## NEUROSCIENCE

# Task learning increases information redundancy of neural responses in macaque visual cortex

Shizhao Liu, Anton Pletenev, Ralf M. Haefner*†, Adam C. Snyder†

**Full article and list of author affiliations:** https://doi.org/10.1126/science.adw7707

**INTRODUCTION:** How does the brain transform sensory input into perception and behavior? The classic model guiding most of neuroscience and modern deep learning views perception as a largely feedforward process: Sensory signals are transformed from early to higher visual areas to make behaviorally relevant information more explicit. Feedback connections are thought to merely fine-tune this process—enhancing relevant features or suppressing noise during attention and learning.

An alternative framework, generative inference, posits that sensory processing is fundamentally bidirectional. In this view, neurons represent beliefs about causes in the external world, continuously updated by the exchange of information between sensory evidence (feedforward) and prior expectations (feedback).

**RATIONALE:** A recent theoretical prediction from the generative inference framework offers a decisive way to empirically distinguish these two models. Generative inference models predict an increased sharing of task-related information among sensory neurons while learning a perceptual decision-making task—manifesting as higher redundancy in their responses. This prediction directly opposes the classic model, which holds that learning and attention reduce redundancy and correlated variability to improve coding efficiency.

To test these conflicting predictions, we measured changes in information redundancy among neurons in visual area V4 of two macaque monkeys as they learned to discriminate between two orientations in two separate tasks (cardinal and oblique). Neural activity was recorded chronically using Utah arrays over weeks of training. We quantified information redundancy as the difference between the linear Fisher information carried by the intact population activity and that carried by the same population after removing correlations.

**RESULTS:** At the start of learning, redundancy was near zero, indicating largely independent neural responses. Over the course of training, redundancy increased, ultimately reaching levels where roughly half of each neuron's information was shared with other recorded neurons. Redundancy also increased dynamically within trials, over hundreds of milliseconds, consistent with the gradual accumulation and sharing of information. Increased redundancy did not result in a loss of information in the population but, instead, the individual-neuron information increased—both predicted by generative inference. Learning-related changes in redundancy were stronger during task performance compared with passive viewing on the same day, which suggests that the increase in redundancy owing to a redistribution of information depends on active task engagement.

**CONCLUSION:** Learning a perceptual task increased information redundancy among sensory neurons—a result that contradicts conventional understanding of the roles of learning and attention. Rather than eliminating correlated variability, learning appears to redistribute information across neurons through feedback and recurrent interactions, enabling consistent beliefs about the sensory world. These findings suggest that cortical sensory processing is best understood as a dynamic inference process—one that integrates prior expectations and sensory evidence—challenging the long-held assumption of a fundamentally unidirectional information flow during sensory processing in the brain. □
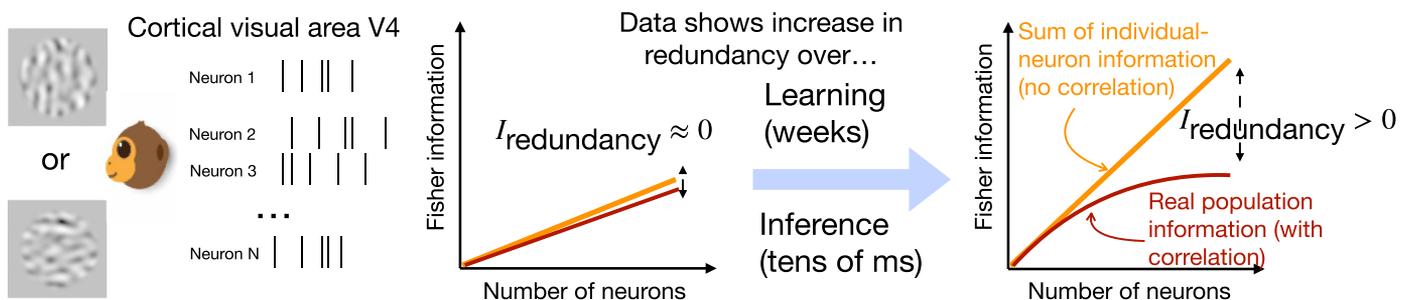
**Redundancy and information-limiting correlations among visual neurons increased during learning discrimination tasks.**

## NEUROSCIENCE

# Task learning increases information redundancy of neural responses in macaque visual cortex

Shizhao Liu[1,2], Anton Pletenev[1,2], Ralf M. Haefner[1,2,3,4]*†, Adam C. Snyder[1,2,5]†

How does the brain optimize sensory information for decision-making in new tasks? One hypothesis suggests that learning reduces redundancy in neural representations to improve efficiency, whereas another, based on Bayesian inference, predicts that learning increases redundancy by distributing information across neurons. We tested these hypotheses by tracking population responses in macaque cortical area V4 as monkeys learned visual discrimination tasks. We found strong support for the Bayesian predictions: Task learning increased redundancy in neural responses over weeks of training and within single trials. This redundancy did not reduce information but instead increased the information carried by individual neurons. These insights suggest that sensory processing in the brain reflects a generative rather than discriminative inference process.

Two fundamentally different perspectives on visual processing guide current models of biological and machine vision. In one perspective, information on the retina is transformed from a pixel basis to one that makes behaviorally relevant information readily decodable by downstream decision circuits (1–3). Within this classic perspective, learning and attention are viewed as mechanisms that optimize representations for downstream decoding accuracy by reducing redundancy, including information-limiting correlations, to improve behavioral performance (4–11). In the other perspective, visual processing is assumed to invert an internal generative model of the relationship between possible causes in the world and visual observations [referred to as "unconscious inference" (12) and "analysis by synthesis" (13)]. We refer to this perspective as generative inference (14).

A key difference between these two frameworks lies in the role of signals carried by feedback connections. Although known to be ubiquitous in the brain from anatomical studies (15), the role of feedback connections remains a matter of debate. In classic models of perception, feedback signals serve to improve intermediate representations. On short timescales, in the form of attention, representational improvements supported by feedback connections include sharpening tuning to behaviorally relevant stimulus features and suppressing harmful noise (7, 8, 11, 16–21), both reducing redundancy. On long timescales, such as in learning, representational improvements are also assumed to reduce redundancy (4–6, 11). In these models, redundancy reflects inefficiency in the neural representation that decreases the information in the entire population, implicitly assuming that the information each neuron can carry is fixed, limited by, for example, metabolic costs.

In generative inference models, feedback signals communicate probabilistic prior beliefs between different areas (14, 22–24). The key feature of generative inference is that neurons in the sensory cortex represent so-called posterior beliefs that integrate these prior beliefs with incoming sensory evidence. Thus, prior beliefs share information among different neurons (25) and have been predicted to increase information redundancy, as reflected in correlated variability, over the course of learning a psychophysical task (26). This prediction of an increase in information redundancy over the course of learning contradicts the classic perspective and is, therefore, ideal for distinguishing between these two influential frameworks.

### Differential predictions of feedforward and generative inference

When performing a visual discrimination task, the observer's choices are based on the activity of their sensory neurons (1). The information about stimulus identity carried by populations of neurons is limited by their response variability and the correlation structure present in that variability (Fig. 1A). If variability is uncorrelated across neurons, then information grows linearly with the number of neurons but grows only sublinearly if the correlation structure is information limiting (9)—i.e., information is redundant (Fig. 1B) (27). We quantify the degree of redundancy by computing the difference between the actual linear Fisher information in the population and the linear Fisher information after destroying all correlations by shuffling trials: $I_{redundancy} = I_{shuffle} - I_{real}$. We asked how does sensory representation change over the course of learning—does it become more or less redundant?

The classic perspective predicts redundancy to decrease over learning, either as the result of rewiring the feedforward pathway to reduce redundancies in input introduced by the task stimulus (4–6) or as the result of attentional processes that learn to suppress task-relevant correlated variability, which is considered noise in the classic framework (7, 8, 11, 17, 19). By contrast, prior work has shown that generative inference predicts an increase in redundancy in neural responses as the result of learning a task, owing to correlated variability induced by task-specific feedback from decision-making areas (Fig. 1C) (23, 26).

### Animals learned two tasks over weeks

We trained two rhesus macaque monkeys on two coarse-orientation discrimination tasks each: The cardinal task was to discriminate between 0° and 90° orientation, and the oblique task required the animals to discriminate between 45° and 135° (Fig. 2A). We used dynamic noise stimuli whose orientation energy was varied to manipulate signal strength (28–30) (Fig. 2B and materials and methods). Stimulus duration was 1.6 s, after which the monkeys reported a choice by making a saccade to one of two targets (Fig. 2, A and C). We call each learning phase an epoch, for a total of four epochs across two monkeys (two tasks by two monkeys).

After training, both monkeys successfully performed both tasks (Fig. 2D). Decision strategies at the end of three of the four epochs approximated those of the ideal observer (Fig. 2E); monkey G's strategy for the oblique task was more consistent with performing a detection rather than a discrimination task (its responses were mostly explained by the presence or absence of orientation power at 135°).

To track the animal's state of learning over time, we defined the learning index—the product of (i) similarity of empirical and ideal observer strategy and (ii) how well animal choice was explained by the stimulus (materials and methods). Learning index tightly correlated with behavioral accuracy at individual coherence levels (fig. S1) and allowed us to track learning even as coherence levels changed with idiosyncratic variability owing to interruptions in training (Fig. 2F).

### Information redundancy increased with task learning

We recorded V4 population activity with 96-channel Utah arrays. Each session yielded on average $34.2 \pm 9.8$ units for monkey R ($23.4 \pm 7.8$ single units, $10.8 \pm 4.9$ multiunits) and $82.8 \pm 16.5$ units for monkey G ($33.3 \pm 12.0$ single units, $49.5 \pm 13.6$ multiunits; materials and methods). The peristimulus spike time histograms (PSTHs) showed significant stimulus tuning signal (fig. S2), and about 40% of individual units

[1]Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA. [2]Center for Visual Science, University of Rochester, Rochester, NY, USA. [3]Department of Computer Science, University of Rochester, Rochester, NY, USA. [4]Department of Physics and Astronomy, University of Rochester, Rochester, NY, USA. [5]Department of Neuroscience, University of Rochester, Rochester, NY, USA. *Corresponding author. Email: ralf.haefner@rochester.edu †These authors contributed equally to this work.

**Fig. 1. Illustration of experiment and key predictions.** (**A**) Response of two example neurons to a horizontal and a vertical stimulus of intermediate signal strength (coherence). Each neuron's response is variable (middle), and this variability is correlated. In this example, positive correlations ($\rho = 0.32$, Pearson correlation) align with the task-tuning axis (both neurons prefer 0° stimulus), reducing the decoding performance (right). (**B**) Linear Fisher information as a function of population size for two example sessions: during the task early in learning, during the task late in learning, and during passive viewing late in learning. $I_{\text{real}}$ (orange) denotes actual linear Fisher information, and $I_{\text{shuffle}}$ (yellow) represents the linear Fisher information after shuffling trials, thereby destroying noise correlations. (**C**) Differential predictions for changes in redundancy over the course of task learning.
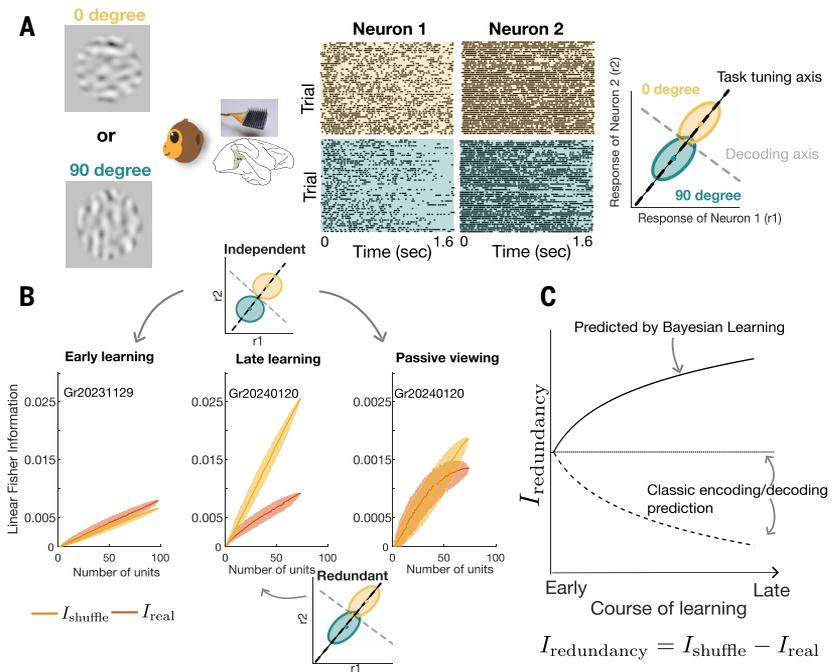


**Fig. 2. Experimental design and behavior.** (**A**) (Top) Task design. After animals fixated on the center to start a trial, choice targets were shown, followed by the stimulus. After stimulus offset, animals indicated their choice by a saccade to one target. (Bottom) Study timeline. (**B**) Example stimulus images for four orientations and two coherence levels. (**C**) Spatial layout of the fixation point (blue dot), stimulus (dashed circle), and choice targets (black and white dots). The red numbers (not shown during the experiment) indicate the stimulus orientation corresponding to each choice target. (**D**) Psychometric curves of late-learning sessions. (**E**) Orientation kernels of late-learning sessions. Dashed lines indicate ideal observer predictions, solid lines indicate across-sessions averages, and shaded areas indicate the standard error of the mean across sessions. The bars on the right indicate the bias averaged across sessions, and the error bars indicate the standard error. (The temporal kernels are shown in fig. S21.) (**F**) Time course of the learning index for two animals and two tasks. Solid lines and shading represent medians and 68% confidence intervals across 1000 bootstrap samples, respectively. Gray vertical lines separate each epoch into early learning and late learning. (Time course with real experiment dates can be found in fig. S22.)
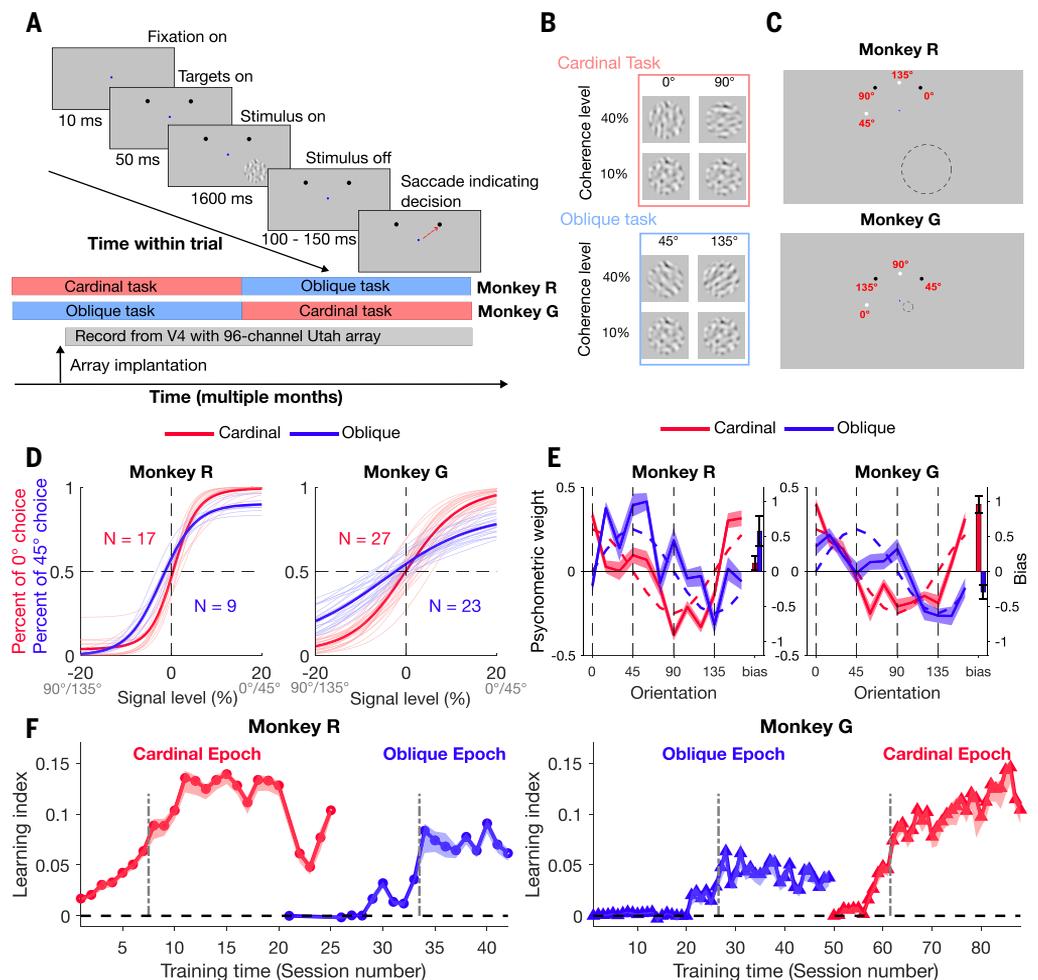
were significantly tuned to task-relevant orientations in both early- and late-learning sessions (fig. S3, B and C). The main text shows results combining both single and multiunits. Analyses based on only single units, or based on only multiunits, showed consistent agreement (fig. S4).

We quantified redundancy, $I_{redundancy}$, as the difference between linear Fisher information about the task-relevant stimulus carried by the neurons without correlations ($I_{shuffle}$) and with correlations (actual population response, $I_{real}$). To track linear Fisher information across many sessions with variable sample sizes of neurons, we randomly subsampled a constant number of units for each session within each epoch (monkey R, cardinal: 17 units; monkey R, oblique: 18 units; monkey G, oblique: 44 units; monkey G, cardinal: 55 units). See materials and methods and results for full population in fig. S5.

$I_{redundancy}$ was close to zero at the beginning of training for each of the four epochs, in agreement with prior empirical (31–33) and theoretical work (9, 34). $I_{redundancy}$ increased with task training as expected from hierarchical Bayesian inference (23, 26) and was significantly smaller for early-learning sessions compared with late-learning sessions for each of the four epochs (Fig. 3C, "Task") [unpaired $t$ test: monkey R, cardinal: $t(21) = 4.04$, $P = 5.9 \times 10^{-4}$; monkey R, oblique: $t(12) = 5.39$, $P = 1.6 \times 10^{-4}$; monkey G, oblique: $t(46) = 5.25$, $P = 3.8 \times$

$10^{-6}$; monkey G, cardinal $t(36) = 3.06$, $P = 4.1 \times 10^{-3}$]. Furthermore, $I_{redundancy}$ significantly correlated with learning index such that better performance was associated with stronger, not weaker, information-limiting correlations (Fig. 3B) [Spearman rank correlation: monkey R, cardinal: $\rho(21) = 0.74$, $P = 9.3 \times 10^{-5}$; monkey R, oblique: $\rho(12) = 0.70$, $P = 0.0069$; monkey G, oblique: $\rho(46) = 0.61$, $P = 6.7 \times 10^{-6}$; monkey G, cardinal: $\rho(36) = 0.65$, $P = 1.6 \times 10^{-5}$]. Accordingly, for each epoch, information saturated at smaller population sizes late in learning compared with early in learning (fig. S6). By contrast, average noise correlations did not change systematically across four epochs (fig. S7, A and C).

To investigate whether the relationship between information redundancy and learning differed between more and less task-informative neurons, we computed $I_{redundancy}$ separately by neural sensitivity ($d'$) by dividing our population into two halves: the 50% of neurons with higher $d'$ (more task-informative) versus the 50% with lower $d'$ (less task-informative). The relationship was stronger for more task-informative neurons (fig. S8A and table S1), which had higher $I_{redundancy}$ late, but not early, in learning (fig. S8B and table S2).

The increase of $I_{redundancy}$ was related to active task engagement: When we computed $I_{redundancy}$ from passive-viewing trials collected at the end of each session, we did not see a consistent increase with task performance (Fig. 3C, "Passive," and fig. S9). Also, in the late-learning stage, $I_{redundancy}$ during task performing was significantly greater than during passive viewing in three of four epochs (Fig. 3C) [paired $t$ test: monkey R, cardinal: $t(12) = 6.99$, $P = 1.45 \times 10^{-5}$; monkey R, oblique: $t(8) = 1.41$, $P = 0.195$; monkey G, oblique: $t(22) = 3.94$, $P = 6.97 \times 10^{-4}$; monkey G, cardinal: $t(26) = 5.21$, $P = 1.92 \times 10^{-5}$]. This was not the case in the early-learning stage (Fig. 3C) [paired $t$ test: monkey R, cardinal: no passive-viewing data collected in early learning; monkey R, oblique: $t(4) = -1.32$, $P = 0.256$; monkey G, oblique: $t(24) = -1.05$, $P = 0.303$; monkey G, cardinal: $t(10) = 2.01$, $P = 0.0718$]. Because passive-viewing data and task-performing data contained exactly the same population of neurons and were recorded on the same day, these results provide strong evidence that the increase of $I_{redundancy}$ was flexible and linked to active task performing and not the result of long-term changes in the feedforward pathway.

We ensured that the relationship between $I_{redundancy}$ and behavioral performance was not confounded by changes in coherence levels across training by showing that the relationship held for many individual fixed coherence levels (fig. S10). This relationship was robust after controlling for eye movements (materials and methods, figs. S11 to S13, and tables S3 to S6).

