

Linking correlated variability to flexible behavior

Marlene R. Cohen and Brent Doiron

Project Summary

Improvements in perceptual performance (such as those associated with stimulus contrast, attention, learning, task switching, or arousal/motivation) are typically accompanied by decreases in the correlated trial-to-trial variability of responses from pairs of sensory neurons to repeated presentations of the same stimulus. This widespread observation is at odds with recent work suggests that, contrary to the hypothesis that dominates the field, correlated variability seems ill positioned to impact the information encoded in a neural population and rather is inconsequential to neuronal codes. Stimulus properties can in principle be decoded from neural populations uncorrupted by correlated variability because correlated variability largely fluctuates along one or a small number of axes in neural population space (akin to a single, linear combination of neurons (which we refer to as the correlated variability axis), and hence can be easily ignored by an appropriate (optimal) decoder.

We propose to test a new hypothesis, that correlated variability is a signature of the signals that make their way out of, rather than are encoded in, populations of visual neurons. Biologically, this could correspond to a fixed set of output neurons that flexibly communicate visual information in a task- and context-dependent way. This hypothesis makes essentially opposite predictions from the alternative/classical hypothesis that cognitive processes modulate correlated variability because it is deleterious to neuronal codes. The biggest challenge to our central hypothesis is that it requires a control signal to flexibly modulate visual responses such that the correct, task-relevant information is communicated to output neurons. In principle, this could require control signals that are so intricate and non-robust as to be implausible.

We propose to test this hypothesis using an intertwined experimental and theoretical/modeling strategy. Experimentally, we will record from groups of neurons in visual area V4 during tasks that require subjects to use visual information in flexible ways. Theoretically, we will use biologically constrained circuit models to explore the relationship between correlated variability, flexible behavior, and the control signals that guide activity. We use a unified analysis approach for data and models to measure the relationships between correlated variability and visual, task, and premotor signals. We will test three predictions of our central hypothesis, that 1) observers will be better able to discriminate stimuli whose representations happen to be aligned with the correlated variability axis (Aim 1), 2) stimulus representations (signals) will be better aligned with the correlated variability axis when they are task relevant than irrelevant (Aim 2), and 3) correlated variability will be better aligned with representations of action plans than the perceptual judgment (Aim 3).

The proposed studies will test a novel hypothesis that can resolve a longstanding paradox about why neuronal variability is so tightly linked to behavior. Doing so will change the way neural populations are measured, modeled, and interpreted, and could have broad implications for efforts to repair and enhance cognitive flexibility.

Project Narrative

This project is relevant to public health because essentially all brain disorders that affect flexible behavior including depression, schizophrenia, autism, and attention disorders involve large networks of neurons. Our goal is to test a novel hypothesis about how interactions between neurons influence the activity of downstream populations and ultimately guide behavior. Understanding the complex relationship between large groups of neurons, perceptual and cognitive processes, and flexible behavior will be critical for efforts to enhance and repair cognitive flexibility.

Specific Aims

Many studies, including ours, have demonstrated that improvements in perceptual performance (such as those associated with stimulus contrast, attention, learning, task switching, or arousal/motivation) are typically accompanied by decreases in mean correlation between the responses pairs of sensory neurons to repeated presentations of the same stimulus^{1–23}. We recently showed that correlated variability explains virtually all of the choice-predictive signals in mid-level visual area V4^{13,24}. Links between correlated variability and behavior have generated widespread interest for translational and basic science reasons. Correlated variability is easy to modulate using natural cognitive processes, neurotransmitters, or pharmaceuticals^{1,8,14,25}, making it an attractive therapeutic target. For basic science, correlated variability offers a powerful constraint on mechanistic models of a variety of perceptual and cognitive processes, including our own^{26–31}.

However, these observations comprise a paradox because recent work suggests that population-wide correlated variability should be for the most part irrelevant to neuronal coding. Indeed, stimulus properties can be decoded from neural populations with significant population-wide variability because any correlated variability fluctuates along one or a small number of axes in neural population space (akin to a linear combination of neurons, which we refer to as the correlated variability axis^{27,32–37}, and can be ignored by an optimal decoder^{33,38–42}. Thus, manipulating variability seems unimpactful from a neuronal coding perspective.

We propose to test a new hypothesis informed by our recent observations, that through its relationship to tuning properties and modulation by cognition, the axis of correlated variability reflects the signals that are communicated out of, rather than are encoded in, populations of visual neurons. Biologically, this could correspond to a fixed set of projection neurons that communicate visual information in a task-dependent way. This hypothesis makes opposite predictions from the alternative framework where cognitive processes modulate correlated variability because it harms information coding. The biggest challenge to our hypothesis is that it requires a control signal to modulate responses such that task-relevant information is communicated by the population. In principle, this could require control signals that are too intricate to be plausible.

Our goals are to test this hypothesis using an intertwined experimental and theoretical strategy. Experimentally, we will record from groups of neurons in visual area V4, where many of where most of our previous observations were made^{3,4,13–17,18,24,30}, while monkeys use visual information in flexible ways that allow us to test different predictions of our hypothesis. Theoretically, we will use biologically constrained circuit models to explore the relationship between correlated variability, flexible behavior, and the top-down control signals that guide it. We use a unified analysis approach for data and models to measure the relationships between correlated variability and visual, task, and premotor signals. We will test the following predictions:

Prediction/Aim 1: Observers will be better able to discriminate stimuli whose representations are aligned with the correlated variability axis. We will measure V4 responses while monkeys make veridical (continuous) estimates of the medial axis curvature of objects that vary in task-irrelevant features like shape, color, and orientation. The alignment between the representation of the curvature and the correlated variability axes is object dependent. Experiment: Does the degree of alignment between curvature representation and the correlated variability axis predict the accuracy of perceptual judgments? Theory: **In a model where connectivity depends on tuning similarity is the correlated variability axis aligned with the neural coding axis?**

Prediction/Aim 2: Representations of stimulus features will be better aligned with the correlated variability axis when they are used to guide behavior than when they are not. Monkeys and models will alternate between discriminating the shape or color of two stimuli, a type of task switching known to modulate V4 responses^{43–46}. Experiment: Do these modulations serve to align the representation of the task-relevant feature with the correlated variability axis? Theory: **Can biologically plausible control signals (e.g. to groups of neurons) change how responses are aligned with the correlated variability axis to enable task switching?**

Prediction/Aim 3: Correlated variability will be better aligned with representations of action plans than the perceptual judgment. We will dissociate the curvature judgment from the action plan by requiring the animals and models to communicate their decision by making eye movements to a target with trial variable location. Experiment: Does the representation of the action plan (which needs to be communicated to motor neurons) or the perceptual inference better align with the correlated variability axis? Theory: **Can biologically plausible control signals align the action plans with the correlated variability axis?**

The proposed studies will test a novel neural coding hypothesis that can resolve a longstanding paradox about why noise is so tightly linked to behavior. Doing so will change the way neural populations are measured, modeled, and interpreted, and could have broad implications for efforts to repair and enhance cognitive flexibility.

SIGNIFICANCE

Importance of the Problem

In recent years, a staggering number of studies have related different aspects of flexible, sensory-guided behavior to a very simple measure of shared variability in a population of sensory neurons: the mean correlation between the spike count responses of pairs of neurons to repeated presentations of the same stimulus (termed spike count noise correlations, or r_{SC} ¹⁻²²). There are good reasons that so many studies have focused on correlated variability in sensory neurons:

- It is an obvious thing to measure as the field shifts from single neurons to populations.
- It is flexibly modulated by essentially every process that modulates perception including contrast, adaptation, learning, attention, task-switching, and arousal/motivation¹⁻²³.
- It is linked to behavior on a trial-by-trial basis^{13,24}.
- It can be changed by microinfusing neurotransmitters⁸, and pharmaceuticals¹⁴.
- It provides a powerful constraint on mechanistic models^{24,26-30,47,48}.

Together, these observations that correlated variability is so intertwined with flexible behavior suggest a tantalizing possibility: that understanding the relationship between correlated variability and behavior could be key to repairing and enhancing cognition in healthy subjects, patients, and even in artificial networks.

We need a new hypothesis. The dominant hypothesis that correlated variability corrupts information coding is inconsistent with recent observations. Theoretical work shows that neuronal population decoders can easily ignore shared population trial-to-trial variability^{24,33,38-42}. Decoders for the specific task can ignore such correlated variability because it does not corrupt the population response along the relevant dimensions of neuronal response space for stimulus decoding. Indeed, we recently demonstrated experimentally that despite the fact that it changes correlated variability, spatial attention does not strongly modulate the amount of visual information that is encoded in a population of visual neurons¹⁹. Further, we showed that in a change detection task, the axis in neuronal population space that explains the most mean correlated variability explains virtually all of the choice-predictive signals in visual area V4^{13,24}.

These observations comprise a paradox: correlated variability need not be related to behavior, but it is.

Central hypothesis: the structure of correlated variability reflects the neuronal (sub)population that (directly or indirectly) guides behavior, not the quality of information encoded within the whole population. In other words, we hypothesize that the relationship between correlated variability and behavior stems not from its impact of information coding but because it reflects the subset of neural information that will be ultimately used to guide behavior (**Fig. 1**). This could be thought of as the activity of a set of projection neurons with correlated inputs that projects to a downstream area involved in decision-making. If validated, our central hypothesis can reconcile the apparent paradox between variability, coding, and behavior.

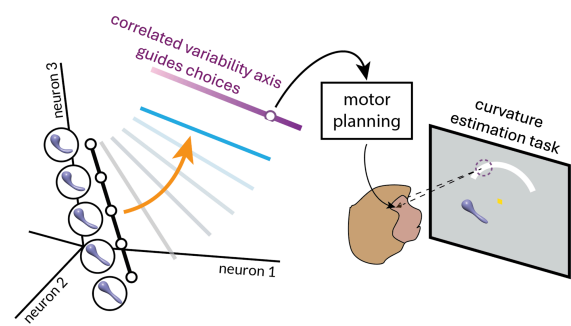


Figure 1. Schematic of our central hypothesis. Visual features like curvature are represented as a trace (black line) in the population space in which each dimension represents the response of one V4 neuron. We hypothesize that small response modulations associated with attention, task-switching, surround modulation, and/or motor planning transform that representation such that task-relevant information aligns with the axis that accounts for the most correlated variability in the population (purple line). This transformation causes projection neurons to communicate task-relevant subset of information to the neurons involved in planning actions.

