Nat Neurosci. Author manuscript; available in PMC 2021 April 23.

Published in final edited form as:
Nat Neurosci. 2020 November 01; 23(11): 1410–1420. doi:10.1038/s41593-020-0696-5.

Prefrontal cortex exhibits multi-dimensional dynamic encoding during decision-making

Mikio C. Aoi[*,1], **Valerio Mante**[2], **Jonathan W. Pillow**[1]

[1]Department of Psychology and Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey, U.S.A. [2]Institute of Neuroinformatics, University of Zurich and ETH Zurich, Zurich, Switzerland

Abstract

Recent work has suggested that prefrontal cortex (PFC) plays a key role in context-dependent perceptual decision-making. Here we address that role using a new method for identifying task-relevant dimensions of neural population activity. Specifically, we show that PFC has a multi-dimensional code for context, decisions, and both relevant and irrelevant sensory information. Moreover, these representations evolve in time, with an early linear accumulation phase followed by a phase with rotational dynamics. We identify the dimensions of neural activity associated with these phases, and show that they do not arise from distinct populations, but of a single population with broad tuning characteristics. Finally, we use model-based decoding to show that the transition from linear to rotational dynamics coincides with a plateau in decoding accuracy, revealing that rotational dynamics in PFC preserve sensory choice information for the duration of the stimulus integration period.

Introduction

A large body of work has aimed to identify the precise computational roles of various brain regions during perceptual decision-making[1–8]. Recent interest has centered on prefrontal cortex (PFC), which has been shown to carry a wide range of sensory, cognitive, and motor signals relevant for integrating sensory information and making decisions[1;4;6;7;9–12]. A barrier to understanding PFC's functional role, however, is that PFC neurons exhibit mixed selectivity, characterized by heterogeneous tuning to multiple task variables[13]. These idiosyncratic single-neuron responses make it difficult to gain insight into the population-level representation of different sensory and cognitive variables.[14–16].

[*]Corresponding author.

Author Contributions

M.C.A and J.W.P. developed the model and performed data analysis. V.M. conceived and conducted the experiments and collected the data. All authors helped with the interpretation of analysis and writing of the paper.

Competing Financial Interests Statement

The authors declare no competing interests.

Europe PMC Funders Author Manuscripts

Here we analyze the population-level representation of information in PFC using *model-based targeted dimensionality reduction* (mTDR), a general method for identifying the dimensions of population activity that encode information about different task variables over time. We applied this method to electrophysiology data recorded during a context-dependent perceptual decision-making task[1], in which a context cue determined what kind of sensory information (color or motion) should be used for making a binary decision (Figure 1a,b). In contrast to previous findings, our analysis revealed that the encoding of decisions, context, and relevant as well as irrelevant stimulus variables exhibited rotational dynamics in a multi-dimensional subspace, involving modulation of two or more orthogonal neural activity patterns over time.

We also introduce a new unsupervised method, *sequential principal components analysis*, for decomposing multidimensional representations into a sequence of axes that reflect the temporal order in which information about each variable becomes available. This method reveals that multi-dimensional trajectories can be decomposed into an early phase with linear dynamics, followed by a later phase with rotational dynamics. We used model-based decoding under the mTDR framework to show that the transition between these phases corresponded to a saturation in decoding accuracy for sensory as well as decision information, suggesting that the population did not continue to accumulate sensory information during the rotational phase.

Taken together, these results substantially extend the prevailing picture of decision encoding in PFC: rather than integrating evidence along a single dimension of population activity, with amplitude that reflects accumulated evidence[17], neural population activity enters a phase of rotational dynamics that maintains information about the choice as well as relevant and irrelevant sensory information over the entire course of a single trial.[18–20].

Results

Model-based targeted dimensionality reduction

To characterize population-level representations in PFC, we introduce a new method, *model-based targeted dimensionality reduction* (mTDR), which seeks to identify a set of dimensions of population activity that carry information about distinct task variables. We illustrate the basic intuition for mTDR with a hypothetical 3-neuron population in a perceptual decision-making task (Figure 2). For this example, there are two task variables of interest: a sensory stimulus x_s and a binary decision variable x_c. These variables modulate the firing rates in different ways producing a diverse pattern of population responses across conditions (Figure 2a).

The population-level response can be described as trajectories in a 3-dimensional state space, where the coordinates of each axis correspond to the firing rates of each of the neurons (Figure 2b). Although the full space is 3-dimensional, the trajectories exhibit low-dimensional structure that is not apparent from the firing rates alone. Specifically, the population activity is confined to a 2D plane defined by two axes: a 1D “stimulus axis” (blue arrow) captures information about the stimulus strength, while a 1D “decision axis” (red arrow) captures information about the choice. Projecting the population response onto each

Europe PMC Funders Author Manuscripts

of these axes reveals a timecourse of information about stimulus level and choice, respectively (Figure 2c).

The goal of mTDR is to identify these encoding subspaces from neural population data. For our three-neuron example, the mTDR model describes the time evolution of the population response $\mathbf{y}(t)$, a vector of 3 neural firing rates, as:

$$\mathbf{y}(t) = x_{\text{stim}} \cdot \mathbf{w}_{\text{stim}} \cdot s_{\text{stim}}(t) + x_{\text{choice}} \cdot \mathbf{w}_{\text{choice}} \cdot s_{\text{choice}}(t) + \text{noise}, \tag{1}$$

where x_{stim} is the stimulus variable, which takes one of six values from [−3, −2, −1, +1, +2, +3] indicating the level of sensory evidence, and x_{choice} denotes the decision variable, which takes on values of ±1, indicating a positive or negative choice. The activity vectors $\mathbf{w}_{\text{stim}}$ and $\mathbf{w}_{\text{choice}}$ are patterns of activity across the three neurons specifying the stimulus and choice axes (blue and red arrows in Figure 2b), and the time-varying functions $sw_{\text{stim}(t)}$ and $s_{\text{choice}}(t)$ are temporal profiles for the activity along stimulus and choice axes, respectively (Figure 2c). Noise is added to each firing rate to account for variability not due to variations in the trial type.

Although the choice and decision subspaces in this example are both 1-dimensional, the mTDR model easily generalizes to higher dimensionality and for an arbitrary number of task variables. Let Y denote a *neurons* × *time* matrix of firing rates for a single condition defined by task variables $\{x_1, \ldots x_P\}$. The mTDR model decomposes population activity as:

$$\mathbf{Y} = x_1 \mathbf{W}_1 \mathbf{S}_1^{\top} + \cdots + x_P \mathbf{W}_P \mathbf{S}_P^{\top} + \text{noise} \tag{2}$$

where $\mathbf{W}_p$ is *neurons* × r_p matrix whose columns span a r_p-dimensional encoding subspace for task variable x_p, and $\mathbf{S}_p$ is a *time* × r_p matrix of temporal profiles that describe the timecourse of population activity within this subspace (Supplementary Math Note Figure 1). This model-based formalism represents a generalization of targeted dimensionality reduction (TDR)[1], which allows us to identify both the number of activity patterns used to encode different variables and the timecourses with which these patterns are recruited. (For details see Methods and Supplementary Note 2).

Population coding of task variables in PFC

To investigate population-level coding in PFC, we applied mTDR to neural data recorded from an area in and around the frontal eye fields (FEF) of two monkeys performing a context-dependent decision-making task[1] (see Methods, Experimental details) In this task, monkeys were presented with a visual stimulus that contained colored, moving dots on each trial (Figure 1a). A context cue (yellow square or blue cross) appeared before each trial and instructed the monkeys to attend either to the color (red vs. green) or the motion (left vs. right) of the dots. In the color context, the animal had to attend to color and ignore motion, making a saccade to the red (green) target if the majority of the dots were red (green). In the motion context, the animal had to attend to motion and ignore color, e.g. making a left (right) saccade if the dot motion was left (right).

Europe PMC Funders Author Manuscripts

Task difficulty was controlled by varying the fraction of red vs. green (and coherently moving) dots, across 6 levels of coherence for each stimulus dimension (Figure 1b). Following a randomized delay the monkey was cued to indicate its decision by making a saccade to one of the two targets.

Classical approaches to analyzing data of this type involve analyzing average firing rates, or peristimulus time histograms, for different task conditions (conditional PSTH's) such as "all trials with the strong rightward motion and a rightward choice". For this dataset, the conditional PSTHs of individual neurons exhibited heterogeneous tuning to the different task variables[1] (Figure 1c). This hetereogeneity, and the fact that each neuron encodes several task variables, makes it difficult to obtain a clear picture of the population-level representation of task variables from an examination of single-neuron PSTHs.

To overcome these limitations, we used mTDR to determine the dimensionality of population-level representations of the task variables. The mTDR model included a regressor for each of 6 task variables: color strength, motion strength, context, and choice, as well as two additional terms for the absolute values color and motion strength. Absolute value terms were included due to the observation that some neurons displayed nonlinear encoding of stimuli, consistent with observations of nonlinear mixed selectivity[13]. The model also included a term for the condition-independent firing rate, which reflects temporal modulation not due to the task variables (see Methods for details). To determine the dimensionality of the encoding of each task variable, we used a greedy selection method based on the Akaike information criterion[21] (AIC) that added dimensions based on their contribution to the model prediction performance[22]. We validated this approach with simulation experiments and with cross validation on the real data, which we found to slightly underestimate dimensionality due to the need to divide data into training and test sets (Figure 3c).

We found that population-level representations of all task variables were at least two-dimensional, and at least three-dimensional in monkey A (Figure 3; Supplementary Table Supplementary Table 1). Figure 3a shows the variable-specific components revealed by mTDR for an example neuron. The first three columns show the timecourse of this neuron's activity within the first three dimensions of the corresponding variable's encoding space. The timecourses represent the columns of the temporal component matrices S_p, scaled by the levels of each of the task variables x_p (eq. 2). Thus, each trace represents the inferred contribution of each dimension to the neuron's PSTH from the different settings of the associated task variable. The rightmost column of Figure Figure 3a shows the model-based estimate of the neuron's net time-varying response to each task variable. Summing these responses gives the model-based reconstruction of the neuron's PSTH ("model PSTH") for each task condition; this matches the neuron's true PSTH to high accuracy (bottom). Because each neuron weights each dimension independently, the fitted model collectively accounts for a wide variety of conditional PSTHs (Figure 3b). Note that the data were not temporally smoothed and no smoothness constraints were included in the model, indicating that the smoothness of the timecourses is a property of the data.

To examine whether the model was flexible enough to capture the diverse response profiles observed across the population, we calculated the R^2 of the conditional PSTH for each condition using held-out data. We found that the R^2 of PSTH reconstructions increased with firing rate, with some neurons achieving R^2 greater than .9 (Figure 3d). The dependence of R^2 on firing rate likely reflects higher signal-to-noise ratio in higher firing-rate neurons.

We also measured how much of the variance from held-out trials could be explained by each of the learned subspaces alone (Figure 3e). We defined each subspace by a set of orthonormal vectors ordered by the amount of variance explained (for details see Methods and Supplementary Note 4.1). We found that all dimensions contributed to the variance of at least some neurons, but that different neurons had their variance distributed differently across components. For example, for the decomposition in Figure 3a, dimension 3 of the abs(motion) axis contributes more than dimension 1 despite the first dimension describing most of the variance across the population. These findings verify that the mTDR model captures high-variance dimensions and that the model describes a large fraction of the variance of the PSTHs for most neurons, despite the model being relatively low dimensional.

State-space trajectories reveal dynamic encoding

To explore the dynamics of population-level encoding during decision formation, we examined projections of neural activity from held-out trails onto the estimated subspaces (Figure 4a–d; Extended Data 1, Extended Data 3; Supplementary Videos). In contrast to previous findings[1;8], we found that the encoding of the stimulus variables (motion and color) was not transient, but persisted throughout the recording epoch. Projections of population activity onto a single motion or color axis identified with classic TDR suggested that sensory axis projections decay rapidly after stimulus onset[1]. However, mTDR revealed that stimulus information persists by rotating within multi-dimensional motion and color subspaces (Figure 4a,b).

More generally, we observed that for nearly all subspaces, the neural trajectories on nearly all task conditions initially moved outward along a single axis and then began rotating in a consistent direction (Figure 4a–d). This observation prompted us to identify the precise orientation of this initial axis and when, or if, the trajectories curved into a second dimension. We therefore sought a procedure that would identify an orthogonal set of axes ordered by the times at which population activity first projects onto them. The resulting method, which we call “sequential principle components analysis” (seqPCA), identifies the direction the trajectories are moving and the time at which a change in direction occurs (see Methods, Supplementary Note 9). We used seqPCA to obtain an interpretable set of axes for the subspaces identified by mTDR.

Using seqPCA, we identified a orthonormal basis for each subspace, with axes that we labelled as “early,” “middle,” and “late”, based on the times at which they became active during the task period. (Figure 4a–d). By definition, the early axis accounted for the majority of the variance in neural trajectories during the time period immediately after stimulus onset. Variance that was not described by the early axis but emerges sometime after stimulus onset is captured by the middle axis. The late axis accounts for activity that is not accounted for by the early and middle axes, but is present as the epoch transitions from the stimulus presentation to the delay period. Projections onto the seqPCA axes show clear times at which task variable information becomes available onto each axis (right-side panels in Figure 4a–d). For all subspaces, we found that the early epoch is characterized by loading of the projections almost exclusively onto a single axis. In contrast, the middle and late epochs were two dimensional, or higher.

We found that the transience of the early stimulus axes resembles that of the stimulus encodings using the TDR method[1]. Indeed, we found that our early axis was well correlated with the TDR axes (see Supplementary Note 10). It is therefore apparent that the middle and late seqPCA axes permit the stimulus information to persist. We compared projections onto the subspaces learned by mTDR with the 1D axes of TDR (see Supplementary Note 10) and found that while stimulus information appeared transient for TDR, the mTDR projections were both larger and more persistent (Figure 4e, Extended Data 4). Finally, the population-level representation of choice, context, abs(motion), and abs(color) also exhibited multi-dimensional structure (Figure 4c,d; Extended Data 1). We describe this structure and discuss its consequences in subsequent sections.

Trajectories exhibit rotational dynamics

The projections of neural population activity onto motion, color, choice, and context exhibited rotations after early-axis activity reached a peak and middle-axis activity began to increase (Figure 4a–d). This observation is supported by the fact that the trajectories are ≥ 2 dimensional during this period (Figure 4a–d, right panels). While rotations are inherently ≥2-dimensional, the fact that we found trajectories to be ≥2-dimensional need not imply rotations. We therefore identified the plane of greatest rotation of the trajectories using jPCA[18] (Extended Data 2, Extended Data 6), and observed clear rotational structure. The two dimensions of the jPCA plane accounted for a relatively large amount of the variance for all task variables (Supplementary figure 1). Condition-shuffled projections yielded no apparent sequential or rotational structure (Supplementary Note 5, Supplementary figure 2, Supplementary figure 3).

In order to rigorously examine the presence of rotational dynamics, we examined the angle of rotation that the trajectories traversed from the beginning of the middle epoch to the end of stimulus viewing (Figure 4f). We reasoned that for trajectories to be consistent with rotational dynamics they would have to have monotonically changing angle of rotation. We compared the angle of rotation to samples from the null distribution corresponding to the maximum entropy distribution with the same second order moments as the data[23] (Figure 4f, see Supplementary note 5 for details). We found evidence for rotational dynamics in motion, color, choice, and context subspaces, although rotations were less consistent with the trajectories of the color encoding for monkey F (Extended Data 3, Extended Data 6. These results indicate that rotational dynamics are not trivially present in these data and that we observe them in most of the linear subspaces examined.