## Information redundancy results from hierarchical Bayesian inference

The empirical finding that $I_{redundancy}$ increased with task learning is unexpected in light of classic models that cast learning as increasing representational efficiency (5, 6) or the role of attention as suppressing information-limiting covariability (7, 8, 10), both decreasing redundancy. However, if neurons represent posterior beliefs in a hierarchical model, learning a typical perceptual decision-making task will increase their redundancy because neurons share information with
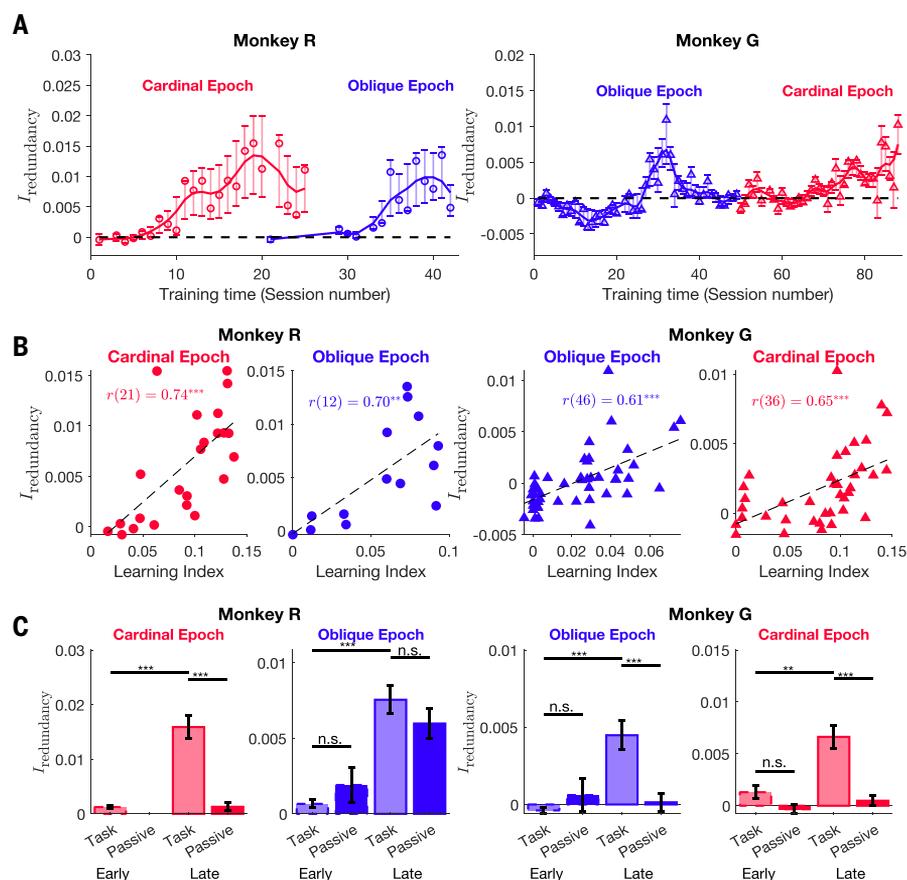


**Fig. 3. Redundancy increases over the course of learning.** (A) Time course of $I_{redundancy}$ during the course of learning using subsampled populations with consistent number of units (monkey R: cardinal $n = 17$, oblique $n = 18$; monkey G: oblique $n = 44$, cardinal $n = 55$). Lines and error bars are means and 68% confidence intervals across bootstrapped samples, respectively. (B) Scatter plots between the learning index and $I_{redundancy}$ for two tasks and two animals. Each dot represents a session. Black dashed lines are linear regression fits across sessions. (C) We separated sessions into two groups, early and late, using the boundary shown in Fig. 2F. The bar plots averaged $I_{redundancy}$ across sessions in each learning stage when the animals were performing the task or passively viewing the stimulus. Error bars indicate the standard error of the mean across sessions. Significance levels: n.s., not significant; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

each other. In a so-called typical perceptual decision-making task, the brain must combine information across many sensory neurons representing different aspects of sensory input and often accumulate information over time (1). Although neurons may represent information that is, on average, independent during natural vision, the task stimuli typically induce dependencies in space and time. For instance, in our case, the orientation in one location of the stimulus was predictive of orientation throughout the stimulus, and the orientation at one moment during the trial was predictive of orientation throughout the trial. Generative inference thus predicts that after learning, but not before, this information is shared between neurons (Fig. 4A). In particular, it predicts that redundancy increase does not reflect a decrease in total information but rather an increase in the marginal information present in each individual neuron—something that we confirmed using a previously published model (23). Simulating three stages of learning (before, during, and after), we found that for subsets of model neurons, the increase in $I_{redundancy}$ was explained by the increase in $I_{shuffle}$ not a decrease in $I_{real}$ (Fig. 4B and materials and methods) (23). Because $I_{shuffle}$ was simply the sum over the marginal information in each neuron, its increase reflected an increase in the average information carried by individual neurons. This information was shared between individual neurons and drove the increase in redundancy in the population. The information in the entire population of 256 neurons in our model, $I_{real}$, decreased as expected from the fact that computation cannot create, but only destroy, information (Fig. 4C). However, for a downstream brain area that reads out this information from a subset of all sensory neurons, this decrease is more than compensated for by the benefits of redistribution (35)—the fewer neurons read out, the larger the benefit (Fig. 4C).

We tested these predicted changes of $I_{real}$ and $I_{shuffle}$ in our empirical data. In agreement with our predictions, both $I_{real}$ and $I_{shuffle}$ increased during the course of task learning, with $I_{shuffle}$ growing faster than $I_{real}$ (Fig. 4D). During all four task epochs, $I_{shuffle}$ significantly correlated with learning index (Fig. 4E) [Spearman rank correlations: monkey R, cardinal: $\rho(21) = 0.74$, $P = 8.68 \times 10^{-5}$; monkey R, oblique: $\rho(12) = 0.75$, $P = 3.25 \times 10^{-3}$; monkey G, oblique: $\rho(46) = 0.34$, $P = 0.0186$; monkey G, cardinal: $\rho(36) = 0.78$, $P = 2.01 \times 10^{-7}$]. $I_{real}$ also significantly correlated with learning index in three of four epochs (Fig. 4E) [monkey R, cardinal: $\rho(21) = 0.68$, $P = 4.66 \times 10^{-4}$; monkey R, oblique: $\rho(12) = 0.73$, $P = 4.13 \times 10^{-3}$; monkey G, cardinal: $\rho(36) = 0.58$, $P = 1.66 \times 10^{-4}$], but no significant relationship was found in the oblique epoch of monkey G [$\rho(46) = -0.15$, $P = 0.30$]. Our empirical data sampled only a small subset of all sensory neurons and therefore should be compared with the results from subsets of neurons in our model (Fig. 4B). After learning, responses in that small subset will also contain information from other, unrecorded, neurons, which is why we expect an increase of $I_{real}$ with learning in the data. Analyzing passive-viewing data recorded during the same sessions did not show a consistent increase in $I_{real}$ and $I_{shuffle}$ with task performance, which suggests that active task engagement facilitates information redistribution (fig. S14).

The increase in $I_{real}$ during training raises the possibility that the observed rise in redundancy with learning does not reflect an increase in redistribution but rather stems directly from the increase in $I_{real}$ itself. To exclude this possibility, we confirmed that the average percentage of information per neuron that is redundant increased over learning (from ~0 to 50%; fig. S15).

To directly compare information in neural responses with behavioral performance, we converted the latter into Fisher information ($I_{behav}$; materials and methods). After learning the cardinal task, $I_{behav}$ was higher than $I_{real}$ (Fig. 4D and fig. S16A), which suggests that animals extracted information from other unobserved neurons to perform the task. For the oblique task, $I_{behav}$ was smaller than or equal to $I_{real}$ (Fig. 4D and fig. S16A), which suggests that the oblique effect that we observed (36, 37) may partially be due to less efficient use of neural information for behavior in oblique tasks. $I_{real}$ and $I_{behav}$ were correlated, except for monkey G's oblique epoch, which is unsurprising because they are both explained by the same information redistribution process [monkey R, cardinal: $\rho(22) = 0.59$, $P = 2.62 \times 10^{-3}$; monkey R, oblique: $\rho(12) = 0.92$, $P = 0.00$; monkey G, cardinal: $\rho(35) = 0.38$, $P = 1.96 \times 10^{-2}$; monkey G, oblique: $\rho(46) = 0.06$, $P = 6.88 \times 10^{-1}$].
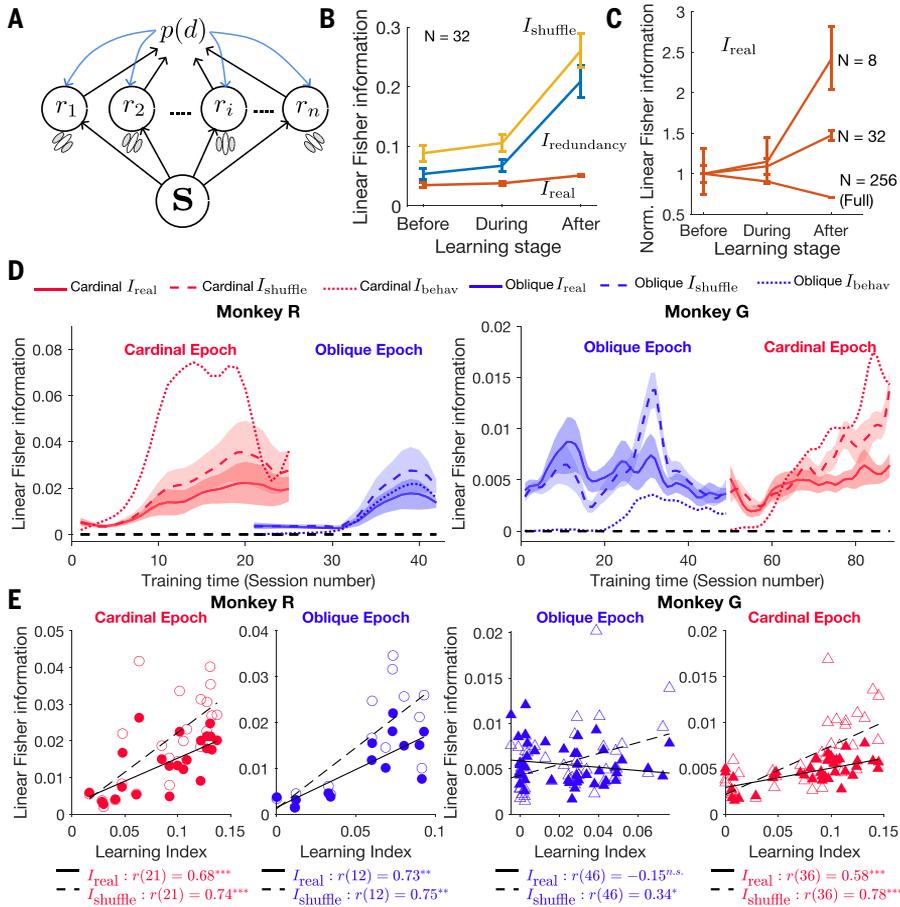


**Fig. 4. Increase in redundancy is explained by increase in $I_{shuffle}$ not a decrease in $I_{real}$.** (**A**) Information sharing during generative inference. The activities of sensory neurons, $r_i$, represent posterior beliefs about the intensity of oriented edges in the input stimulus. A decision-making area computes a belief about a decision variable, $p(d)$. This belief, based on the information in all sensory neurons, informs all sensory neurons through feedback signals (blue arrows). (**B** and **C**) Model (23) simulations at different stages of learning. Error bars indicate 68% confidence intervals across random samples of 32 (B) and 8, 32, and 256 (C) neurons out of a total of 256 model neurons. In (C), for each population size, the values were normalized by the mean $I_{real}$ before learning. (**D**) Time course of $I_{real}$, $I_{shuffle}$, and behavioral Fisher information ($I_{behav}$) during learning in empirical data. Shaded areas indicate 68% confidence intervals across bootstrapped samples. (**E**) Correlation between learning index $I_{real}$ and $I_{shuffle}$. One dot per session. Black lines are a linear regression fit across sessions, solid for $I_{real}$ and dashed for $I_{shuffle}$. Significance levels: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.
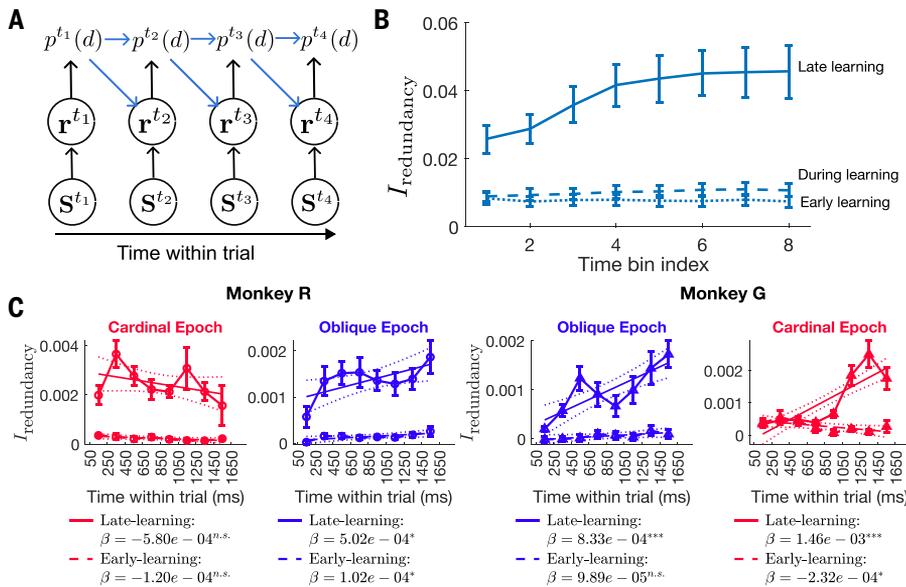
**Fig. 5. Redundancy increases within a trial.** (**A**) Temporal propagation of beliefs during generative inference. $t_1$ to $t_4$ represent times within a single trial. (**B**) Change in $I_{redundancy}$ across eight time bins within a trial for three stages of learning (model). Error bars are 68% confidence intervals across 256 random subsamples (size = 32) of model neurons. (**C**) $I_{redundancy}$ for each 200-ms time bin within a trial in empirical data, controlled for population size. Solid lines indicate averages across late-learning sessions, and dashed lines indicate averages across early-learning sessions. β is the regression coefficient: $I_{redundancy}$ versus time bin. Error bars indicate the standard error of the mean across sessions. Significance levels: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Comparing neurometric and psychometric thresholds instead of Fisher information yielded the same conclusions (fig. S17 and table S7).

## Redundancy increases within a trial

During generative inference, redundancy should not just increase over learning but also within a trial for our stimulus. Because orientation from early video frames predicts orientation later, neurons that represent posterior beliefs will incorporate information from those earlier frames in the form of prior expectations (Fig. 5A) (23, 38). Information in each neuron will thus accumulate over the course of the trial, and because that information is shared among all neurons, redundancy will increase too.

We first confirmed this prediction using synthetic data from the same model as before (Fig. 5A). Late in learning (but not early in learning), $I_{redundancy}$ increased over the course of a simulated trial, split into eight time bins (Fig. 5B). This effect was explained by an increase in information per individual neuron not a decrease in information across the population (fig. S18A).

To test this prediction in our empirical data, we divided each trial into eight nonoverlapping 200-ms-duration time bins. In three of four epochs, redundancy significantly increased over the eight bins within a trial during late-learning sessions (Fig. 5C, solid lines) (linear regression: monkey R, oblique: β = $5.0160 \times 10^{-4}$, $P = 0.021997$; monkey G, oblique: β = $8.3294 \times 10^{-4}$, $P = 1.1432 \times 10^{-5}$; monkey G, cardinal: β = $1.4571 \times 10^{-3}$, $P = 3.5195 \times 10^{-13}$). The slope of this dependency increased between early-learning and late-learning sessions.

The only exception was the monkey R cardinal epoch, in which no increase of $I_{redundancy}$ was observed across the eight time bins (β = $-5.8008 \times 10^{-4}$, $P = 0.17042$ for late-learning sessions). Because $I_{redundancy}$ was already relatively high in the first time bin, we hypothesized that the predicted increase might have happened over a shorter timescale within the first bin. Repeating the analysis with a finer timescale (50 ms) revealed that $I_{redundancy}$ in the monkey R cardinal epoch did consistently increase within the first 450 ms (fig. S19). Despite

general agreement with Bayesian model predictions, the differences between time courses suggest heterogeneity in temporal details of task strategies.

Finally, we confirmed that the within-trial increase of $I_{redundancy}$ was explained by a faster increase of $I_{shuffle}$ rather than a decrease in $I_{real}$ (fig. S18). Unexpectedly, $I_{real}$ in late-learning data increased less than expected, or even decreased, within trials (fig. S18B, top). Examining before-learning empirical data revealed a marked decrease in both $I_{real}$ and $I_{shuffle}$ within a trial (fig. S18B, bottom), presumably owing to adaptation (39), consistent with the empirical PSTH (fig. S2). Our work suggests that the redistribution of information from early to later in the trial may compensate for this adaptation-related reduction in information.

## Discussion

We tested competing predictions for two dominant theoretical frameworks guiding the field's understanding of sensory processing. Predicted by generative inference, but unexpected from the classic perspective, we found that redundancy in sensory responses increased over weeks while monkeys learned two perceptual discrimination tasks. Expressed differently, information-limiting correlations grew as behavioral performance improved. The growth of these correlations did not imply a decrease in overall information in the population in our data but reflected an increase in the marginal information in each neuron as the brain learned to propagate task-related information between neurons. The same process played out over the course of tens to hundreds of milliseconds within a trial: As the monkeys accumulated sensory evidence, the redundancy in sensory responses increased.

Our findings have implications for models of sensory processing and the interpretation of empirical data. Unlike assuming that "all information is on the retina" and sensory processing simply reformats it to "untangle" behaviorally relevant variables (1, 2), our results support the hypothesis that sensory neurons compute posterior beliefs that incorporate all possible information about the features that they represent (22, 23, 26, 38, 40, 41). An important case is prior information, which classically has been assumed to be combined with sensory information at the top level of the sensory processing hierarchy (42–46) but which growing evidence implies is represented at all stages of sensory processing (25, 29, 47–49).

Since Zohary et al. (50) found that information saturates in populations of correlated sensory neurons, many theoretical (9, 51–53) and experimental (7, 8, 11, 19, 21, 54–61) studies have investigated correlations, including their dependence on learning and behavioral state, such as expectation and attention (10). Almost all of these studies have interpreted neural covariability within the classic framework, assuming that its effect on information mediates behavior. This has led to the overarching conclusion that behavioral improvement through learning and attention is mediated by a decrease in harmful correlations (11). Our work demonstrates the opposite—that information-limiting correlations become stronger as behavior improves—and suggests that they are mere by-products of information redistribution that increases information in individual neurons rather than decreasing information in the population.

Although we interpret the redundancy increase as a side effect of computing posteriors during generative inference in a Bayesian framework, our findings are also compatible with suggestions that increased redundancy may facilitate robust and nonspecific decoding (35) and

faster learning (*62–64*). Redundant activity between neurons might also improve readout by increasing temporal consistency between presynaptic inputs (*65*). The dynamic by which hierarchical Bayesian inference increases redundancy is mathematically similar to that described by Chadwick *et al.* (*66*), who studied optimal evidence integration in a recurrent neural network. Whereas in their model, information redistribution results from recurrent connectivity, in our model, it is the result of top-down feedback signals, which better explains the context-dependent nature of this redistribution (*38*, *54*) evident in our data in the difference between passive viewing and active task engagement (Fig. 3C).

Although our interpretation differs from previous works, our data are compatible with them. We found that information-limiting correlations were weak before learning, in agreement with both theoretical considerations (*9*, *34*) and empirical findings: Prior studies that detected them in the absence of a task were based on at least an order of magnitude more neurons than our study (*32*, *33*, *67*). By contrast, results in animals performing a task also found them to be stronger and specific to that task, which suggests that they were the result of performing the task (*68*). They were also shown to change with the task, likely involving feedback signals (*29*). Studies that documented changes over learning generally did not focus on information-limiting correlations directly but typically reported changes in average correlations (*11*, *55*). For instance, Ni *et al.* (*11*) found an overall decrease in the average pairwise noise correlation in the visual cortex over learning a perceptual task.

Information redundancy and information-limiting correlations are closely linked and are not generally related to average noise correlations (*9*). In our study, average noise correlations decreased in one epoch, increased in another epoch (possibly related to recording quality declining over long training time), and did not change in the other two epochs (fig. S7). One previous study, which compared shuffled and real Fisher information (*31*), appears to agree with our findings (*31*); however, the effect remained unquantified and was interpreted by the authors as a floor effect of low information before learning. Our finding that not just absolute but also relative redundancy increased over learning contradicts this interpretation and is further evidence for an increase in redistribution of information after learning.

The information redundancy that we investigated was measured within a single sensory cortical area. Recent investigations of how information is shared among neurons during perceptual decision-making have instead focused on neural populations in the association cortex (*65*) or populations that spanned multiple areas that included mixtures of sensory and decision-making populations (*69*). However, even the classical framework predicts an increase in redundancy with learning across areas because performing the task is impossible without sensory neurons and decision-making neurons sharing information. The divergent predictions of the classical and generative inference frameworks about information redundancy only apply within a single sensory population.

Multiple studies have reported an increase in the information represented by individual neurons (*70–73*). Our results suggest that these changes are not explained by classic representation learning but are more parsimoniously understood by the brain having learned a new prior over sensory inputs in the task context (*26*).

Our results address a long-standing puzzle in systems neuroscience: How can it be that individual neurons contain as much information about the task as the behavior of the entire animal, whose choices are informed by many such neurons (*74–76*)? Our population neurometric thresholds were close to psychometric thresholds (table S7), consistent with previous studies with single units. Yet even our population was a small portion of neurons in the brain that could provide task information. Our results imply that much of the information in each neuron is actually shared with other neurons, including many unobserved ones, as predicted by the generative inference framework after learning (Fig. 4C).

Mechanistically, the observed within-trial increase in redundancy likely depends on recurrent interactions within the sensory cortex and/or top-down feedback from higher-order areas. Feedback projections from the parietal or prefrontal cortex could broadcast accumulated beliefs about the trial category as task-specific priors (*23*). Alternatively, horizontal connections within the cortex might support lateral propagation of information between similarly tuned neurons, amplifying shared content over time. Both mechanisms are supported by anatomical and physiological data (*59*, *77*, *78*). The timescale, task dependence, and differences in the redundancy increase between active task engagement and passive viewing argue against fixed structural sources and instead favor dynamic and state-dependent processing.

