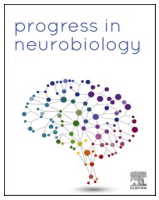
Progress in Neurobiology 251 (2025) 102794

Contents lists available at ScienceDirect

Progress in Neurobiology

journal homepage: www.elsevier.com/locate/pneurobio



Neuronal encoding of recognition memory for numerical quantities in macaque intraparietal and prefrontal cortices

Tobias Machts, Julia Grüb, Andreas Nieder [1,*]

Animal Physiology, Institute of Neurobiology, University of Tuebingen, Auf der Morgenstelle 28, Tuebingen 72076, Germany

ARTICLE INFO

Keywords:
Recognition memory
Single neurons
Monkey
Prefrontal cortex
Intraparietal cortex

ABSTRACT

The parieto-frontal number network in primates is vital for extracting and memorizing numerical information. However, how neurons in these regions retain abstract numerical categories to recognize target numbers amidst ongoing numerical input is unclear. To explore this, single neurons were recorded from the ventral intraparietal cortex (VIP) and lateral prefrontal cortex (PFC) of two male macaques trained to memorize and recognize target numerosities while viewing sequences of irrelevant numerosities. In the VIP, neuronal selectivity for both target and irrelevant numerosities declined rapidly, making it unable to distinguish relevant from irrelevant quantities. Conversely, PFC neurons maintained selective tuning for target numerosities over time but not for irrelevant ones, enabling the distinction between sought and irrelevant quantities. Match enhancement effects, where firing increased for repeated target numerosities, were observed only in the PFC. In contrast, match suppression effects, involving reduced firing for repeated target numerosities, occurred in both the VIP and PFC. These findings suggest the VIP primarily encodes displayed numerosities, while the PFC is specialized for processing, storing, and recognizing numerical quantities by enhancing familiar numerosities. This highlights the PFC's key role in recognition memory, contrasting with the transient coding observed in the VIP.

1. Introduction

We usually encounter important objects or events not only once in our lives, but several times. The cognitive function that enables us to recognize that something currently perceived has been experienced before is called recognition memory, a critical component of explicit memory (Winters et al., 2008). While recognition memory has been extensively studied for arbitrary sensory stimuli (Brown and Aggleton, 2001), its exploration in the context of abstract categories, such as numbers, has been neglected. Recognition memory of numbers is essential for tasks such as remembering item prices or accurately dosing medications. Recognition memory of numerosities — the number of items in a set — also enables animals to efficiently navigate their environment, make informed decisions, and engage in behaviors critical for survival and reproduction (Nieder, 2020).

Studies of sensory recognition memory report neurons whose activity indicates whether a stimulus is familiar or novel (Rainer and Miller, 2000; Brown and Aggleton, 2001). Most studies have identified neurons in high-level cortical areas whose response magnitudes decrease with repetition and thus familiarity, a phenomenon known as 'repetition suppression' (Fahy et al., 1993; Li et al., 1993; Miller and Desimone, 1994; Riches et al., 1991; Xiang and Brown, 1998; Ng et al., 2014). However, other mechanisms for signaling familiarity have also been observed (Meyer and Rust, 2018), such as more selective tuning (Rainer and Miller, 2000; Freedman et al., 2006; Anderson et al., 2008;), tuning peak enhancement (Woloszyn and Sheinberg, 2012; Lim et al., 2015), or even increases in firing rate (Tamura et al., 2017). Understanding which signals correlate with perceptual memories and how these neuronal correlates differ along the cortical processing hierarchy remains an area of active research.

While sensory recognition memory has been studied using arbitrary pictogram stimuli, the memory necessary for recognizing abstract categories, such as numerical quantities, in a sequence of numerosities has not been investigated. The special feature of numerical quantities in contrast to arbitrary visual stimuli is that they are semantically related to each other by ordinal magnitude (e.g., three is greater than two but less than four). This relationship is often described using the metaphor of the 'mental number line,' where numbers are arranged in an ordered

* Corresponding author.
E-mail address: andreas.nieder@uni-tuebingen.de (A. Nieder).
[1] ORCID: 0000-0001-6381-0375

https://doi.org/10.1016/j.pneurobio.2025.102794
Received 6 December 2024; Received in revised form 1 May 2025; Accepted 10 June 2025
Available online 13 June 2025

sequence (Galton, 1880; Dehaene et al., 1993; Zorzi et al., 2002; Nieder, 2020). Recognition memory of numerical quantities therefore heavily depends on the representation of, or tuning to, target numerosities as opposed to neighboring non-target numerosities as a unique feature of parameterized quantities.

A key behavioral phenomenon related to numerical cognition based on the number line is the 'numerical distance effect' (Moyer and Landauer, 1967; Buckley and Gillman, 1974). This effect highlights that it is more challenging to differentiate between numbers that are close together (e.g., 2 and 3) compared to those that are further apart (e.g., 2 and 8). This effect is mirrored in the neurons' response properties. In the parieto-frontal number network of primates, numerosity-selective neurons are specifically responsive to the presentation and short-term memorization of particular numerical values (Nieder et al., 2002, 2006; Sawamura et al., 2002; Nieder and Miller, 2003; 2004; Nieder and Merten, 2007; Okuyama et al., 2015; Nieder, 2016). Like sensory and motor neurons that are tuned to a specific feature or parameterized coding space, numerosity-selective neurons exhibit a bell-shaped tuning curve. This means they have a maximum firing rate for their preferred numerosity and show progressively lower activity as the presented numbers deviate from this preferred value. The amplitude and width of these tuning curves reflect the neurons' selectivity for specific numerosities (Nieder, 2023). This tuning selectivity is a key aspect not only of sensory numerosity representations but also for the maintenance of numerical recognition memory; neurons whose tuning deteriorates with time and numerosity repetition would be unsuitable as a correlate of recognition memory for numerosity.

In this study, we investigated where and how numerosity-selective neurons in the parieto-frontal number network can serve as a neuronal representation of recognition memory for numerical quantity. Both the areas within the intraparietal sulcus (IPS) and the lateral prefrontal cortex (PFC) have been repeatedly found to encode numerical information, forming the core number network in primates (Nieder, 2024). In contrast, such information is largely absent in the anterior inferotemporal cortex (Nieder and Miller, 2004).

We trained two rhesus macaques to perform a sequential delayed-match-to-numerosity task with variable numbers of non-matching visual numerosity stimuli between sample and matching test displays. The delayed match-to-numerosity task involves a working memory component that interacts closely with recognition memory (Miller et al., 1996). While working memory temporarily stores and manipulates information, recognition memory retrieves and identifies previously encountered information. Both memory systems are crucial for accurate decision-making and memory-based tasks, often working together to support effective cognitive processing.

The monkeys were required to form an internal representation or template, akin to a search image, for recognizing and identifying specific cued numerical quantities. While the monkeys performed this task, we simultaneously recorded single-neuron activity from the lateral PFC and the ventral intraparietal area (VIP) in the intraparietal cortex, two key number processing areas in the primate brain (Nieder and Miller, 2004; Viswanathan and Nieder, 2013; 2015,; Nieder and Dehaene, 2009; Nieder, 2016). To explore the neuronal correlate for recognition memory of numerical quantity categories, we quantified and compared the numerosity tuning curve characteristics and population decoding accuracies for repeated presentations of both matching and non-matching numerosities in the two cortical areas under the exact same recording conditions.

2. Results

2.1. Behavioral performance

Two monkeys were trained on a modified version of the delayed match-to-numerosity task. They were presented with sample numerosities 1–4, followed by a delay period, and then exposed in equal proportions of up to four different sequential test phases (see Materials and Methods; Fig. 1A). The monkeys were required to memorize the variable sample numerosity and to search the sequence of test displays for a numerosity that matched the sample numerosity, and they were rewarded for responding whenever one of the displays in the test periods showed a matching numerosity (correct trials). This encouraged them to inspect each test stimulus and to keep searching for the matching numerosity.

Both monkeys mastered the task, with mean performances across sessions significantly exceeding the chance level of 25 % (monkey 1: 27 sessions, 61.32 ± 1.36 %, $p < 0.001$; monkey 2: 33 sessions, 60.95 ± 0.58 %, $p < 0.001$, Binomial test) (see Fig. S1 for chance level calculation). To evaluate the monkeys' ability to identify the correct matching numerosity while disregarding the nonmatching numerosities, behavioral performance functions were computed for each session. These functions reflect the probability that a monkey judged the numerosities in the test periods as equal to the sample numerosity. For all sample numerosities, both monkeys more frequently selected the matching numerosity over the nonmatching numerosities (Fig. 1B,D). The best performance was achieved for numerosity 1 (peak: monkey 1, 95.02 ± 1.12 %; monkey 2, 96.01 ± 0.52 %) and declined with increasing numerosity up to numerosity 4 (peak: monkey 1, 56.31 ± 1.96 %; monkey 2, 64.57 ± 1.99 %). As expected, based on the numerical distance effect, which suggests that neighboring numerosities are the most challenging to discriminate (Nieder and Miller, 2003; Merten and Nieder, 2009), both monkeys exhibited the highest error rates for nonmatching numerosities adjacent to the respective target numerosity, with noticeable improvements for more distant nonmatch numerosities. Additionally, the performance functions systematically widened (i.e., became less selective) for increasing target numerosities, reflecting the numerical size effect, which indicates that, at a given numerical distance, discriminating target numerosities becomes more difficult as their values increase (Nieder, 2020). Thus, the monkeys' performance exhibits the characteristics of the approximate number system (ANS), a non-symbolic estimation system that also emerges in humans when they are prevented from counting and rely on non-symbolic numerosity estimation (Merten and Nieder, 2009; Nieder, 2024).

We computed the averaged performance for each trial type (i.e., no nonmatch, one nonmatch, two nonmatches, three nonmatches) across the four sample numerosities (Fig. 1C,E). As anticipated, the number of inserted nonmatch numerosities influenced the monkeys' ability to recall and detect the matching numerosity (two-way ANOVA with the factors 'number of nonmatches' and 'sample numerosity'; monkey 1: $F(3431) = 60.1$, $p < 0.001$; monkey 2: $F(3527) = 66.49$, $p < 0.001$). Compared to the pure matching trials in which no nonmatch numerosity was shown, the monkeys' overall performance progressively decreased with the number of nonmatches shown in a trial. The observed decline in behavioral performance with increasing nonmatch trials reflects the growing cognitive demands on recognition memory, a critical aspect of the study. Monkey 1, but not Monkey 2, shows a mild, idiosyncratic performance increase for numerosities following three nonmatches. However, this pattern is uncorrelated with Monkey 2's behavior and likely reflects an individual, unknown strategy. Furthermore, behavioral performance depended on the presented sample numerosity within each trial type (significant interaction of trial type and presented sample numerosity; monkey 1 $F(9, 431) = 5.16$, $p < 0.001$; monkey 2 $F(9, 527) = 4.27$, $p < 0.001$). As a reflection of the numerical size effect, (balanced) trials in which the monkeys had to search for numerosity 1 were the easiest, while trials in which numerosity 4 was the target challenged the monkeys the most (Fig. 1**C,E**). Nevertheless, the averaged performances for all trial types and sample numerosities were significantly higher than chance (for all sample numerosities and trial types, $p < 0.05$, Binomial test), revealing that the monkeys were able to find the correct numerosity most of the time.

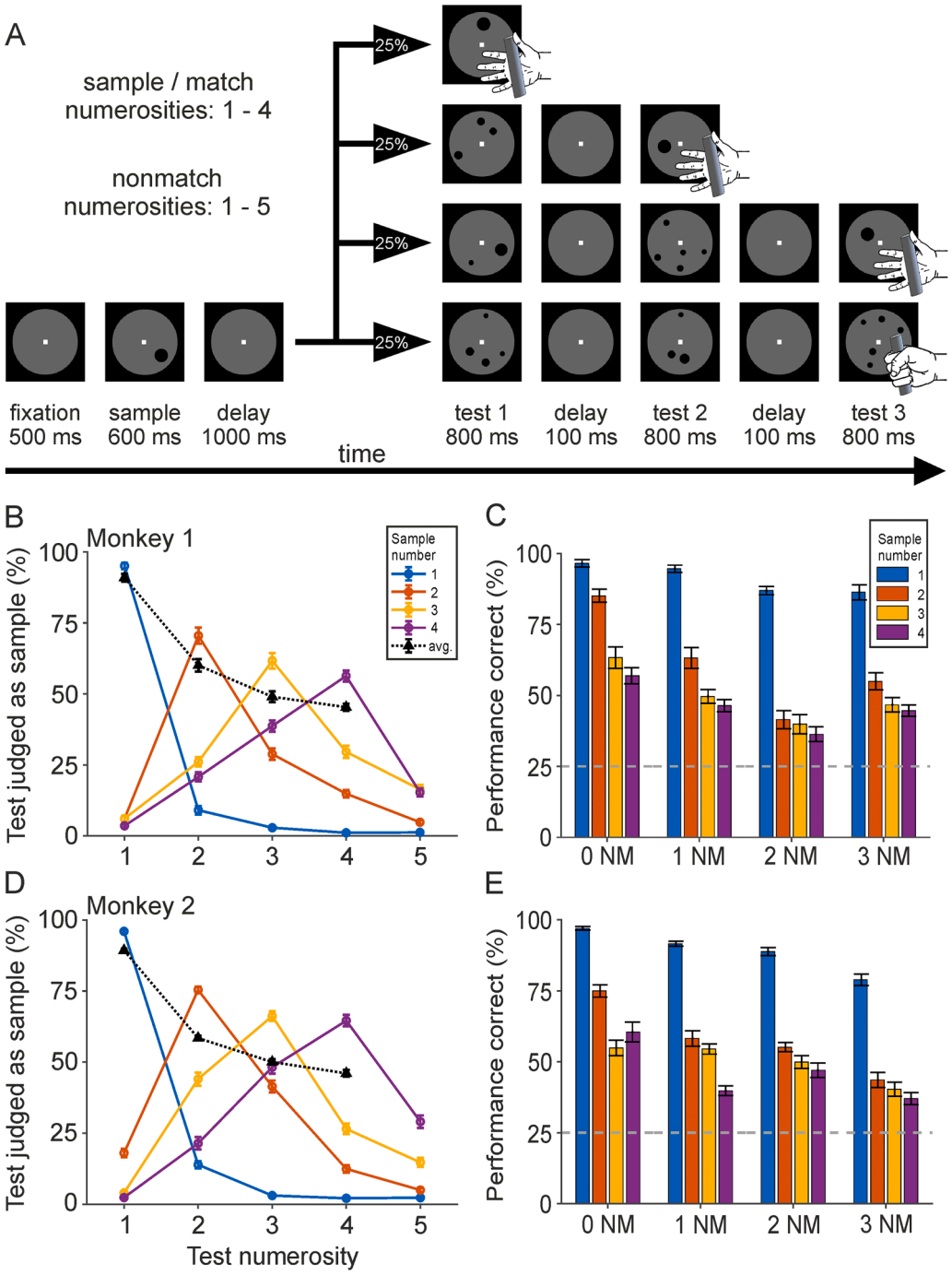


Fig. 1. Task protocol and behavioral performance. (A) Monkeys performed a delayed match-to-sample task with dot numerosities. A trial started when the monkey grabbed the response bar and fixated the fixation target; both actions were required to be maintained throughout the task until the monkey responded by releasing the bar. Numerosities ranging from 1 to 4 were presented during a sample phase that was followed by a delay period. The monkey had to respond whenever a numerosity that matched the sample was presented in one of the three subsequent Test phases. The monkeys had to refrain from responding for as long as non-matching numerosities (ranging from 1 to 5) were presented in the Test phases. Up to three sequential Test-phases (Test1-Test3) were presented pseudo-randomly in different trial conditions; in each Test-phase, a match or nonmatch-numerosity could appear. If the last Test3 period showed a nonmatch, the monkey was required to maintain the bar until display-offset. This resulted in a total four different trial conditions. Sample numerosities as well as trial conditions were presented in a balanced way. (B, D) Behavioral performance functions of the two monkeys to sample numerosities 1–4 (monkey 1: 27 sessions; monkey 2: 33 sessions). The functions reflect the probability that a monkey judged displays in the test periods as containing the same number of items as the sample numerosity (indicated in four colors, solid line). The peak data points of each colored function indicate correct sample matching. Data points in the flanks reflect error responses to nonmatch numerosities. Black dotted line indicates averaged performance for each sample numerosity. Error bars (very small) indicate the SEM across sessions. (C, E) Average correct performance of the two monkeys to the four sample numerosities and the four Test conditions (NM = nonmatch). Grey dashed line indicates chance level at 25 %. Error bars indicate the SEM across sessions.

