CellPress

Neuron
Article

# Complementary Roles for Primate Frontal and Parietal Cortex in Guarding Working Memory from Distractor Stimuli

Simon Nikolas Jacob[1,2,*] and Andreas Nieder[1]
[1]Animal Physiology, Institute of Neurobiology, University of Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany
[2]Department of Psychiatry and Psychotherapy, Charité Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany
*Correspondence: simon.jacob@charite.de
http://dx.doi.org/10.1016/j.neuron.2014.05.009

## SUMMARY

**Prefrontal cortex (PFC) and posterior parietal cortex are important for maintaining behaviorally relevant information in working memory. Here, we challenge the commonly held view that suppression of distractors by PFC neurons is the main mechanism underlying the filtering of task-irrelevant information. We recorded single-unit activity from PFC and the ventral intraparietal area (VIP) of monkeys trained to resist distracting stimuli in a delayed-match-to-numerosity task. Surprisingly, PFC neurons preferentially encoded distractors during their presentation. Shortly after this interference, however, PFC neurons restored target information, which predicted correct behavioral decisions. In contrast, most VIP neurons only encoded target numerosities throughout the trial. Representation of target information in VIP was the earliest and most reliable neuronal correlate of behavior. Our data suggest that distracting stimuli can be bypassed by storing and retrieving target information, emphasizing active maintenance processes during working memory with complementary functions for frontal and parietal cortex in controlling memory content.**

## INTRODUCTION

Cognitive control involves the grouping of stimuli into meaningful categories, online storage in working memory, and selection of behaviorally relevant over irrelevant information (Baddeley, 2012). Because working memory has a limited capacity, relevant information needs to be protected against distracting representations (Luck and Vogel, 1997; Vogel et al., 2005). The prefrontal cortex (PFC) and posterior parietal cortex (PPC) are thought to be the major nodes that enable us to selectively attend to target stimuli while filtering distracting information.

A large body of experimental evidence suggests that PFC and PPC adopt specialized functions in working memory and selective attention. PPC neurons represent the most recent stimulus irrespective of its relevance to the current task (Constantinidis and Steinmetz, 1996), and thus fully encode distractors (Bisley and Goldberg, 2003, 2006; Suzuki and Gottlieb, 2013). In contrast, PFC has been associated with controlling lower-level visual areas and gating access to working memory (Anderson and Green, 2001; Feredoes et al., 2011; McNab and Klingberg, 2008). The ability to resist interfering stimuli is compromised in monkeys (Malmo, 1942; Suzuki and Gottlieb, 2013) and humans with lateral PFC lesions (Chao and Knight, 1995, 1998). Responses of single PFC neurons in the monkey are diminished for unattended targets (Everling et al., 2002). Compared to parietal cells, PFC neurons respond little to the presentation of distractors (di Pellegrino and Wise, 1993; Lennert and Martinez-Trujillo, 2011; Qi et al., 2010; Suzuki and Gottlieb, 2013). These studies collectively suggest that attentional filtering performance in primates relies on the ability of PFC neurons to suppress interfering stimuli.

Our current knowledge of the resistance to memory interference stems from experiments that measured neuronal responses to comparatively simple spatial stimuli placed in the visual periphery. Thus, there is currently insufficient data to determine whether prefrontal inhibition of distractors is a general principle of cognitive functioning or rather restricted to particular situations. Specifically, it is unknown whether prefrontal suppression of interfering stimuli is found when more complex features that typically drive PFC neurons, such as abstract cognitive categories, have to be filtered. To investigate the neuronal mechanisms of maintaining abstract category information in the light of interference, we trained two rhesus monkeys to memorize the number of visual items (numerosity) while resisting other distracting numerosities. We then simultaneously recorded single-unit activity from the PFC and the ventral intraparietal area (VIP) of the PPC, two key areas for numerosity processing that contain high proportions of quantity-selective neurons (Nieder and Miller, 2004; Nieder et al., 2002, 2006; Viswanathan and Nieder, 2013).

We considered two hypotheses. Task-irrelevant distractor numerosities could be processed primarily by parietal neurons, whereas prefrontal neurons might remain largely unaffected by interfering information. Alternatively, PFC target representations could break down in the face of strong distracting stimulation, possibly requiring this area to regenerate target information following the interference to solve the task. We found evidence for the latter. PFC readily represented the distractor but subsequently recovered target information, while,

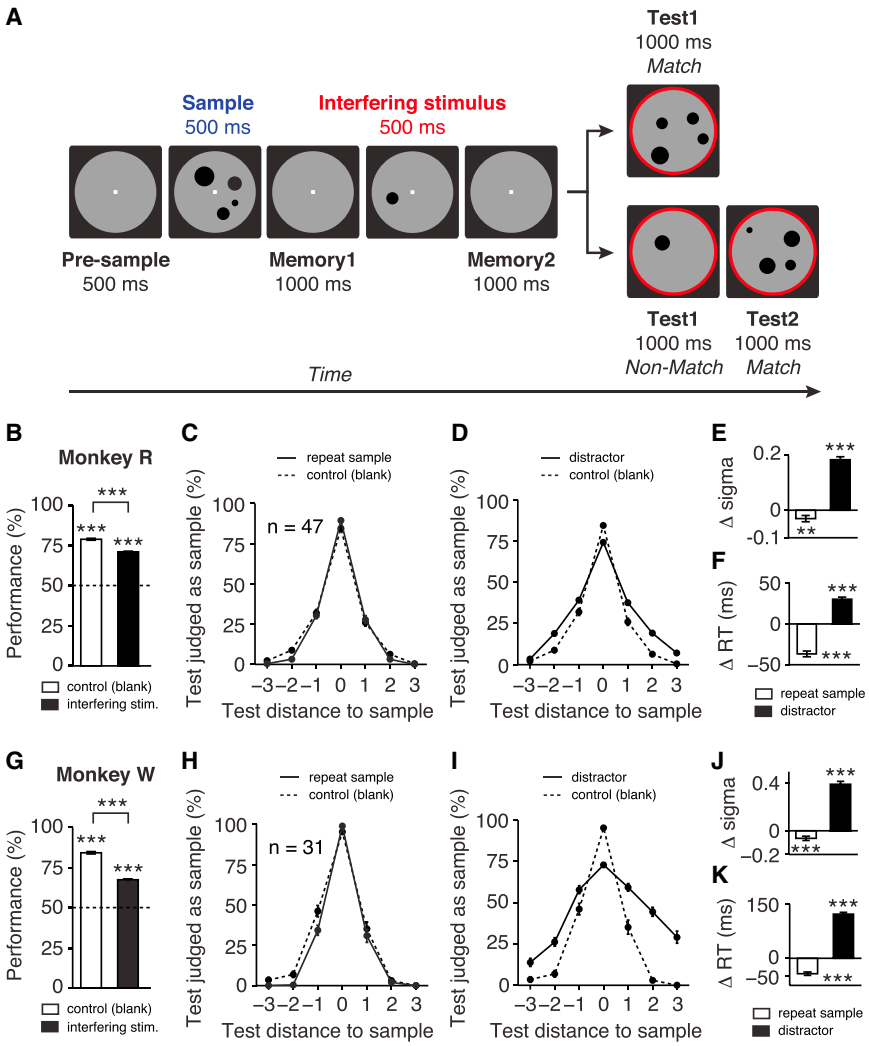


**Figure 1. Task Protocol and Behavioral Performance**

(A) Delayed-match-to-sample task. Monkeys had to release a bar if the sample and first test display contained the same number of items (match) and had to continue holding it if they did not (non-match). A task-irrelevant, interfering numerosity presented in the working memory period had to be resisted.

(B–F) Behavioral performance for monkey R (n = 47 sessions).

(B) Mean performance in trials without (control, blank) and with interfering stimuli. The dashed line denotes chance level.

(C) Performance curves for trials without interfering stimuli (control; dashed line) and for trials where the interfering numerosity was identical to, i.e., repeated the sample (solid line). The monkeys' performance for all sample test combinations is plotted against numerical distance between test and sample numerosity. The peak represents the percentage of correct match trials, and other data points mark the percentage of errors in nonmatch trials.

(D) Performance curves for trials without interfering stimuli (control; dashed line) and for trials where the interfering numerosity was not identical to the sample, i.e., a true distractor (solid line).

(E) Performance curve width was used as a measure of the precision of sample numerosity representation. Data are presented as the difference in width compared to the control condition for trials where the interfering numerosity repeated the sample (left bar) and for trials where the interfering numerosity represented a true distractor (right bar).

(F) Difference in RTs (correct match trials) compared to the control condition for trials where the interfering numerosity repeated the sample (left bar) and for trials with a true distractor (right bar).

(G–K) Same convention as in (B)–(F) or monkey W (n = 31 sessions). Error bars, SEM across sessions. $^{**}p < 0.01$; $^{***}p < 0.001$.

surprisingly, target memories were maintained to a significantly greater extent in VIP neurons. Our results differ from previous studies by showing that neuronal suppression of interfering stimuli in PFC is not necessary to overcome distractors, and suggest different mechanisms by which the frontoparietal network controls working memory content to guide goal-directed behavior.

## RESULTS

### Behavioral Performance

Two monkeys performed a modified version of a delayed-match-to-numerosity task (Nieder et al., 2002) (sample numerosities 1–4), in which a task-irrelevant, interfering numerosity was embedded in the working memory period (Figure 1A). The 500-ms-duration interfering stimulus (ranging from 1 to 4 items) was presented during the memory interval on 80% of the trials (20% of the trials each with numerosity 1, 2, 3, and 4). In the remaining 20% of the trials, a blank gray background circle of equal duration replaced the interfering numerosity, i.e., no task-irrelevant stimulus was shown (standard delayed-match-to-numerosity task). These trials served as control trials. Low-level visual features were controlled and could not systematically influence task performance (Nieder et al., 2002).

Both animals had previously received extensive training in the standard delayed-match-to-numerosity task (Nieder et al., 2006). Within 3 to 5 months of gradually introducing the interfering numerosity, performance also stabilized in these trials (see Experimental Procedures). As expected, performance in trials with interfering stimuli was lower compared to control trials (monkey R: 71% ± 0.5% versus 79% ± 0.6% [n = 47 sessions], $p < 0.001$, Wilcoxon signed-rank test, Figure 1B; monkey W: 67% ± 0.5% versus 84% ± 0.8% [n = 31 sessions], $p < 0.001$, Figure 1G). Importantly, both animals performed significantly above chance level in trials with interfering stimuli ($p < 0.001$, Wilcoxon signed-rank test, for monkey R, Figure 1B, and monkey W, Figure 1G). Successful filtering of the interfering stimulus was evident in match trials where the interfering

numerosity differed from the sample and test numerosity (Figure 1A, upper trial branch; monkey R: 71% ± 0.7%, $p < 0.001$ versus chance level, Wilcoxon signed-rank test; monkey W: 71% ± 1.0%, $p < 0.001$). The reduction in performance induced by the interfering numerosity was most pronounced in these trials (monkey R: −11%, monkey W: −23% compared to control match trials). Performance decreased to a smaller extent in all other trial types (e.g., in nonmatch trials where the interfering numerosity matched the first test stimulus; Figure 1A, lower trial branch; monkey R: −6%, monkey W: −13% compared to control nonmatch trials). The reduction in performance was in the range of and comparable to recent studies using other interfering stimuli (Lennert and Martinez-Trujillo, 2011; Suzuki and Gottlieb, 2013). Finally, there were almost no trials in which the animals mistakenly responded to the task-irrelevant stimulus instead of to the first test numerosity, which was marked by a red ring (monkey R: < 0.01%, monkey W: < 0.2%; Figure 1A). Collectively, these results showed that both animals had learned to respond to the sample numerosity despite memory interference by the task-irrelevant stimulus.

To quantify the precision of sample representation in working memory, we plotted the extent to which animals judged the test as equal in number to the sample as a function of the numerical distance between test and sample numerosity. Peaked curves were obtained in trials with interfering stimuli, as in control trials, which confirmed that both animals continued to correctly compare the test to the sample despite the presence of the interfering numerosity (Figures 1C, 1D, 1H, and 1I). The impact of the interfering stimulus on performance depended on its numerosity. A small increase in performance was observed when the interfering stimulus was equal in number to, i.e., repeated, the sample numerosity ("repeat-sample trials") (Figures 1C and 1H). Compared to control trials, performance curves were slightly sharper in repeat-sample trials (monkey R: △ sigma [difference of performance curve widths] −0.03 ± 0.01, $p < 0.01$, Wilcoxon signed-rank test, Figure 1E; monkey W: △ sigma −0.06 ± 0.02, $p < 0.001$, Figure 1J). In contrast, the monkeys made more errors when the interfering numerosity was unrelated to the sample, i.e., a true distractor (Figures 1D and 1I). Compared to control trials, behavioral performance curves were wider in distractor trials (monkey R: △ sigma 0.18 ± 0.01, $p < 0.001$, Figure 1E; monkey W: △ sigma 0.39 ± 0.02, $p < 0.001$, Figure 1J), indicating more errors on trials in which sample and interfering numerosity differed. As a second behavioral parameter, we analyzed reaction times (RTs), which were shorter in repeat-sample trials compared to control trials (monkey R: △ RT −37 ± 4 ms, $p < 0.001$, Wilcoxon signed-rank test, Figure 1F; monkey W: △ RT −43 ± 5 ms, $p < 0.001$, Figure 1K) and increased in true distractor trials (monkey R: △ RT 30 ± 3 ms, $p < 0.001$, Figure 1F; monkey W: △ RT 122 ± 5 ms, $p < 0.001$, Figure 1K). Thus, the monkeys' errors were not the result of a general lack of understanding of the task but instead reflected the power of an interfering numerosity to specifically compete with the sample for representation in working memory. For simplicity, the interfering stimulus (comprising both repetitions of the sample and true distractors) will from now on be referred to as the distractor.