Although we found strong and consistent evidence that information redundancy increases with task learning, we also observed variability in the detailed pattern of results across the four task epochs. For example, the generative inference framework in its most straightforward form predicts a monotonic increase in $I_{redundancy}$, $I_{real}$, and $I_{shuffle}$ over learning. Although we indeed found such monotonic increases for three of the four task epochs, for monkey G's oblique task epoch, we found a more complex time course of information changes (Fig. 3A and Fig. 4D). Our Bayesian model for generative inference is a simplified one, and thus it should not be expected to account for all empirical observations. First, it assumes continual learning and monotonic increases in performance and is thus unable to account for the much more idiosyncratic time courses of learning in each of the four epochs. The strong and highly significant correlations between information redundancy and learning index account for this nonmonotonicity in learning performance. Second, although our simplified Bayesian model always, and only, performs our orientation tasks by design, this is clearly not the case for the brain. It is an empirical question whether the task-related generative inference requires active task engagement or whether the brain always unconsciously performs it after learning. Moreover, the model assumes uniform task engagement within a trial and thus predicts a consistent increase of $I_{redundancy}$ within a trial (Fig. 5B). Because our stimulus was quite long (1.6 s), the animals' task engagement likely varied throughout the trial, with the exact timing differing from epoch to epoch or even from session to session. This could account for variability in the within-trial dynamic of $I_{redundancy}$.

Our Bayesian generative model also assumes no misalignment between the learned task and the task defined by the experimenter. It therefore predicts perfect alignment between feedback signals and sensory neurons' feedforward task-relevant information (*26*), whereas our empirical results show that the task-relevant information is distributed more widely. Information redundancy among task-sensitive neurons was higher than that among nonsensitive neurons when sensitivity was computed from task data (fig. S8 and table S1). However, when we separated units by sensitivity computed from passive-viewing data, we observed little difference between groups (fig. S20 and table S1). Simulating task misalignment in the model (*23*), we found statistically indistinguishable redundancy for lower and higher $d'$ populations for a task misalignment of as little as 20°. Because the brain has to learn the task from trial and error, some mismatch is expected and indeed evident in our orientation psychometric kernels (Fig. 2E). Moreover, our synthetic neurons are only tuned to orientation, whereas in more realistic scenarios, V4 neurons are also tuned along many dimensions beyond orientation, so we expect that such populations would be more susceptible to even smaller task misalignment.

Our findings challenge the prevailing view that learning in the sensory cortex universally promotes more efficient, less redundant representations. Instead, learning in our task increased information redundancy in the macaque visual cortex, consistent with the predictions of a generative inference framework in which sensory neurons encode posterior beliefs. These redundancy increases were behaviorally relevant, strongest for task-informative neurons, and evident both across weeks of training and within individual trials—suggesting that

they reflect dynamic inference processes rather than fixed circuit properties. By linking learning-dependent changes in correlated variability to feedback-driven belief propagation, our results call for a reassessment of how sensory representations are shaped by task demands.

## REFERENCES AND NOTES

1. A. J. Parker, W. T. Newsome, *Annu. Rev. Neurosci.* **21**, 227–277 (1998).
2. J. J. DiCarlo, D. D. Cox, *Trends Cogn. Sci.* **11**, 333–341 (2007).
3. N. C. Rust, J. J. DiCarlo, *J. Neurosci.* **30**, 12978–12995 (2010).
4. F. Attneave, *Psychol. Rev.* **61**, 183–193 (1954).
5. H. B. Barlow, in *Sensory Communication*, W. A. Rosenblith, Ed. (1961), pp. 217–233.
6. E. P. Simoncelli, B. A. Olshausen, *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).
7. J. F. Mitchell, K. A. Sundberg, J. H. Reynolds, *Neuron* **63**, 879–888 (2009).
8. M. R. Cohen, J. H. Maunsell, *Nat. Neurosci.* **12**, 1594–1600 (2009).
9. R. Moreno-Bote *et al.*, *Nat. Neurosci.* **17**, 1410–1417 (2014).
10. A. Kohn, R. Coen-Cagli, I. Kanitscheider, A. Pouget, *Annu. Rev. Neurosci.* **39**, 237–256 (2016).
11. A. M. Ni, D. A. Ruff, J. J. Alberts, J. Symmonds, M. R. Cohen, *Science* **359**, 463–465 (2018).
12. H. von Helmholtz, *Handbuch der Physiologischen Optik* (Leopold Voss, 1867).
13. A. Yuille, D. Kersten, *Trends Cogn. Sci.* **10**, 301–308 (2006).
14. B. Peters *et al.*, arXiv:2401.06005 [q-bio.NC] (2024).
15. D. J. Felleman, D. C. Van Essen, *Cereb. Cortex* **1**, 1–47 (1991).
16. T. Moore, M. Zirnsak, *Annu. Rev. Psychol.* **68**, 47–72 (2017).
17. M. R. Cohen, J. H. Maunsell, *Neuron* **70**, 1192–1204 (2011).
18. D. A. Ruff, M. R. Cohen, *Nat. Neurosci.* **17**, 1591–1597 (2014).
19. A. C. Snyder, M. J. Morais, M. A. Smith, *J. Neurophysiol.* **116**, 1807–1820 (2016).
20. A. S. Nandy, J. J. Nassi, J. H. Reynolds, *Neuron* **93**, 235–246 (2017).
21. G. H. Denfield, A. S. Ecker, T. J. Shinn, M. Bethge, A. S. Tolias, *Nat. Commun.* **9**, 2654 (2018).
22. T. S. Lee, D. Mumford, *J. Opt. Soc. Am. A* **20**, 1434–1448 (2003).
23. R. M. Haefner, P. Berkes, J. Fiser, *Neuron* **90**, 649–660 (2016).
24. X. Pitkow, D. E. Angelaki, *Neuron* **94**, 943–953 (2017).
25. C. Findling *et al.*, *Nature* **645**, 192–200 (2025).
26. R. D. Lange, R. M. Haefner, *PLOS Comput. Biol.* **18**, e1009557 (2022).
27. B. B. Averbeck, P. E. Latham, A. Pouget, *Nat. Rev. Neurosci.* **7**, 358–366 (2006).
28. H. Nienborg, B. G. Cumming, *J. Neurosci.* **34**, 3579–3585 (2014).
29. A. G. Bondy, R. M. Haefner, B. G. Cumming, *Nat. Neurosci.* **21**, 598–606 (2018).
30. R. D. Lange *et al.*, *J. Neurophysiol.* **129**, 1021–1044 (2023).
31. M. Sanayei *et al.*, *Nat. Commun.* **9**, 4238 (2018).
32. O. I. Rumyantsev *et al.*, *Nature* **580**, 100–105 (2020).
33. M. Kafashan *et al.*, *Nat. Commun.* **12**, 473 (2021).
34. I. Kanitscheider, R. Coen-Cagli, A. Pouget, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6973–E6982 (2015).
35. X. Pitkow, S. Liu, D. E. Angelaki, G. C. DeAngelis, A. Pouget, *Neuron* **87**, 411–423 (2015).
36. E. Mach, *Sitzungsber. Kais. Akad. Wiss.* **43**, 215–224 (1861).
37. S. Appelle, *Psychol. Bull.* **78**, 266–278 (1972).
38. R. D. Lange, A. Chattoraj, J. M. Beck, J. L. Yates, R. M. Haefner, *PLOS Comput. Biol.* **17**, e1009517 (2021).
39. A. Kohn, *J. Neurophysiol.* **97**, 3155–3164 (2007).
40. D. H. Ackley, G. E. Hinton, T. J. Sejnowski, *Cogn. Sci.* **9**, 147–169 (1985).
41. J. Fiser, P. Berkes, G. Orbán, M. Lengyel, *Trends Cogn. Sci.* **14**, 119–130 (2010).
42. J. I. Gold, M. N. Shadlen, *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
43. T. D. Hanks, M. E. Mazurek, R. Kiani, E. Hopp, M. N. Shadlen, *J. Neurosci.* **31**, 6339–6352 (2011).
44. Y. Niv, *Nat. Neurosci.* **22**, 1544–1553 (2019).
45. V. Rao, G. C. DeAngelis, L. H. Snyder, *J. Neurosci.* **32**, 10063–10074 (2012).
46. R. Nogueira *et al.*, *Nat. Commun.* **8**, 14823 (2017).
47. P. Berkes, G. Orbán, M. Lengyel, J. Fiser, *Science* **331**, 83–87 (2011).
48. R. Jardri, S. Duverne, A. S. Litvinova, S. Denève, *Nat. Commun.* **8**, 14218 (2017).
49. J. Park, S. Kim, H. R. Kim, J. Lee, *Sci. Adv.* **9**, eadg4156 (2023).
50. E. Zohary, M. N. Shadlen, W. T. Newsome, *Nature* **370**, 140–143 (1994).
51. L. F. Abbott, P. Dayan, *Neural Comput.* **11**, 91–101 (1999).
52. H. Sompolinsky, H. Yoon, K. Kang, M. Shamir, *Phys. Rev. E* **64**, 051904 (2001).
53. A. S. Ecker, P. Berens, A. S. Tolias, M. Bethge, *J. Neurosci.* **31**, 14272–14283 (2011).
54. M. R. Cohen, W. T. Newsome, *Neuron* **60**, 162–173 (2008).
55. Y. Gu *et al.*, *Neuron* **71**, 750–761 (2011).
56. T. Z. Luo, J. H. Maunsell, *Neuron* **86**, 1182–1188 (2015).
57. A. S. Ecker *et al.*, *Science* **327**, 584–587 (2010).
58. A. S. Ecker *et al.*, *Neuron* **82**, 235–248 (2014).
59. A. C. Snyder, M. J. Morais, A. Kohn, M. A. Smith, *J. Neurosci.* **34**, 11222–11227 (2014).
60. A. C. Snyder, M. J. Morais, C. M. Willis, M. A. Smith, *Nat. Neurosci.* **18**, 736–743 (2015).
61. A. Umakantha *et al.*, *Neuron* **109**, 2740–2754.e12 (2021).
62. T. A. Engel, W. Chaisangmongkon, D. J. Freedman, X.-J. Wang, *Nat. Commun.* **6**, 6454 (2015).
63. M. R. Nassar, D. Scott, A. Bhandari, *J. Neurosci.* **41**, 6740–6752 (2021).
64. C. Haimerl, D. A. Ruff, M. R. Cohen, C. Savin, E. P. Simoncelli, *Nat. Commun.* **14**, 7879 (2023).
65. M. Valente *et al.*, *Nat. Neurosci.* **24**, 975–986 (2021).
66. A. Chadwick *et al.*, *Neuron* **111**, 106–120.e10 (2023).
67. R. Bartolo, R. C. Saunders, A. R. Mitz, B. B. Averbeck, *J. Neurosci.* **40**, 1668–1678 (2020).
68. N. C. Rabinowitz, R. L. Goris, M. Cohen, E. P. Simoncelli, *eLife* **4**, e08998 (2015).
69. S. Ebrahimi *et al.*, *Nature* **605**, 713–721 (2022).
70. A. Schoups, R. Vogels, N. Qian, G. Orban, *Nature* **412**, 549–553 (2001).
71. H. Z. Adab, R. Vogels, *Curr. Biol.* **21**, 1661–1666 (2011).
72. Y. Yan *et al.*, *Nat. Neurosci.* **17**, 1380–1387 (2014).
73. A. G. Khan *et al.*, *Nat. Neurosci.* **21**, 851–859 (2018).
74. K. H. Britten, M. N. Shadlen, W. T. Newsome, J. A. Movshon, *J. Neurosci.* **12**, 4745–4765 (1992).
75. S. J. Prince, A. D. Pointon, B. G. Cumming, A. J. Parker, *J. Neurosci.* **20**, 3387–3400 (2000).
76. S. R. Allred, B. Jagadeesh, *J. Neurophysiol.* **98**, 1263–1277 (2007).
77. N. T. Markov *et al.*, *J. Comp. Neurol.* **522**, 225–259 (2014).
78. D. D. Stettler, A. Das, J. Bennett, C. D. Gilbert, *Neuron* **36**, 739–750 (2002).
79. S. Liu, A. Pletenev, R. M. Haefner, A. Snyder, Chronic Utah array recordings from macaque V4 during learning of orientation discrimination tasks, Zenodo (2025).

# Science

**AAAS**

## Supplementary Materials for

### Task learning increases information redundancy of neural responses in macaque visual cortex

Shizhao Liu *et al*.

Corresponding author: Ralf M. Haefner, ralf.haefner@rochester.edu

**The PDF file includes:**

Materials and Methods
Supplementary Text
Figs. S1 to S22
Tables S1 to S7
References

**Other Supplementary Material for this manuscript includes the following:**

MDAR Reproducibility Checklist

## Materials and Methods

We follow the ARRIVE guidelines and a checklist is provided in the supplementary materials.

### Experiment procedures

**Ethical oversight**   Experimental procedures were approved by the University Committee on Animal Resources of the University of Rochester (protocol number: 102016/ 2018-001) and were performed in accordance with the United States National Research Council's *Guide for the Care and Use of Laboratory Animals* (*80*).

**Subjects**   We used two (the minimal number for replication) male adult rhesus macaques (*Macaca mulatta*) for this study. Rhesus macaques possess visual cognitive abilities and a hierarchical visual cortical organization comparable to those of humans, making them an appropriate model for our research question and enhancing the translational relevance of this study (*81*). Surgeries were performed under isoflurane anesthesia using aseptic technique and perioperative opiate analgesics. To immobilize the head during experiments, a titanium head post was attached to the skull by titanium screws for monkey R and by bone cement and ceramic screws for monkey G. After the animals were trained to consistently perform one of the orientation discrimination tasks (the cardinal task for monkey R and the oblique for monkey G), we implanted a 96-channel Utah array in V4 on each of their left hemispheres. More details about array implantation surgery can be found in Supplementary Text.

**Microelectrode array recording and spikesorting**   The Utah arrays were chronically implanted, and the exact sample size of neurons was a random surgical outcome. Based on our experience with multielectrode array recordings (*19, 81–83*), we anticipated recording from 30–60 neurons from each animal, which has been shown to characterize substantial population activity structure (*84, 85*). Signals from the Utah arrays were band-pass filtered (250 - 7500 Hz), digitized at 30 kHz, and amplified by a Grapevine system (Ripple). Segments of 52 samples (1.73 ms) around threshold-crossing were stored for offline processing. The threshold was set as a 3-4 multiple of the running median estimate of the spike-bandwidth signal. Spike sorting was performed with customized software (*86*). First, an artificial neural network was applied to distinguish action potential waveforms from noise. Then, we manually separated units based on waveform shapes and clusters from principal component analysis. Manual spike sorting was blind to all experimental conditions except for the learning stage. This initial spike sorting procedure yielded $36.7 \pm 10.1$ units each session for monkey R, and $92.1 \pm 18.2$ units for monkey G.

**RF mapping**   On the first recording session after array implantation surgery, we mapped the receptive fields of spiking neurons recorded by the arrays by presenting small sinusoidal gratings (radius = 1.6°of visual angle) of four orientations at a matrix of positions that covered putative receptive fields based on anatomical locations of the implant. We then used stimuli that roughly

covered the aggregated RF area and kept the location and size constant throughout the experiment for each animal. For monkey R, the center of the stimulus was (right 6.4° and down 14.2°) with radius of 6.4°. For monkey G, the stimulus was centered at (right 2.1° and down 0.3°) and had a radius of 1.3°.

**Visual stimuli**    Visual stimuli were presented on a $1920 \times 1080$ - pixel (subtended 47.2° - 31.3° of visual angle) ViewPixx/3D monitor (VPixx Technologies, Saint-Bruno, QC, Canada) at 120 Hz. The animals were located 49 cm away from the monitor. The stimulus was 2-D bandpass-filtered dynamic white noise, as used in previous studies (*28–30*). To generate stimulus, a white noise image was first generated given a noise seed, then filtered in the Fourier domain. Spatial frequency was controlled by the radial coordinate that followed the Rice density, whose mean and standard deviation were determined in a parameter mapping session to maximize task sensitivity of the recorded neuron population.

For monkey R, the mean of spatial frequency was $0.62$ cycles degree$^{-1}$ and the standard deviation was $0.31$ cycles degree$^{-1}$. For monkey G, the mean of spatial frequency was $1.24$ cycles degree$^{-1}$ with standard deviation of $0.62$ cycles degree$^{-1}$. The orientation was controlled by the angular coordinate that followed the von Mises density for which the mean could take one of the orientations to discriminate, i.e. 0° or 90° for the cardinal task, 45° or 135° for the oblique task. The concentration parameter $\kappa$ modulated the orientation bandwidth, i.e. the coherence level ($c$) of orientation signals, which controlled the task difficulty. The coherence level could range from 0% to 100%. A value of $c = 0\%$ corresponds to infinite bandwidth with mixture of all orientations equally likely, whereas $c = 100\%$ corresponds to pure gratings at one of the four task-relevant orientations. A 2-D Gaussian envelope was then applied to the images.

Because our stimuli were filtered from random noise, the random seeds used to generate noises would alter the stimulus image. To control the feedforward input, we chose 16 random seeds from a pool that maximized population task sensitivity and kept them constant for each task epoch. First, we randomly generated a pool of 300 integers as seeds. Each seed corresponds to one white noise image, on which we could apply filters to introduce orientation and spatial signal. The spatial frequency and spatial frequency range of filters were already determined separately, as mentioned elsewhere in the Methods section. For each seed/noise image, we created a 3-second movie stimulus consisting of 10 periods. In each period, the signal level changed from -15% to 15% with a 5% step size, and each signal level corresponded to one image frame, presented for 25 ms. We used such a stimulus because, traditionally, mapping task-sensitivity for a few hundred seeds would not be realistic due to the time limitation of behaving animals. Each neuron's sensitivity for each seed was quantified as the strength of modulation on its response by the periodic signal level change. To do so, we applied the Fast Fourier transform to spiking activity and obtained the normalized energy around the frequency of signal level change (3.33 cycles/second). We then used a random walk searching algorithm to find a subset of seeds (16 seeds for each task) that maximized the average modulation strength across the population of neurons. To generate variation in input images to infer the animals' task-solving strategy, we randomly permuted the presenting order of the seeds (each

seed corresponded to one image, shown for 100 ms) in each trial.

**Behavioral task**    We used MATLAB (The MathWorks, Inc) and Psychtoolbox (Brainard, 1997; Kleiner, 2007) to present visual stimuli and control the experiments. Gaze of the animals was tracked by an infrared eye tracking system (EyeLink 1000; SR Research, Ottawa, Ontario) and was monitored online by our software. Each trial started with presenting a fixation dot, which was presented at the center of the screen for monkey G, and 6.4° above the screen center for monkey R (to accommodate the eccentricity of the recorded RF). After the animal fixated on the dot, two choice targets that were equal-distance to fixation center were presented first, each corresponding to one orientation (See Fig. 2c for target setup). After a delay of 50 ms, 16 stimulus frames (each 100 ms duration) appeared sequentially, covering the receptive field. After stimulus presentation, the animals were required to wait for at least another 100 ms (fixed to be 100 ms for monkey R. uniformly sampled from 100 to 150 ms with 10 ms increment for monkey G) until fixation point was off. Then they had 1000 ms to indicate their choice of orientation by making a saccade to the corresponding choice target. The animal was rewarded with several drops of juice or water for making the correct choice of orientation and was rewarded randomly with 50% chance for 0% coherence trials. To better engage the animals, we increased amount of reward every consecutive time they finished the trial correctly (but no more than three times of the initial amount). If the gaze left 1.5° (for monkey R) or 1.7° (for monkey G) window around the fixation point prior to the fixation point offset, the trial was aborted and was not included in our analysis.

One experiment session was composed of several coherence levels of different probability of appearance, determined based on the subjects' performance in previous sessions. Early training sessions were mainly composed of high coherence trials (30% or 60% coherence) but still had low coherence trials with $c \leq 20\%$, allowing us to compare neural response between early training and after-learning sessions. As the animals improved their performance level, we gradually increased the percentage of low coherence trials relative to high coherence trials. For each session, a set of coherence levels and their percentage were chosen such that (1) there were 10% - 15% zero-coherence trials for purposes outside the scope of this paper; (2) there were middle coherence levels ($c \in [15\%, 30\%]$) trials to provide clearer guidance of the task for the animals; (3) and there were low coherence levels ($c \in [3\%, 10\%]$) trials. The percentage of low coherence trials increased as the animal's performance improved so that the overall accuracy and the reward amount were kept relatively constant across training. In each trial, the orientation and coherence level were randomly sampled. The number of sessions in each learning epoch was determined by animal's learning progress: we recorded until the animals finished about 5–10 sessions after their performance plateaued.

**Passive viewing sessions**    At the end of each session (except for the first 11 sessions of monkey R cardinal epoch), we recorded a passive viewing session. In each trial, the animal was required to fixate at the fixation point while 3 or 4 stimuli were presented. Each stimulus was 400 ms, with 150 ms inter-stimulus-interval. The choice targets were not shown and the animals were not

required to make decisions about stimulus orientation. Instead, a blue saccade target was shown at a random location at the end of each trial, and the animal was required to make a saccade to the target in order to get reward. Only the direction of target was randomized; the saccade distance was fixed at 4.83 visual degrees. The purpose of the saccade requirement was to keep the animal engaged during passive-viewing.

The stimulus location, size, spatial frequency and spatial frequency range of stimuli during passive-viewing were the same as those of stimuli during task-performing. The noise seeds were selected from the 16-seeds pool used in the task. The composition of orientation and coherence levels varied across sessions: for monkey R, we used 12 orientations tiled uniformly from 0 to 165 degrees, with only 15% coherence level at first; then we added 7.5% coherence level. for monkey G, we started with 8 orientations but reduced to only four task relevant ones (i.e., 0, 90, 45 and 135 degrees). The coherence levels were [0%, 7.5%, 15%]. Due to constraints of animals' working motivation, the number of repeats ($T$) was smaller in passive viewing sessions compared to task-performing, ranging from 20 to 100.

### Behavioral measurements

**Psychometric curve and threshold**    The PSIGNIFIT library (*87*) for MATLAB was used to compute psychometric curves: Fig. 2 shows the maximum a posteriori (MAP) fits. Psychometric thresholds corresponding to 75% accuracy were inferred from the psychometric function.