2.2. General neuronal response properties

We recorded a total of 372 neurons from area VIP in the fundus of the intraparietal sulcus (IPS) and 509 neurons from the dorsolateral prefrontal cortex (dlPFC, centered around the principal sulcus) from two monkeys (Fig. 2A). The proportions of neurons for each monkey can be seen in Table S1. Throughout the paper, we focused our analyses on activity during the presentation of sample and test stimuli, including matching test stimuli to which the monkeys responded quickly by releasing a bar.

To determine the starting point of the neuronal analysis window, we utilized the neurons' visual response latencies. Visual response latencies were measurable in 51 % of the VIP neurons (190/372) and in 64 % of the PFC neurons (324/509). VIP neurons exhibited significantly shorter latencies compared to PFC neurons (median VIP: 111 ms, median PFC: 124 ms, Wilcoxon rank sum test, $Z = 2.6619$, $p = 0.0078$, two-sided) (Fig. 2B). We standardized the analysis window by using the first 25th percentile of the VIP (86 ms) and PFC latency distributions (100 ms) as the standard latency for all neurons, including those for which we could not determine individual latencies based on our criteria.

To ensure that neuronal responses were not contaminated by motor activity — that is, activity related to neurons involved in preparing and executing the lever release movement — we used the monkeys' reaction time (RT) distributions of correct responses to the matching test stimulus as endpoints of the duration of the neuronal analysis windows. In 84 % (monkey 1) and 89 % (monkey 2) of the cases, respectively, the monkeys responded after 400 ms or longer to the match. Therefore, we used 400 ms as the analysis window endpoint to ensure that neuronal activity is analyzed in equal time windows between the two monkeys and analyzed all correct trials in which the monkeys responded with an RT of at least 400 ms (trials with shorter RTs were excluded). Thus, VIP neurons were analyzed in a time window from 86 ms to 400 ms after sample/test onset, and PFC neurons from 100 ms to 400 ms. Qualitatively similar results in numerosity tuning were achieved by using 300 ms as the endpoint (cutoff) of the analysis window (see Fig. S2), confirming that the results were not due to motor-preparatory activity.

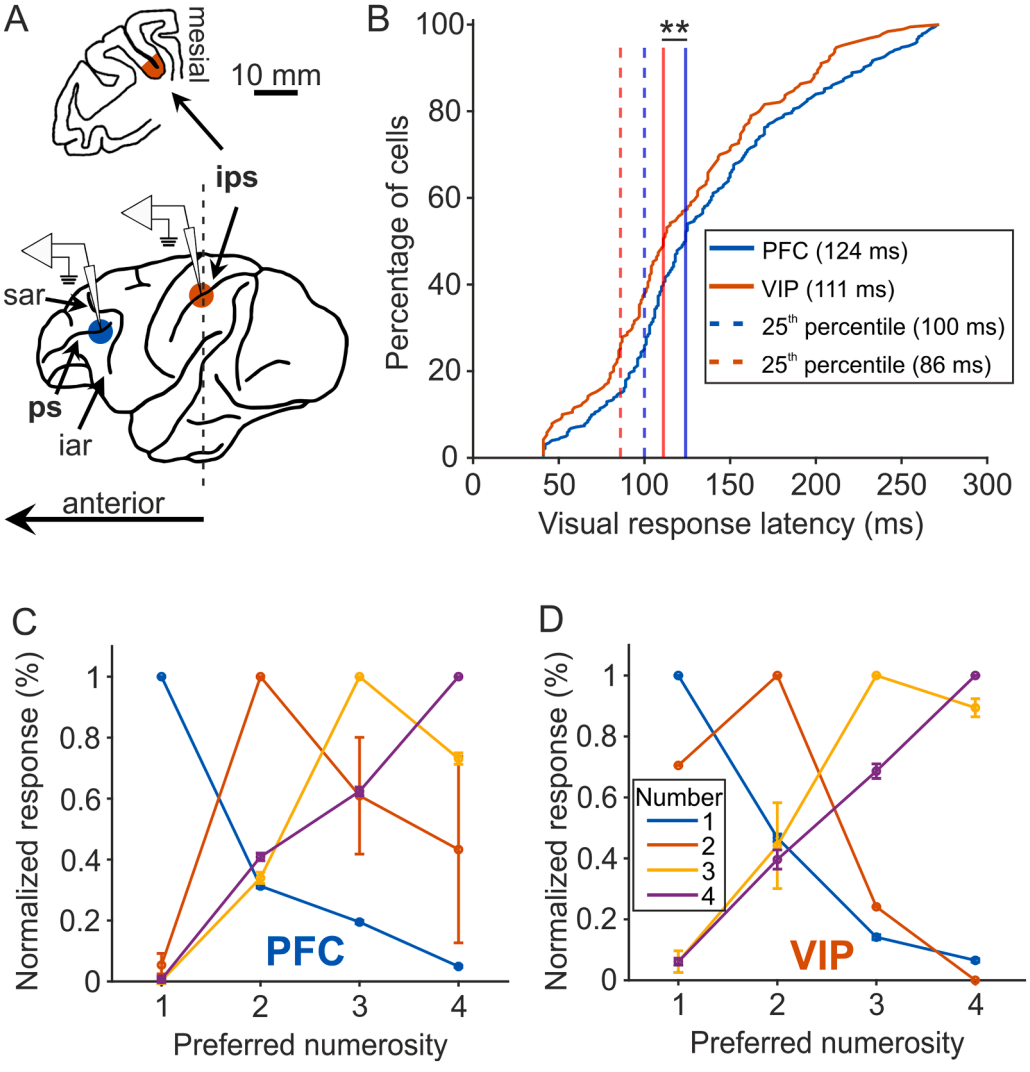


Fig. 2. Recording sites, visual response latencies and population tuning curves of PFC and VIP neurons. (A) Schematic diagram of the macaque brain illustrating the locations in PFC (blue) and VIP (red) where recordings were performed. Abbreviations: ips, intraparietal sulcus; ps, principal sulcus; sar, superior arcuate sulcus; iar, inferior arcuate sulcus. (B) Cumulative latency distributions across all neurons recorded in PFC (N = 324, blue) and VIP (N = 190, red) that showed significant firing rate change compared to baseline activity after sample onset. Numbers next to labels indicate the median latencies and latencies at the 25th percentiles in PFC and VIP. The vertical solid lines indicate the medians and the vertical dashed lines show the 25th percentiles. **$p < 0.01$. (C) Population tuning curves for numerosity-selective neurons in PFC (n = 68) during the sample period. Population tuning curves were obtained by averaging the normalized tuning curves of all single units preferring the same sample numerosity (see color-code). Error bars indicate the SEM. Numerosity 1, 2, 3, 4: $n = 35, 2, 11, 20$. (D) Population tuning curves for numerosity-selective neurons in VIP (n = 34) during the sample period. Same layout as in (C). Numerosity 1, 2, 3, 4: $n = 19, 1, 3, 11$.

2.3. Numerosity selective neurons

Using this time window, we initially identified neurons that exhibited numerosity selectivity (i.e., tuning to numerosity) during sample presentation. Since we were specifically interested in tuning changes over time, we only selected neurons with at least 18 correct trials per sample numerosity and 3 correct trials per Test numerosity (Test 1–3 periods) for further analysis. Thus, the analyzed population comprised a total of 369 PFC neurons and 254 VIP neurons that met all inclusion criteria.

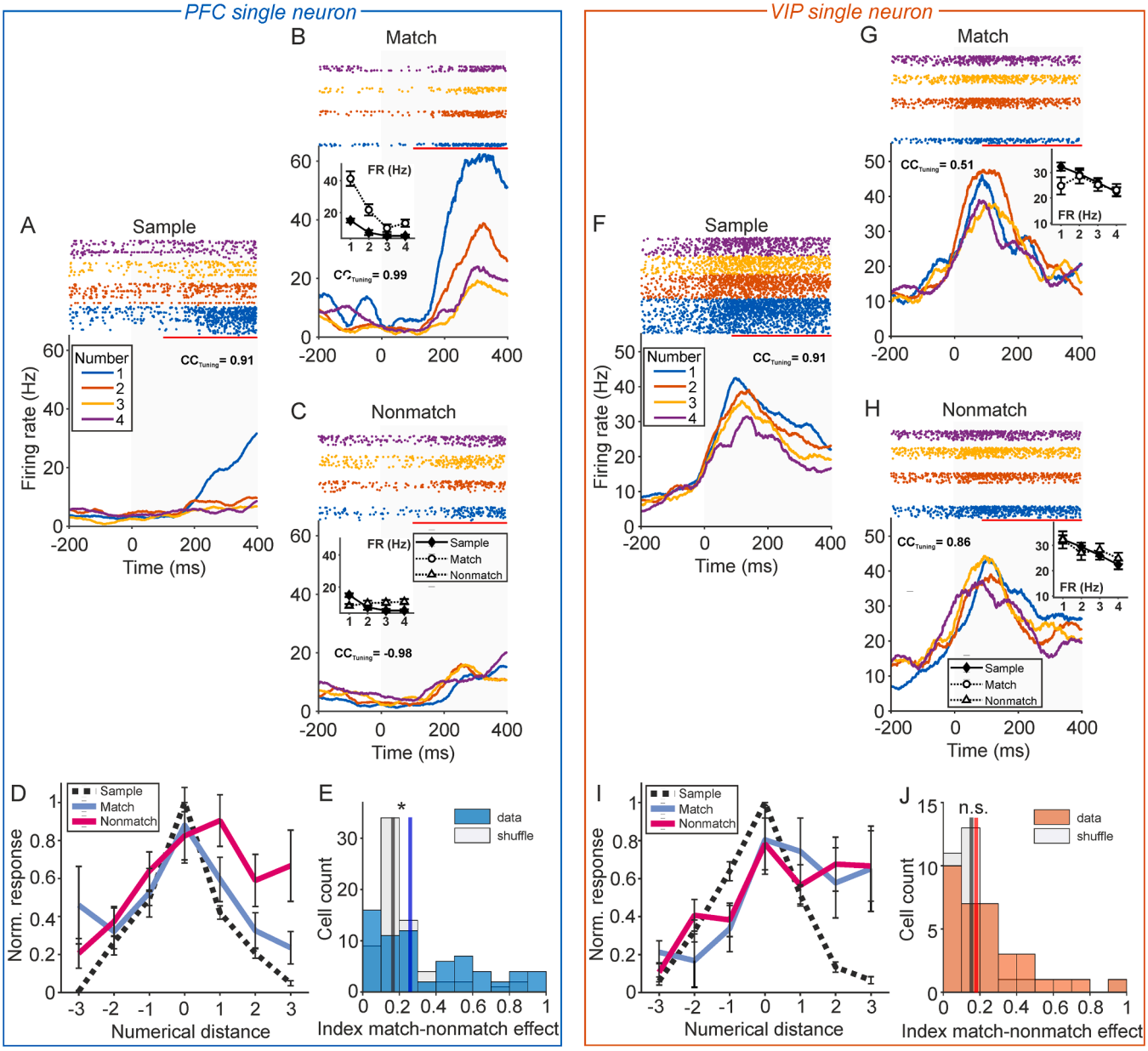


Fig. 3. Numerosity selectivity of single neurons during the sample and Test1 match and nonmatch periods. (A-C) Responses of an exemplary PFC neuron preferring numerosity 1 in the sample period (A), as match (B), and as nonmatch numerosities (c) during Test1. Top panels show dot-raster histograms (each line represents a trial; each dot represents an action potential) and bottom panels show smoothed spike-density functions (activity averaged and smoothed with a 75-ms Gaussian kernel). The colors of dots and spike density functions correspond to the numerosity of the sample/test stimulus. Time 0 is the onset of the stimulus in the sample (A) and Test1 period (B,C), indicated by the gray shaded areas. Insets show tuning curves calculated during the analysis window indicated by the red bar plotted at the top of the spike-density functions. The tuning function for the sample period (solid line) is plotted together with the tuning functions of match situation (circle, dotted line) in (B), or nonmatch situation (triangle, dotted line) in (c). CC_{Tuning}-values in spike-density histograms represent the cross-correlation coefficients between cross-validated tuning curves during the sample (A), a comparison between sample and match tuning curves (b), and a comparison between sample and nonmatch tuning curves (c). Error bars show the SEM. (D) Normalized and averaged numerosity tuning curves of all numerosity-selective neurons in PFC (N = 68) during the sample, match, and nonmatch phases. Prior to averaging, neuronal activity for each neuron was aligned to the preferred numerosity at center value 0 and plotted as a function of numerical distance. (E) Distribution of indices quantifying the differences between responses to match and nonmatch numerosities during Test1 in PFC (N = 68). The median of the real data (blue) indicated by the colored vertical line were tested against a shuffled distribution (grey) indicated by a black line. *$p < 0.05$. (F-H) Responses of an exemplary VIP neuron preferring numerosity 1 in the sample period (F) as match (G), and as nonmatch numerosities (H) during Test1. Same layout as in (A-C). (I) Normalized and averaged numerosity tuning curves of all numerosity-selective neurons in VIP (N = 34) during the sample, match, and nonmatch phases. Same layout as in (D). (J) Distribution of indices showing the strength of the match-nonmatch effect in VIP (N = 34). Real data (orange) were tested against a shuffled distribution (grey). n.s. not significant. Layout as in (E).

To identify numerosity-selective neurons, we applied a two-way ANOVA with main factors of 'numerosity' and 'protocol', evaluated at $p < 0.01$. We only considered neurons that were exclusively numerosity-selective, showing a main effect for numerosity but no main effect for protocol or interaction. By combining both selection criteria, we found 34 numerosity-selective neurons with a sufficient number of trials in VIP (34/254; 13.4 % of the recorded neurons) and 68 neurons in PFC (68/369; 18.4 % of the recorded neurons). In PFC, only 3.7 % of all neurons were non-exclusivelynumerosity selective, i.e. they showed in addition to a main effect 'numerosity', a main effect 'protocol' and/or an interaction between 'numerosity' and 'protocol'. In VIP, only 2.2 % of all neurons were non-exclusively numerosity selective. The proportions of exclusively numerosity-selective neurons reported in the current study are comparable to what we found in previous studies with other monkeys (on average, VIP: 11 %; PFC: 19 %) (Diester and Nieder, 2007; Viswanathan and Nieder, 2015; 2020). The average tuning functions for selective neurons in the two regions are shown in Fig. 2C,D.

To assess the robustness of numerosity tuning, we cross-validated the tuning functions of individual selective neurons by correlating tuning curves derived from odd versus even trials. In both brain areas, we found cross-correlation coefficients close to 1 (PFC: $CC_{median} = 0.89$, VIP: $CC_{median} = 0.90$), indicating strong and stable numerosity tuning in both PFC and VIP. The correlation coefficients did not differ significantly between PFC and VIP (Wilcoxon rank-sum test, $Z = -0.0674$, $p = 0.9462$, two-sided), suggesting that tuning stability was comparable between neurons in both regions (Suppl. Fig. S3).

We next evaluated the numerosity tuning of sample-selective neurons during the delay working memory period. Consistent with previous findings (Nieder et al., 2002), a similar proportion of sample-selective numerosity neurons were also tuned to numerosity during the first delay period after the sample in both the PFC and VIP (PFC: 25/68 = 37 %; VIP: 15/34 = 44 %). Moreover, numerosity tuning between the sample and the first delay period was correlated in PFC neurons but not in VIP neurons (Figure S4).

After calculating neuronal tuning functions for each of these neurons, we explored the stability of their tuning relative to sample presentation in two dimensions. First, tuning stability was examined at the identical trial time when Test1 was either a match or a nonmatch. We hypothesized that tuning remained stable to different degrees in VIP and PFC when the monkey was presented with the matching numerosity it was searching for, but deteriorated when the monkey saw a deviating numerosity for what it was not seeking in a given trial. Second, tuning stability was investigated across all three succeeding test periods (Test1 to Test3) when the respective test periods were a match. We anticipated that tuning to the matching numerosity would remain stable across time when the monkey was searching for it throughout the trial, potentially with differences between VIP and PFC. We start with the first question: tuning as a function of matching or nonmatching numerosity.