## Distractor Numerosities Are Fully Represented in PFC

We simultaneously recorded single-unit activity from lateral PFC ($n = 473$; 319 from monkey R, 154 from monkey W) and parietal VIP ($n = 386$; 278 from monkey R, 108 from monkey W) while the monkeys performed the task (Figures 2A–2E). Almost all neurons in both areas significantly modulated their firing rate during the course of the trial (task-related neurons; one-way ANOVA across trial epochs, evaluated at $p < 0.05$; PFC: $n = 448$ [95%]; VIP: $n = 361$ [94%]); Figures 2F and 2G, left). Of these task-related neurons, a large proportion encoded the sample and/or distractor numerosity in at least one trial epoch (two-way ANOVA with factors sample [1–4] and distractor numerosity [1–4], evaluated at $p < 0.01$; PFC: total $n = 234$ [52%]; VIP: total $n = 130$ [36%]); Figures 2F and 2G, right; see Table S1 for details). The proportion of number-selective neurons in PFC and VIP observed here was comparable to and even exceeded results from previous studies (Nieder et al., 2002; Nieder and Miller, 2004; Tudusciuc and Nieder, 2007).

We quantified how much information about the sample and distractor numerosity was carried by the discharge rates of the entire population of task-related PFC and VIP neurons ($\omega^2$ explained variance) (Buschman et al., 2011; Siegel et al., 2009) (one-way ANOVA for sample and distractor numerosity [Siegel et al., 2009; Warden and Miller, 2007]), Figures 2H and 2I; identical results that also included the interaction term were obtained using a two-way ANOVA with factors sample and distractor numerosity, Figure S1). In PFC, sample information increased sharply after sample presentation and reached a second peak immediately prior to distractor presentation (Figure 2H). Surprisingly, sample information then strongly decreased (sample percent explained variance [$\omega^2$ PEV] in 100 ms windows after distractor onset and offset: 2.1 ± 0.2 versus 0.8 ± 0.1, respectively, $p < 0.001$, Wilcoxon signed-rank test) and was largely replaced during the distractor period by information about the distractor stimulus (bin-wise Wilcoxon signed-rank test; Figure 2H). Distractor information reached peak levels comparable to previous maximal values for sample information. In the second memory period, i.e., after distractor offset, sample information recovered but was paralleled by increasing distractor information, i.e., sample and distractor numerosity were represented to the same extent (Figure 2H).

A different picture emerged for VIP, where information about the sample numerosity was maintained throughout the trial (sample $\omega^2$ PEV in 100 ms windows after distractor onset and offset: 1.0 ± 0.1 versus 1.0 ± 0.1, respectively, $p = 0.93$, Wilcoxon signed-rank test; Figure 2I). Sample information in VIP was always larger than for the distractor, even during distractor presentation (bin-wise Wilcoxon signed-rank test; Figure 2I). Moreover, VIP activity clearly favored the sample over the distractor numerosity during the second memory period prior to presentation of the test stimulus. Collectively, these results demonstrate that the distractor numerosity had a more disturbing effect on maintenance of sample information in prefrontal cortex, not posterior parietal cortex.

Because the distractor numerosity systematically influenced the animals' behavior (Figure 1), we performed the analysis of sample information separately for repeat-sample, control (blank), and true distractor trials (Figures 2J and 2K). In both

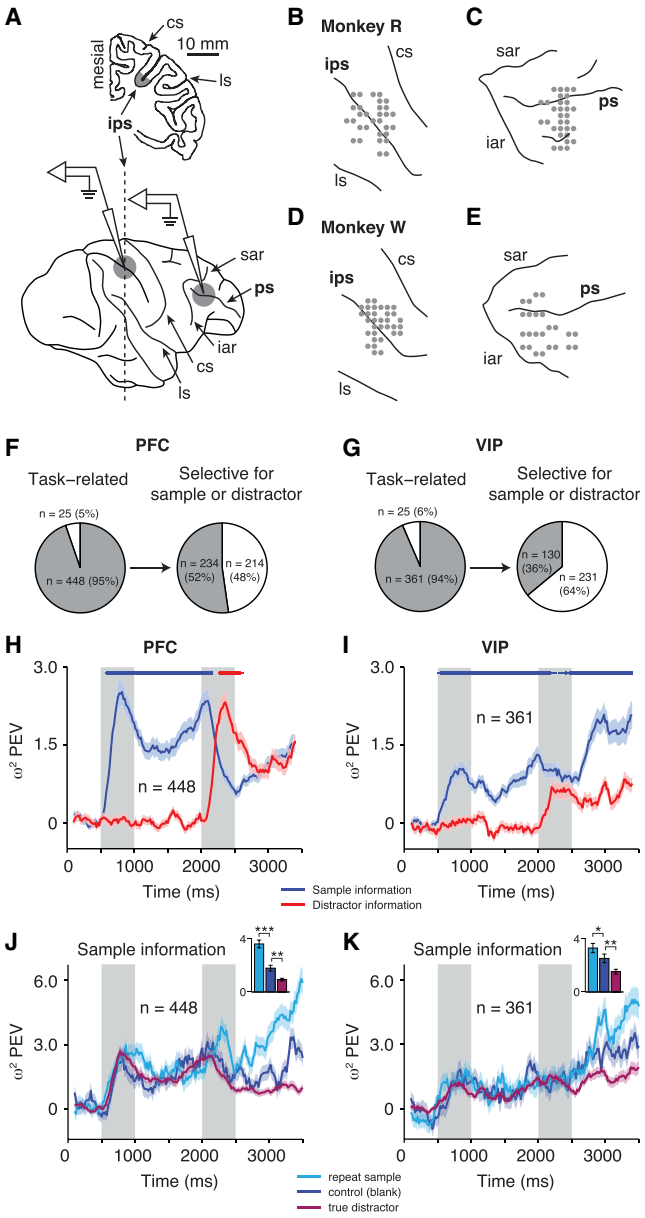


**Figure 2. Electrophysiological Recordings and Neuronal Selectivity for Sample and Distractor Numerosities**

(A) Schematics of a rhesus monkey brain depicting the location of simultaneous single-unit recordings in the prefrontal and parietal (area VIP) cortices.

(B and C) Anatomical surface reconstruction of the recording penetrations in parietal cortex (fundus of the intraparietal sulcus, VIP; B) and lateral PFC (C) of monkey R.

(D and E) Same convention as in (B) and (C) for monkey W.

(F and G) Percentage of recorded neurons that were task related (left) and selective for the sample or distractor numerosity (right) in PFC (F) and VIP (G).

(H and I) Sliding-window $\omega^2$ PEV quantifying the information about the sample and distractor numerosity present across task-related neurons recorded in prefrontal (H) and parietal cortex (I). Bars above the curves denote the time bins where sample information was larger than distractor information (blue; thick: $p < 0.01$; thin: $p < 0.05$) or smaller (red).

(J and K) Sliding-window $\omega^2$ PEV quantifying the information about the sample numerosity across task-related neurons recorded in prefrontal (J) and parietal cortex (K) for repeat-sample, control, and true distractor trials. Inset: mean $\omega^2$ PEV across the second memory period for each trial type. Error bars and bands, SEM across neurons; ips, intraparietal sulcus; cs, central sulcus; ls, lateral sulcus; ps, principal sulcus; sar, superior arcuate sulcus; iar, inferior arcuate sulcus; ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. See also Figure S1.

areas, sample information in the second memory period was greatest for repeat-sample trials (facilitation), while true distractors reduced sample information below the level of control trials (Wilcoxon signed-rank tests; inset in Figures 2J and 2K).

## Neuronal Population Dynamics in State Space

To gain further insight into the temporal dynamics of the recorded neuronal populations, we asked how individual numerosities were represented across time. We therefore decomposed ensemble activity in PFC and VIP using multidimensional state-space analysis (neuronal circuit trajectories) (Harvey et al., 2012; Stokes et al., 2013) (Figures 3A and 3B). At each point in time, the activity of n recorded neurons can be defined by a point in n-dimensional space, with each dimension representing the activity of a single neuron. Different trajectories are traversed for different neuronal states, i.e., they represent the encoding of different stimuli in working memory. Dimensionality is effectively reduced using factor analysis. We first sorted trials by sample numerosity. Prefrontal trajectories separated after presentation of the sample and parametrically spanned state space as a function of sample numerosity (Figure 3A); adjacent numerosities were represented by adjacent trajectories, forming an ordered layout of neuronal states. Thus, similar target numerosities were laid down in similar patterns of neuronal ensemble activity. Trajectories in PFC were maximally separated after the first memory delay, but collapsed almost completely when the distractor was presented (compare time points four and five). Separation partially returned in the second memory delay. Population activity in VIP initially followed a similar time course (Figure 3B). Notably, however, the distance between sample trajectories was reduced to a lesser extent in the distractor period (compare time points four and five).

The layout of sample trajectories in state space reflected two characteristics of the analog representation of quantities: the distance between two trajectories scaled with the numerical distance between their corresponding numerosities (distance effect); at equal numerical distances, two trajectories were closer as the magnitudes of their corresponding numerosities increased (magnitude effect) (Nieder, 2013) (Figures 3C and 3D). We used the mean intertrajectory distance ($\triangle$d) as a proxy for the strength of neuronal selectivity for a stimulus (Figures 3E and 3F). Trajectories were calculated using trials sorted either by sample or by distractor numerosity (Figures S2A and S2B). PFC ensemble activity strongly encoded the sample numerosity until the distractor period, when the distractor started to override the sample numerosity (at distractor peak bin: sample $\triangle$d 0.25 ± 0.08, distractor $\triangle$d 0.71 ± 0.23, $p < 0.05$, Wilcoxon signed-rank test; Figure 3E). This prominent reversal was not observed in VIP (at distractor peak bin: sample $\triangle$d 0.33 ± 0.12, distractor $\triangle$d 0.33 ± 0.14, $p = 1.0$, Wilcoxon signed-rank test). Unlike in PFC, VIP sample trajectories diverged significantly more than distractor trajectories in the second memory period (mean across time bins: PFC, sample $\triangle$d 0.24 ± 0.08, distractor $\triangle$d

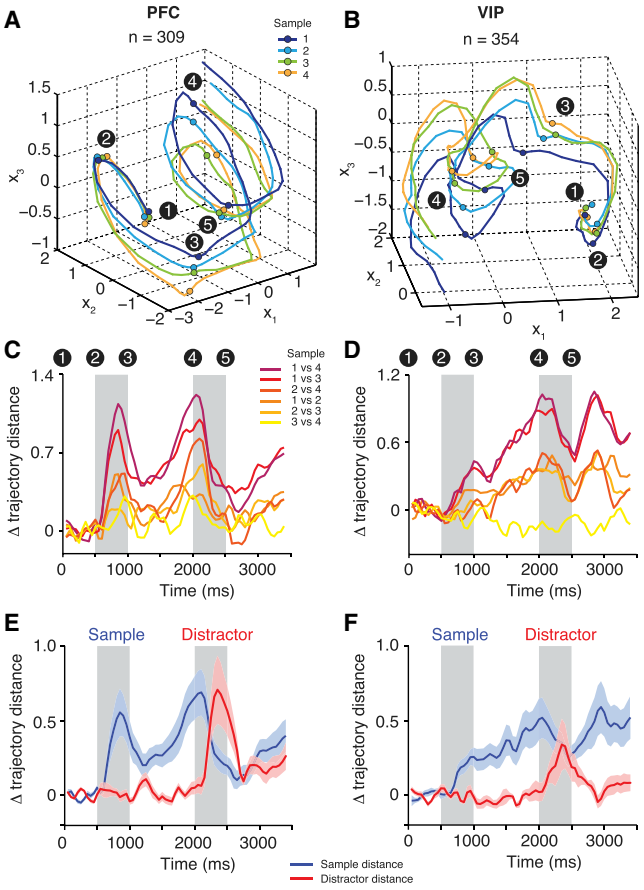


**Figure 3. Neuronal Population Trajectories for Sample and Distractor Numerosities**

(A) Factor analysis describing the state space of neuronal population activity in PFC (n = 309) for each sample across time, plotted for the first three common factors. Time points mark the onset of the (1) fixation (presample), (2) sample, (3) first memory, (4) distractor, and (5) second memory period. Trajectories represent the mean across single trials.

(B) Same analysis for the population of VIP neurons (n = 354).

(C) Intertrajectory Euclidean distance across time as a measure of neuronal stimulus selectivity for all sample-sample combinations in PFC.

(D) Same layout as in (C) for VIP.

(E) Mean intertrajectory distance across time for the PFC population. Blue and red curves represent the distance averaged across all sample-sample and distractor-distractor trajectory combinations, respectively.

(F) Same layout as in (E) for VIP. Error bands, SEM across trajectories. See also Figures S2 and S3.

0.24 ± 0.07, p = 1.0, Wilcoxon signed-rank test; VIP, sample △d 0.37 ± 0.13, distractor △d 0.06 ± 0.09, p < 0.05; Figures 3E and 3F).

Distractor trajectories were equally well described in state space defined by sample or distractor numerosities (Figures S2C and S2D). To further quantify the trajectory divergence, we used a classifier based on the distance from an individual trial trajectory to the mean sample and distractor trajectories at individual time points (Figures S3Aa and S3B). It was possible to decode from PFC population activity in single correct trials the numerosity currently held in memory at better than chance levels in all trial periods (Figure S3A). Following distractor presentation, it was more likely to correctly decode the distractor than the sample numerosity in PFC. Single-trial classification accuracy was lower in VIP but, importantly, higher for the sample compared to the distractor throughout the trial (Figure S3B). These results show that ensemble activity in PFC and VIP can be described using divergent, numerosity-specific trajectories through state space and provide support for the explained variance data (Figures 2H and 2I) using a conceptually distinct analysis.