Projections onto the subspaces for the absolute values of motion and color (abs(motion), abs(color)) were qualitatively different from those of the linear terms (Extended Data 1). While they clearly encoded the absolute values of the stimuli, evidence for rotational dynamics was not significant (Figure 4f, Extended Data 2, Extended Data 1).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Characterizing neural selectivity across subspaces

We used the mTDR model and seqPCA to examine the how tuning properties of these cells changed over time and the relationships between tuning between cells. Individual neurons exhibited complex mixtures of early, middle and late responses (Figure 5a). While the population tuning of some task variables (abs(motion), abs(color)) were dominated by the early response none of the task variables were found to display clustering, but a continuous distribution of tuning across all three seqPCA axes. Late axes tended to explain less of the population variance, especially for color, choice, and abs(motion), but were responsible for explaining the majority of the variance for at least some neurons. This is evident from the dirth of units in the “late” vertices of Figure Figure 5a.

Also notable was the low density of cells near the early/late axis (i.e. left-arm of the ternary plots in Figure 5a). A low density of cells near the early/late edge of the plot indicates that there are few cells that encode a task variable at the beginning and end of stimulus viewing but lose sensitivity to a task variable in the middle of stimulus viewing. The lack of cells with a gap in the timing of encoding implies that individual cells tend to encode task variables in continuous epochs, even if only transiently.

The subspaces identified by mTDR for motion, color, and choice, were positively correlated (Figure 5b, c). More specifically, the weights defining the motion and color bases were correlated with the choice weights but not with one another, indicating that motion and color representations both contributed to the choice encoding but that there was little cross-stimulus interference between representations.

Accurate stimulus decoding coincided with the onset of rotational dynamics

The mTDR model provides a framework for population decoding by maximum likelihood (see Supplementary Note 6). This framework allows our decoding analysis to be consistent with the results of dimensionality reduction. We can therefore investigate how and when the features of the low-dimensional trajectories translate into putatively perceived stimuli and behavior, and whether or not these features may be read out by downstream populations. While decoding of task variables does not imply a causal role for the encoded variables in PFC function, decoding analysis can provide a clearer picture of the dynamics and fidelity of task variable encoding.

For decoding analyses with monkey A, we used 4-fold cross validation in which the held out trials were used to produce 100 pseudosamples (with replacement) for decoding (for monkey F we used 2-fold cross validation with similar results). The resulting decoded values were averaged over pseudosamples and cross validation folds.

Stimuli could be accurately decoded within ≈150ms of stimulus onset for the motion stimulus and within ≈200ms for the color stimulus, roughly corresponding to the time of transition between the early and middle seqPCA axes (Figure 6a). The values of the decoded stimuli were constant by the start of the middle epoch for both contexts and the variance of the decoding decreased dramatically up to this time (Figure 6b). Thus, the change in population dynamics (early-to-middle transition) within the stimulus subspace was consistent with decoding accuracy and stability. The decoded values are slightly biased toward zero in the irrelevant context, suggesting some gating of information across contexts. Moreover, we found that the decoded stimulus values of the same sign were more closely spaced than the true stimuli. This effect is consistent with observations of stimulus encoding in FOF[6] (a rodent analogue for the FEF) where accumulated evidence was found to be nonlinearly encoded.

To examine the content of stimulus encoding it is often informative to examine error trials. We therefore examined the decoded stimulus for error trials using the weakest stimulus strengths (dashed lines, Figure 6c, Extended Data 7c). For these data only the weakest stimulus strengths had enough error trials to provide reliable statistical analysis[1]. For monkey A, we found that the decoded stimulus values on error trials were similar to correct-trial decoding but were opposite in sign; suggesting that the origin of errors was (on average) an incorrect percept.

Choice Decoding

In order examine how the decision may have evolved over the course of stimulus viewing we next studied how and when decision information became available in PFC and the dynamics of choice encoding. While we encoded stimuli as continuous-valued variables in our model, choice was encoded as binary. Therefore, we examined the log likelihood ratio (LLR) (for details, see Supplementary Note 6.3) over time of pseudotrials sampled from held-out data between the likelihood of a preferred, versus an anti-preferred, choice (Figure 7a).

The magnitude of the LLRs increased monotonically over time indicating an increasing strength of the decision signal. However, the magnitude of the LLR did not differ strongly with respect to context, direction of decision, stimulus strength, or whether the trials were correct or error trials (dashed lines, Figure 7a). By transforming the LLRs into decision probabilities (Figure 7b, see Supplementary Note 6.3) we could examine a moment-by-moment probability of the animal's choice and estimate when the decisions were unequivocal. We found that the choices could be discriminated with better than 95% accuracy as early as 300-350ms following stimulus onset (Figure 7b). This timing corresponded to the time of transition between the early and middle seqPCA axes for choice. Similar results were observed for monkey F (Extended Data 8). These results suggest that on average the animals had made their decisions well before stimulus offset regardless of the stimulus coherence and that decisions were coincident with a change in dynamics from a linear to rotational within the choice subspace.

Restricting the choice subspace to only the early, middle, or late axes the LLRs displayed the same invariance to choice, stimulus strength, context, and correct/error trial identity as the full model. For both monkeys, the early axis provided the majority of the available information about the decision and early axis decoding alone is nearly as accurate as the full model (Figure 7d, Extended Data 8d). However, the middle and late axes also displayed information about the choice later during stimulus viewing.

Because we can decode the animals' decisions with the early axis alone, it would seem as though the middle and late axis information is redundant and it is unclear what the purpose of these axes are. Similar multidimensional encoding of decision has been observed previously in premotor cortex[24].

Context Decoding

We examined the context signal using the same LLR method as our analysis of choice (Extended Data 9a, Extended Data 10a). The context evidence did not differ strongly across decision, stimulus strength, or whether the animal provided a correct or incorrect response. Transforming the LLRs into a probability of the perceived context (Extended Data 9b, Extended Data 10b) showed that the correct context could be identified for both monkeys on the majority of pseudotrials from the first time point, which is consistent with the fact that the context cue was presented 650 ms before stimulus onset[1]. These patterns hold for LLRs of error pseudotrials as well as for decoding restricted to only the early, middle, or late subspaces (Extended Data 9c,d, Extended Data 10c,d). These findings demonstrate that accurate context information was available in PFC for the vast majority of both correct and error trials, suggesting that confusion about context was not a significant source of errors.

Discussion

Our analyses have shown that PFC encodes individual task variables in distinct multidimensional subspaces within which the representation changes over time. The population activity patterns representing each task variable tended to follow a stereotyped pattern of 1D/linear encoding, followed by rotational dynamics. Our ability to make these observations was enabled by a new method of dimensionality reduction that is based on a generative model of the data.

We found that the dynamic nature of encodings in PFC requires multiple dimensions of neural population activity for accurate characterization. In particular, only multidimensional encoding, as opposed to 1D encoding, captures the persistence of stimulus information in PFC throughout the stimulus-viewing epoch (Figure 4a,b). This finding complements the original report of these data[1], suggesting that previously reported transient stimulus encoding in PFC is only consistent with the early encoding axis (Figure 4e). While the mechanisms of selection and integration proposed in Mante et al.[1] are consistent with the early evolution of trajectories revealed here, by themselves they cannot readily explain the following rotational dynamics. Our observations resemble multi-dimensional stimulus coding that mixes transient and persistent components[24] as well as population code “morphing”[16], where the optimal weights for decoding from population activity change over time, although the results shown here are on a time scale that is nearly an order of magnitude faster than previously reported.

While we validated our method for identifying the “true” dimensionality of the data using simulation experiments, it is unclear if the dimensionality would differ under different experimental conditions. Specifically, the dimensionalities we learned are likely to be influenced by a variety of factors[25] including the sample size, the fraction of neurons observed, the intrinsic model dynamics, and the task complexity. Some of these factors may explain the differences in dimensionality between the two animals in the present study, where the dimensionalities of monkey F were lower than monkey A in correspondence with

Europe PMC Funders Author Manuscripts

smaller sample sizes and fewer recorded cells. However, we emphasize that during the early encoding, nearly all trajectories are 1D and only afterward are ≥2D. The fact that trajectories are multidimensional after the first transition may be a reflection of their rotational nature during this epoch. Rotations are inherently ≥2D since they require both sine and cosine parts for each axis of rotation.

The mTDR method is distinct from unsupervised dimensionality reduction methods like PCA or factor analysis in that it uses information about the experimental variables of interest on each trial. The method is also distinct from previously proposed supervised methods[1;26–29] in its use of an explicit generative model to describe the transformation from task variables to neural activity patterns. This distinction not only allows us to make predictions of population responses to experimental contingencies not observed in the data (something not possible for methods based on the conditional PSTHs like dPCA without model-based interpolations[27]) but it allows us to apply the tools of probabilistic modeling and inference to estimate both the model parameters and the dimensionality of the encoding.

Our approach (eq. 5) is similar in principle to that used by Mante et al.[1] and other linear regression models used previously (see examples[30–33]). However, ours is distinguished in its explicit specification of low-rank regression parameters and neuron-specific noise variance. Future iterations of our model may be improved by accounting for nonlinear mapping of stimuli onto neuronal responses[30], by modeling of noise correlations between simultaneously recorded neurons, and accounting for variable trial lengths.

Much theoretical development has rested on the notion that single-neuron spike rates map onto an evidence accumulator but recent evidence in the frontal orienting field (FOF, a rodent analogue of the FEF) has challenged this view[6], suggesting that this region can be better described as maintaining a running motor plan (saccade for FEF, orienting for FOF) based on the evidence accumulated[6;7]. While our analysis does not aim to suggest a causal role of FEF, the results of the present study could be interpreted as supporting this view, where the early dynamics represent an evolving decision and the rotational dynamics indicate an evolving motor plan, but more work is needed to determine the precise role of FEF, and PFC more generally.

Functional significance of sequential subspaces

Our analysis revealed temporally segregated dynamics with early-axis activity transitioning to middle-and late-axes, with rotations dominating at around 200–400 ms after stimulus onset (Figure 4, Fig. Extended Data 2). The temporal separation of the early/linear and rotational subspaces suggest that these are subspaces within which distinct computations are evolving[18;20;34] or have independent sets of downstream targets[19].

With the present data we can only speculate about what the nature of these different computations must be but the present analysis indicates the possibility that the early-epochs are concomitant with the temporal window that decision making is performed. For example, the timing of transition between early and middle epochs is consistent with the timing of accurate decoding of the animals’ decisions from single pseudotrials (Figure 7, Extended Data 8). This time frame is consistent with the timing of saturation of the chronometric curve for the traditional kinematogram task[35–37], with the distribution of step times in the stepping model of evidence accumulation[38], and with early weighting of evidence in visual discrimination tasks[39]. This evidence suggests that the transition from linear to rotational dynamics is a correlate of decision commitment.

A similar sequence of dynamics has been observed in population activity from premotor cortex that corresponds to distinct "preparatory" and "movement" epochs[18–20;34]. However, in these studies the transitions in dynamics could be linked directly to an overt action (arm movement) while our animals would not have made an overt action (saccade to target) until 300-1,500 ms after the end of our analysis window[1]. Therefore, if the animal has made its decision then it would have done so only covertly.

These distinctions, however, may be superficial. The qualitative features of our results reflect those in motor cortex strikingly well[18–20;34] suggesting that common mechanisms may be at work in both motor execution and decision making. Indeed, FEF is defined as a region that elicits eye movement under stimulation[40;41] and has been implicated as a region important for visual decision making[1;4;6;9–12;14;15], oculomotor planning[42], and covert visuospatial attention[43;44]. Thus, we may think of FEF as itself a premotor area responsible for visuospatial attention and motor planning associated with decision making[4;6;7;9]. The dynamic transitions in our analysis could be interpreted as signaling decision commitment[6], or as signaling a covert action (saccade preparation), in analogy with the transitions observed between preparatory and movement periods seen in premotor cortex[18;19;34]. Single-trial population analysis and analysis of delay and saccade epochs of these experiments may shed light on how the dynamics we observe reflect the animals' decisions.

Some subspaces lacked a distinct late component (eg. color and choice subspaces for Monkey A, Figure 4a,c). However, it is possible that the middle seqPC for some task variables served a similar role as the late seqPC for others; preparing the network for a new set of targets or storing the memory of the stimuli as persistent activity over the course of the delay period. The number of seqPC's needed to describe the population activity may reflect the rate that trajectories rotate into new encoding directions, and therefore correspond to a quantitative rather than a qualitative difference in encoding. Future work should be aimed at identifying the significance of the dimensionality of the encoding relative to the sequential dynamics.

The nature of dynamic encoding for the context variable remains mysterious. Context encoding for both animals displayed clear and consistent dynamics (Figure 4d, Extended Data 3d) including rotations (Figure 4f). Furthermore, while most of the predictive capacity of the context encoding lies in the early subspace (Extended Data 9, Extended Data 10), where context is encoded throughout the stimulus viewing period, context encoding at the single-neuron level is broadly distributed across the early, middle, and late axes (Figure 5b, Extended Data 5), indicating that some neurons do not encode context until well after stimulus onset. Further work is needed to determine what, if any, function these dynamics serve in decision making and memory. The uniqueness of these phenomena to the present setting is an active area of research.

Differences in encoding between animals

The two monkeys in this study displayed similar, but not identical, encoding properties. For example, the encodng trajectories for motion are similar (Figure 4a, Extended Data 3a) but we found obvious differences between the encoding trajectories for color (Figure 4b versus Extended Data 3b). For monkey A the color trajectories closely resemble the trajectories for motion (Figure 4a,b) while for monkey F the color trajectories do not display obvious rotations (Extended Data 3b). Choice and context trajectories in monkey F appear to be similar to those of monkey A (Figure 4e,g and Extended Data 3e,g) but display less pronounced rotations, (Extended Data 2, Extended Data 6). These across-animal differences verify that rotational dynamics are not trivially present in these data and while it is unclear precisely what function they serve they are a potentially important feature of encoding in PFC.

While the reason for differing dynamics between the color encoding for monkey F and the other stimulus encodings is unclear we do have some behavioral clues as to its effect. For example, the color-context psychometric curve for monkey F was somewhat more shallow than for motion as well as for both motion or color for monkey A (Extended data Fig. 2d in[1]), and motion served as more of a distraction during the color task for monkey F than for monkey A, suggesting that color discrimination task was more difficult for monkey F. Furthermore, we found that the decoding accuracy for color in monkey F was worse than for monkey A (Figure 6, Extended Data 7) suggesting that color information was more poorly represented in PFC for monkey F. Although not definitive, together these results suggest that monkey F may have had more difficulty with color coherence perception and that indistinct encoding features are a correlate of perceptual uncertainty. Future experiments could be aimed at examining this hypothesis.

Decoding of error trials suggests sources of errors

There are three ways that the animals may commit an error: the animal perceived the wrong stimulus (e.g. perceived left motion on a right-motion trial); the animal was confused about the context (e.g., made its decision using the color information in the motion context); or the animal made a random choice (i.e., a “lapse” trial). The results of this analysis for monkey A at the weakest stimulus strengths, indicate that the animal perceived the wrong stimulus. The decoded context on average was the correct context, (Extended Data 9), ruling out whether or not the animal was confused about which stimulus it was supposed to attend. Lapse errors are also unlikely to contribute significantly to the animal’s behavior. The psychophysical curves of monkey A suggest a small lapse rate, if any[1], and stimulus decoding indicates that the perceived relevant stimulus on error trials was of the opposite sign as the stimulus that was presented (Figure 6c). Together, these observations indicate that most error trials are based on an incorrect perception of the relevant stimulus. A more direct trial-by-trial analysis of simultaneously recorded neurons would be useful in confirming this hypothesis.

The results for monkey F are more difficult to interpret. The decoded stimuli for error trials appear to be close to 0, indicating an ambiguous stimulus (Extended Data 7b). Furthermore, the choice signal on error trials appears to be present earlier on average than on correct trials and is present on some trials as early as the first time point (Extended Data 8b) suggesting that the animal may have made its decision before even viewing the stimulus, and that a significant source of errors for monkey F were lapses.

Given the present data, it may be impossible to distinguish the neural correlates of decision making from those of planning for the eventual saccade. Recent work has shown that there may be independent cortical signals for evidence accumulation and decision commitment in other cortical areas[39]. It may be unlikely using these data to distinguish between a deliberate effort to make a stimulus discrimination and the formation of a motor plan[31].