2.4. Tuning stability to match and nonmatch numerosity in the Test1 period

We initially examined the tuning functions of single neurons. Fig. 3A-C illustrates the detailed discharges and the tuning functions (insets) of an example PFC neuron tuned to numerosity 1 during sample presentation for the match and nonmatch Test1 period. When a matching numerosity was displayed, this neuron's tuning remained stable and exhibited enhanced firing rates for the match period (Fig. 3B). In contrast, this neuron's numerosity selectivity was significantly diminished when a nonmatch numerosity was presented in the same Test1 period, resulting in a more or less flat tuning function (Fig. 3C). Similar stability in tuning to match numerosities, but degraded tuning for nonmatch numerosities, was observed for the majority of PFC neurons.

In contrast, VIP neurons displayed less stable tuning. Fig. 3F-H portrays an example VIP neuron tuned to numerosity 1. Despite clear tuning to the sample numerosities (Fig. 3F), this neuron exhibited poorer tuning to the match numerosity (Fig. 3G) and similarly deteriorated tuning to the nonmatching numerosities (Fig. 3H).

To quantify tuning stability as a function of match versus nonmatch numerosity, we calculated normalized tuning curves with firing rates during sample presentation as the reference for all stimulus presentation phases. Subsequently, the normalized tuning functions were averaged according to the respective preferred sample numerosities for match and nonmatch conditions. Finally, the normalized tuning functions for match and nonmatch conditions were aligned relative to the preferred sample numerosities and plotted as a function of numerical distance from the respective preferred numerosities, with the preferred numerosities set to numerical distance 0. This approach allowed us to assess how neuronal tuning varied across different stimulus conditions and numerical distances from the preferred numerosities.

In PFC, the population-tuning curve for the match condition was comparable to the sample tuning curve and showed a corresponding preferred numerosities (Fig. 3D). Interestingly, this was not the case for the nonmatch presentation; here, the population-tuning curve was asymmetric, and the peak deviated from the preferred numerosity. We have previously shown that numerosity tuning curves are asymmetric on a linear number scale (e.g., Nieder and Miller, 2003), with this effect being less visible in sharp tuning functions during the sample period and more pronounced as neurons become less selective in the Test1 period. To quantify this difference between sample and Test1 periods, we fitted Gauss functions to the individual tuning curves and derived the sigma-value as a measure of tuning width, or tuning selectivity, respectively.

The Gauss fits described the tuning functions well. In both PFC and VIP, the goodness-of-fit (R^2) for the fits during sample period (the reference period) were close to 1 (PFC: median $R^2 = 0.95$; VIP: median $R^2 = 0.91$). The R^2 values were higher during the sample period compared to the Test periods, but no difference was found between the match (PFC: median $R^2 = 0.71$; VIP: median $R^2 = 0.57$) and nonmatch periods (PFC: median $R^2 = 0.78$; VIP: median $R^2 = 0.65$), neither in PFC (Wilcoxon signed-rank test, $Z = -0.5214$, $p = 0.6021$, two-sided) nor in VIP (Wilcoxon signed-rank test, $Z = -0.9704$, $p = 0.3318$, two-sided).

In PFC, no difference in tuning width was found for sample and match conditions (median sample: $\sigma = 1.24$, median match: $\sigma = 1.45$, Wilcoxon signed rank test, $Z = -1.8392$, $p = 0.1977$, two-sided, with Bonferroni correction), indicating that PFC neurons maintained their numerosity selectivity between these two trial phases. In contrast, the tuning functions in the nonmatch condition were significantly wider, both compared to the match condition (median match: $\sigma = 1.45$, median nonmatch: $\sigma = 2.01$, Wilcoxon signed rank test, $Z = -2.4686$, $p = 0.0407$, two-sided, with Bonferroni correction, $n = 68$) and the sample period (median sample: $\sigma = 1.23$, median nonmatch: $\sigma = 2.01$, Wilcoxon signed rank test, $Z = -4.3261$, $p < 0.001$, two-sided, with Bonferroni correction). Thus, tuning selectivity of PFC neurons specifically decreased if a nonmatch numerosity was shown.

In VIP, in contrast, the population-tuning curves for both the match and nonmatch conditions were deteriorated relative to the sample tuning curve (Fig. 3I). Gauss function fitted to the tuning curves showed that the tuning widths for both the match condition (median sample: $\sigma = 1.42$, median match: $\sigma = 2.35$, Wilcoxon signed rank test, $Z = -2.9491$, $p = 0.0096$, two-sided, with Bonferroni correction) and the nonmatch condition (median sample: $\sigma = 1.42$, median nonmatch: $\sigma = 2.58$, Wilcoxon signed rank test, $Z = -3.1372$, $p = 0.0051$, two-sided, with Bonferroni correction) were significantly larger compared to the sample period. In addition, no tuning widths difference was observed between match and nonmatch conditions (median match: $\sigma = 2.34$, median nonmatch: $\sigma = 2.58$, Wilcoxon signed rank test, $Z = 0.6240$, $p = 1$, two-sided, with Bonferroni correction, $n = 34$). This speaks for a deterioration of tuning selectivity in VIP from sample to Test1 that is independent of whether a match or a nonmatch numerosity is presented.

We quantified the magnitude of firing rate differences to the preferred numerosity between match and nonmatch conditions in both

brain areas by calculating an index capturing the match-versus-nonmatch effect (Miller et al., 1996; see Methods). The higher the index, the stronger the effect (values of 0 indicate equal responses to preferred numerosities in matches and nonmatches). We compared the medians of each distribution within each brain area with a shuffled distribution representing chance values. In PFC, the real index (median 0.26) was significantly higher than the shuffled index (0.17) ($Z = 0.9567$, $p = 0.0209$, Wilcoxon signed rank test, $n = 68$) (Fig. 3E). In VIP, in contrast, the real index (median 0.18) was not different from the shuffled index (0.15) ($Z = 0.9935$, $p = 0.3387$, Wilcoxon signed rank test, $n = 34$) (Fig. 3**J**). Moreover, a direct comparison between brain areas showed that the real indices for PFC were significantly larger compared to VIP ($Z = 1.6824$, $p = 0.0462$, Wilcoxon rank sum test). In sum, match-nonmatch effects were prominent in PFC, but absent in VIP.

A matching numerosity in the Test1 period could either enhance the firing rates to the preferred numerosity, leading to a positive index, or suppress the firing rates to the preferred numerosity, causing a negative index (Miller et al., 1996). We therefore categorized the neurons into match-enhancement and match-suppression cells and compared the indices for the two separate classes in PFC and VIP. Negative indices of match-suppression cells were rectified by using the absolute values for better comparison, thus absolute index values larger than 0 indicate differential responses to preferred numerosities in matches and nonmatches.

The indices for neurons exhibiting a match-enhancement effect in the PFC ($n = 31$) and VIP ($n = 14$), respectively, are shown in Figs. 4A and 4B. The median enhancement index for the PFC was 0.23, compared to 0.15 in the VIP. We compared these medians within each brain area using a shuffled distribution representing chance values. In the PFC, the real index (median 0.23) was significantly higher than the shuffled index (0.13) ($Z = 2.4785$, $p = 0.0132$, Wilcoxon signed-rank test, $n = 31$). In contrast, the real index in the VIP (median 0.15) was not significantly different from the shuffled index (0.12) ($Z = 1.5380$, $p = 0.1240$, Wilcoxon signed-rank test, $n = 14$). Furthermore, direct comparisons between brain areas showed that the real indices for the PFC tended to be larger than those for the VIP ($Z = 1.2871$, $p = 0.0990$, Wilcoxon rank-sum test).

The indices for neurons showing a match-suppression effect in the PFC ($n = 37$) and VIP ($n = 20$) are depicted in Figs. 4C and 4D. In both the PFC and VIP, the real indices (median 0.30 and median 0.22) were significantly higher than the shuffled indices (0.14 for both) ($Z = 3.6199$, $p < 0.001$, Wilcoxon signed-rank test, $n = 37$ for PFC; $Z = 2.6157$, $p = 0.0089$, Wilcoxon signed-rank test, $n = 20$ for VIP). Furthermore, we found no significant difference when comparing the distributions between the two brain areas ($Z = 1.1455$, $p = 0.1260$, Wilcoxon rank-sum test). In summary, significant match enhancement effects were found only in the PFC, while significant match suppression effects occurred in both the PFC and VIP with similar magnitudes.

Finally, to explore the information contained in all PFC and VIP neurons, regardless of their numerosity tuning, we conducted a time-resolved percent explained variance (PEV, ω^2) analysis (PFC: $n = 369$, VIP: $n = 254$; see Methods for details). Percent explained variance (PEV) is a statistical measure that quantifies how much of the variability in our neuronal dataset is explained by the factor 'match' or 'nonmatch'. Both PFC and VIP neurons represented match information throughout the sample and delay periods (greater than the significance threshold of the shuffled distribution at $p < 0.01$), while nonmatch information was still absent ($p < 0.01$) (Fig. 5). Nonmatch information was only represented by neuron populations during the Test1 period, with important differences between PFC and VIP. In PFC, match information continued to increase and consistently remained higher than nonmatch information ($Z = 3.8263$, $p < 0.001$, Wilcoxon signed-rank test, $n = 369$). In VIP, however, match information remained constant during Test1, showing no significant difference from nonmatch information ($Z = 1.6477$, $p = 0.0994$, Wilcoxon signed-rank test, $n = 254$). This suggests that PFC neurons, but not VIP neurons, irrespective of numerosity tuning, discriminate between match and nonmatch information. These effects were also observed in PFC and VIP neurons of individual monkeys. In monkey 2, however, the match information seems to be more strongly maintained in the population code during the delay and does not appear to be updated in VIP, whereas the non-match information is (Suppl. Fig. S5).

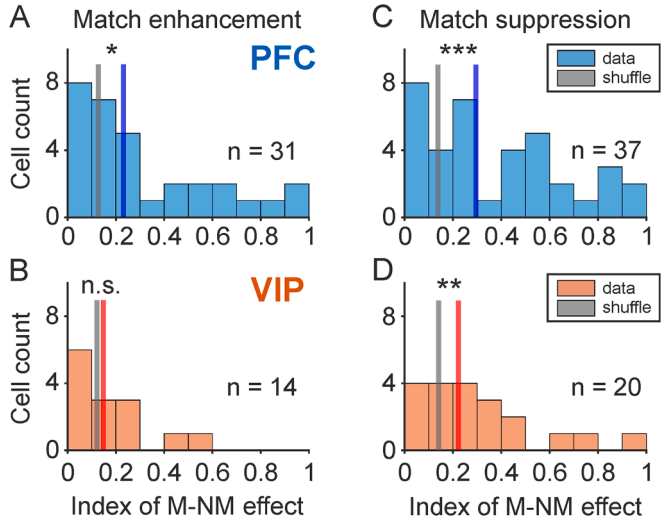


Fig. 4. Distribution of match-nonmatch indices showing the strength of the match enhancement effect and match suppression effect in PFC and VIP in the Test1 period. (A) Indices for neurons exhibiting a match enhancement effect in the PFC. The index is calculated as the absolute value of the difference between match and non-match responses, divided by their sum. The median of the real data, indicated by the blue vertical line, was tested against a shuffled distribution, whose median is indicated by a black line. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; n.s. not significant. (B) Indices for neurons exhibiting a match enhancement effect in VIP. Layout as in (A). (C) Indices for neurons exhibiting a match suppression effect in PFC. Layout as in (A). (D) Indices for neurons exhibiting a match suppression effect in VIP. Layout as in (A).

2.5. Population decoding of match and nonmatch numerosity in the Test1 period

So far, our analysis of single-neuron responses suggests that numerosity-selective PFC neurons have a privileged role in coding and maintaining sought after numerical information. We further tested this by a population decoding analysis using a support vector machine (SVM) classifier. SVM classifiers offer a powerful approach to decoding neural activity by simultaneously considering the combined activity of multiple neurons, capturing complex multidimensional patterns and non-linear relationships underlying population coding. This makes classifier decoding more robust to noise and complexity than simple averaging approaches. We trained the classifier on the firing rates of numerosity-selective neurons (balanced neuron numbers for the two brain areas) during the sample period. Next, we tested the trained classifier on its accuracy to discriminate numerosities during the match or nonmatch Test1 period (6-fold cross validated, 100 times resampled; see Methods).

Fig. 6A depicts the mean decoding accuracy for the sample (as a reference), the match Test1 period, and the nonmatch Test1 period in both PFC and VIP, respectively. In all conditions and brain areas, the decoding was significantly above chance (larger than the 95th percentile of the respective shuffled distribution; PFC: Sample, Match, Nonmatch [36.11 %, 33.33 %, 31.94 %]; VIP: Sample, Match, Nonmatch [35.42 %, 31.25 %, 32.64 %]).

We next compared the classifier's decoding quality between PFC and VIP neurons. In the sample period, decoding accuracy was significantly higher in PFC (mean 60 ± 0.66 %) than in VIP (53 ± 0.66 %) (permutation test, $p < 0.001$, with alpha correction, $n = 100$). As expected,

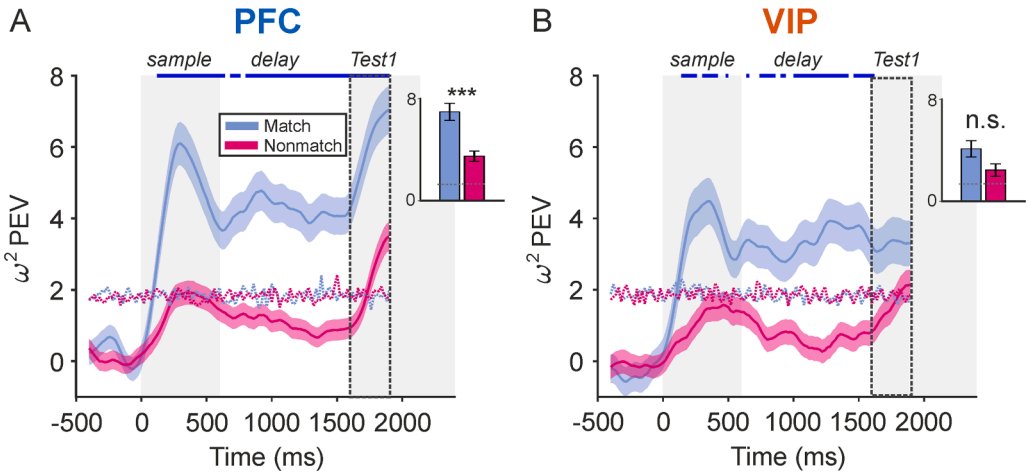


Fig. 5. Sliding-window ω^2 PEV from sample to Test1 period quantifying the encoding of match (same as sample) and nonmatch (different from sample) numerosity. Sliding-window ω^2 PEV was computed for all neurons recorded with at least 18 correct trials per sample numerosity and 3 correct trials per test numerosity in each test phase. (A) PFC neurons (n = 369), (B) VIP neurons (n = 254). Bars above the curves represent time bins where match information exceeded nonmatch information (blue: $p < 0.05$). *Insets*: Mean ω^2 PEV across the Test1 period for match and nonmatch conditions (derived from inside dashed rectangle in the main figure). Dotted lines indicate the significance threshold ($p < 0.01$, one-sided) from shuffled trial labels. Error bars and shaded regions indicate SEM across neurons; ***$p < 0.001$; n.s., not significant.

accuracy dropped in both brain areas for the subsequent Test1 periods. Averaged decoding accuracy in the match condition dropped to 42 % (± 0.81) in PFC, and 37 % (± 0.64) in VIP. Still, classifier accuracy in the match condition remained significantly higher for PFC neurons compared to VIP neurons (permutation test, $p < 0.001$, with alpha correction)

Important differences in decoding emerged for match and nonmatch conditions within each brain area. In PFC, the decoding accuracy for the match period (42 ± 0.81 %) was significantly higher than in the nonmatch period (35 ± 0.64 %) (permutation test, $p < 0.001$, with alpha correction). This was not the case for VIP with a similarly low decoding performance for match and nonmatch conditions (match: 37 ± 0.64 %, nonmatch: 38 ± 0.60 %, permutation test, $p = 0.283$, with alpha correction). A second classifier model, trained and tested with nearly double the number of trials, yielded qualitatively identical results (see Fig. S6). Overall, decoding performance of the classifier based on an equal number of numerosity-selective neurons shows that the PFC represents the sought after matching numerosity better than the nonmatching numerosity, and overall better than VIP neurons which did not differentiate between match and nonmatch numerosities.