**Psychometric kernels and learning index**    To quantify each animals' behavioral strategy in each task, we performed logistic regression predicting choices from the empirical orientation energy in the images assuming a spatio-temporal kernel of rank-1 that factorizes in the spatial and temporal dimensions (*30*). This yielded a (normalized) orientation kernel $\hat{\mathbf{w}}_\theta$, a temporal kernel $\mathbf{w}_f$, and a bias term. The orientation kernel represents how much weight the animal puts on each orientation to make decisions, and the temporal kernel represents the contribution of each stimulus frame to the final decision. Details of this analysis can be found in the Supplementary Text. We defined the learning index as

$$\text{learning index} = \sum_f w_f \sum_\theta w_\theta \hat{w}_\theta$$

where $\mathbf{w}_\theta$ is the optimal kernel ($\mathbf{w}_\theta = \sin(2\theta+90°)$ for the cardinal task, $\mathbf{w}_\theta = \sin(2\theta)$ for the oblique task). Intuitively, the learning index is the product of how well the empirical strategy matches that of an ideal observer (the dot product between $\hat{\mathbf{w}}_\theta$ and $\mathbf{w}_\theta$), and how much the observer's choice is based on the stimulus signal.

**Behavioral Fisher information**    In order to compare neural Fisher information to behavioral performance using equivalent units, we computed behavioral Fisher information. The behavioral Fisher information can be computed as

$$I_{\text{behav}} = \frac{(z^+ - z^-)^2}{4x^2} \tag{S1}$$

5

where $x$ is the stimulus coherence, and $z^{+/-}$ denotes the $z$-score of probability of choosing the "negative orientation" when the signal level is positive ($z^+$) or negative ($z^-$). See Supplementary Text for derivation of Eq. S1.

We used Eq. S1 to compute $I_{\text{behav}}$ for each coherence level. The $I_{\text{behav}}$ of the whole session is computed as their weighted average, with the number of trials as the weights.

## Neural data analysis

**Inclusion criteria**     In order to avoid potential confounds, we excluded unstable units, identified as those with a firing rate coefficient of variation (CV) larger than 1 on a slow time-scale (computed across time bins of 15 min duration). Furthermore, we only included units whose average firing rates during stimulus presentation were larger than 1 spike second$^{-1}$ and whose activity was significantly modulated by visual stimulus for at least 20% of the stimulus presentation time (independent $t$-test against the baseline, use $p < 0.05$ as significance threshold). After exclusion, we kept, on average, $34.2 \pm 9.8$ units for monkey R and $82.8 \pm 16.5$ units for monkey G.

For each unit, we computed a signal-to-noise ratio (SNR), defined as the amplitude of the average action potential waveform divided by twice the standard deviation of waveform noise. The SNR was $2.82 \pm 0.86$ for monkey R and $2.53 \pm 0.55$ for monkey G. We used 2.5 as SNR threshold and yielded $23.4 \pm 7.8$ single-units and $10.8 \pm 4.9$ multi-units from monkey R, and $33.3 \pm 12.0$ single-units and $49.5 \pm 13.6$ multi-units from monkey G. We included both multi-units and single-units in results shown in the main text. Subsets of the analysis were also done using only single units, or only multi-units.

Exclusion of units was intentional, and the criteria were pre-established based on previous experience (*82*), rather than on the distributions of our data.

**Tuning properties of units**     We first examined sensitivity to stimulus and choice of included units. For each unit and each coherence level, we computed the tuning index for task-relevant orientations, stimulus $d'$, as well as choice $d'$.

$$\text{Tuning index} = \frac{\text{mean}(\mathbf{r}_{\text{pref stim}}) - \text{mean}(\mathbf{r}_{\text{anti-pref stim}})}{\text{mean}(\mathbf{r}_{\text{pref stim}}) + \text{mean}(\mathbf{r}_{\text{pre stim}})}$$

$$\text{Stimulus } d' = \frac{\text{mean}(\mathbf{r}_{\text{pref stim}}) - \text{mean}(\mathbf{r}_{\text{anti-pref stim}})}{\sqrt{\left(\text{var}(\mathbf{r}_{\text{pref stim}}) + \text{var}(\mathbf{r}_{\text{anti-pref stim}})\right)/2}}$$

$$\text{Choice } d' = \frac{\text{mean}(\mathbf{r}_{\text{pref choice}}) - \text{mean}(\mathbf{r}_{\text{anti-pref choice}})}{\sqrt{\left(\text{var}(\mathbf{r}_{\text{pref choice}}) + \text{var}(\mathbf{r}_{\text{anti-pref choice}})\right)/2}}$$

where $\mathbf{r}_{\text{pref stim}}$ and $\mathbf{r}_{\text{anti-pref stim}}$ are spike counts in response to preferred and anti-preferred orientations, respectively. Note that these two orientations are always 90° apart by task design. $\mathbf{r}_{\text{pref choice}}$ and $\mathbf{r}_{\text{anti-pref choice}}$ are spike counts from trials where the animal chose the orientation corresponding to or opposite to one unit's preference. Importantly, when computing stimulus $d'$,

6

we removed the effect of choice by first computing $d'$ conditioned on choice, then averaging them. Similarly, the effect of the stimulus was removed from the choice $d'$ by conditioning on the stimulus.

**Computing linear Fisher Information**   We computed empirical linear Fisher Information, and uncertainty about it, using a previously published bias correction for finite number of trials (*88*). For this current study, we extended the published formulas to account for different numbers of trials under two conditions (i.e., two orientations).

Consider a coherence level $c$ and two orientations to be discriminated represented by signal levels $\pm c$. The trial numbers of $\pm c$ are $T_1$, $T_2$, respectively. $N$ denotes the number of neurons. The naive estimator of linear Fisher information is defined as

$$\hat{I}_{\text{naive}} = \frac{d\boldsymbol{\mu}^{\mathsf{T}}}{d\theta}\mathbf{S}^{-1}\frac{d\boldsymbol{\mu}}{d\theta}$$

where $\mathbf{S}$ is the covariance matrix averaged across two orientations, $d\boldsymbol{\mu}$ denotes the difference in neural responses between two orientations, and $d\theta = 2c$. The bias-corrected estimator of linear Fisher information is given by:

$$\hat{I}_{\text{bc}} = \hat{I}_{\text{naive}}\frac{T_1 + T_2 - N - 3}{T_1 + T_2 - 2} - \frac{(T_1 + T_2)\,N}{T_1 T_2 d\theta^2} \tag{S2}$$

and its variance is

$$\text{var}\,\hat{I}_{\text{bc}} = (\alpha + 2\beta)I_{\text{bc}}^2 + (6\alpha + 12\beta + 4)\gamma I_{\text{bc}} + (3\alpha + 6\beta + 2)\gamma^2 N \tag{S3}$$

where

$$\alpha = \frac{2}{(v - N)(v - N - 3)}\;;\quad \beta = \frac{v - N - 1}{(v - N)(v - N - 3)}\;;\quad \gamma = \frac{T_1 + T_2}{T_1 T_2 d\theta^2}\;;\quad v = T_1 + T_2 - 2$$

For derivations and more details see Supplementary Text.

**Estimating Fisher information from empirical data**   We computed linear Fisher Information estimates ($\hat{I}_{\text{bc}}$), and the associated uncertainties (var $\hat{I}_{\text{bc}}$) separately for each stimulus coherence level, before combining the estimates weighted by the inverse of their uncertainties. Only trials completed by the animals (i.e., they kept fixation until the stimulus offset and made a saccade to indicate orientation choice) were included in the analysis. Only trials with coherence level less than 20% were included to best ensure local linearity of orientation tuning curves.

We took into account typical response latency of V4 neurons, and used spike counts from 50 ms to 1650 ms after stimulus onset to compute Fisher information. To examine evolution of $I_{\text{redundancy}}$ within a trial, we split the trial into 8 non-overlapping time bins, 200 ms each, and applied the above analysis to each of them.

**Controlling size of neural population**   The number of units yielded varied from session to session. To avoid potential confounds introduced by the varying population size on $I_{\text{real}}$, $I_{\text{shuffle}}$, and $I_{\text{redundancy}}$, we performed all Fisher information analyses on a constant population size by randomly sub-sampling $n_{\text{min}}$ units from all sessions where $n_{\text{min}}$ was determined by the session with the smallest number of units. The exception was the cardinal epoch of monkey R, where we excluded a session with only 10 units and based all analyses on the second-smallest unit number ($n = 17$). To sample units from each session, we generated up to 1000 unique combinations of $n_{\text{min}}$ units. If fewer than 1000 combinations were possible, we included all possible combinations. The final unit numbers used in the main analysis were: monkey R, cardinal: 17 units; monkey R, oblique: 18 units; monkey G: oblique: 44 units; monkey G, cardinal: 55 units.

In fig. S5 and fig. S16, we present $I_{\text{redundancy}}$, $I_{\text{real}}$ and $I_{\text{shuffle}}$ computed using the whole populations which varied session by session. The results were consistent with those based on size-controlled populations (Fig. 3 and Fig. 4). Results in fig. S13 and fig. S19 were computed using the whole populations for the sake of computational time. All other figures with Linear Fisher information estimation used size-controlled populations.

**Average noise correlations**   We computed noise correlations as the Pearson correlation of spike counts between pairs of neurons over repeated presentations of the same condition, separately for each signal level $\leq 20\%$. Uncertainty estimates were obtained using a leave-one-out approach. Results from all signal levels were linearly combined weighted by the inverse of the uncertainty associated with each estimate.

**Population neurometric thresholds**   The neurometric threshold was computed similarly to psychometric threshold using the PSIGNIFIT library (*87*). We computed orientation decoding accuracy for each coherence level with a linear decoder, cross-validated using a leave-one-out approach. The weights of the linear decoder were analytically computed using trials from the training set: $\mathbf{w} = \mathbf{S}^{-1}\mathbf{f}'$, in which $\mathbf{f}'$ denotes the difference in mean spike counts to two orientations, and $\mathbf{S}$ is the average of two covariance matrices, $\mathbf{S_1}$ and $\mathbf{S_2}$, corresponding to each orientation respectively.

**Analyzing passive-viewing data**   Since passive viewing trials were shorter than task trials, we applied stricter inclusion criteria to ensure statistically robust estimates: the average firing rates needed to be larger than 2 spikes second$^{-1}$ and during at least 20% of the stimulus presentation time the response had to be significantly modulated by the stimulus. This yielded on average $31.2 \pm 9.63$ units for monkey R (min = 10, max = 49) and $72.4 \pm 14.57$ units for monkey G (min = 34, max = 121). We also controlled the population size of passive-viewing data analysis. The used unit numbers were monkey R, cardinal: 17 units; monkey R, oblique: 13 units; monkey G: oblique: 34 units; monkey G, cardinal: 51 units. To account for latencies, analyses were based on spike counts from 50 ms after stimulus onset to 450 ms after stimulus onset to compute Fisher information. The first ten sessions in monkey R cardinal epoch did not contain passive viewing trials and are excluded from this analysis. For the comparison between task-performing and passive viewing data

(Fig. 3f), we split the task-performing trials into four time bins (400 ms each) to match the bin size of the passive-viewing data.

## Statistical tests

Spearman's Rank correlation quantified all correlational relationships between measurements of interest, e.g., the correlation between information redundancy and learning index, because it is more robust to non-normally distributed data compared to Pearson's Linear correlation (*89*).

We applied linear regression to quantify whether information redundancy or raw information increases with time in a task trial, with coefficient significance assessed by F-tests ($\alpha$=0.05).

The within-session comparisons, for which data points could be paired according to session index, used the paired t-test. Examples include comparing redundancy during passive viewing against task-performing, as well as comparing redundancy within units of higher task sensitivity and those of lower task sensitivity. Comparing observations during the early-learning and late-learning stages employed the independent t-test.

A permutation-based test was used to test whether information redundancy within higher task sensitivity units was more strongly correlated with behavioral performance than that within lower task sensitivity units. Specifically, we first computed the true differences in correlation/regression coefficients between two groups, then randomly shuffled the group labels 10000 times to obtain the null distribution of these differences.

## Hierarchical Bayesian inference model

**Model description**  We generated synthetic data using a hierarchical Bayesian inference model previously introduced in (*23*) (`https://github.com/haefnerlab/sampling_decision`). This model integrates a linear Gaussian representation of sensory neural responses (*90*) with an ideal observer model for a two-choice task. This model has been shown to accurately predict behavior, as well as noise and choice correlations in sensory neurons (*23, 29*). The model learns a task-specific prior and dynamically updates its belief about the stimulus category over the course of a trial, leveraging feedforward, recurrent, and feedback connections.

To simulate different learning stages, we varied the model parameter $\delta \in \{0, 0.02, 0.08\}$, which modulates the strength of the task-specific prior (*23*). The model performs hierarchical inference via neural sampling, though the predicted increases in redundancy—both across learning and within a trial—are independent of this choice (*26*). Stimulus difficulty was manipulated across 7 contrast levels, with 1024 trials simulated per condition (*23*). At $\delta = 0$, model performance remained at chance level, whereas at $\delta = 0.08$, it approached near-perfect accuracy ($\sim 100\%$) for a contrast level of 10%. (see (*23*) for details).

**Applying Fisher information estimation methods to synthetic data**  Analyses on the synthetic data were performed analogously to those on the empirical data, computing $I_{\text{real}}$, $I_{\text{shuffle}}$ and $I_{\text{redundancy}}$ for each contrast level and then taking their weighted average. We computed linear Fisher

information with various population sizes: the whole population (256 neurons, model variable $\mathbf{x}$), randomly sub-sampling 32 neurons or 8 neurons. We sub-sampled each session 256 times to estimate uncertainties. We used spike counts simulated from 96 Markov chain Monte Carlo samples of $\mathbf{x}$ to compute linear Fisher information, and split them evenly into 8 non-overlapping time bins to examine the dynamic of $I_{\text{real}}$, $I_{\text{shuffle}}$ and $I_{\text{redundancy}}$ within a trial.

## Control analyses

**Eye movement**   The gaze of the animals was recorded by EyeLink 1000 and our custom Matlab software. For each trial, we obtained the trajectory of gaze (position on the horizontal and vertical axes of the screen; sampling rate was 1000 Hz) from stimulus onset to stimulus offset. We smoothed the eye velocity (smoothing window = 6 samples) and reconstructed the trajectory from the initial position and the smoothed velocity.

We aimed to test whether systematic changes of eye movement variables over learning drove the increase in information redundancy. To do so, we examined five eye dynamic variables: position on the x-axis and y-axis, velocity on the x-axis and y-axis, as well as pupil size. For each session, we computed the mean and variance of each variable across trials, thus obtaining ten quantities. First, we calculated correlations between these variables and the learning index to check if any of them systematically evolved with learning. But more importantly, to test whether the relationship between information redundancy and behavior was robust when controlling for eye movements. Moreover, we considered whether using all ten quantities as control variables for partial correlations would affect our conclusions, as this would be the most conservative test. Directly doing so could be problematic due to relatively high dimensionality, small sample size, and collinearity between variables. Therefore, we applied principal component analysis to these eye dynamic quantities and used their PCs as the control variables.

It is possible that the eye movement variable evolved in conjunction with learning, but was not linked to behavior or Fisher information in the neural population. To test this, we first attempted to decode the animals' choice with the five variables: x-position, y-position, x-velocity, y-velocity, and pupil size. We ran linear discriminant analysis with 10-fold cross-validation. Decoding was done for each signal level with enough ($> 25$) trials for both choices, and trials of the majority choice were randomly subsampled to ensure that the two classes were balanced. The decoding accuracies of multiple signal levels were simply averaged. We examined whether decoding accuracy systematically changes with learning, and more importantly, the partial correlation between $I_{\text{redundancy}}$ and learning index with decoding accuracy as control.

The average eye position of each trial shifts the average stimulus input and thus may affect the total spike count. To more directly assess whether our finding of a correlation between redundancy and learning index could be attributed eye movement, we computed the Spearman correlation coefficient between $I_{\text{redundancy}}$ and the learning index using subsets of trials based on their average eye position. First, we computed each session's grand average position (session-wise eye position center), and the Euclidean distance between each trial's average position and this center. We next compared the correlations coefficients obtained from trials with average eye position close to the

session average compared to trials far from the average. Since the value of an empirically measured correlation coefficient depends on the amount of data, we selected 90%, 80%, 70%, 60%, 50%, and 40% of trials closest to the center, as well as the same percentage of trials furthest away from the center. If the correlation between $I_{\text{redundancy}}$ and the learning index were explained by variability in eye position, we would expect the correlation coefficients between these two groups to diverge as the subset size decreased, contrary to our actual findings (fig. S13a).

We performed a similar analysis was done to check for the effect of eye velocity on our results. The average eye velocity of each trial was used as a proxy for whether and how strongly microsaccades happened during each trial. We again chose different percentages of trials of lowest and highest eye speed and found no systematic dependence on eye speed that could explain our main finding of a robust relationship between redundancy and learning index (fig. S13b).

Despite that some of the ten analyzed eye movement metrics changed with training (fig. S11, table S3) and contained information about upcoming choices in some epochs (fig. S12, table S6), partial correlations between $I_{\text{redundancy}}$ and learning index remained significant for all epochs (1) with each metric as control variable (table S4),(2) with principal components that explained most eye-movement variance as control (table S5), and (3) with decoding accuracy of choice as control (table S6). We found no significant difference in terms of correlation between $I_{\text{redundancy}}$ and learning index when simply partitioning trials by eye position or eye velocity (fig. S13). Thus, eye movement variables are not likely a substantial confound for our main conclusions.

**Single and multi- units**    It is widely known that the signal quality of Utah arrays tends to decrease with time, leading to more multi-units activity later in the recording, and that the average noise correlations between multi-units are usually higher (*91*). To account for signal quality change, we repeated our analysis for single units and multi-units separately, using SNR > 2.5 as the classifying threshold. The range of number of units was $10 - 54$ for monkey R ($6 - 40$ single units and $4 - 27$ multi-units), and $44 - 134$ for monkey G ($15 - 71$ single units and $18 - 97$ multi-units). Again, to control for population size, we chose the largest number of units for each epoch such that most sessions could be included because they had more units. For the single-unit analysis, we used $N = 11$ for monkey R cardinal, $N = 12$ for oblique, $N = 20$ for monkey G oblique, and $N = 15$ for cardinal. For multi-unit analysis, we used $N = 9$ for monkey R cardinal, $N = 9$ for oblique, $N = 22$ for monkey G oblique, and $N = 18$ for cardinal. Note that this resulted in the removal of 7 monkey R cardinal sessions and 10 monkey R oblique sessions. Fisher information results with only single or multi-units are shown in fig. S4.

## Supplementary Text

### Enrichment of the animal

The macaque monkeys were housed individually or in pairs at the primate facility of the University of Rochester, on a 12-hour light-dark cycle (light from 7 am to 7 pm). The enrichment includes toys and videos played on television during the day, selected to match monkeys' interests. They were

provided with a nutritionally balanced diet, monitored by on-site veterinarians and animal facility staff. Their water intake was regulated, ensuring a minimum of 20 ml/kg/day during experimental sessions and 60 ml/kg/day on non-experimental days. Regular health assessments were performed by trained personnel, including onsite veterinarians and experienced laboratory members familiar with primate care.

**Array implantation surgery**

Surgery was performed under isoflurane anesthesia using aseptic technique, with perioperative opioid analgesics and postoperative antibiotics. To expose the target cranial site, we incised the scalp or removed bone cement as necessary. A bone drill was used to create $\sim 20$ mm craniotomies.

Implant location was confirmed via visual reference to sulcal landmarks: rostral to the lunate sulcus, caudal to the superior temporal sulcus, and dorsal to the ascending ramus of the inferior occipital sulcus. We then introduced a $\sim 10$ mm durotomy, positioned the array and array connector (secured with cortical screws), and inserted the array to a depth of 1 mm using a pneumatic array inserter (Blackrock Microsystems).

One reference wire was inserted under the dura, and another was inserted between the dura and the bone. The durotomy was then sutured closed and the dura was covered with surgical cellulose foam. We then sealed the craniotomy with bone cement (monkey G) or titanium straps secured with cortical screws (monkey R). Finally, the scalp was sutured closed.

**Compute psychometric kernel**

To quantify the animals' behavioral strategy in each task, we followed prior work (*30*) and ran a logistic regression that predicts the animal's choice for each trial based on the empirical "orientation energy" (**E**) in the images. **E** is an array with dimensions trials ($t$) $\times$ orientations ($\theta$) $\times$ frames ($f$), where each entry, $e_{t,\theta,f}$, was computed as the dot product between an orientation-dependent template ($\mathbf{M}(\theta)$) and the signal power of one frame in the Fourier domain ($||\mathcal{F}(\mathbf{I}_{t,f})||^2$):

$$e_{t,\theta,f} = \sum_{\rho,\phi} ||\mathcal{F}(\mathbf{I}_{tf})||^2_{\rho\phi} \mathbf{M}(\theta)_{\rho\phi}$$

The orientation template is given by

$$M(\theta)_{\rho\phi} = Rice(\rho; \mu_\rho, \sigma_\rho)\mathbb{I}_{tf}\{|\theta - \phi| < 7.5°\}$$

where $Rice()$ denotes he density function of Rician distribution, and $\rho$, $\phi$ are the spatial frequency and orientation polar coordinates in the Fourier domain. $\mu_\rho$ and $\sigma_\rho$ represent the mean and standard deviation of spatial frequency, as empirically used for each subject (see materials and methods). We used 12 templates, with $\theta$ spaced from 0° to 165° in 15° increments.

To standardize regressors, we generated zero-coherence stimuli with matched radius, spatial frequency, and spatial frequency standard deviation. All entries of **E** were $z$-scored with respect to mean and standard deviations of the orientation energy in zero-coherence stimuli, ensuring comparability across tasks and animals.

The psychophysical kernel model included 31 parameters: the orientation kernel ($\mathbf{w}_\theta$) with 12 orientation weights, the temporal kernel ($\mathbf{w}_f$) with 16 frame weights (assuming independence between $\mathbf{w}_\theta$ and $\mathbf{w}_f$), a bias term ($w_b$), and two lapse parameters ($\lambda_l$ and $\lambda_h$). The model was defined as follows:

$$z_t = w_b + \sum_{f,\theta} w_f w_\theta e_{f,\theta}$$

$$q_t = \lambda_l + (1 - \lambda_l - \lambda_h)\frac{1}{1 + e^{-z_t}}$$

$$\text{negative log likelihood} = -\sum_{\substack{t \\ c_t=+1}} \log(q_t) - \sum_{\substack{t \\ c_t=-1}} \log(1 - q_t)$$

where $c_t$ is the animal's choice on trial $t$, and $q_t$ is the predicted probability of choosing the "positive" orientation ($0°$ or $45°$) on trial $t$.