Nevertheless, the results presented here demonstrate the utility of mTDR for the analysis of neuronal population data and provide a description of PFC dynamics that should serve as important constraints on future models of the mechanisms of PFC function.

Methods

Detailed description of model

High-dimensional description of observations—Our model describes trial-by-trial neuronal activity with a linear regression with respect to the task variables. We assume that the activity of the i^{th} neuron $y_{i,k}(t)$ at time t on trial k can be described by a linear combination of P task variables $x_k^{(p)}$, $p = 1, \ldots, P$ (eg. stimulus variables, behavioral outcomes, and nonlinear combinations thereof), such that

$$y_{i,k}(t) = x_k^{(1)}\beta_{i,1}(t) + x_k^{(2)}\beta_{i,2}(t) + \cdots + x_k^{(P)}\beta_{i,P}(t) + \epsilon_{i,k}(t)\,. \tag{3}$$

where the P values of the task variables $x_k^{(p)}$ are known, the $\beta_{i,p}(t)$ are unknown coefficients, and $\epsilon_{i,k}(t)$ is noise. This basic model structure is identical to that of the regression model used in[1] and has been successfully employed in characterizing neuronal activity of single neurons in other studies of perceptual decision making[4746]. In cases where we include a time-varying mean rate that is independent of the task variables, we define $x_k^{(P)} \equiv 1$ for all k, and the P^{th} component becomes the time-varying mean.

To represent all neurons simultaneously, we concatenate the responses into a vector $\mathrm{y}_k(t)$ and write

$$\mathbf{y}_k(t) = x_k^{(1)}\,\boldsymbol{\beta}_1\,(t) + x_k^{(2)}\,\boldsymbol{\beta}_2\,(t) + \cdots + x_k^{(P)}\,\boldsymbol{\beta}_P\,(t) + \,\boldsymbol{\epsilon}_k\,(t), \tag{4}$$

where $\mathbf{y}_k(t) = (y_{1,k}(t), \ldots, y_{n,k}(t))^\top$, $\boldsymbol{\beta}_p(t) = (\beta_{1,p}(t), \ldots, \beta_{n,p}(t))^\top$, and $\boldsymbol{\epsilon}_k(t) = (\epsilon_{1,k}(t), \ldots, \epsilon_{n,k}(t))^\top$.

For trial epochs of duration T we can regard all observations on a given trial to be a matrix, $\mathrm{Y}_k = (\mathrm{y}_k(1), \ldots, \mathrm{y}_k(T))$, giving the observation model

$$\mathbf{Y}_k = x_k^{(1)}\mathbf{B}_1 + x_k^{(2)}\mathbf{B}_2 + \cdots + x_k^{(P)}\mathbf{B}_P + \mathbf{E}_k, \tag{5}$$

where $\mathbf{E}_k = (\epsilon_k(1), \ldots, \epsilon_k(T))$, and $\mathbf{B}_p = (\beta_p(1), \ldots, \beta_p(T))$. For the present study, we assume the noise is normally distributed $\epsilon_k(t) \sim \mathcal{N}(0, \mathbf{D}^{-1})$ for all trials k and times t, where $\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is a $n \times n$ diagonal matrix of noise precisions.

Low-dimensional description of observations—With no additional constraints our observation model (5) is extremely high dimensional and is effectively a separate linear regression for each neuron at every time point. This would only be a sensible model if we believed that neurons were not in fact coordinating activity between each other or across time. To define our low-dimensional model we can describe each $\mathbf{B}_p$ by a low-rank factorization, i.e. $\mathbf{B}_p = \mathbf{W}_p \mathbf{S}_p$, where $\mathbf{W}_p$ and $\mathbf{S}_p$ are $n \times r_p$ and $r_p \times T$ respectively, where $r_p = \mathrm{rank}(\mathbf{B}_p)$. Equivalently, we can say that r_p is the dimensionality of the encoding of task variable p. This is equivalent to saying that the characteristic response of each neuron to the p^{th} task variable can be expressed as a linear combination of r_p weighted basis functions $\beta_i^p(t) = \sum_{j=1}^{r_p} w_{i,j}^{(p)} s_j^{(p)}(t)$, where r_p is the dimensionality of the encoding, $\{s_j^{(p)}(t)\}_{j=1}^{r_p}$ are a common set of time-varying basis functions, and $\{w_{i,j}^{(p)}\}_{j=1}^{r_p}$ are neuron-dependent mixing weights.

The mTDR model does not impose any orthogonality between task variables or task variable subspaces. This permits accurate recovery of subspaces even when the encoding dimensions are correlated, which can result in correlations between task variable representations, as we demonstrate in Supplementary Note 8. It is desirable therefore to be able to visualize the part of the encoding of each task variable that is unmixed. We therefore orthogonalize the subspaces with respect to correlated subspaces for visualization in Figure Figure 4.

Marginal estimation of model parameters

The goal of inference is to estimate the factors of $\mathbf{B}_p$ and the ranks r_p. Our proposed estimation strategy, for computational and statistical efficiency, is to estimate only one set of factors ($\{\mathbf{W}_p\}$ or $\{\mathbf{S}_p\}$). This is possible when we integrate out one set of factors. For example, if we define a prior probability density over the mixing weights $p(\mathbf{W})$, then for data likelihood $p(\mathbf{Y}|\mathbf{W}, \mathbf{S})$ the marginal likelihood of the matrix of time-varying basis functions $\mathbf{S}$ can be obtained by

$$p(\mathbf{Y} \mid \mathbf{S}, \lambda) = \int_{-\infty}^{\infty} p(\mathbf{Y} \mid \mathbf{W}, \mathbf{S}, \lambda) p(\mathbf{W}) d\mathbf{W}. \quad (6)$$

In principle, either set of factors may be selected for marginalization. In practice however the set of factors with lowest dimension should be selected to keep computational costs low. In this paper we focus on the case where $T \ll n$ and we therefore will estimate the set of weights $\{\mathbf{S}_p\}$ while integrating over $\{\mathbf{W}_p\}$. The fact that either set of factors may be determined in this way means that there is a duality between rows and columns imposed by this model that is similar in principle to the duality between factors and latent states for probabilistic principle components analysis[48].

Europe PMC Funders Author Manuscripts

If we let the noise distribution and prior distribution of **W** both be Gaussian then we can use standard Gaussian identities to derive the marginal density $p(\mathbf{Y}|\mathbf{S},\boldsymbol{\lambda})$ and the corresponding posterior density $p(\mathbf{W}|\mathbf{Y}, \mathbf{S}, \boldsymbol{\lambda})$. A simple starting assumption would be to let all elements of **W** to be independent standard normal, (i.e., $w \sim \mathcal{N}(0, I_{\tilde{r}n})$ where $\tilde{r} = \sum_p \mathrm{rank}(\mathbf{B}_p)$). We therefore assume that the weights are a priori independent and that the noise variance is independent across both neurons and time. In principle, our framework supports the application of more structured priors and noise covariances, but we will not explore more elaborate models in this paper. Further details are developed in Supplemental Section 2

Experimental details

A detailed description of these data have been published previously[1]. Briefly, two adult male rhesus monkeys were trained to perform a context-dependent 2-alternative forced-choice visual discrimination task. At the beginning of each trial the monkeys were cued (Figure 1a) to respond to either the motion or the color parts of the stimulus. After the context-cue presentation two targets appear for 350 ms, followed by 750 ms presentation of the stimulus. The stimulus was then followed by a randomized 300–1500 ms delay after which the monkey was cued to indicate its decision with a saccade to either of the two targets. The position of red and green targets was randomized on each trial.

Electrophysiological data were recorded from tungsten electrodes implanted in the arcuate sulcus in and around the frontal eye field (FEF). Electrodes were lowered two at a time into adjacent grid holes and were advanced until at least one single-unit could be isolated, although some trials yielded multiunit activity. Data were recorded using the MAP data–acquisition system (Plexon Inc., Dallas, TX). All recorded units were included in the analysis. Spike sorting was conducted by clustering based on principle components analysis using the Plexon offline sorter (Plexon Inc., Dallas, TX). Each isolated cluster was functionally treated as a unit. Some clusters did not correspond to well discriminated, single-unit activity and were therefor deemed multi-unit activity.

All analyses presented in this paper used spike counts binned at 50ms (for model fitting and decoding) or 12.5 ms (for display of projections, jPCA, and PSTHs). Analysis windows for both monkeys started 100 ms after stimulus onset and continued for 100 ms after stimulus onset[1]. Color coherence was transformed into position evidence based on the location of the red and green target. All data were analyzed with custom scripts written in MATLAB (The MathWorks, Inc., Natick, MA).

Model structure

Inclusion of linear terms for color, motion, choice, and context were substantiated by prior work[1]. Examination of the PSTHs revealed that stimulus encoding was asymmetric (eg. unit 2 in Figure 1c), such that the encoding of the stimulus strength was stronger in one direction than the other. This suggested that the absolute value of the stimulus strengths should be jointly modeled with the linear encoding of the stimuli. Model fits using terms for the absolute value of the stimuli resulted in smaller AIC than model fits with only linear terms (Monkey A: $\mathrm{AIC}_{\mathrm{linear}} = 9.79 \times 10^7$, $\mathrm{AIC}_{\mathrm{abs}} = 7.33 \times 10^7$, Monkey F: $\mathrm{AIC}_{\mathrm{linear}} = 8.065 \times 10^7$, $\mathrm{AIC}_{\mathrm{abs}} = 8.0628 \times 10^7$).

In general, we suggest that investigators proceed with task variable inclusion in the same way that one would when performing traditional linear regression. This process should include careful consideration of the phenomenology of encoding in the population they are studying and principled model selection metrics. For large numbers of putative task variables investigators should consider model selection via sparse priors on regression coefficients.

Cross validated variance explained

To asses the variance in the population responses that is explained by our method we conducted 4-fold cross-validation (CV) where, on each fold of CV, we used a randomly selected sample of 75% of the trials as training data to estimate the parameters of the model. Using the remaining 25% of the trials as test data, we made PSTHs for every possible task variable contingency for correct trials (total of 144 conditions). The reported variance explained was averaged over the four CV folds.

When assessing variance explained, the population PSTH's for each condition was averaged over all extraneous task variables. For example, to assess the variance explained by the motion subspaces we averaged the PSTHs over all task variables except motion. We therefore had 6 sets of PSTHs for each neuron that were projected onto the motion subspace.

To determine if the variance that was explained by the estimated subspaces was greater than chance we compared the observed variance explained to the distribution of variance explained obtained by random projections. As a serrogate null distribution we generated 500 samples for each task variable of random projection weights from a normal distribution and calculated the explained variance for each sample. We then asked what the probability was of the observed explained variance being larger than the explained variance of the random projections for each neuron. We found that many neurons exceeded the 95% Bonferroni-corrected significance threshold across nearly all dimensions.

Sequential PCA (seqPCA)

The seqPCA algorithm identifies an orthogonal basis on which variance of a D-dimensional trajectory is sequentially explained. The algorithm starts by calculating the variance explained by the first singular vector of a sequence of $D \times t$ data matrices $\mathbf{Y}_t$, where t indicates the number of time points included in the data. As the number of data points increases, the first singular vector explains a larger proportion of the variance, $p_{1,t}$, until trajectories change direction, after which $p_{1,t}$ decreases. The t at which $p_{1,t}$ reaches its peak is considered a transition time and the left singular vector at this time is considered the first seqPC. Variability explained by this axis is subtracted from the data and the procedure is repeated to identify the 2nd seqPC, and so on. For details, see Supplementary Math Note 9.

The seqPCA algorithm displays some sensitivity to noise by making peaks in $p_{1,t}$, difficult to identify. However, moderate smoothing (Gaussian window, 50ms width) of the trajectories appeared to mitigate this effect. Greater robustness may be offered by translation of this algorithm into an optimization framework[49]. A related method has been developed for identification of sequential motifs of spike rasters[50].

Europe PMC Funders Author Manuscripts

Statistics

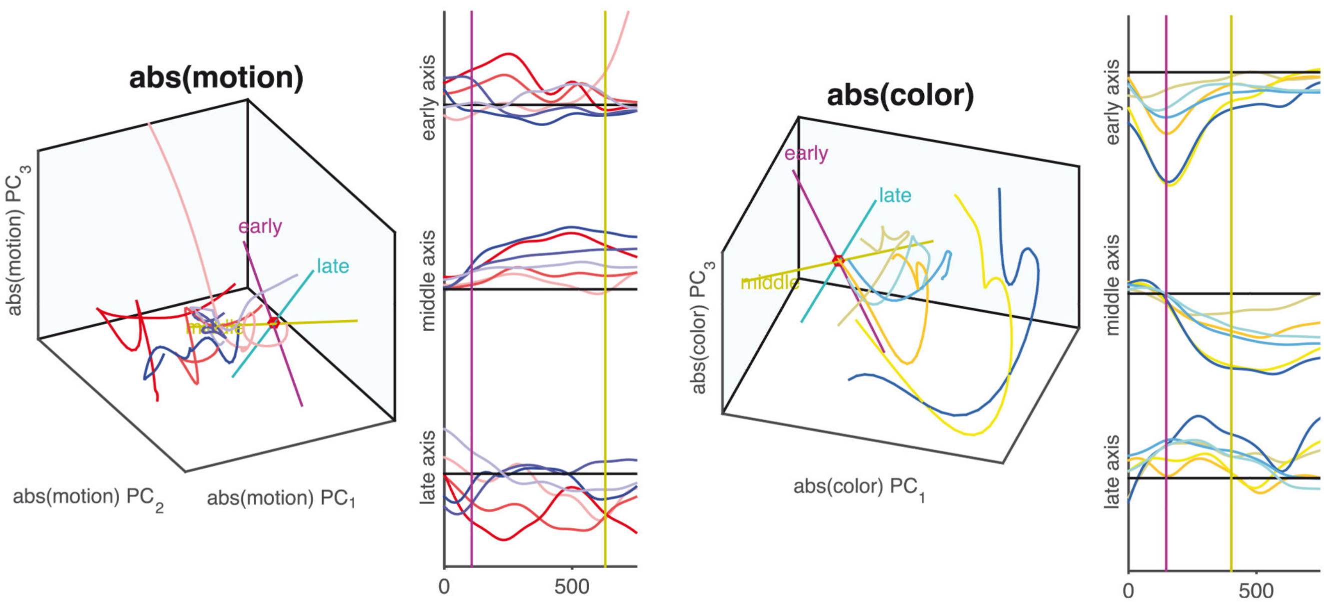
No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications[1]. Data collection and analysis were not performed blind to the conditions of the experiments.

Tests for rotational dynamics by the method of Elsayed & Cunningham[23] depend on the distribution coming from a special form of tensor-variate normal distribution that corresponds to the maximum entropy distribution for tensor-valued data with independent tensor dimensions. Data distribution was assumed to be normal but this was not formally tested.

Significance of the magnitude of inner products (Figures Figure 5, Extended Data 5) was determined by a null distribution based on the positive half-Gaussian with zero-mean and standard deviation $\sigma_0 = 1/n$, where n is the number of neurons ($n = 762$ for monkey A, $n = 640$ for monkey F), and controlled by the positive false discovery rate[45]. We conducted the same procedure using standard deviations from bootstrap samples but found that the asymptotic formula ($\sigma_0 = 1/n$) was slightly more conservative as the σ_0 was a few percent larger. The null distribution was used following Kobak et al.[27] and validated by the 1-sample Kolmogorov-Smirnov test with 1000 permutations of the weight indices as samples from the surrogate null distribution ($P = 0.78$, $D_n = 0.0207$).