Furthermore, we calculated the average performance of the classifier as a function of distance from the real numerosity and used, for each iteration, the sigma-value derived from Gauss fits as a measure of tuning width analogous to the population-tuning curves. Within both brain areas, the tuning functions during the sample period were significantly narrower compared to match and nonmatch conditions, respectively (sample PFC: median $\sigma = 1.01$, sample VIP: median $\sigma = 1.27$, $p < 0.001$). Importantly, and similar to the previous results with averaged tuning curves, the PFC classifiers' tuning functions in the nonmatch condition were significantly wider than those calculated for the match situation (Fig. 6B, median match: $\sigma = 1.86$, median nonmatch: $\sigma = 2.68$, Wilcoxon rank sum test, $Z = -7.2556$, $p < 0.001$, two-sided, with Bonferroni correction, $n = 100$). This was not the case for tuning functions of the VIP (Fig. 6C) for which the width in match and nonmatch conditions were not significantly different (median match: $\sigma = 2.50$, median nonmatch: $\sigma = 2.26$, Wilcoxon rank sum test, $Z = 2.0732$, $p = 0.1145$,

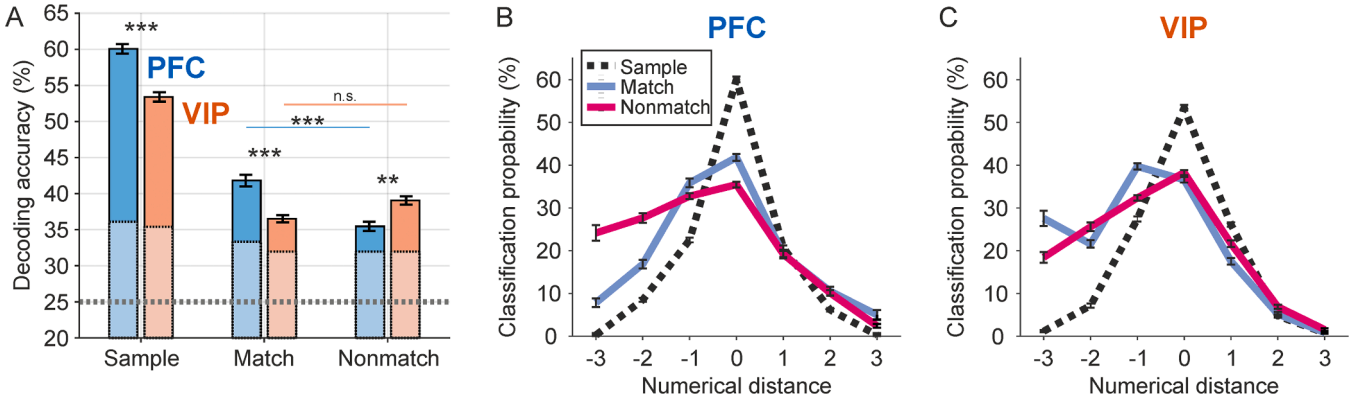


Fig. 6. Decoding numerosity with a SVM classifier from numerosity-selective neurons in the sample and match and nonmatch Test1 periods. (A) Mean decoding accuracy for sample, match, and nonmatch phases. The classifier was trained on sample activity and then tested on sample activity (odd versus event trials cross validation), match activity in Test1, and nonmatch activity in Test1. PFC and VIP cell counts were balanced by creating two pseudo populations of equal neuron numbers. Grey dashed line indicates chance level at 25 %. Dashed lines within the bars depicts significance threshold ($p < 0.05$, one-sided). Error bars show the SEM across 100 resamples. ***$p < 0.001$; **$p < 0.01$; n.s. not significant. (B) Normalized and averaged numerosity tuning curves derived from classifier performance on PFC neurons (N = 34) during the sample, match, and nonmatch phases. Classification probability for each resample is aligned to the preferred numerosity at center value 0 and plotted as a function of numerical distance. Error bars show the SEM across 100 resamples. (C) Normalized and averaged numerosity tuning curves derived from classifier performance on VIP neurons (N = 34) during the sample, match, and nonmatch phases. Layout as in (B).

two-sided, with Bonferroni correction, n = 100).

2.6. Tuning stability to matching numerosity across time and test periods

After comparing tuning to match and nonmatch, we switched to investigating tuning stability across time. To that aim, we explored tuning of single numerosity-selective neurons in PFC and VIP to matching numerosities along the three succeeding test periods (Test1 to Test3). Fig. 7A-D shows the detailed discharges and the tuning functions (insets) of an example PFC neuron tuned to sample numerosity 4 across stimulus repetitions. This PFC neuron's tuning remained stable for all three succeeding test phases until the matching numerosity was presented. Similar stability in tuning was seen for the majority of PFC neurons. In contrast, such a robust encoding was rarely observed in VIP neurons. Fig. 7E-F depicts an example VIP neuron tuned to numerosity 1. This neuron showed a deteriorated tuning with the presentation of more test numerosities and had lost tuning during test periods (Fig. 7G, H). Inspection of single neuron activity indicated that the PFC robustly encodes the target numerosity across repetitions, whereas VIP deteriorated with repetitions. This suggests that VIP neurons are primarily involved during the early stages of cortical processing, while PFC neurons maintain the target number across temporal delays during delayed searches.

To explore tuning stability as a function of test period repetition across neurons, we again computed normalized tuning curves (see description above) and plotted them aligned to the preferred numerosities of each neuron.

In PFC, population-tuning curves remained stable and showed rather symmetric peak curves in all three test periods following the sample presentation (Fig. 8A). Compared to the reference sample period, the preferred numerosity in PFC remained the same across the three subsequent test periods ($peak_{sample} = 1.00 \pm 0.00$, $peak_{test1} = 0.88 \pm 0.20$, $peak_{test2} = 0.79 \pm 0.14$, $peak_{test3} = 0.80 \pm 0.13$, permutation test, Sample/Test1: $p = 0.6795$, Sample/Test2: $p = 0.2957$, Sample/Test3: $p = 0.2672$, $n = 68$). Furthermore, we compared again the tuning width of the individual tuning curves. We found no difference between the sigma-values derived from the three test periods (median Test1: $\sigma = 1.45$, median Test2: $\sigma = 1.84$, median Test3: $\sigma = 1.79$, Wilcoxon signed rank test, Test1/Test2: $Z = -1.7842$, $p = 0.2975$, Test1/Test3: $Z = -1.9981$, $p = 0.1828$, Test2/Test3: $Z = -0.6416$, $p = 1$, with Bonferroni correction, $n = 68$).

In VIP, a different picture emerged (Fig. 8B). While VIP population tuning curves were still comparable for the second stimulus presentation (Test1) ($peak_{sample} = 1.00 \pm 0.00$, $peak_{test1} = 0.80 \pm 0.18$, permutation test, $p = 0.2971$, $n = 34$), both the amplitude and the shape of the tuning functions deteriorated with more stimulus repetitions. For the

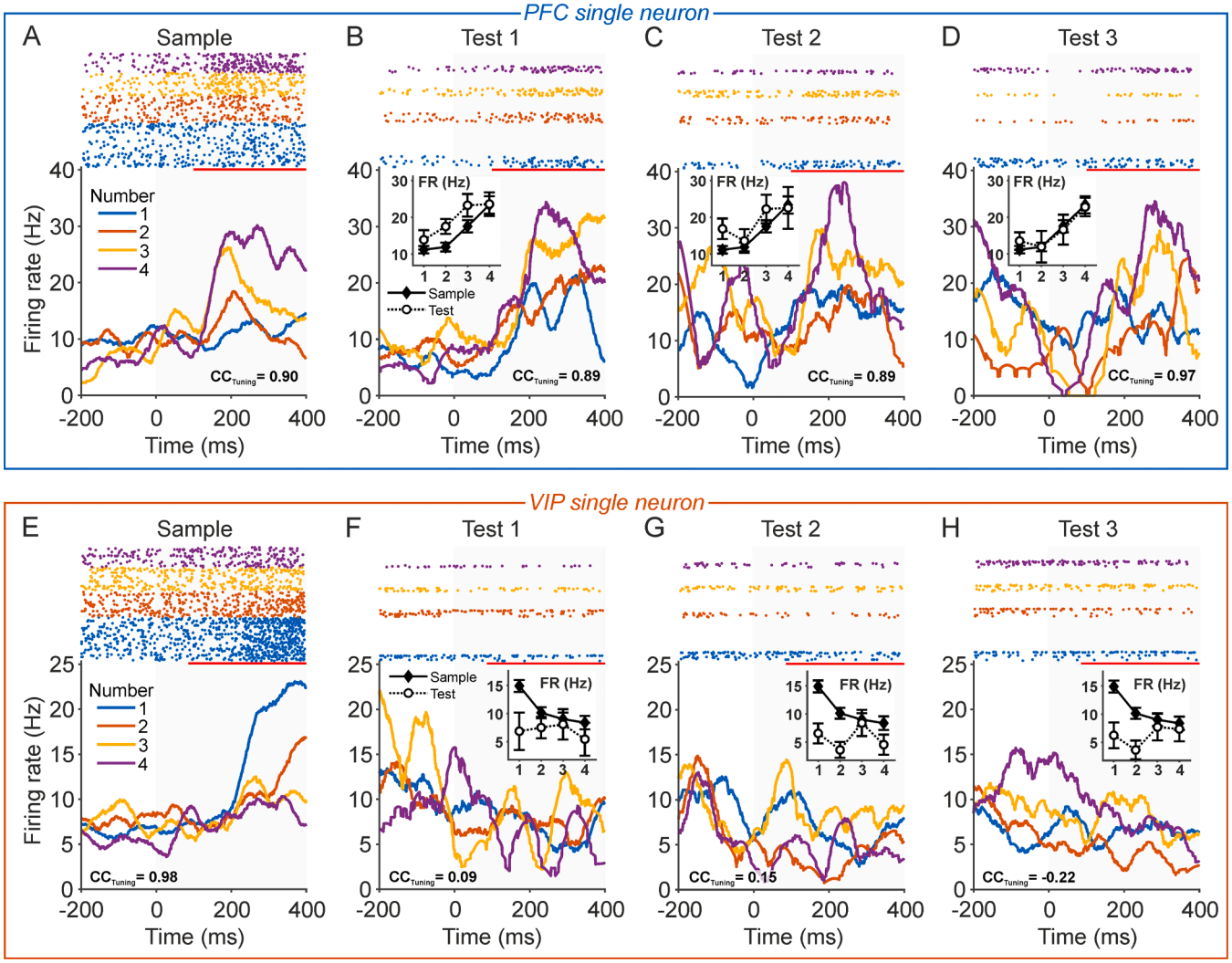


Fig. 7. Numerosity selectivity of single neurons during the sample and Test1–3 match periods. (A-D) Responses of an exemplary PFC neuron preferring numerosity 4 in the sample period (A) to match numerosities in Test1 (B), Test2 (C), and Test3 (D) periods. Same layout as in Fig. 3a-c. (E-H) Responses of an exemplary VIP neuron preferring numerosity 1 in the sample period (E) to match numerosities in Test1 (F), Test2 (G), and Test3 (H) periods. Same layout as in Fig. 3E-G.

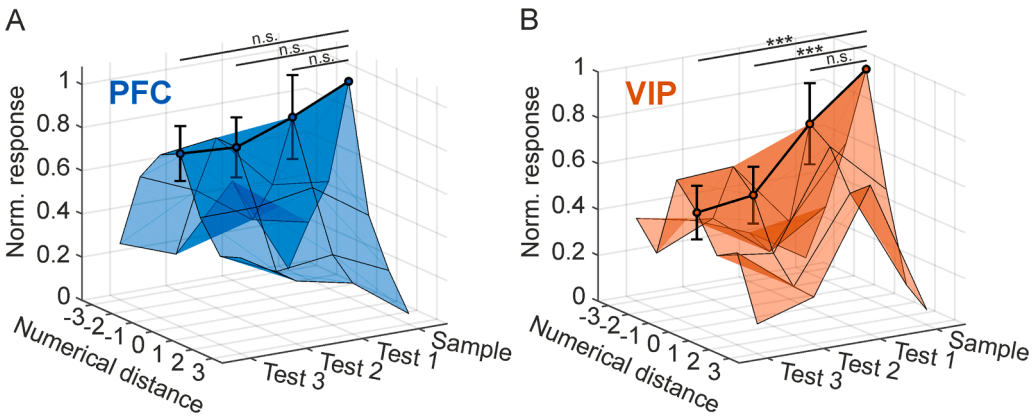


Fig. 8. Time course of tuning in numerosity-selective neurons across the sample and the three successive Test periods. (A) Normalized and averaged numerosity tuning curves of all numerosity-selective neurons in PFC (N = 68) during the sample, Test1, Test2, and Test3 phases mapped as a 3D-surface plot. Prior to averaging, neuronal activity for each neuron was aligned to the preferred numerosity at center value 0 and plotted as a function of numerical distance. Tuning remained stable across Test periods. Error bars indicate SEM across neurons. n.s. not significant. (B) Normalized and averaged numerosity tuning curves of all numerosity-selective neurons in VIP (N = 34) during the sample, Test1, Test2, and Test3 phases mapped as a 3D-surface plot. Same layout as in (A). Tuning deteriorated across Test periods. Error bars indicate SEM across neurons. ***p < 0.001; n.s. not significant.

Test2 and Test3 periods, the VIP population-tuning curves were attenuated, asymmetric, and the peaks were off-center. The normalized peak values in Test2 and Test3 were significantly different to the peak values of the sample period ($peak_{sample} = 1.00 \pm 0.00$, $peak_{test2} = 0.53 \pm 0.13$, $peak_{test3} = 0.50 \pm 0.12$, permutation test, Sample/Test1: $p < 0.001$, Sample/Test2: $p < 0.001$, $n = 34$). These results support the initial finding that PFC neurons maintain their tuning across multiple match presentations, whereas the numerosity tuning of VIP neurons degrades.

Next, we extended the index calculations introduced during the Test1 period (Fig. 3E,J) to the Test2 and Test3 periods (Fig. 9). Besides comparing the neuronal response between the match and nonmatch of the Test1 period, we investigated also how the indices are distributed when using the neuronal response during the other match periods (match Test2 vs. nonmatch Test1; match Test3 vs. nonmatch Test1).

In PFC, we found that real indices to all Test periods (median Test1: 0.26; Test2: 0.37; Test3: 0.30) were significantly higher compared to the shuffled chance values (0.17, 0.17, 0.18, respectively) (match Test1 vs. nonmatch Test1: $Z = 0.9567$, $p = 0.0209$; match Test2 vs. nonmatch Test1: $Z = 3.6973$, $p < 0.001$; match Test3 vs. nonmatch Test1: $Z = 2.6333$, $p = 0.0085$; Wilcoxon signed rank test, $n = 68$) (Fig. 9A,C,E).

In VIP, this was not the case. The real indices to all Test periods (median Test1: 0.18; Test2: 0.14; Test3: 0.18) were not significantly higher compared to the shuffled chance values (0.15, 0.14, 0.15, respectively) (match Test1 vs. nonmatch Test1: $Z = 0.9935$, $p = 0.3387$; match Test2 vs. nonmatch Test1: $Z = 0.5213$, $p = 0.6022$; match Test3 vs. nonmatch Test1: $Z = 1.5087$, $p = 0.1314$; Wilcoxon signed rank test, $n = 34$) (Fig. 9B,D,F).