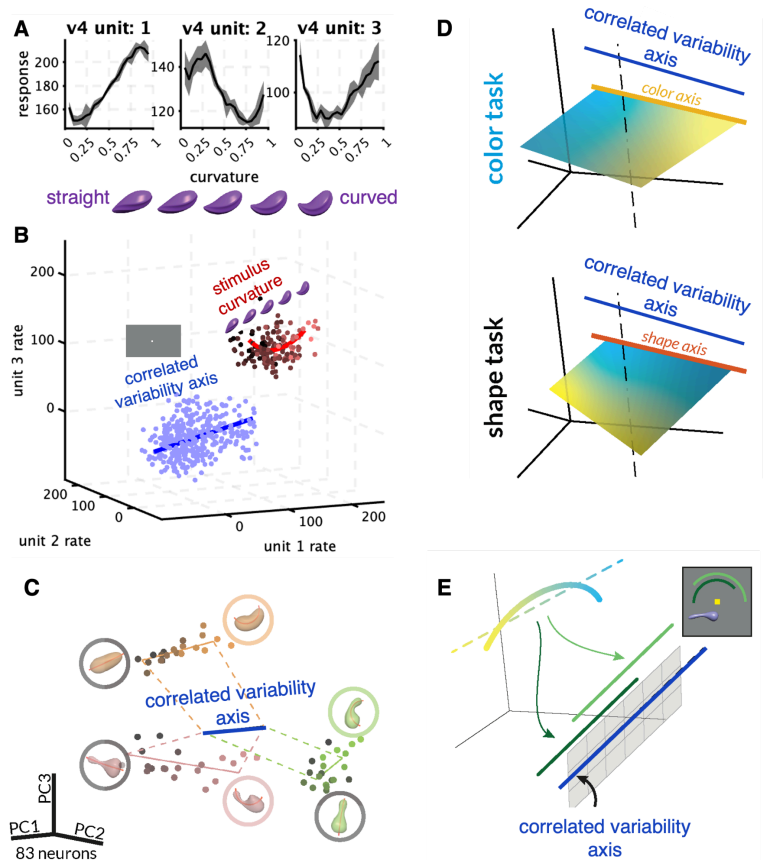
Both the sensory information used to guide behavior (**Fig. 2A**) and correlated variability can be approximated as a linear combination of the responses of sensory neurons. For example, consider a task in which a subject must estimate the value of one feature, such as the ripeness of a banana or, in our study, the medial axis curvature of a three-dimensional shape (purple icons in **Fig. 2A, B**). The response of a population of visual neurons, each of which is tuned for the feature of interest, can be visualized as a point in a high dimensional space in which each dimension is the spike count response of one neuron. Varying the feature of interest traces out a smooth path in that space (red points in **Fig. 2B** for the responses of three example V4 units; this is also true in simulations with Gaussian or monotonic tuning). Even though these paths are somewhat curved, the relationship between neurons and behavior is well-approximated by projecting that curve onto a line (e.g. a

linear combination of neural responses^{49,50}). Thus, neuronal activity in laboratory tasks is low dimensional (lie on a line), because the behaviors themselves are low dimensional (e.g. selecting from discrete choices).

Similarly, correlated variability in populations of sensory neurons is also typically low dimensional^{27,32–37}. As such, shared variability is well described by a low number of ‘latent variables’, rather than a high number of descriptors where each pair of neurons would have an exclusive source of correlation. Furthermore, cognitive processes primarily affect these low dimensional variables (approximately rank one^{27,32}), and leave marginal statistics (i.e variance) unaffected. For simplicity, we refer to the linear combination of neurons that explains the greatest variance in correlated variability as the **correlated variability axis** (blue line in Fig. 2B).

Our central hypothesis is that the correlated variability axis is approximately the same as the linear combination of neuronal responses that best explain the choices in perceptual tasks. We will test three predictions of this central hypothesis (one in each Aim; schematized in Fig. 2C-E).

Figure 2. Predictions of our central hypothesis. **A.** Curvature tuning curves of three example V4 units (purple icons represent five curvatures of one example 3D shape). Error bars represent standard error of the mean firing rate. **B.** Responses of those same example units plotted in a space where the response of each unit is one dimension. Blue points represent baseline responses recorded while monkeys fixated a blank screen. We define the correlated variability axis (blue line) as the first principal component of this cloud of points. Colors from red to black represents the curvature of one 3D stimulus. Each point is one trial. **C.** Hypothesis for Aim 1. For illustration, the responses of 83 V4 units are projected into the first three principal components. (All analyses are done in the full space whose dimensionality is the number of neurons.) Each point is the mean response to one shape at one curvature (colored to black points represent most to least curved shapes). Shapes and curvatures are randomly interleaved. The curvature representations of each shape are poorly aligned in neural population space. We will test the hypothesis that the monkeys’ ability to estimate curvature is related to the projection of the curvature representation onto the correlated variability axis (blue line). **D.** Hypothesis for Aim 2. Monkeys will alternate discriminating the color or shape of two stimuli. We will test the hypothesis that context-related control signals will rotate the representation of color and shape such that the representation of the task-relevant feature is aligned with the correlated variability axis (blue line). **E.**



Hypothesis for Aim 3. Monkeys will do a task variant that requires them to have flexible mappings between the curvature representation (yellow to blue curve) and the eye movement used to communicate the decision (green lines). We will test the hypothesis that since the motor plan is what guides behavior, control (premotor) signals transform the curvature representation such that the correct motor plan is aligned with the correlated variability axis (blue line).

Rationale for our hypothesis

Four lines of indirect evidence support our hypothesis.

1: Signal and noise are inextricably linked.

Essentially every study that measures correlated variability observes a relationship between signal and noise correlation: neurons with more similar tuning have higher noise correlations^{1,23,25,33,51–59}. From a mechanistic perspective, this is perhaps unsurprising. Neural responses come from the synaptic circuit in which the population is embedded (both feedforward and recurrent inputs). As such, the circuit defines the structure of both the trial-averaged and trial-variable population response; indeed, synaptic activity is agnostic to experimenter defined ideas of ‘signal’ and ‘noise’. The observed relationship between signal and noise may reflect a higher proportion of common inputs and/or a higher probability of direct connections between neurons with similar tuning^{23,60,61}. We therefore expect that the correlated variability axis is similar to the one that

accounts for the most signal, as defined as the responses of a population of neurons to stimuli that vary in every feature to which they are tuned.

2: A relatively fixed subset of neurons project out of a sensory population to a downstream target.

Only a small subset of neurons in any population project to any downstream target, much less a particular target that mediates a specific behavior. Recent studies in multiple species and using multiple experimental techniques have demonstrated that the sensory and task information that is communicated by any neural population and a downstream target is only a subset of that encoded in the whole population^{20,62–65}. Further, our previous theoretical work demonstrates that changes in noise correlations are positioned to have a much greater impact on the information that is communicated by the sub-population that projects out of a brain region compared to the information determined by the entire population^{26,28,31,66}.

3: Signal and noise dimensions are aligned in a change detection task.

Using V4 recordings and circuit models, we recently demonstrated that the dimensions of neural population space that best explain correlated variability align with the dimensions of population space that encode stimuli in a change detection task²⁴. This suggests that at least in this task and with these stimuli, it might be optimal to base behavior on noise dimensions because they are the same as signal dimensions.

However, this explanation might have been lucky: we tested it in one task in which the action was indistinguishable from the perceptual inference about stimuli that varied in only one feature. Whether this explanation generalizes to the rich and flexible relationships between natural stimuli and behavior is unknown.

4: Area V4 neurons are positioned to mediate multiple forms of flexible, visually-guided behavior.

Previous studies, including ours, show that neurons in area V4 are modulated by a wide range of visual and cognitive processes associated with the ability to flexibly use visual information in many ways. V4 neurons are modulated by spatial and feature attention^{67–71}, learning^{13,72–76}, task-switching^{43–45}, arousal/alertness^{15,77,78}, motor planning^{79–86}, objects in the receptive field surround^{87–93}, and stimulus contrast^{94–98}. These processes all typically affect single neuron responses, most typically by multiplicatively scaling tuning curves (gain changes^{71,99,100}). All of these processes are also all associated with changes in correlated variability^{1–23}.

Although efforts to analyze the relationship between these modulations on a population level and behavior are in early stages, our preliminary work suggests that these modulations are sufficient to support key predictions of our hypothesis. We found that despite the fact that spatial attention changes correlated variability, the same linear combination of the responses of visual neurons predicted choices in different attention conditions^{19,20,24}. Furthermore, in a study with a different focus but that produced substantial preliminary data for this proposal⁵⁰, we found that we could use stable linear decoders to a) predict judgments of one visual feature of stimuli that differ in task-irrelevant features, b) predict judgments about different features in a task-switching context, and c) predict different movements used to communicate a perceptual judgment.

Together, these observations suggest that a relatively fixed linear combination of V4 responses is most closely associated with behavior amid changes in stimuli, tasks, and internal cognitive processes. Here, we propose to test the hypothesis that this linear combination is the combination most aligned with correlative variability.

Challenges to our central hypothesis

There are two core challenges to our central hypothesis, which we propose to address using experimental and theoretical approaches, respectively. We will use both approaches to address questions in each Aim.

1: Identifying a general relationship between correlated variability and visually-guided behavior requires a controlled but flexible behavioral paradigm.

Our central hypothesis is very general, positing that correlated variability reflects the information that will guide behavior amid a wide range of stimuli, tasks, and behaviors used to communicate perceptual judgments. A core challenge is the requirement to measure the relationship between correlated variability and behavior in a set of task and stimulus conditions that are similar enough to be compared but varied enough to test the hypothesis that correlated variability reflects task-relevant information in many conditions.