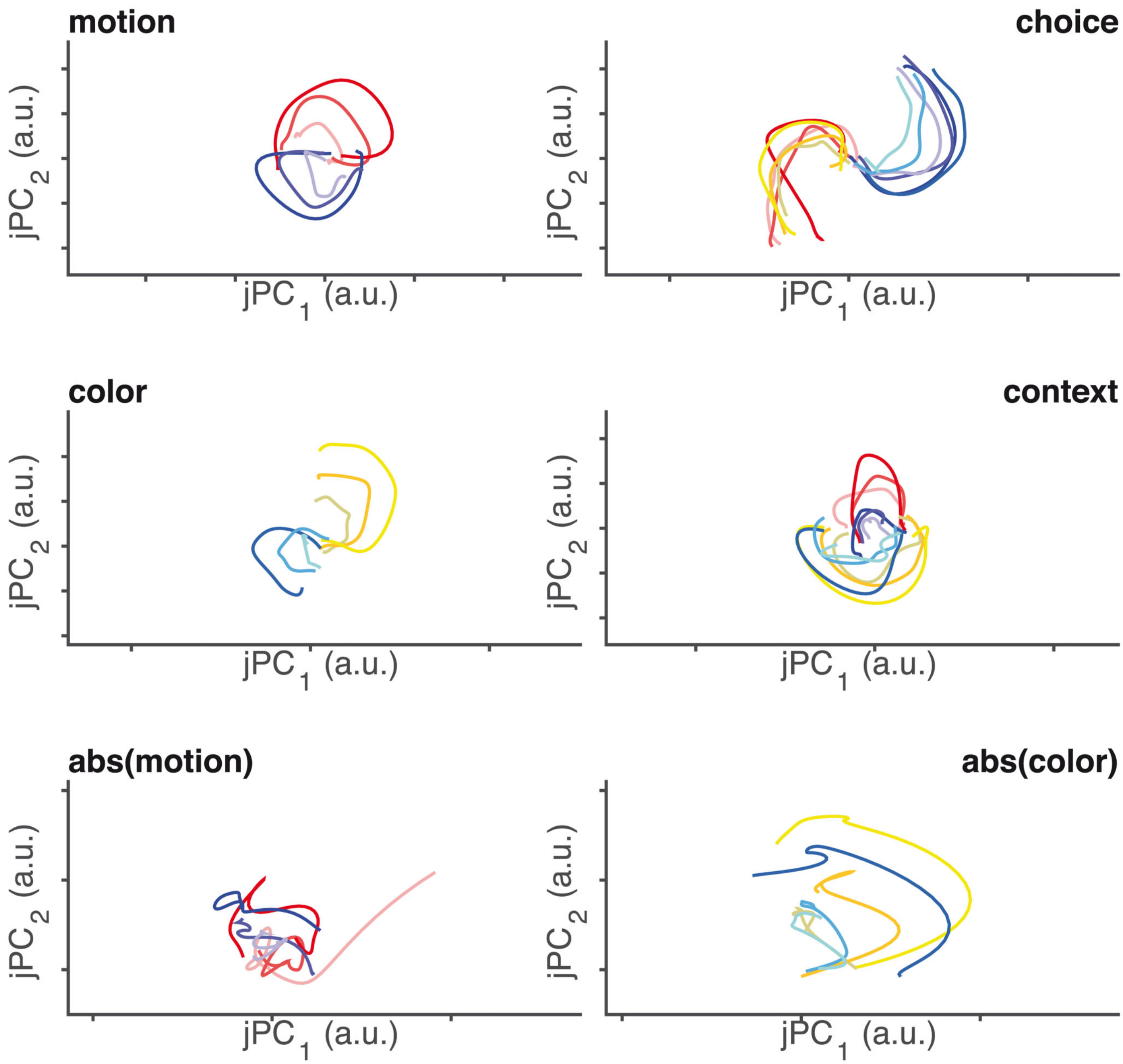
Permutation tests for canonical correlations (Supplementary Math Note Math Note Figure 3) were performed with 200 uniformly randomized permutations.

Extended Data



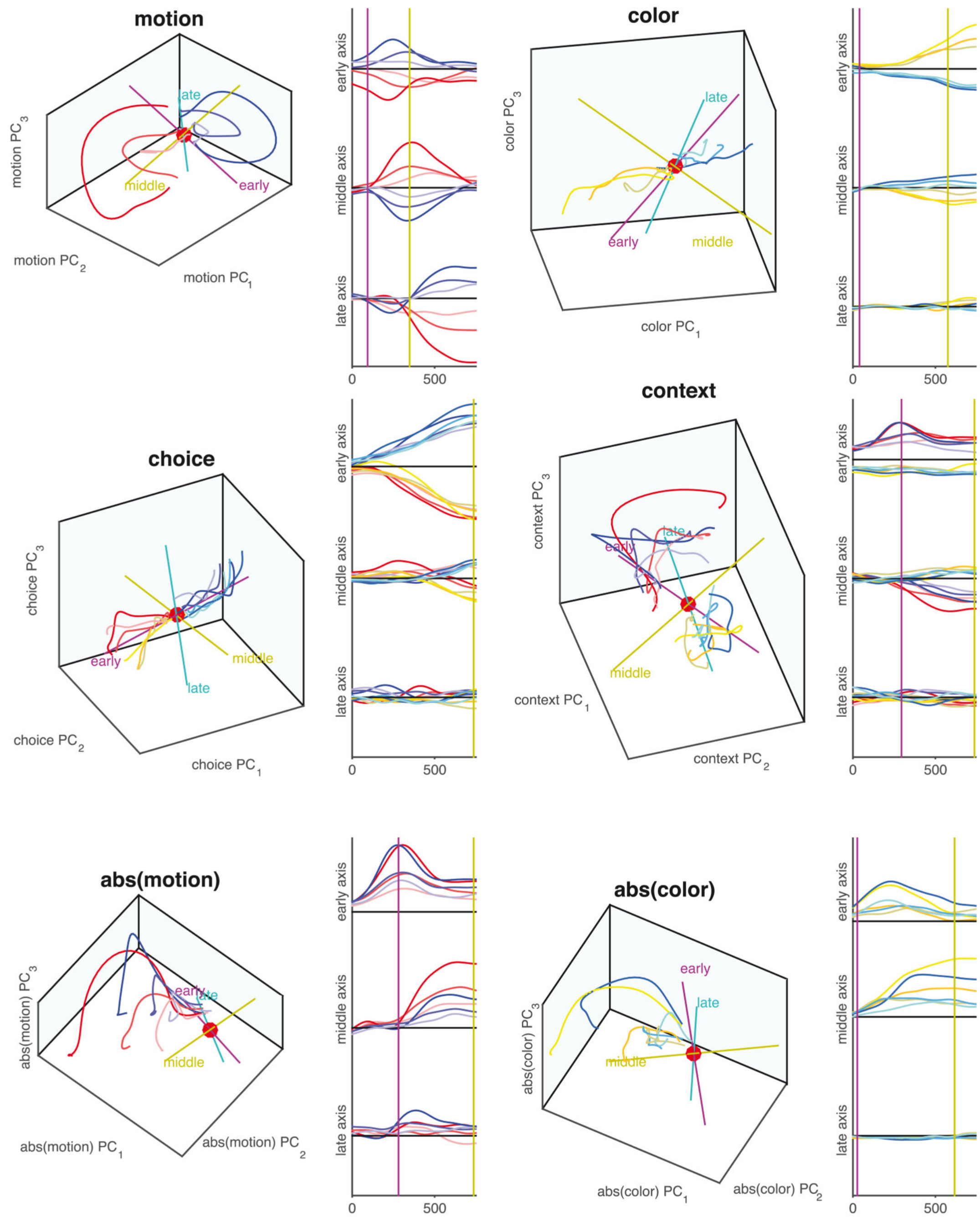
Extended Data Fig. 1. Projections of population PSTH’s onto the first, second, and third PC-axes for monkey A

a) The abs(motion) and **b)** abs(color) subspaces. Subspaces have been orthogonalized with respect to the first dimension of the choice subspace. The monkey gave the correct response for all trials used. Colored axes indicate dominant axes in the early, middle, and late periods of the stimulus epoch, as determined by the methods described in Supplementary section 9. Purple vertical lines indicate transition from the early to middle epochs. Yellow vertical lines indicate transition from the middle to late epochs as in Figure Figure 4. Plotting colors are the same as those in Figure 4. Units of the ordinate are arbitrary but all axes are on the same scale.



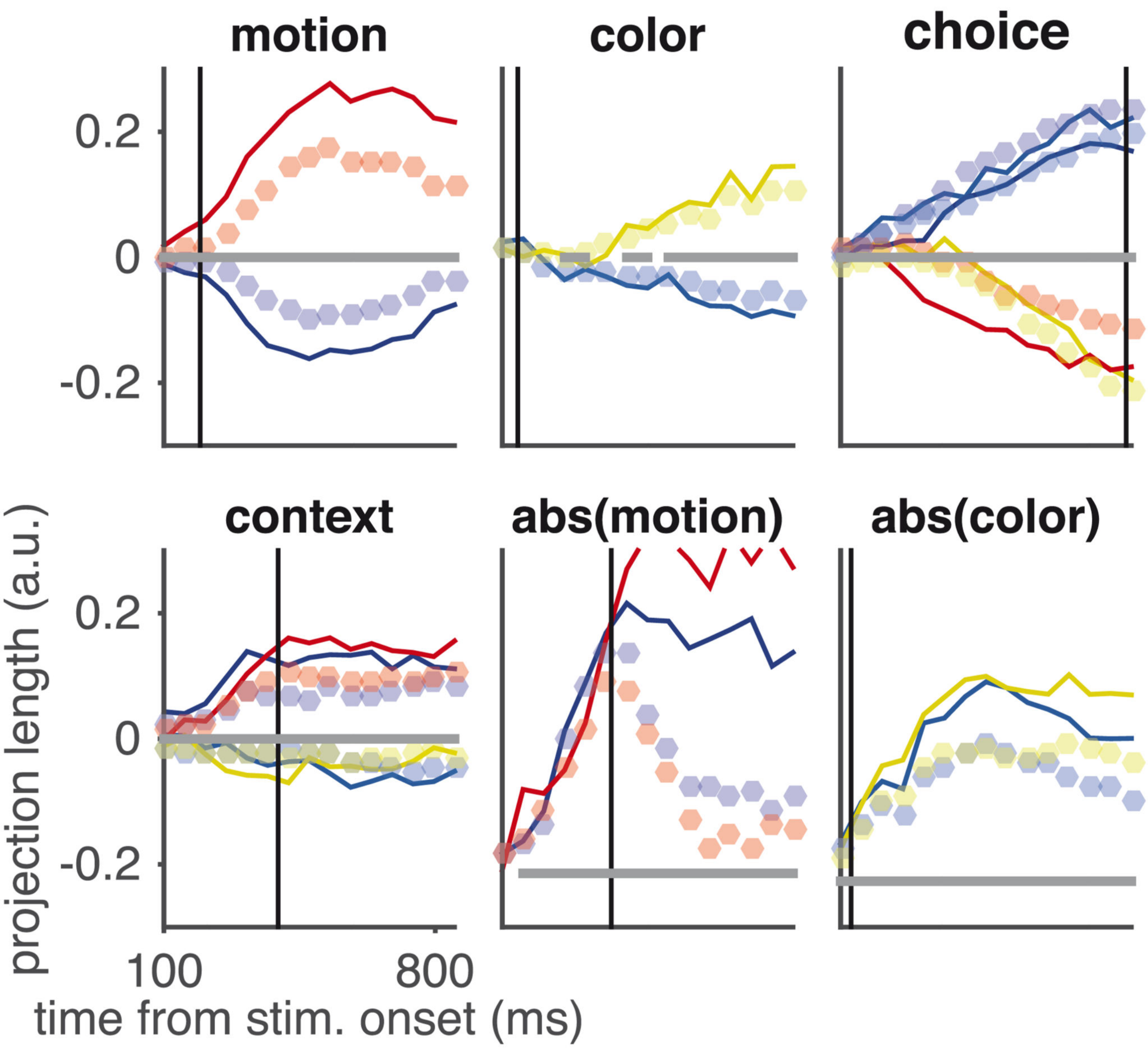
Extended Data Fig. 2. Projections of population PSTH's onto jPCA axes for monkey A.

Projections are onto the first two jPCA axes identified by the trajectories shown in Figure 4. The jPCA axes reveal strongly rotational dynamics for motion, color, choice, and context subspaces.



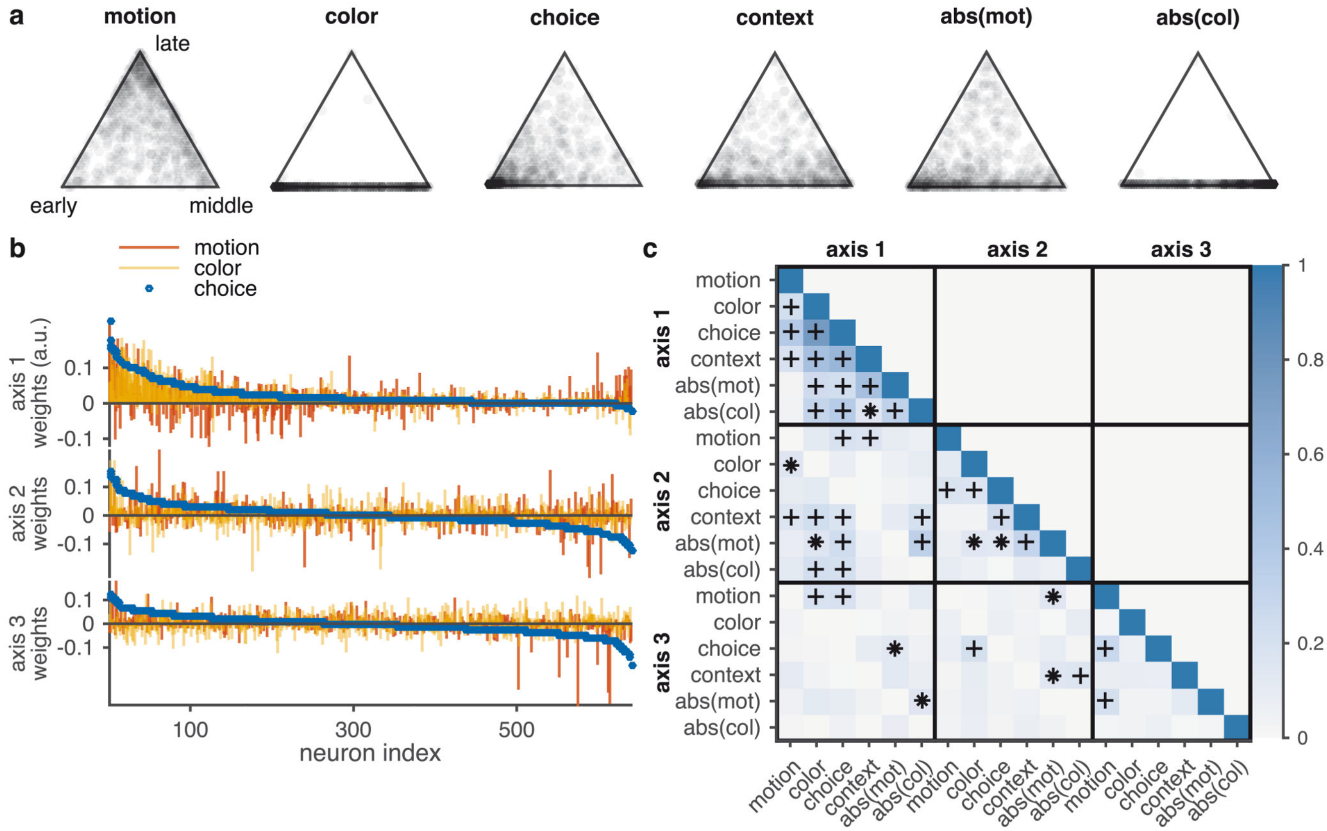
Extended Data Fig. 3. Projections of population PSTH's for monkey F onto the first, second, and third PC-axes of all task variables subspaces.

Plotting conventions and analyses are the same as those for Figure 4. Projected data is averaged over 2-folds of cross validated projections where a random sampling of half of the data was used to estimate parameters and the remaining half used to make projections.