Furthermore, the indices were significantly higher in PFC compared to VIP for the first two Test periods (match Test1 vs. nonmatch Test1:

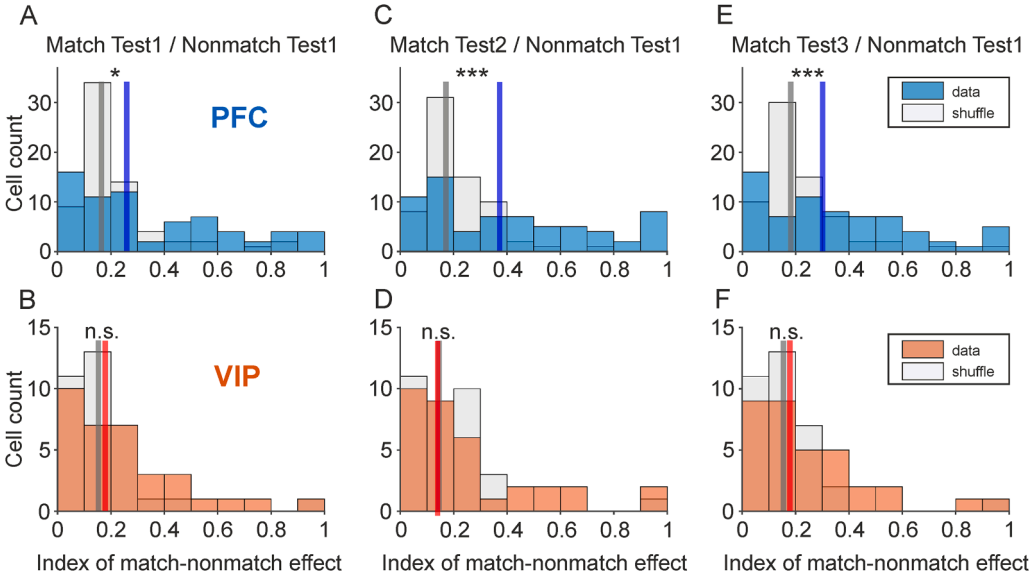


Fig. 9. Indices quantifying the differences between responses to Test1–3 match and Test1 nonmatch numerosities. (A,C,E) Distribution of indices showing the strength of the match-nonmatch effects across the three Test periods relative to the nonmatch responses in the Test1 period in PFC (N = 68). The distribution of indices is shown for the Test1 (A, same as in Fig. 3*E*), Test2 (c), and Test3 periods (E). The median of the real data indicated by the colored vertical line in each Test period was tested against a shuffled distribution indicated by a black line. ***p < 0.001; **p < 0.01; *p < 0.05; n.s. not significant. (B,D,F) Distribution of indices showing the strength of the match-nonmatch effects across the three Test periods relative to the nonmatch responses in the Test1 period in VIP (N = 34). The distribution of indices is shown for the Test1 (B, same as in Fig. 3*J*), Test2 (D), and Test3 periods (F). Same layout as in (A,C,E).

Z = 1.6824, p = 0.0462; match Test2 vs. nonmatch Test1: Z = 2.7652, p = 0.0028), and tended to be larger in the last Test period (match Test3 vs. nonmatch Test1: Z = 1.5335, p = 0.0626; Wilcoxon rank sum test). Thus, in PFC, higher tuning indices indicated more pronounced differences between match and nonmatch across all three test periods, an effect that was absent in VIP.

We compared the tuning properties in more detail by cross-correlating the entire tuning curves of the sample period (as a reference) with those of the subsequent test periods to derive the cross-correlation coefficients (CC_{Tuning}) (Diester and Nieder, 2007, see Material and Methods). The higher the coefficients, the more similar are the tuning functions in sample and in the respective test phases, i.e. a CC_{Tuning} of 1 would indicate identical tuning profiles, whereas a CC_{Tuning} of −1 would signify tuning functions exactly inverted relative to each other. As ground truth, we also determined the CC_{Tuning} for the sample itself by comparing the tuning functions derived from odd trial numbers with that derived from even trial numbers.

Example data can be seen for individual neurons in Fig. 7; here, the example PFC neuron shows a stable tuning across the test phases (Fig. 7A-D). The tuning functions are highly similar (i.e., overlapping), resulting in a high CC_{Tuning} near 1. In contrast, the example VIP neuron shows deteriorated tuning with increasing test phase repetitions, resulting in close-to-zero CC_{Tuning} (Fig. 7E-H).

In PFC, the CC_{Tuning} of all numerosity-selective neurons in all three test periods were significantly higher than zero ($median_{test1}$ = 0.7091, $median_{test2}$ = 0.5518, $median_{test3}$ = 0.5285; bootstrapping the medians, 95th confidence intervals Test1 [0.38, 0.82], Test2 [0.27, 0.70], Test3 [0.30, 0.74]) (Fig. 10B-D). This again suggests that the initial sample tuning remains stable over the three test phases.

In VIP, in contrast, the CC_{Tuning} in numerosity-selective VIP neurons showed overall smaller coefficient values in the test periods ($median_{test1}$ = 0.5602, $median_{test2}$ = 0.2846, $median_{test3}$ = 0.3310, Fig. 10F-H). The median CC_{Tuning} for the Test1 and Test2 periods were still significantly higher than zero (bootstrapping the medians, 95th confidence intervals Test1 [0.13, 0.74], Test2 [0.01, 0.67]), whereas the median for the Test3 period were not significantly different from zero (Fig. 10H; 95th confidence intervals Test3 [-0.09, 0.67]). Thus, the neurons entire tuning functions were more robustly encoded across the test periods in PFC neurons than in VIP neurons.

2.7. Population decoding of matching numerosity across time and test periods

Finally, we applied a population decoding analysis as previously used for the match-nonmatch comparison using a support vector machine (SVM) classifier. We trained the classifier on the firing rates of (area-balanced) numerosity-selective neurons during the sample period and tested the trained classifier during the Test1–3 periods again only for match conditions. Fig. 11 shows the decoding accuracy for the sample period and the three test periods for both brain areas. As expected, accuracy dropped in both brain areas for the subsequent test periods, but interesting differences between PFC and VIP emerged.

In PFC, average decoding accuracy dropped from 60 % (± 0.66) during sample presentation to 42 % (± 0.81), 43 % (± 0.70), and 40 % (± 0.73) during Test1, Test2, and Test3, respectively (Fig. 11A). In all cases, the classifier performances were significantly higher than chance level (larger than the 95th percentile of the shuffled distribution [33.33 %, 31.94 %, 33.33 %]) and stable across the test periods. The decoding accuracies in PFC between the Test1 and Test2, as well as between Test1 and Test3 were not significantly different (permutation test, Test1/Test2: p = 1, Test1/Test3: p = 0.396, with alpha correction), whereas the decoding success for Test2 period was slightly better than Test3 (permutation test, p = 0.01, with alpha correction).

In VIP, average decoding accuracy decreased from 53 % (± 0.66) during sample presentation to 37 % (± 0.70), 37 % (± 0.71) and 32 % (± 0.64) during Test1, Test2, and Test3, respectively (Fig. 11A). While accuracy was still above chance during Test1 and Test2 (larger than the 95th percentile of the shuffled distribution [31.94 %, 33.33 %]), accuracy dropped to chance level for Test3 (smaller than the 95th percentile of the shuffled distribution [34.03 %]). Decoding performance in VIP for Test1 and Test2 were comparable (permutation test, p = 1, with alpha correction) but relative to those dropped for the Test3 period (permutation test, Test1/Test3: p < 0.001, Test2/Test3: p < 0.001, with alpha correction).

When comparing PFC and VIP, we found that decoding accuracy in PFC was significantly higher in all three test periods (permutation test,

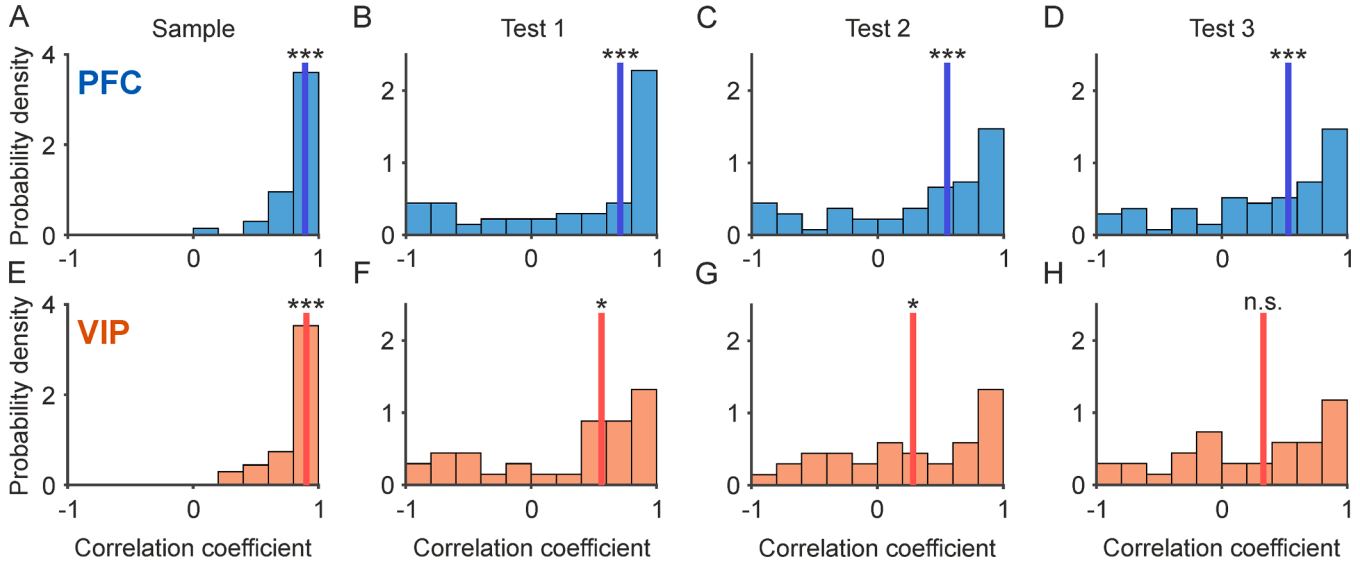


Fig. 10. Comparison of numerosity tuning based on tuning curve cross-correlations. (A-D) Probability density functions of cross-correlation coefficients (CC_{Tuning}) calculated for comparisons of tuning curves of numerosity selective PFC neurons (N = 68) during the sample (cross validation of odd versus even trials) (A), match Test1 (sample versus match Test1) (B), match Test2 (sample versus match Test2) (C), and match Test3 period (sample versus match Test3) (D). Note different y-axes for sample and test period. The medians (indicated by the vertical lines) were tested against zero (no tuning). ***p < 0.001; *p < 0.05; n.s. not significant. (E-H) Probability density functions of cross-correlation coefficients (CC_{Tuning}) calculated for numerosity selective VIP neurons (N = 34) across Test periods. Same layout as in (A-D).

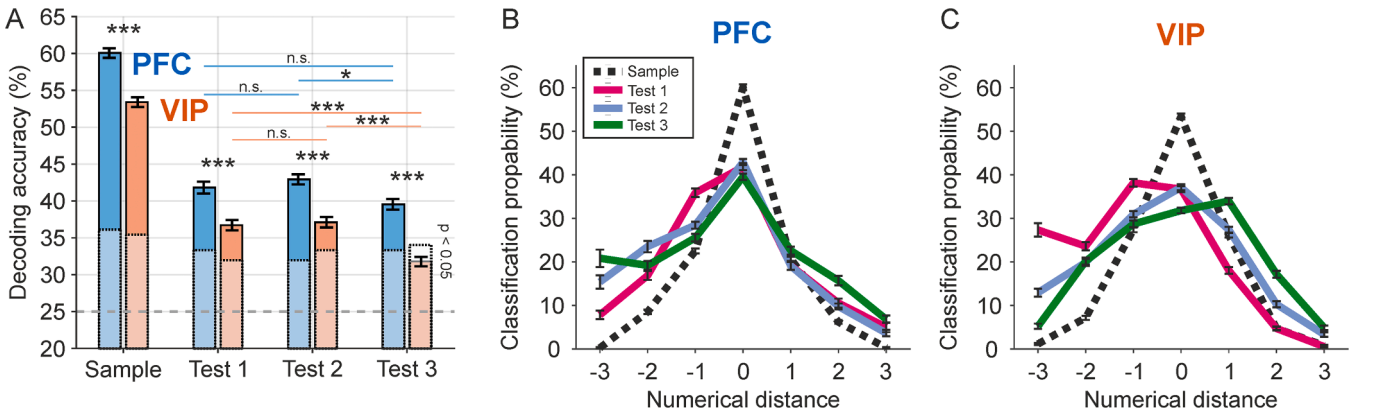


Fig. 11. Decoding numerosity with a SVM classifier from numerosity-selective neurons across match Test1 to Test3 periods. (A) Mean decoding accuracy for sample phase, and match Test1 to Test3 phases. The classifier was trained on sample activity and then tested on sample activity (odd versus event trials cross validation), and match activity in Test1, Test2 and Test3 phases. PFC and VIP cell counts were balanced by creating two pseudo populations of equal neuron numbers. In PFC, decoding accuracy remains stable and significantly better than in VIP throughout the Test periods. In VIP, numerosity decoding was no longer significant in the Test3 period. Layout as in Fig. 5A. (B) Normalized and averaged numerosity tuning curves derived from classifier performance on PFC neurons (N = 34) during the sample and match conditions of Test1, Test2, and Test3 phases. Classification probability for each resample is aligned to the preferred numerosity at center value 0 and plotted as a function of numerical distance. Error bars show the SEM across 100 resamples. Layout as in Fig. 5B. (C) Tuning functions from classifier performance on VIP neurons. Layout as in Fig. 5C.

for all test periods: $p < 0.001$, with alpha correction).

Finally, we derived tuning curves from the classifier performance, plotted them as a function of numerical distance aligned to the preferred numerosity, fitted Gauss functions to the tuning curves, and compared the sigma-values as a measure of tuning width (Fig. 11B,C). During the sample, Test1 and Test2 period, VIP showed significantly wider tuning functions compared to PFC (median sample PFC: $\sigma = 1.01$, median sample VIP: $\sigma = 1.27$, Wilcoxon rank sum test, $Z = -6.8171$, $p < 0.001$; median Test1 PFC: $\sigma = 1.86$, median Test1 VIP: $\sigma = 2.50$, Wilcoxon rank sum test, $Z = -5.5893$, $p < 0.001$; median Test2 PFC: $\sigma = 1.96$, median Test2 VIP: $\sigma = 2.26$, Wilcoxon rank sum test, $Z = -2.6987$, $p = 0.007$), with a tendency for sharper PFC tuning in the Test3 period (median Test3 PFC: $\sigma = 2.34$, median Test3 VIP: $\sigma = 2.58$, Wilcoxon rank sum test, $Z = -1.7702$, $p = 0.0767$)

3. Discussion

Previous studies have identified neural signatures for purely sensory recognition memory in a variety of high-level cortical areas in primates, such as the inferior temporal cortex (Baylis and Rolls, 1987; Riches et al., 1991; Eskandar et al., 1992; Miller et al., 1991, 1993; Li et al., 1993; Miller and Desimone, 1993, 1994; Vogels et al., 1995; Anderson et al., 2008; Woloszyn and Sheinberg; 2012; Meyer and Rust, 2018), the medial temporal lobe areas (Riches et al., 1991; Fahy et al., 1993), the posterior parietal cortex (Constantinidis and Steinmetz, 2001; Rawley and Constantinidis, 2010), and the lateral prefrontal cortex (Miller et al., 1996; Rainer and Miller, 2000; Freedman et al., 2003; Ng et al., 2014; Parto Dezfouli et al., 2021). In these regions, many neurons respond differently to test stimuli based on whether they match the sample; in other words, whether the monkey is memorizing the sample and attentively looking for it across time.

In the current study, we compared and contrasted the neuronal mechanisms of recognition memory involved in match versus nonmatch computation of abstract numerical quantities in monkeys. To perform the delayed match-to-numerosity task with variable frequencies of non-matching numerosities, the monkey had to retain a memory of the sample quantity and determine whether one of a variable number of test stimuli matches it. Within an ecological context, animals benefit from such a search image for numerosities, allowing them to identify cued numerical quantities — such as food, friends, or foes (Ishii and Shimada, 2010) — within a stream of numerical values (Nieder, 2020). Using analyses of single-neuron tuning and population decoding approaches, we identified aspects of recognition memory for quantities in the lateral PFC and the VIP within the intraparietal sulcus. The observed PFC effects are unlikely to be related to long-term memory, as the monkeys evaluated different numerosities on a trial-by-trial basis in the current DMS task. Instead, the sustained numerosity-selective activity in PFC likely reflects its role in working memory, consistent with evidence that PFC neurons maintain information over short delays to guide behavior in tasks requiring temporary numerical storage and manipulation. Below, we discuss the prominent differences between these two brain areas, which are known to form the primate core number network (Nieder and Miller, 2004; Nieder and Dehaene, 2009; Nieder, 2016).