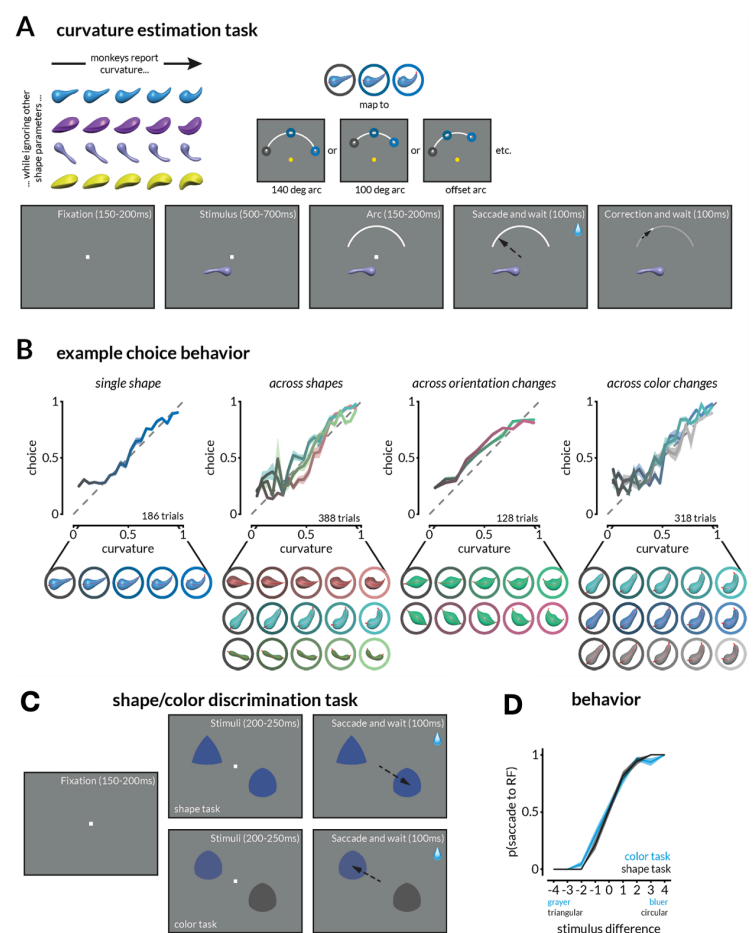
We have designed a behavioral paradigm that requires monkeys to exhibit many forms of flexibility in the relationship between visual information and behavior⁵⁰. They:

- make continuous perceptual judgments, which enhances our ability to link neurons to behavior
- generalize across stimuli with many task-irrelevant features to make a perceptual judgment about a single visual feature (important to **Aim 1**),

- switch between making judgments about each of two visual features (important for **Aim 2**), and
- maps each perceptual judgment onto one of several actions (important for **Aim 3**).

Monkeys will perform a curvature estimation task (**Fig. 3A**) while we record neural activity from area V4 using chronically implanted multi-electrode arrays^{3,13}. Monkeys fixate a central point on an otherwise blank screen for 150-200 ms, during which time we record spiking responses to calculate the correlated variability axis (**Fig. 1B**). While they maintain fixation, we present a three-dimensional stimulus in the joint receptive fields of the recorded neurons. The monkeys indicate their curvature inference at a corresponding position on a target arc presented in the upper hemifield. The arc has different sizes and is positioned at different locations. A saccade to the leftmost point of the arc indicates a straight stimulus (curvature = 0), a saccade to the right rightmost end corresponds to a maximally curved (curvature = 1) stimulus, and the monkeys can indicate any intermediate value in between. Monkeys are rewarded according to the accuracy of their estimate. The curvature, task-irrelevant features, and positions and locations are all randomly interleaved across trials. We generate the stimuli by randomizing the cross-sectional shape, aspect ratio, color, 3D rotation, twist along the axis, gloss, thickness profile, and other features of an oblong object, and monkeys are able to ignore these task-irrelevant features (**Fig. 3B**). In some sessions, monkeys perform a two-alternative, forced choice task in which they switch between discriminating the shape or color of two stimuli on randomly interleaved trials (**Fig. 3C,D**).

Figure 3. Tasks and behavior. **A.** Schematic of the continuous curvature estimation task (Aims 1 and 3). Stimuli that vary in curvature and task-irrelevant features are presented in the joint receptive fields of V4 neurons while monkeys fixate. After 550-800ms, a target arc is presented in the upper hemifield, and monkeys are rewarded for making a saccade to a location on the arc that corresponds to the stimulus curvature. The reward amount is inversely related (with a threshold) to the error in curvature judgment. In the experiments for Aim 3, the radial position, angular position, and length of the target arc are varied pseudorandomly. **B.** Monkeys report medial axis curvature while ignoring other stimulus features. Example continuous estimation behavior on four sessions during which the curvature of one or many shapes was estimated on interleaved trials. Shading indicates the standard error of the mean (SEM). **C.** Schematic of the two-feature (shape/color) discrimination task (Aim 2). The monkey is rewarded for making a saccade to one of the stimuli (one in the joint receptive fields of the recorded V4 neurons and one in the opposite hemifield). During the shape task, the colors of the two stimuli are the same and the more circular stimulus is rewarded. During the color task, the stimuli are the same shape and the bluer stimulus is rewarded. **D.** Example psychometric curves for the shape (black) and color (blue) task. The plot depicts the proportion of trials in which the monkey chose the stimulus in the receptive fields of the recorded V4 neurons as a function of the relevant feature of the stimulus that was in that receptive field. These data are from a single experimental session (314 total trials, 89% correct overall; 176 color task trials, 89% correct; 138 shape task trials, 90% correct).



2: Our central hypothesis requires control signals to modulate sensory neurons in a way that enables the flexible use of sensory information.

Our hypothesis relies on a top down control signal (e.g. associated with attention, task context, surround modulation, or motor planning) that modulates the responses of visual neurons such that the representation of the appropriate information is aligned with the correlated variability axis. We previously modeled spatial attention as a top down input that affects a spatially segregated sub-population of neurons²⁶. It is easy to imagine that this is biologically feasible, because there are many feedback connections that are spatially/feature organized. A goal of this grant is to determine whether biologically feasible control signals could perform the more subtle transformations required to enable different forms of flexible behavior.

We have developed a theoretical framework that provides the modulations needed to change the information available to a decoder of population activity²⁸. The theory is circuit-based, so it will also provide how control signals applied to the network should be arranged to maximize code performance. We expand below.

Sub-population codes. The performance of any decoder depends on a tradeoff between the sensitivity of the population response to a stimulus feature (i.e. signal) and population-wide response variability (i.e. noise). Both of these properties are affected by stimulus/task conditions. However, if network activity can be approximated as a linear system, then the signal-to-noise (SNR) of the full population response will be the same regardless of the condition (**Fig. 4A**). This is because if a condition increases the neuronal gain through recurrent circuit interactions, the same circuitry also increases the noise. Thus, there is no net change in SNR^{26,28,31}.

The invariance of SNR with condition is at odds with our expectation that stimulus structure, task conditions, and cognitive states should affect information coding, since they affect animal behavior. This perceived conflict of the invariance of SNR to stimulus/task conditions requires that the decoder has access to all of the neurons in the population. This is an unreasonable assumption as most downstream regions only receive inputs from a select sub-population of projection neurons from the upstream area.

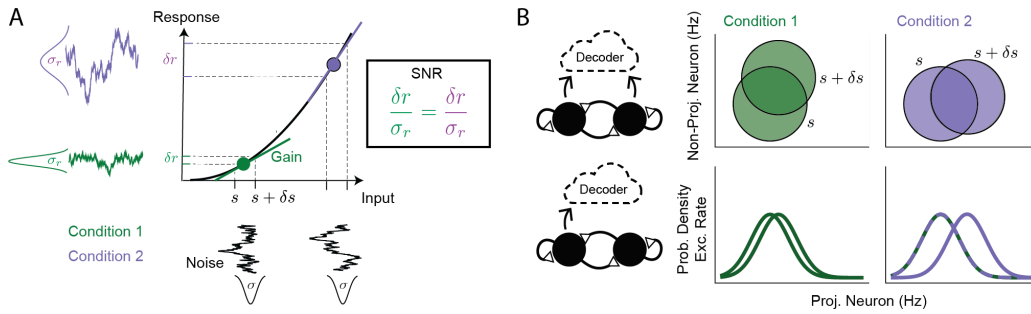
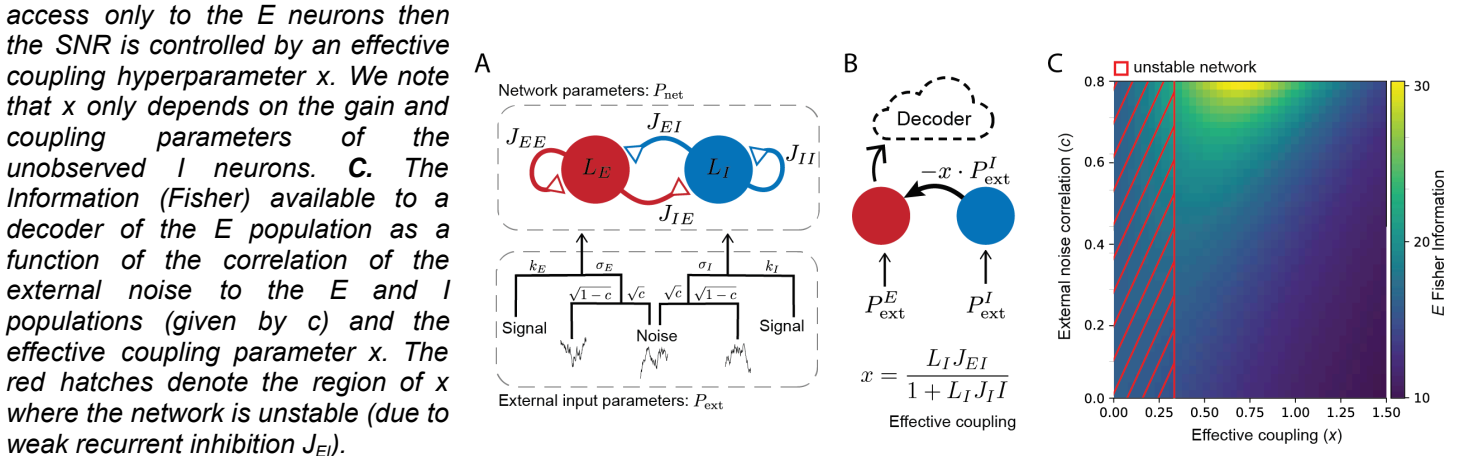


Figure 4: Impact of stimulus/task condition on information. **A.** Single neuron example of stimulus discrimination between stimulus s and $s + \delta s$ in the presence of input noise of variance σ . For condition 1 (green) the neuron is in a low gain state, so that both response difference δr and response variance σ_r are small.

By contrast, in condition 2 (purple) the neuron is in a high gain state so that δr and σ_r are large. However, the linear(ized) transfer causes the SNR $\delta r/\sigma_r$ to be identical for the low and high gain states. **B.** (Top) Same as A for a two neuron population. Circles represent the population variability, and their overlap denotes stimulus ambiguity. The condition rotates the responses yet the overlap in response remains unchanged, and hence information is also unchanged. (Bottom) If a decoder has access to only one neuron, the condition modulation changes the information communicated to the decoder.

