

Integrated neural dynamics for behavioral decisions and attentional competition in the prefrontal cortex

Short title: Prefrontal neural dynamics and attentional control

Yaara Erez^{1,2,3*}, Mikiko Kadohisa², Philippe Petrov⁴, Natasha Sigala^{5,6}, Mark J. Buckley⁴,
Makoto Kusunoki², and John Duncan^{2,4}

1 Faculty of Engineering, Bar-Ilan University, Ramat-Gan 5290002, Israel

2 MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge CB2 7EF, UK

3 Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat-Gan 5290002,
Israel

4 Department of Experimental Psychology, University of Oxford, Oxford OX2 6GG, UK

5 Brighton and Sussex Medical School

6 Sackler Center for Consciousness Science, University of Sussex, Brighton BN1 9RR, UK

*Correspondence: yaara.erez@biu.ac.il (Y.E.)

Acknowledgements: This work was supported by MRC intramural programme MC_UU_00005/6, a research award 220020081 from the James S. McDonnell Foundation, and grant 101092/Z/13/Z from the Wellcome Trust. Y.E. was supported by a Royal Society Dorothy Hodgkin Research Fellowship DH130100. We thank Daniel Mitchell for useful comments and advice throughout this study.

Keywords: Cognitive control; Attention; Prefrontal cortex; Non-human primate; Population coding; Single units; Dynamics; Dimensionality reduction

ABSTRACT

In the behaving monkey, complex neural dynamics in the prefrontal cortex contribute to context-dependent decisions and attentional competition. We used demixed principal component analysis to track prefrontal activity dynamics in a cued target detection task. In this task, the animal combined identity of a visual object with a prior instruction cue to determine a target/nontarget decision. From population activity, we extracted principal components for each task feature, and examined their time course and sensitivity to stimulus and task variations. For displays containing a single choice object in left or right hemifield, object identity, cue identity and decision were all encoded in population activity, with different dynamics and lateralization. Object information peaked at 100-200 ms from display onset and was largely confined to the contralateral hemisphere. Cue information was weaker, and present even prior to display onset. Integrating information from cue and object, decision information arose more slowly, and was bilateral. Individual neurons contributed independently to coding of the three task features. The analysis was then extended to displays with a target in one hemifield and a competing distractor in the other. In this case, the data suggest that each hemisphere initially encoded the identity of the contralateral object. The distractor representation was then rapidly suppressed, with the final target decision again encoded bilaterally. The results show how information is coded along task-related dimensions while competition is resolved and suggest how information flows within and across frontal lobes to implement a learned behavioral decision.

ABBREVIATIONS

dPCA: demixed principal component analysis

T: target

D: Distractor

N: Neutral

PSTH: peri-stimulus time histogram

INTRODUCTION

Flexible and adaptive processing of information is a hallmark of goal-directed behavior. With a constantly changing environment, varying task demands determine how input information will be processed and selected, with a key role played by the prefrontal cortex (Norman and Shallice, 1980; Dehaene et al., 1998; Botvinick et al., 2001; Duncan, 2001; Miller and Cohen, 2001). In non-human primates, responses of individual frontal neurons are tuned to multiple task variables including cues, stimuli, categorical information, and decisions (Freedman et al., 2001; Wallis et al., 2001; Rigotti et al., 2013; Kadohisa et al., 2019). Tuning is adjusted by behavioral relevance, with information more strongly encoded when it is relevant to a current decision (Sakagami and Niki, 1994; Rainer et al., 1998; Freedman et al., 2001; Everling et al., 2002; Duong et al., 2019). Similar flexible and adaptive coding has been observed in the human frontal cortex using neuroimaging, with representation of behaviorally-relevant task events such as cues, stimulus information, categorical distinctions, and responses (Jiang et al., 2007; Li et al., 2007; Liu et al., 2011; Woolgar et al., 2011b, 2011a; Lee et al., 2013; Erez and Duncan, 2015; Jackson et al., 2016; Wen et al., 2019).

To organize task performance, multiple variables must be combined into the correct computational structure or control program (Machens et al., 2005; Rigotti et al., 2010; Duncan et al., 2020). Correct computational structures must be based on the animal's long-term learning, indicating how environmental and other variables can be combined for goals to be achieved. To analyze such structures, methods are needed to extract multiple variables from prefrontal population activity, and to study dynamics as these variables are appropriately combined (Machens et al., 2005; Kadohisa et al., 2013; Mante et al., 2013; Kobak et al., 2016; Hunt et al., 2018).

In a recent study from our group, Kadohisa et al. (Kadohisa et al., 2013) asked how prefrontal population activity in non-human primates develops when items compete for attention. In a cued target detection task, a cue at the start of each trial indicated the target object to be detected in a subsequent choice display. If a target was present in the choice display, the animal awaited a subsequent go signal, then made a saccade to the target location. Importantly, some displays contained two items presented in opposite hemifields, the target on one side and a nontarget on the other. For these two-object displays, measures of population activity showed intriguing dynamics. Initially, responses reflected the item presented contralaterally to the recorded hemisphere, with activity resembling the response to the contralateral item presented alone. Thus, information concerning the ipsilateral item – clearly visible when this item appeared alone – disappeared in the bilateral display. Other tasks have revealed a similar contralateral dominance in both prefrontal (Duong et al., 2019) and inferotemporal (Chelazzi et al., 1998) neurons. In line with well-established normalization models of attention (Reynolds et al., 1999; Lee and Maunsell, 2009; Reynolds and Heeger, 2009), these results suggest that, when two objects are present in the visual field, they compete to drive neural responses. Later, however – in the period leading up to the go signal – activity approached response to the most relevant item (target) presented alone, whether contralateral or ipsilateral. Dominance by the target was achieved more slowly when the nontarget was itself a target on other trials, consistent with long-established behavioral findings showing that such nontargets are hard to ignore (Shiffrin and Schneider, 1977).

In a further analysis of data from just the single-object displays of this study, Kadohisa et al. (Kadohisa et al., 2015) considered two analysis windows, early (50-250 ms from display onset) and late (300-500 ms). In the early window, strong responses to a contralateral stimulus were

accompanied by selectivity of single cells for its identity and target/nontarget category. Later, category information extended to encompass both contralateral and ipsilateral stimuli.

In this task, following the presentation of the choice display, cue and object information are combined to form a final decision (Mante et al., 2013; Stokes et al., 2013; Duong et al., 2019). Here, we examine the same data set to unravel detailed population coding dynamics of each task feature for each type of display. To this end we use a recently developed dimensionality reduction technique, demixed principal component analysis (dPCA) (Kobak et al., 2016). We ask whether information coding along the task-related dimensions exhibits distinct, temporally variable trajectories; how these are modulated by the hemifield of the presented objects and their behavioral priority; and whether coding of different task features is driven by separate or overlapping neuronal sub-populations.

We find that all three task features are represented in prefrontal activity, with different, characteristic time courses. Task features also differ in lateralization; while identity is strongly coded only for objects in the contralateral hemifield, the decision code is bilateral, suggesting exchange of information between hemispheres. Matching common findings of heterogeneous responses in prefrontal neurons (Rigotti et al., 2013), single neurons are not specialized for coding different task features. Instead, across the neural population, distributions of coding strength are independent for different features. Our analyses also illuminate attentional competition beyond the findings of Kadohisa et al. (Kadohisa et al., 2013) where overall differences in population coding were demonstrated but without detailed investigation of separate task features. The data suggest that, with a target in one hemifield and a distractor in the other, each object identity is initially coded in the contralateral hemisphere. Distractor identity is then rapidly suppressed, with the target

decision finally reflected in both hemispheres. The results suggest a schematic model of information flow within and between frontal lobes, implementing a prefrontal control program for a learned behavioral decision.

MATERIALS AND METHODS

Subjects

Subjects were two male rhesus monkeys (*Macaca mulatta*) weighing 11 (monkey A) and 10 kg (monkey B). All the experimental procedures were conducted in accordance with the Animals (Scientific Procedures) Act 1986 of the UK. All the procedures were in compliance with the guidelines of the European Community for the care and use of laboratory animals (EUVD, European Union directive 86/609/EEC) and were licensed by a Home Office Project License obtained after review by Oxford University's Animal Care and Ethical Review committee.

Task

A cued target detection task was used (Figure 1A) (Kadohisa et al., 2013, 2015). Based on training prior to recording, two cues (cue 1, cue 2) were associated with one target object each (object 1, object 2), and a third neutral object (object 3) was not associated with any cue. Figure 1A shows the cue-target pairs for monkey A; different items were used for monkey B. A red dot at the center of the screen marked the start of a trial; the monkey was required to fixate throughout the trial until the saccadic response at the end (fixation window: 5° x 5° for monkey A, 4° x 4° for monkey B). Following fixation for 1,000 ms, a cue stimulus (2° x 2°) was presented at the center of the screen for 500 ms, indicating the target for the current

trial, followed by a randomly varying delay (400-600 ms for monkey A and 400-800 ms for monkey B). Next, a choice display was presented (500 ms), including either a single object or two objects. Objects ($2^\circ \times 2^\circ$) were always centered on the horizontal meridian. In single-stimulus displays, the object was presented 6° to the left or right of fixation, determined randomly. The position of the objects was therefore either contralateral or ipsilateral to the recorded hemisphere. The object could be either a target (T) if associated with the preceding cue, a distractor (D) if associated with the other cue, or a neutral stimulus (N) if not associated with any cue and therefore never a target (object 3). To test for attentional competition, two-stimulus displays contained one stimulus in each hemifield, with combinations of target and neutral objects (T + N, low competition), target and distractor objects (T + D, high competition), and distractor and neutral objects (D + N, target absent trials). Objects were presented 6° to the left and right of fixation, and the right-left or left-right configuration was randomly determined. To avoid response bias, the frequencies of the choice display types were adjusted to include a target object in half of all single-stimulus and half of all two-stimulus displays. The frequencies of the main display types were otherwise the same. A trial was terminated without reward if there was a premature saccadic response outside the fixation window during the trial, and these trials were excluded from the analysis (Kadohisa et al., 2013, 2015).

The choice display was followed by an additional delay period (randomly varying, 100-150 ms for monkey A, 300-500 ms for monkey B). The fixation dot then turned green, indicating the go signal and response interval. For trials that included a target in the display (target present), the monkey had to saccade to the remembered T location (target window: $6^\circ \times 6^\circ$ for monkey A, $3.5^\circ \times 3.5^\circ$ for monkey B) and was rewarded immediately for a correct saccade with a drop

of liquid. For trials that did not include a target in the display (target absent), the monkey had to fixate for the whole response interval (1,000 ms) and was then rewarded either for a further saccade (monkey A) or immediately (monkey B).

For monkey A, cues varied randomly between trials in some sessions, while other sessions included alternating short blocks of fixed cues (15-20 trials per block). Data from these two types of sessions were combined (Kadohisa et al., 2013). For monkey B, cues always varied randomly between trials.