## Single VIP Neurons Encode Sample Information, Not Distractor Information

We next determined whether the observed population responses were obtained by averaging over heterogeneous groups of neurons or whether they reflected the activity of typical single units (Figure S4). A representative prefrontal neuron increased its firing rate when either the sample or distractor numerosity was presented (Figure 4A). When trials were sorted by sample, firing rates in this cell diverged during the sample period, but not during the distractor period (Figure 4A, top panel). Conversely, activity levels discriminated distractor, but not sample, numerosities when trials were sorted by distractor (Figure 4A, bottom panel). Thus, this neuron encoded the most recently presented stimulus, irrespective of its relevance for solving the task (Figure 4B). Another example PFC neuron increased its firing rate primarily during the memory periods (Figures 4C and 4D). In this neuron, distractor information temporarily replaced the memory of the sample numerosity. VIP neurons that encoded the distractor were rare (see Figures S5A and S5B for an example). Typical single neurons in this area encoded the sample numerosity at various stages of the trial, including during the presentation of the distractor, but never represented the distracting stimulus (Figures 4E–4H).

We explored the time course of sample or distractor selectivity for all task-related single neurons in both areas (sliding-window analysis of explained variance; Figure 5). Similar to previous studies (Nieder and Miller, 2004; Viswanathan and Nieder, 2013), VIP neurons represented sample numerosities on average earlier than PFC neurons (VIP: 158 ± 4 ms, PFC: 171 ± 1 ms; Figure S6), although this result did not reach statistical significance in the present data set (p = 0.47, Wilcoxon rank sum test). A total of 209 individual PFC neurons reached the criteria for sample selectivity at some point during the trial (see Experimental Procedures). A total of 110 PFC neurons represented the distractor (inset, Figure 5A). Note that the number of time bins during which a neuron might become selective for the sample was twice that for the distractor. In VIP, 144 neurons represented the sample, and only 23 neurons represented the distractor (inset, Figure 5B). Significantly more neurons encoded both the sample and the distractor in PFC compared to VIP (n = 80/209 versus n = 15/144, p < 0.001, Fisher's exact test).

To integrate sample and distractor preference into a single measure, we calculated a time-resolved stimulus selectivity index (SSI) for all sample-selective neurons: positive values indicate that the neuron's firing-rate variance is better explained by differences in the sample numerosity, and negative values indicate that the neuron's activity correlates better with the

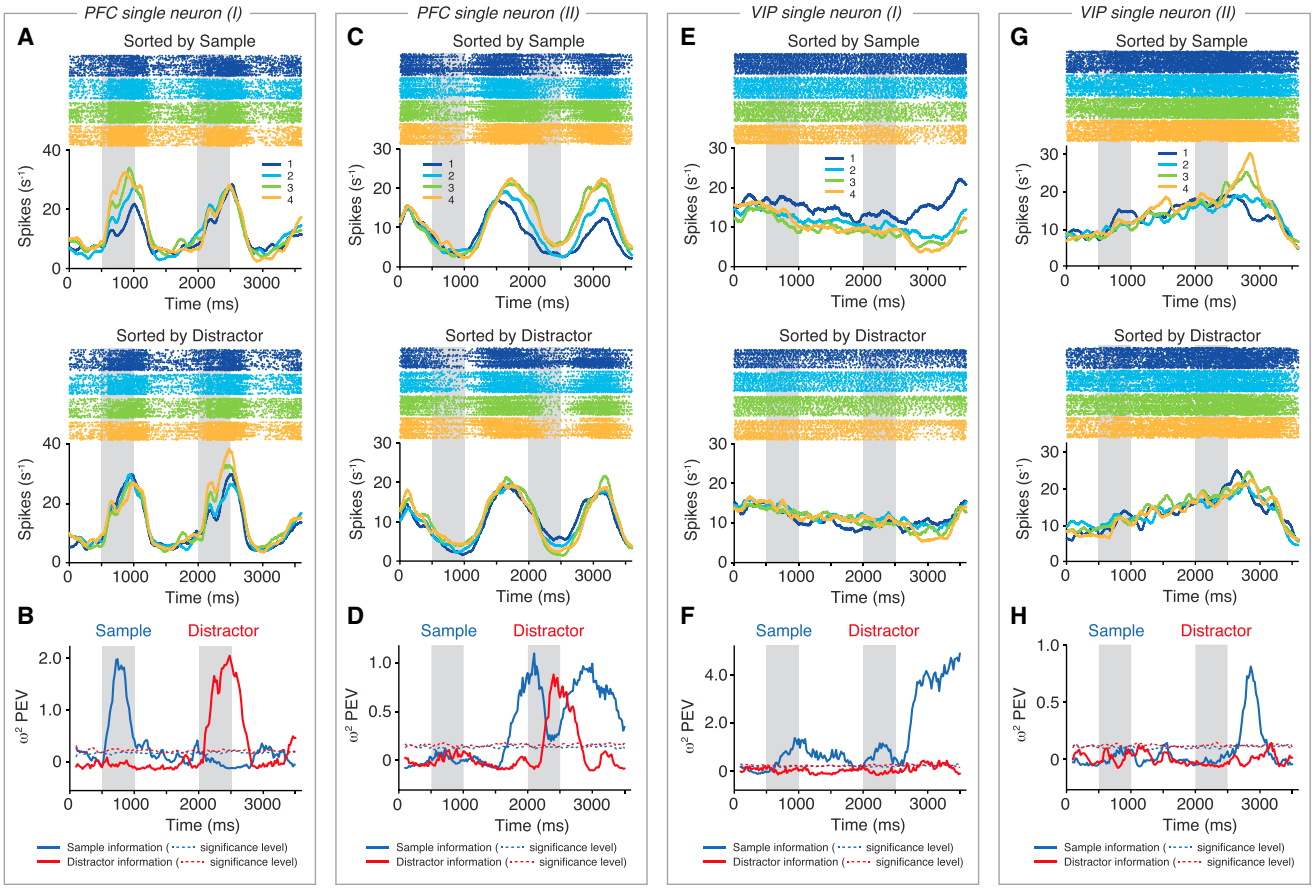


**Figure 4. Example PFC and VIP Single Neurons**

(A) Raster plots and spike-density histograms for an example PFC neuron. Trials are sorted by sample (top panel) or distractor numerosity (bottom panel).

(B) Sliding-window $\omega^2$ PEV quantifying the information about the sample and distractor numerosity for the neuron in (A). Dashed lines mark the significance threshold ($p = 0.01$).

(C) Raster plots and spike-density histograms for a different example PFC neuron.

(D) Sliding-window PEV for the neuron in (C).

(E and F) Same layout as in (A) and (B) for an example VIP neuron. Note the difference in y axis scaling in (F) compared to (B) or (D).

(G and H) Same layout as in (E) and (F) for a different example VIP neuron. See also Figures S4 and S5.

distractor (SSI values were normalized to the maximum for illustration purposes) (Figures 5C and 5D). In PFC, the distractor was encoded by neurons that became sample selective at various different time points during the trial (Figure 5C). The distractor was particularly well represented in neurons with short sample latencies that started to encode the sample shortly after it was presented (example neuron in Figures 4A and 4B). These neurons' SSI time courses following sample and distractor onset were mirror images of each other, i.e., these cells did not distinguish at all between task-relevant and irrelevant stimuli. In VIP, few of these neurons were found (example neuron in Figures S5A and S5B). The vast majority of cells did not represent the distractor at any time point. Instead, sample-selective neurons were found in all trial periods, including during distractor presentation (Figure 5D).

To determine the extent to which single neurons changed their coding preference from the sample to the distractor, we calculated cross-correlations between sample and distractor numerosity tuning curves (Diester and Nieder, 2007). For all neurons that started to encode the sample in either the sample or first memory period (one-way ANOVA, evaluated at $p < 0.01$; PFC: $n = 177$, VIP: $n = 68$), trials were sorted by sample numerosity in the sample and first memory period and by distractor in the distractor and second memory period (Figure 6A). Numerosity tuning curves were derived from time windows at equivalent positions in the first (sample-sorted) and second (distractor-sorted) halves of the trial and cross-correlated. The higher the cross-correlation coefficient (CC), the more a neuron switched from encoding the sample numerosity to representing the distractor. CC values were particularly high in PFC neurons that encoded the sample early on (Figures S7A and S7B). To account for potential differences in response latency between PFC and VIP neurons, we compared CC values at peak. Mean peak CCs were significantly higher in PFC compared to VIP neurons (PFC: 0.53 ± 0.01, VIP: 0.43 ± 0.02, $p < 0.001$, Wilcoxon rank sum test, Figure 6B). As a control for differences in cell counts

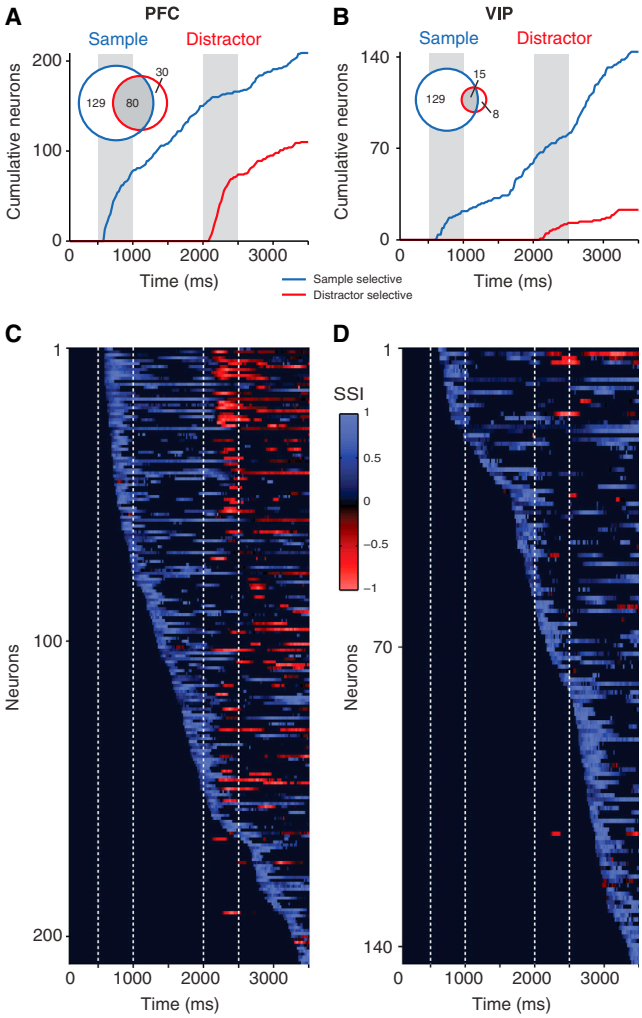


**Figure 5. Representation of Sample and Distractor Numerosities in PFC and VIP Neurons**

(A) Cumulative number of PFC neurons reaching the significance threshold for sample or distractor selectivity as determined by a sliding-window analysis of PEV (Figure 4). Inset: Venn diagram of total sample- and distractor-selective neurons (n = 209 and n = 110, respectively; overlap: n = 80).

(B) Same layout as in (A) for VIP neurons (sample: n = 144, distractor: n = 23; overlap: n = 15).

(C) SSI for all sample-selective PFC neurons across time. Values of 1 (blue) indicate a neuron encodes only the sample, values of −1 (red) signal complete representation of the distractor. Neurons are sorted according to the latency of sample encoding. White dashed lines mark the sample and distractor trial periods.

(D) Same layout as in (C) for VIP neurons. See also Figure S6.

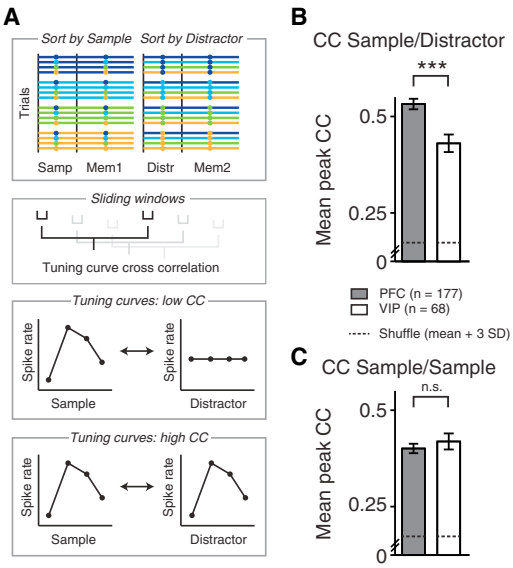


**Figure 6. Tuning Curve Cross-Correlations**

(A) Analysis schematic. Trials were sorted by sample in the sample and first memory period (colored lines) and by distractor in the distractor and second memory period (colored dots). In a sliding-window analysis, numerosity tuning curves were derived from time windows at equivalent positions in the first (sample sorted) and second half (distractor sorted) of the trial and cross-correlated. Low values indicate that selectivity differs for sample and distractor; high values signal that sample and distractor are represented similarly, i.e., neurons switched from coding the sample to representing the distractor.

(B) Mean peak CCs between sample and distractor tuning curves for PFC and VIP neurons that were sample selective in the sample and/or first memory period (n = 177 and n = 68, respectively).