When a decoder only receives projections from a subset of the cortical population, a control signal can impact the SNR by decoupling modulations in gain from those of response variability. This can come from rotations, translations, or dilations of population activity that, when viewed from the perspective of the projection neuron sub-population, can increase discriminability (**Fig. 4B**). This framework also identifies how control signals must be applied to the circuit to enact an increase (or decrease) in SNR. **Specifically, only modulations of the neurons that are not projection neurons can affect the SNR of the sub-population of projection neurons³¹.**

Figure 5: Controlling information in sub-population codes. **A.** Simple E-I network driven by external signal and noise (given by parameters P_{ext}) and with recurrent network interactions (given by parameters P_{net}). **B.** When a decoder has access only to the E neurons then the SNR is controlled by an effective coupling hyperparameter x . We note that x only depends on the gain and coupling parameters of the unobserved I neurons. **C.** The Information (Fisher) available to a decoder of the E population as a function of the correlation of the external noise to the E and I populations (given by c) and the effective coupling parameter x . The red hatches denote the region of x where the network is unstable (due to weak recurrent inhibition J_{EI}).



As an example, consider a simple excitatory (E) and inhibitory (I) network with input signal and noise distributed across the populations (**Fig. 5A**). Top-down control of the I population activity changes the SNR available to a decoder with access only to

the E population. This control is described by an effective coupling (x) that only depends on the I neuron gain (L_I) and synaptic coupling from and onto I neurons (**Fig. 5B**). The SNR (measured by the Fisher information) of the E population depends on the structure of the input noise (given by input correlation c) and the effective coupling (x) (**Fig. 5C**). Thus, a control signal that affects the I neuron gain L_I can modulate the SNR (through x). This framework generalizes to linear(ized) networks of arbitrary size and structure, with the effective coupling and its dependence on the coupling and gains of non-projecting neurons being easily computed.

Nevertheless, a challenge remains. Our current models involve simple control signals applied symmetrically to all non-projecting neurons. A major goal of this study is to extend this framework by experimenting with different control signals to determine whether transformations that can enable task switching (**Aim 2**) or a mapping between perception and action (**Aim 3**) can be instantiated using biologically plausible control signals.

Significance of the Research Outcomes

The proposed studies use a combination of experimental and theoretical approaches to test the hypothesis that correlated variability reflects the information that is communicated out of a population of sensory neurons to guide behavior. If true, this hypothesis would have broad implications for basic science, because we could use correlated variability in any brain area to investigate the flexible contribution of a neural population to a behavior. It would also have translational implications, because it is straightforward to modulate correlated variability using existing pharmaceuticals as well as by natural cognitive processes^{8,14,72}.

INNOVATION

Conceptual innovations: A new neural coding hypothesis. The proposed studies will test a novel neural coding hypothesis that can unify past results by resolving a longstanding paradox about why noise is so tightly linked to behavior. If true, this hypothesis would upend how we think about noise in the brain, changing how response variability is measured, modeled, and interpreted.

Experimental innovations: A combination of tasks to study the neural basis of many forms of flexible, visually-guided behavior. Our behavioral framework enables us to measure the neural basis underlying many forms of flexibility in the relationship between visual stimuli and behavior. These experiments will produce rich data sets suitable for testing the hypotheses in this proposal and beyond.

Theoretical innovations: Using biophysically realistic circuit models to understand control signals that could guide flexible behavior. We previously used biophysically realistic models of cortical circuits to understand how variability arises in and is communicated between neural populations^{24,26–28,30,31}. Here, we extend this work to understand control signals that could enable these circuits to mediate flexible behavior.

APPROACH

Our overall objective is to use intertwined behavioral, physiological, and theoretical approaches to understand the relationship between the activity of populations of visual neurons and behavior and to resolve the seemingly paradoxical relationship between correlated variability and behavior. Our central hypothesis is that correlated variability reflects signals that guide behavior via selective communication out of visual cortex. We propose to test this hypothesis experimentally by recording from groups of V4 neurons while monkeys use visual information in flexible ways. Theoretically, we will use biologically constrained circuit models to explore the relationship between correlated variability, flexible behavior, and the control signals that guide it. We use a unified analysis approach for data and models to measure the relationships between correlated variability and visual, task, and premotor signals. Each Aim tests one prediction of our central hypothesis using complementary experimental and theoretical approaches.

Experimental approach

Subjects: Our subjects will be adult rhesus monkeys (*Macaca mulatta*) between 4 and 15 kg and of either sex. Because typical studies involve two subjects (statistical power comes from the number of neurons or behavioral sessions, and experimental groups are task conditions), we do not analyze differences between groups of animals. However, in every experiment, we collect tuning data and simple behavioral data that will allow us to perform a meta study of sex as a biological variable using animals from multiple studies.

General approach: We will record from groups of neurons in V4 using chronically implanted multielectrode arrays (Blackrock Microsystems, 256 channels) while monkeys perform tasks (**Fig. 3**). We place the stimulus within the joint receptive fields of the recorded neurons and position the target arc outside the response fields.

Rigor and reproducibility: A cornerstone of our approach is using continuous estimation tasks. Our previous computational work showed that continuous tasks provide substantially more statistical power with which to

assess the relationship between neurons and behavior¹⁰¹. We used methods described in¹⁰² to assess the number of experimental sessions necessary for each Aim given effect sizes comparable to those in our preliminary data. We estimate that we will need approximately 40 experiments per Aim to establish relationships between behavior and correlated variability with 80% power. We anticipate that we will need 45-60 minutes to collect sufficient data for the primary recording experiments, and another 5-10 minutes to collect tuning data and supplementary measurements.

Many of our analysis methods rely on some form of linear regression. All of these analyses will be cross validated, using different subsets of trials to find weights and test performance. In particular, we will use the initial fixation period of each trial (when the monkeys fixate a blank screen) to identify the correlated variability axis and responses to the stimulus or stimulus + arc to measure stimulus representations.

Theoretical approach

Model: We will build and analyze networks of spiking neuron models guided by our past work^{26–28,30,31,66}. Populations of excitatory (*E*) and inhibitory (*I*) neurons will be modeled with an adaptive integrate-and-fire formalism, and synaptic coupling will obey known kinetics. Networks will have *N* neurons with *E* neurons having $4N/5$ members, and *I* neurons having $N/5$. Neurons will receive feedforward input carrying information about stimulus features like curvature, color, texture, blur, orientation, depth, etc. (**Fig. 6A**). We assume *K* stimulus features: s^1, \dots, s^K which define inputs to neurons that have a prescribed feature selectivity s_i^1, \dots, s_i^K (for neuron *i*). The recurrent wiring in the network will be probabilistic with connections statistics that depends on the difference of preferred features $s_i^1 - s_j^1, \dots, s_i^K - s_j^K$ (**Fig. 6A**). These *E* - *I* networks produce irregular spiking dynamics that is consistent with cortical recordings (**Fig. 6B₁,B₂**). Importantly, neurons show a clear stimulus tuning (**Fig. 6C₁**) that has a broad distribution of response selectivity (**Fig. 6C₂**).

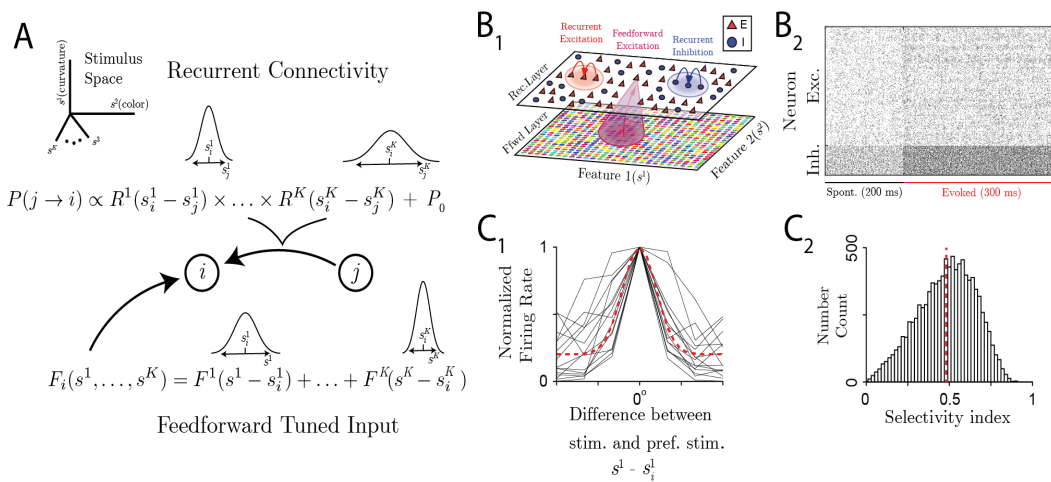


Figure 6: Circuit Model with feature-based connectivity. **A.** Schematic of the feedforward and feedback coupling in the model. Each *E* neuron *i* receives feedforward tuned input F_i from all *K* stimulus pathways (F is a Gaussian with a given spread σ). The probability of a recurrent connection from neuron *j* to neuron *i* obeys a product of Gaussians *R*, each centered at the difference of feature selectivity between neuron *i* and *j*. **B.** The *E* and *I* neurons have feature-dependent

coupling (B_1), and the network produces Poisson-like spiking statistics (B_2). **C.** Neurons have tuned responses to each feature (C_1), and a broad range of tuning strength (C_2).

Rigor and reproducibility: Solutions of network models are often criticized for being fragile, in part because they require fine-tuning of model parameters to capture a desired aspect of the data. However, networks with large excitation and inhibition that maintain a dynamic balance are famously robust to parameter changes. To ensure that balanced solutions are obtained we will in addition to having feature dependent components in the recurrent coupling we will also have a feature independent component (term P_0 in **Fig. 7A**, top). This will allow a global balance condition to the network and give robust solutions (**Fig. 7B** and **C** use this with $K=3$).

A benefit of our modeling framework is that the output of our models are spike trains, as in the experiments. This means that we will apply the same analysis techniques for both experiment and model responses. This is often not possible for models with firing rate equations, where often (stochastic) interpretations of rate variables must be made to compare directly with data.

Prediction/Aim 1: Observers will be better able to discriminate stimuli whose representations happen to be aligned with the correlated variability axis.

Objectives and rationale: There is considerable heterogeneity in observers' abilities to judge even the same feature of different stimuli. For example, observers might have an easier time judging the shade of yellow of a banana than a sunflower. Similarly, there is considerable heterogeneity in the way the same feature of different stimuli is encoded in the brain⁵⁰, meaning that the color of bananas might trace out a path in neural population space that is different in shape and/or alignment from the path corresponding to the color of flowers.