In the PFC, we observed only mild modification of neuronal activity when the same preferred numerosity was repeated. Specifically, there was no degradation in tuning selectivity when the same sample numerosity was presented again as a matching Test stimulus. This indicates that PFC neurons maintained their numerosity selectivity between these trial phases. This held true even when the presentation of the target numerosity was interrupted and delayed by one or more non-matching numerosities. However, decoding accuracy for the population of selective PFC neurons dropped during the subsequent matching Test period, indicating some loss of selectivity with time. However, despite this initial drop, accuracy remained significantly and constantly above chance even when the target numerosity presentation was interrupted and postponed by non-matching numerosities. The observed reduction in neuronal activity when the sample stimulus is immediately repeated as the matching test stimulus is consistent with the phenomenon known as repetition suppression, which has been well documented for visual stimuli (Miller et al., 1991, 1993; Miller and Desimone, 1994; Steinmetz and Constantinidis, 1995; Constantinidis and Steinmetz, 2001). Our findings demonstrate that this effect also applies to abstract quantitative categories, such as the number of dots in a set.

Unlike the stable tuning to sought-after target numerosities, the tuning selectivity of PFC neurons specifically decreased when the same numerosities were presented as non-matches during the test periods. This decline was evident in both tuning selectivity and population decoding accuracy. The decoding analysis shows that only PFC neurons discriminate match from nonmatch numerosities, highlighting the PFC's sensitivity to behavioral relevance (Viswanathan and Nieder, 2015) and its role in distinguishing task-relevant quantities. When numerosities are being searched for as matches, PFC neurons respond with high selectivity and are finely tuned to them. However, tuning and decoding accuracy significantly decay when the same numerosities are not the ones

memorized and attended to by the monkey.

In contrast to PFC neurons, VIP neurons displayed three notable distinctions when monkeys were seeking a matching stimulus within a sequence of numerosities. First, VIP neurons exhibited an overall diminished responsiveness to numerosities across all trial periods, evident from the lower accuracy in population decoding analysis.

Second, the tuning selectivity of VIP neurons notably waned with repetitions of numerosities, irrespective of whether they were presented as matching or non-matching test stimuli. This indicates a lack of distinctions between match and non-match numerosities in VIP neurons. Consequently, the robust match-versus-nonmatch effect observed in PFC, as indicated by the match-versus-nonmatch index, was notably absent in VIP. VIP neurons seem to possess limited capabilities in discerning sought-after versus irrelevant numerosities.

Third, we observed differences in the mechanisms of how PFC and VIP neurons responded to repeated numerosities. In our study with numerical categories, significant match enhancement effects were found only in the PFC. Match enhancement is the effect of increased neuronal activity when a stimulus matches a previously encountered stimulus that has been stored in memory, an active mechanism that engages in favor of sought-after stimuli, ensuring that repetitive, irrelevant non-match stimuli are ignored (Engel and Wang, 2011). The presence of match enhancement effects exclusively in the PFC, alongside match suppression effects in both the VIP and PFC, supports the notion that the PFC plays a specialized role in enhancing the recognition of familiar numerosities. Match enhancement has been individually observed in the prefrontal (Miller et al., 1996; Freedman et al., 2003; Qi et al., 2012; Parto Dezfouli et al., 2021), temporal (Miller and Desimone, 1994), and parietal (Rawley and Constantinidis, 2010) cortices. Conversely, significant match suppression effects occurred in both the PFC and VIP, with similar magnitudes in these two brain areas. This reduction may help select PFC and VIP neurons filter out familiar numerosities, allowing them to focus on novel stimuli. Together, match enhancement and match suppression create a balance that enables PFC neurons, in particular, to efficiently process, store, and respond to numerical quantity by enhancing recognition of familiar numerosities while suppressing redundancy.

Overall, VIP activity lacks the capacity to discern and encode sought-after versus irrelevant numerosities and appears unable to track the numerosity the monkey is searching for. In contrast, PFC signals distinguish target versus irrelevant numerosities, serving as a signature for recognition memory. The observed stability of the tuning of PFC neurons during match conditions, in contrast to the overall less stable tuning of VIP neurons, highlights the specialized role of the PFC in maintaining precise numerical representations. The observed neuronal response latencies between these regions (see also Viswanathan and Nieder, 2013, 2015), and corroborating previous findings on earlier numerosity selectivity in VIP (Nieder and Miller, 2004) suggests that the VIP functions upstream of the PFC. This temporal sequencing underscores the hierarchical processing dynamics where VIP inputs precede and likely influence the activity within the PFC, thereby suggesting a feedforward mechanism in the neural circuitry. This aligns with established neuroanatomical pathways and functional connectivity patterns (Chafee and Goldman-Rakic, 2000; Crowe et al., 2013), reinforcing the role of VIP in preliminary numerical integration before the executive processing in the PFC. Overall, the data suggest that VIP neurons handle the initial encoding of numerosity in early cortical processing, routing numerical information to PFC neurons, which maintain the target number across temporal delays for recognition in delayed searches.

Conversely, the PFC is more involved in executive functioning with categories (Roy et al., 2010; Antzoulatos and Miller, 2011). This is also evident in the current study on number categories, in which monkeys had to indicate whether two sequentially presented numerosities were the same or different, and numerosity tuning for targets was stable across time whereas it deteriorated for non-matches. This correlates with a previous finding in a decision task in which PFC neurons initially represented numerical values with high precision before the neuronal codes became transformed during the course of the abstract decision process (Viswanathan et al., 2024). Together, these observations suggest that the PFC may function as a search image for numbers, enabling primates to swiftly and accurately identify specific numerical quantities within a sequence of numbers (Ishii and Shimada, 2010).

The neuronal patterns observed in the current study in PFC and VIP contrast with those from a previous numerosity distractor task (Jacob and Nieder, 2014; Jacob et al., 2018). In the classic distractor task, an irrelevant stimulus is passively viewed at a predefined trial period. However, in the current study, monkeys were presented with a variable count of sequential test displays, requiring active evaluation of each numerosity presentation to determine if it matched the number of dots. The absence of distractors in our study is a strength, as it isolates the core processes involved without the additional complexity of attention-related factors. This difference in task design may explain why PFC neuron tuning was affected by a passively viewed distractor (Jacob and Nieder, 2014; Jacob et al., 2016; Lin et al., 2023), whereas it remained stable in the current study where numerosity presentations were unpredictable. Conversely, VIP neuron tuning was unaffected by a passive distractor (Jacob and Nieder, 2014), but in the current task, VIP neurons appear unable to track the numerosity that the monkey is actively searching for. This suggests the sequential search task engages different cognitive processes and memory systems, emphasizing continuous comparison and evaluation of each display rather than suppressing an interfering stimulus.

The fronto-parietal network plays a crucial role in recognizing auditory numerical information. Both macaques and humans represent the number of tones in intraparietal and dorsal premotor areas, while tone repetition patterns are processed in ventral prefrontal cortex and basal ganglia (Wang et al., 2015). Macaques can detect first-order sequence violations in auditory cortex and second-order violations in a frontoparietal network, indicating a hierarchical predictive coding system similar to humans (Uhrig et al., 2014). These neural mechanisms are essential for recognition memory, enabling the encoding, maintenance, and retrieval of auditory information. The intraparietal and dorsal premotor areas contribute to the initial representation and maintenance of numerical and auditory information, while the ventral prefrontal cortex and basal ganglia are involved in processing and updating tone repetition patterns. The ability to detect sequence violations reflects the brain's capacity to predict and recognize auditory patterns, supporting the function of the recognition memory system.

Our study confirms that the IPS, particularly areas like LIP and MIP, play key roles in perceptual decision-making. In studies by Freedman and co-workers with monkeys performing a delayed match-to-category task involving visual motion direction, LIP neurons encoded category information with greater strength and faster latency than PFC neurons (Freedman and Assad, 2006; Swaminathan et al., 2012). While PFC primarily guided match/non-match decisions, LIP and MIP were more involved in stimulus evaluation and motor planning (Zhou et al., 2021). Pharmacological inactivation studies confirmed LIP's causal, task-independent role in memory-based visual categorical decisions, regardless of task structure or response modality (Zhou et al., 2023).

While these studies confirm the involvement of IPS regions in categorization and decision-making, they have not tested explicit recognition memory with multiple sequential choices, as examined in the current study. Additionally, the role of VIP, a key focus of the current work, has not been explored in motion direction categorization. Another crucial difference lies in the stimuli: Freedman and colleagues used arbitrary, learned motion direction categories, while we employed empirically ordered categories—the number of items in a set—aligned along a "number line." Despite these differences, both approaches highlight the importance of the IPS within the fronto-parietal categorization network.

Research on neuronal responses in the monkey ventral intraparietal

area (VIP) suggests a complex relationship with motor behavior. VIP neurons show sensitivity to visual motion and heading direction, with responses modulated by attention and eye movements (Cook and Maunsell, 2002; Zhang and Britten, 2010, 2011). These findings collectively suggest that VIP neuronal responses are influenced by motor behavior and task context.

The macaque ventral intraparietal area (VIP) is involved in various sensorimotor and cognitive functions, including motion processing, multisensory integration, head peripersonal space processing, defensive behavior, and numerosity coding (Foster et al., 2022). In humans, VIP has evolved into a "VIP complex" comprising three distinct parietal areas with specialized functions related to self-representation and spatial processing. A proposed unifying functional principle is "prediction in space and time," linking VIP to state estimation, potentially including numerical recognition memory, as a key parietal sensorimotor function (Foster et al., 2022).

In our proposed hierarchical model, VIP neurons are responsible for the initial encoding of numerosity, while PFC neurons maintain target numbers across temporal delays for recognition. To further enrich this framework, we propose incorporating the hippocampus and MTL structures as critical components in binding numerical information with contextual details, thereby facilitating the formation and retrieval of robust numerical memories. The hippocampus supports recognition memory by retrieving studied details, while the MTL cortex provides familiarity signals (Norman and O'Reilly, 2003). Both the MTL and PFC are crucial for visual short-term memory of numerical information (Kutter et al., 2018), with their interactions underlying working memory mechanisms (Wu and Buckley, 2022), including arithmetic memory (Menon, 2016). This integration is consistent with established neuroanatomical pathways and functional connectivity patterns, underscoring the collaborative roles of these regions in numerosity processing and recognition memory.

Besides the fronto-parietal number network, medial temporal lobe (MTL) areas, including the hippocampus, are known to support sensory recognition memory (Riches et al., 1991; Fahy et al., 1993; Brown and Aggleton, 2001). In humans, MTL neurons also encode numerical values (Kutter et al., 2018, 2023, 2024) and working memory-related arithmetic rules (Kutter et al., 2022). Although traditionally linked to long-term memory, hippocampal neurons can show sustained, feature-selective delay activity predicting successful working memory retrieval (Kornblith et al., 2017; Kamiński et al., 2017). While not directly assessed here, the hippocampus and MTL likely contribute to a broader network, working alongside the PFC and VIP to support numerical cognition and recognition memory.

While numerous computational models of recognition memory have been proposed (Sohal and Hasselmo, 2000; Norman and O'Reilly, 2003; Bogacz and Brown, 2003; Cortes et al., 2010; Tyulmankov et al., 2022), most of these models treat familiarity discrimination as a binary problem for unrelated stimuli — determining whether an arbitrary stimulus is novel or familiar. However, the situation is more complex for numerical quantities, which are represented in an orderly fashion on a mental number line. The bell-shaped tuning of numerosity-selective neurons and the resulting neuronal distance effect reflect the inherent relatedness between numerical values that are part of recognition memory for quantities. Neuronal network models have successfully simulated the emergence of numerosity tuning (Dehaene and Changeux, 1993; Verguts and Fias, 2004; Nasr et al., 2019; Nasr and Nieder, 2021; Mistry et al., 2023) and the counting of stimulus occurrence frequencies (Dasgupta et al., 2022). However, the neural circuit responsible for storing numerosity recognition memory has yet to be modeled.

The current physiological data are essential for advancing models of the numerosity system, as they help us move beyond the mere extraction of numerical quantity to understanding how the brain uses numerical information for goal-directed tasks, such as recognizing target numbers after a delay. Future studies could explore how task difficulty and training influence the neural correlates of numerosity recognition memory. Additionally, investigating the role of other brain areas, like the MTL regions or anterior inferotemporal cortex, would be a valuable direction for further research.

4. Methods

4.1. Experimental animals and surgery

Two male adult rhesus monkeys (*Macaca mulatta*) were implanted with two recording chambers each, centered over the principal sulcus in the dorsolateral prefrontal cortex (PFC) and the intraparietal sulcus (IPS) in the posterior parietal cortex (monkey 1 and monkey 2, both left hemisphere). Chamber implantation was guided by anatomical magnetic resonance imaging (MRI) and stereotaxic measurements. All surgeries were performed under sterile conditions while the monkeys were under general anesthesia. The monkeys received postoperative antibiotics and analgesics. All procedures were performed in accordance with the guidelines for animal experimentation approved by local authorities, the Regierungspräsidium Tübingen, Germany.

4.2. Experimental apparatus and stimuli

Monkeys were seated inside primate chairs in a chamber 57 cm away from a 15" flat screen monitor (with a resolution of 1024 by 768 pixels and a refresh rate of 60 Hz). The NIMH Cortex software (Bethesda, MD) was used to present the stimuli and behavioral data acquisition. The monkeys' eye gaze was tracked by an infrared eye tracking system (ISCAN, Cambridge, MA).

We used random dot arrays as numerosity stimuli ranging between 1 and 5 dots (diameter range from 0.5 ° to 0.9 ° of visual angle). Each unique dot pattern was presented on a gray background circle subtending 5.4 ° of visual angle and was used as sample and test stimulus. A custom-written MATLAB (Version R2020b, MathWorks Inc., Natick, MA) program generated the stimuli anew before each recording session to prevent the monkeys from memorizing stimulus sequences. Sample and test stimulus of the same countable numerosity were never identical within a trial and session. In order to control for low-level, non-numerical visual features in half of the trials, dot diameters were selected randomly (standard protocol). In the other half of the trials, dot density and total occupied area were equated across stimuli (control protocol). Trials of the standard and control protocol were pseudo-randomly and unpredictably presented.

4.3. Behavioral protocol

We trained two macaque monkeys on a modified version of the delayed match-to-numerosity task in which the monkeys had to match test numerosities to sample numerosities (Fig. 1A). Nonmatching numerosities that were presented between sample and matching test stimulus had to be ignored. Up to three sequential numerical nonmatches were presented, i.e. the task consisted of up to three test phases and in each test phase the right or wrong numerosity could occur. A trial was initiated by the monkeys when they grabbed a bar and maintained eye fixation within a window of 3.5° of visual angle, centered around a central fixation point, i.e., ±1.75° from the center. Trials were immediately aborted (blue screen appears) and excluded from further analysis if the animals broke fixation. At first, a gray circular background was shown for 500 ms (pure fixation period). Subsequently, the sample stimulus containing 1–4 dots appeared for 600 ms. This sample stimulus had to be memorized over a delay of 1000 ms (again gray background circle) and had to compared in a following sequence of test numerosities. Each test stimulus was presented for 800 ms with a 100 ms delay between each other. Since single trials with three test displays lasted nearly 5 s, during which the monkeys were required to maintain eye fixation, extending the delay intervals would have further prolonged the trials and led to a significant increase in fixation breaks, making it impractical

to incorporate longer delays.

For the case that the numerosity in test stimulus was same as in the sample (match situation, 1 – 4), the monkeys had to respond with a bar release. If the numerosities were different (nonmatch situation, 1 – 5), the animals continued to hold the bar. A nonmatch numerosity was never shown twice per trial. Sample numerosities and trial types (i.e. if no nonmatch, one nonmatch, two nonmatch or three nonmatch numerosities were presented) were balanced across conditions and displayed with equal probability (25 % each, pseudo-randomly intermixed). Thus, the monkeys could not predict when or whether the match stimulus would appear. Correct decision per trial was rewarded with a drop of water. In 25 % of the trials, when presenting three nonmatch numerosities in sequence, the monkeys were rewarded when holding the bar throughout the whole trial until display-offset. Chance level of 25 % was achieved by averaging the probabilities for each trial type (for details see Figure S1). Since the number of trials a monkey can perform during an experimental session is limited (on average monkey 1 completed 483 trials and monkey 2 534 trials), not all possible combinations of sample (1 – 4) and test numerosities (1 – 5) could be presented. Therefore, a predefined collection of 64 unique sample-test-combinations was shown.