Overall, the task included six main choice display types (three single-stimulus and three two-stimulus displays), and a total of 24 conditions defined by combinations of cue, display type and hemifield (Kadohisa et al., 2013). Twelve single-stimulus conditions arise from combination of 2 cues, 3 objects and 2 hemifields. Twelve two-stimulus conditions arise from 2 cues, 3 object combinations (T + D, T + N, D + N) and 2 spatial arrangements. Single-stimulus N trials were not included in the analysis, along with a small number of D + D trials (the same D stimulus appearing in both hemifields) present in some sessions.

Recordings

Each monkey was implanted with a custom-designed titanium head holder and recording chamber(s) (Max Planck Institute for Biological Cybernetics, Tübingen, Germany), fixed on the skull with stainless steel screws. Chambers were placed over the lateral prefrontal cortex of the left (AP = 25.3, ML = -20.0; AP, anterior-posterior; ML, medio-lateral) and right (AP = 31.5, ML = 22.5) hemispheres for monkey A and the right hemisphere (AP = 30.0, ML = 24.0) for monkey B (Kadohisa et al., 2013, 2015). Recording locations (Figure 1B) were in dorsal and ventral regions of the posterior lateral prefrontal cortex, around the posterior third of the

principal sulcus for each animal. Under each chamber, a craniotomy was made for physiological recording. All surgical procedures were aseptic and carried out under general anesthesia. We used arrays of tungsten microelectrodes (FHC, Bowdoin, ME, USA) mounted on a grid (Crist Instrument, Hagerstown, MD, USA) with 1 mm spacing between adjacent locations inside the recording chamber. The electrodes were independently controlled by a hydraulic, digitally controlled microdrive (Electrodes Drive, NAN Instruments, Nof Hagalil, Israel, for monkey A; Multidrive 8 Channel System, FHC for monkey B). Neural activity was amplified, filtered, and stored for offline cluster separation and analysis with the Plexon MAP system (Plexon Inc, Dallas, Texas, USA). Eye position was sampled using an infrared eye tracking system (120 Hz, ASL, Bedford, MA, USA, for monkey A; 60 Hz, Iscan, Woburn, MA, USA, for monkey B) and stored for offline analysis. Data were recorded over a total of 140 daily sessions. Before starting the task, microelectrodes were advanced until neuronal activity could be isolated. Neurons were not preselected for task-related responses. At the end of the experiments, animals were deeply anaesthetized with barbiturate and then perfused through the heart with heparinized saline followed by 10% formaldehyde in saline. The brains were removed for histology and recording locations were confirmed on dorsal and ventral frontal convexities and within the principal sulcus.

Data and analysis

Only data from successfully completed trials (fixation held until go signal) were analyzed, and only units with a minimum of four trials in each condition were included in the analysis. All statistical analyses were conducted using MATLAB (MathWorks Inc.). We used customized code for the analysis and the demixed PCA toolbox (Kobak et al., 2016) to construct the low-dimensionality subspaces.

For all analyses, spike data of each unit were first smoothed with a Gaussian kernel of SD=20 ms, cutoffs ± 1.5 SD. This kernel is consistent with previous papers that analyzed this dataset (Kadohisa et al., 2013, 2015) and was chosen as a balance between temporal precision and noise reduction. The analysis focused on the choice display epoch of a trial as during this epoch the cue and stimulus information are combined to reach a decision. We therefore used data from each trial from -200 ms to 600 ms from choice display onset, throughout which the animals kept fixation. To avoid over-weighting of neurons with high spike rates, data for each neuron were normalized across all trials, time points and conditions by subtracting the mean and dividing by SD prior to any analysis. The normalized data were then used for the dPCA.

Low-dimensional neural state space

For any given condition and time point, neural activity can be described as a point in an N-dimensional space, where N is the number of recorded neurons. In principal component analysis (PCA), the aim is to reduce dimensionality; activity is now described in an X-dimensional subspace, $X \ll N$, with each dimension or axis of the subspace defined by a weighted sum of activity in multiple neurons. Dimensionality reduction is useful if neurons share firing properties to some extent, e.g., strong response to a given stimulus, or similar time-courses of activity.

We used a recently developed dimensionality reduction technique, demixed PCA (dPCA), which is described in detail in Kobak et al. (Kobak et al., 2016). As usual in PCA, it decomposes the high dimensional space of population activity into a small number of components. Importantly, and unlike a standard PCA, the components identified by dPCA are task-related, i.e., correspond to pre-defined task variables such as stimulus, decision and cue, while the error is minimized between the reconstructed signal and average activity for each level of the

different task variables. The resulting compressed subspace both captures the majority of variance in the data, and critically, it aims to keep the task-related components demixed, such that each component captures variance that is related mostly to one task feature only (such as object and decision). The method is thus useful for separating and tracking neural representation of each task feature.

We applied dPCA to construct low-dimensionality subspaces capturing the three task variables (cue, object identity, and decision), using the four single-stimulus T and D displays that cross these variables: cue1-object1-T, cue2-object2-T, cue2-object1-D, cue1-object2-D. Note that, in this task, the target/nontarget decision is confounded with motor output (saccade or hold fixation). However, we conducted the analysis on time windows during which fixation was required and there was no eye movement. Furthermore, as the data suggested no activity specifically related to the eye movement (Kadohisa et al., 2013), we refer to this variable as target/nontarget decision rather than movement preparation. For the dPCA, trial data for the four conditions could be grouped according to object identity (object 1 or 2), decision (T or D) and cue (cue 1 or 2, the interaction of object and decision) to allow the extraction of the relevant components. Because strong hemifield effects have been previously demonstrated in the prefrontal cortex (Ungerleider et al., 1989; Chelazzi et al., 1998; Buschman et al., 2011; Kadohisa et al., 2013, 2015; Matsushima and Tanaka, 2014), we constructed separate subspaces for the contralateral and ipsilateral stimulus displays.

Calculation of explained variance by each of the dPCA components was conducted using the dPCA toolbox and is described in detail in Kobak et al. (Kobak et al., 2016). Briefly, explained variance was computed in a standard way as the fraction of variance explained in peri-stimulus time histograms (PSTHs), subtracting the unexplained variance (difference between

the PSTHs and the reconstructed signal) from the total variance of the PSTHs and dividing by the total. Using the PSTHs rather than single trial data was required because data were sequentially recorded in sessions over different days. The explained variance was further decomposed into the contribution of the different task variables by using the marginalized PSTH across these variables.

Population coding of task-relevant information

We tracked the dynamics of cue, object and decision information within the task-related subspace and how it evolves when two items compete for attention. To investigate these temporal trajectories, we projected the data onto the extracted subspaces. For each task condition (each combination of cue + stimulus), these projections show movements along each axis of the subspace. Comparing movements along the axis for different conditions shows the time-course of representing each task feature. To ensure the generalizability of the compressed subspaces and avoid over-fitting, we employed a split-half cross-validation approach. We split the data for each neuron and condition into two halves (odd and even trials, alternating through the session). Within each half, data for each condition were averaged across trials in 1 ms bins to create a PSTH. We then computed separate subspaces for each half, and projected data from each half onto the subspace constructed from the other half. Results from these two cross-validated projections were then averaged. Signs of neuronal weights for each component were reversed when required to ensure that all the four subspaces (two halves, two hemifields) were compatible.

Statistical analyses

To measure coding of task-relevant information across the subspace, we subtracted the averaged projections of one level of each task variable on the corresponding component from the averaged projections of the other level. For example, for object information, the average projection of object 2 (across the two cues/decisions) on the object component was subtracted from the average projections of object 1 (across the two cues/decision). Similar measures were obtained for decision information (T present – T absent) and cue information (cue 1 – cue 2). Information was computed for each time point (1 ms bins) and averaged across subspaces for the two halves of the data.

We then used permutation analysis to determine statistical significance of these information measures, with cluster size correction for multiple comparisons across time points. For each comparison, a null distribution was generated by permuting condition labels for the trials within each half of the data (odd and even trials) then repeating exactly the same procedure as the main test: computing PSTHs, projecting the PSTHs onto the relevant component in the other half subspace, computing the difference using the appropriate contrast to measure coding and averaging across halves. This was repeated 1000 times. To estimate the distribution of cluster sizes expected by chance, we used a leave-one-iteration out approach. For each iteration, we marked time points at which the measured difference was greater than the value in all remaining iterations, then selected the cluster with the maximal number of such time points. To establish significance in the real data, we first marked time points at which the measured difference in the real data was greater than the values in all the permuted data and considered as significant only clusters larger than all maximum clusters in the permuted data (i.e., $\alpha = 0.001$, one-tailed). Permutation analysis to compare information

in two-stimulus and single-stimulus displays was done in a similar way, permuting condition labels between the appropriate conditions for each comparison. Because the subspaces and statistical tests were done using split-half cross-validation, comparisons between any two or more conditions were not prone to biases depending on whether they were used to construct the subspace (i.e., single-stimulus displays) or not (i.e., two-stimulus displays).

Correlations of neuronal weights

We used correlations across the neural population to test for dependence between neuronal dPCA encoder weights between subspaces and axes. For dependence between axes (objects, decision, and cue), we averaged the component weights across the subspaces of the two halves of the data, rectified them, and correlated all three possible pairs of axes using Spearman's rank correlation, for each hemifield subspace separately. The weights were rectified because the measure of interest for this analysis is the relative contribution of each neuron as reflected in the absolute weight, rather than its preference as reflected in the sign (e.g., preference for object 1 or object 2). Accordingly, a Spearman's rank correlation was used because of the rectification of the weights. To test for the dependence of representations across the contralateral and ipsilateral subspaces, for each subspace we averaged weights from the two halves of the data, and Pearson-correlated them across the two hemifields. Similarly, to test for the reliability of each set of neuronal weights within a subspace, we used Pearson's correlation between weights extracted from the two independent halves of the data.

RESULTS

For single-stimulus displays, accuracy levels were high for target (average 84%; 86% monkey A, 83% monkey B) and neutral trials (average 85%; 83% monkey A, 87% monkey B) (Kadonaga et al., 2013, 2015). The relatively high difficulty level for distractor trials (which may be targets on other trials) was evident in a much lower accuracy (average 60%; 56% monkey A, 63% monkey B). Most (72%) of the errors in distractor trials were a saccade to the stimulus location following the go signal, further confirming the difficulty of ignoring distractors. Accuracy levels were similar for T + D (average 76%; 70% monkey A, 82% monkey B) and T + N (average 77%; 77% monkey A, 76% monkey B) displays, but while many errors in the T + D displays were saccades to the distractor location (59%), the most common error for T + N displays was to keep fixation (79%). Accuracy levels for the D + N displays were slightly lower (average 67%; 60% monkey A, 73% monkey B), as expected when targets are absent in visual search.