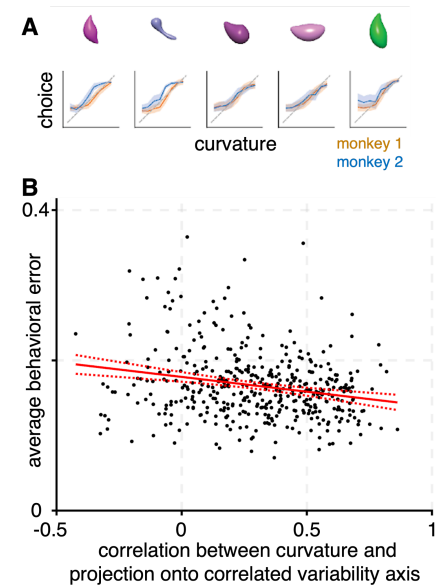
The goal of this Aim is to test the hypothesis that there is a relationship between behavioral and neuronal heterogeneity. Experimentally, we will test the prediction of our hypothesis that observers will make more accurate judgments of the curvature of objects whose representation in area V4 aligns better with the correlated variability axis (e.g. Fig. 1C). This is the opposite prediction of the alternative hypothesis, which posits that correlated variability harms information coding, meaning observers should make less accurate judgments of shapes whose curvature representation is aligned with the correlated variability axis. Theoretically, we will test the related hypothesis that with biophysically realistic patterns of network connectivity, the optimal decoding of one stimulus feature from population response will be aligned with the correlated variability axis, regardless of variation in task-irrelevant features.

Experimental design and preliminary results: We will record V4 responses while monkeys perform the curvature estimation task (**Fig. 3A**). As in all Aims, we use responses during the initial fixation period to identify the correlated variability axis. On randomly interleaved trials, monkeys judge the curvature of shapes that vary in task-irrelevant features (color, orientation, aspect ratio, and other shape parameters; **Fig. 7A**). We will assess psychophysical performance by calculating the average absolute value difference between the reported and actual curvature of each shape, averaged across all curvatures (at least 20 curvatures per shape).

Our primary analysis method is to project responses onto the correlated variability axis and calculate the correlation between those projections and the actual curvature of each stimulus (calculated separately for each shape; the sign of the correlation is calculated for one set of trials and then tested on another[MC2]). A strong alignment between the curvature representation and the correlated variability axis would result in a strong correlation; orthogonal representations would produce zero correlation.

Our preliminary data support the prediction of our central hypothesis. **Fig. 7B** shows the behavioral error as a function of the correlation between correlated variability and shape curvature (each point is one shape; each recording session contributes 3-6 shapes). There is a negative relationship between these behavioral and neuronal measurements, indicating that poor psychophysical performance (high error) is associated with worse alignment between the curvature representation and the correlated variability axis.

Figure 7. Preliminary data indicating that perception is better for shapes whose curvature representation is more strongly aligned with the correlated variability axis. **A.** Psychometric curves (chosen vs. actual stimulus curvature) for five example shapes. There is shape-to-shape heterogeneity in performance that is common to the two monkeys (colors). **B.** Preliminary data indicating that shapes for which the curvature representation is more strongly aligned with the correlated variability axis (x-axis) are associated with lower behavioral error in the curvature estimate task (y-axis). We summarize performance as average behavioral error (mean absolute difference between chosen and actual curvature). Each point represents one shape. During each recording session, trials using 3-6 shapes are randomly interleaved.



This relationship is notable for two reasons. First, a key prediction of the alternative hypothesis that correlated variability corrupts stimulus coding is that the key signal dimensions should be orthogonal to the correlated variability. However, the correlation between stimulus curvature and the projection onto the correlated variability axis is non-zero for most shapes (x-axis in **Fig. 4B**) indicating substantial alignment between signal and noise. Second, several experimental factors should dilute the relationship between alignment and behavior: there is day to day variability in neurons and in the monkeys' internal states, the correlated variability axis is computed during a different time period than the neural and behavioral responses to the stimuli, our ability to estimate the curvature representations and correlated variability axis are data-limited, etc. The fact that this relationship persists despite these factors is strong preliminary support for our hypothesis.

Theoretical design and preliminary results: The theoretical goal of this Aim is to use our network model to explore how feature-based wiring in the cortical network establishes a low-dimensional space in which

fluctuations are confined, and how this relates to the axis of stimulus representation. We will build our spiking model network with K stimuli features (see **Fig. 6**). If $K \ll N$ (as will always be the case in our work) we expect noise correlations to be low dimensional^{103,104}, inheriting this from the rank K connectivity component due to feature-based wiring (the $P(j \rightarrow i)$ expression in **Fig. 6A**). We previously analyzed firing rate networks with $K=1$ feature-based connectivity and indeed a component of internally generated noise correlations were aligned with the stimulus space¹⁰⁵. In this network, when a stimulus s_1 elicits a response in neurons tuned to $s_1 = s_{N/2}$, then noise correlations are positive for similarly tuned neurons and negative for differently tuned neurons, owing to fluctuations in the Gaussian shaped response (**Fig. 8A,B**). In our spiking network model we will measure the covariability axis (technically a K -dim plan for $K>1$) in the absence of stimulus (modeling the fixation period). We will test that the covariability axis is not greatly shifted in evoked states.

We expect that the low rank structure of the noise correlations (of which the principal component is the covariability axis) will be compromised if K is too large, since correlations should be spread along many feature defined pathways in the circuit. This is because for large K then the recurrent wiring $P(j \rightarrow i)$ (**Fig. 6A**) will become less specific through the successive multiplications of shifted Gaussians associated with each stimulus feature. As such, a one dimensional covariability axis will not be possible, which is at odds with our recordings (**Fig. 2B**). Consequently, we predict that the number of features whose composition represents an object must be small enough so that structured connectivity is not diffuse (in the graph). This model will be an essential tool for all Aims.

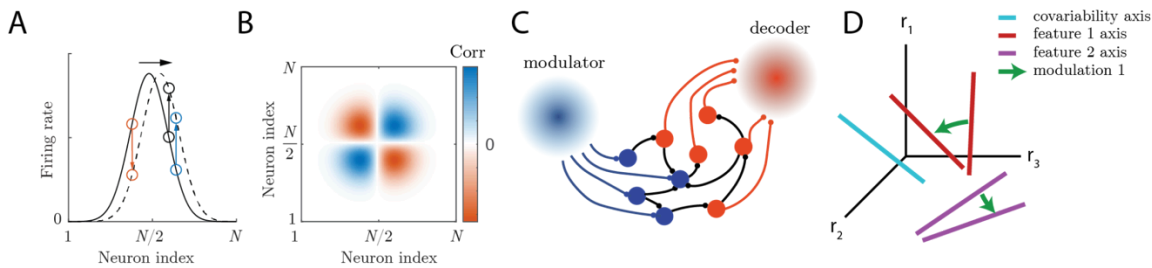


Figure 8. Stimulus aligned covariability in model networks. **A.** Responses of neurons to a stimulus s . We organize our neuron indices according to their tuning for s . **B.** Fluctuations in

population activity create positive correlations for similarly tuned and negative correlations for dissimilarly tuned neurons (**A** and **B** are from¹⁰⁵). **C.** Schematic of the modulated (blue) and the decoded (red) sub-populations of our recurrent circuit model. **D.** Schematic of a modulation (performed by a control signal) that aligns the population representation of feature 1 with the covariability axis, yet does not align other representations (feature 2) with the covariability axis.

Potential pitfalls and alternatives: The physiological, psychophysical, theoretical, and data analysis methods for this Aim are well established in our labs, and we have completed preliminary experiments and models, so we do not anticipate difficulties completing the goals of this Aim. Experimentally, the biggest challenge is obtaining sufficient data that are suitable for calculating the correlated variability axis. By design, this experiment includes many stimuli that vary in curvature as well as task-irrelevant features. This means that unlike some of our previous studies in which we used a small number of very simple stimuli^{3,13,19,24}, we do not repeat any one stimulus enough to calculate correlated variability. For this reason and for cross-validation, we identify the correlated variability axis from responses during a completely different time period (the initial 200 ms after the monkey attains fixation and before the stimuli are presented) than the period when we calculate responses to different stimulus properties. This is sound experimental design, but it reduces the probability that noise will be related to signal. The most notable challenge is that estimates of both signal and noise are dominated by the activity of the highest responding cells, which will be different during baseline and stimulus periods^{1,106}. Our preliminary results (**Fig. 7**) suggest that these factors will not prevent us from observing a relationship between signal and correlated variability, but if it becomes a problem, we will compare results to a subset of experiments in which we present many more repeated trials of a much smaller subset of stimuli. Further, the model circuit may not produce covariability that is sufficiently low dimensional to identify a covariability axis. In this case we will provide low rank input noise (correlated) to constrain population activity.

Prediction/Aim 2: Representations of stimulus features will be better aligned with the correlated variability axis when they are used to guide behavior than when they are not.

Objectives and rationale: Like neurons in all sensory areas, V4 neurons are tuned to multiple features, including which include curvature, color, texture, blur, center of mass, orientation, depth, and many other shape attributes (for review, see (71, 72)). In general, features that do not covary in natural stimuli are represented independently (refs, mixed selectivity). Put another way, the tuning of a neuron to curvature does not tell us anything about its tuning for color. In neural population space, this means that we expect the representations of color and curvature to be orthogonal. Monkeys and humans can easily switch between discriminating different

features of an object, such as reporting curvature on some trials and color on others. Orthogonal representations form a challenge to our central hypothesis that the information used to guide choices is aligned to the correlated variability axis. This requires that some control signal modulates V4 responses such that curvature information, but not color, is aligned to the correlated variability axis when the monkey makes curvature judgments, and the opposite for color judgments.

The goal of this Aim is to test those predictions. Experimentally, we will test the prediction that the representation of each feature is more aligned with correlated variability when that feature guides behavior than when it does not. This prediction is at odds with the alternative hypothesis that correlated variability harms information coding, so any task-dependent modulations should reduce the alignment between the representation of the relevant feature and correlated variability. Theoretically, we will test the hypothesis that simple, biologically plausible control signals can rotate the representations of different features in the circuit model so that the representation of the task-relevant feature is aligned with the correlated variability axis.