4.4. Neurophysiological recordings

In each session, arrays of up to eight glass-coated 1 MΩ tungsten microelectrodes (Alpha Omega LTD, Israel) were inserted in each recording chamber using custom-made screw microdrives and a grid (Crist Instruments, USA) with 1 mm spacing. Stable and well isolated neurons were randomly selected; no attempt was made to preselect neurons according to response properties. To access the IPS, electrodes were passed along the course of it to a depth of 9–10 mm (Fig. 2A). A MAP Plexon system was used for signal acquisition, amplification, filtering and digitalization. Waveform separation was performed offline (Plexon Systems, USA).

4.5. Data analysis

All analyses were conducted in MATLAB (version R2021a, MathWorks Inc.) using custom-written scripts. All values in the main text and figures refer to the mean ± standard error of the mean (SEM), unless stated otherwise. SEM was calculated as the standard deviation divided by the square root of the number of samples.

Unless specified otherwise, neurons were included in the below listed analyses if the following criteria were met: First, their average firing rate across trials was at least 0.5 spikes/s; second, they were recorded for at least 18 correct trials per numerosity during the sample period (4 sample numerosities); and third, they were recorded for at least 3 correct trials per numerosity in each of the three test periods where a match numerosity occurred (match Test1 to Test3), as well as in the first test period when a nonmatch numerosity was presented (nonmatch Test1). To have the same numerical range between match and nonmatch situation (1 – 4 match vs. 1 – 4 nonmatch), we excluded all trials that contained nonmatch numerosity 5 in Test1 period. A total of 369 neurons in PFC and 254 neurons in VIP fulfilled these criteria.

4.6. Behavioral data analysis

For each recording sessions behavioral tuning functions were computed to describe the percentage of trials for which a test stimulus was judged as being equal in number to the sample. Mean tuning functions were computed as the average over all individual functions per session. Furthermore, the averaged performance for the four trial types (i.e. no nonmatch, one nonmatch, two nonmatch or three nonmatch test stimuli) according to the four sample numerosities was calculated to see an influence on the behavior. For each correct trial, in which the monkey had to release the bar (75 % of the trials), the reaction time was measured after match stimulus onset.

4.7. Visual response latency analysis

The visual response latency analysis was accomplished for all neurons with an average firing rate of at least 0.5 spike/s across trials and at least 5 correct trials per sample numerosity and stimulus protocol (4 sample numerosities x 2 stimulus protocols). For each cell the firing rate was averaged over all correct trials in 20 ms windows in 1 ms steps. Mean and standard deviation (SD) of baseline activity was measured for the last 200 ms of the presample period plus the first 40 ms of the sample period. The latency analysis was restricted to 40 ms till 300 ms after sample onset and was defined as the first time bin within this window in which the activity was higher or lower than 2 SDs compared to the baseline activity for more than 20 consecutive windows. Latencies estimated by visual inspection coincided with latencies determined in this way in most of the cases. We compared the latency distributions between VIP and PFC with a Wilcoxon rank sum test (two-sided). Furthermore, we calculated the 25th percentile of the respective latency distribution to take this value as the 'main' latency for all subsequent analyses. VIP neurons were analyzed from now on with a latency of 86 ms and PFC neurons with a latency of 100 ms after sample or test onset.

4.8. Neuronal selectivity and tuning analysis

Neuronal activity was analyzed separately for the sample period, for the first test period in which a nonmatch was presented and the three subsequent test periods showing the match situation. Since the monkeys were required to respond with a bar release when they detected a match numerosity, we only used correct trials in which the animals responded after 400 ms to ensure that the endpoint of the analysis window was not contaminated by motor activity. This was the case for 84.4 % of all possible match trials in monkey 1 and for 89.1 % in monkey 2. Trials with a shorter reaction time were excluded from further analysis. In the nonmatch situation such limitation was not necessary, since the monkeys had to withhold the bar throughout the nonmatch presentation. Thus, the time window for all analyses started with the 'main' latency of the respective brain area (see above) and ended 400 ms after sample or test onset.

To determine numerosity-selectivity during the sample period, a two-factorial analysis of variance (ANOVA) was performed with the factors numerosity (sample numerosity 1 – 4) and protocol (standard and control). Neurons with a significant main effect for the factor numerosity ($p < 0.01$), but no main effect for the factor protocol or interaction, were classified as exclusively numerosity-selective. This strict significance threshold of $p < 0.01$ was applied only for assessing firing rate differences using the ANOVA to avoid false positive numerosity selective neurons; for all other statistics, the conventional significance threshold of $p < 0.05$ was used. For numerosity-selective neurons, we created neuronal tuning functions for the sample, nonmatch Test1 period and match Test1 to Test3 period within the above described time window by averaging the firing rates across trials. The numerosity eliciting the highest discharge rate was defined as "most-preferred numerosity" of a given cell, whereas the numerosity provoking the lowest firing rate was referred as "least-preferred numerosity". Using this approach, we were also able to compute average and normalized neuronal tuning functions by averaging and normalizing (between 0 % and 100 %) all individual tuning functions with the same preferred sample numerosity.

4.9. Population-tuning analysis

To derive averaged numerosity-filter functions for the sample, for the nonmatch Test1 period and for the match Test1 to Test3 period, the individual tuning functions of each neuron were normalized by the ac-

tivity between its most-preferred and least-preferred numerosity determined during the sample period. Thus, the activity at the preferred numerosity for the sample period is set to 1. Furthermore, this maximum activity was used as the reference to compute the averaged numerosity-filter functions for the other test periods. In other words, the initially defined most-preferred numerosity remains the same across all test periods, even though the preferred numerosity may not necessarily correspond to the maximum activity response. We pooled the resulting normalized tuning curves across such neurons which were defined in the sample period as exclusively numerosity-selective by a two-way ANOVA ($p < 0.01$). The time windows to determine the neuronal tuning functions had equal length in sample and the respective test period ('main' visual response latency till 400 ms after sample or test onset). We compared the peak values of each averaged numerosity-filter function by using a permutation test (normalized firing rates, 10.000 permutations with replacement). Furthermore, we quantified tuning changes by fitting each normalized individual tuning function by a symmetric Gaussian distribution, described by the following formula:

$$f(x) = amp * e^{-\left(\frac{x-xc}{\sigma}\right)^2} + y0$$

To acquire the best fit to the data, σ (standard deviation of the Gaussian distribution), and *amp* (amplitude) were adjustable during the fitting procedure. The peak function's yaxis offset $y0$ was set to zero, and the center of the fitted distributions xc was fixed at the preferred numerosity that was determined during the sample period. The sample period served as the reference for determining each neuron's preferred numerosity, as there were no task demands during this time to influence tuning. Tuning curves from subsequent periods are aligned to this reference, and any deviation from the preferred numerosity in later test periods indicates distorted tuning. We used a Wilcoxon signed rank test (with Bonferroni correction, two-sided) to compare the Gaussian widths (σ) determined for the match and nonmatch situation, as well as for the match situations among themselves.

4.10. Index for tuning strength variations

To quantify tuning strength variations between the match and nonmatch test situation, we calculated for each cell an index for the most preferred numerosity determined during the sample period. The index can be calculated by subtracting the mean neuronal response to the match presentation (FR_{match}) from those to the nonmatch presentation (FR_{nonmatch}) and dividing the absolute value of the difference by the sum of the two means. It can be described by following formula:

$$Index = \frac{(FR_{match} - FR_{\text{nonmatch}})}{(FR_{\text{match}} + FR_{\text{nonmatch}})}$$

Neurons that responded stronger to a numerosity when displayed as a match compared to when displayed as sample showed a match-enhancement effect. Neurons that responded stronger to a numerosity when displayed as a sample compared to when displayed as nonmatch showed a match suppression effect. For a better visualization, we calculated for cells that show a match suppression effect the absolute value, whereby the index can range only between 0 and 1. Values close to 0 indicate equal responses to the two respective test situations.

We used this calculation for two approaches: First, we pooled the absolute indices to see general tuning strength variations between the match and nonmatch situation of the Test1 period (match Test1 vs. nonmatch Test1), but also for comparing the match situation of the Test2 and Test3 period with the nonmatch situation of the Test1 period (match Test2 vs. nonmatch Test1, match Test3 vs. nonmatch Test1). Second, we went more into detail and calculated the indices separately for the match enhancement and match suppression effect of the Test1 and Test2 period (match vs. nonmatch). To see which tuning strength variation (match enhancement or match suppression) can explained by chance, we calculated a shuffled distribution by dissolving the association between the respective test situations (i.e. match or nonmatch) and the trial-by-trial firing rates of each neuron. We repeated this process 10,000 times, calculated the mean of these single values and compared the median of the new distribution with the median of the real data by a Wilcoxon signed rank test (two-sided). Furthermore, we tested with a one-sided Wilcoxon rank sum test if the median PFC index is larger than the median VIP index. For this analysis we considered just neurons, which were in the sample period determined as exclusively numerosity-selective by a two-way ANOVA ($p < 0.01$). Firing rates were derived from time windows with equal length in sample and the respective test period ('main' visual response latency till 400 ms after sample or test onset).

4.11. Tuning curve cross-correlation analysis

Cross-correlations were performed between neuronal tuning functions (Diester and Nieder, 2007) calculated during the sample period and the subsequent three match periods. This kind of analysis provides a measure to quantify the extent to which single neurons changes their initial coding (i.e. representing the sample numerosities) during the presentation of test numerosities. Furthermore, we calculated the cross-correlation for the sample period for itself by comparing the tuning functions derived from odd trial numbers with that evaluated for even trial numbers. We considered just neurons, which were in the sample period determined as exclusively numerosity-selective by a two-way ANOVA ($p < 0.01$). Tuning curves were derived from time windows with equal length in sample and the respective test period ('main' visual response latency till 400 ms after sample or test onset). The cross-correlations were calculated by following formula:

$$\text{CCTuning} = \frac{\sum_{n=1}^{4}\left(FR_{\text{sample}(n)} - \overline{FR}_{\text{sample}}\right) * \left(FR_{\text{test}(n)} - \overline{FR}_{\text{test}}\right)}{\sqrt{\sum_{n=1}^{4}\left(FR_{\text{sample}(n)} - \overline{FR}_{\text{sample}}\right)^2} * \sqrt{\sum_{n=1}^{4}\left(FR_{\text{test}(n)} - FR\text{t}_{\text{es}}FR\text{t}\right)^2}},$$

where $FR_{\text{sample}}(n)$ is the average firing rate for numerosity $n \in \{1,2,3,4\}$ during the sample period, $\overline{FR}_{\text{sample}}$ the average firing rate across all numerosities during the sample period, $FR_{\text{test}}(n)$ the average firing rate for numerosity $n \in \{1,2,3,4\}$ during the test period, and $\overline{FR}_{\text{test}}$ the average firing rate across all numerosities during the test period. Thus, the higher the cross-correlation coefficients, the more stable is the initial encoding across the respective test period. We performed this analysis for VIP and PFC neurons separately, resulting in four different cross-correlation coefficient distributions per brain area (sample and three match periods). Each distribution was tested against zero by bootstrapping the medians 10,000 times with replacement. The significance threshold was set to $p < 0.05$, i.e. the 5th and 95th percentile. For a better comparison between the brain areas, the probability density function was plotted as a histogram.

4.12. SVM classifier

We trained a linear multiclass support vector machine (SVM) classifier (one-versus-one classification, Chang and Lin, 2011) on trial firing rates during the sample period and tested it for the discrimination of numerosities on spiking activity derived from the sample period itself, from the nonmatch Test1 period and from the subsequent three match periods. This resulted in five analyses per brain area. We only used neurons which were determined as exclusively-numerosity selective during the sample period. A six-fold cross-validation was performed, resulting in 15 trials for training and 3 trials for testing per numerosity. Within each cross-validation loop a new training and test subset was drawn. However, through the limitation of trials, it was possible that for the test subset derived from the test periods the same trials were drawn multiple times. Trial firing rates were z-scored on the parameters from the training subset. We repeated training and testing 100 times, selecting with each iteration a new set of trials randomly. The tested SVM

classifier provides the predicted class labels for the trials in the test subset. This allows us to compute the main decoding success as well as to construct in combination with the true class labels of the test subset confusion matrices. Both possess an accuracy measure in percent. From each confusion matrix we calculated the classification probability as a function of distance from the real numerosity derived from the diagonals (averaged performance parallel to the mean diagonal). The resulting classification probability curve was fitted by a symmetric Gaussian distribution to derive sigma-values as a measure of tuning width for each iteration. It can be described by the following formula:

$$f(x) = amp * e - \left(\frac{x - xc}{\sigma}\right)^2 + y0$$

The standard deviation of the Gaussian distribution (σ) was adjustable during the fitting procedure. The peak function's yaxis offset $y0$ was set to zero, the amplitude (amp) corresponded to the classification probability of the mean diagonal and the center (xc) of the fitted distribution was fixed at zero. We used a Wilcoxon rank sum test (two-sided) to compare the Gaussian widths (σ) determined for each situation.

We evaluated the decoding performance of the SVM classifier in balanced VIP and PFC pseudo-populations by drawing with each iteration a new set of neurons with equal size (decisive factor was the lowest neuron number of the respective brain area). In order to create a chance level performance distribution, we randomly shuffled the association between firing rates and trial class labels (numerosity) with each iteration. The significance threshold was set to $p < 0.05$ (one-sided), i.e. the actual decoding success value was required to be larger than 95 % of values in the shuffled distribution. We tested decoding success distributions between VIP and PFC as well as between sample, match and nonmatch with a permutation test (with alpha correction, 10.000 permutations with replacement) for each stimulus period.

4.13. PEV analysis

To quantify the information about the match (same as sample) or nonmatching (other than sample) numerosity that was carried by a neuron over the time course, we performed the percent explained variance (ω^2 PEV) anaylsis. As other studies already showed (e.g. Buschman et al., 2011; Jacob and Nieder, 2014), the PEV analysis is suitable to measeure the amount of information carried by neuronal populations. In our case, PEV reflects how much of the variance in a neuron's firing rate can be explained by the match and nonmatch numerosity. For this, we considered all neurons that were recorded with at least 18 correct trials per sample numerosity and 3 correct trials per test numerosity in each test phase (match/nonmatch). 369 PFC neurons and 254 VIP neurons fullfilled this criterion. We used a one-way categorical sliding-window ANOVA (200 ms window with 20 ms step size) over the relevant time window of the trial (baseline onset until 400 ms after test onset) to yield the sum-of-squares. From these, we calculated ω^2 according to the following formula:

$$\omega^2 = \frac{SS_{Groups} - df * MSE}{SS_{Total} + MSE} * 100,$$

where SS_{Groups} denotes the sum-of-squares between groups (numerosities), SS_{Total} the total sum-of-squares, df the degrees of freedom, and MSE the mean squared error. The number of trials in each group was balanced to 3 trials. For this, a random subset of trials was drawn and the statistic was calculated. This process was repeated 25 times, and the overall statistic was taken to be the mean of the stratified values. The resulting values were averaged over units to extract the population PEV as a function of time for match and nonmatch numerosities. Furthermore, we calculated the mean PEV during the Test1 period for match and nonmatch condition for a better comparison. For this, we used firing rates which were derived from time windows between the 'main' visual response latency for the respective brain area and 400 ms after test onset. Then, the same calculation process as described above followed. In order to create a baseline PEV, we randomly shuffled the association between firing rates and numerosity. This process was repeated 1000 times, and the shuffled PEV was taken to be the mean of the individual values. The significance threshold was set to $p < 0.01$ (one-sided), i.e. the PEV was required to be larger than 99 % of values in the shuffled distribution. Bin-wise Wilcoxon signed rank tests (evaluated at $p < 0.05$) were used to compare match and nonmatch numerosity PEV within the population of PFC and VIP neurons.

Author contributions

Conceptualization and Methodology: TM, AN; Investigation, Data curation, Analysis, Software and Visualization: TM, JG; Writing – Original draft: TM, AN; Writing - Review/Editing: TM, JG, AN; Funding acquisition and Supervision: AN.

CRediT authorship contribution statement

Julia Grüb: Writing – review & editing, Investigation, Formal analysis. **Tobias Machts:** Writing – review & editing, Writing – original draft, Software, Investigation, Formal analysis, Conceptualization. **Andreas Nieder:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by DFG grant NI 618/10–1 to A.N.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.pneurobio.2025.102794.

