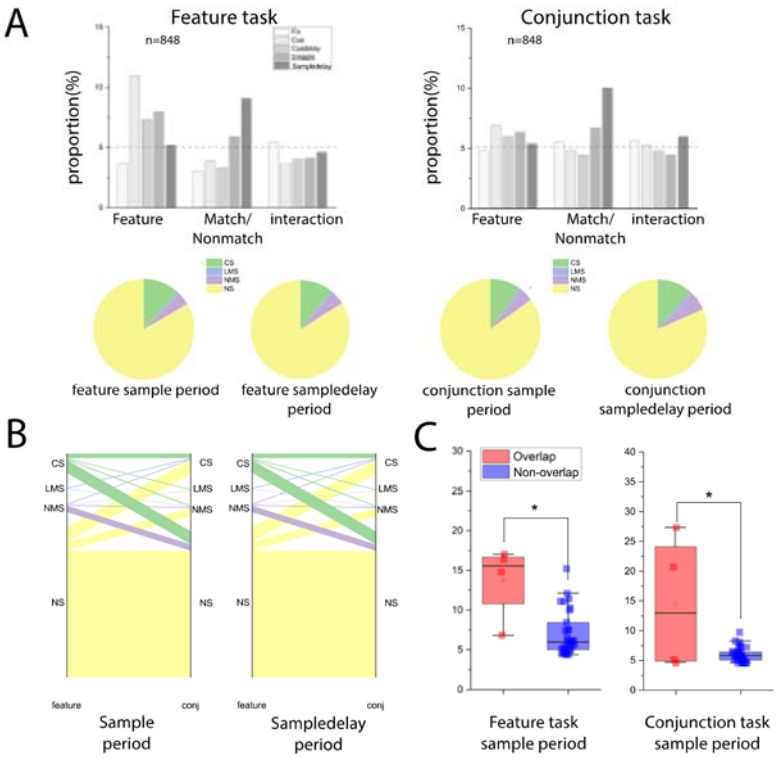


620

621 Figure. 4 Neuron selectivity modulated by task requirements. (A) Comparison of the
622 proportion of cells with different selectivity in passive vs. active spatial task after
623 training. The same population was included for the comparison in two conditions. (B)
624 Informative cells in two tasks are largely nonoverlapping. The degree of mixed
625 selectivity does not differ in overlapping and nonoverlapping populations of NMS cells.
626 (C) Within and cross-task decoding for sample location and match/nonmatching using
627 overlapping NMS population.

628

629



630

631 Figure. 5 Neuron selectivity modulated by task context. (A) The proportion of different

632 selectivity cells in feature and conjunction tasks in the same population of cells, after

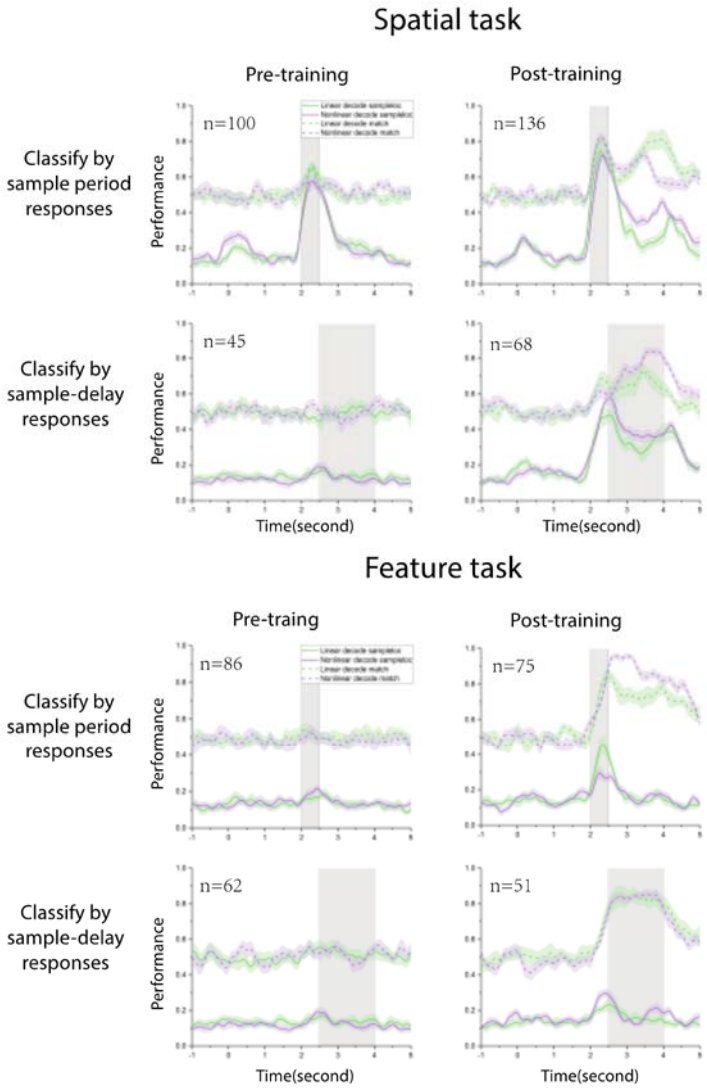
633 controlling for trial number and stimuli pairs used. (B) cell selectivity category mapping