Neurons were recorded on the lateral frontal surface across three hemispheres of the two monkeys. All analyses of neurophysiological data were based just on correct trials. The analysis included data from 337 neurons that had a minimum of four correct trials per condition for all conditions (Monkey A: N = 140 right, N = 71 left; monkey B: N = 126 right). The average number of trials per condition was 14.5.

[Figure 1 about here]

Low-dimensional state space captures coding of task-relevant variables across the neural population

Using data from 4 single-stimulus conditions (cue1-object1-T, cue2-object2-T, cue2-object1-D, cue1-object2-D; see Materials and Methods), we applied dPCA to construct low-dimensional subspaces capturing coding of the three task variables (cue, object identity, and decision; examples of responses of single cells are shown in Supplementary Figure 1). Because we expected somewhat different responses to contralateral and ipsilateral stimuli (Kadohisa et al., 2015), separate subspaces were created for stimuli in each hemifield.

For each of the four subspaces (2 hemifields, 2 halves), scree plots showing variance explained by each of the first 20 components are shown in Supplementary Figure 2. As expected, each early component captured variance primarily related to a single variable – time, object, decision or cue – with ‘time’ components capturing the time-course of activity across the analysis period, irrespective of other variables (Kobak et al., 2016). For the contralateral subspace, across the top 20 components, the object, decision and cue explained 7.3%, 7.7%, and 4.9% of the variance, respectively, of which 41%, 35%, and 38% was explained by the first component that was assigned to each of the variables. For the ipsilateral subspace, across the top 20 components, the object, decision and cue explained 6.2%, 8.3%, and 5.5% of the variance, respectively, of which 32%, 50%, and 40% was explained by the first component that was assigned to each of the variables. On average, the first matching component of each of the three task variables (object, decision and cue) explained 2.14 times more variance than the second component associated with the variable. Therefore, for simplicity, for each hemifield, we constructed just a three-dimensional subspace with axes comprised of the first object, decision and cue components.

To investigate the temporal trajectory of information coding for each of the task variables, we projected population responses onto each component (axis) of the subspace. Critically, the projection on each component in the subspace reflects coding along this component, *demixed* from coding of other task variables. To avoid over-fitting, data from each half of the trials were projected onto the subspace of the other half, and projections were averaged across halves. For any given task condition, e.g., cue1-object1-T, each projection shows the weighted mean response of neurons contributing to the corresponding principal component. As weights can be positive or negative, projections can also be positive or negative. For example, on the object component, positive and negative projections reflect coding of object 1 and 2, respectively. To calculate projections for stimuli on each side, contralateral or ipsilateral, we used components of the corresponding subspace. Thus, projections for contralateral stimuli show movements within the contralateral-stimulus subspace, and vice versa for ipsilateral stimuli.

For the object axis, cross-validated projections for the four critical single-stimulus conditions (cue1-object1-T, cue2-object2-T, cue2-object1-D, cue1-object2-D) are shown in Figure 2A. For contralateral stimuli (left), there was clear separation of object 1 and object 2 for both target and distractor conditions, as expected for a component designed to maximize this separation. Beyond this expected separation, the object component reveals the temporal progression of object information coding. Time points at which this separation was statistically significant are shown by the horizontal gray line. Significant coding started from around 50 ms post stimulus onset with a peak around 140 ms which then dropped and remains sustained throughout the trial. Object coding was much weaker for ipsilateral stimuli (Figure 2A, right),

with significant coding from around 100 ms until just after 340 ms, but without the clear peak shown by contralateral stimuli.

[Figure 2 about here]

For these same single-stimulus displays, a strikingly different temporal trajectory was observed for decision coding (Figure 3A). For both contralateral and ipsilateral stimuli, the data showed gradual development of the target-distractor distinction, beginning around 100 ms and 250 ms, respectively, and increasing over time. If anything, target present/absent separation was stronger for ipsilateral stimuli, probably because, for contralateral stimuli, the selected dPCA component captured a smaller part of the decision-related activity (see above).

[Figure 3 about here]

Lastly, projections onto the cue axis revealed some weak cue coding (Figure 4A). Cue information was stronger for a contralateral stimulus, beginning around stimulus onset. As the cue was the same whether the subsequent stimulus was contralateral or ipsilateral, and was known before stimulus onset, a tentative interpretation is reinstatement of a cue signal as a contralateral stimulus is processed. For ipsilateral stimuli, cue coding was only significant for a brief period around stimulus onset.

[Figure 4 about here]

Altogether, the single-stimulus data show that task-relevant information is captured by low-dimensional subspaces, with distinct temporal profiles for the representation of the elements that comprise the control program of the task – object, decision and cue. Most prominently, an early phase of strong object coding along with weak cue coding is dominated by the contralateral stimulus. In contrast, a gradual build-up of the decision state is seen for both contralateral and ipsilateral stimuli, with partial overlap of decision coding for stimuli on the two sides.

Stimulus and decision information in attentional competition

We next used the same subspaces (derived from single-stimulus trials) to investigate population activity when items compete for attention. For comparability with the analyses of single-stimulus data, we again divided trials into two halves, projected data for each half onto the single-stimulus subspace derived from the other half, and averaged the results. Data were analyzed separately for low competition (target/distractor + neutral, T/D + N) and high competition (target + distractor, T + D) displays.

Results for low-competition displays are shown in Figures 2B, 3B and 4B. Comparison with results for a single T/D stimulus (Figures 2A, 3A, 4A) suggests only modest effects of the added N. For object coding (Figure 2B), there was again strong discrimination only for a contralateral stimulus (target or distractor), with separate statistical tests for T + N and D + N displays showing similar coding for targets and distractors even when accompanied by an ipsilateral

N. With the accompanying contralateral N, object coding for an ipsilateral stimulus was not significant at any time point (Figure 2B, right). Decision coding was again seen for both contralateral and ipsilateral stimuli (Figure 3B). Cue coding again was weak, with only scattered points of significance for an ipsilateral T/D stimulus (Figure 4B). Statistical comparisons between single-stimulus and low-competition displays (see Materials and Methods) showed no time points with significant differences in either object, decision, or cue coding, for either contralateral or ipsilateral T/D stimuli.

In the high-competition conditions, a target and distractor appeared in the same display (T + D). For these displays, projections of interest in contralateral and ipsilateral subspaces are shown in Figure 5.

[Figure 5 about here]

For each of the four possible T + D displays, Figure 5A shows projections onto the object axis of the contralateral subspace. For comparison, data for single contralateral stimuli are copied from Figure 2A. The data shows that, for each T + D display, responses closely resembled responses just to the single object in the contralateral field, when this was presented alone. The most striking exception was that, when the contralateral object was a distractor (Figure 5A, solid light blue and orange), its object-selective response was short-lived, significant only around 100-180 ms from display onset (Figure 5A, pale grey line), suggesting a rapid shut-off of object coding for a contralateral D. Statistical comparison of the contralateral object discrimination in single-stimulus and T + D displays suggested a brief difference at around 200

ms, significant ($p < 0.001$) for individual time points but not surviving cluster correction. A similar analysis of movements in the ipsilateral subspace showed little coding (data not shown), in line with weak object coding for ipsilateral objects even when these were presented alone (Figure 2A).

For decision coding (Figure 5B, C), responses to T + D displays are shown along with responses to a single contralateral or ipsilateral target (dotted lines, copied from Figure 3A) for comparison. When the target in a T + D display was contralateral (Figure 5B), activity in the contralateral subspace closely followed the activity for a contralateral target presented alone. Thus, unlike responses on the object dimension, activity on the decision dimension was driven by the target stimulus, whose presence determined the required response. When the target in a T + D display was ipsilateral (Figure 5C), results were rather different. Now movements along the decision axis were delayed and weakened compared with movements for the same ipsilateral T presented alone (Figure 5C; significant difference between single-stimulus and T + D displays shown by pale gray line). Thus, population activity indicating an ipsilateral target was significantly suppressed by the presence of a competing, contralateral distractor.

For the cue dimension, finally, responses on T + D trials again suggested only weak and occasional cue discrimination, resembling that seen in single-stimulus displays (data not shown).

Distributed, mixed and independent task-related representations

The neuronal weights associated with each axis of a subspace reflect the strength of contribution of each neuron's activity to the overall population coding, as well as its

preference depending on the sign (e.g., object 1 or 2 for object coding). For each axis, the distribution of weights across neurons was unimodal and centered around 0 (Figure 6A), demonstrating that coding of task-relevant variables was distributed with varying levels of contributions across neurons, rather than driven by a small sub-population. Weights on each axis were correlated across contralateral and ipsilateral subspaces (Figure 6A), with the largest correlation for the decision component where coding was relatively strong in both contralateral and ipsilateral subspaces. Correlations were significant, however, even for object and cue, despite weak coding of object in the ipsilateral subspace and cue in both subspaces. These correlations demonstrate that the representations in the two hemispaces are related.

A second question is whether coding across the neural population of different task variables is driven by separate sub-populations of neurons, or whether computation is done within the same neural circuit for all task-relevant variables. We addressed this question by correlating the rectified neuronal weights of pairs of task variables (Figure 6B). Weights were rectified in order to consider the magnitude of contribution of each neuron rather than its preference (e.g, to object 1 or 2). Correlations were generally low, except for a positive correlation between decision and object weights in the ipsilateral subspace. This result shows the dominance of mixed selectivity, with a neuron's contribution to coding of each task variable rather independent of its contribution to others.

[Figure 6 about here]

Reliability

We constructed the low-dimensional subspaces by using a split-half cross-validated approach. To investigate the reliability of representations, we correlated weights derived from each half of the data (Figure 7). Reliabilities were highly significant, particularly for the components that showed strong coding: the decision variable in both hemifields ($r = 0.43$ and $r = 0.53$ for contralateral and ipsilateral subspaces, respectively, both $p < 0.000001$) and the object variable in the contralateral subspace ($r = 0.49$, $p < 0.000001$). Even for other variables, however, these data confirm some stability in coding in each subspace.

[Figure 7 about here]

DISCUSSION

Prefrontal coding of task-related information has been demonstrated across a wide range of variables, including visual properties (e.g., motion color), tactile input, cues, reward value, belief updating and decision (Romo et al., 1999; Romo and Salinas, 2003; Rushworth et al., 2011; Mante et al., 2013; Cai and Padoa-Schioppa, 2014; Hunt et al., 2018; Aoi et al., 2020). In this study we investigated prefrontal coding during a context-dependent decision-making task. We used dPCA to construct a low-dimensional subspace that separates coding of cue, object and decision, and tracked their dynamics through the course of a trial. With this approach we aimed to analyze how a learned cognitive operation is implemented in neural population activity. The results suggest how information flows within and between the two frontal lobes as a behavioral decision is reached.