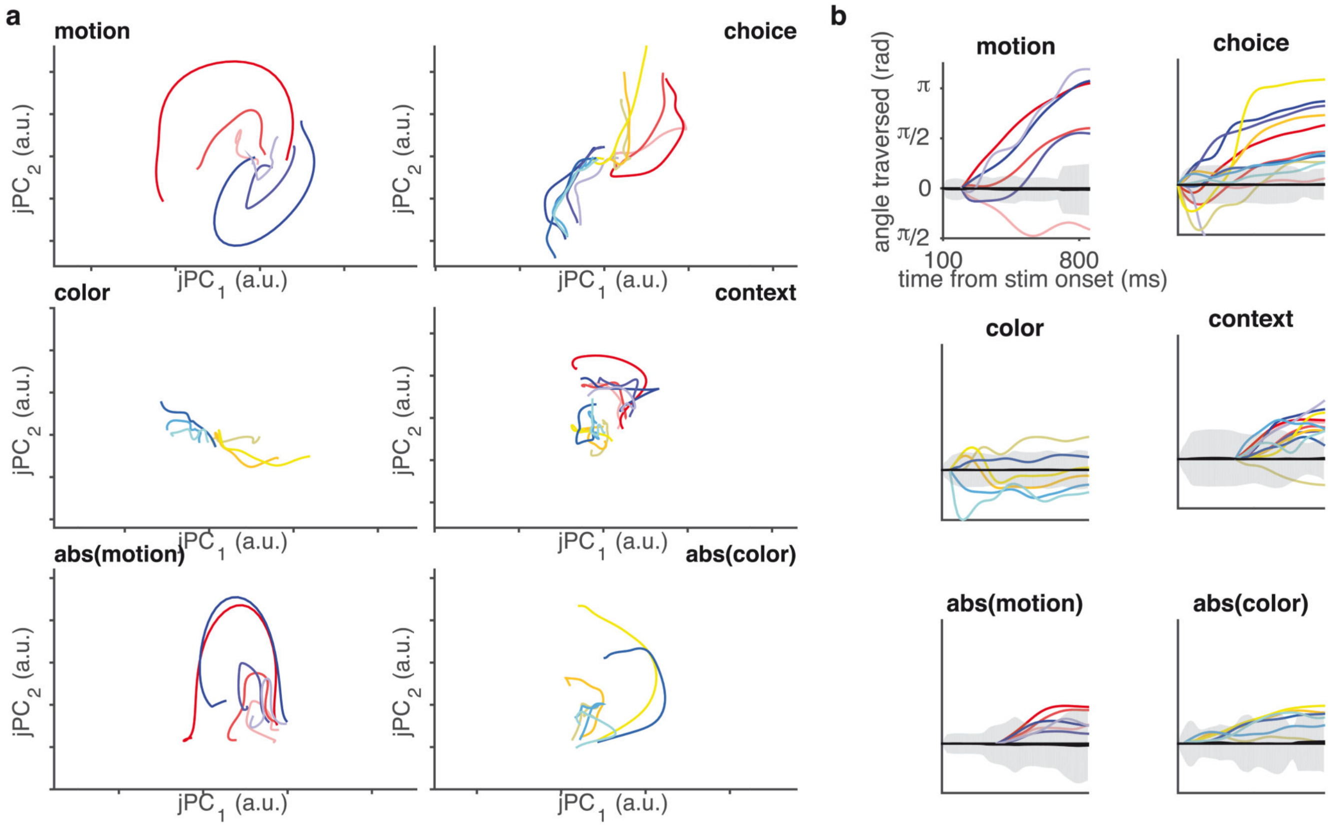
Extended Data Fig. 4. Encoding strength of population pseudosamples for monkey F onto the first three axes of all task variables subspaces.

Plotting conventions and analyses are the same as those for Figure 4. Projected data is averaged over 2-folds of cross validated projections where pseudosamples were drawn from held-out trials. Grey bars at y = 0 indicate time points where the mTDR projections had significantly stronger encoding across all stimulus levels than the 1D projections (left-tailed Wilcoxon signed-rank test, pFDR[45] controlled at .01).



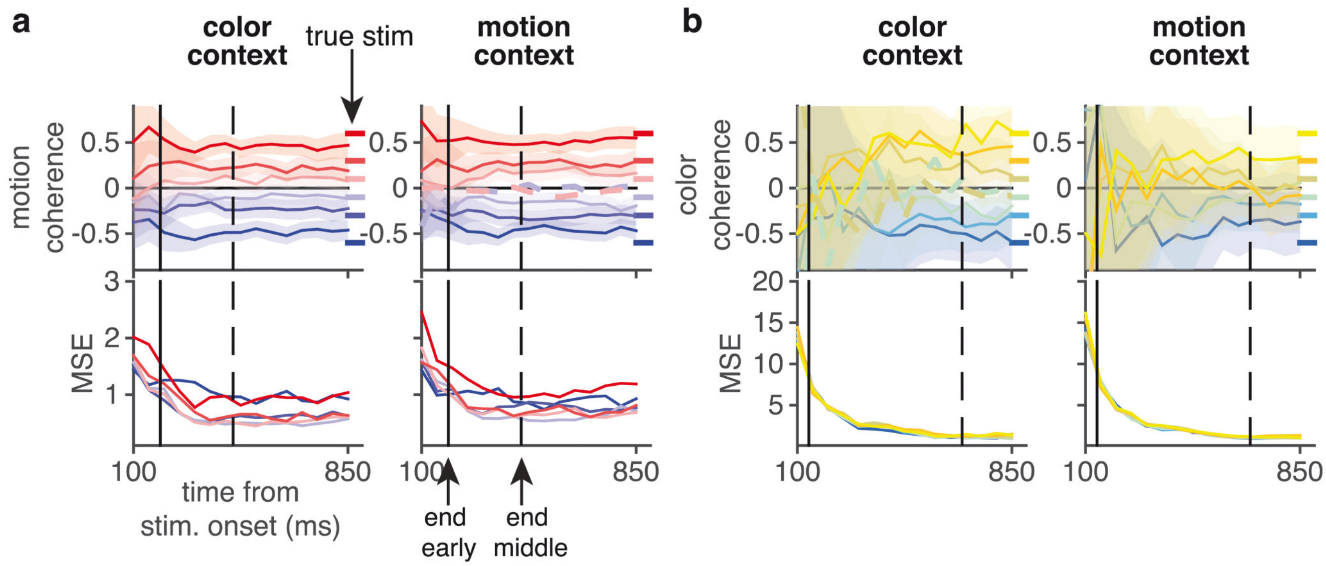
Extended Data Fig. 5. Distribution of variance among seqPCA axes. Monkey F

Plotting conventions are the same as for Figure 5. (**a**) Proportion of variance among seqPCA axes. Each marker corresponds to one neuron. The position of each neuron indicates the distribution of variance from PSTHs across corresponding early, middle, and late axes. e.g. a point that lies closer to the "early" vertex of the motion plot has more of its motion-specific variance explained by the early axis while a point in the middle of the simplex has variance equally distributed across all axes. Darker regions indicate higher density of points. Colored dots correspond to cells displayed in Figure Figure 3. (**b**) Weights of the top (in terms of variance explained) 3 axes for all cells for motion, color, and choice subspaces. Cell indexes are sorted according to the choice weights from most positive to most negative. (**c**) Magnitude of the Pearson correlation between top 3 subspace axes. The magnitude is used because the axes are only identifiable up to a sign. Markers indicate significant correlations controlled by the positive false discovery rate[45](* Q < .01, +Q < .01). Null distribution is based on the positive half-Gaussian with zero-mean and standard deviation σ0 = 1/n, where n = 640 is the number of neurons. Significant correlations are most consistent between color-choice and motion-choice pairs.



Extended Data Fig. 6. Rotational dynamics of subspace projections for Monkey F.

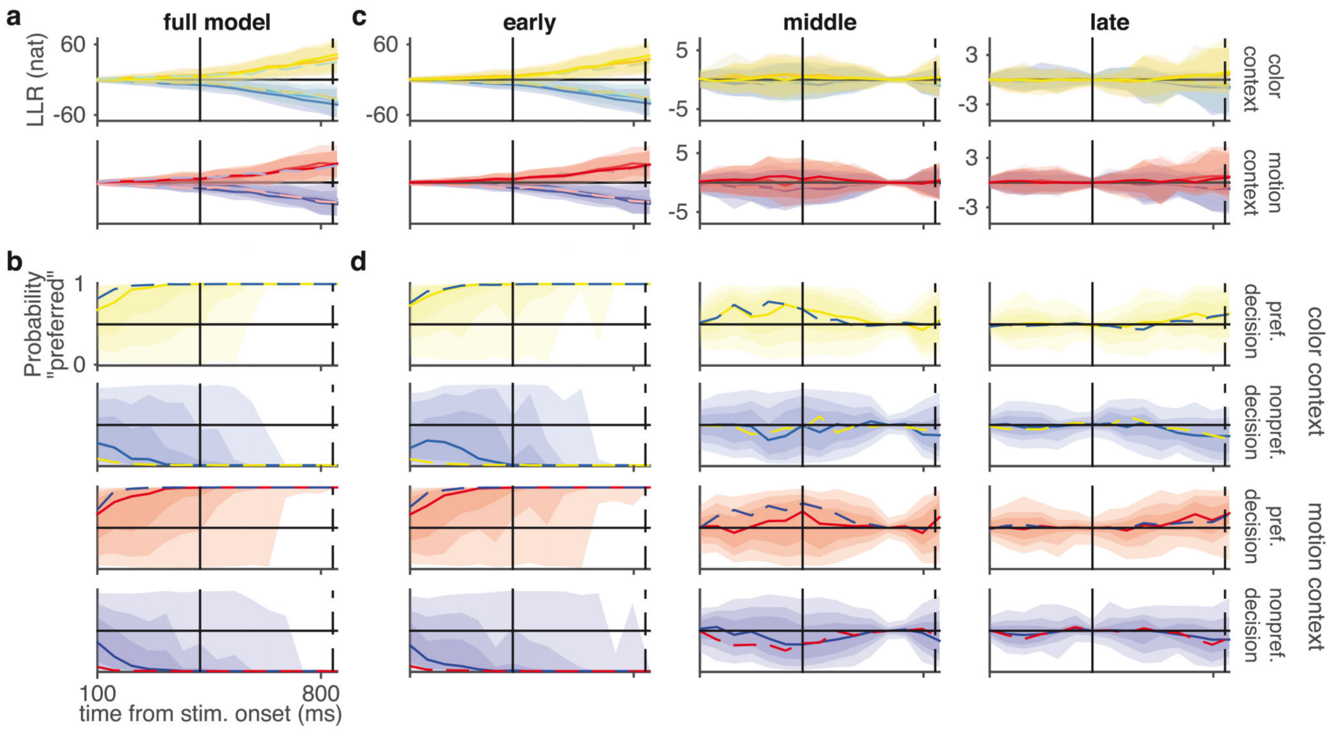
(**a**)Projections of population PSTH's for monkey F onto the jPC-axes of all task variables subspaces. Plotting conventions and analyses are the same as those for Figure 4. Projected data is averaged over 2-folds of cross validated projections where a random sampling of half of the data was used to estimate parameters and the remaining half used to make projections. (**b**)Angle of rotation over time for low-D trajectories of monkey F. Rotation angle traversed through rotational projection using jPCA. Angle was calculated starting from time when the projection transitions between the early and middle epochs. Coherent traversal across stimulus strengths that is consistent and monotonically increasing is an indication of rotation. Shaded areas are 95% confidence regions calculated using a maximum entropy method[23] (n = 100 samples) under the null hypothesis of no population structure other than the empirical means and covariances across time, neurons, and task conditions.



Extended Data Fig. 7. Instantaneous decoding of stimulus for monkey F.

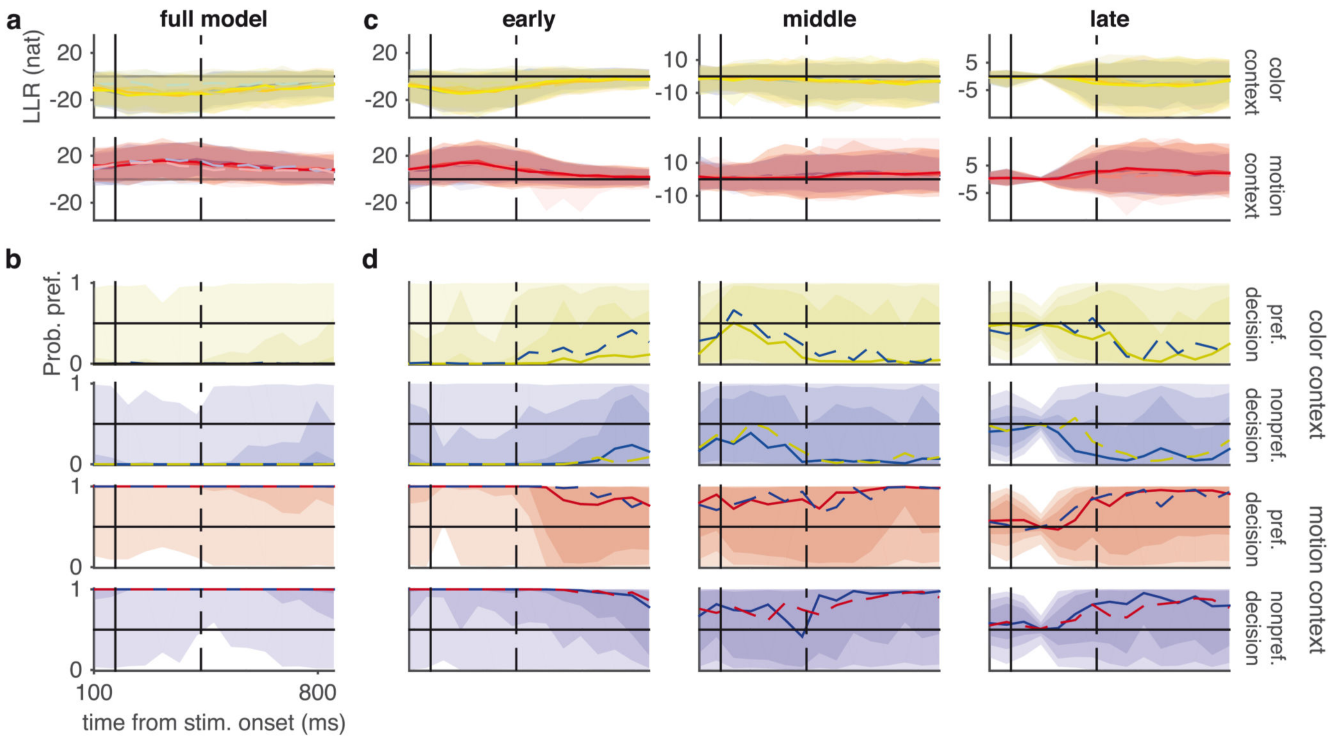
Plotting conventions and analyses are the same as for Figure 6 **a**) Top: Decoded motion coherence by mTDR model in both contexts. Bottom: Mean squared error (MSE) over time of motion coherence decoding across stimulus levels and context. MSE decreases precipitously, and then stabilize around the time of the first transition. **b**) Same as **a**) for color coherence decoding. Shaded regions indicate 50% confidence intervals. Dashed lines indicate error trials from the corresponding context for the lowest stimulus strengths. 100 pseudotrials for each of 2-fold cross validation used for each analysis. Solid vertical lines indicate the time of early/middle axis transition for the corresponding stimulus subspace projections. Dashed vertical lines indicate the time of middle/late transition.