Experimental design and preliminary results: We will test this prediction experimentally by recording from groups of V4 neurons while monkeys perform a simplified two-alternative forced choice task in which they must discriminate either the color (gray to blue) or the curviness (triangle to circle) of the displayed shape (**Fig. 3C**) on randomly interleaved trials. We present one stimulus inside the joint receptive fields of the recorded neurons' and one in the opposite hemifield. The monkey will be rewarded for selecting the bluer stimulus (color task) or the more circular one (shape task). In preliminary experiments, the recorded neurons had a range of tuning preferences and displayed similar patterns and magnitudes of gain modulation that depended on the neuron's tuning and the stimulus outside the receptive field as in previously published accounts of task switching, feature attention, and surround modulation^{50,72,100}.

The key prediction is that while color and curvature are encoded in different dimensions of neural population spaces, choices will always be aligned with the correlated variability axis. Therefore, if choices are based on different features on randomly interleaved trials, those gain changes must reformat the representations of color and curvature so that the relevant information is aligned with the correlated variability axis on each trial. In contrast, the alternative hypothesis suggests that correlated variability harms information coding, so it predicts that the representation of the task-relevant feature should be *less* aligned with the correlated variability axis than the representation of the task-irrelevant feature.

Fig. 9 shows preliminary support for this hypothesis. The analysis method is similar to Aim 1: we identify the correlated variability axis from periods of fixation before stimuli presentation and then assess the extent to which we can decode stimulus features from projections of responses onto that axis. In our preliminary experiments, our ability to distinguish the curvature of the stimulus in the joint receptive fields of the recorded neurons from projections on to the correlated variability axis was better on trials when that feature was task-relevant (x-axis) than on trials when curvature was irrelevant (y-axis). These results provide preliminary support for our central hypothesis that the representation of a stimulus feature (e.g. curvature) is more aligned with correlated variability when it is task-relevant and that the task-dependent gain changes coordinate across the population to modulate this alignment.

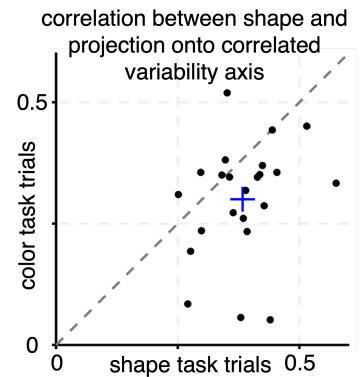


Figure 9. Preliminary support for the hypothesis that information about a visual feature becomes more aligned with the correlated variability axis. The plot shows the correlation between the shape of the stimulus in the V4 neurons' joint receptive field and projections onto the correlated variability axis while monkeys discriminate the color (y-axis) or shape (x-axis) of the pair of stimuli in the 2AFC task (**Fig. 2C**). Each point represents one recording session in one monkey. Most points lie below the diagonal, indicating that the representation of shape is better aligned with the correlated variability axis when that feature is task-relevant.

Theoretical design and preliminary results: We will use our model from **Aim 1** to deduce which control signals can enable task switching by aligning appropriate information with the correlated variability axis. In our network of N neurons, we assume that $P < N$ excitatory neurons project to a downstream decoder (**Fig. 8C**, red), while the remaining $M=N-P$ neurons do not (this population includes all I neurons and the E neurons that we do not decode). Our preliminary theory suggests that the M non-projecting neurons are viable targets for modulation that will affect the stimulus information in the projection population (**Fig. 4, 5**). We will model a control signal as a top-down bias to the non-projecting neurons through the input vector $m = [m_1, \dots, m_M]$ (**Fig. 8C**, blue). Our past work considered a homogeneous modulation (i.e. $m_i=m_j$ for all i and j)^{26,27}, but for the proposed studies, heterogeneous perturbations will better control population response.

Since we know the gain of the non-projecting neurons and how it is modulated via m , then we can calculate the effective coupling parameter x (as in **Fig. 4B**); x will depend on the gains and synaptic weights (in and out) of the M non-projecting neurons³¹. Given x , we can calculate the modulated linear Fisher information of the projecting population (**Fig. 8C**). The control signal m will be chosen to maximize the information about the stimulus s^1 in the projecting population. We hypothesize that the optimal modulation will significantly increase the gain to stimulus s^1 , while also (necessarily) increasing their noise correlations. Nevertheless, the benefit of increased gain will outweigh the cost of increased noise correlations so as to maximize information about the stimulus in the subpopulation code. We will test the hypothesis that the modulation that optimizes information in the projecting population will also align the representation of s^1 with the covariability axis (**Fig. 8D, red**). Further, while the modulations will also shift the population code for the other stimuli features (s^2, \dots, s^K) we hypothesize that they will not align with the covariability axis (**Fig. 8D, purple**).

Potential pitfalls and alternatives: The physiological, psychophysical, theoretical, and data analysis methods for this Aim are well established in our labs, and we have completed preliminary experiments and models, so we do not anticipate difficulties completing the goals of this Aim. Experimentally, the biggest challenge is that training monkeys to switch between discriminating two features is difficult and time consuming. For this reason, we did our preliminary experiments using a simplified two alternative, forced-choice task in which the monkey can use a combined strategy (look at the curvier, bluer strategy), and we switch the relevant feature by changing the relative properties of the two stimuli (e.g. we create color discrimination using stimuli with identical curvature). In this task, gain changes might be attributed to task-switching, surround modulation, or a combination. The primary impact of an insufficient task would be a lack of difference between the two task conditions. Our preliminary data suggest that this task will be sufficient to see context-dependent changes in the alignment between the representations of different features and correlated variability. However, if future data suggest that this task will be insufficient to test our hypotheses, we can train our animals to switch between the continuous curvature estimation task we use in Aims 1 and 3 and an analogous continuous color estimation task. We have extensive experience training animals on this sort of task switching^{46,107}.

In the network model, the linear Fisher information may be relatively insensitive to any modulation. If so, we input external sources of correlation that will need to be canceled by the recurrent network. That cancellation will be controlled by the modulation signal (see **Fig. 5C**). In this scenario, the same modulation affects the Fisher information to a much larger degree when $c=0.8$ compared to $c=0$.

Prediction/Aim 3: Correlated variability will be better aligned with representations of action plans than the perceptual judgment.

Objectives and rationale: Like other primarily sensory areas, V4 is known for its role in perceptual decision-making, much more than for its modest modulations by motor planning and other extra-retinal processes like surround modulation or directing spatial attention outside its receptive field. For this reason, a strong test of our central hypothesis that correlated variability reflects the signals that will guide behavior is something unintuitive: that correlated variability should be more strongly related to the actions used to communicate a perceptual judgment than to the judgment itself.

The goal of this Aim is to test this prediction by dissociating the perceptual judgment from the action that is used to communicate it. Experimentally, we will train monkeys to report curvature judgment by making an eye movement to the appropriate relative location on an arc whose position and length varies from trial to trial in an uncued way. We will test the prediction that the axis that best predicts the location of the upcoming eye movement is more aligned with correlated variability than the axis that best predicts the curvature judgment on a relative scale. Theoretically, we will extend our models to guide a flexible action and determine whether biologically plausible control signals could align the appropriate action plan with the correlated variability axis.

Experimental design and preliminary results: We will test this prediction by recording from V4 neurons while monkeys perform a variation of the curvature estimation task (**Fig. 3A**) in which the arc that the monkeys use to communicate their curvature judgment changes location and/or length on randomly interleaved trials in an uncued way. The arc appears at a random location and length after the stimulus has been on for 550-800 ms, and the monkeys must plan an appropriate eye movement that they execute after the fixation point disappears.

Our preliminary results support the prediction that the representation of the upcoming eye movement is more aligned with the correlated variability axis than the representation of the curvature of the stimulus. In the full population space (where the response of each neuron represents one dimension), our ability to linearly decode the stimulus curvature and the direction of the upcoming eye movement (saccade) is similar (**Fig. 10A**, left two

bars, and **Fig. 10B**). However, when projected onto the correlated variability axis, performance plummets for curvature but not for saccade decoding (**Fig. 10A**, right bars, and **Fig. 10B**).

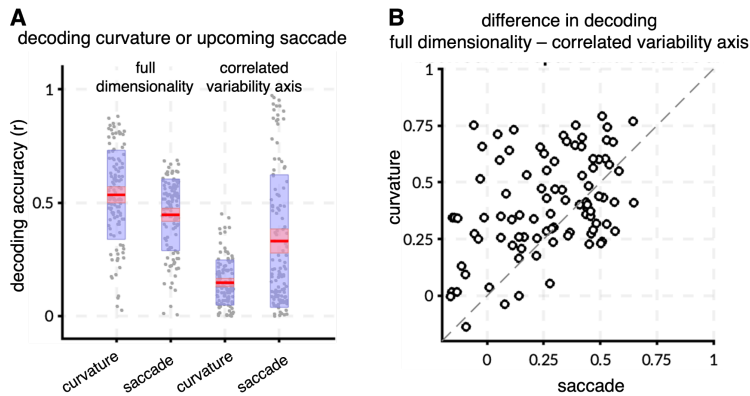


Figure 10. Preliminary support for the hypothesis that information about the upcoming action is more aligned with the correlated variability axis than stimulus information. **A.** Ability to decode stimulus curvature and the direction of the upcoming eye movement (saccade) in the full population space are similar (left two bars). Curvature decoding is much worse when projected onto the correlated variability axis, while saccade decoding is affected less (right bars). Each recording session contributes one point to each bar. **B.** Replotting of the data in (A), showing that for most sessions, the difference between decoding in the full population space and along the correlated variability axis is greater for curvature (y-axis) than saccade (x-axis; most points lie above the unity line). Each session is one point.

Theoretical design and preliminary results: To model the coding of eye movements in V4 activity we will take a one dimensional collapse of the activity of the projecting population. This dimension corresponds to the angle of the eye movement that the monkeys use to report their curvature estimate. The control signal will still be defined using the whole modulation population (of size M) as in Aim 2. We will choose the projection weights from the P ($\gg 1$) projecting neurons to maximize the Fisher information about feature value s^i obtained from the output dimension. This massive dimensionality reduction (from P to 1) represents a strong nonlinearity that will in principle reduce the available information about any stimulus feature. However, we hypothesize that the optimization of information about s^i will cause an even better alignment with the correlated variability axis. Indeed, to maximize information in this single dimension, even more value will be placed on high gain neuronal responses which are aligned with the axis of covariability. High gain is needed due to the low dimension of the output, necessitating mechanisms to best separate the mean responses to stimuli s^i and $s^i + \Delta$ (Δ is a small difference; needed to compute Fisher information), since fluctuations will overlap significantly in the output dimension and will be difficult to organize to maximize information (unlike in high dimensions).