634 in two tasks. (C) overlapping and nonoverlapping populations differ in degree of

635 nonlinear mixed selectivity.

636

637



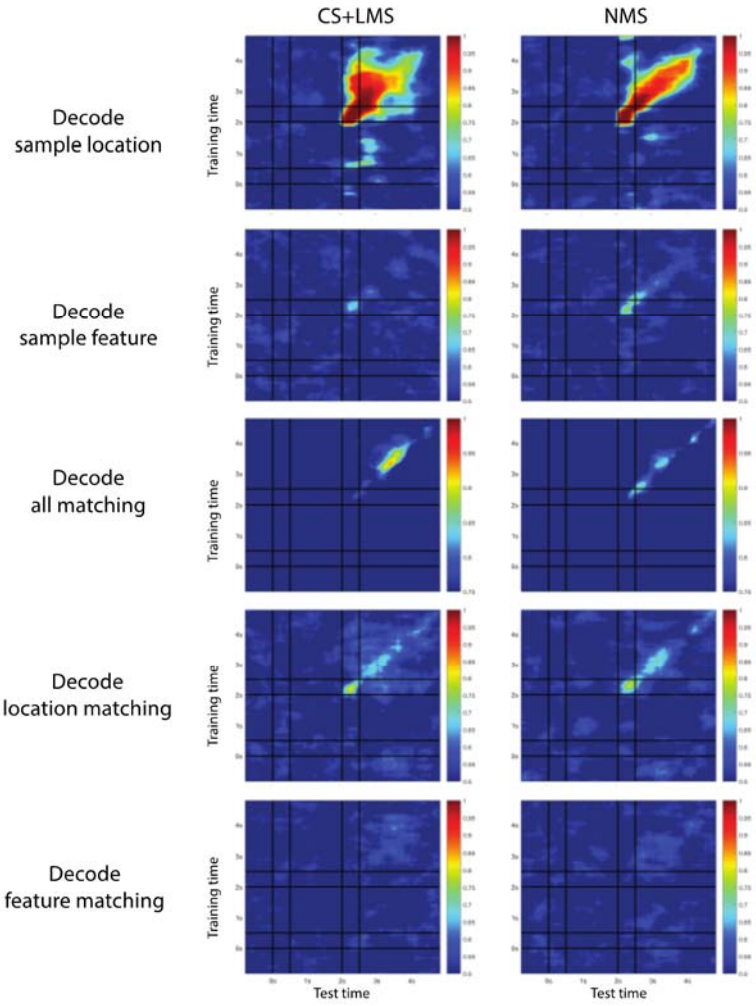
638

639 Figure.6 Decoding for stimuli and matching status in the linear and nonlinear selective

640 cell, before and after training.

641

642



643

644 Figure.7 Cross temporal decoding for different task variables in conjunction task using

645 linear and nonlinear mixed cells reveals different coding dynamics between linear and

646 nonlinear populations.