Europe PMC Funders Author Manuscripts



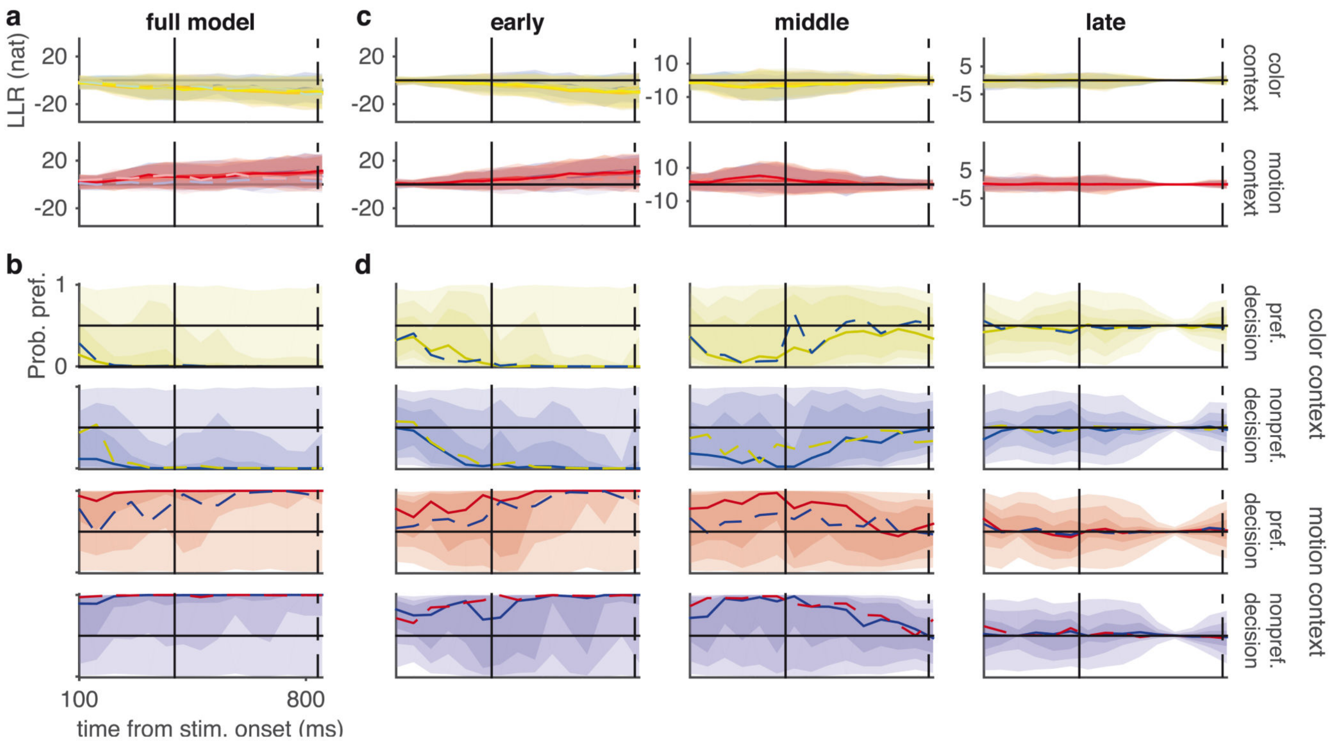
Extended Data Fig. 8. Instantaneous decoding of decision for monkey F.

Plotting conventions and analyses are the same as for Figure 6 **a**) Log-likelihood ratios (LLR's) in favor of a preferred choice using single pseudotrials from color - context (gold-blue, sorted by color coherence) and motion - context (red-violet, sorted by motion coherence) trials. Shaded regions indicate 95% quantile intervals for each stimulus strength. Solid lines indicate the median of correct trials. Dashed lines indicate median of error trials. **b**) Probability of a preferred choice based on corresponding LLRs combined over all stimulus strengths (see section 6.3 for details). Solid lines indicate median of correct trials. Dashed lines indicate median of error trials. Shaded regions indicate quantile coverage intervals of correct trials (light-to-dark: 95%,75%,50%). 100 pseudotrials for each of 2-fold cross validation folds used for all analyses. **c**) LLRs for in favor of a preferred choice where the choice subspace has been restricted to only the early, middle, or late axes. **d**) Probability of a preferred choice based on LLRs from (c).



Extended Data Fig. 9. Instantaneous decoding of context for monkey A.

a) LLRs for monkey A in favor of the motion context using single pseudotrials, sorted by color coherence. Shaded regions indicate 95% quantile intervals for each stimulus strength. Solid lines indicate the median over correct trials. Dashed lines indicate median of error trials. **b**) Probability of the motion context based on corresponding LLRs combined over all stimulus strengths. Solid lines indicate median of correct trials. Dashed lines indicate median of error trials. Shaded regions indicate quantile intervals of correct trials (light-to-dark: 50%, 75%, 95%). Color conventions are the same as in Figure 4. 100 pseudotrials for each of 4-fold cross validation folds used for all analyses.



Extended Data Fig. 10. Instantaneous decoding of context for monkey F.

Plotting conventions are the same as in Extended Data 9. 100 pseudotrials for each of 2-fold cross validation folds used for all analyses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by grants from the McKnight Foundation, Simons Collaboration on the Global Brain (SCGB AWD543027, M.C.A. & J.W.P.), the NIH BRAIN initiative (R01EB026946 and NS104899, J.W.P.), the NSF CAREER Award (IIS-1150186, J.W.P.), and a U19 NIH-NINDS BRAIN Initiative Award (5U19NS104648, M.C.A. & J.W.P.). V.M. was supported by the Swiss National Science Foundation (SNSF Professorship PP00P3-157539), the Simons Foundation (to William T Newsome and Valerio Mante, Award 328189), the Swiss Primate Competence Center in Research, Howard Hughes Medical Institute (through William T Newsome, Investigator), and the DOD | USAF | AFMC | Air Force Research Laboratory (AFRL): William T Newsome, agreement number FA9550-07-1-0537

Data Availability

Data are available for download at https://www.ini.uzh.ch/en/research/groups/mante/data.html

Code Availability

Demo code for the mTDR method is available for Matlab at http://www.mikioaoi.com/samplecode/RDRdemo.zip

Europe PMC Funders Author Manuscripts

References

[1]. Mante, Valerio; Sussillo, David; Shenoy, Krishna V; Newsome, William T. Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature. 2013; 503(7474):78–84. [PubMed: 24201281]

[2]. Raposo, David; Kaufman, Matthew T; Churchland, Anne K. A category-free neural population supports evolving demands during decision-making. Nature neuroscience. 2014

[3]. Roy, Jefferson E; Buschman, Timothy J; Miller, Earl K. Pfc neurons reflect categorical decisions about ambiguous stimuli. Journal of cognitive neuroscience. 2014; 26(6):1283–1291. [PubMed: 24405188]

[4]. Siegel, Markus; Buschman, Timothy J; Miller, Earl K. Cortical information flow during flexible sensorimotor decisions. Science. 2015; 348(6241):1352–1355. [PubMed: 26089513]

[5]. Katz, Leor N; Yates, Jacob L; Pillow, Jonathan W; Huk, Alexander C. Dissociated functional significance of decision-related activity in the primate dorsal stream. Nature. 2016; 535(7611):285–288. [PubMed: 27376476]

[6]. Hanks, Timothy D; Kopec, Charles D; Brunton, Bingni W; Duan, Chunyu A; Erlich, Jeffrey C; Brody, Carlos D. Distinct relationships of parietal and prefrontal cortices to evidence accumulation. Nature. 2015; 520(7546):220–223. [PubMed: 25600270]

[7]. Erlich, Jeffrey C; Brunton, Bingni W; Duan, Chunyu A; Hanks, Timothy D; Brody, Carlos D. Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. Elife. 2015; 4 e05457

[8]. Goard, Michael J; Pho, Gerald N; Woodson, Jonathan; Sur, Mriganka. Distinct roles of visual, parietal, and frontal motor cortices in memory-guided sensorimotor decisions. Elife. 2016; 5 e13764 [PubMed: 27490481]

[9]. Bruce, Charles J; Goldberg, Michael E. Primate frontal eye fields. i. single neurons discharging before saccades. Journal of Neurophysiology. 1985; 53(3):603–635. [PubMed: 3981231]

[10]. Kim JN, Shadlen MN. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat Neurosci. 1999 Feb; 2(2):176–185. [PubMed: 10195203]

[11]. Ding, Long; Gold, Joshua I. Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. Cerebral Cortex. 2011 bhr178

[12]. Stokes, Mark G; Kusunoki, Makoto; Sigala, Natasha; Nili, Hamed; Gaffan, David; Duncan, John. Dynamic coding for cognitive control in prefrontal cortex. Neuron. 2013; 78(2):364–375. [PubMed: 23562541]

[13]. Rigotti, Mattia; Barak, Omri; Warden, Melissa R; Wang, Xiao-Jing; Daw, Nathaniel D; Miller, Earl K; Fusi, Stefano. The importance of mixed selectivity in complex cognitive tasks. Nature. 2013 May; 497(7451):585–590. [PubMed: 23685452]

[14]. Purcell, Braden A; Heitz, Richard P; Cohen, Jeremiah Y; Schall, Jeffrey D; Logan, Gordon D; Palmeri, Thomas J. Neurally constrained modeling of perceptual decision making. Psychological review. 2010; 117(4):1113. [PubMed: 20822291]

[15]. Heitz, Richard P; Schall, Jeffrey D. Neural mechanisms of speed-accuracy tradeoff. Neuron. 2012; 76(3):616–628. [PubMed: 23141072]

[16]. Parthasarathy, Aishwarya; Herikstad, Roger; Bong, Jit Hon; Medina, Felipe Salvador; Libedinsky, Camilo; Yen, Shih-Cheng. Mixed selectivity morphs population codes in prefrontal cortex. Nature neuroscience. 2017; 20(12):1770–1779. [PubMed: 29184197]

[17]. Beck JM, Ma WJ, Kiani R, Hanks T, Churchland AK, Roitman J, Shadlen MN, Latham PE, Pouget A. Probabilistic population codes for bayesian decision making. Neuron. 2008; 60(6):1142–1152. [PubMed: 19109917]

[18]. Churchland, Mark M; Cunningham, John P; Kaufman, Matthew T; Foster, Justin D; Nuyujukian, Paul; Ryu, Stephen I; Shenoy, Krishna V. Neural population dynamics during reaching. Nature. 2012; 487(7405):51–56. [PubMed: 22722855]

[19]. Kaufman, Matthew T; Churchland, Mark M; Ryu, Stephen I; Shenoy, Krishna V. Cortical activity in the null space: permitting preparation without movement. Nature neuroscience. 2014; 17(3):440–448. [PubMed: 24487233]

Europe PMC Funders Author Manuscripts

[20]. Lara AH, Cunningham JP, Churchland MM. Different population dynamics in the supplementary motor area and motor cortex during reaching. Nature communications. 2018; 9(1):2754.

[21]. Akaike, Hirotugu. A new look at the statistical model identification. IEEE transactions on automatic control. 1974; 19(6):716–723.

[22]. Aoi, Mikio; Pillow, Jonathan W. Model-based targeted dimensionality reduction for neuronal population data. Advances in neural information processing systems. 2018:6690–6699. [PubMed: 31274967]

[23]. Elsayed, Gamaleldin F; Cunningham, John P. Structure in neural population recordings: an expected byproduct of simpler phenomena? Nature neuroscience. 2017; 20(9):1310. [PubMed: 28783140]

[24]. Rossi-Pool, Román; Zainos, Antonio; Alvarez, Manuel; Zizumbo, Jerónimo; Vergara, José; Romo, Ranulfo. Decoding a decision process in the neuronal population of dorsal premotor cortex. Neuron. 2017; 96(6):1432–1446. [PubMed: 29224726]

[25]. Williamson, Ryan C; Cowley, Benjamin R; Litwin-Kumar, Ashok; Doiron, Brent; Kohn, Adam; Smith, Matthew A; Byron, M Yu. Scaling properties of dimensionality reduction for neural populations and network models. PLoS computational biology. 2016; 12(12) e1005141 [PubMed: 27926936]

[26]. Cunningham, John P; Byron, M Yu. Dimensionality reduction for large-scale neural recordings. Nature neuroscience. 2014; 17(11):1500–1509. [PubMed: 25151264]

[27]. Kobak, Dmitry; Brendel, Wieland; Constantinidis, Christos; Feierstein, Claudia E; Kepecs, Adam; Mainen, Zachary F; Qi, Xue-Lian; Romo, Ranulfo; Uchida, Naoshige; Machens, Christian K. Demixed principal component analysis of neural population data. eLife. 2016; 5 e10989 [PubMed: 27067378]

[28]. Machens, Christian K. Demixing population activity in higher cortical areas. Frontiers in Computational Neuroscience. 2010; 4(0)

[29]. Machens CK, Romo R, Brody CD. Functional, but not anatomical, separation of “what” and “when” in prefrontal cortex. The Journal of Neuroscience. 2010; 30(1):350–360. [PubMed: 20053916]

[30]. Romo, Ranulfo; Brody, Carlos D; Hernández, Adrián; Lemus, Luis. Neuronal correlates of parametric working memory in the prefrontal cortex. Nature. 1999 Jun; 399(6735):470–473. [PubMed: 10365959]

[31]. Hernández, Adrián; Zainos, Antonio; Romo, Ranulfo. Temporal evolution of a decision-making process in medial premotor cortex. Neuron. 2002; 33(6):959–972. [PubMed: 11906701]

[32]. Romo, Ranulfo; Hernández, Adrián; Zainos, Antonio; Lemus, Luis; Brody, Carlos D. Neuronal correlates of decision-making in secondary somatosensory cortex. Nature neuroscience. 2002; 5(11):1217–1225. [PubMed: 12368806]

[33]. Romo, Ranulfo; Hernández, Adrián; Zainos, Antonio. Neuronal correlates of a perceptual decision in ventral premotor cortex. Neuron. 2004; 41(1):165–173. [PubMed: 14715143]

[34]. Elsayed, Gamaleldin F; Lara, Antonio H; Kaufman, Matthew T; Churchland, Mark M; Cunningham, John P. Reorganization between preparatory and movement population responses in motor cortex. Nature Communications. 2016; 7 13239

[35]. Gold JI, Shadlen MN. The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. J Neurosci. 2003 Jan; 23(2):632–651. [PubMed: 12533623]

[36]. Kiani, Roozbeh; Hanks, Timothy D; Shadlen, Michael N. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. J Neurosci. 2008 Mar; 28(12):3017–3029. [PubMed: 18354005]

[37]. Meister, Miriam LR; Hennig, Jay A; Huk, Alexander C. Signal multiplexing and single-neuron computations in lateral intraparietal area during decision-making. The Journal of neuroscience. 2013; 33(6):2254–2267. [PubMed: 23392657]

[38]. Latimer, Kenneth W; Yates, Jacob L; Meister, Miriam LR; Huk, Alexander C; Pillow, Jonathan W. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. Science. 2015; 349(6244):184–187. [PubMed: 26160947]

[39]. Yates, Jacob L; Park, Il Memming; Katz, Leor N; Pillow, Jonathan W; Huk, Alexander C. Functional dissection of signal and noise in mt and lip during decision-making. Nature neuroscience. 2017; 20(9):1285. [PubMed: 28758998]

[40]. Vernet, Marine; Quentin, Romain; Chanes, Lorena; Mitsumasu, Andres; Valero-Cabré, Antoni. Frontal eye field, where art thou? anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations. Frontiers in Integrative Neuroscience. 2014; 8(66):1–24. [PubMed: 24474908]

[41]. Bruce, Charles J; Goldberg, Michael E; Bushnell, M Catherine; Stanton, GregoryB. Primate frontal eye fields. ii. physiological and anatomical correlates of electrically evoked eye movements. Journal of neurophysiology. 1985; 54(3):714–734. [PubMed: 4045546]

[42]. Rizzolatti, Giacomo; Riggio, Lucia; Dascola, Isabella; Umiltá, Carlo. Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. Neuropsychologia. 1987; 25(1):31–40. [PubMed: 3574648]

[43]. Thompson, Kirk G; Biscoe, Keri L; Sato, Takashi R. Neuronal basis of covert spatial attention in the frontal eye field. Journal of Neuroscience. 2005; 25(41):9479–9487. [PubMed: 16221858]

[44]. Schall, Jeffrey D. On the role of frontal eye field in guiding attention and saccades. Vision research. 2004; 44(12):1453–1467. [PubMed: 15066404]

[45]. Storey, John D. A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002; 64(3):479–498.