(C) Same layout as in (B) for tuning curves derived for sample numerosities in all trial epochs. Error bars, SEM across neurons; ***$p < 0.001$; n.s., not significant. See also Figure S7.

or possibly coding strength (tuning curve width), we calculated CCs between tuning curves derived for the sample numerosity in all epochs. Importantly, no differences were found between PFC and VIP neurons (PFC: 0.40 ± 0.01, VIP: 0.42 ± 0.02, p = 0.62, Wilcoxon rank sum test, Figure 6C), excluding the possibility that unequal selectivity had biased our cross-correlation. Our analyses of single-neuron responses collectively suggest that sample and distractor representations shared significantly larger resources in PFC compared to VIP. Irrelevant distractor stimuli were readily encoded by single prefrontal neurons, whereas the distractor did not propagate well in area VIP.

## Sample Memory Strength in VIP Predicts Behavior

So far, we reported data from correct trials. To investigate to what extent the sample and distractor memory traces were correlated with successful completion of the task, we analyzed error trials (Figure 7). We first explored whether the strength of sample numerosity representation in either area predicted whether the animals would make errors (Figures 7A and 7C). To ensure a robust analysis, we selected neurons with a sufficient number of error trials from the population of sample- or distractor-selective cells described in Figures 5A and 5B (PFC: n = 182/239; VIP: n = 114/152; see Experimental Procedures). Information about the sample stimulus was quantified using $\omega^2$ explained variance as in Figures 2H and 2I. We averaged information across neurons in individual trial periods and compared correct with error trials in both areas. In PFC, sample information was identical in early trial periods and decreased to the same extent following presentation of the distractor (Figure 7A). The only difference between correct and error trials—and thus indicator of a behavioral mistake—was observed in the second memory period after distractor offset, where sample information

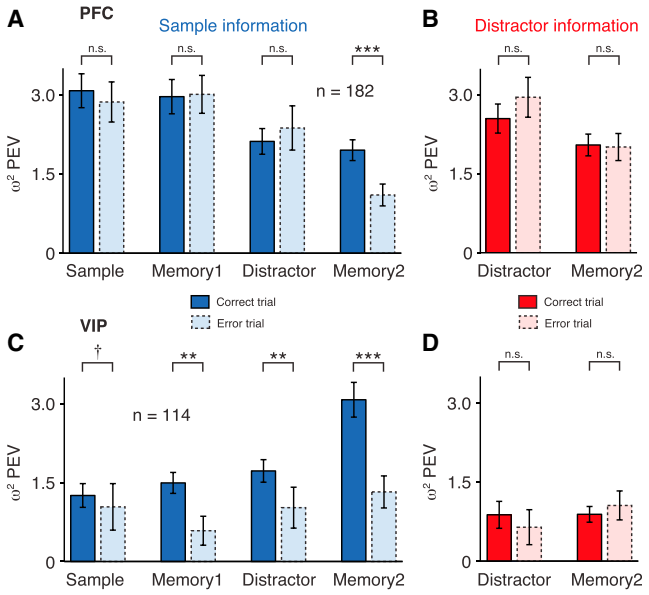


**Figure 7. Error Trial Analysis**

(A and B) $\omega^2$ PEV in the sample, first memory, distractor, and second memory periods for sample-selective PFC neurons that were recorded during at least four error trials for each sample and each distractor (n = 182). Data are presented for correct trials (saturated colors, solid outlines) and for error trials (unsaturated colors, dashed outlines).

(C and D) Same layout as in (A) and (B) for sample-selective VIP neurons (n = 114). Error bars, SEM across neurons; †$p < 0.1$; **$p < 0.01$; ***$p < 0.001$; n.s., not significant. See also Figure S8.

recovered less in error trials (second memory period: $\omega^2$ 1.95 ± 0.2 versus 1.1 ± 0.21 for correct and error trials, respectively, $p < 0.001$, Wilcoxon signed-rank test). In contrast, the strength of sample representation in VIP predicted behavioral success at a much earlier stage: the animals made errors when sample information failed to accumulate in the course of the trial, starting as early as during sample presentation (sample period: $\omega^2$ 1.26 ± 0.23 versus 1.04 ± 0.44, $p = 0.09$; first memory period: $\omega^2$ 1.5 ± 0.2 versus 0.58 ± 0.28, $p < 0.01$; distractor period: $\omega^2$ 1.72 ± 0.21 versus 1.02 ± 0.39, $p < 0.01$; second memory period: $\omega^2$ 3.08 ± 0.33 versus 1.32 ± 0.3, $p < 0.001$; Figure 7C). Similar results were obtained in an analysis of firing rates for preferred numerosities (Figure S8).

Finally, we determined whether the amount of distractor information was correlated with the rate of successfully completed trials (Figures 7B and 7D). Interestingly, the strength of distractor representation was identical in correct and error trials in both PFC and VIP. This suggests that the distractor was not suppressed to solve the task. In summary, lack of sample information in area VIP was the earliest predictor of forthcoming errors, but we did not observe a correlation between performance and filtering of the distractor in either PFC or VIP.

## DISCUSSION

We hypothesized that irrelevant distractor numerosities would be processed primarily by parietal neurons, whereas prefrontal neurons would largely suppress distractor information to solve the task. We report two main findings that violated this prediction. First, most PFC neurons did not resist interference, despite good behavioral filtering performance in both animals. Surprisingly, distractors affected working memory less in VIP. Second, neuronal representation of sample information, in particular in VIP, was the best predictor of behavioral outcome. We did not observe a correlation of distractor information with performance in either area.

### Lack of Prefrontal Distractor Suppression

In PFC, information about the interfering stimulus carried by neuronal population activity reached peak levels comparable to the sample memory (Figures 2 and 3). Many single prefrontal neurons represented both target and distractor memories (Figures 4A–4D, 5A, 5C, and 6B). Importantly, the strength of distractor representation in PFC was unrelated to the animals' performance (Figure 7B). These results differ markedly from those of recent electrophysiological studies showing that distractor suppression in PFC is tightly correlated with attentional filtering performance (Lennert and Martinez-Trujillo, 2011; Suzuki and Gottlieb, 2013). For example, in an oculomotor delayed-response task with a similar stimulus sequence as in the present protocol, saccade targets placed in the receptive fields of PFC neurons elicited strong responses, while responses were significantly weaker when distractors were presented at the same locations (Suzuki and Gottlieb, 2013).

Why was no suppression of the interfering stimulus observed here? In computational models of persistent working memory activity, protection from interference is only seen as long as external stimulation does not overwhelm the network; when stimulation intensity is too strong, distractors disrupt sample-related activity, and the network is able to maintain only a memory of the most recent stimulus (Brunel and Wang, 2001). Indeed, the numerical stimuli in our study represent highly abstract categories and strongly drive prefrontal neurons (Nieder et al., 2002; Nieder and Miller, 2004). In contrast, most previous experiments probed neuronal activity in the domain of spatial locations (Katsuki and Constantinidis, 2012; Lennert and Martinez-Trujillo, 2011; Qi et al., 2010; Suzuki and Gottlieb, 2013). The representation of locations in space is a typical feature of parietal cortex neurons (Chafee and Goldman-Rakic, 1998; Bisley and Goldberg, 2003, 2006). The use of spatial targets and distractors could have freed up resources in PFC, enabling this area to assume cognitive control functions such as suppression of interfering stimuli. Thus, filtering of complex features might require distinct processing in prefrontal and parietal cortex compared to the previously studied spatial filtering. It is also possible that the total number of stimuli presented to the monkeys (one sample and one distractor) was below the visual working memory capacity in PFC, but not VIP (Buschman et al., 2011), which could have reduced the need for prefrontal neurons to filter out distracting information.

Another common feature of previous studies was that animals maintained central fixation but covertly attended to and monitored the periphery for visual changes, e.g., the presentation of target or distractor stimuli (Rainer et al., 1998; Everling et al., 2002; Lennert and Martinez-Trujillo, 2011; Stokes et al., 2013;

Suzuki and Gottlieb, 2013). The degree of target/nontarget discrimination was therefore dependent on the locus of attention and the result of attentional filtering (Rainer et al., 1998; Everling et al., 2002). In contrast, we presented numerosities at the center of gaze, possibly putting more pressure on the animals to process the presented stimuli. This task design is shared by a previous report showing that prefrontal selectivity for sample images decreased markedly during presentation of task-irrelevant stimuli and was not superior to selectivity in inferior temporal cortex (ITC) (Miller et al., 1996). Contrary to ITC, however, target-selective activity in PFC returned after distractor offset, an observation we also made (Miller et al., 1996) (Figures 2, 3, 4C and 4D). Together, our results suggest that, beyond distractor suppression, the ability to regenerate target information following a strong interference is an equally important property of PFC that allows this area to guide behavior. This notion is in line with the concept of PFC as a flexible neuronal network for providing relevant information on demand (and not necessarily throughout the entire trial) (Stokes et al., 2013). Indeed, we found that target memory strength in the second memory period preceding the test stimulus was predictive of a successful trial (Figure 7A), indicating that sample information in PFC, delivered at the crucial moment, is required to solve the task. The source of this recovery is currently unclear. It could arise from residual sample information in PFC itself, which was reduced, but not lost completely (Figures 2H and 2J). Alternatively, relocation of memories could also occur from protected storage in VIP (see discussion below). Finally, even in correct trials, prefrontal target information did not exceed the strength of distractor representations prior to the test period (Figures 2 and 3). Thus, a crucial executive control operation performed by the PFC in this task period was likely stimulus selection (rather than suppression) for the upcoming response. Recent experiments have shown that oscillatory activity across neuronal populations could provide a means for disambiguating information about multiple items in working memory (Siegel et al., 2009; Dipoppa and Gutkin, 2013). Our results are therefore in good agreement with emerging evidence that mixed rather than pure selectivity for sensory information in prefrontal neurons is one of the fundamental principles of PFC organization (Rigotti et al., 2013).

### Working Memory for Sample Information in VIP

Compared to PFC, distractors had less impact on VIP processing. Information about the sample stimulus contained in VIP population activity was larger than for the distractor at all times (Figures 2 and 3). Very few single neurons encoded both the sample and the distractor, and distractor information did not propagate well in VIP neurons (Figures 4, 5, and 6). It is crucial to note that the measures investigating the impact of distractors on sample information are relative by necessity, i.e., determined within areas. Differences in the number of selective neurons or their selectivity (see next paragraph) preclude direct interarea comparisons and would otherwise bias the results.

Area VIP is an extensively connected multimodal association area in the parietal (dorsal) stream—medial to area LIP in the fundus of the intraparietal sulcus (IPS)—that responds to visual, somatosensory, vestibular, and auditory stimulation (Bremmer et al., 2002; Colby et al., 1993; Duhamel et al., 1998; Schlack et al., 2005). Several studies have shown that PPC is involved in representing visual working memory and complex abstract visual categories that emerge independently of feedback from PFC (Swaminathan and Freedman, 2012). VIP in particular is considered a major hub of numerical information processing in the primate brain, surpassed only by PFC regarding the number of selective neurons (Nieder and Miller, 2004; Nieder et al., 2006; Nieder, 2012; Viswanathan and Nieder, 2013). The proportion of numerosity-encoding neurons in VIP found here even exceeds previous reports. The importance of parietal target memory strength for solving the present task is clearly demonstrated by the fact that the amount of VIP—not PFC—sample information was the earliest reliable indicator of a successful trial and predicted the monkeys' choice even before the distractor was presented (Figure 7C).

The finding that distractors affected VIP to a lesser extent than PFC was unexpected. Parietal cortex (LIP) readily responds to attention-capturing distractor stimuli and is thought to function as a saliency map and novelty detector (Bisley and Goldberg, 2003, 2006; Suzuki and Gottlieb, 2013). However, the present results now suggest that in situations where PFC does not maintain target information throughout the trial, parietal cortex (VIP) might serve as a protected storage area instead. An interesting question for future exploration is whether PFC might be providing a gating signal to lower-level parietal areas to control access to working memory (Feredoes et al., 2011; McNab and Klingberg, 2008).

### Distinct Roles for Frontoparietal Cortex in Distractor Resistance

We can exclude that the animals failed to give preference to the sample over the distractor as an explanation for the present results, because, first, performance was above chance level in trials with interfering numerosities (Figure 1); second, sample (not distractor) information was predictive of behavior (Figure 7); and third, our findings are in accord with a previous report on prefrontal susceptibility to distractors (Miller et al., 1996). We can also exclude trial timing as a major factor (our task did not require sustained attention, since stimulus timing was not jittered and therefore known to the animals), because tasks using fixed timing have also produced distractor suppression in PFC (Everling et al., 2002).

Instead, we endorse the view that several mechanisms may be conceived of to explain how the brain identifies important targets in the face of distracting stimuli. Our data suggest that there is no universally optimal approach, but that the frontoparietal network adapts to current task demands and shuttles flexibly between strategies. While some distractors might be eliminated by suppression and attentional filtering (di Pellegrino and Wise, 1993; Lennert and Martinez-Trujillo, 2011; Qi et al., 2010; Suzuki and Gottlieb, 2013), we show here that they can also be efficiently bypassed by maintaining and recovering target information. Prefrontal and parietal cortex assumed distinct and specialized functions to accomplish these tasks. Our data support current models of working memory that emphasize active maintenance of stimuli in memory by activating their stored representations (in contrast to passively retained sensory traces) (Baddeley, 2012; Scott et al., 2012). Future studies should

therefore address the question of how PFC and PPC interact to create the observed selectivity patterns. Our experiments provide testable hypotheses regarding the degree and direction of functional coupling within the frontoparietal circuit that enables the primate brain to store, protect, and select behaviorally relevant information (Crowe et al., 2013).