The prior on $w_b$ was Gaussian, while lapse parameters had an exponential prior. To prevent overfitting, we implemented smoothing priors for $\mathbf{w}_\theta$ and $\mathbf{w_f}$:

$$\text{negative log prior} = \frac{\text{ori}_{\text{ar2}}}{2}\mathbf{w}_\theta{}^\top \mathbf{AR2_{ori}}\mathbf{w}_\theta + \frac{\text{time}_{\text{ar0}}}{2}\mathbf{w_f}{}^\top\mathbf{w_f} + \frac{\text{time}_{\text{ar2}}}{2}\mathbf{w_f}{}^\top \mathbf{AR2_{time}}\mathbf{w_f}$$

$$+ \frac{\text{bias}_{\text{var}}}{2}w_b^2 + \frac{\lambda_l + \lambda_h}{\text{lapse}_{\text{mean}}}$$

where hyperparameters were set as:

$$[\text{ori}_{\text{ar2}}, \text{time}_{\text{ar0}}, \text{time}_{\text{ar2}}, \text{bias}_{\text{var}}, \text{lapse}_{\text{mean}}] = [0.5, 6, 60, 0.15, 0.06]$$

Here, $\mathbf{AR2_{ori}}$ and $\mathbf{AR2_{time}}$ denote the second-order finite-difference matrix with kernel $[1, -2, 1]$.

Parameter inference was performed using MATLAB built-in optimization function FMINCON, which minimized the negative log-posterior via the sequential quadratic programming algorithm. Lapse parameters were constrained between 0 and 1, and temporal kernel parameters were constrained to be positive. The L-2 norm of $\mathbf{w}_\theta$ was constrained to 1.

We employed a two-step fitting procedure: first, the temporal kernel was estimated while assuming fixed orientation kernel weights ($w_\theta = +1$ for $\theta = 0°$ or $45°$; $w_\theta = -1$ for $\theta = 90°$ or $135°$). The resulting temporal weights were then used to initialize the joint fitting of spatial and temporal kernels. Initial values for $\mathbf{w}_\theta$ and $w_b$ were set to zero, while $\lambda_l$ and $\lambda_h$ were initialized at 0.01.

**Behavioral Fisher information**

To directly compare behavior with neural Fisher information using equivalent units, we computed "behavioral Fisher information" ($I_{\text{behav}}$). At each coherence level $x$, we assumed that the subjects represented each orientation with a Gaussian distribution characterized by a mean $\mu$ and standard deviation $\sigma$ (shared between the two orientations), and made decision based on a criteria $c$.

The probability of choosing the "negative" ($\theta = 90°$ or $135°$) orientation when the stimulus orientation was "negative" or "positive" ($\theta = 90°$ or $135°$) can be computed from the behavioral

data, where $y$ denotes choice and $s$ denotes stimulus):

$$p(y = -1|s = -x) = \frac{\sum_t \mathbb{I}\{y_t = -1, s_t = -x\}}{\sum_t \mathbb{I}\{s_t = -x\}}$$

$$p(y = -1|s = x) = \frac{\sum_t \mathbb{I}\{y_t = -1, s_t = x\}}{\sum_t \mathbb{I}\{s_t = x\}}$$

Then, we transform these probabilities into $z$-scores:

$$z^- = \Phi^{-1}(p(y = -1|s = -x)), \quad z^+ = \Phi^{-1}(p(y = -1|s = x))$$

where $\Phi^{-1}$ represents the inverse of the cumulative function of a Gaussian distribution with mean $\mu$ and variance $\sigma^2$.

Since the $z$-score is given by

$$z = \frac{c - \mu}{\sigma}$$

we obtain the following equations:

$$\sigma = \frac{\mu + c}{z^-}, \quad \sigma = \frac{\mu + c}{z^+}.$$

Solving for $c$ and $\sigma$, we get

$$c = \mu \frac{z^+ - z^-}{z^+ + z^-}, \quad \sigma = \frac{-2\mu}{z^+ - z^-}$$

We also define the behavioral $f'$ as

$$f' = \frac{2\mu}{2x} = \frac{\mu}{x}$$

Thus, behavioral Fisher information is given by

$$I_{\text{behav}} = \frac{f'^2}{\sigma^2} = \frac{(z^+ - z^-)^2}{4x^2} \tag{S4}$$

Although we did not have direct access to the subjects' internal representations of stimuli ($\mu$, $\sigma$) or their decision criteria ($c$), these unknown variables canceled out in the process of computing $I_{\text{behav}}$.

We used Eq. S4 to compute $I_{\text{behav}}$ for each coherence level. The overall $I_{\text{behav}}$ for a session was calculated as a weighted average across coherence levels, with the number of trials serving as weights.

**Computing linear Fisher Information**

We estimated linear Fisher Information using analytical methods using bias correction (*88*). That work assumed equal trial counts for both stimulus conditions, but in our empirical data, animals performed different numbers of trials per condition. We therefore modified the formulas to account for different trial counts, $T_1$ and $T_2$, corresponding to the two orientations being discriminated.

**Bias-corrected estimation of $I_{\text{real}}$**   On each trial, a stimulus from one of two classes (positive or negative orientation) was presented. Let $\mathbf{r}_t^+$ and $\mathbf{r}_t^-$ denote the population responses to positive and negative stimuli on trial $t$ respectively. We follow prior work (*88*) in assuming (1) Gaussian variability, (2) equal covariance across conditions, and (3) locally linear tuning curves with respect to signal level. Model responses can be written as:

$$\mathbf{r_t^\pm} \sim \mathcal{N}\left(\mathbf{f}(\theta \pm d\theta), \mathbf{\Sigma}\right) \approx \mathcal{N}\left(\mathbf{f}(\theta) \pm d\theta \mathbf{f}'(\theta), \mathbf{\Sigma}\right)$$

In our task, the two orientations were always 90 degrees apart, $\theta$ denotes signal level of stimulus, which controlled the discrimination difficulty (See materials and methods). The tuning curves were assumed linear for signal levels close to zero (we constrained our analysis to trials with $|\theta| \le 20\%$). $T_1$ denotes number of trials for the positive class and $T_2$ for the negative classes. The empirical mean and covariance are given by:

$$\boldsymbol{\mu}^+ = \frac{1}{T_1} \sum_{t=1}^{T_1} \mathbf{r_t^+}, \quad \boldsymbol{\mu}^- = \frac{1}{T_2} \sum_{t=1}^{T_2} \mathbf{r_t^-}$$

$$\mathbf{S} = \frac{\sum_{t=1}^{T_1} \left(\mathbf{r_t^+} - \boldsymbol{\mu}^+\right)\left(\mathbf{r_t^+} - \boldsymbol{\mu}^+\right)^\top + \sum_{t=1}^{T_2} \left(\mathbf{r_t^-} - \boldsymbol{\mu}^-\right)\left(\mathbf{r_t^-} - \boldsymbol{\mu}^-\right)^\top}{T_1 + T_2 - 2}$$

The sampling distributions of the empirical mean of two stimuli are:

$$\boldsymbol{\mu}^+ \sim \mathcal{N}\left(\mathbf{f}(\theta) + \frac{d\theta}{2}\mathbf{f}'(\theta), \frac{\mathbf{\Sigma}}{T_1}\right)$$

$$\boldsymbol{\mu}^- \sim \mathcal{N}\left(\mathbf{f}(\theta) - \frac{d\theta}{2}\mathbf{f}'(\theta), \frac{\mathbf{\Sigma}}{T_2}\right)$$

From

$$\frac{d\boldsymbol{\mu}}{d\theta} = \frac{\boldsymbol{\mu}^+ - \boldsymbol{\mu}^-}{d\theta} \tag{S5}$$

it follows that the unbiased estimator in Eq. (S5) is sampled from:

$$\frac{d\boldsymbol{\mu}}{d\theta} \sim \mathcal{N}\left(\mathbf{f}'(\theta), \frac{T_1 + T_2}{T_1 T_2}\frac{\mathbf{\Sigma}}{d\theta^2}\right) \tag{S6}$$

We assume equal covariance $\mathbf{\Sigma}$ under two stimulus conditions, so $\mathbf{r_t^+} - \boldsymbol{\mu}^+$ and $\mathbf{r_t^-} - \boldsymbol{\mu}^-$ are drawn from an N-variate normal distribution

$$\mathbf{r_t^+} - \boldsymbol{\mu}^+ \sim \mathcal{N}_N(\mathbf{0}, \mathbf{\Sigma})$$

$$\mathbf{r_t^-} - \boldsymbol{\mu}^- \sim \mathcal{N}_N(\mathbf{0}, \mathbf{\Sigma})$$

Therefore, $S$ follows Wishart's distribution

$$\mathbf{S} \sim W_N\left(\frac{\mathbf{\Sigma}}{T_1 + T_2 - 2}, T_1 + T_2 - 2\right)$$

The expectation of the inverse of $S$ is:

$$\langle \mathbf{S^{-1}} \rangle = \frac{T_1 + T_2 - 2}{T_1 + T_2 - N - 3} \mathbf{\Sigma}^{-1} \tag{S7}$$

The naive estimator of linear Fisher information is defined as

$$\hat{I}_{\text{naive}} = \frac{d\boldsymbol{\mu}^\top}{d\theta} \mathbf{S^{-1}} \frac{d\boldsymbol{\mu}}{d\theta}$$

Following (88), the expectation value of this naive estimator is:

$$\langle \hat{I}_{\text{naive}} \rangle = \left\langle \frac{d\boldsymbol{\mu}^\top}{d\theta} \mathbf{S^{-1}} \frac{d\boldsymbol{\mu}}{d\theta} \right\rangle = \text{Tr} \left( \left\langle \frac{d\boldsymbol{\mu}}{d\theta} \frac{d\boldsymbol{\mu}^\top}{d\theta} \mathbf{S^{-1}} \right\rangle \right) = \text{Tr} \left( \left\langle \frac{d\boldsymbol{\mu}}{d\theta} \frac{d\boldsymbol{\mu}^\top}{d\theta} \right\rangle \langle \mathbf{S^{-1}} \rangle \right) \tag{S8}$$

Substituting Eq. (S6) and Eq. (S7) into Eq. (S8), we obtain:

$$\begin{aligned} \langle \hat{I}_{\text{naive}} \rangle &= \frac{T_1 + T_2 - 2}{T_1 + T_2 - N - 3} \text{Tr} \left( \left( \mathbf{f'f'}^\top + \frac{T_1 + T_2}{T_1 T_2} \frac{\mathbf{\Sigma}}{d\theta^2} \right) \mathbf{\Sigma}^{-1} \right) \\ &= \frac{T_1 + T_2 - 2}{T_1 + T_2 - N - 3} \left( I + \frac{T_1 + T_2}{T_1 T_2} \frac{N}{d\theta^2} \right) \end{aligned} \tag{S9}$$

Therefore, we can solve the bias-corrected estimator of linear Fisher information:

$$\hat{I}_{\text{bc}} = \hat{I}_{\text{naive}} \frac{T_1 + T_2 - N - 3}{T_1 + T_2 - 2} - \frac{(T_1 + T_2) N}{T_1 T_2 d\theta^2} \tag{S10}$$

**Bias-corrected estimation of $I_{\text{shuffle}}$**    We also need to derive the unbiased estimator for $I_{\text{shuffle}}$ to compute $I_{\text{redundancy}}$, the difference between $I_{\text{shuffle}}$ and $I_{\text{real}}$. $I_{\text{shuffle}}$ can be treated as the sum of linear Fisher information from individual neurons, assuming they are independent.

$$I_{\text{naive, shuffle}} = \sum_{i=1}^{N} I_i = \sum_{i=1}^{N} \frac{\left( \frac{d\mu_i}{d\theta} \right)^2}{s_i^2}$$

The distributions of empirical mean responses are identical to Eq. (S6). The distributions and expectations for each neuron's variance can are:

$$s_i^2 \sim \left( \frac{\sigma_i^2}{T_1 + T_2 - 2}, T_1 + T_2 - 2 \right)$$

with the expectation of the inverse of variances:

$$\left\langle \frac{1}{s_i^2} \right\rangle = \frac{T_1 + T_2 - 2}{(T_1 + T_2 - 4) \sigma_i^2} \tag{S11}$$

16

Again, following (88), the expectation of the naive $I_{\text{shuffle}}$ is

$$
\begin{aligned}
\langle \hat{I}_{\text{naive,shuffle}} \rangle &= \sum_{i=1}^{N} \left\langle \left( \frac{d\mu_i}{d\theta} \right)^2 \right\rangle \left\langle \frac{1}{s_i^2} \right\rangle \\
&= \sum_{i=1}^{N} \left( f_i'^2 + \frac{T_1 + T_2}{T_1 T_2} \frac{\sigma_i^2}{d\theta^2} \right) \frac{T_1 + T_2 - 2}{(T_1 + T_2 - 4)\sigma_i^2} \\
&= \frac{T_1 + T_2 - 2}{T_1 + T_2 - 4} \left( I_{\text{shuffle}} + \frac{T_1 + T_2}{T_1 T_2} \frac{N}{d\theta^2} \right)
\end{aligned}
\tag{S12}
$$

Thus, the bias-corrected estimator of $I_{\text{shuffle}}$ is

$$
\hat{I}_{\text{bc,shuffle}} = \hat{I}_{\text{naive,shuffle}} \frac{T_1 + T_2 - 4}{T_1 + T_2 - 2} - \frac{(T_1 + T_2)\,N}{T_1 T_2 d\theta^2}
\tag{S13}
$$

**Variance of the bias-corrected linear Fisher information** To quantify the uncertainty of Fisher information estimation for individual sessions, we employed the analytical method of (88) to calculate the variance of bias-corrected Fisher information, making minor adjustments to account for different trial numbers across conditions. The variance variance of $I_{bc}$ is given by

$$
\operatorname{var} \hat{I}_{\text{bc}} = (\alpha + 2\beta)I^2 + (6\alpha + 12\beta + 4)\gamma I + (3\alpha + 6\beta + 2)\gamma^2 N
\tag{S14}
$$

where

$$
\alpha = \frac{2}{(v - p)(v - p - 3)}, \quad \beta = \frac{v - p - 1}{(v - p)(v - p - 3)}
$$

$$
\gamma = \frac{T_1 + T_2}{T_1 T_2 d\theta^2}, \quad v = T_1 + T_2 - 2, \quad p = N
$$

Next, we compute the variance of bias-corrected $I_{\text{shuffle}}$. Since $\hat{I}_{\text{shuffle}} = \sum_{i=1}^{N} \hat{I}_i$, we have

$$
\operatorname{var} \hat{I}_{\text{shuffle}} = \sum_{i=1}^{N} \operatorname{var} \hat{I}_i
\tag{S15}
$$

where $\operatorname{var} \hat{I}_i$ can be computed with Eq. (S14) plugging in $p = 1$.
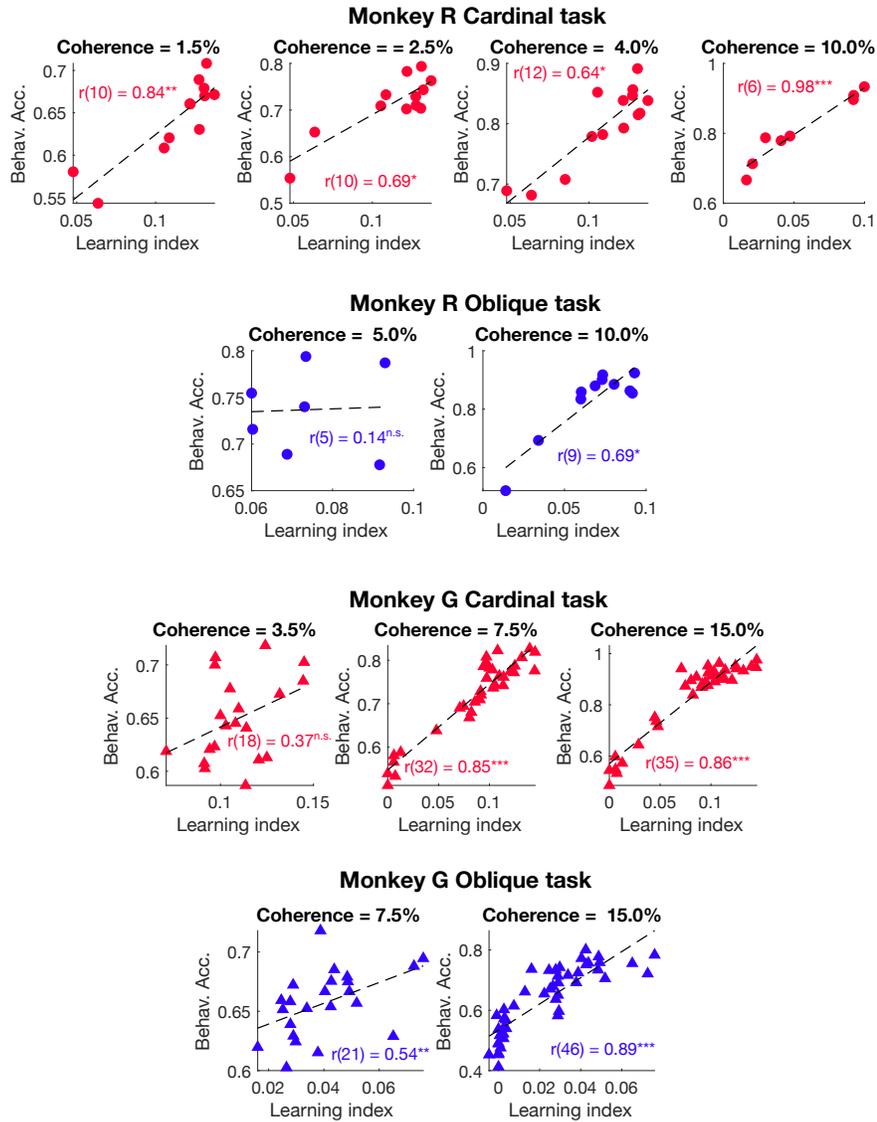
**Fig. S1: Correlation between learning index and behavioral accuracy across coherence levels.** For each epoch, coherence levels that appeared in multiple sessions were selected. We examined the correlation between the behavioral accuracy of each coherence level and the learning index for each session. Overall, strong positive correlations were observed between these two measures of behavioral performance. Note that the learning index was calculated using data from multiple coherence levels, so deviations from $r = 1$ are expected. (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).

# a

**Late-learning**

**Monkey R**



**Monkey G**



**Early-learning**

**Monkey R**



**Monkey G**



# b

**Late-learning**

**Monkey R**



**Monkey G**



**Early-learning**

**Monkey R**



**Monkey G**



— Stimulus: preferred; Choice: preferred    — Stimulus: anti-preferred; Choice: preferred

---- Stimulus: preferred; Choice: anti-preferred    ---- Stimulus: anti-preferred; Choice: anti-preferred

— Significant stimulus tuning    — Significant choice tuning    ---- Stimulus on/off

**Fig. S2:** See caption on the next page.

**Fig. S2: Peristimulus time histogram (PSTH). a**: Normalized neural response from 50 ms before stimulus onset to 100 ms after stimulus offset. Dashed vertical lines indicate stimulus onset and offset. To normalize firing rates of each unit, we computed the mean and standard deviation of firing rates during the baseline period (50 ms before stimulus onset) of all analyzed trials of a session, and used them to $z$-score firing rates of each time window (Bin size = 50 ms. A Gaussian-weighted moving average filter was applied). Then we grouped trials into four conditions based on each unit's preferred orientation and the chosen orientation. For each unit, the normalized firing rates within each group were averaged to yield the mean response for each condition (represented by the blue/orange lines). The shaded areas indicate the standard error of the mean. The top panels display the peristimulus spike-time histogram (PSTH) for four epochs during the late-learning stage, and the bottom panels show the PSTH for the early learning stage (the definitions of the late and early learning stages are provided in Fig. 2f). Horizontal maroon and green bars indicate time periods when significant stimulus or choice tuning was identified with cluster-based permutation test. **b**: Same as panel **a** but presenting un-normalized firing rate.
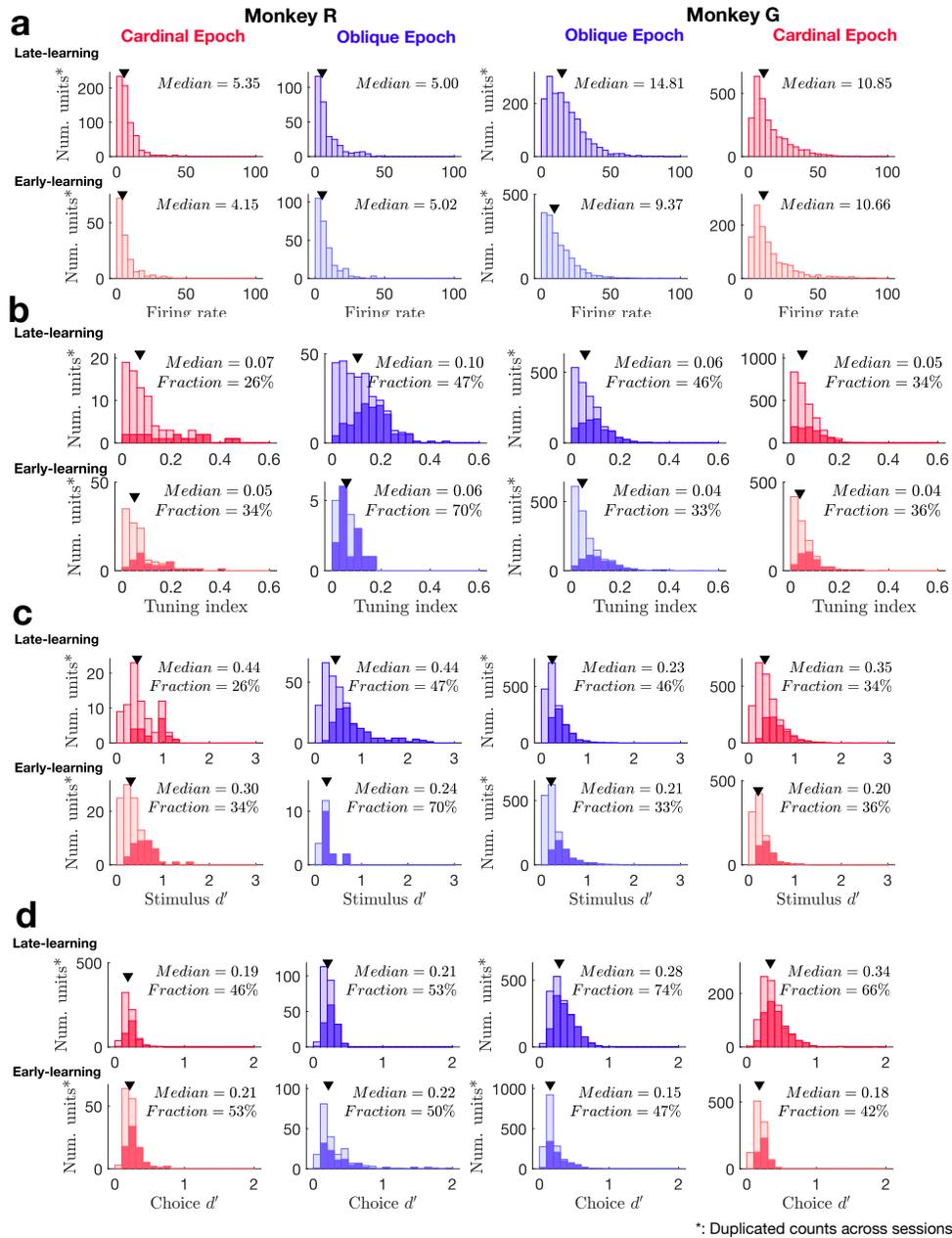
**Fig. S3: Histograms of basic properties of analyzed units.** In each panel, units from late-learning sessions are presented on top while units from early-learning sessions are at the bottom. The triangles indicate the median values. **a**: Averaged firing rates in response to all coherence levels. **b**: Tuning index computed with the highest coherence level for each animal (monkey R: 10%; monkey G: 15%). **c**: Stimulus $d'$ computed with the highest coherence level for each animal. **d**: Choice $d'$, computed for each coherence level, then averaged across coherence levels. In panels **b** and **c**, the filled histograms mark individual units with significant stimulus tuning or choice tuning (independent samples t-test).