[46]. Brody, Carlos D; Hernández, Adrián; Zainos, Antonio; Romo, Ranulfo. Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. Cerebral cortex. 2003; 13(11):1196–1207. [PubMed: 14576211]

[47]. Roitman JD, Shadlen MN. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. Journal of Neuroscience. 2002; 22(21):9475. [PubMed: 12417672]

[48]. Lawrence N. Probabilistic non-linear principal component analysis with gaussian process latent variable models. The Journal of Machine Learning Research. 2005; 6:1816.

[49]. Cunningham, John P; Ghahramani, Zoubin. Linear dimensionality reduction: Survey, insights, and generalizations. Journal of Machine Learning Research. 2015; 16(89):2859–2900.

[50]. Mackevicius, Emily L; Bahle, Andrew H; Williams, Alex H; Gu, Shijie; Denisenko, Natalia I; Goldman, Mark S; Fee, Michale S. Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience. Elife. 2019; 8 e38471 [PubMed: 30719973]

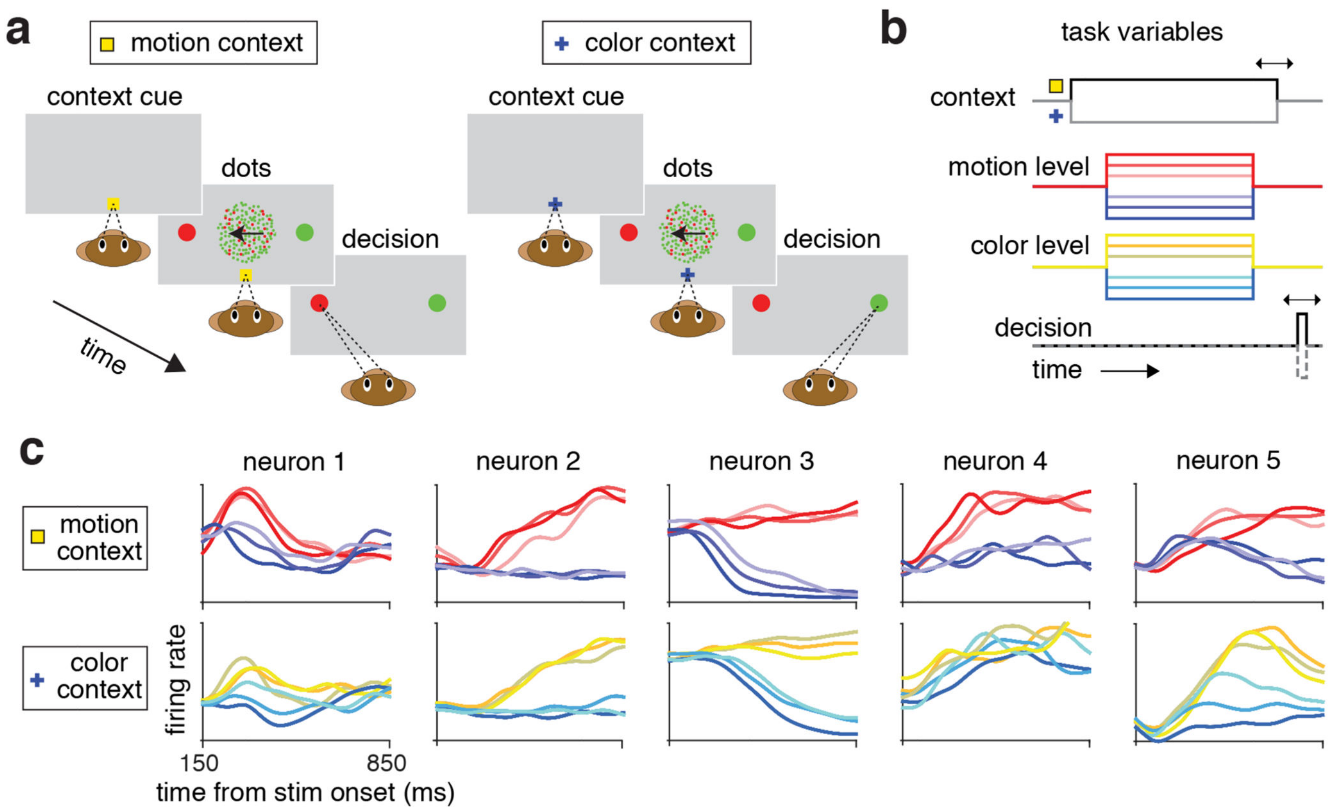


Figure 1. Context-dependent decision-making task and neural responses.

a) On each trial, the animal was presented with a context cue (yellow dot or blue cross) indicating which dimension of the stimulus the animal is to attend to, followed by a stimulus of colored, moving dots. On motion context trials the animal is cued to respond to the dominant dot motion direction. In color context trials the animal is cued to respond to the dominant color of the dots. **b**) The strength of both the color (red / green) and motion (left / right) stimulus was displayed with one of six possible degrees of coherence, making for many possible task contingencies (2 choices × 2 contexts × 6 motion strengths × 6 color strengths = 144 possible combinations). **c**) PSTHs of representative neurons for monkey A. Motion context PSTHs were sorted by motion coherence and averaged over color coherence. Color context PSTH's were sorted by color coherence and averaged over motion coherence. Red–indigo color scale indicates motion coherence where red indicates the preferred motion direction. Gold–blue color scale indicates color coherence where gold indicates the preferred color direction. Bolder colors indicate stronger coherence.

Europe PMC Funders Author Manuscripts

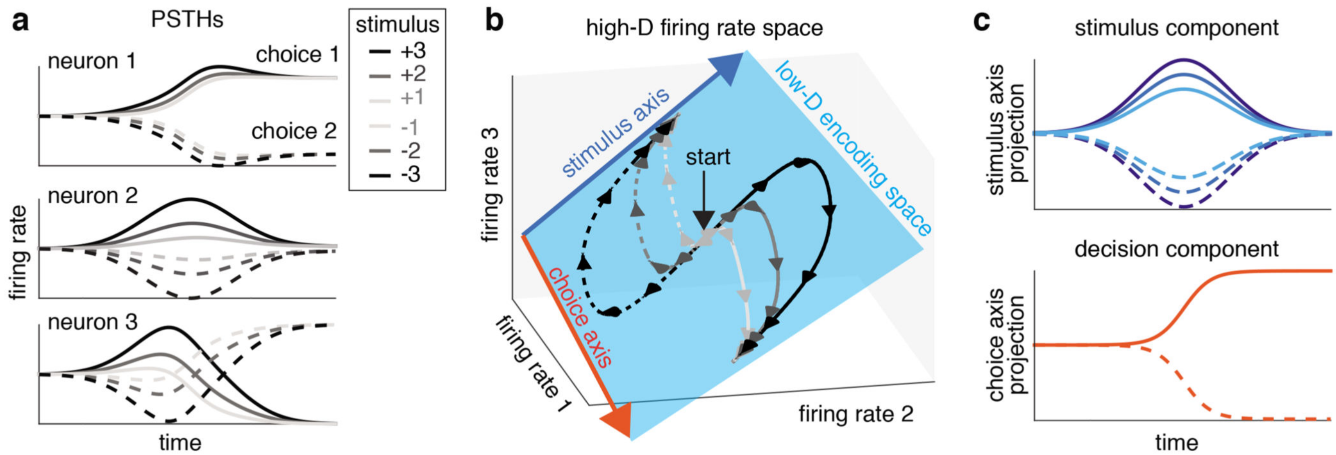


Figure 2. Schematic illustrating low-dimensional population-level encoding in a binary sensory decision-making task.

(a). Conditional PSTHs for three neurons that exhibit mixed selectivity to a stimulus variable (taking on six different values) and a choice variable (taking on two values). **(b)** Modulations of the PSTHs by the task variables span a 2-dimensional "encoding subspace", which is low-dimensional relative to the 3-dimensional space of firing rates. In this case, a 1D stimulus-encoding subspace (blue arrow) captures all information about the stimulus value, while a 1D choice-encoding subspace (red arrow) captures all information about the decision. Note, for example, that the neuron 2 firing rate axis is nearly orthogonal to the choice axis, meaning that neuron 2 carries almost no information about choice. **(c)**. Projections onto the stimulus and choice subspaces reveal the time-course of information about stimulus and choice, respectively. These timecourses can be seen as temporal basis functions for the single-neuron PSTHs shown in (**a**). mTDR aims to recover these encoding subspaces even in the presence of additional components that take neural activity outside the plane spanned by these two axes, and is not restricted to 1D subspaces.

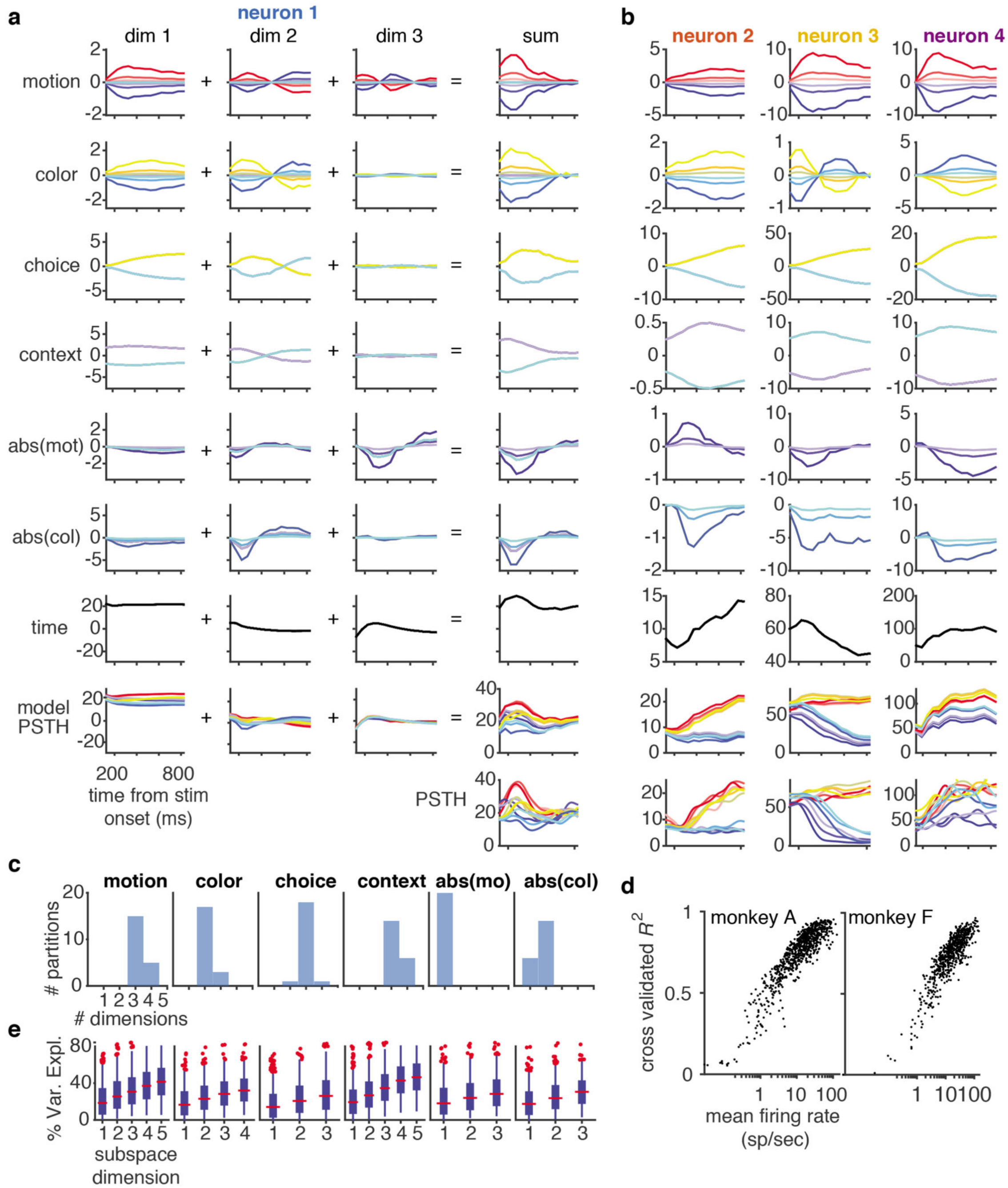


Figure 3. Model fit for monkey A.

a) Example of a neuron's fitted responses composed of a set of weighted basis functions (same as neuron 1 from Figure 1c). These basis functions are shared by the whole population but are weighted differently for each neuron. Weighted basis functions are summed to form the neuron's response to each task variable. The responses for each task variable are then added together to give the model reconstructed PSTHs (model PSTH). The conditional PSTHs of this neuron are shown for comparison. **b**) Summed responses for three additional example neurons (same as neurons 2–4, from Figure 1c) which display a diversity of dynamics. **c**) Dimension estimation based on 5x 4-fold (20 estimates) cross validation. Dimensionality is slightly smaller than estimated using all data but is tightly distributed around a single estimate. **d**) R^2 of the model reconstructions for the PSTHs as a function of mean firing rate for each neuron. **e**) Percent variance explained for PSTHs of each neuron (n = 762) by projection onto each subspace dimension. Red horizontal bars indicate the median. Box edges indicate 25th and 75th percentiles. Whiskers indicate positions of furthest points from median not considered outliers. Red dots indicate outliers with respect to a normal distribution. Dots have been horizontally jittered to aid with visualization. Results have been averaged for each neuron over 4 CV folds. Colors in title text for (**a**) and (**b**) correspond to colors of markers in Figure 5.

Europe PMC Funders Author Manuscripts

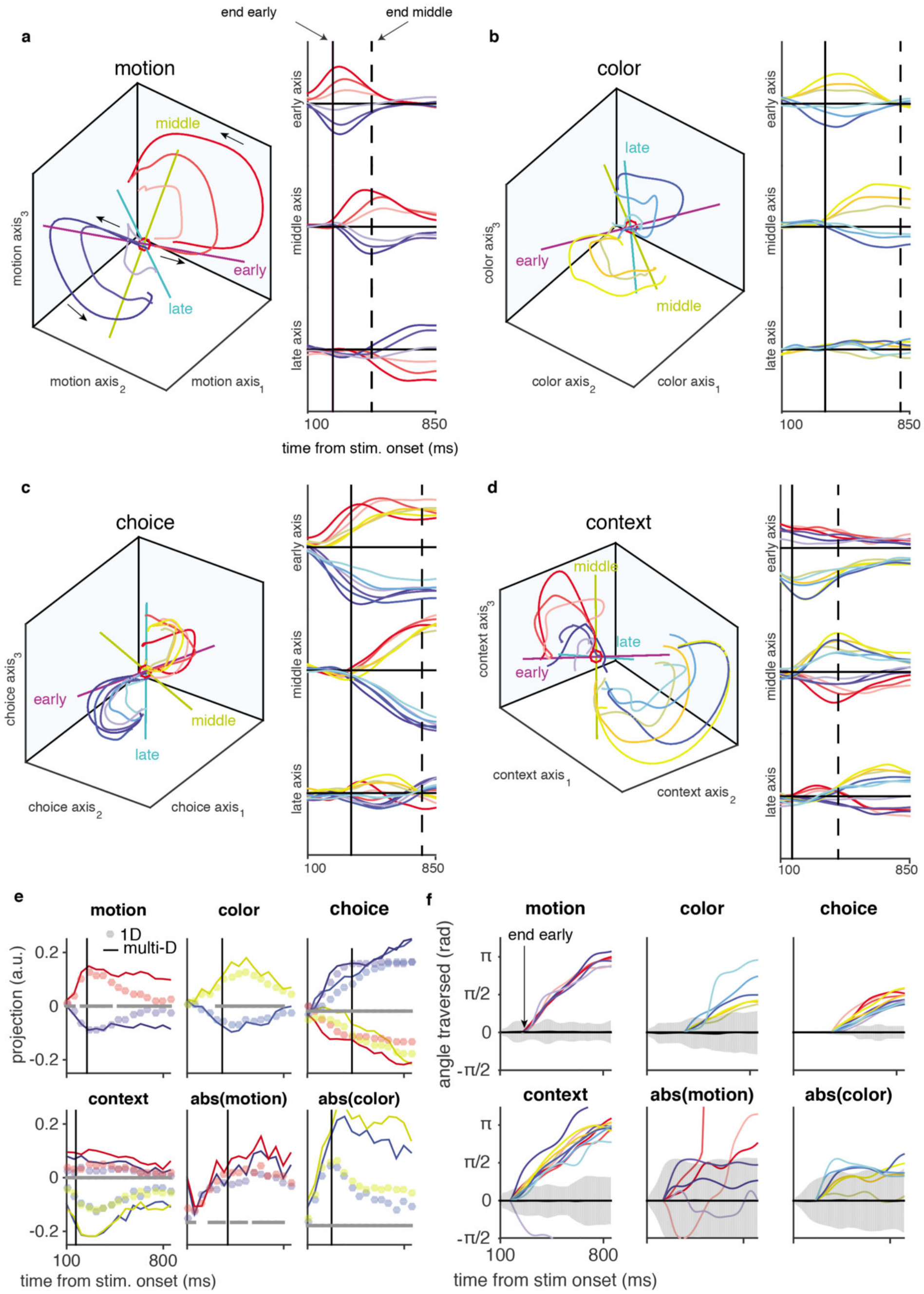


Figure 4. Projections of population PSTH's onto latent encoding subspaces.