## EXPERIMENTAL PROCEDURES

### Surgical Procedures

Two adult male rhesus monkeys (*Macaca mulatta*; monkey R and monkey W) were implanted with two right-hemispheric recording chambers centered over the principal sulcus of the lateral PFC and the VIP in the fundus of the IPS. All experimental procedures were in accordance with the guidelines for animal experimentation approved by the local authority: the Regierungspräsidium Tübingen.

### Task and Stimuli

The animals grabbed a bar to initiate a trial and maintained eye fixation (ISCAN, Woburn, MA) within 1.75° of visual angle of a central white dot. Stimuli were presented on a centrally placed gray circular background subtending 5.4° of visual angle. Following a 500 ms presample (pure fixation) period, a 500 ms sample stimulus containing one to four dots was shown. The monkeys had to memorize the sample numerosity for 2,500 ms and compare it to the number of dots (one to four) presented in a 1,000 ms test stimulus. Test stimuli were marked by a red ring surrounding the background circle. If the numerosities matched (50% of trials), the animals released the bar (correct match trial). If the numerosities were different (50% of trials), the animals continued to hold the bar until the matching number was presented in the subsequent image (correct nonmatch trial). Match and nonmatch trials were pseudorandomly intermixed. Correct trials were rewarded with a drop of water. In 80% of trials, a 500 ms interfering numerosity of equal numerical range was presented between the sample and test stimulus. The interfering numerosity was not systematically related to either the sample or test numerosity and therefore was not required to solve the task. In 20% of trials, a 500 ms gray background circle without dots was presented instead of an interfering stimulus, i.e., trial length remained constant (control condition, blank). Trials with and without interfering numerosities were pseudorandomly intermixed. Stimulus presentation was balanced; a given sample was followed by all interfering numerosities with equal frequency, and vice versa.

Low-level, nonnumerical visual features could not systematically influence task performance (Nieder et al., 2002): in half of the trials, dot diameters were selected at random. In the other half, dot density and total occupied area were equated across stimuli. CORTEX software (NIMH, Bethesda, MD) was used for experimental control and behavioral data acquisition. New stimuli were generated before each recording session to ensure that the animals did not memorize stimulus sequences.

### Electrophysiology

Up to eight 1 MΩ glass-isolated tungsten electrodes (Alpha Omega, Israel) per chamber and session were acutely inserted through an intact dura with 1 mm spacing. Stable and well-isolated neurons were recorded at random (Figure S4). To access VIP, electrodes were passed along the course of the IPS to a depth of 9–13 mm below the cortical surface (Nieder and Miller, 2004; Nieder et al., 2006; Vallentin et al., 2012). Prior to recording neuronal activity in VIP, correct positioning of the electrodes was ensured by physiological criteria (response to moving visual stimuli and tactile stimulation). Signal acquisition, amplification, filtering, and digitalization were accomplished with the MAP system (Plexon, Dallas, TX). Waveform separation was performed offline (Plexon).

### Data Analysis

Analysis was performed with MATLAB (Mathworks, Natick, MA). Unless specified otherwise, neurons were included in the analysis if the following criteria were met: first, their average firing rate across trials was at least 1 spike/s; second, they were recorded for at least 1 correct trial in all 20 conditions (4 sample numerosities × 5 interfering numerosities including the control [0] condition); and third, they modulated their firing rate in the course of the trial (task-related neurons, one-way ANOVA with average firing rates in the presample [fixation], sample, first memory, interfering stimulus, and second memory periods; evaluated at $p < 0.05$).

#### *Behavioral Data*

Behavioral tuning functions were used to describe the percentage of trials (y axis) for which a test stimulus (x axis, units of numerical distance to sample numerosity) was judged as being equal in number to the sample. For each condition and session, Gaussian curves were fitted to the tuning functions. The location of the mean (0) and the amplitude (percent correct trials) were fixed, and the Gaussian's width (sigma) was free. Differences in width to the control condition were averaged across sessions and tested for significance against zero median using Wilcoxon signed-rank tests.

For each condition and session, RTs were determined for match trials only (in nonmatch trials, the second test image following the nonmatch was always a match and therefore predictable). RT differences to the control condition were averaged across sessions and tested for significance against zero median using Wilcoxon signed-rank tests.

#### *Neuronal Information*

To quantify the information about the sample or interfering numerosity that was carried by a neuron's firing rate, we used the $\omega^2$ PEV measure (Buschman et al., 2011; Hentschke and Stüttgen, 2011; Puig and Miller, 2012). $\omega^2$ was derived from a categorical ANOVA and reflects how much of the variance in a neuron's firing rate can be explained by the numerosity of a particular stimulus (see Supplemental Experimental Procedures). Neurons were separately tested for sample and interfering numerosity PEV. Bin-wise Wilcoxon paired signed-rank tests (evaluated at $p < 0.05$ and $p < 0.01$) were used to compare sample and interfering numerosity PEV within the population of PFC and VIP neurons.

To determine at which point in the trial a neuron carried significant information about either the sample or interfering stimulus (Figure 5), we used a permutation test in a sliding-window analysis. For every analysis window (200 ms duration, 20 ms step), we created a null distribution of PEV values by randomly shuffling the association between firing rates and numerosities and calculating $\omega^2$ (repeated 1,000 times). The significance threshold was set to $p < 0.01$ (one-sided), i.e., the actual PEV value was required to be larger than 99% of values in the null distribution. To control for multiple comparisons, a neuron was said to significantly encode the sample and/or interfering stimulus if it crossed the respective thresholds for five consecutive windows. The onset of the first of these windows was taken to be the neuron's response latency. For the high-resolution comparison of sample selectivity latency in PFC and VIP (Figure S6), we restricted the analysis to the sample phase (50 ms window, 1 ms step, 25 consecutive windows).

For the analysis of $\omega^2$ PEV in error trials, we included neurons that were recorded during at least four error trials across each sample and interfering numerosity. For every neuron included, correct and error trial $\omega^2$ PEV values were averaged for the sample, first memory, interfering stimulus, and second memory periods and compared using a Wilcoxon paired signed-rank test.

In the same population of neurons, we performed an error trial analysis with firing rates to preferred numerosities, defined as the numerosity that elicited maximal firing in a 500 ms time window aligned to an individual neuron's sample selectivity latency. Firing rates in correct and error trials were normalized (maximum and minimum firing rate set to one and zero, respectively), averaged for the sample, first memory, interfering stimulus, and second memory periods and compared using Wilcoxon paired signed-rank tests. The analysis was performed separately in both monkeys.

#### *Stimulus-Selectivity Index*

We calculated a time-resolved SSI to determine whether a neuron's firing rate carried more information about the sample or interfering numerosity. The SSI was determined for all bins where a neuron significantly encoded the sample and/or interfering numerosity. Positive values indicate that the neuron's firing rate carries more information about the sample numerosity; negative values indicate that discharge rates vary more strongly with the interfering numerosity (see Supplemental Experimental Procedures). For illustration purposes, the SSI was normalized by the maximum of the absolute values of each neuron.

### *Tuning Curve Cross-Correlation*

Cross-correlations were calculated for neurons that were selective for the sample numerosity in the sample and/or first memory period as determined by a one-way ANOVA ($p < 0.01$). We sorted trials by sample numerosity in the sample and first memory period and by interfering numerosity in the interfering numerosity and second memory period (Figure 6A). Tuning curves were derived from time windows (200 ms duration, 20 ms step) at equivalent positions in the sample-sorted and interfering-numerosity-sorted part of the trial and cross-correlated as described previously (Diester and Nieder, 2007) (see Supplemental Experimental Procedures). The higher the CCs, the more a neuron switched from encoding the sample to representing the interfering numerosity. Shuffle predictors for each region were calculated by creating a null distribution of CCs by randomly shuffling the association between firing rates and numerosities (1,000 repetitions). Shuffled values were centered on zero (mean $< 10^{-3}$). Figures 6B and 6C show mean values plus three standard deviations (across neurons and time windows).

### *Factor Analysis*

Factor analysis was performed using MATLAB toolboxes (Yu et al., 2009). It extracts low-dimensional neuronal trajectories from noisy spiking activity and represents these in state space, where each data point corresponds to the instantaneous firing rate of the population of neurons at a given point in time. Firing rates in correct trials were processed using five latent dimensions, a bin width of 50 ms, and a Gaussian smoothing kernel of 50 ms. To obtain a comparable number of pseudosimultaneously recorded PFC and VIP neurons for the analysis of sample (interfering numerosity) trajectories, we included PFC neurons with at least 33 (24) trials per numerosity and VIP neurons with at least 19 (15) trials per numerosity. For each sample and interfering numerosity, we picked the minimum number of trials shared across neurons from one recording area and calculated averages across single-trial population trajectories. The Euclidean distance between two trajectories at corresponding time points was used as a measure of the difference in the population's activation state between individual numerosities. The distance in the presample (fixation) period was used as a baseline and subtracted from distances calculated at subsequent time points.

Classification of single trials was performed based on distances to the individual numerosities' mean trajectories. For each trial and time point, the distance between the test and each of the four reference trajectories was calculated. A trial was classified correctly if the test trajectory's numerosity was equal to the numerosity of the closest reference trajectory. Classification performance is expressed as the mean accuracy across trials (chance level 25%). Classification was performed using leave-one-out cross-validation, i.e., mean reference trajectories were calculated excluding the test trajectory (Harvey et al., 2012).

## SUPPLEMENTAL INFORMATION

Supplemental Information includes eight figures, one table, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2014.05.009.

## AUTHOR CONTRIBUTIONS

S.N.J. designed and performed experiments, analyzed data, and wrote the paper. A.N. designed experiments and wrote the paper.

## ACKNOWLEDGMENTS

This work was funded by grants from the Deutsche Forschungsgemeinschaft to S.N.J. (JA 1999/1-1) and A.N. (NI 618/4-1). S.N.J. was supported by the Charité Clinical Scientist Program funded by the Charité Universitätsmedizin Berlin and the Berlin Institute of Health.

Accepted: May 2, 2014
Published: July 2, 2014

## REFERENCES

Anderson, M.C., and Green, C. (2001). Suppressing unwanted memories by executive control. Nature *410*, 366–369.

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

Bisley, J.W., and Goldberg, M.E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. Science *299*, 81–86.

Bisley, J.W., and Goldberg, M.E. (2006). Neural correlates of attention and distractibility in the lateral intraparietal area. J. Neurophysiol. *95*, 1696–1717.

Bremmer, F., Klam, F., Duhamel, J.-R., Ben Hamed, S., and Graf, W. (2002). Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). Eur. J. Neurosci. *16*, 1569–1586.

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

Buschman, T.J., Siegel, M., Roy, J.E., and Miller, E.K. (2011). Neural substrates of cognitive capacity limitations. Proc. Natl. Acad. Sci. USA *108*, 11252–11255.

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

Chao, L.L., and Knight, R.T. (1995). Human prefrontal lesions increase distractibility to irrelevant sensory inputs. Neuroreport *6*, 1605–1610.

Chao, L.L., and Knight, R.T. (1998). Contribution of human prefrontal cortex to delay performance. J. Cogn. Neurosci. *10*, 167–177.

Colby, C.L., Duhamel, J.R., and Goldberg, M.E. (1993). Ventral intraparietal area of the macaque: anatomic location and visual response properties. J. Neurophysiol. *69*, 902–914.

Constantinidis, C., and Steinmetz, M.A. (1996). Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task. J. Neurophysiol. *76*, 1352–1355.

Crowe, D.A., Goodwin, S.J., Blackman, R.K., Sakellaridi, S., Sponheim, S.R., MacDonald, A.W., III, and Chafee, M.V. (2013). Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. Nat. Neurosci. *16*, 1484–1491.

Diester, I., and Nieder, A. (2007). Semantic associations between signs and numerical categories in the prefrontal cortex. PLoS Biol. *5*, e294.

di Pellegrino, G., and Wise, S.P. (1993). Visuospatial versus visuomotor activity in the premotor and prefrontal cortex of a primate. J. Neurosci. *13*, 1227–1243.

Dipoppa, M., and Gutkin, B.S. (2013). Flexible frequency control of cortical oscillations enables computations required for working memory. Proc. Natl. Acad. Sci. USA *110*, 12828–12833.

Duhamel, J.R., Colby, C.L., and Goldberg, M.E. (1998). Ventral intraparietal area of the macaque: congruent visual and somatic response properties. J. Neurophysiol. *79*, 126–136.

Everling, S., Tinsley, C.J., Gaffan, D., and Duncan, J. (2002). Filtering of neural signals by focused attention in the monkey prefrontal cortex. Nat. Neurosci. *5*, 671–676.

Feredoes, E., Heinen, K., Weiskopf, N., Ruff, C., and Driver, J. (2011). Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. Proc. Natl. Acad. Sci. USA *108*, 17510–17515.

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

Hentschke, H., and Stüttgen, M.C. (2011). Computation of measures of effect size for neuroscience data sets. Eur. J. Neurosci. *34*, 1887–1894.

Katsuki, F., and Constantinidis, C. (2012). Early involvement of prefrontal cortex in visual bottom-up attention. Nat. Neurosci. *15*, 1160–1166.

Lennert, T., and Martinez-Trujillo, J. (2011). Strength of response suppression to distracter stimuli determines attentional-filtering performance in primate prefrontal neurons. Neuron *70*, 141–152.

Luck, S.J., and Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. Nature *390*, 279–281.

Malmo, R.B. (1942). Interference factors in delayed response in monkeys after removal of frontal lobes. J. Neurophysiol. *5*, 295–308.

McNab, F., and Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. Nat. Neurosci. *11*, 103–107.

Miller, E.K., Erickson, C.A., and Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. J. Neurosci. *16*, 5154–5167.

Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. Proc. Natl. Acad. Sci. USA *109*, 11860–11865.

Nieder, A. (2013). Coding of abstract quantity by 'number neurons' of the primate brain. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. *199*, 1–16.