Data availability

Data will be made available on request.

References

Anderson, B., Mruczek, R.E., Kawasaki, K., Sheinberg, D., 2008. Effects of familiarity on neural activity in monkey inferior temporal lobe. Cereb. Cortex 18 (11), 2540–2552.
Antzoulatos, E.G., Miller, E.K., 2011. Differences between neural activity in prefrontal cortex and striatum during learning of novel abstract categories. Neuron 71 (2), 243–249.
Baylis, G.C., Rolls, E.T., 1987. Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. Exp. Brain Res. 65, 614–622.
Bogacz, R., Brown, M.W., 2003. Comparison of computational models of familiarity discrimination in the perirhinal cortex. Hippocampus 13 (4), 494–524.
Brown, M.W., Aggleton, J.P., 2001. Recognition memory: what are the roles of the perirhinal cortex and hippocampus? Nat. Rev. Neurosci. 2, 51–61.
Buckley, P.B., Gillman, C.B., 1974. Comparisons of digits and dot patterns. J. Exp. Psychol. 103 (6), 1131–1136.
Buschman, T.J., Siegel, M., Roy, J.E., Miller, E.K., 2011. Neural substrates of cognitive capacity limitations. Proc. Natl. Acad. Sci. 108, 11252–11255.
Chafee, M.V., Goldman-Rakic, P.S., 2000. Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. J. Neurophysiol. 83 (3), 1550–1566.
Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 27.
Constantinidis, C., Steinmetz, M.A., 2001. Neuronal responses in area 7a to multiple stimulus displays: II. Responses are suppressed at the cued location. Cereb. Cortex 11, 592–597.

Cook, E.P., Maunsell, J.H., 2002. Attentional modulation of behavioral performance and neuronal responses in middle temporal and ventral intraparietal areas of macaque monkey. J. Neurosci. 22 (5), 1994–2004.
Cortes, J.M., Greve, A., Barrett, A.B., van Rossum, M.C., 2010. Dynamics and robustness of familiarity memory. Neural Comput. 22 (2), 448–466.
Crowe, D.A., Goodwin, S.J., Blackman, R.K., Sakellaridi, S., Sponheim, S.R., MacDonald 3rd, A.W., Chafee, M.V., 2013. Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. Nat. Neurosci. 16 (10), 1484–1491.
Dasgupta, S., Hattori, D., Navlakha, S., 2022. A neural theory for counting memories. Nat. Commun. 13, 5961.
Dehaene, S., Bossini, S., Giraux, P., 1993. The mental representation of parity and number magnitude. J. Exp. Psychol. Gen. 122, 371–396.
Dehaene, S., Changeux, J.P., 1993. Development of elementary numerical abilities: a neuronal model. J. Cogn. Neurosci. 5 (4), 390–407.
Diester, I., Nieder, A., 2007. Semantic associations between signs and numerical categories in the prefrontal cortex. PLoS Biol. 5 (11), e294.
Engel, T.A., Wang, X.J., 2011. Same or different? A neural circuit mechanism of similarity-based pattern match decision making. J. Neurosci. Official J. Soc. Neurosci. 31 (19), 6982–6996.
Eskandar, E.N., Richmond, B.J., Optican, L.M., 1992. Role of inferior temporal neurons in visual memory. I. Temporal encoding of information about visual images, recalled images, and behavioral context. J. Neurophysiol. 68, 1277–1295.
Fahy, F.L., Riches, I.P., Brown, M.W., 1993. Neuronal activity related to visual recognition memory: long-term memory and the encoding of recency and familiarity information in the primate anterior and medial inferior temporal and rhinal cortex. Exp. Brain Res. 96 (3), 457–472.
Foster, C., Sheng, W.A., Heed, T., Ben Hamed, S., 2022. The macaque ventral intraparietal area has expanded into three homologue human parietal areas. Prog. Neurobiol. 209, 102185.
Freedman, D.J., Assad, J.A., 2006. Experience-dependent representation of visual categories in parietal cortex. Nature 443, 85–88.
Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2003. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J. Neurosci. 23, 5235–5246.
Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2006. Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. Cereb. Cortex 16 (11), 1631–1644.
Galton, F., 1880. Visualised numerals. Nature 21, 252–256.
Ishii, Y., Shimada, M., 2010. The effect of learning and search images on predator–prey interactions. Popul Ecol. 52, 27–35.
Jacob, S.N., Hähnke, D., Nieder, A., 2018. Structuring of abstract working memory content by fronto-parietal synchrony in primate cortex. Neuron 99 (3), 588–597 e5.
Jacob, S.N., Nieder, A., 2014. Complementary roles for primate frontal and parietal cortex in guarding working memory from distractor stimuli. Neuron 83 (1), 226–237.
Jacob, S.N., Stalter, M., Nieder, A., 2016. Cell-type-specific modulation of targets and distractors by dopamine D1 receptors in primate prefrontal cortex. Nat. Commun. 7, 13218.
Kamiński, J., Sullivan, S., Chung, J.M., Ross, I.B., Mamelak, A.N., Rutishauser, U., 2017. Persistently active neurons in human medial frontal and medial temporal lobe support working memory. Nat. Neurosci. 20, 590–601.
Kornblith, S., Quian Quiroga, R., Koch, C., Fried, I., Mormann, F., 2017. Persistent single-neuron activity during working memory in the human medial temporal lobe. Curr. Biol. 27, 1026–1032.
Kutter, E.F., Boström, J., Elger, C.E., Mormann, F., Nieder, A., 2018. Single neurons in the human brain encode numbers. Neuron 100 (3), 753–761 e4.
Kutter, E.F., Boström, J., Elger, C.E., Nieder, A., Mormann, F., 2022. Neuronal codes for arithmetic rule processing in the human brain. Curr. Biol. CB 32 (6), 1275–1284 e4.
Kutter, E.F., Dehnen, G., Borger, V., Surges, R., Mormann, F., Nieder, A., 2023. Distinct neuronal representation of small and large numbers in the human medial temporal lobe. Nat. Hum. Behav. 7 (11), 1998–2007.
Kutter, E.F., Dehnen, G., Borger, V., Surges, R., Nieder, A., Mormann, F., 2024. Single-neuron representation of nonsymbolic and symbolic number zero in the human medial temporal lobe. Curr. Biol. CB 34 (20), 4794–4802 e3.
Li, L., Miller, E.K., Desimone, R., 1993. The representation of stimulus familiarity in anterior inferior temporal cortex. J. Neurophysiol. 69 (6), 1918–1929.
Lim, S., McKee, J.L., Woloszyn, L., Amit, Y., Freedman, D.J., Sheinberg, D.L., Brunel, N., 2015. Inferring learning rules from distributions of firing rates in cortical neurons. Nat. Neurosci. 18 (12), 1804–1810.
Lin, X.X., Nieder, A., & Jacob, S.N. (2023). The neuronal implementation of representational geometry in primate prefrontal cortex. Science advances, 9(50), eadh8685.
Menon, V., 2016. Memory and cognitive control circuits in mathematical cognition and learning. Prog. Brain Res. 227, 159–186.
Merten, K., Nieder, A., 2009. Compressed scaling of abstract numerosity representations in adult humans and monkeys. J. Cogn. Neurosci. 21 (2), 333–346.
Meyer, T., Rust, N.C., 2018. Single-exposure visual memory judgments are reflected in inferotemporal cortex. eLife 7, e32259.
Miller, E.K., Desimone, R., 1994. Parallel neuronal mechanisms for short-term memory. Science 263, 520–522.
Miller, E.K., Erickson, C.A., Desimone, R., 1996. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. J. Neurosci. 16, 5154–5167.
Miller, E.K., Li, L., Desimone, R., 1991. A neural mechanism for working and recognition memory in inferior temporal cortex. Science 254, 1377–1379.
Miller, E.K., Li, L., Desimone, R., 1993. Activity of neurons in anterior inferior temporal cortex during a short-term memory task. J. Neurosci. 13, 1460–1478.
Mistry, P.K., Strock, A., Liu, R., Young, G., Menon, V., 2023. Learning-induced reorganization of number neurons and emergence of numerical representations in a biologically inspired neural network. Nat. Commun. 14 (1), 3843.
Moyer, R.S., Landauer, T.K., 1967. Time required for judgements of numerical inequality. Nature 215 (2), 1519–1520.
Nasr, K., Nieder, A., 2021. Spontaneous representation of numerosity zero in a deep neural network for visual object recognition. iScience 24 (11), 103301.
Nasr, K., Viswanathan, P., Nieder, A., 2019. Number detectors spontaneously emerge in a deep neural network designed for visual object recognition. Sci. Adv. 5 (5) eaav7903.
Ng, C.W., Plakke, B., Poremba, A., 2014. Neural correlates of auditory recognition memory in the primate dorsal temporal pole. J. Neurophysiol. 111 (3), 455–469.
Nieder, A., 2016. The neuronal code for number. Nat. Rev. Neurosci. 17 (6), 366–382.
Nieder, A., 2020. The adaptive value of numerical competence. Trends Ecol. Evol. 35 (7), 605–617.
Nieder, A., 2023. Convergent circuit computation for categorization in the brains of primates and songbirds. Cold Spring Harb. Perspect. Biol. 15 (12), a041526.
Nieder A. (2024). The Calculating Brain. Physiological reviews, 10.1152/physrev.00014.2024. Advance online publication. https://doi.org/10.1152/physrev.00014.2024.
Nieder, A., Dehaene, S., 2009. Representation of number in the brain. Annu. Rev. Neurosci. 32, 185–208.
Nieder, A., Diester, I., Tudusciuc, O., 2006. Temporal and spatial enumeration processes in the primate parietal cortex. Science 313 (5792), 1431–1435.
Nieder, A., Freedman, D.J., Miller, E.K., 2002. Representation of the quantity of visual items in the primate prefrontal cortex. Science 297 (5587), 1708–1711.
Nieder, A., Merten, K., 2007. A labeled-line code for small and large numerosities in the monkey prefrontal cortex. J. Neurosci. 27 (22), 5986–5993.
Nieder, A., Miller, E.K., 2003. Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. Neuron 37 (1), 149–157.
Nieder, A., Miller, E.K., 2004. A parieto-frontal network for visual numerical information in the monkey. Proc. Natl. Acad. Sci. 101 (19), 7457–7462.
Norman, K.A., O'Reilly, R.C., 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol. Rev. 110 (4), 611–646.
Norman, K.A., O'Reilly, R.C., 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol. Rev. 110 (4), 611–646.
Okuyama, S., Kuki, T., Mushiake, H., 2015. Representation of the numerosity 'zero' in the parietal cortex of the monkey. Sci. Rep. 5, 10059.
Parto Dezfouli, M., Zarei, M., Constantinidis, C., Daliri, M.R., 2021. Task-specific modulation of PFC activity for matching-rule governed decision-making. Brain Struct. Funct. 226 (2), 443–455.
Qi, X.-L., Meyer, T., Stanford, T.R., Constantinidis, C., 2012. Neural correlates of a decision variable before learning to perform a match/non-match task. J. Neurosci. 32, 6161–6169.
Rainer, G., Miller, E.K., 2000. Effects of visual experience on the representation of objects in the prefrontal cortex. Neuron 27 (1), 179–189.
Rawley, J.B., Constantinidis, C., 2010. Effects of task and coordinate frame of attention in area 7a of the primate posterior parietal cortex. J. Vis. 10 (1), 1–16.
Riches, I.P., Wilson, F.A., Brown, M.W., 1991. The effects of visual stimulation and memory on neurons of the hippocampal formation and the neighboring parahippocampal gyrus and inferior temporal cortex of the primate. J. Neurosci. 11, 1763–1779.
Riches, I.P., Wilson, F.A., Brown, M.W., 1991. The effects of visual stimulation and memory on neurons of the hippocampal formation and the neighboring parahippocampal gyrus and inferior temporal cortex of the primate. J. Neurosci. Official J. Soc. Neurosci. 11 (6), 1763–1779.
Roy, J.E., Riesenhuber, M., Poggio, T., Miller, E.K., 2010. Prefrontal cortex activity during flexible categorization. J. Neurosci. Official J. Soc. Neurosci. 30 (25), 8519–8528.
Sawamura, H., Shima, K., Tanji, J., 2002. Numerical representation for action in the parietal cortex of the monkey. Nature 415 (6874), 918–922.
Sohal, V.S., Hasselmo, M.E., 2000. A model for experience-dependent changes in the responses of inferotemporal neurons. Network 11 (3), 169–190.
Steinmetz, M.A., Constantinidis, C., 1995. Neurophysiological evidence for a role of posterior parietal cortex in redirecting visual attention. Cereb. Cortex 5, 448–456.
Swaminathan, S.K., Freedman, D.J., 2012. Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. Nat. Neurosci. 15, 315–320.
Tamura, K., Takeda, M., Setsuie, R., Tsubota, T., Hirabayashi, T., Miyamoto, K., Miyashita, Y., 2017. Conversion of object identity to object-general semantic value in the primate temporal cortex. Science 357 (6352), 687–692.
Tyulmankov, D., Yang, G.R., Abbott, L.F., 2022. Meta-learning synaptic plasticity and memory addressing for continual familiarity detection. Neuron 110 (3), 544–557 e8.
Uhrig, L., Dehaene, S., Jarraya, B., 2014. A hierarchy of responses to auditory regularities in the macaque brain. J. Neurosci. Official J. Society Neurosci. 34 (4), 1127–1132.
Verguts, T., Fias, W., 2004. Representation of number in animals and humans: a neural model. J. Cogn. Neurosci. 16 (9), 1493–1504.
Viswanathan, P., Nieder, A., 2013. Neuronal correlates of a visual "sense of number" in primate parietal and prefrontal cortices. Proc. Natl. Acad. Sci. 110 (27), 11187–11192.
Viswanathan, P., Nieder, A., 2015. Differential impact of behavioral relevance on quantity coding in primate frontal and parietal neurons. Curr. Biol. CB 25 (10), 1259–1269.

Viswanathan, P., Nieder, A., 2020. Spatial neuronal integration supports a global representation of visual numerosity in primate association cortices. J. Cogn. Neurosci. 32 (6), 1184–1197.
Viswanathan, P., Stein, A.M., Nieder, A., 2024. Sequential neuronal processing of number values, abstract decision, and action in the primate prefrontal cortex. PLoS Biol. 22 (2), e3002520.
Vogels, R., Sary, G., Orban, G.A., 1995. How task-related are the responses of inferior temporal neurons? Vis. Neurosci. 12, 207–214.
Wang, L., Uhrig, L., Jarraya, B., Dehaene, S., 2015. Representation of numerical and sequential patterns in macaque and human brains. Curr. Biol. CB 25 (15), 1966–1974.
Winters, B.D., Saksida, L.M., Bussey, T.J., 2008. Object recognition memory: neurobiological mechanisms of encoding, consolidation and retrieval. Neurosci. Biobehav. Rev. 32 (5), 1055–1070.
Woloszyn, L., Sheinberg, D.L., 2012. Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex. Neuron 74 (1), 193–205.
Wu, Z., Buckley, M.J., 2022. Prefrontal and medial temporal lobe cortical contributions to visual short-term memory. J. Cogn. Neurosci. 35 (1), 27–43.
Xiang, J.Z., Brown, M.W., 1998. Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. Neuropharmacology 37 (4-5), 657–676.
Zhang, T., Britten, K.H., 2010. The responses of VIP neurons are sufficiently sensitive to support heading judgments. J. Neurophysiol. 103 (4), 1865–1873.
Zhang, T., Britten, K.H., 2011. Parietal area VIP causally influences heading perception during pursuit eye movements. J. Neurosci. Official J. Society Neurosci. 31 (7), 2569–2575.
Zhou, Y., Rosen, M.C., Swaminathan, S.K., Masse, N.Y., Zhu, O., Freedman, D.J., 2021. Distributed functions of prefrontal and parietal cortices during sequential categorical decisions. eLife 10, e58782.
Zhou, Y., Zhu, O., Freedman, D.J., 2023. Posterior parietal cortex plays a causal role in abstract memory-based visual categorical decisions. J. Neurosci. 43 (23), 4315–4328.
Zorzi, M., Priftis, K., Umiltà, C., 2002. Brain damage: neglect disrupts the mental number line. Nature 417, 138–139.