21
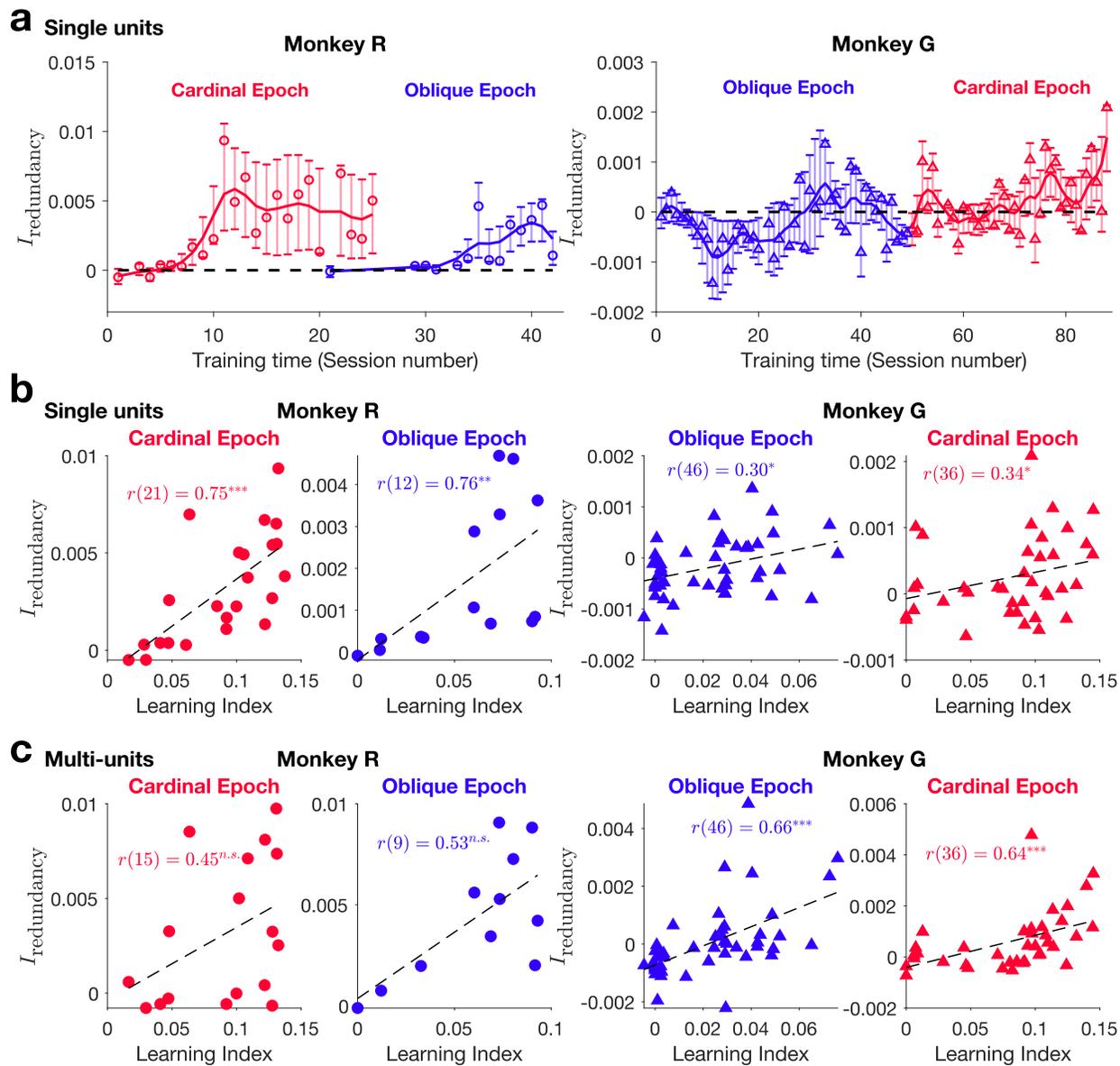
**Fig. S4:** See caption on the next page

**Fig. S4:** $I_{\text{redundancy}}$ **computed with single units or multi-units.** Same as Fig. 3, but with only single or multi- units. Spearman rank correlation between $I_{\text{redundancy}}$ computed with single units and the learning index: monkey R, cardinal: $\rho(21) = 0.75$, $p = 5.3 \times 10^{-5}$; monkey R, oblique: $\rho(12) = 0.76$, $p = 0.0023$; monkey G, oblique: $\rho(46) = 0.30$, $p = 0.037$; monkey G, cardinal: $\rho(36) = 0.34$, $p = 0.035$. Spearman rank correlation between $I_{\text{redundancy}}$ computed with multi-units and the learning index: monkey R, cardinal: $\rho(15) = 0.45$, $p = 0.074$; monkey R, oblique: $\rho(9) = 0.53$, $p = 0.10$; monkey G, oblique: $\rho(46) = 0.66$, $p = 8.7 \times 10^{-7}$; monkey G, cardinal: $\rho(36) = 0.64$, $p = 2.3 \times 10^{-5}$. The positive relationship between $I_{\text{redundancy}}$ and the learning index remains robust when using either single units (**b**) or multi-units (**c**). The only exception is the monkey R , where the correlation is only marginally significant ($p = 0.07$ or $p = 0.10$). This could be due to the exclusion of sessions and the low number of multi-units. For monkey R, the array signal quality was high (see figure. S7b for SNR), resulting in an insufficient yield of multi-units (we removed sessions that have fewer than 9 multi-units for this analysis).
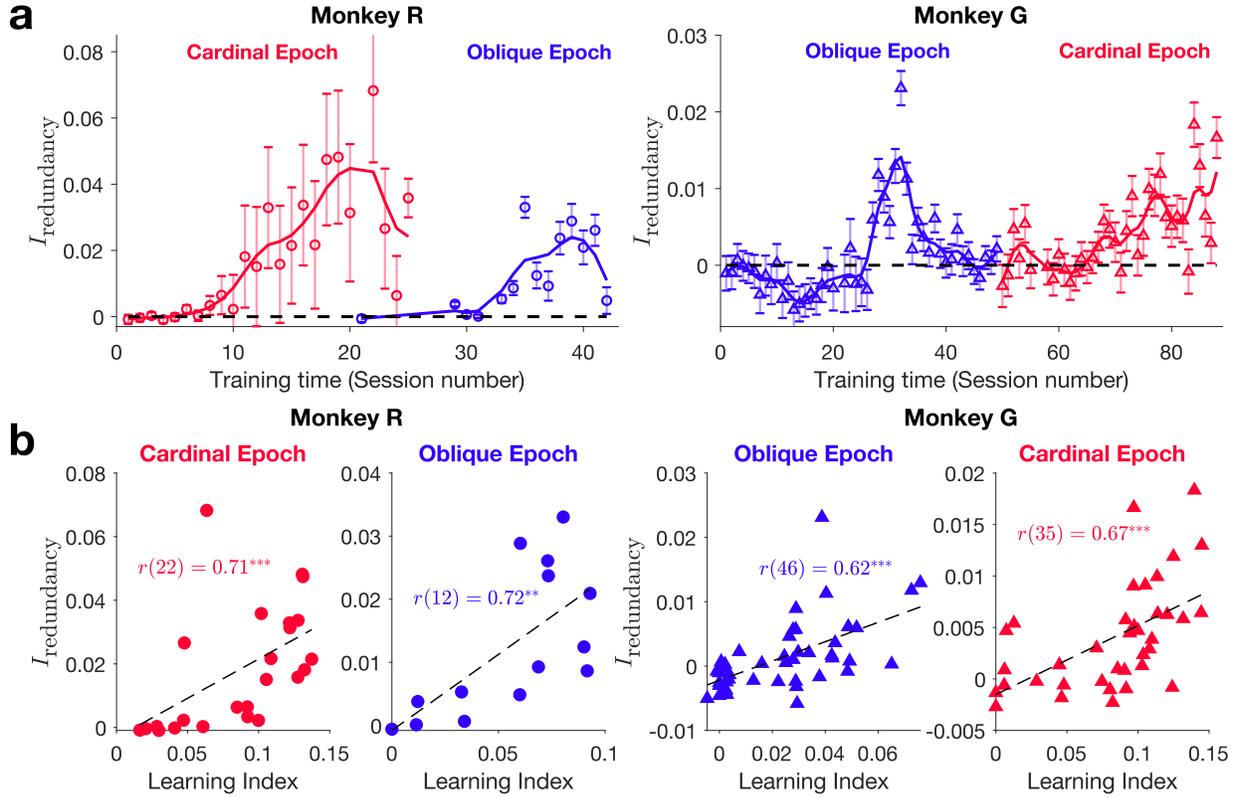
**Fig. S5:** $I_{\text{redundancy}}$ **computed using the whole neural populations over the course of learning.** Same as Fig. 3, but for the whole populations, and that the error bars in panel **a** indicate analytically computed standard deviations (materials and methods). Spearman rank correlations between $I_{\text{redundancy}}$ and the learning index are reported for each case: monkey R, cardinal task: $\rho(22) = 0.71$, $p = 1.6 \times 10^{-4}$; monkey R, oblique task: $\rho(12) = 0.72$, $p = 0.0048$; monkey G, oblique task: $\rho(46) = 0.62$, $p = 4.6 \times 10^{-6}$; monkey G, cardinal task: $\rho(35) = 0.67$, $p = 1.2 \times 10^{-5}$. (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$)
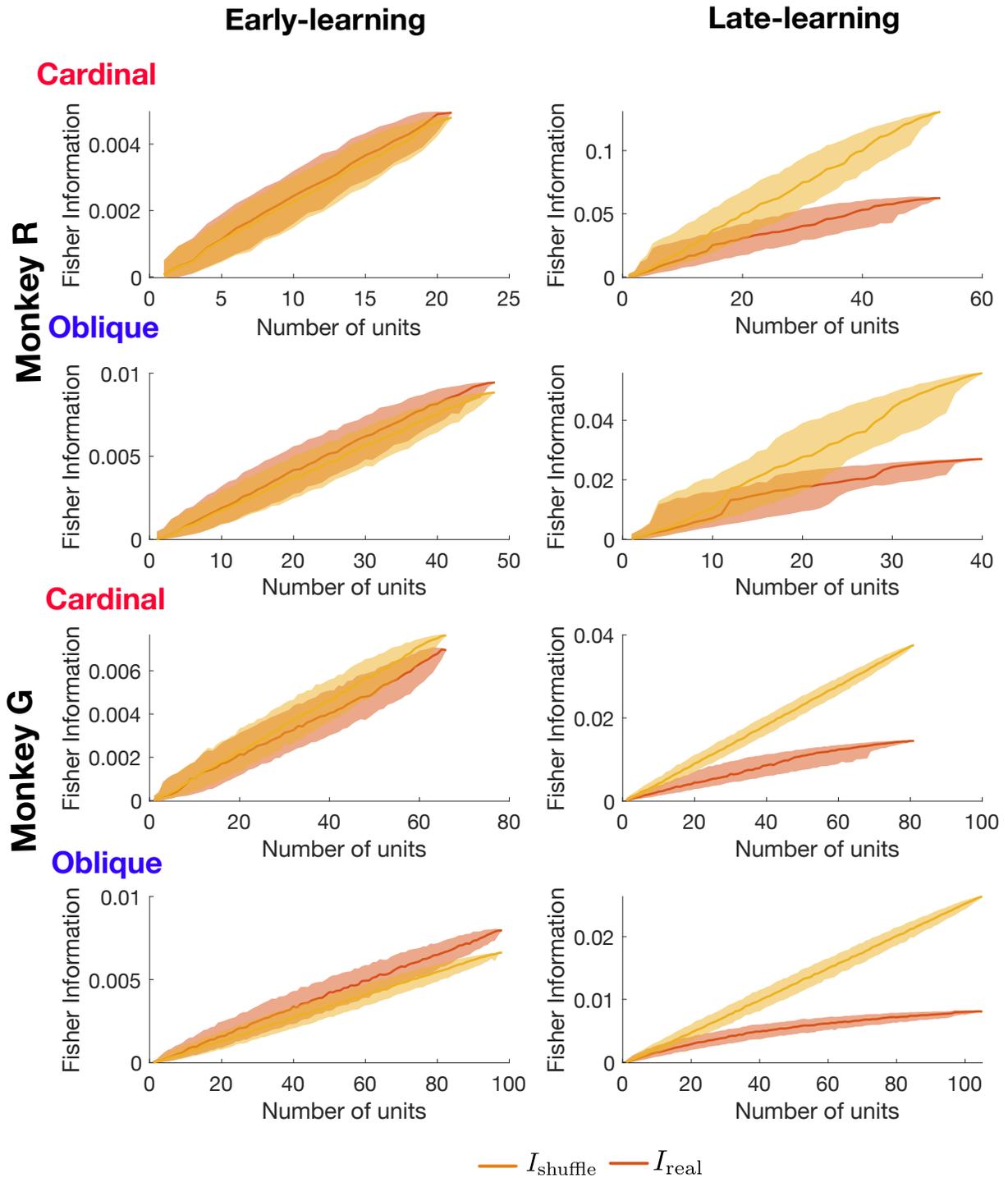
**Fig. S6: $I_{\text{real}}$ and $I_{\text{shuffle}}$ as a function of neural population size in example sessions.** One session before learning and one session after learning were selected for each task epoch. For each population size, units were bootstrapped 1,000 times or as many times as allowed by the total population size. Solid lines indicate the mean across samples, and shaded areas represent the 68% confidence intervals.

**a**

**Monkey R**

Cardinal Epoch    Oblique Epoch

**Monkey G**

Oblique Epoch    Cardinal Epoch

**b**

**Monkey R**

Cardinal Epoch    Oblique Epoch

**Monkey G**

Oblique Epoch    Cardinal Epoch

**c**

**Monkey R**

Cardinal Epoch    Oblique Epoch

$r(22) = -0.68^{***}$    $r(12) = 0.40^{n.s.}$

**Monkey G**

Oblique Epoch    Cardinal Epoch

$r(46) = 0.52^{***}$    $r(36) = -0.11^{n.s.}$
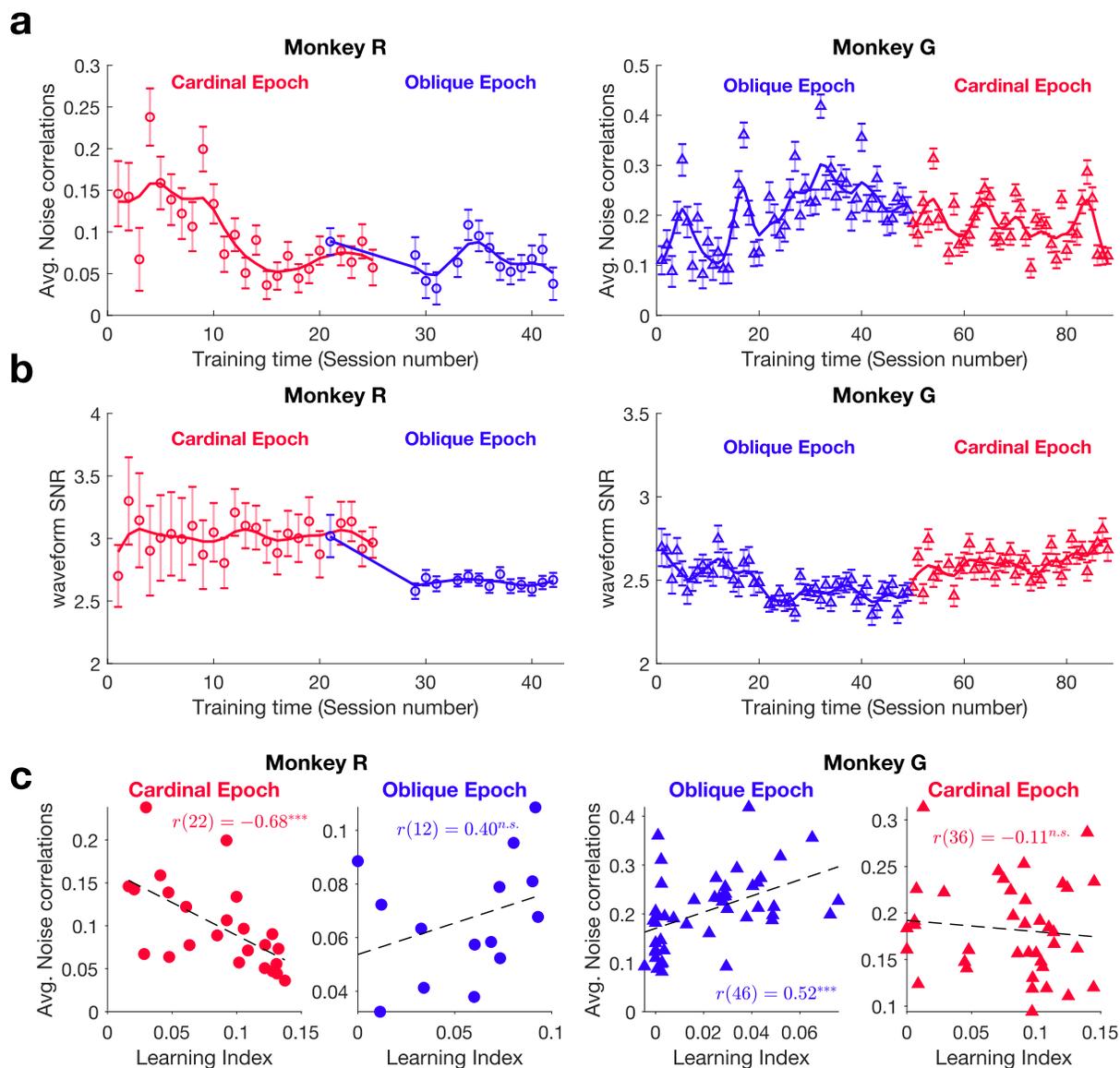
**Fig. S7:** See caption on the next page