Nieder, A., and Miller, E.K. (2004). A parieto-frontal network for visual numerical information in the monkey. Proc. Natl. Acad. Sci. USA *101*, 7457–7462.

Nieder, A., Freedman, D.J., and Miller, E.K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. Science *297*, 1708–1711.

Nieder, A., Diester, I., and Tudusciuc, O. (2006). Temporal and spatial enumeration processes in the primate parietal cortex. Science *313*, 1431–1435.

Puig, M.V., and Miller, E.K. (2012). The role of prefrontal dopamine D1 receptors in the neural mechanisms of associative learning. Neuron *74*, 874–886.

Qi, X.-L., Katsuki, F., Meyer, T., Rawley, J.B., Zhou, X., Douglas, K.L., and Constantinidis, C. (2010). Comparison of neural activity related to working memory in primate dorsolateral prefrontal and posterior parietal cortex. Front. Syst. Neurosci. *4*, 12.

Rainer, G., Asaad, W.F., and Miller, E.K. (1998). Selective representation of relevant information by neurons in the primate prefrontal cortex. Nature *393*, 577–579.

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

Schlack, A., Sterbing-D'Angelo, S.J., Hartung, K., Hoffmann, K.-P., and Bremmer, F. (2005). Multisensory space representations in the macaque ventral intraparietal area. J. Neurosci. *25*, 4616–4625.

Scott, B.H., Mishkin, M., and Yin, P. (2012). Monkeys have a limited form of short-term memory in audition. Proc. Natl. Acad. Sci. USA *109*, 12237–12241.

Siegel, M., Warden, M.R., and Miller, E.K. (2009). Phase-dependent neuronal coding of objects in short-term memory. Proc. Natl. Acad. Sci. USA *106*, 21341–21346.

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

Suzuki, M., and Gottlieb, J. (2013). Distinct neural mechanisms of distractor suppression in the frontal and parietal lobe. Nat. Neurosci. *16*, 98–104.

Swaminathan, S.K., and Freedman, D.J. (2012). Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. Nat. Neurosci. *15*, 315–320.

Tudusciuc, O., and Nieder, A. (2007). Neuronal population coding of continuous and discrete quantity in the primate posterior parietal cortex. Proc. Natl. Acad. Sci. USA *104*, 14513–14518.

Vallentin, D., Bongard, S., and Nieder, A. (2012). Numerical rule coding in the prefrontal, premotor, and posterior parietal cortices of macaques. J. Neurosci. *32*, 6621–6630.

Viswanathan, P., and Nieder, A. (2013). Neuronal correlates of a visual "sense of number" in primate parietal and prefrontal cortices. Proc. Natl. Acad. Sci. USA *110*, 11187–11192.

Vogel, E.K., McCollough, A.W., and Machizawa, M.G. (2005). Neural measures reveal individual differences in controlling access to working memory. Nature *438*, 500–503.

Warden, M.R., and Miller, E.K. (2007). The representation of multiple objects in prefrontal neuronal delay activity. Cereb. Cortex *17* (*Suppl 1*), i41–i50.

Yu, B.M., Cunningham, J.P., Santhanam, G., Ryu, S.I., Shenoy, K.V., and Sahani, M. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. J. Neurophysiol. *102*, 614–635.

**Neuron, Volume *83***

# Supplemental Information

# Complementary Roles for Primate Frontal and Parietal Cortex in Guarding Working Memory from Distractor Stimuli

**Simon Nikolas Jacob and Andreas Nieder**

# Figure S1, related to Figure 2

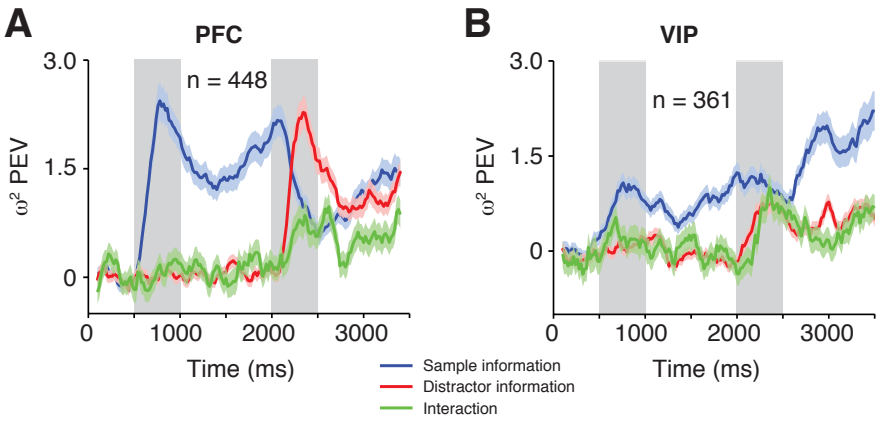


**Figure S1** Explained variance analysis using two-way ANOVA. (**A, B**) Sliding window percent explained variance ($\omega^2$ PEV) quantifying the information about the sample and distractor numerosity as well as the interaction term across task-related neurons recorded in prefrontal (A) and parietal cortex (B). Error bands, s.e.m. across neurons.

# Figure S2, related to Figure 3

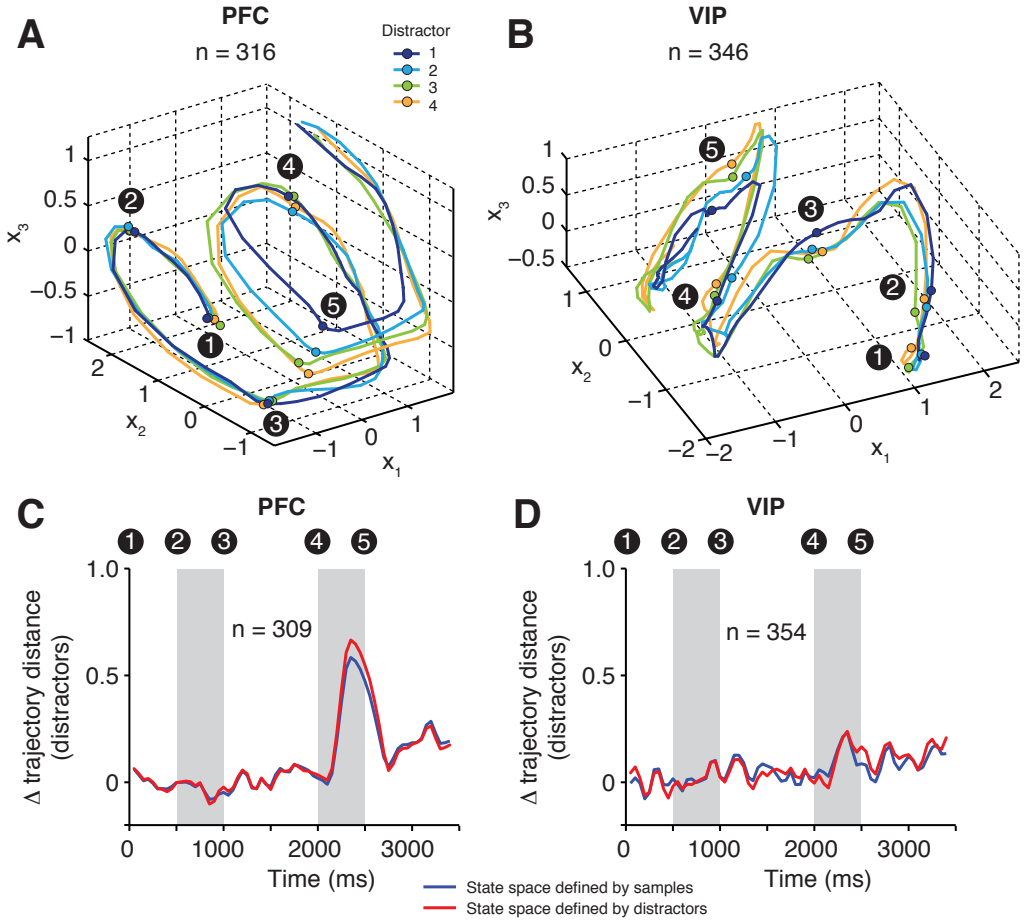


**Figure S2** Neuronal population trajectories for distractor numerosities. (**A**) Factor analysis describing the state space of neuronal population activity in PFC (n = 316) for each distractor across time, plotted for the first three common factors. Time points mark the onset of the (1) fixation (pre-sample), (2) sample, (3) first memory, (4) distractor and (5) second memory period. Trajectories represent the mean across single trials. (**B**) Same analysis for the population of VIP neurons (n = 346). (**C**) Mean inter-trajectory Euclidean distance across all distractor-distractor combinations in PFC as a measure of neuronal stimulus selectivity (same population as in Fig. 3A). Blue and red curves represent the distance calculated using the state space defined by sample and distractor numerosities, respectively. (**D**) Same layout as in (C) for VIP (same population as in Fig. 3A).

## Figure S3, related to Figure 3

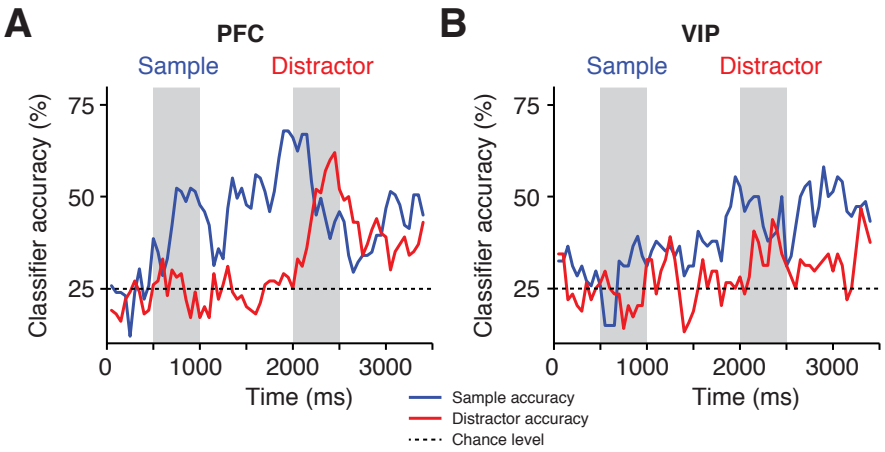


**Figure S3** Single-trial decoding of sample and distractor numerosities. (**A**) Time-resolved single-trial decoding of sample and distractor numerosities from PFC population activity (n = 309 neurons). Trials were sorted according to sample or distractor numerosity and subjected to factor analysis (Fig. 3). For each single trial trajectory and time bin, the distance to the 4 mean trajectories was calculated. Classification accuracy reflects the percentage of trials that were correctly classified as being closest to the actual sample or distractor trajectory (chance level P = 0.25). (**B**) Same layout for the VIP population (n = 354 neurons).

## Figure S4, related to Figure 4

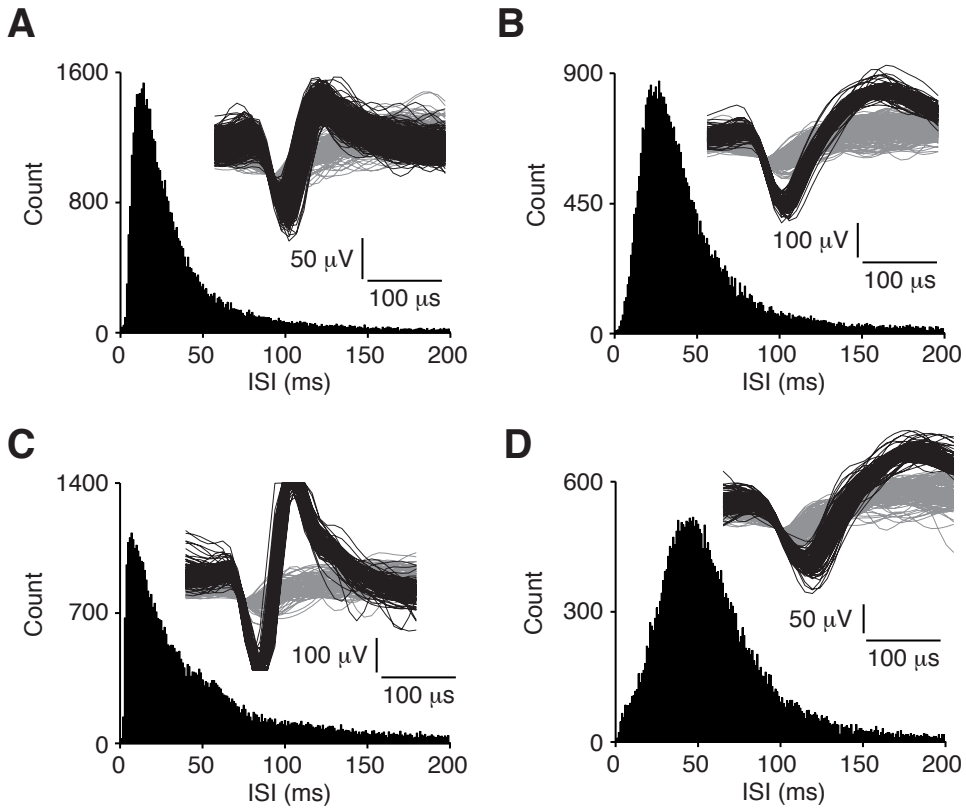


**Figure S4** Extracellular single-unit recordings. (**A**) Distribution of inter-spike intervals for the single neuron in Fig. 4A together with action potential waveforms extracted from a representative 16 s extracellular voltage trace (inset; noise marked in gray). (**B**) Same layout for the single neuron in Fig. 4C. (**C**) Same layout for the single neuron in Fig. 4E. (**D**) Same layout for the single neuron in Fig. 4G.