To summarize our findings, and the processing dynamics they suggest, we use the scheme in Figure 8. In the task-related subspaces for single stimulus displays (Figure 8A), coding along the cue and object axes (top) as well as the decision axes (bottom) is shown for the two hemispheres relative to the stimulus presentation hemifield. While we recorded activity in a single hemisphere for contralateral and ipsilateral stimuli, in the figure, findings are re-cast to show inferred activity in each hemisphere for a single stimulus. Activity at different loci within the object-cue subspace are shown by colored dots (top, colors matching plots in Figures 2-5). Colored arrows between top and bottom sections show suggested dynamic shifts from object-cue coding to decision coding, leading to movements along the decision axis (vertical gray arrows, bottom). Suggested inter-hemispheric transfer of information is shown in horizontal curvy pale gray arrows (bottom). Extension to an example high-competition display is shown in Figure 8B (see below).

[Figure 8 about here]

For a single stimulus (Figure 8A), our results show object coding in the contralateral hemisphere which rose to a peak around 150 ms from display onset, followed by sustained coding throughout the rest of the trial. Cue coding in this hemisphere was weak, but followed a generally similar time course. Figure 8A illustrates the locus of activity in the object-cue subspace for the relatively stable period from 150 ms onwards, with colored dots for the four single-stimulus conditions in our task. Integrated object and cue information then drives coding along the decision axis (solid colored arrows) towards the correct behavioral decision,

either 'target present' or 'target absent' (dark and pale gray arrows, respectively), with a gradually increasing strength of coding over time. In line with previous reports, these coding patterns reflect context-dependent decision signals in a prefrontal circuit (Sakagami and Niki, 1994; Machens et al., 2005; Mante et al., 2013; Stokes et al., 2013), dominated by the contralateral stimulus (Kadohisa et al., 2013; Duong et al., 2019).

In contrast, in the ipsilateral hemisphere there was only weak object and cue coding, as illustrated with the four colored dots close to the center of the two axes in Figure 8A (top right). Object coding for a single ipsilateral stimulus, although weak, was sustained for a time window similar to the peak of coding for a contralateral stimulus, consistent with our previous report for the same dataset where a small proportion of cells were selective for an ipsilateral object identity (Kadohisa et al., 2015). Importantly, despite the weak object and cue coding in the ipsilateral hemisphere, which suggests only limited effect on the decision in this hemisphere (dotted colored arrows), our results show that a strong decision state in fact developed. Indeed, ipsilateral decision coding was similar in strength and dynamics to coding on the contralateral hemisphere. Given weak object and cue information in the ipsilateral hemisphere, a putative mechanism to drive a strong decision trajectory is cross-hemisphere transfer of decision information from the opposite hemisphere (pale gray bi-directional arrows between the decision axes). This suggested bi-directional exchange of decision information between the hemispheres then further reinforces the development of a coherent decision state in both hemispheres, either 'target present' or 'target absent'.

Within each hemisphere, moderate but highly significant correlations between contralateral and ipsilateral dPCA weights (Figure 6A) revealed some overlapping, though not identical, coding of each task variable for contralateral and ipsilateral stimuli. Especially important is

the decision axis, which in each hemisphere showed strong coding for both contralateral and ipsilateral stimuli. Partial overlap of the two axes suggests an integrated 'target present' or 'target absent' decision. Partial separation likely reflects the different responses (saccade directions) required for contralateral and ipsilateral stimuli. As noted earlier, the absence of saccadic activity in this cell population (Kadohisa et al., 2013) suggests an abstract response decision rather than specific motor preparation. Mostly negligible weight correlations across different task variables (Figure 6B) demonstrate mixed selectivity of prefrontal neurons (Rigotti et al., 2013; Xie et al., 2022), such that a neuron's contribution to any one representational axis was largely independent of its contribution to others. Notably absent were negative correlations, which might be expected if dedicated neural populations coded object and cue, then feeding into a further dedicated population for decision. Instead, overlapping populations could contribute to coding each task variable (Sigala et al., 2008), as would be expected, for example, in a recurrent network (Rigotti et al., 2010; Mante et al., 2013).

Moving on from single-stimulus display, we analyzed information dynamics when items compete to determine the behavioral decision. The importance of levels of competition to behavioral outcome is well established in human visual search experiments (Shiffrin and Schneider, 1977; Schneider and Fisk, 1982). Items that are not currently targets but have been frequently experienced as targets throughout learning history impose a large conflict and compete strongly with target items for attention. In contrast, less conflict is evoked by items that can always be ignored. A similar manipulation of competition was applied in our study, with distractor objects strongly competing for attention, and therefore expected to introduce

a substantial effect on coding within the state space, and neutral items posing only little competition with potentially little effect on coding.

Indeed, adding a low-competition neutral item to a target or distractor had very little effect on coding of object, decision and cue information. One striking effect, however, is that object coding in the ipsilateral hemisphere – already weak for a single stimulus – was now completely eliminated by the accompanying neutral item in the opposite hemifield (Figure 2). Despite the absence of object coding in the ipsilateral hemisphere, strong decision coding still developed (Figure 3). This finding adds weight to our proposal that the decision is first computed in the contralateral hemisphere, then communicated from one hemisphere to the other.

A different picture was observed for displays containing both a target and distractor, when attentional competition was high. Movements in the representational space for this case are illustrated in Figure 8B, for an example display with cue 1 followed by a target in one hemisphere and a distractor in the other. Note that, within each hemisphere, there are potential movements along two decision axes, one for the contralateral and one for the ipsilateral stimulus, but given the partial overlap of these two axes, here for simplicity they have been collapsed into one. Figure 8B shows arrows just for the active paths, i.e., those driven by the particular combination of stimuli present in the example display.

The results show that, in each hemisphere, object coding was dominated by the contralateral stimulus, either a target or distractor (Figure 5A). Given the connections inferred for the single stimulus case (Figure 8A), these two separate object representations should attempt to drive decision coding in the two hemispheres in opposite directions (dark blue and orange arrows). For the hemisphere contralateral to the target, decision activity evolved as expected (Figure

5B), but for the hemisphere contralateral to the distractor, a very different picture emerged. First, the data suggest a rapid shut-off of the distractor object representation (Figure 5A). Then decision coding moves, not towards the 'target absent' decision as should have occurred without competition from the other hemisphere (demonstrated by the orange arrow), but slowly and weakly towards the (task-appropriate) 'target present' decision. To achieve this outcome, Figure 8B shows suppression from the 'target present' end of the decision axis in one hemisphere to the object representation in the opposite hemisphere (zigzag arrow), implied and supported by the results. The findings suggested that, as this suppression develops, object representation is lost in the hemisphere contralateral to the distractor, leaving the 'target present' decision to be transferred from the opposite hemisphere. Though the zigzag arrow in Figure 8B suggests direct interaction between the two frontal lobes, this is only one possible way for the 'target present' decision in one hemisphere to suppress object information in the other. A different possibility would be for detection of the task-relevant target to feed back to earlier cortical levels, biasing competition to this target and suppressing the accompanying visual representation of the distractor. In either case, as the logic of the task requires, the object that should control the decision comes finally to dominate neural activity of both hemispheres along the decision axis, though in the hemisphere contralateral to the distractor, this dominance is achieved only slowly and incompletely, as observed in the results, presumably reflecting a remaining influence from initial distractor processing. These results are broadly consistent with attentional competition in visual areas, where early sensory-driven responses are followed by responses dominated by the most behaviorally relevant stimulus (Luck et al., 1997; Chelazzi et al., 1998; Roelfsema et al., 1998; Ghose and Maunsell, 2008).

Figure 8 exemplifies the kind of computational structure required for control of complex behavior. In prefrontal cortex, such structures must be assembled by learning; for example, learning that cue and object information must be combined to determine oculomotor choices, and that the cued target should dominate in a two-stimulus display. In the broader context, this is an example of just one task that can be learned and coded in the prefrontal cortex. Modeling studies have demonstrated how such computation can be reproduced within a neural network (Mante et al., 2013; Chaisangmongkon et al., 2017). Furthermore, one network can implement many such computations for different tasks that are learned through experience, creating a task space from which task representations can be composed to facilitate learning of new task structures (Yang et al., 2019). We demonstrate here an implementation of one building block within such a task space and how it can be used for neural computation in an extended task with competing items. More broadly, dimensionality reduction techniques such as dPCA offer a window onto such learned computational structures and their neural dynamics.

DATA AND CODE AVAILABILITY

The raw neuronal data and custom analysis scripts that support the findings in this study will be made freely available for download at a public repository upon publication. The code for the dPCA toolbox is freely available for download (Kobak et al., 2016).

AUTHOR CONTRIBUTIONS

Y.E. conceived the study, analyzed data and wrote the manuscript. M.Ka., P.P. and M.Ku. collected data. M.Ka., P.P., N.S., M.B., M.Ku and J.D. contributed to original task design and project administration. J.D. conceived the study, supervised all aspects of the project and wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

- Aoi MC, Mante V, Pillow JW (2020) Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. *Nature Neuroscience* 23:1410–1420.
- Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict Monitoring and Cognitive Control. *Psychological Review* 108:624–652.
- Buschman TJ, Siegel M, Roy JE, Miller EK (2011) Neural substrates of cognitive capacity limitations. *PNAS* 108:11252–11255.
- Cai X, Padoa-Schioppa C (2014) Contributions of Orbitofrontal and Lateral Prefrontal Cortices to Economic Choice and the Good-to-Action Transformation. *Neuron* 81:1140–1151.
- Chaisangmongkon W, Swaminathan SK, Freedman DJ, Wang XJ (2017) Computing by Robust Transience: How the Fronto-Parietal Network Performs Sequential, Category-Based Decisions. *Neuron* 93:1504-1517.e4.
- Chelazzi L, Duncan J, Miller EK, Desimone R (1998) Responses of Neurons in Inferior Temporal Cortex During Memory-Guided Visual Search. *Journal of Neurophysiology* 80:2918–2940.
- Dehaene S, Kerszberg M, Changeux J-P (1998) A neuronal model of a global workspace in effortful cognitive tasks. *PNAS* 95:14529–14534.
- Duncan J (2001) An adaptive coding model of neural function in prefrontal cortex. *Nature reviews Neuroscience* 2:820–829.
- Duncan J, Assem M, Shashidhara S (2020) Integrated Intelligence from Distributed Brain Activity. *Trends in Cognitive Sciences* 24:838–852.
- Duong L, Leavitt M, Pieper F, Sachs A, Martinez-Trujillo J (2019) A Normalization Circuit Underlying Coding of Spatial Attention in Primate Lateral Prefrontal Cortex. *eNeuro* 6:ENEURO.0301-18.2019.
- Erez Y, Duncan J (2015) Discrimination of Visual Categories Based on Behavioral Relevance in Widespread Regions of Frontoparietal Cortex. *J Neurosci* 35:12383–12393.
- Everling S, Tinsley CJ, Gaffan D, Duncan J (2002) Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nat Neurosci* 5:671–676.
- Freedman D, Riesenhuber M, Poggio T, Miller E (2001) Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science* 291:312–312.
- Ghose GM, Maunsell JHR (2008) Spatial Summation Can Explain the Attentional Modulation of Neuronal Responses to Multiple Stimuli in Area V4. *J Neurosci* 28:5115–5126.