647

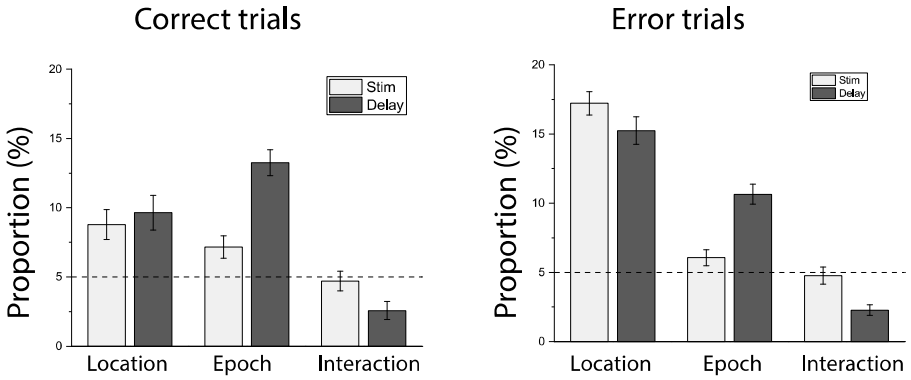


Figure. 8 Comparison of cell selectivity in correct and error trials in the same population for spatial match to sample task, after controlling for trial number and location pairs used.

REFERENCES

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

Baddeley, A. (2012). Working memory: theories, models, and controversies. Annu Rev Psychol *63*, 1-29.

Barak, O., Rigotti, M., and Fusi, S. (2013). The sparseness of mixed selectivity neurons controls the generalization-discrimination trade-off. J Neurosci *33*, 3844-3856.

Bherer, L., Kramer, A.F., Peterson, M.S., Colcombe, S., Erickson, K., and Becic, E. (2008). Transfer effects in task-set cost and dual-task cost after dual-task training in older and younger adults: further evidence for cognitive plasticity in attentional control in late adulthood. Exp Aging Res *34*, 188-219.

Buonomano, D.V., and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. Nat Rev Neurosci *10*, 113-125.

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

Constantinidis, C., and Klingberg, T. (2016). The neuroscience of working memory capacity and training. Nat Rev Neurosci *17*, 438-449.

Constantinidis, C., and Procyk, E. (2004). The primate working memory networks. Cogn Affect Behav Neurosci *4*, 444-465.

Constantinidis, C., and Qi, X.L. (2018). Representation of Spatial and Feature Information in the Monkey Dorsal and Ventral Prefrontal Cortex. Front Integr Neurosci *12*, 31.

Cortese, S., Ferrin, M., Brandeis, D., Buitelaar, J., Daley, D., Dittmann, R.W., Holtmann, M., Santosh, P., Stevenson, J., Stringaris, A., *et al.* (2015). Cognitive training for attention-deficit/hyperactivity disorder: meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. J Am Acad Child Adolesc Psychiatry *54*, 164-174.

Curtis, C.E., and D'Esposito, M. (2004). The effects of prefrontal lesions on working memory performance and theory. Cogn Affect Behav Neurosci *4*, 528-539.

Dux, P.E., Tombu, M.N., Harrison, S., Rogers, B.P., Tong, F., and Marois, R. (2009). Training improves multitasking performance by increasing the speed of information processing in human prefrontal cortex. Neuron *63*, 127-138.

Foster, J.J., Bsales, E.M., Jaffe, R.J., and Awh, E. (2017). Alpha-Band Activity Reveals Spontaneous Representations of Spatial Position in Visual Working Memory. Curr Biol *27*, 3216-3223 e3216.

Fukuda, K., Awh, E., and Vogel, E.K. (2010). Discrete capacity limits in visual working memory. Curr Opin Neurobiol *20*, 177-182.

Fusi, S., Miller, E.K., and Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. Curr Opin Neurobiol *37*, 66-74.

Jaeggi, S.M., Buschkuehl, M., Jonides, J., and Perrig, W.J. (2008). Improving fluid intelligence with training on working memory. ProcNatlAcadSciUSA *105*, 6829-6833.

Johnston, W.J., Palmer, S.E., and Freedman, D.J. (2020). Nonlinear mixed selectivity supports reliable neural computation. PLoS Comput Biol *16*, e1007544.

Klingberg, T., Fernell, E., Olesen, P., Johnson, M., Gustafsson, P., Dahlstrîm, K., Gillberg, C.G., Forssberg, H., and Westerberg, H. (2005). Computerized Training of Working Memory in Children with ADHD - a Randomized, Controlled Trial. J Am Acad Child Adolesc Psychiatry *44*, 177-186.

Klingberg, T., Forssberg, H., and Westerberg, H. (2002). Training of working memory in children with ADHD. J Clin Exp Neuropsychol *24*, 781-791.