Potential pitfalls and alternatives: As in the other Aims, the experimental and theoretical methods for this Aim are underway in our labs, so we do not anticipate problems completing the primary goals of this Aim. The biggest challenge is that interleaving trials with different arc locations and sizes makes the task considerably more difficult for the monkeys. To moderate the difficulty and maintain the monkey's motivation in our preliminary experiments, we had to make the differences in the location and size of the response arcs somewhat subtle. This limits our ability to dissociate the representation of curvature from the representation of the eye movement plan. Despite this limitation, in preliminary experiments, we were able to make the dissociation in our neural data (**Fig. 10**). If this is a problem in future experiments, we could achieve the goals of this Aim via alternative behavioral strategies, like teaching the monkeys to wait longer before initiating the eye movement or making the perceptual part of the task easier (e.g. by relaxing the accuracy needed to obtain a reward). We can also maximize our chances of success by picking the exact arc manipulations (e.g. specific locations or lengths) that are maximally discriminable in the responses of the recorded neurons.

The theory aspects of this Aim are new: our previous work has never modeled action selection. We are enthusiastic about the potential of these studies to guide our future efforts to model a variety of behaviors. If we do not identify control signals that can explicitly guide eye movements, we will be able to accomplish the primary goals of this Aim by compromising biological realism and modeling the behavior as a back end recurrent network module (comparable to those we have used to model other behaviors^{50,107}).

Research impact

The proposed studies will test a novel neural coding hypothesis that, counter to the hypotheses that have driven decades of work, correlated variability reflects rather than corrupts the signals that guide behavior. The results of our studies can resolve a longstanding paradox about why noise is so tightly linked to behavior. Doing so will change the way neural populations are measured, modeled, and interpreted in a range of species, brain areas, and tasks. The resulting discoveries will provide insights into the mechanisms by which sensory information is flexibly perceived, remembered, and acted upon. These efforts could therefore have broad implications for efforts to repair and enhance cognitive flexibility.

REFERENCES

1. Cohen, M. R. & Kohn, A. Measuring and interpreting neuronal correlations. *Nat Neurosci* vol. 14 811–9 (2011).
2. Nirenberg, S. & Latham, P. E. Decoding neuronal spike trains: how important are correlations? *Proc. Natl. Acad. Sci. U. S. A.* **100**, 7348–53 (2003).
3. Cohen, M. R. & Maunsell, J. H. Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* vol. 12 1594–600 (2009).
4. Cohen, M. R. & Maunsell, J. H. Using neuronal populations to study the mechanisms underlying spatial and feature attention. *Neuron* vol. 70 1192–204 (2011).
5. Gregoriou, G. G., Gotts, S. J., Zhou, H. & Desimone, R. High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* **324**, 1207–10 (2009).
6. Gregoriou, G. G., Rossi, A. F., Ungerleider, L. G. & Desimone, R. Lesions of prefrontal cortex reduce attentional modulation of neuronal responses and synchrony in V4. *Nat. Neurosci.* **17**, 1003–11 (2014).
7. Gu, Y. *et al.* Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* **71**, 750–61 (2011).
8. Herrero, J. L., Gieselmann, M., Sanayei, M. & Thiele, A. Attention-induced variance and noise correlation reduction in macaque V1 is mediated by NMDA receptors. *Neuron* **78**, 729–39 (2013).
9. Luo, T. Z. & Maunsell, J. H. R. Neuronal Modulations in Visual Cortex Are Associated with Only One of Multiple Components of Attention. *Neuron* **86**, 1182–1188 (2015).
10. Mayo, J. P. & Maunsell, J. H. Graded Neuronal Modulations Related to Visual Spatial Attention. *J Neurosci* **36**, 5353–61 (2016).
11. Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* vol. 63 879–88 (2009).
12. Nandy, A. S., Nassi, J. J. & Reynolds, J. H. Laminar Organization of Attentional Modulation in Macaque Visual Area V4. *Neuron* **93**, 235–246 (2016).
13. Ni, A. M., Ruff, D. A., Alberts, J. J., Symmonds, J. & Cohen, M. R. Learning and attention reveal a general relationship between population activity and behavior. *Science* vol. 359 463–465 (2018).
14. Ni, A. M., Bowes, B. S., Ruff, D. A. & Cohen, M. R. Methylphenidate as a causal test of translational and basic neural coding hypotheses. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2120529119 (2022).
15. Ruff, D. A. & Cohen, M. R. Global cognitive factors modulate correlated response variability between V4 neurons. *J Neurosci* vol. 34 16408–16 (2014).
16. Ruff, D. A. & Cohen, M. R. Attention can either increase or decrease spike count correlations in visual cortex. *Nat. Neurosci.* **17**, 1591–1597 (2014).
17. Ruff, D. A. & Cohen, M. R. Attention Increases Spike Count Correlations between Visual Cortical Areas. *J Neurosci* vol. 36 7523–34 (2016).
18. Ruff, D. A. & Cohen, M. R. Stimulus Dependence of Correlated Variability across Cortical Areas. *J Neurosci* vol. 36 7546–56 (2016).
19. Ruff, D. A. & Cohen, M. R. Simultaneous multi-area recordings suggest that attention improves performance by reshaping stimulus representations. *Nat Neurosci* vol. 22 1669–1676 (2019).
20. Srinath, R., Ruff, D. A. & Cohen, M. R. Attention improves information flow between neuronal populations without changing the communication subspace. *Curr Biol* **31**, 5299–5313 e4 (2021).
21. Yan, Y. *et al.* Perceptual training continuously refines neuronal population codes in primary visual cortex. *Nat. Neurosci.* **17**, 1380–7 (2014).
22. Zénon, A. & Krauzlis, R. Attention deficits without cortical neuronal deficits. *Nature* **489**, 434–437 (2012).
23. Poort, J. *et al.* Learning and attention increase visual response selectivity through distinct mechanisms. *Neuron* **110**, 686–697.e6 (2022).
24. Ni, A. M., Huang, C., Doiron, B. & Cohen, M. R. A general decoding strategy explains the relationship between behavior and correlated variability. *eLife* **11**, e67258 (2022).

25. Nienborg, H., Cohen, M. R. & Cumming, B. G. Decision-Related Activity in Sensory Neurons : Correlations Among Neurons and with Behavior. *Annu. Rev. Neurosci.* **35**, 463–483 (2012).
26. Kanashiro, T., Ocker, G. K., Cohen, M. R. & Doiron, B. Attentional modulation of neuronal variability in circuit models of cortex. *Elife* vol. 6 (2017).
27. Huang, C. *et al.* Circuit Models of Low-Dimensional Shared Variability in Cortical Networks. *Neuron* vol. 101 337–348.e4 (2019).
28. Negrón, A., Getz, M. P., Handy, G. & Doiron, B. The mechanics of correlated variability in segregated cortical excitatory subnetworks. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2306800121 (2024).
29. Rosenbaum, R., Smith, M. A., Kohn, A., Rubin, J. E. & Doiron, B. The spatial structure of correlated neuronal variability. *Nat. Neurosci.* **20**, 107–114 (2017).
30. Rouse, T. C., Ni, A. M., Huang, C. & Cohen, M. R. Topological insights into the neural basis of flexible behavior. *bioRxiv* 2021.09.24.461717 (2021) doi:10.1101/2021.09.24.461717.
31. Getz, M. P., Huang, C. & Doiron, B. Subpopulation Codes Permit Information Modulation Across Cortical States. 2022.09.28.509815 Preprint at <https://doi.org/10.1101/2022.09.28.509815> (2022).
32. Rabinowitz, N. C., Goris, R. L., Cohen, M. & Simoncelli, E. P. Attention stabilizes the shared gain of V4 populations. *Elife* **4**, e08998 (2015).
33. Kohn, A., Coen-cagli, R., Kanitscheider, I. & Pouget, A. Correlations and Neuronal Population Information. *Annu. Rev. Neurosci.* **39**, 237–256 (2016).
34. Lin, I. C., Okun, M., Carandini, M. & Harris, K. D. The Nature of Shared Cortical Variability. *Neuron* **87**, 644–56 (2015).
35. Ecker, A. S. *et al.* State dependence of noise correlations in macaque primary visual cortex. *Neuron* **82**, 235–48 (2014).
36. Williamson, R. C. *et al.* Scaling Properties of Dimensionality Reduction for Neural Populations and Network Models. *PLoS Comput Biol* **12**, e1005141 (2016).
37. Scholvinck, M. L., Saleem, A. B., Benucci, A., Harris, K. D. & Carandini, M. Cortical state determines global variability and correlations in visual cortex. *J Neurosci* **35**, 170–8 (2015).
38. Kafashan, M. *et al.* Scaling of sensory information in large neural populations shows signatures of information-limiting correlations. *Nat. Commun.* **12**, 473 (2021).
39. Kanitscheider, I., Coen-Cagli, R., Kohn, A. & Pouget, A. Measuring Fisher information accurately in correlated neural populations. *PLoS Comput Biol* **11**, e1004218 (2015).
40. Moreno-Bote, R. *et al.* Information-limiting correlations. *Nat. Neurosci.* **17**, 1410–1417 (2014).
41. Pitkow, X., Liu, S., Angelaki, D. E., DeAngelis, G. C. & Pouget, A. How Can Single Sensory Neurons Predict Behavior? *Neuron* **87**, 411–423 (2015).
42. Rummyantsev, O. I. *et al.* Fundamental bounds on the fidelity of sensory cortical coding. *Nature* **580**, 100–105 (2020).
43. Grothe, I. *et al.* Attention Selectively Gates Afferent Signal Transmission to Area V4. *J. Neurosci. Off. J. Soc. Neurosci.* **38**, 3441–3452 (2018).
44. Chen, J., Wu, S. & Li, F. Cognitive Neural Mechanism of Backward Inhibition and Deinhhibition: A Review. *Front. Behav. Neurosci.* **16**, 846369 (2022).
45. Premereur, E., Janssen, P. & Vanduffel, W. Functional MRI in Macaque Monkeys during Task Switching. *J. Neurosci. Off. J. Soc. Neurosci.* **38**, 10619–10630 (2018).
46. Xue, C., Kramer, L. E. & Cohen, M. R. Dynamic task-belief is an integral part of decision-making. *bioRxiv* 2021.04.05.438491 (2021).
47. Beiran, M., Meirhaeghe, N., Sohn, H., Jazayeri, M. & Ostojic, S. Parametric control of flexible timing through low-dimensional neural manifolds. *Neuron* **111**, 739–753.e8 (2023).
48. Jazayeri, M. & Ostojic, S. Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Curr. Opin. Neurobiol.* **70**, 113–120 (2021).
49. Kramer, L. E., Konkle, T. & Cohen, M. R. Contributions of the early visual system to high-level visual distinctions. *J. Vis.* **21**, 2412–2412 (2021).