- Hunt LT, Malalasekera WMN, de Berker AO, Miranda B, Farmer SF, Behrens TEJ, Kennerley SW (2018) Triple dissociation of attention and decision computations across prefrontal cortex. *Nature Neuroscience* 21:1471–1481.
- Jackson J, Rich AN, Williams MA, Woolgar A (2016) Feature-selective Attention in Frontoparietal Cortex: Multivoxel Codes Adjust to Prioritize Task-relevant Information. *Journal of Cognitive Neuroscience* 29:310–321.
- Jiang X, Bradley E, Rini RA, Zeffiro T, Vanmeter J, Riesenhuber M (2007) Categorization training results in shape- and category-selective human neural plasticity. *Neuron* 53:891–903.
- Kadohisa M, Kusunoki M, Petrov P, Sigala N, Buckley MJ, Gaffan D, Duncan J (2015) Spatial and temporal distribution of visual information coding in lateral prefrontal cortex. *Eur J Neurosci* 41:89–96.
- Kadohisa M, Petrov P, Stokes M, Sigala N, Buckley M, Gaffan D, Kusunoki M, Duncan J (2013) Dynamic construction of a coherent attentional state in a prefrontal cell population. *Neuron* 80:235–246.
- Kadohisa M, Watanabe K, Kusunoki M, Buckley MJ, Duncan J (2019) Focused Representation of Successive Task Episodes in Frontal and Parietal Cortex. *Cerebral Cortex*:bhz202.
- Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, Qi X-L, Romo R, Uchida N, Machens CK (2016) Demixed principal component analysis of neural population data. *eLife* 5:e10989.
- Lee J, Maunsell JHR (2009) A normalization model of attentional modulation of single unit responses. *PLoS ONE* 4:e4651.
- Lee SH, Kravitz DJ, Baker CI (2013) Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nat Neurosci* 16:997–999.
- Li S, Ostwald D, Giese M, Kourtzi Z (2007) Flexible coding for categorical decisions in the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27:12321–12330.
- Liu T, Hospadaruk L, Zhu DC, Gardner JL (2011) Feature-specific attentional priority signals in human cortex. *J Neurosci* 31:4484–4495.
- Luck SJ, Chelazzi L, Hillyard SA, Desimone R (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J Neurophysiol* 77:24–42.
- Machens CK, Romo R, Brody CD (2005) Flexible Control of Mutual Inhibition: A Neural Model of Two-Interval Discrimination. *Science* 307:1121–1124.
- Mante V, Sussillo D, Shenoy KV, Newsome WT (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503:78–84.

- Matsushima A, Tanaka M (2014) Different Neuronal Computations of Spatial Working Memory for Multiple Locations within versus across Visual Hemifields. *J Neurosci* 34:5621–5626.
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202.
- Norman DA, Shallice T (1980) *Attention to Action: Willed and Automatic Control of Behavior*. San Diego: University of California, Center for Human Information Processing.
- Rainer G, Asaad WF, Miller EK (1998) Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* 393:577–579.
- Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci* 19:1736–1753.
- Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61:168–185.
- Rigotti M, Barak O, Warden MR, Wang X-J, Daw ND, Miller EK, Fusi S (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* 497:1–6.
- Rigotti M, Ben Dayan Rubin D, Wang X-J, Fusi S (2010) Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. *Front Comput Neurosci* 4:24.
- Roelfsema PR, Lamme VA, Spekreijse H (1998) Object-based attention in the primary visual cortex of the macaque monkey. *Nature* 395:376–381.
- Romo R, Brody CD, Hernández A, Lemus L (1999) Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399:470–473.
- Romo R, Salinas E (2003) Flutter Discrimination: neural codes, perception, memory and decision making. *Nature Reviews Neuroscience* 4:203–218.
- Rushworth MFS, Noonan MAP, Boorman ED, Walton ME, Behrens TE (2011) Frontal Cortex and Reward-Guided Learning and Decision-Making. *Neuron* 70:1054–1069.
- Sakagami M, Niki H (1994) Encoding of behavioral significance of visual stimuli by primate prefrontal neurons: relation to relevant task conditions. *Exp Brain Res* 97:423–436.
- Schneider W, Fisk AD (1982) Concurrent automatic and controlled visual search: Can processing occur without resource cost? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8:261–278.
- Shiffrin RM, Schneider W (1977) *Controlled and Automatic Human Information Processing: II. Perceptual Learning, Automatic Attending, and a General Theory*. *Psychol Rev* 84:64–64.

- Sigala N, Kusunoki M, Nimmo-Smith I, Gaffan D, Duncan J (2008) Hierarchical coding for sequential task events in the monkey prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America* 105:11969–11974.
- Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J (2013) Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78:364–375.
- Ungerleider LG, Gaffan D, Pelak VS (1989) Projections from inferior temporal cortex to prefrontal cortex via the uncinate fascicle in rhesus monkeys. *Exp Brain Res* 76:473–484.
- Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature* 411:953–956.
- Wen T, Duncan J, Mitchell DJ (2019) The time-course of component processes of selective attention. *Neuroimage* 199:396–407.
- Woolgar A, Hampshire A, Thompson R, Duncan J (2011a) Adaptive coding of task-relevant information in human frontoparietal cortex. *J Neurosci* 31:14592–14599.
- Woolgar A, Thompson R, Bor D, Duncan J (2011b) Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *NeuroImage* 56:744–752.
- Xie Y, Hu P, Li J, Chen J, Song W, Wang X-J, Yang T, Dehaene S, Tang S, Min B, Wang L (2022) Geometry of sequence working memory in macaque prefrontal cortex. *Science* 375:632–639.
- Yang GR, Joglekar MR, Song HF, Newsome WT, Wang X-J (2019) Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience* 22:297–306.

Figure legends:

A

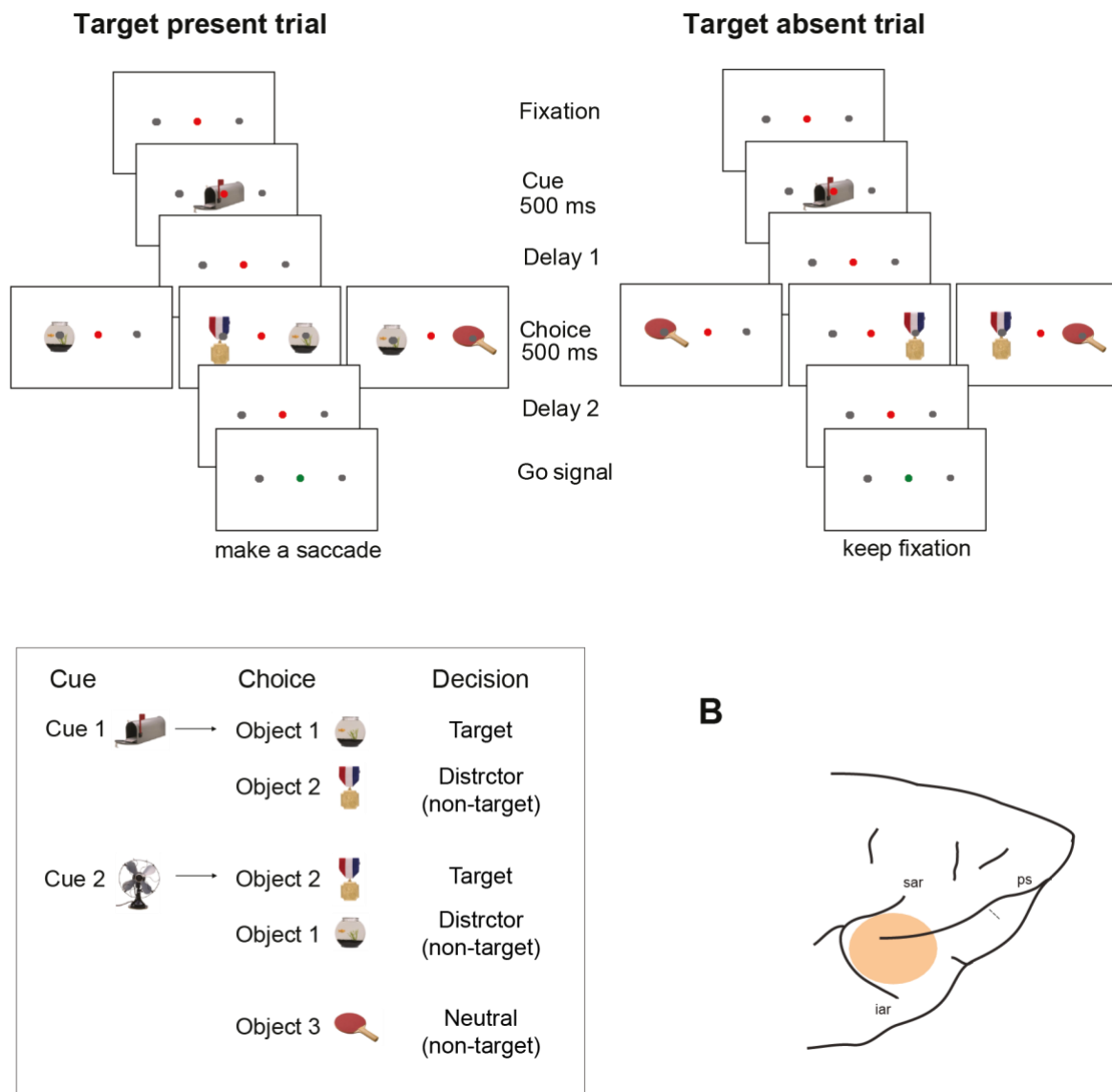


Figure 1: Task and recording locations. A: Task. Each trial began with fixation on a central dot, followed by a cue (500 ms). After a delay (variable delay 1, see Materials and Methods for details), a choice display appeared (500 ms) with either one or two stimuli. Based on previous training, for each animal, each of two cues was associated with a single target object. Cues and cue choice objects for one animal are illustrated in the inset on the right. Depending on the preceding cue, the animals had to make a decision whether a target (the associated object) appeared in the choice display (target present trial) or not (target absent trial).

Following another delay (variable delay 2, see Materials and Methods for details), the color of the fixation dot was changed to green, indicating the go signal. On target present trials, animals made a saccade to the target location and were then rewarded with a drop of liquid for a successful trial. On target absent trials, animals had to hold fixation during the response interval and were rewarded at the end of a wait period. Choice display objects could be the object associated with the preceding cue (target), the object that was associated with the other cue (distractor), or a third object that was not associated with any cue (neutral). Stimuli could appear either to the right or left of the fixation dot (i.e., contralateral or ipsilateral to the recorded hemisphere). Two-stimulus displays were target + distractor (T + D), target + neutral (T + N) and distractor + neutral (D + N). **B: Approximate recording locations in the right hemisphere.** For one animal, additional recordings were made in a similar region of the left hemisphere. ps: principal sulcus; sar: superior arcuate sulcus; iar: inferior arcuate sulcus. Modified from Kadohisa et al. (Kadohisa et al., 2013).