# Figure S5, related to Figure 4

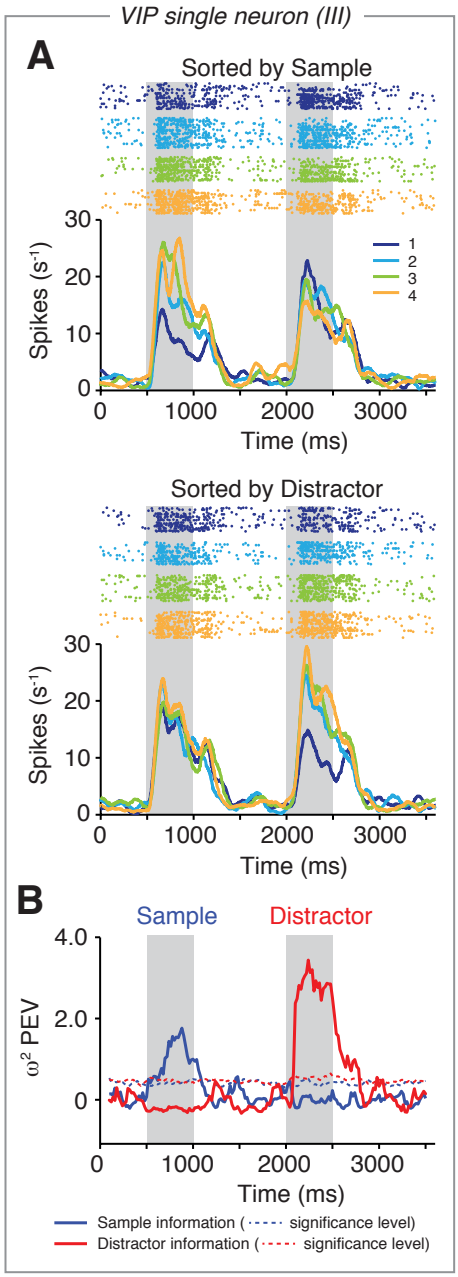


**Figure S5** Rare VIP neuron encoding both sample and distractor. (**A**) Raster plots and spike density histograms for a single VIP neuron. Trials are sorted according to sample (top panel) or distractor numerosity (bottom panel). (**B**) Sliding window percent explained variance ($\omega^2$ PEV) quantifying the information about the sample and distractor numerosity for the neuron in (A). Dashed lines mark the significance threshold (P = 0.01).

# Figure S6, related to Figure 5

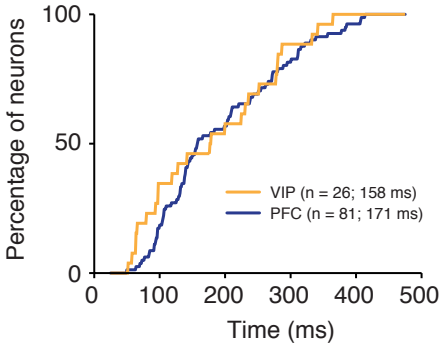


**Figure S6** Time course of frontoparietal numerosity selectivity. Cumulative latency distribution of PFC and VIP neurons determined by a sliding window percent explained variance ($\omega^2$ PEV) analysis quantifying the onset of sample information in the sample period.

# Figure S7, related to Figure 6

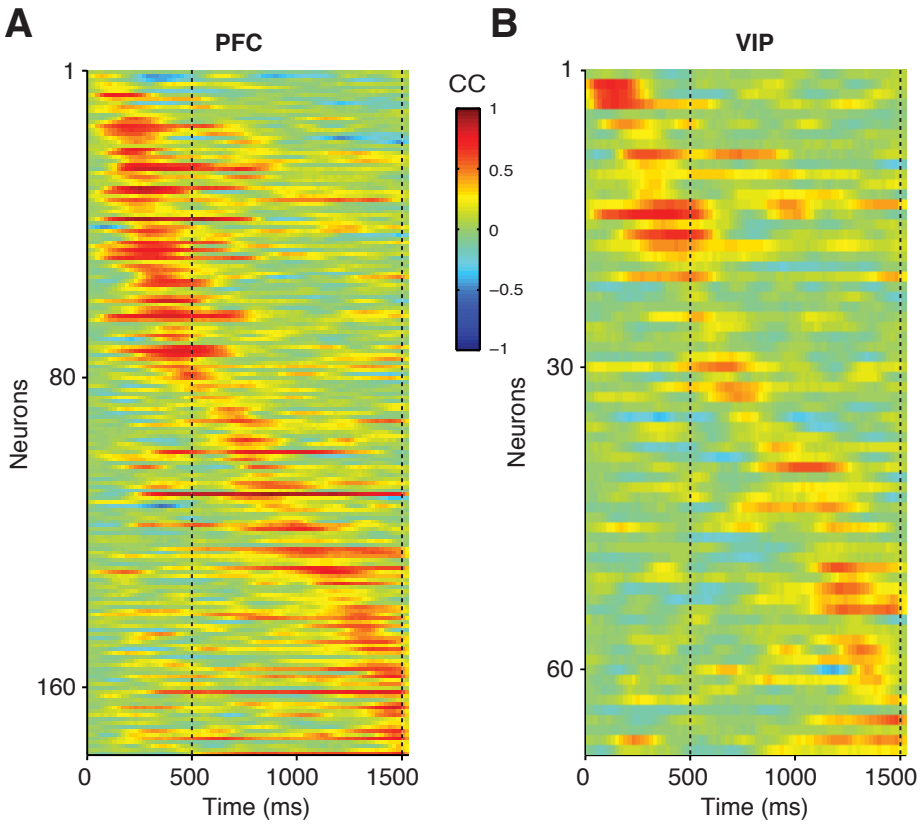


**Figure S7** Sliding window cross-correlation. (**A, B**) Cross-correlation coefficients (CC) between sample and distractor tuning curves, derived as explained in Fig. 6A, for PFC (A) and VIP (B) neurons that were sample selective in the sample and/or first memory period (n = 177 and n = 68, respectively). Neurons are sorted by time of maximal correlation. Data are smoothed by a five-point running average rectangular filter. The first dashed line represents the offset of the stimulus (sample/distractor) period, the second dashed line indicates the offset of the memory (first/second) period.

# Figure S8, related to Figure 7

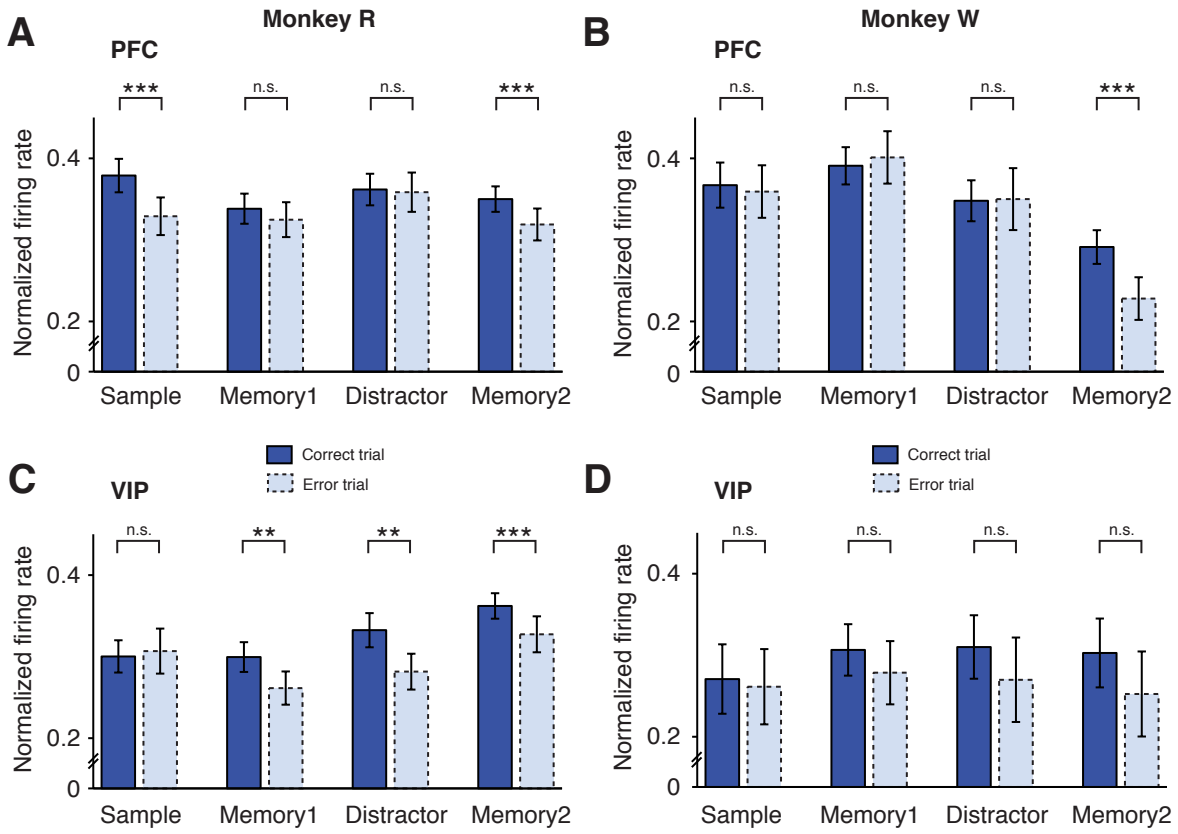


**Figure S8** Error trial analysis. (**A**, **B**) Normalized firing rates for preferred numerosities in the sample, first memory, distractor and second memory periods for the same population of sample-selective PFC neurons as in Fig. 7A plotted separately for monkey R (A) and monkey W (B). Data are presented for correct trials (saturated colors, solid outlines) and for error trials (unsaturated colors, dashed outlines). (**C**, **D**) Same layout as in (A, B) for the same sample-selective VIP neurons as in Fig. 7C. Error bars, s.e.m. across neurons; **, $P < 0.01$; ***, $P < 0.001$; n.s., not significant.

# Table S1, related to Figure 2

Table S1A. Number of task-related neurons showing selectivity in PFC.

| | Sample | Memory1 | Distractor | Memory2 |
|---|---|---|---|---|
| Main effect of sample | 88 | 98 | 43 | 73 |
| Main effect of distractor | | | 76 | 73 |
| Interaction s. x d. | | | 18 | 26 |

two-way ANOVA evaluated at $P < 0.01$

Table S1B. Number of task-related neurons showing selectivity in VIP.

| | Sample | Memory1 | Distractor | Memory2 |
|---|---|---|---|---|
| Main effect of sample | 27 | 36 | 24 | 64 |
| Main effect of distractor | | | 24 | 28 |
| Interaction s. x d. | | | 21 | 13 |

two-way ANOVA evaluated at $P < 0.01$

## SUPPLEMENTAL EXPERIMENTAL PROCEDURES

### Surgical procedures

Two adult male rhesus monkeys (*Macaca mulatta*, monkey R and monkey W) were implanted with a titanium head post and two right-hemispheric recording chambers centered over the principal sulcus of the lateral prefrontal cortex (PFC), anterior to the frontal eye fields, and over the ventral intraparietal area (VIP) in the fundus of the intraparietal sulcus guided by anatomical MRI and stereotaxic measurements. Chambers were angled to be able to penetrate the cortical surface perpendicularly. Surgery was conducted using aseptic techniques under general anesthesia. Structural magnetic resonance imaging to locate anatomical landmarks was performed before implantation. All experimental procedures were in accordance with the guidelines for animal experimentation approved by the local authority, the Regierungspräsidium Tübingen.

### Task and stimuli

The monkeys were trained to match visually presented non-symbolic set sizes (numerosities) while suppressing a salient task-irrelevant, interfering numerosity (Fig. 1A). The animals grabbed a bar to initiate a trial and maintained eye fixation within 1.75 ° of visual angle of a central white dot. An infrared-based eye tracking system monitored ocular position (ISCAN, Woburn, MA). Trials were immediately aborted and excluded from further analysis if the animals broke fixation. Stimuli were presented on a centrally placed gray circular background subtending 5.4 ° of visual angle. Following a 500 ms pre-sample (pure fixation) period, a 500 ms sample stimulus containing 1 to 4 dots was shown. The monkeys had to memorize the sample numerosity for 2,500 ms and compare it to the number of dots (1 to 4) presented in a 1,000 ms test stimulus. Test stimuli were marked by a red ring surrounding the background circle. If the numerosities matched (50 % of trials), the animals released the bar (correct Match trial). If the numerosities were different (50 % of trials), the animals continued to hold the bar until the matching number was presented in the subsequent image (correct Non-match trial). Match and non-match trials were pseudo-randomly intermixed. Correct trials were rewarded with a drop of water. In 80 % of trials, a 500 ms interfering numerosity of equal numerical range was presented between the sample and test stimulus. The interfering numerosity was not systematically related to either the sample or test numerosity and therefore not

required to solve the task. In 20 % of trials, a 500 ms gray background circle without dots was presented instead of an interfering stimulus, i.e. trial length remained constant (control condition, blank). Trials with and without interfering numerosities were pseudo-randomly intermixed. Stimulus presentation was balanced. That is, across a set of trials in which a given numerosity was used as the sample, it was followed by all interfering numerosities with equal frequency. Similarly, a given interfering numerosity was preceded by all sample numerosities with equal probability. Thus, when trials are sorted by sample (or interfering numerosity), any measure of neuronal activity is directly attributable to that stimulus, because the influence of the interfering numerosity (or sample) is factored out across trials.