**Fig. S7: Average noise correlation and spiking waveform signal-to-noise (SNR) over the course of learning. a**: Time course of average noise correlations for each epoch across all included units (materials and methods). Error bars indicate the standard error across unit pairs. **b**: Time course of the SNR of unit waveforms for each epoch. Error bars indicate the standard error cross units. **c**: Scatter plots of the learning index versus average noise correlations for two tasks and animals. Dots represent sessions. Black dashed lines indicate linear regression fits. (Significance levels: $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.) Average noise correlations did not consistently increase with task learning. In monkey R's cardinal epoch, they decreased ($\rho(23) = -.68, p < 0.001$) consistent with (*11*). in monkey G's oblique epoch, they increased ($\rho(47) = .52, p < 0.001$). However, this task epoch spanned a particularly long period (Fig. **??**), and signal quality from the array may have declined over time. For the other two epochs, no significant relationship was found.
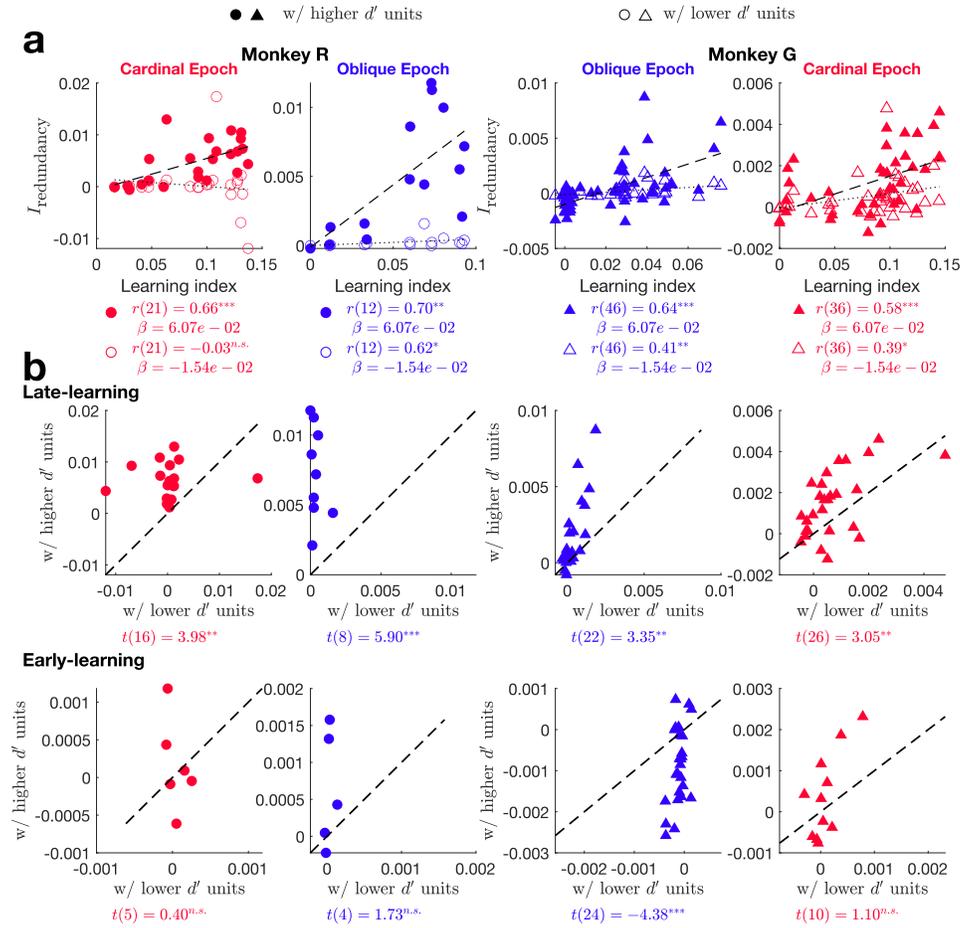
**Fig. S8: Information redundancy within population of high or low stimulus $d'$ based on task-performing data.** Units from each session were separated into two halves based on their stimulus $d'$ (Methods), and $I_{\text{redundancy}}$ was computed separately for each of the two populations, consisting of higher stimulus $d'$ units, and lower stimulus $d'$ units, respectively. Size of population was kept constant within each epoch (monkey R, cardinal: N = 8, monkey R, oblique: N = 9; monkey G, oblique: N = 22; monkey G cardinal: N = 27) **a**: Scatter plots between $I_{\text{redundancy}}$ of higher $d'$/lower $d'$ population (represented by filled/empty symbols) and learning index. **b**: Scatter plots comparing $I_{\text{redundancy}}$ for higher stimulus $d'$ population with $I_{\text{redundancy}}$ for lower $d'$ population. Each dot represents a session. The late-learning sessions are shown in the top row while the early-learning sessions are in the bottom row.
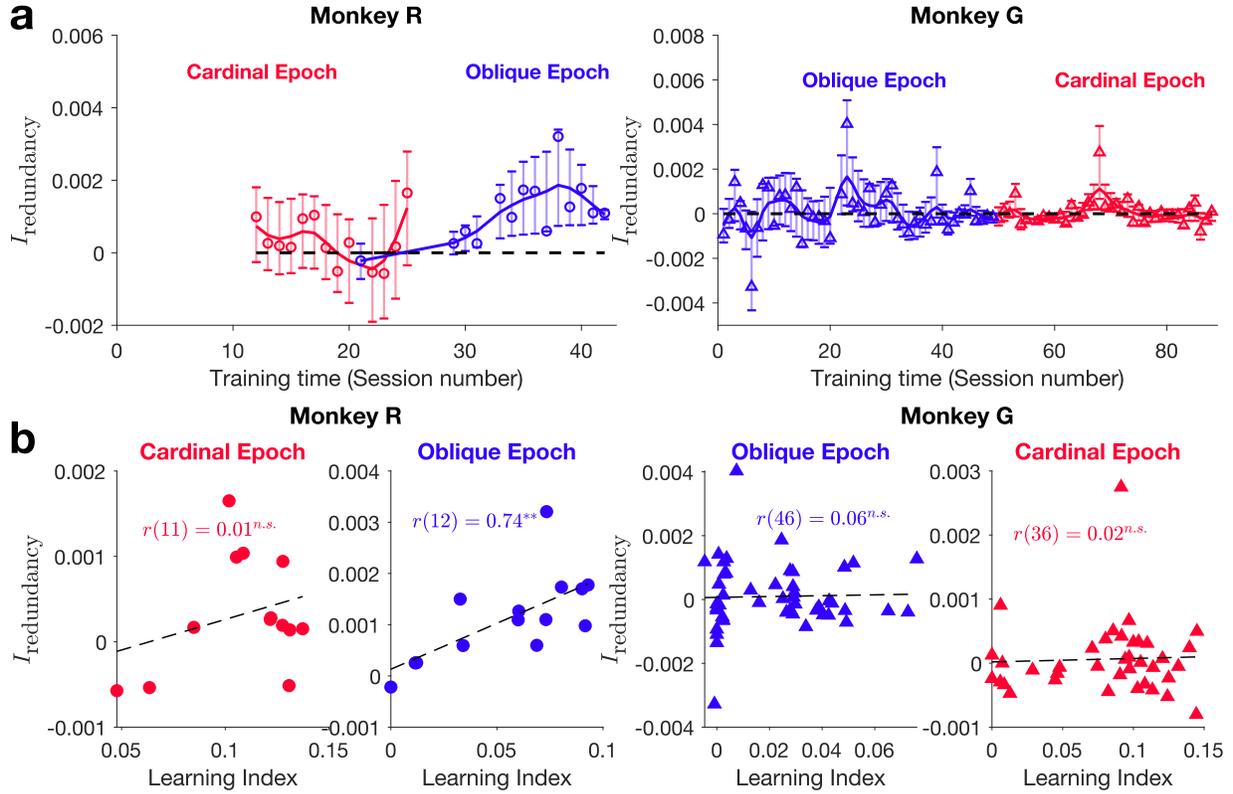
**Fig. S9:** $I_{\text{redundancy}}$ **in passive-viewing data.** Same as Fig. 3 but computed with neural responses during passive-viewing. Spearman rank correlation between $I_{\text{redundancy}}$ and the learning index: monkey R, cardinal: $\rho(11) = 0.01$, $p = 0.98$; monkey R, oblique: $\rho(12) = 0.74$, $p = 0.0038$; monkey G, oblique: $\rho(46) = 0.06$, $p = 0.67$; monkey G, cardinal: $\rho(36) = 0.02$, $p = 0.92$. This figure shows that $I_{\text{redundancy}}$ in passive-viewing data did not consistently increase with learning.
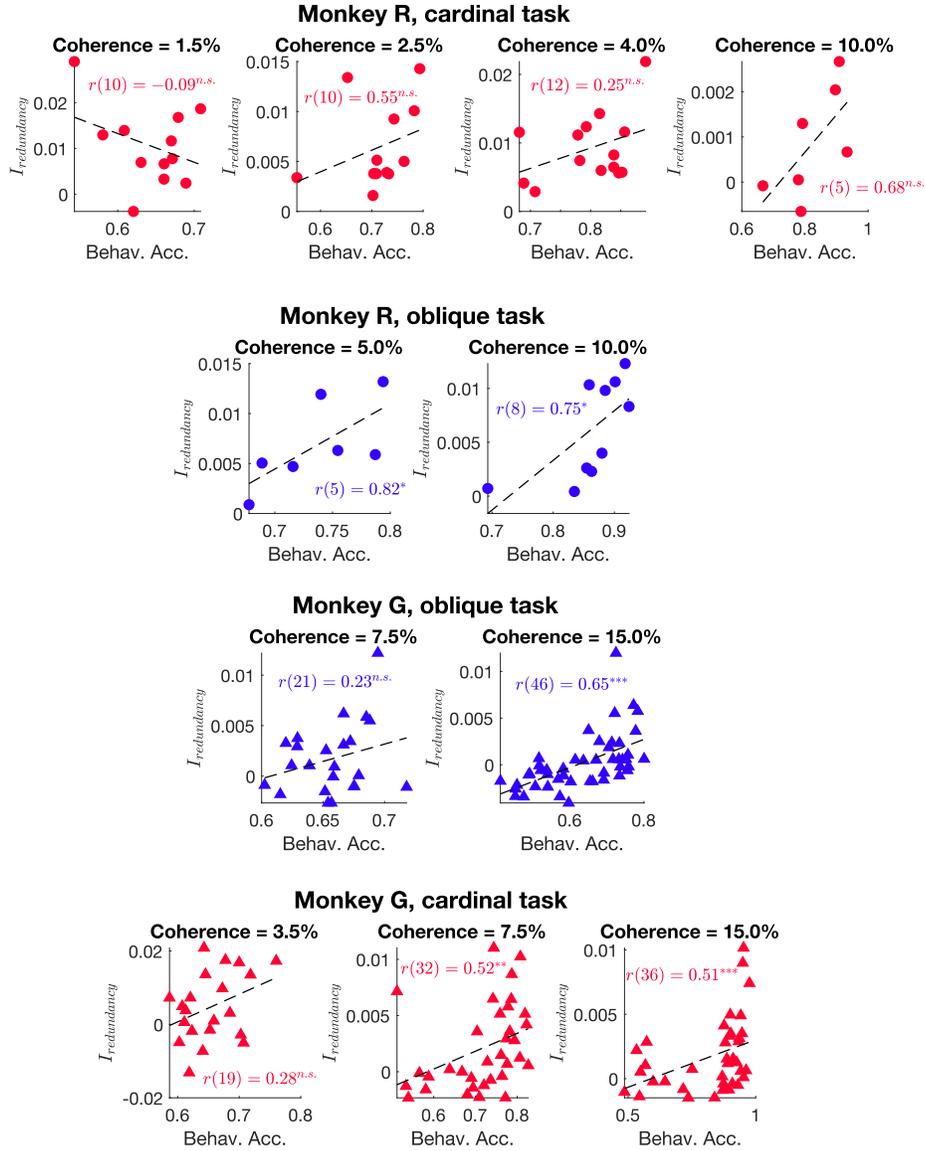
**Fig. S10: Correlation between $I_{\text{redundancy}}$ and behavioral accuracy for individual coherence levels.** For each epoch, a subset of coherence levels that appeared in multiple sessions was selected to examine the correlation between $I_{\text{redundancy}}$ and behavioral accuracy. Overall, positive correlations were observed, though some did not reach statistical significance, likely due to the limited number of sessions. This analysis supports the conclusion that the increase in $I_{\text{redundancy}}$ with learning, as shown in Fig. 3, cannot be attributed to changes in coherence composition during learning. (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$)
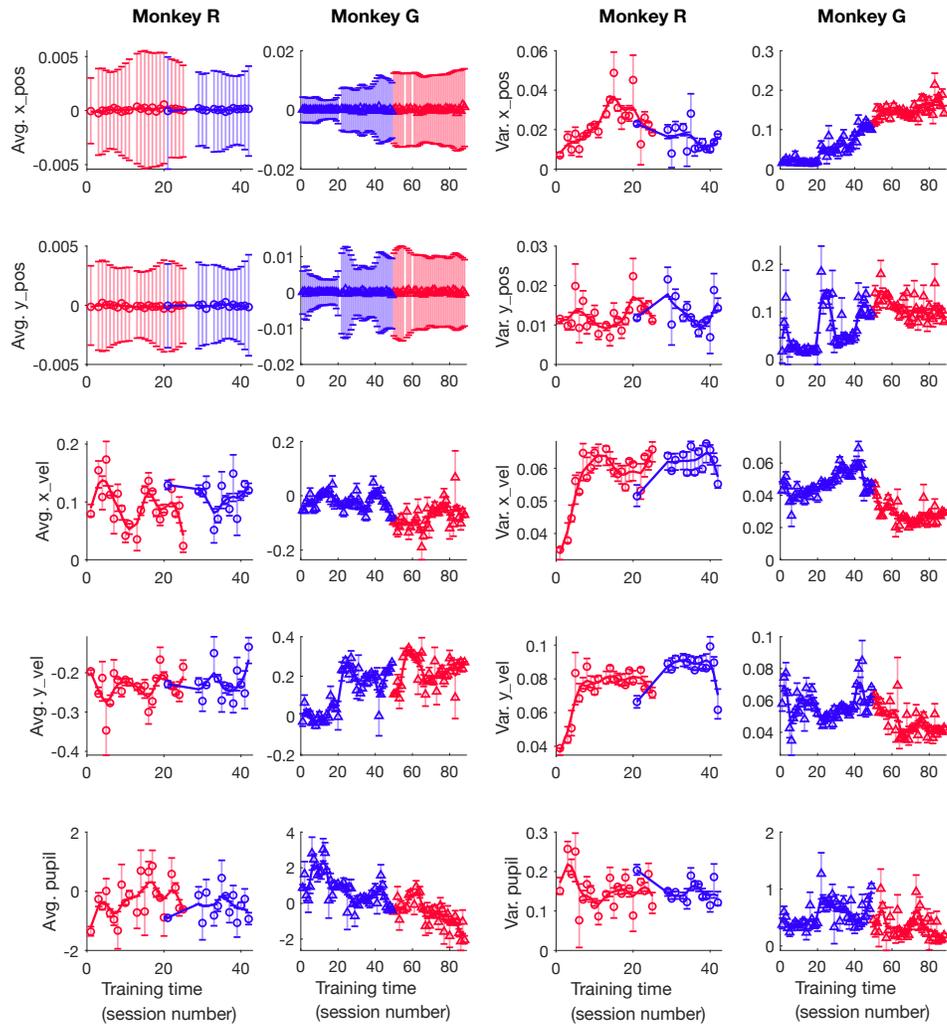
**Fig. S11: Time course of eye dynamic variables.** We examined five eye dynamic variables: position projected on the x-axis and y-axis, velocity along x-axis and y-axis, as well as pupil size. For each variable, we computed the mean values of each trial, then the average values (shown in the left two columns) and variance (the right two columns) across trials within each session. Error bars indicate 68% confidence interval estimated from 1000 Bootstraps of trials for each session. Position and velocity are in visual degrees, while pupil size is in arbitrary units and was normalized for each animal.
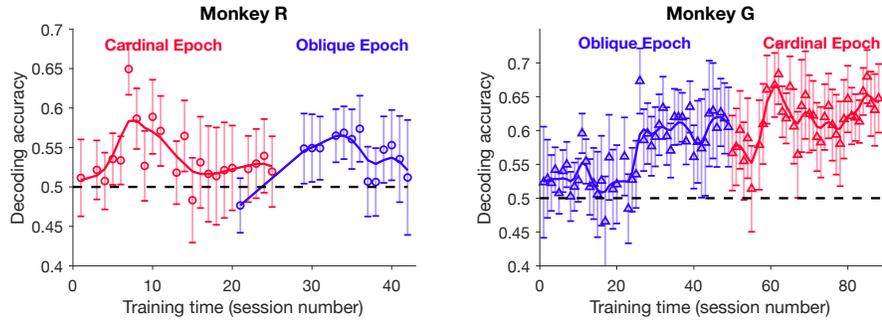
**Fig. S12: Decoding accuracy of animals' choice from eye dynamic variables**
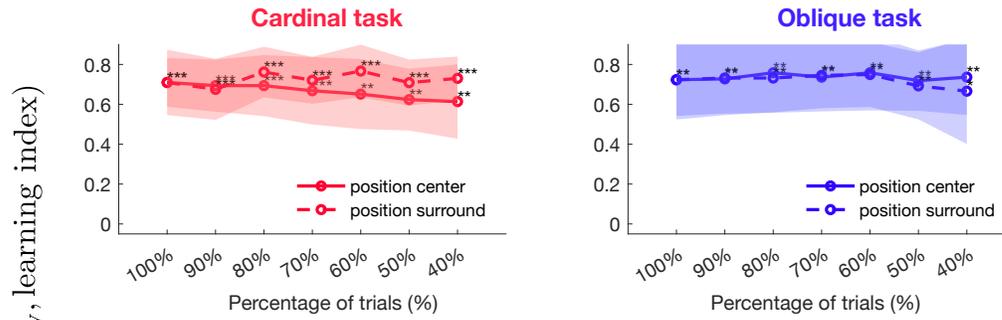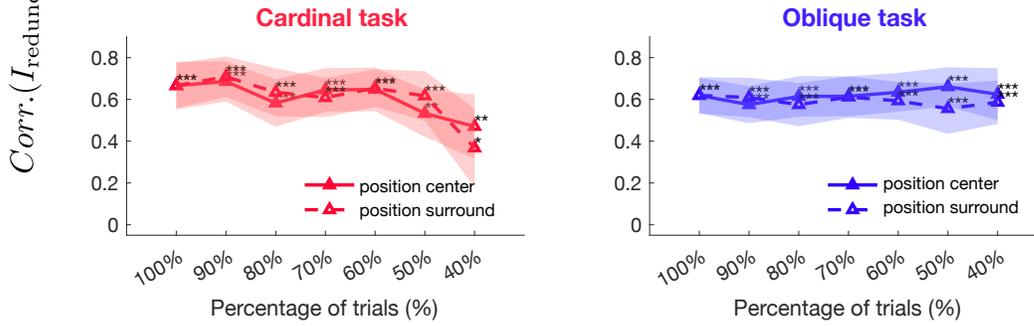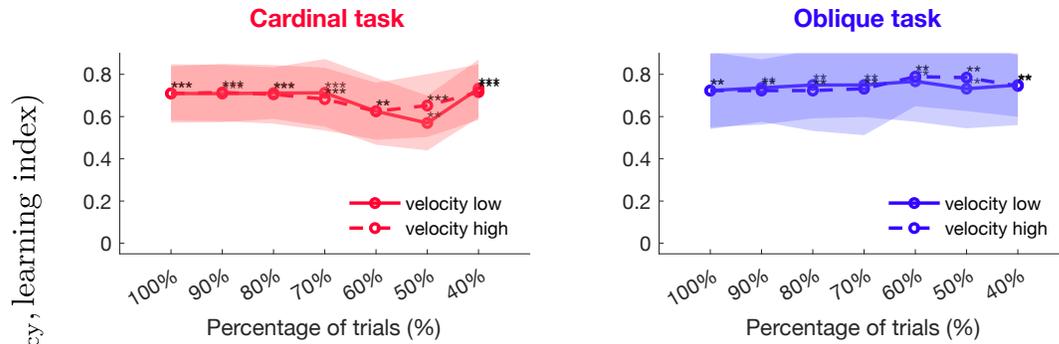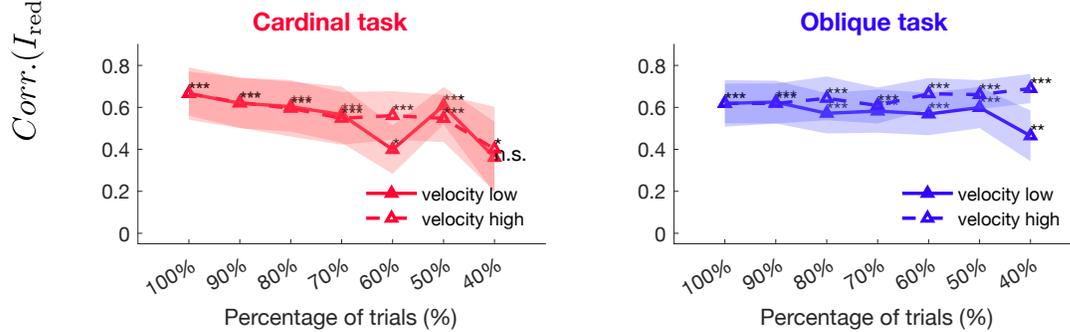
**Fig. S13:** See caption on next page.