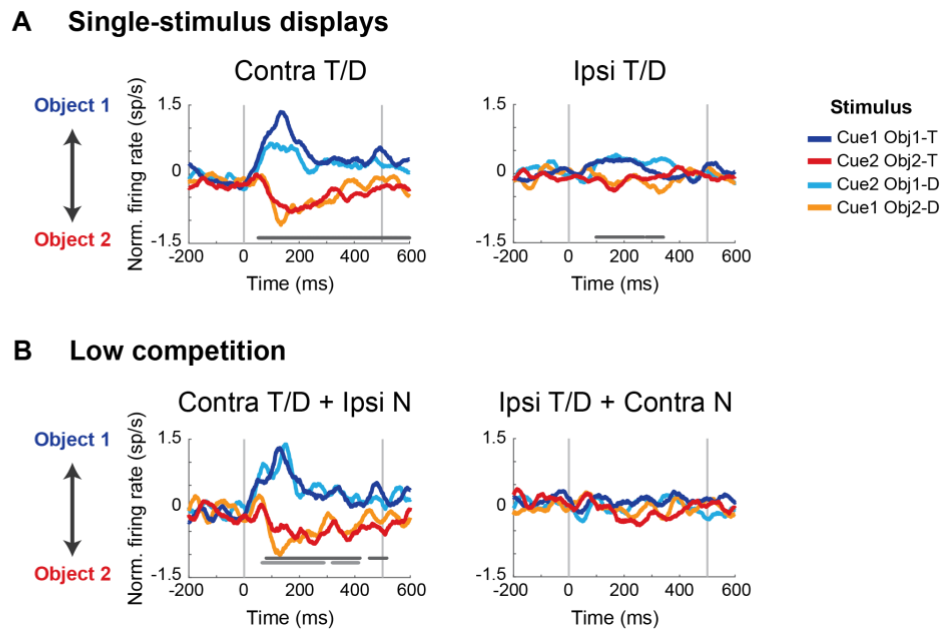


Figure 2: Object information for single-stimulus and low competition conditions. A: Projections on the dPCA object axis for single target (T) and distractor (D) stimuli. Population activity relative to choice display onset is projected onto the first object component of each subspace (contralateral subspace for contralateral stimuli; ipsilateral subspace for ipsilateral stimuli), with positive firing rates indicating a representation of object 1, and negative firing rates indicating object 2. Cross-validated responses are shown as averages of the two halves of the data (odd and even trials), each projected onto the subspace of the other half of the data. Gray vertical lines indicate stimulus onset (0 ms) and offset (500 ms). Horizontal line at the bottom indicates significant difference between object 1 and object 2, averaged across target and distractor, reflecting object information. Significance is determined using a permutation approach with cluster-based correction across time points. **B:** Object information for low competition conditions with an added neutral (N) stimulus (T + N and D + N). Horizontal lines at the bottom indicate significant difference between object 1 and object 2 for T + N (dark gray) and D + N (light gray).

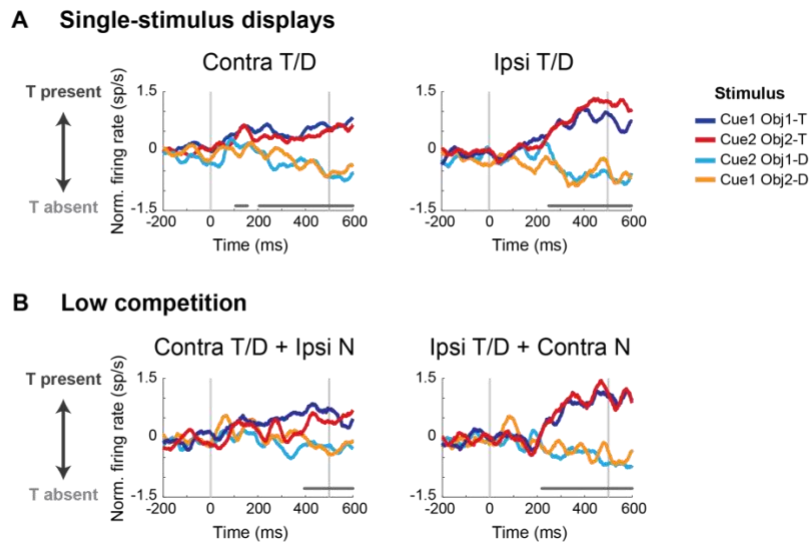


Figure 3: Decision information for single-stimulus and low competition conditions. A: Projections on the dPCA decision axis for single target (T) and distractor (D) stimuli. Population activity relative to choice display onset is projected onto the first decision component of each subspace (contralateral subspace for contralateral stimuli; ipsilateral subspace for ipsilateral stimuli), with positive firing rates indicating ‘target present’, and negative firing rates indicating ‘target absent’. Cross-validated responses are shown as averages of the two halves of the data (odd and even trials), each projected onto the subspace of the other half of the data. Gray vertical lines indicate stimulus onset (0 ms) and offset (500 ms). Horizontal line at the bottom indicates significant difference between ‘target present’ and ‘target absent’, averaged across objects 1 and 2, reflecting decision information. Significance is determined using a permutation approach with cluster-based correction across time. **B:** Object information for low competition conditions with an added neutral (N) stimulus (T + N and D + N). Gray horizontal line at the bottom indicates significant difference between T + N vs. D + N conditions.

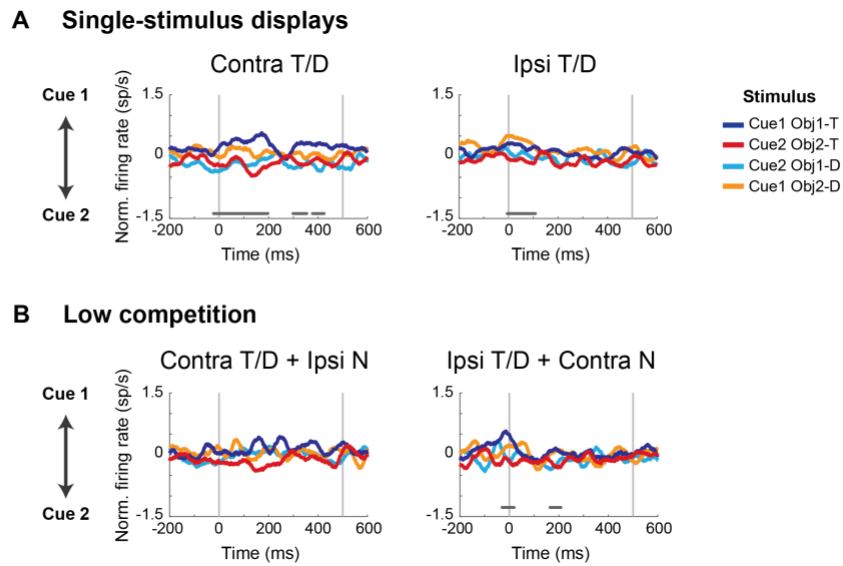
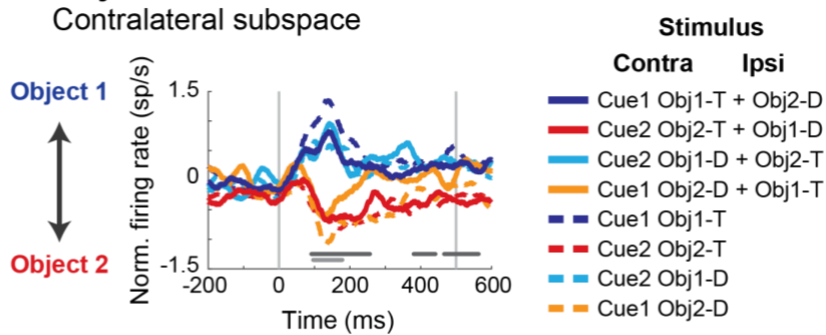


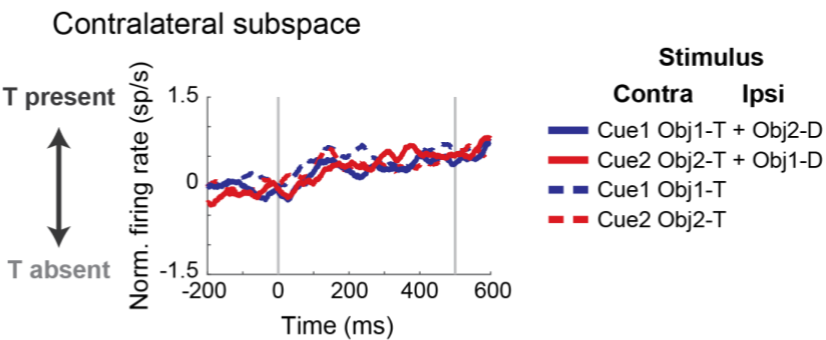
Figure 4: Cue information for single-stimulus and low competition conditions. A: Projections on the dPCA cue axis for single target (T) and distractor (D) stimuli. Population activity relative to choice display onset is projected onto the first cue component of each subspace (contralateral subspace for contralateral stimuli; ipsilateral subspace for ipsilateral stimuli), with positive firing rates indicating a representation of cue 1, and negative firing rates indicating cue 2. Cross-validated responses are shown as averages of the two halves of the data (odd and even trials), each projected onto the subspace of the other half of the data. Gray vertical lines indicate stimulus onset (0 ms) and offset (500 ms). Horizontal line at the bottom indicates significant difference between cue 1 and cue 2, averaged across objects 1 and 2, reflecting cue information. Significance is determined using a permutation approach with cluster-based correction across time points. **B:** Cue information for low competition conditions with an added neutral (N) stimulus.