Both animals had previously been trained to full proficiency in the standard delayed-match-to-numerosity task without task-irrelevant numerosities. Over the course of 3 to 5 months, the interfering numerosity was slowly introduced while carefully monitoring behavioral performance. The red ring surrounding the test stimuli ensured that the animals did not confuse individual stimuli and almost never responded prior to presentation of the test numerosities. During training, there were no abrupt changes in performance to suggest that the animals had switched response strategies. Neuronal recordings commenced when performance in both animals was stable over several weeks.

Low-level, non-numerical visual features could not systematically influence task performance (Nieder et al., 2002): in half of the trials, dot diameters were selected at random. In the other half, dot density and total occupied area were equated across stimuli. CORTEX software (NIMH, Bethesda, MD) was used for experimental control and behavioral data acquisition. All stimuli were produced using MATLAB (The Mathworks, Natick, MA) and generated anew before every recording session to ensure that the animals could not solve the task by memorizing stimulus sequences.

**Electrophysiology**

In each recording session, eight 1 MΩ glass-isolated tungsten electrodes (Alpha Omega, Israel) per chamber were acutely inserted through an intact dura with 1 mm spacing using custom-made screw microdrives. Brain tissue was allowed to settle for at least 30 minutes before recording. Stable and well-isolated neurons were recorded at random (Fig. S4); no attempt was made to preselect neurons according to particular response properties. To access area VIP at the fundus of the intraparietal sulcus (IPS), electrodes were passed along the course of the IPS to a depth of 9 to

13 mm below the cortical surface (Nieder and Miller, 2004; Nieder et al., 2006; Vallentin et al., 2012). Prior to recording neuronal activity in VIP, correct positioning of the electrodes was ensured by physiological criteria (response to moving visual stimuli and tactile stimulation). Signal acquisition, amplification, filtering and digitalization were accomplished with the MAP system (Plexon, Dallas, TX). Timestamps of trial events and action potentials were extracted for analysis. Waveform separation was performed offline using a combination of principal component analysis of waveform traces and other properties of the recorded waveforms (amplitude, peak/trough; Offline Sorter, Plexon).

**Data analysis**

Data analysis was performed with MATLAB. None of the reported analyses depended on the exact choice of trials to include or time windows to analyze. Repeating analyses with a different set of parameters yielded comparable results.

Unless specified otherwise, neurons were included in the analysis if the following criteria were met: first, their average firing rate across trials was at least 1 spike/s; second, they were recorded for at least 1 correct trial in all 20 conditions (4 sample numerosities x 5 interfering numerosities including the control [0] condition); and third, they modulated their firing rate in the course of the trial (task-related neurons, one-way analysis of variance (ANOVA) with average firing rates in the pre-sample (fixation), sample, first memory, interfering stimulus, second memory periods; evaluated at $P < 0.05$). A total of 448 neurons in PFC and 361 neurons in VIP fulfilled these criteria.

*Behavioral data*

Behavioral tuning functions were used to describe the percentage of trials (y axis) for which a test stimulus (x axis, units of numerical distance to sample numerosity) was judged as being equal in number to the sample. A numerical distance of 0 denotes match trials; the data point represents the percentage of correct trials. As the numerical distance increases, there is less confusion of the test with the sample numerosity; the data points represent the percentage of error trials. Tuning curves were calculated separately for trials without interfering numerosities (control (blank); Fig. 1C, D, H, I), for trials in which the sample and interfering numerosity were equal ('repeat sample' condition, Fig. 1C, H) and for trials in which the sample and interfering numerosity were different ('distractor' condition, Fig. 1D, I). For each condition and session, Gaussian curves were fitted to the tuning functions. The

location of the mean (0) and the amplitude (percent correct trials) were fixed, the Gaussian's width (sigma) was free. Differences in width to the control condition were averaged across sessions and tested for significance against zero median using Wilcoxon signed rank tests.

For each condition and session, reaction times (RT) were determined for match trials only (in non-match trials, the second test image following the non-match was always a match and therefore predictable). RT differences to the control condition were averaged across sessions and tested for significance against zero median using Wilcoxon signed rank tests.

*Neuronal information*

To quantify the information about the sample or interfering numerosity that was carried by a neuron's firing rate, we used the percent explained variance ($\omega^2$ PEV) measure (Buschman et al., 2011; Hentschke and Stüttgen, 2011; Puig and Miller, 2012). $\omega^2$ reflects how much of the variance in a neuron's firing rate can be explained by the numerosity of a particular stimulus. It was calculated using

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

where the individual terms are derived from a one-way categorical ANOVA (Buschman et al., 2011; Siegel et al., 2009): $SS_{Groups}$ 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. Balancing was accomplished by stratifying the number of trials in each group to a common value: A random subset of trials was drawn (equal to the minimum trial number across groups) 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. $\omega^2$ is an unbiased, zero-mean statistic when there is no information: when we recomputed the analysis after subtracting the shuffled PEV (Siegel et al., 2009; Puig and Miller, 2012), we obtained identical results (not shown). For every neuron and analysis window (200 ms duration, 20 ms step), $\omega^2$ shuffle was calculated after randomly shuffling the association between firing rate and numerosity. This process was repeated 1,000 times, and the shuffled PEV was taken to be the mean of the individual values. Neurons were separately tested for sample and interfering numerosity PEV. Bin-wise Wilcoxon paired signed rank tests

(evaluated at $P < 0.05$ and $P < 0.01$) were used to compare sample and interfering numerosity PEV within the population of PFC and VIP neurons. Repeating the analysis using a two-way categorical ANOVA with factors sample and interfering numerosity yielded identical results (Fig. S1).

To determine at which point in the trial a neuron carried significant information about either the sample or interfering stimulus (Fig. 5), we used a permutation test in a sliding window analysis. For every analysis window (200 ms duration, 20 ms step), we created a null distribution of PEV values by randomly shuffling the association between firing rates and numerosities and calculating $\omega^2$. This process was repeated 1,000 times. The significance threshold for the amount of information in any given window was set to $P < 0.01$ (one-sided), i.e. the actual PEV value was required to be larger than 99 % of values in the null distribution. To control for multiple comparisons, a neuron was said to significantly encode the sample and/or interfering stimulus if it crossed the respective thresholds for 5 consecutive windows. The onset of the first of these windows was taken to be the neuron's response latency. For the high-resolution comparison of sample selectivity latency in PFC and VIP (Fig. S6), we restricted the analysis to the sample phase (50 ms window, 1 ms step, 25 consecutive windows).

Because the animals performed significantly above chance level, we obtained considerably less error trials than correct trials. Therefore, we could not record any neurons that were present for at least 1 error trial in all possible trial conditions. For the analysis of $\omega^2$ PEV in error trials, we therefore relaxed the selection criteria and included neurons that were recorded during at least 4 error trials across each sample and interfering numerosity. 76 % (n = 182/239) of sample- or distractor-selective neurons in PFC and 75 % (n = 114/152) of sample- or distractor-selective neurons in VIP fulfilled these criteria. For every neuron included, correct and error trial $\omega^2$ PEV values were averaged for the sample, first memory, interfering stimulus and second memory periods and compared using a Wilcoxon paired signed rank test.

In the same population of neurons, we performed an error trial analysis with firing rates to preferred numerosities, defined as the numerosity that elicited maximal firing in a 500 ms time window aligned to an individual neuron's sample selectivity latency. Firing rates in correct and error trials were normalized (maximum and minimum firing rate set to 1 and 0, respectively), averaged for the sample, first memory, interfering

stimulus and second memory periods and compared using Wilcoxon paired signed rank tests. The analysis was performed separately in both monkeys.

*Stimulus-selectivity index*

We calculated a time-resolved stimulus-selectivity index (SSI) to determine whether a neuron's firing rate carried more information about the sample or interfering numerosity. The SSI was determined for all bins where a neuron significantly encoded the sample and/or interfering numerosity (see Neuronal Information). It was calculated by

$$SSI = \frac{SS_{Groups,\ Sample} - SS_{Groups,\ Distractor}}{SS_{Total}},$$

where the terms are derived from one-way ANOVAs: $SS_{Groups,\ Sample}$ and $SS_{Groups,\ Distractor}$ denote the sum-of-squares between groups (numerosities) when trials are sorted by sample or interfering numerosity, respectively, and $SS_{Total}$ is the total sum-of-squares of either ANOVA (identical values for sorting by sample or interfering stimulus). Positive values indicate that the neuron's firing rate carries more information about the sample numerosity; negative values indicate that discharge rates vary more strongly with the interfering numerosity. For illustration purposes, the SSI was normalized by the maximum of the absolute values of each neuron.

*Tuning curve cross-correlation*

Cross-correlation of neuronal tuning curves (Diester and Nieder, 2007) provided a measure to quantify the extent to which single neurons switched from encoding the sample to representing interfering numerosities. We sorted trials by sample numerosity in the sample and first memory period and by interfering numerosity in the interfering numerosity and second memory period (Fig. 6A). Tuning curves were derived from time windows (200 ms duration, 20 ms step) at equivalent positions in the sample-sorted and interfering-numerosity-sorted part of the trial and cross-correlated by calculating

$$CC = \frac{\sum_{n=1}^{4}\left(FR_{Sample}(n) - \overline{FR}_{Sample}\right) * \left(FR_{Distractor}(n) - \overline{FR}_{Distractor}\right)}{\sqrt{\sum_{n=1}^{4}\left(FR_{Sample}(n) - \overline{FR}_{Sample}\right)^2} * \sqrt{\sum_{n=1}^{4}\left(FR_{Distractor}(n) - \overline{FR}_{Distractor}\right)^2}},$$

where $FR_{Sample}(n)$ is the average firing rate for numerosity $n \in \{1,2,3,4\}$ when trials were sorted by sample, $\overline{FR}_{Sample}$ the average firing rate across all numerosities when

trials were sorted by sample, $FR_{Distractor}$ (n) the average firing rate for numerosity n ∈ {1,2,3,4} when trials were sorted by interfering numerosity, and $\overline{FR}_{Distractor}$ the average firing rate across all numerosities when trials were sorted by interfering numerosity. Thus, the higher the cross-correlation coefficients (CCs), the more a neuron switched from encoding the sample to representing the interfering numerosity.

CCs were calculated for neurons that were selective for the sample numerosity in the sample and/or first memory period as determined by a one-way ANOVA ($P < 0.01$; n = 177 in PFC, n = 68 in VIP). A random subset of trials (equal to the minimum trial number across numerosities, 8 trials at most) was drawn for each numerical value and time window (sorted as described above). Thus, trial numbers were equated for all neurons, both regions, and sample and interfering numerosities. Tuning functions were built with the averaged firing rates of these trials, and the CC was calculated. This process was repeated 25 times, and the overall statistic was taken to be the mean of the stratified values. CCs for PFC and VIP neurons were compared using Wilcoxon rank sum tests. To control for differences in cell counts or coding strength, we calculated CCs between tuning curves derived for the sample numerosity in all epochs using the same time windows as for the sample/interfering-numerosity CCs. Shuffle predictors for each region were calculated by creating a null distribution of CCs by randomly shuffling the association between firing rates and numerosities (1,000 repetitions). Shuffled values were centered on zero (mean < $10^{-3}$). Fig. 6B, C shows mean values plus 3 standard deviations (across neurons and time windows).

*Factor analysis*

Factor analysis was used to describe the temporal evolution of activity in a large ensemble of recorded neurons, i.e. the population dynamics. Factor analysis extracts low-dimensional neuronal trajectories from noisy spiking activity and represents these in state space, where each data point corresponds to the instantaneous firing rate of the population of neurons at a given point in time. The trajectory is the path that links the sequence of activation states at each time point (Stokes et al., 2013). Compared to PCA, factor analysis better captures shared fluctuations in activity across the population and more effectively discards independent variability specific to individual neurons (Harvey et al., 2012). Analysis was performed using MATLAB toolboxes (Yu et al., 2009). Firing rates in correct trials were processed using 5 latent dimensions, a bin width of 50 ms and a Gaussian smoothing kernel of 50 ms.

Optimal parameters were determined using cross-validation routines from the toolbox (Yu et al., 2009). To obtain a comparable number of PFC and VIP neurons for the analysis of sample trajectories, we included PFC neurons with at least 33 trials per sample numerosity (n = 309) and VIP neurons with at least 19 trials per sample (n = 354). For the analysis of interfering numerosity trajectories, we included PFC neurons with at least 24 trials per interfering numerosity (n = 316) and VIP neurons with at least 15 trials per interfering numerosity (n = 346). These neurons formed populations of pseudo-simultaneously recorded neurons. For each numerosity, we picked the minimum number of trials shared across neurons from one recording area and calculated single-trial population trajectories. Trajectories were then averaged across trials to represent a mean population response for the given numerosity. The analysis was performed separately for trials sorted by sample and by interfering numerosity. To determine how well the state space defined by sample numerosities captured variability in the neuronal responses to interfering numerosities, we repeated the analysis using the population of neurons used in the sample trajectory analysis (PFC: n = 309; VIP: n = 354) and calculated interfering numerosity trajectories in sample space.

The Euclidean distance between two trajectories at corresponding time points was used as a measure of the difference in the population's activation state between individual numerosities. For each pair, the distance in the pre-sample (fixation) period was used as a baseline and subtracted from distances calculated at subsequent time points.

Classification of single trials was performed based on distances to the individual numerosities' mean trajectories. Trials were sorted either according to sample or to interfering numerosity. For each trial and time point, the distance between the test and each of the 4 reference trajectories was calculated. A trial was classified correctly if the test trajectory's numerosity was equal to the numerosity of the closest reference trajectory. Classification performance is expressed as the mean accuracy across trials (chance level 25 %). Classification was performed using leave-one-out cross validation, i.e. mean reference trajectories were calculated excluding the test trajectory (Harvey et al., 2012).