Projections onto the first, second, and third principle-axes of the (**a**) motion, (**b**) color, (**c**) choice, and (**d**) context subspaces. Motion, color, and context subspaces have been orthogonalized with respect to the first dimension of the choice subspace. The choice subspace has been orthogonalized with respect to the context subspace. The context subspace has also been orthogonalized with respect to the motion and color subspaces. Details of orthogonalization are presented in Supplementary note 4.2. Color conventions are the same as those described in Figure Figure 1. Red dots indicate the origin. Projected

Europe PMC Funders Author Manuscripts

PSTH's made from held-out data not used during parameter estimation. **a**) Projections of PSTHs onto the motion subspace, sorted by motion coherence and averaged over color coherence for trials where the motion stimulus was the active context. **b**) Projections onto the color subspace sorted by color coherence and averaged over motion coherence for trials where the color stimulus was the active context. **c**) Projections onto the choice subspace. Motion context trials are displayed with the same sorting and color conventions as displayed in (**a**). Color context trials are displayed with the same sorting and color conventions as displayed in (**b**). Only correct trials are displayed. **d**) Projections onto the context subspace using the same conventions as displayed in (**c**). Only correct trials are displayed. Colored axes in 3D plots indicate seqPCA axes. Solid vertical lines accompanying time traces indicate the time points where middle-axis variance starts to increase. Dashed vertical lines indicate the time points where late-axis variance starts to increase. Units of the ordinate are arbitrary but all time-trace axes are on the same scale. PSTHs were generated with ≈ 13 ms time bins and smoothed with a Gaussian window with standard deviation of ≈ 50 ms. **e**) Median encoding strength of pseudotrials onto the first three encoding axes of mTDR compared with the 1D subspace estimated by the max-norm method used by Mante et al.[1] (see Supplementary note 10 for details). For clarity, only trials with the strongest stimulus strengths are shown. Grey bars at $y = 0$ indicate time points where the mTDR projections had significantly stronger encoding across all stimulus levels than the 1D projections (left-tailed Wilcoxon signed-rank test, pFDR[45] controlled at .01). Multidimensional mTDR projections are larger than 1D projections at nearly all times for all task variables. **f**) Rotation angle traversed through rotational projection using jPCA. Angle was calculated starting from time when the projection transitions between the early and middle epochs. Coherent traversal across stimulus strengths that is consistent and monotonically increasing is an indication of rotation. Shaded areas are 95% confidence regions calculated using a maximum entropy method[23] ($n = 100$ samples) under the null hypothesis of no population structure other than the empirical means and covariances across time, neurons, and task conditions.

Europe PMC Funders Author Manuscripts

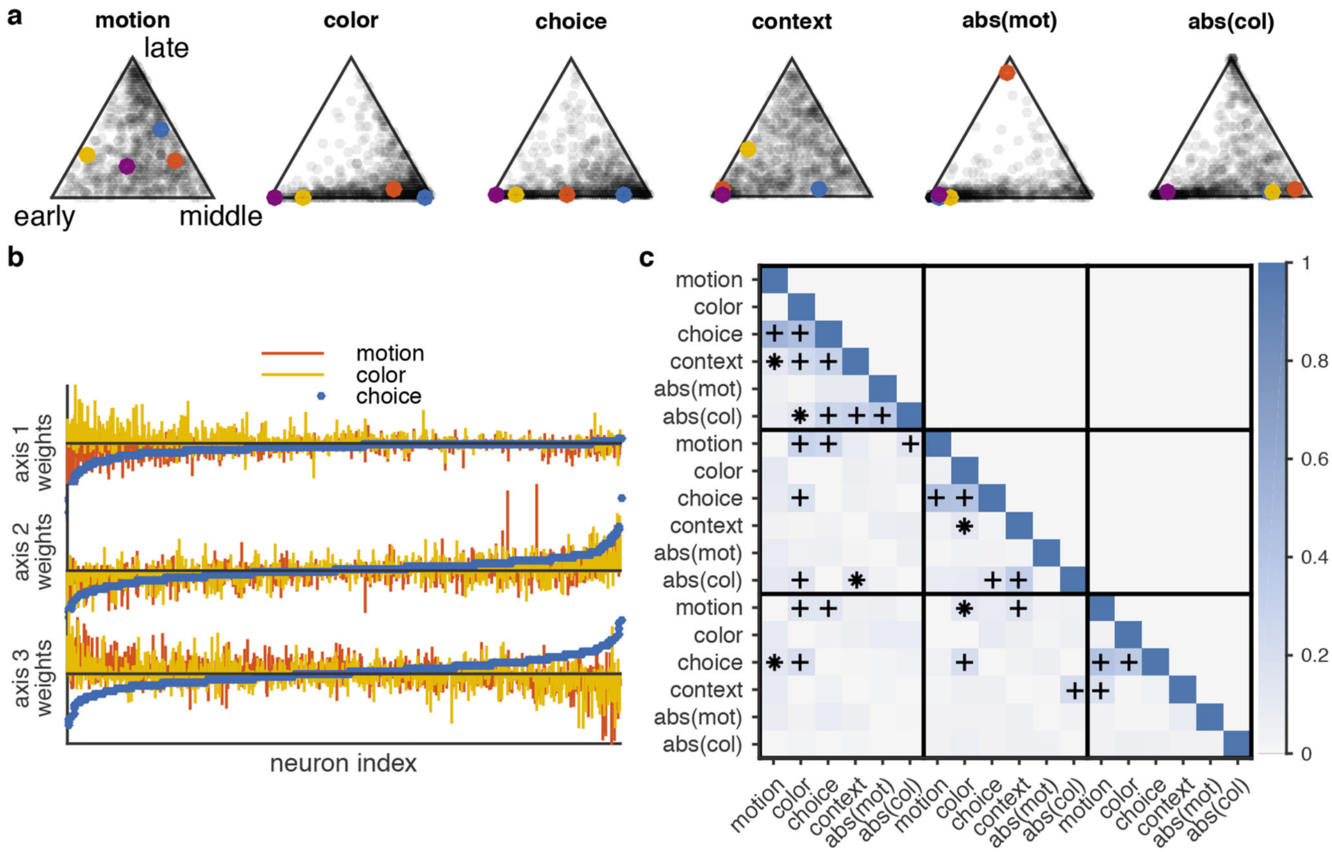


Figure 5. Distribution of variance within and between subspaces.

(**a**) Proportion of variance among seqPCA axes. Each marker corresponds to one neuron. The position of each neuron indicates the distribution of variance from PSTHs across corresponding early, middle, and late axes. e.g. a point that lies closer to the "early" vertex of the motion plot has more of its motion-specific variance explained by the early axis while a point in the middle of the simplex has variance equally distributed across all axes. Darker regions indicate higher density of points. Colored dots correspond to cells displayed in Figure Figure 3. (**b**) Weights of the top (in terms of variance explained) 3 axes for all cells for motion, color, and choice subspaces. Cell indexes are sorted according to the choice weights from most positive to most negative. (**c**) Magnitude of the Pearson correlation between top 3 subspace axes. The magnitude is used because the axes are only identifiable up to a sign. Markers indicate significant correlations controlled by the positive false discovery rate[45])(* $Q < .01$, $+Q < .01$). Null distribution is based on the positive half-Gaussian with zero-mean and standard deviation $\sigma_0 = 1/n$, where $n = 762$ is the number of neurons. Significant correlations are most consistent between color-choice and motion-choice pairs. All tests were 1-sided.

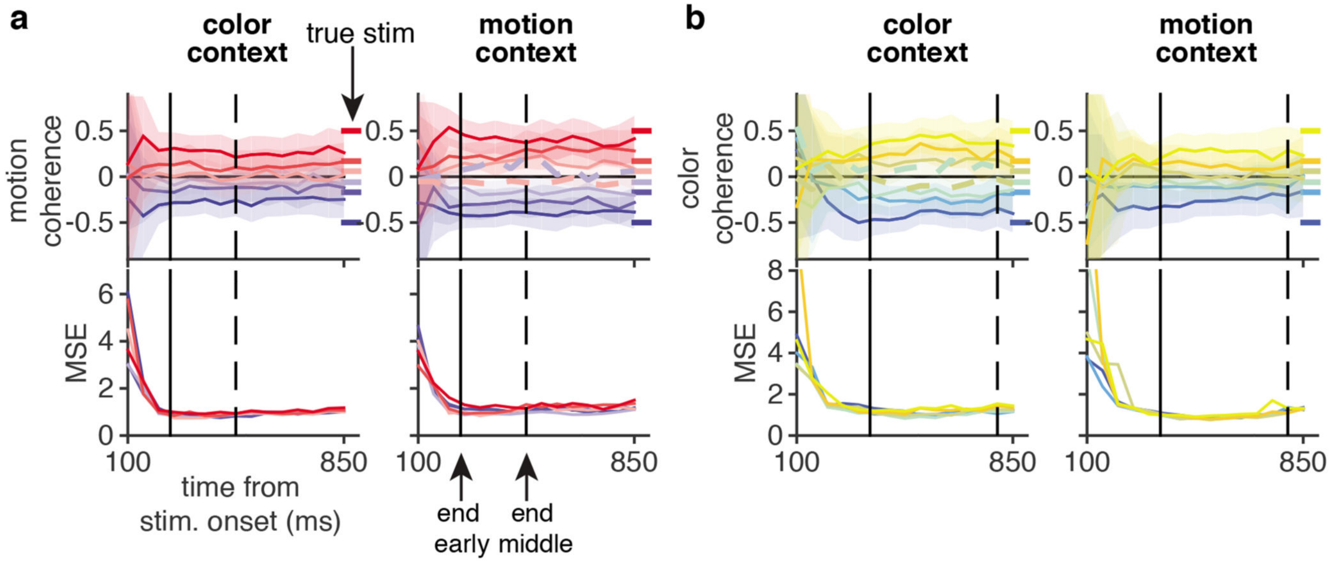


Figure 6. Instantaneous decoding of stimulus for monkey A.

a) Top: Decoded motion coherence by mTDR model in both contexts. Bottom: Mean squared error (MSE) over time of motion coherence decoding across stimulus levels and context. MSE decreases precipitously, and then stabilize around the time of the first transition. **b**) Same as **a**) for color coherence decoding. Color conventions are the same as in Figure 4. Shaded regions indicate 50% confidence intervals. Dashed lines indicate error trials from the corresponding context for the lowest stimulus strengths. 100 pseudotrials for each of 4-fold cross validation (n = 400) used for all analyses. Solid vertical lines indicate the time of early/middle axis transition for the corresponding stimulus subspace projections. Dashed vertical lines indicate the time of middle/late transition.

Europe PMC Funders Author Manuscripts

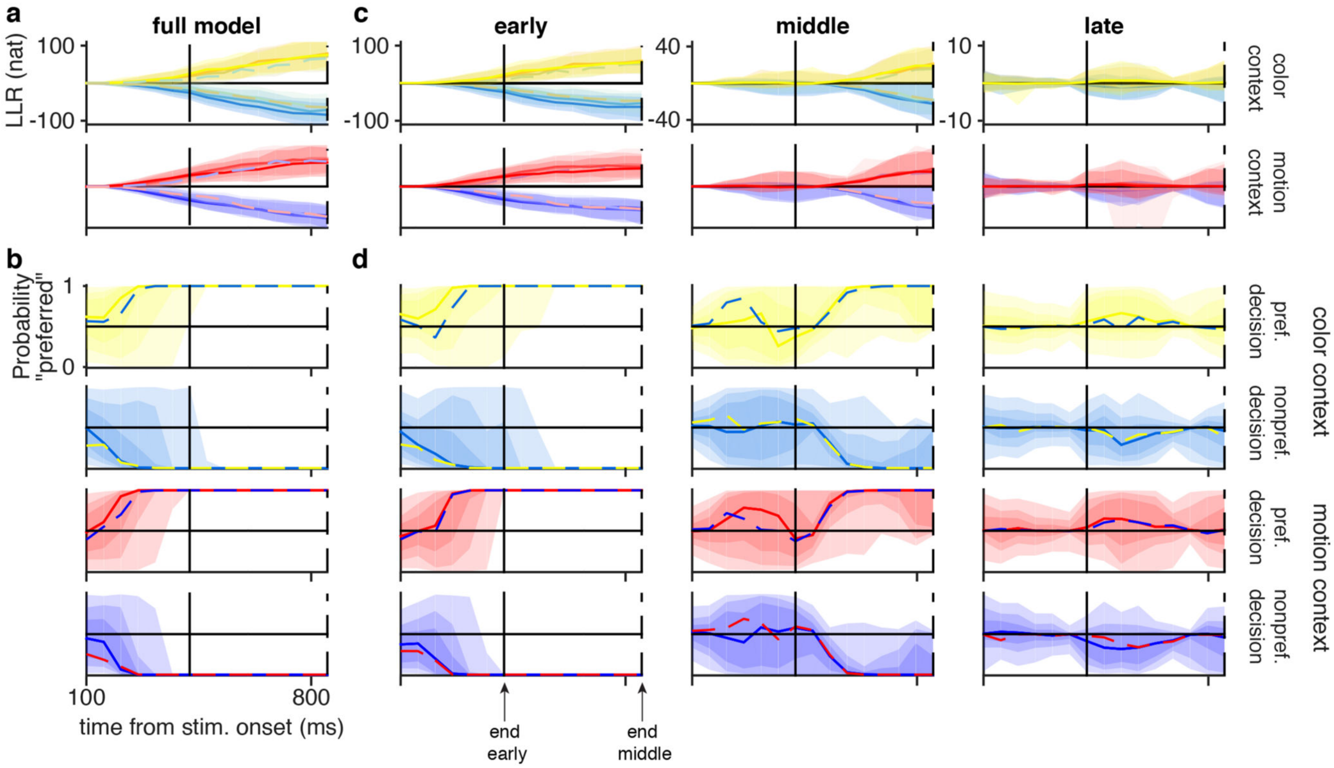


Figure 7. Instantaneous decoding of choice.

a) Log-likelihood ratios (LLR's) for monkey A in favor of a preferred choice using single pseudotrials from color - context (gold-blue, sorted by color coherence) and motion - context (red-violet, sorted by motion coherence) trials. Shaded regions indicate 95% quantile intervals for each stimulus strength. Solid lines indicate the median of correct trials. Dashed lines indicate median of error trials. **b**) Probability of a preferred choice based on corresponding LLRs combined over all stimulus strengths (see section 6.3 for details). Solid lines indicate median of correct trials. Dashed lines indicate median of error trials. Shaded regions indicate quantile coverage intervals of correct trials (light-to-dark: 95%,75%,50%). Color conventions are the same as in Figure Figure 4. 100 pseudotrials for each of 4-fold cross validation folds used for all analyses. **c**) LLRs for in favor of a preferred choice where the choice subspace has been restricted to only the early, middle, or late axes. **d**) Probability of a preferred choice based on LLRs from (c).