High competition T + D

A Object dimension



B Decision dimension



C Decision dimension

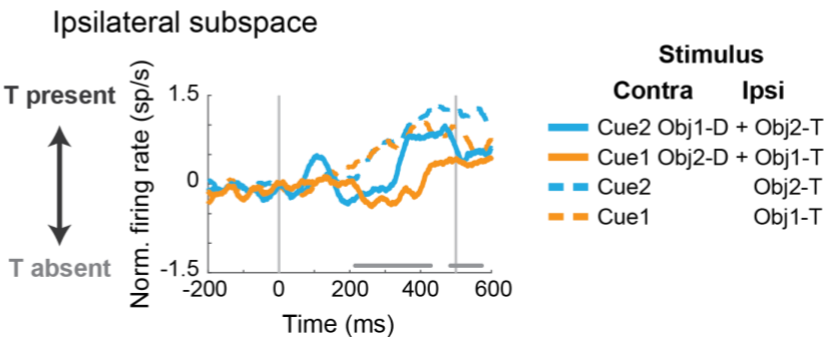


Figure 5: Stimulus and decision coding for high competition conditions (target + distractor,

T + D). A: Object coding on the contralateral subspace. For comparison, projections of single T/D presented contralaterally (copied from Figure 2A) are plotted as dashed lines. Dark and pale gray significance lines at the bottom indicate significant object coding, based on the identity of the contralateral object, for the contra T + ipsi D, and contra D + ipsi T displays, respectively. Statistical comparison between single-stimulus and high-competition displays showed no significant difference in coding the identity of the contralateral object, whether

this was T (dark blue, red) or D (light blue, orange). **B:** Decision coding on the contralateral subspace when targets are presented contralaterally. For comparison, projections of a single T presented contralaterally (copied from Figure 3A) are plotted as dashed lines. There was no difference in decision coding between single-stimulus and high-competition displays. **C:** Decision coding on the ipsilateral subspace when targets are presented ipsilaterally. For comparison, projections of a single T presented ipsilaterally (copied from Figure 3A) are plotted as dashed lines. Pale gray horizontal line at the bottom indicates significant difference in decision coding between single-stimulus and high-competition displays. All other details as in Figures 2-4.

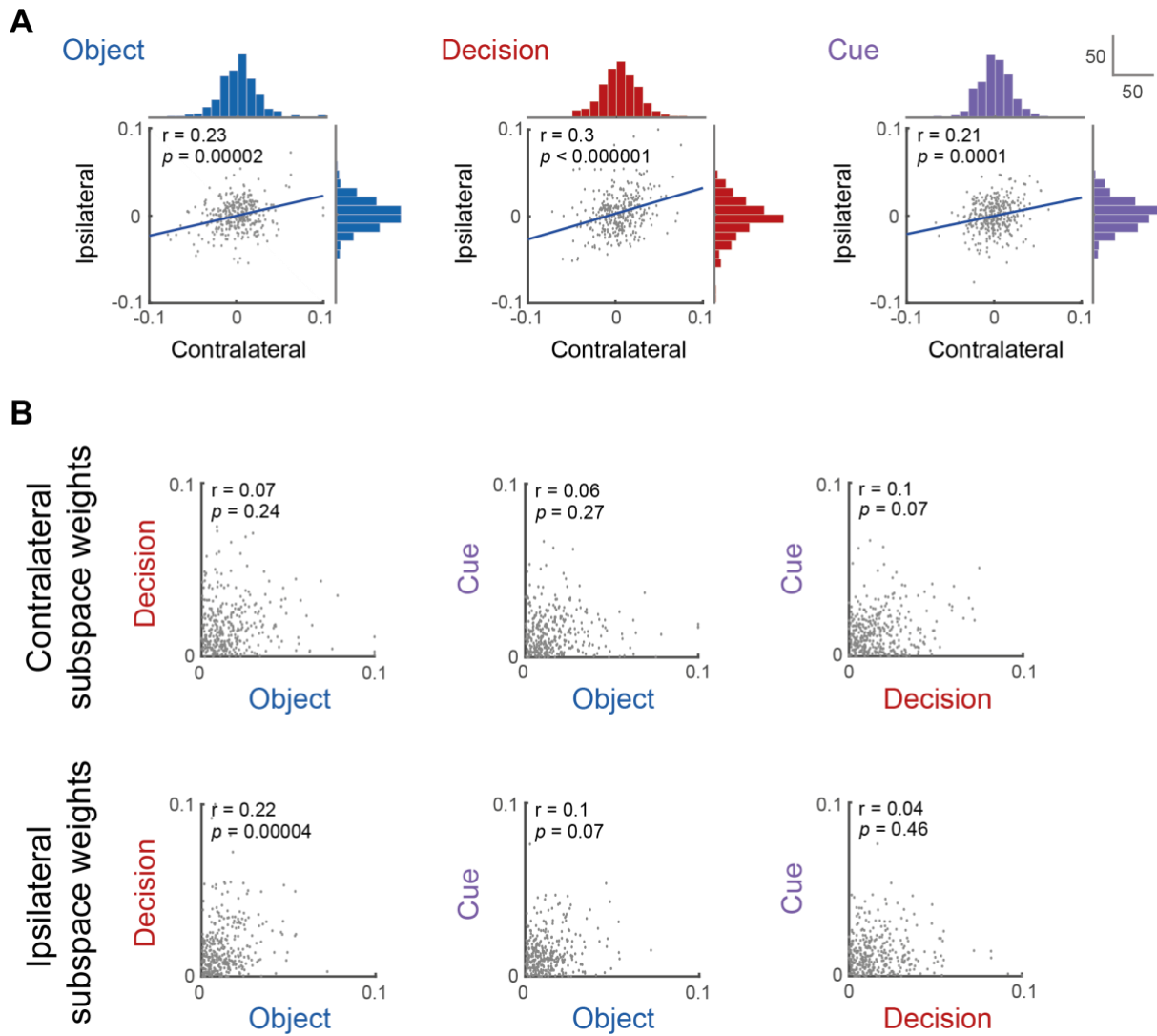


Figure 6: Distributed, mixed and independent task-related representations across the neuronal population. A: Relationship of weights in contralateral and ipsilateral subspaces for each task variable (subspace axis). Each dot is one neuron. Pearson correlation coefficient, p value and a regression line are indicated within each plot. Histograms at top (contralateral subspace) and right (ipsilateral subspace) show weight distribution across neurons. y axis scale for these histograms is shown at top right. **B:** Rectified weights of individual neurons for each pair of axes, for the contralateral (top) and ipsilateral (bottom) subspaces. Absolute weights are plotted to show strength of selectivity regardless of preference. Spearman's rank

correlation coefficient and p value are indicated within each plot. For all plots, weights of each neuron and axis are averaged across the subspaces derived from the two halves of the data.

Contralateral sub-
space weights

Ipsilateral
subspace weights

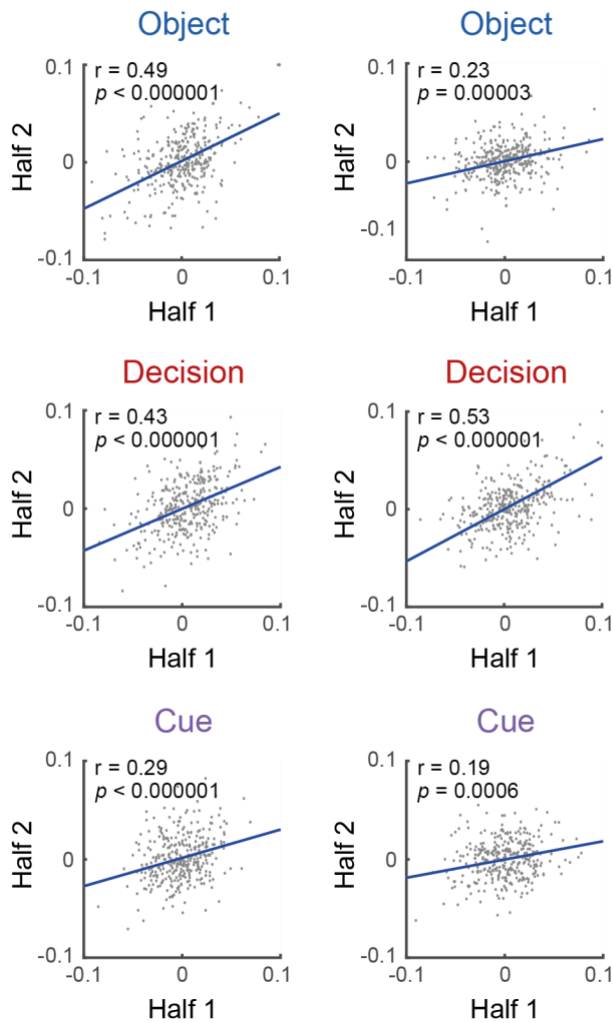


Figure 7: Reliability of coding in each subspace. For each task variable (subspace axis), plots show correlation between weights derived from separate halves of the data. Each dot is one neuron. Pearson correlation coefficient, p value and a regression line are indicated within each plot.