50. Srinath, R., Czarnik, M. M. & Cohen, M. R. Coordinated Response Modulations Enable Flexible Use of Visual Information. *BioRxiv Prepr. Serv. Biol.* 2024.07.10.602774 (2024) doi:10.1101/2024.07.10.602774.
51. Kohn, A. & Smith, M. A. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J. Neurosci.* **25**, 3661–73 (2005).
52. Smith, M. A. & Kohn, A. Spatial and temporal scales of neuronal correlation in primary visual cortex. *J. Neurosci.* **28**, 12591–603 (2008).
53. Ecker, A. S. *et al.* Decorrelated neuronal firing in cortical microcircuits. *Science* vol. 327 584–7 (2010).
54. Huang, X. & Lisberger, S. G. Noise correlations in cortical area MT and their potential impact on trial-by-trial variation in the direction and speed of smooth-pursuit eye movements. *J. Neurophysiol.* **101**, 3012–30 (2009).
55. Cohen, M. R. & Newsome, W. T. Context-dependent changes in functional circuitry in visual area MT. *Neuron* vol. 60 162–73 (2008).
56. Cohen, M. R. & Newsome, W. T. Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *J. Neurosci.* **29**, 6635–48 (2009).
57. Solomon, S. G. & Kohn, A. Moving sensory adaptation beyond suppressive effects in single neurons. *Curr Biol* **24**, R1012-22 (2014).
58. Denman, D. J. & Contreras, D. The structure of pairwise correlation in mouse primary visual cortex reveals functional organization in the absence of an orientation map. *Cereb. Cortex N. Y. N* 1991 **24**, 2707–2720 (2014).
59. Cumming, B. G. & Nienborg, H. Feedforward and feedback sources of choice probability in neural population responses. *Curr Opin Neurobiol* **37**, 126–32 (2016).
60. Hofer, S. B. *et al.* Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nat. Neurosci.* **14**, 1045–1052 (2011).
61. Chadwick, A. *et al.* Learning shapes cortical dynamics to enhance integration of relevant sensory input. *Neuron* **111**, 106-120.e10 (2023).
62. Kohn, A. *et al.* Principles of Corticocortical Communication: Proposed Schemes and Design Considerations. *Trends Neurosci* **43**, 725–737 (2020).
63. Semedo, J. D., Zandvakili, A., Machens, C. K., Yu, B. M. & Kohn, A. Cortical Areas Interact through a Communication Subspace. *Neuron* vol. 102 249-259.e4 (2019).
64. Semedo, J. D., Gokcen, E., Machens, C. K., Kohn, A. & Yu, B. M. Statistical methods for dissecting interactions between brain areas. *Curr Opin Neurobiol* **65**, 59–69 (2020).
65. Gokcen, E. *et al.* Disentangling the flow of signals between populations of neurons. *Nat Comput Sci* **2**, 512–525.
66. Huang, C., Pouget, A. & Doiron, B. Internally generated population activity in cortical networks hinders information transmission. *Sci. Adv.* **8**, eabg5244 (2022).
67. Maunsell, J. H. R. Neuronal Mechanisms of Visual Attention. *Annu Rev Vis Sci* vol. 1 373–391 (2015).
68. Carrasco, M. Visual attention: the past 25 years. *Vision Res* vol. 51 1484–525 (2011).
69. Reynolds, J. H. & Chelazzi, L. Attentional modulation of visual processing. *Annu. Rev. Neurosci.* **27**, 611–47 (2004).
70. Cohen, M. R. & Maunsell, J. Neuronal mechanisms of spatial attention in visual cerebral cortex. in *The Oxford Handbook of Attention* (eds. Nobre, K. & Kastner, S.) (Oxford University Press, Oxford, 2014).
71. Treue, S. Neural correlates of attention in primate visual cortex. *Trends Neurosci* vol. 24 295–300 (2001).
72. Rust, N. C. & Cohen, M. R. Priority coding in the visual system. *Nat. Rev. Neurosci.* (2022) doi:10.1038/s41583-022-00582-9.
73. Gilbert, C. D. & Sigman, M. Brain states: top-down influences in sensory processing. *Neuron* **54**, 677–96 (2007).
74. Sagi, D. Perceptual learning in Vision Research. *Vision Res.* **51**, 1552–1566 (2011).
75. Watanabe, T. & Sasaki, Y. Perceptual learning: toward a comprehensive theory. *Annu. Rev. Psychol.* **66**, 197–221 (2015).

76. Kastner, S. & Ungerleider, L. G. Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* **23**, 315–341 (2000).
77. Boudreau, C. E., Williford, T. H. & Maunsell, J. H. R. Effects of task difficulty and target likelihood in area V4 of macaque monkeys. *J. Neurophysiol.* **96**, 2377–87 (2006).
78. Cowley, B. R. *et al.* Slow Drift of Neural Activity as a Signature of Impulsivity in Macaque Visual and Prefrontal Cortex. *Neuron* **108**, 551–567.e8 (2020).
79. Moore, T., Tolias, A. S. & Schiller, P. H. Visual representations during saccadic eye movements. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8981–8984 (1998).
80. Tolias, A. S. *et al.* Eye movements modulate visual receptive fields of V4 neurons. *Neuron* **29**, 757–767 (2001).
81. Motter, B. C. Saccadic momentum and attentive control in V4 neurons during visual search. *J. Vis.* **18**, 16 (2018).
82. Moore, T. & Chang, M. H. Presaccadic discrimination of receptive field stimuli by area V4 neurons. *Vision Res.* **49**, 1227–1232 (2009).
83. Marino, A. C. & Mazer, J. A. Saccades Trigger Predictive Updating of Attentional Topography in Area V4. *Neuron* **98**, 429–438.e4 (2018).
84. Burrows, B. E., Zirnsak, M., Akhlaghpour, H., Wang, M. & Moore, T. Global selection of saccadic target features by neurons in area v4. *J. Neurosci. Off. J. Soc. Neurosci.* **34**, 6700–6706 (2014).
85. Steinmetz, N. A. & Moore, T. Eye movement preparation modulates neuronal responses in area V4 when dissociated from attentional demands. *Neuron* **83**, 496–506 (2014).
86. Steinmetz, N. A. & Moore, T. Changes in the response rate and response variability of area V4 neurons during the preparation of saccadic eye movements. *J. Neurophysiol.* **103**, 1171–1178 (2010).
87. Popovkina, D. V. & Pasupathy, A. Task Context Modulates Feature-Selective Responses in Area V4. *J. Neurosci. Off. J. Soc. Neurosci.* **42**, 6408–6423 (2022).
88. Pasupathy, A., Popovkina, D. V. & Kim, T. Visual Functions of Primate Area V4. *Annu Rev Vis Sci* **6**, 363–385 (2020).
89. Angelucci, A. & Bressloff, P. C. Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. in *Progress in Brain Research* (eds. Martinez-Conde, S., Macknik, S. L., Martinez, L. M., Alonso, J. M. & Tse, P. U.) vol. 154 93–120 (Elsevier, 2006).
90. Angelucci, A. *et al.* Circuits and Mechanisms for Surround Modulation in Visual Cortex. *Annu. Rev. Neurosci.* **40**, 425–451 (2017).
91. Motter, B. C. Stimulus conflation and tuning selectivity in V4 neurons: a model of visual crowding. *J. Vis.* **18**, 15 (2018).
92. Trott, A. R. & Born, R. T. Input-gain control produces feature-specific surround suppression. *J. Neurosci. Off. J. Soc. Neurosci.* **35**, 4973–4982 (2015).
93. Henry, C. A. & Kohn, A. Feature representation under crowding in macaque V1 and V4 neuronal populations. *Curr. Biol. CB* **32**, 5126–5137.e3 (2022).
94. Namima, T., Yasuda, M., Banno, T., Okazawa, G. & Komatsu, H. Effects of luminance contrast on the color selectivity of neurons in the macaque area v4 and inferior temporal cortex. *J. Neurosci. Off. J. Soc. Neurosci.* **34**, 14934–14947 (2014).
95. Sani, I., Santandrea, E., Golzar, A., Morrone, M. C. & Chelazzi, L. Selective tuning for contrast in macaque area V4. *J Neurosci* **33**, 18583–96 (2013).
96. Wang, X., Nandy, A. S. & Jadi, M. P. Laminar compartmentalization of attention modulation in area V4 aligns with the demands of visual processing hierarchy in the cortex. *Sci. Rep.* **13**, 19558 (2023).
97. Sanayei, M. *et al.* Perceptual learning of fine contrast discrimination changes neuronal tuning and population coding in macaque V4. *Nat. Commun.* **9**, 4238 (2018).
98. Sanada, T. M., Namima, T. & Komatsu, H. Comparison of the color selectivity of macaque V4 neurons in different color spaces. *J. Neurophysiol.* **116**, 2163–2172 (2016).

99. McAdams, C. J. & Maunsell, J. H. R. Effects of Attention on Orientation-Tuning Functions of Single Neurons in Macaque Cortical Area V4. *J. Neurosci.* **19**, 431–441 (1999).
100. Martinez-Trujillo, J. C. & Treue, S. Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol* vol. 14 744–51 (2004).
101. Ruff, D. A., Ni, A. M. & Cohen, M. R. Cognition as a Window into Neuronal Population Space. *Annu Rev Neurosci* vol. 41 77–97 (2018).
102. Brysbaert, M. How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. *J Cogn* **2**, 16 (2019).
103. Mastrogiuseppe, F. & Ostojic, S. Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks. *Neuron* **99**, 609-623.e29 (2018).
104. Landau, I. D. & Sompolinsky, H. Coherent chaos in a recurrent neural network with structured connectivity. *PLoS Comput. Biol.* **14**, e1006309 (2018).
105. Zhang, W.-H., Wu, S., Josić, K. & Doiron, B. Sampling-based Bayesian inference in recurrent circuits of stochastic spiking neurons. *Nat. Commun.* **14**, 7074 (2023).
106. de la Rocha, J., Doiron, B., Shea-Brown, E., Josić, K. & Reyes, A. Correlation between neural spike trains increases with firing rate. *Nature* **448**, 802–6 (2007).
107. Xue, C., Markman, S. K., Chen, R., Kramer, L. E. & Cohen, M. R. Task interference as a neuronal basis for the cost of cognitive flexibility. *BioRxiv Prepr. Serv. Biol.* 2024.03.04.583375 (2024)
doi:10.1101/2024.03.04.583375.