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

Mansouri, F.A., Matsumoto, K., and Tanaka, K. (2006). Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog. J Neurosci *26*, 2745-2756.

Meyer, T., Qi, X.L., and Constantinidis, C. (2007). Persistent discharges in the prefrontal cortex of monkeys naive to working memory tasks. Cereb Cortex *17 Suppl 1*, i70-76.

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

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

Morris, R.G., and Baddeley, A.D. (1988). Primary and working memory functioning in Alzheimer-type dementia. J Clin Exp Neuropsychol *10*, 279-296.

Owen, A.M., Hampshire, A., Grahn, J.A., Stenton, R., Dajani, S., Burns, A.S., Howard, R.J., and Ballard, C.G. (2010). Putting brain training to the test. Nature *465*, 775-778.

Parthasarathy, A., Herikstad, R., Bong, J.H., Medina, F.S., Libedinsky, C., and Yen, S.C. (2017). Mixed selectivity morphs population codes in prefrontal cortex. Nat Neurosci *20*, 1770-1779.

Peijnenborgh, J.C., Hurks, P.M., Aldenkamp, A.P., Vles, J.S., and Hendriksen, J.G. (2015). Efficacy of working memory training in children and adolescents with learning disabilities: A review study and meta-analysis. Neuropsychol Rehabil, 1-28.

Qi, X.L., and Constantinidis, C. (2012a). Correlated discharges in the primate prefrontal cortex before and after working memory training Eur J Neurosci *36*, 3538-3548.

Qi, X.L., and Constantinidis, C. (2012b). Variability of prefrontal neuronal discharges before and after training in a working memory task. PLoS ONE *7*, e41053.

Qi, X.L., Elworthy, A.C., Lambert, B.C., and Constantinidis, C. (2015). Representation of remembered stimuli and task information in the monkey dorsolateral prefrontal and posterior parietal cortex. J Neurophysiol *113*, 44-57.

Qi, X.L., Meyer, T., Stanford, T.R., and Constantinidis, C. (2011). Changes in Prefrontal Neuronal Activity after Learning to Perform a Spatial Working Memory Task. Cereb Cortex *21*, 2722-2732.

Rigotti, M., Barak, O., Warden, M.R., Wang, X.J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature *497*, 585-590.

Rigotti, M., Ben Dayan Rubin, D., Wang, X.J., and Fusi, S. (2010). Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. Front Comput Neurosci *4*, 24.

742 Riley, M.R., and Constantinidis, C. (2016). Role of prefrontal persistent activity in
743 working memory. Front Syst Neurosci *9*, 181.
744 Riley, M.R., Qi, X.L., Zhou, X., and Constantinidis, C. (2018). Anterior-posterior
745 gradient of plasticity in primate prefrontal cortex. Nature communications *9*, 3790.
746 Rossi, A.F., Bichot, N.P., Desimone, R., and Ungerleider, L.G. (2007). Top down
747 attentional deficits in macaques with lesions of lateral prefrontal cortex. J Neurosci *27*,
748 11306-11314.
749 Schwaighofer, M., Fischer, F., and Buhner, M. (2015). Does Working Memory Training
750 Transfer? A Meta-Analysis Including Training Conditions as Moderators. Educ Psychol
751 *50*, 138-166.
752 Shaoyu, Q., Brown, K.A., Orsborn, A.L., Ferrentino, B., and Pesaran, B. (2016).
753 Development of semi-chronic microdrive system for large-scale circuit mapping in
754 macaque mesolimbic and basal ganglia systems. Conf Proc IEEE Eng Med Biol Soc
755 *2016*, 5825-5828.
756 Subramaniam, K., Luks, T.L., Fisher, M., Simpson, G.V., Nagarajan, S., and Vinogradov,
757 S. (2012). Computerized cognitive training restores neural activity within the reality
758 monitoring network in schizophrenia. Neuron *73*, 842-853.
759 Warden, M.R., and Miller, E.K. (2010). Task-dependent changes in short-term memory
760 in the prefrontal cortex. J Neurosci *30*, 15801-15810.
761 Westerberg, H., Jacobaeus, H., Hirvikoski, T., Clevberger, P., Ostensson, M.L., Bartfai,
762 A., and Klingberg, T. (2007). Computerized working memory training after stroke--a
763 pilot study. Brain Inj *21*, 21-29.
764