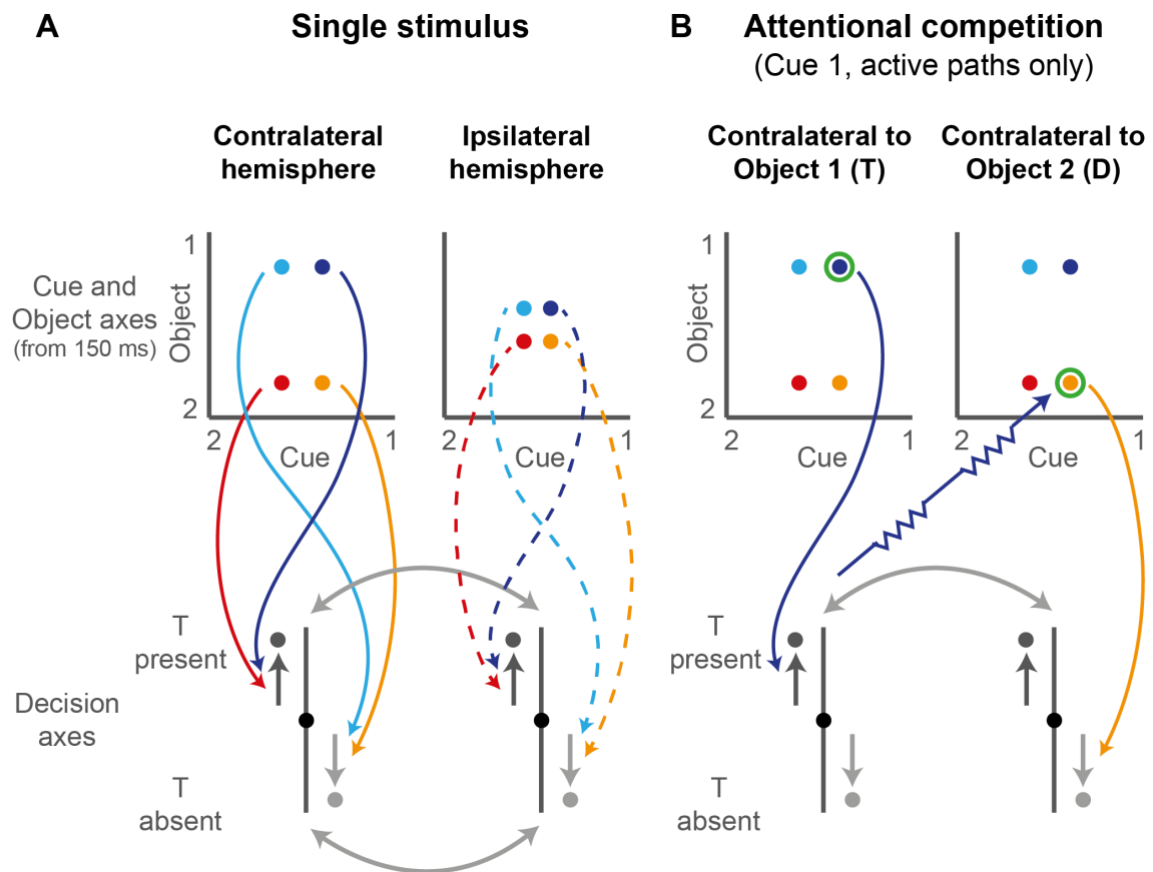
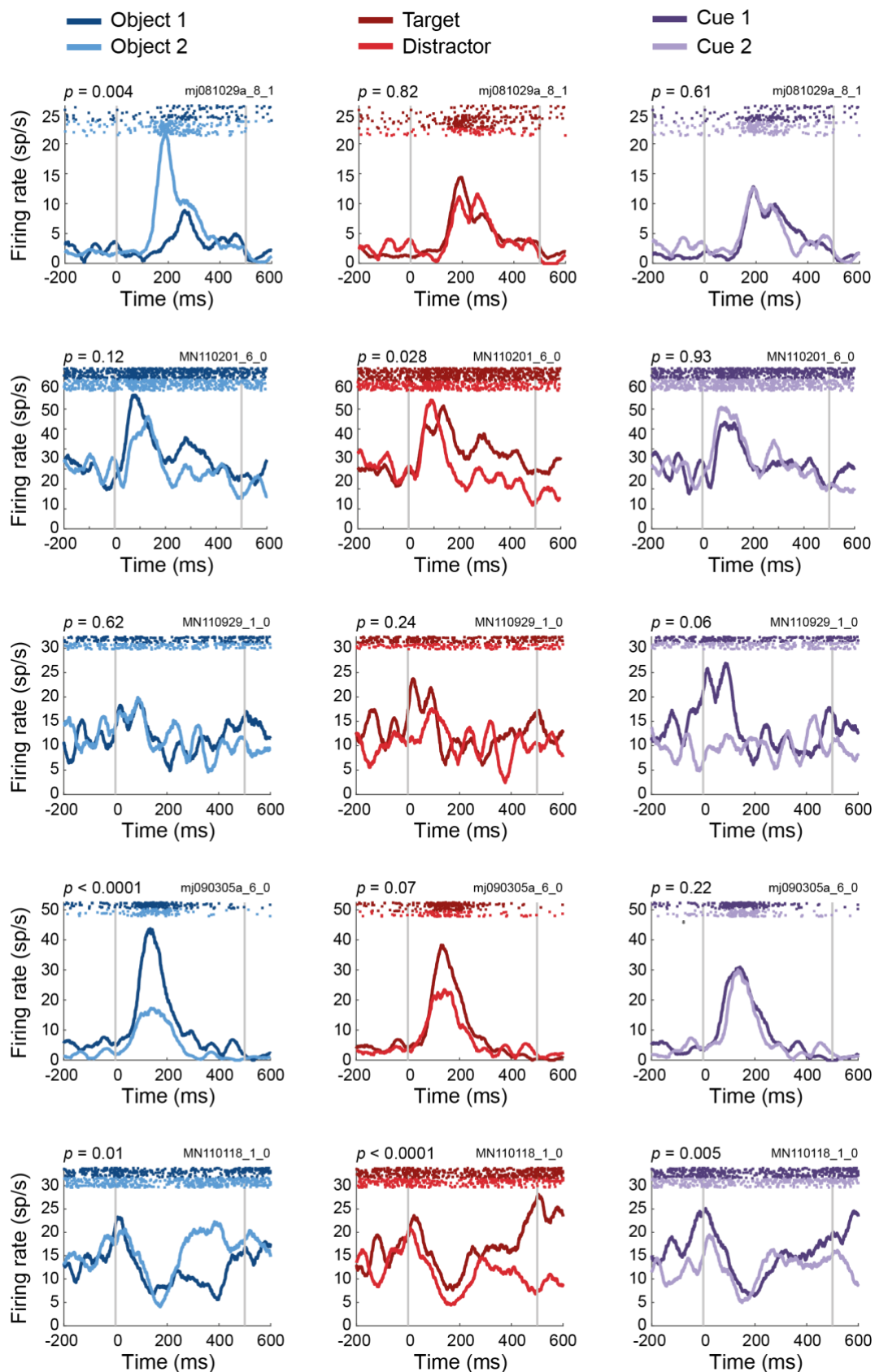


Figure 8: A model for attentional competition in prefrontal cortex. A. Single-stimulus displays. The task subspace for cue and object axes (top) and the decision axis (bottom) is shown for the two hemispheres relative to the stimulus presentation hemifield. Each colored dot shows representation of object and cue information for each of the four single-stimulus conditions, as demonstrated by the results. Shortly after stimulus onset, strong coding emerges in the hemisphere contralateral to the stimulus. Only weak object and cue coding are seen in the ipsilateral hemisphere. Integrated object and cue information then drives coding along the decision axis, mainly in the contralateral hemisphere (colored solid arrows, suggested by the model), and only weakly in the ipsilateral hemisphere (colored dotted arrows, suggested by the model). Coding along the decision axis moves towards the behavioral decision (dark and pale gray arrows beside decision axis, demonstrated by the results), and as it develops is transferred between hemispheres (pale gray arrows between

the decision axes, suggested by the model). **B. Attentional competition.** The task subspaces are shown for competition of a target (T) and distractor (D), with arrangement similar to the single-stimulus displays. For simplicity, within each hemisphere, partially-overlapping decision axes for contralateral and ipsilateral stimuli are collapsed into one (see main text). In this example trial, cue 1 sets object 1 as target and object 2 as distractor (green circles). Subspaces in each hemisphere are shown relative to the contralateral stimulus (colored dots, demonstrated by the results), with colored arrows showing putative active paths of influences driven by this example stimulus. Initially, object coding in each hemisphere is driven by the contralateral stimulus (top). Integration of cue and object then tries to drive the decision state in each hemisphere towards the corresponding behavioral decision (dark blue and orange arrows, suggested by the model). As the decision state approaches 'target present' in one hemisphere, however, this suppresses representation of the distractor object in the opposite hemisphere (zigzag arrow, suggested by the model). Suppression may be direct or indirect (see main text). Finally, the model suggests that 'target present' decision from the hemisphere contralateral to the target is transferred to the other hemisphere, though only slowly and weakly owing to that hemisphere's initial processing of the distractor.

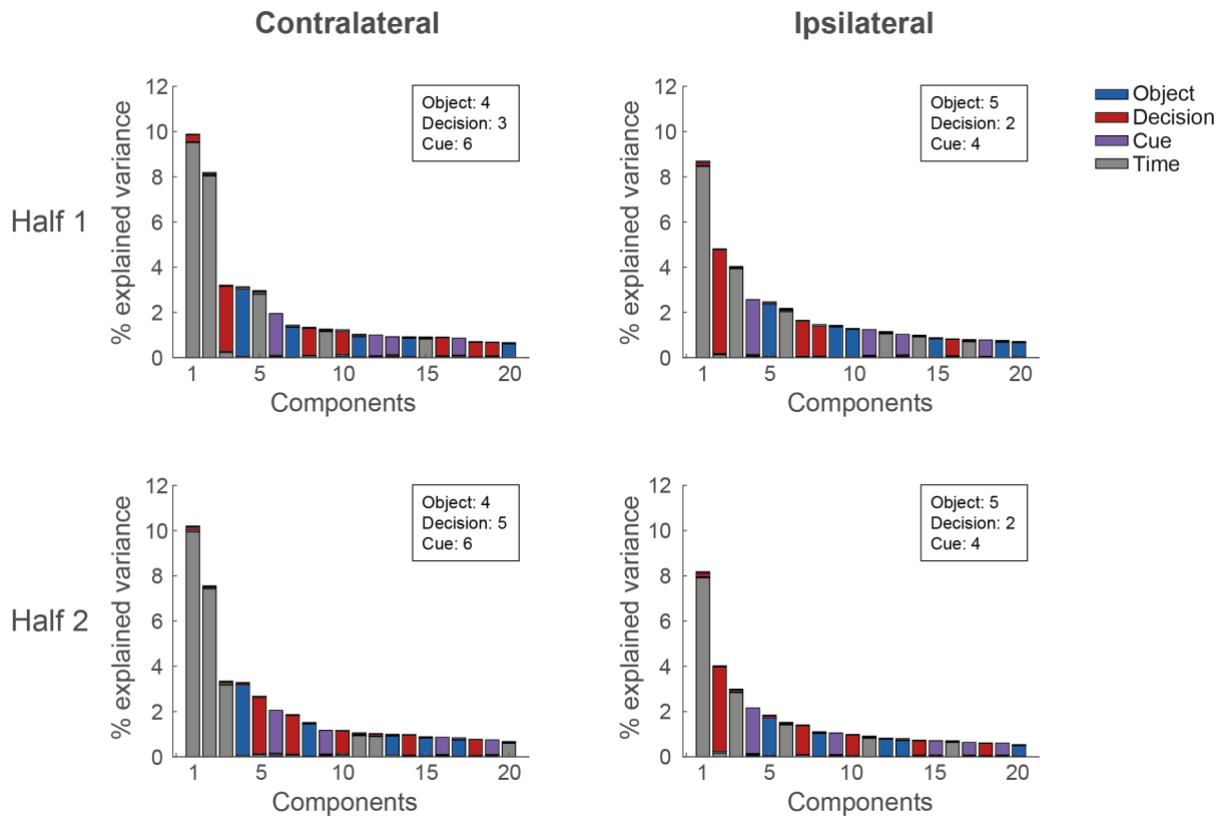
Supplementary information: Integrated neural dynamics for behavioral decisions and

attentional competition in the prefrontal cortex



Supplementary Figure 1: Heterogeneous response profiles of single prefrontal cells.

Examples of five units (one in each row) and their selectivity to object, decision and cue (left, middle and right column, respectively). Data come from single-stimulus trials. Peri-stimulus time histograms show average responses across trials for each of the objects, decisions and cues based on the four single-stimulus contralateral target and distractor conditions. Single-trial raster plots are at the top of each panel. Single cells showed a variety of responses to task-relevant variables, some with selectivity to just one variable (object, decision and cue for units shown in 1st, 2nd and 3rd rows, respectively), and some with selectivity to more than one variable (object and decision in 4th row, and all three variables in 5th row). Gray vertical lines indicate stimulus onset (0 ms) and offset (500 ms). *p* values above each panel indicate selectivity to each task variable as computed by a two-way ANOVA (main effects for object and decision, interaction for cue, based on trial data averaged across all time points from 200 ms before to 600 ms after display onset). Unit ID is shown above each panel on the right.



Supplementary Figure 2: Percentages of variance explained by the first 20 principal components for each subspace. Scree plots for each hemifield, subspace and half. Colors indicate the percentages of explained variance associated with each of the variables. 'Time' indicates variable-independent time-course of activity. Inset: the number of the first component assigned to each variable which was used for the subspaces.