**Fig. S13: Control analysis showing that the correlation between redundancy and learning index is not explained by either eye position or eye velocity.** Spearman correlation between $I_{\text{redundancy}}$ and learning index was computed using subsets of trials selected based on eye position (**a**) and eye velocity (**b**) (Methods). The x-axis represents the percentage of selected trials, and the y-axis shows the Spearman correlation coefficients between $I_{\text{redundancy}}$ and learning index (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).Shaded areas indicate the standard deviation obtained from Bootstrapping sessions. **a**: Solid lines indicate Spearman correlation coefficients computed using trials with average eye positions closer to the grand average of eye position, while dashed lines correspond to trials with eye positions further from the grand average. **b**: Solid lines represent Spearman correlation coefficients computed using trials with lower average eye speed, while dashed lines correspond to trials with higher eye speed.
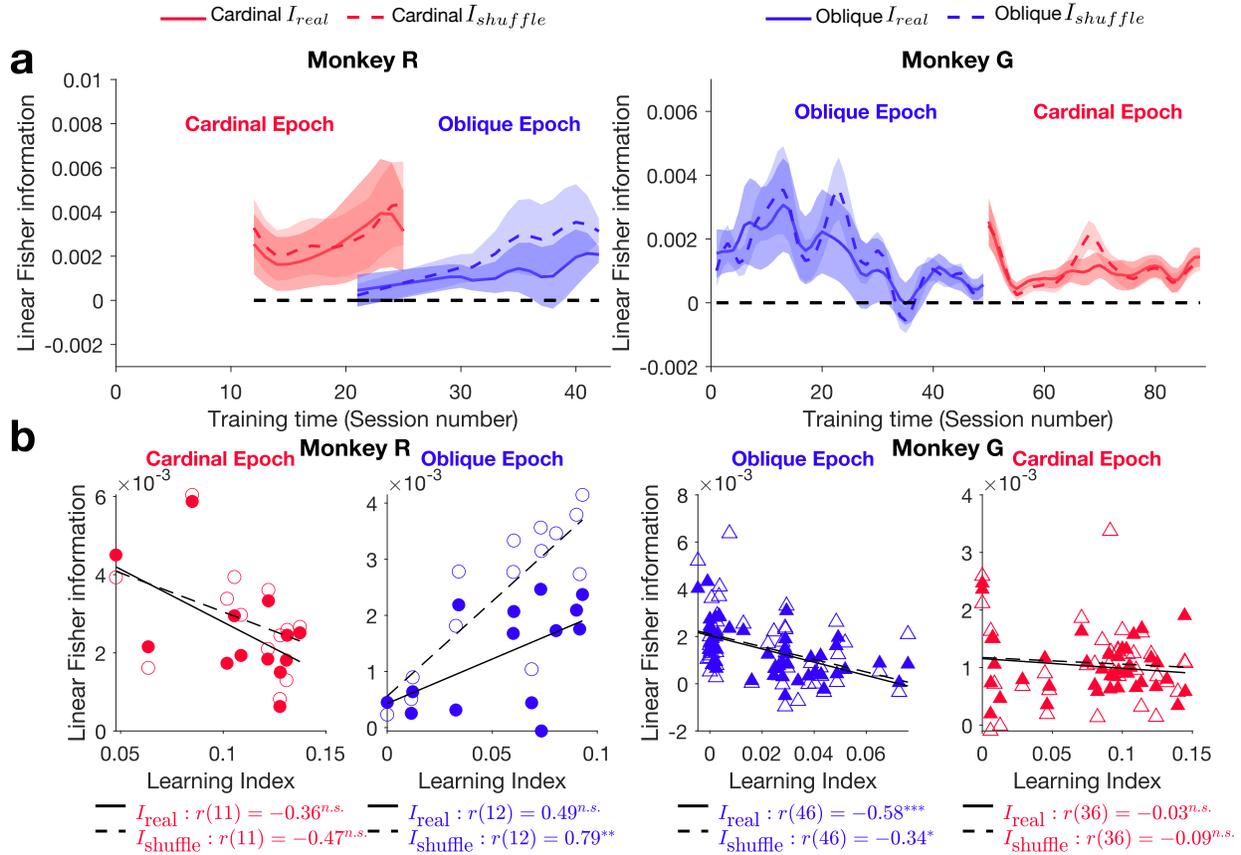
**Fig. S14:** $I_{\text{real}}$ **and** $I_{\text{shuffle}}$ **in passive-viewing data.** Same as Fig. 4 but computed with neural responses during passive-viewing. Spearman correlations between $I_{\text{real}}$ and learning index: monkey R cardinal: $\rho(11) = -0.36$, $p = 0.23$; monkey R, oblique: $\rho(12) = 0.49$, $p = 0.081$; monkey G, oblique: $\rho(46) = -0.58$, $p = 2.5 \times 1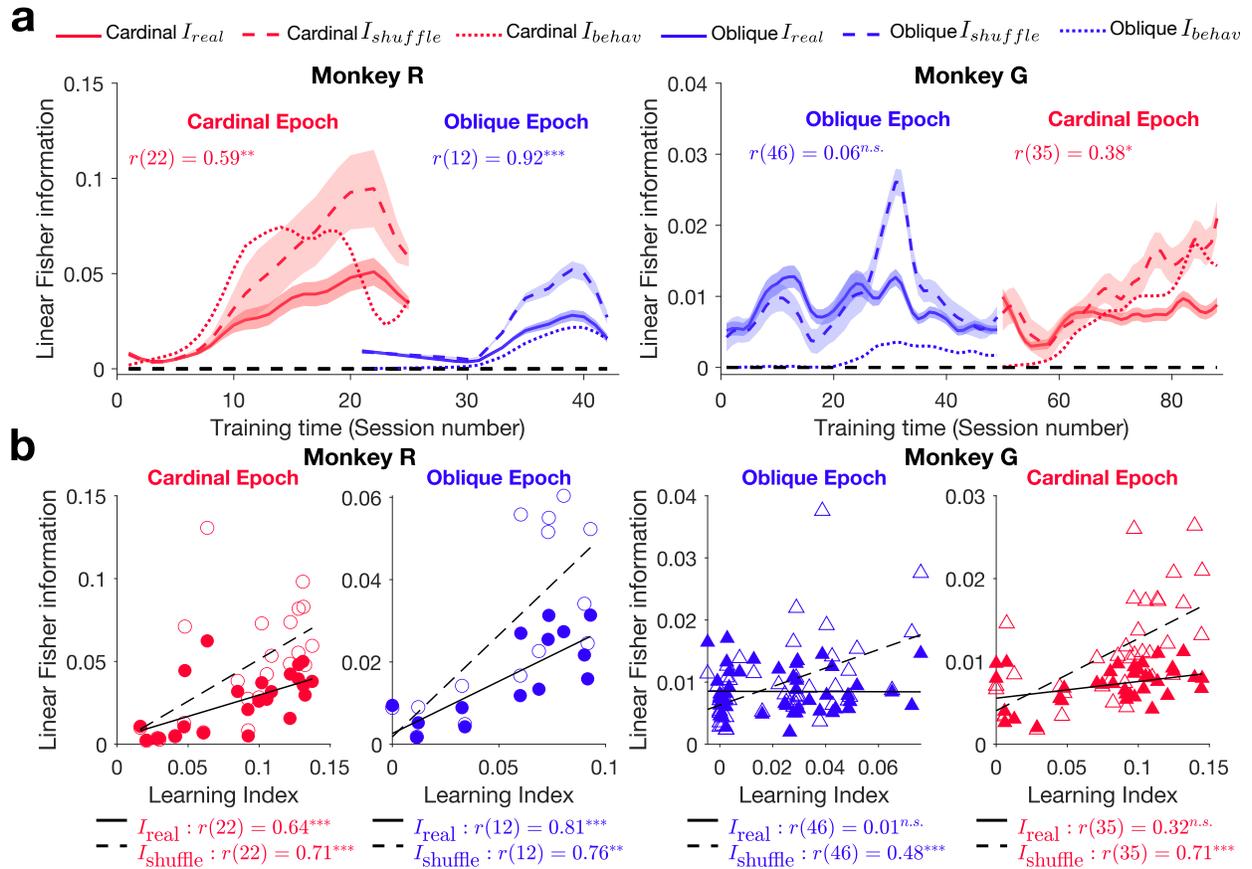0^{-5}$; monkey G, cardinal: $\rho(36) = -0.03$, $p = 0.88$. Spearman correlations between $I_{\text{shuffle}}$ and learning index: monkey R, cardinal: $\rho(11) = -0.47$, $p = 0.11$; monkey R, oblique: $\rho(12) = 0.79$, $p = 0.0013$; monkey G, oblique: $\rho(46) = -0.34$, $p = 0.018$; monkey G, cardinal: $\rho(36) = -0.09$, $p = 0.59$. This figure shows that $I_{\text{real}}$ and $I_{\text{shuffle}}$ in passive-viewing data did not consistently increase with learning.

35
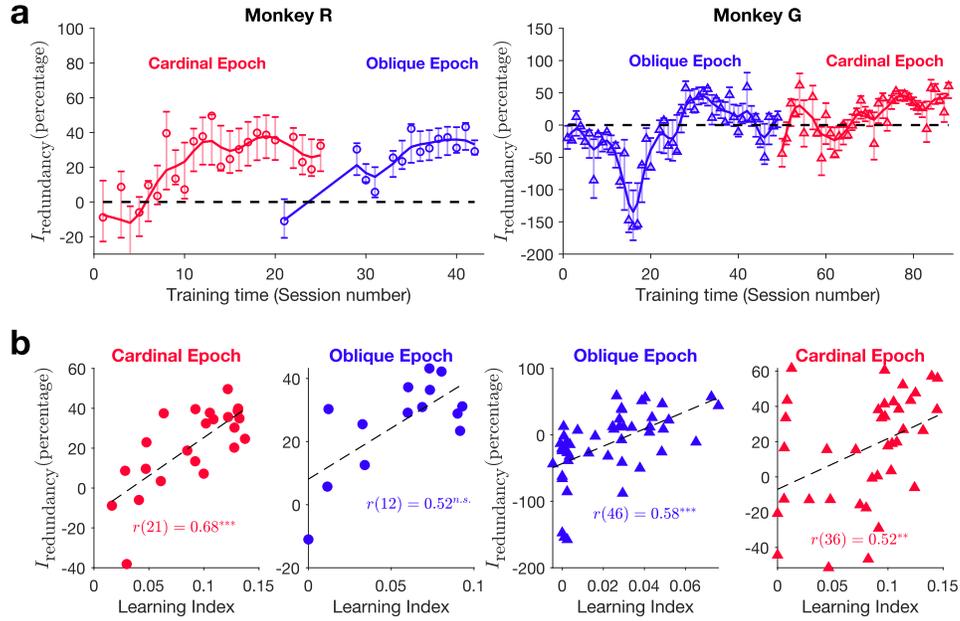
**Fig. S15: Percentage of redundant information** ($I_\text{redundancy}/I_\text{shuffle} \times 100$). **a**: Percentage of redundant information increased with training time. **b**: Scatter plot between this percentage and learning index. Spearman rank correlation: Monkey R, cardinal task: $\rho(21) = 0.68$, $p = 4.9 \times 10^{-4}$; Monkey R, oblique task: $\rho(12) = 0.52$, $p = 0.062$; Monkey G, oblique task: $\rho(46) = 0.58$, $p = 2.0 \times 10^{-5}$; Monkey G, cardinal task: $\rho(36) = 0.52$, $p = 0.0011$.

**Fig. S16:** $I_{\text{real}}$ **and** $I_{\text{shuffle}}$ **over the course of learning computed using the whole neural populations.** Same as Fig. 4, but for the whole populations, and that the error bars in panel **a** indicate analytically computed standard deviations (materials and methods). Spearman correlation coefficients between $I_{\text{real}}$ and $I_{\text{behav}}$ are reported for each condition: monkey R, Cardinal: $\rho(22) = 0.59$, $p = 0.0026$; monkey R, Oblique: $\rho(12) = 0.92$, $p < 2.2 \times 10^{-308}$; monkey G, Cardinal: $\rho(35) = 0.38$, $p = 0.020$; monkey G, Oblique: $\rho(46) = 0.06$, $p = 0.69$. Spearman correlations between $I_{\text{shuffle}}$ and learning index: monkey R, cardinal: $\rho(22) = 0.71$, $p = 1.4 \times 10^{-4}$; monkey R, oblique: $\rho(12) = 0.76$, $p = 0.0025$; monkey G, oblique: $\rho(46) = 0.48$, $p = 7.4 \times 10^{-4}$; monkey G, cardinal: $\rho(35) = 0.71$, $p = 2.0 \times 10^{-6}$. Spearman correlations between $I_{\text{real}}$ and the learning index: monkey R, cardinal: $\rho(22) = 0.64$, $p = 9.9 \times 10^{-4}$; monkey R, oblique: $\rho(12) = 0.81$, $p = 8.1 \times 10^{-4}$; monkey G, oblique: $\rho(46) = 0.01$, $p = 0.93$; monkey G, cardinal: $\rho(35) = 0.32$, $p = 0.055$.

37
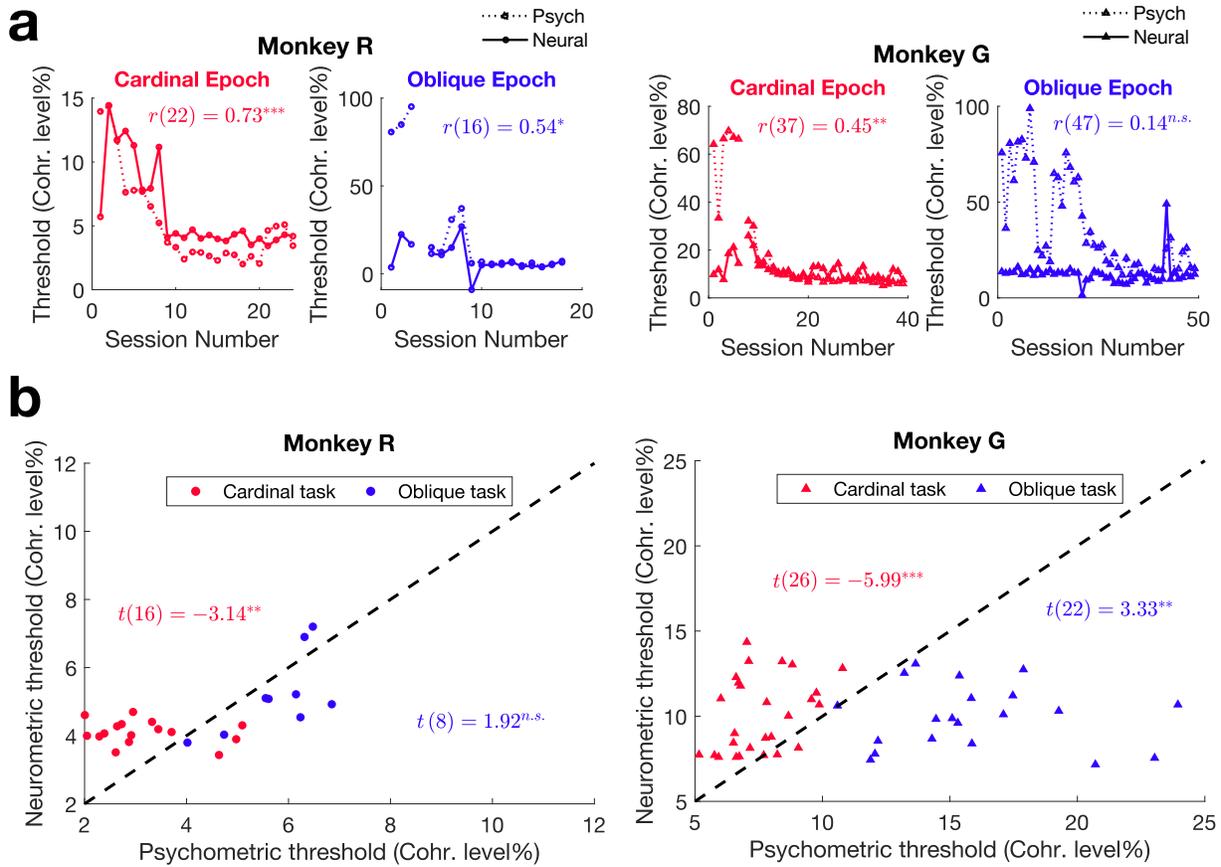
**Fig. S17: Comparison of neurometric and psychometric thresholds. a**: Time course of neurometric and psychometric thresholds throughout learning. **b**: Scatter plot of neurometric versus psychometric thresholds for well-performed sessions.
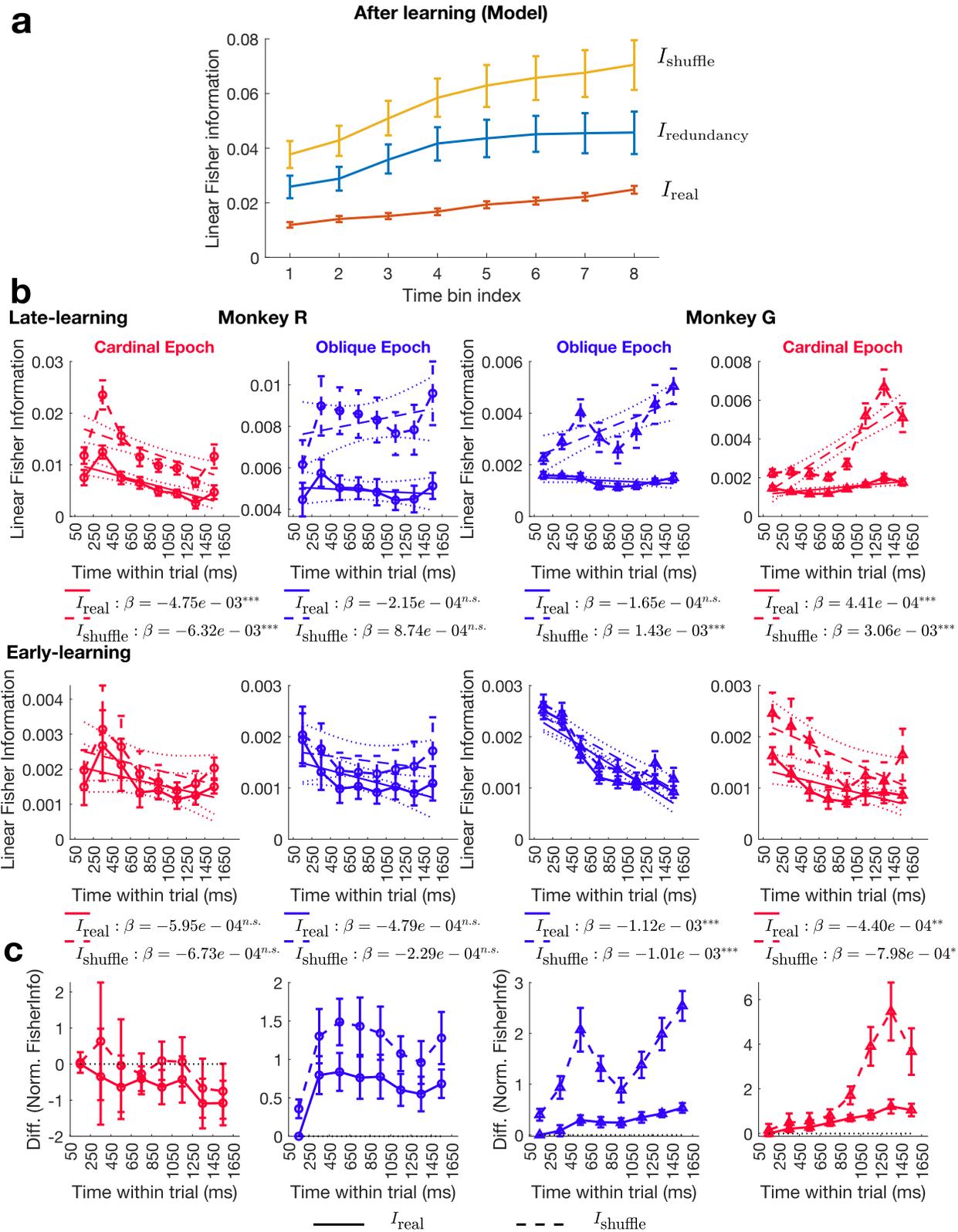
**Fig. S18:** See caption on the next page

**Fig. S18: Within-trial dynamic of $I_{\text{redundancy}}$, $I_{\text{real}}$ and $I_{\text{shuffle}}$.** **a**: Within-trial dynamic of $I_{\text{redundancy}}$, $I_{\text{real}}$ and $I_{\text{shuffle}}$ in the after-learning synthetic dataset. Error bars represent the 68% confidence interval across 256 random samples of synthetic neurons. **b**: Top: Within-trial dynamic of $I_{\text{real}}$ and $I_{\text{shuffle}}$ in late-learning sessions from empirical data. Bottom: Empirical data from early-learning sessions. Solid lines represent $I_{\text{real}}$, while dashed lines represent $I_{\text{shuffle}}$. **c**: Difference between late-learning and early-learning $I_{\text{real}}$ and $I_{\text{shuffle}}$. To ensure comparability across sessions with varying population sizes, data points from each session were normalized by the $I_{\text{real}}$ value of the first time bin.
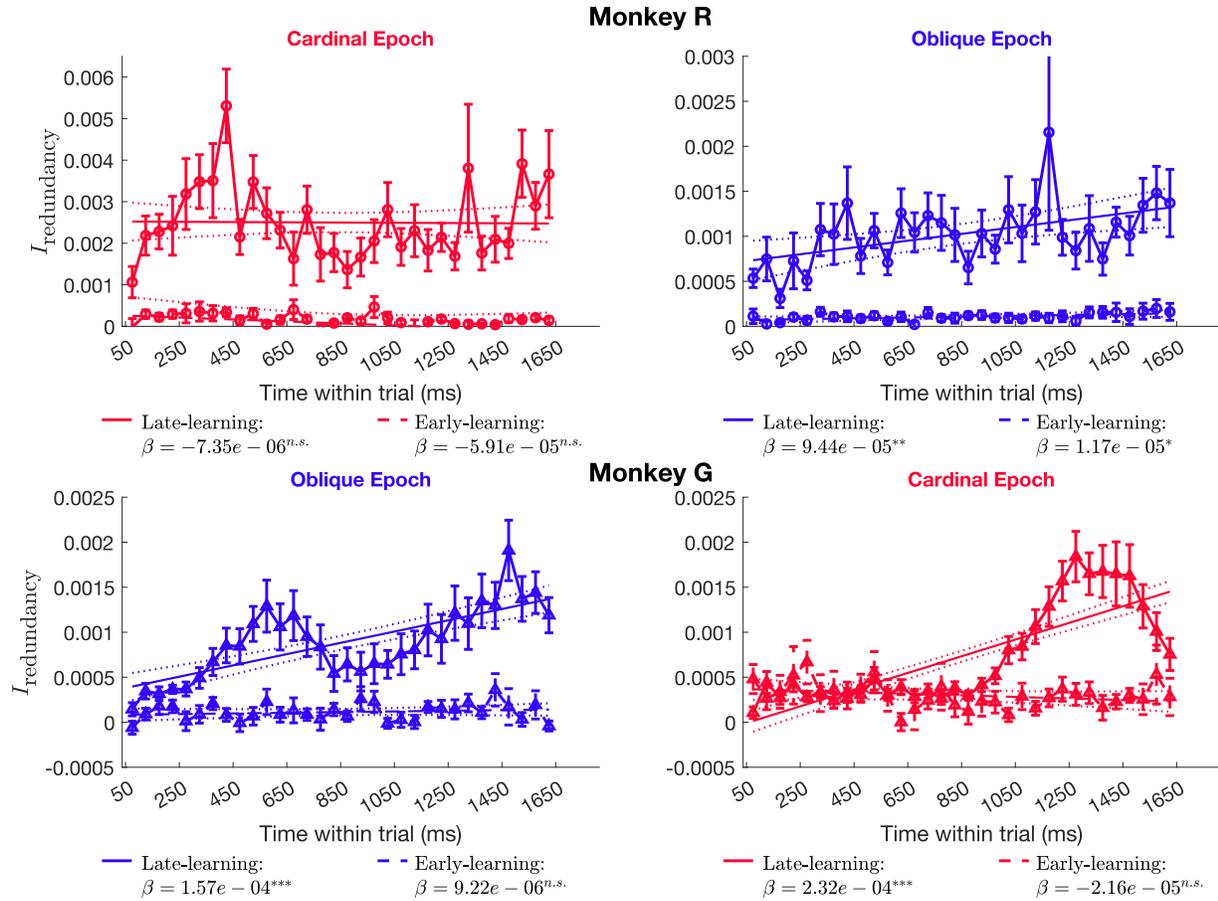
**Fig. S19:** $I_{\text{redundancy}}$ **for each 50 ms time bin within a trial in empirical data.** Same as Fig. 5 except with finer temporal resolution of 50 ms and using the whole neural populations. Top: monkey R. Bottom: monkey G. Linear regression statistics: monkey R, cardinal task: Late-learning: $\beta = -7.3 \times 10^{-6}$, $p = 0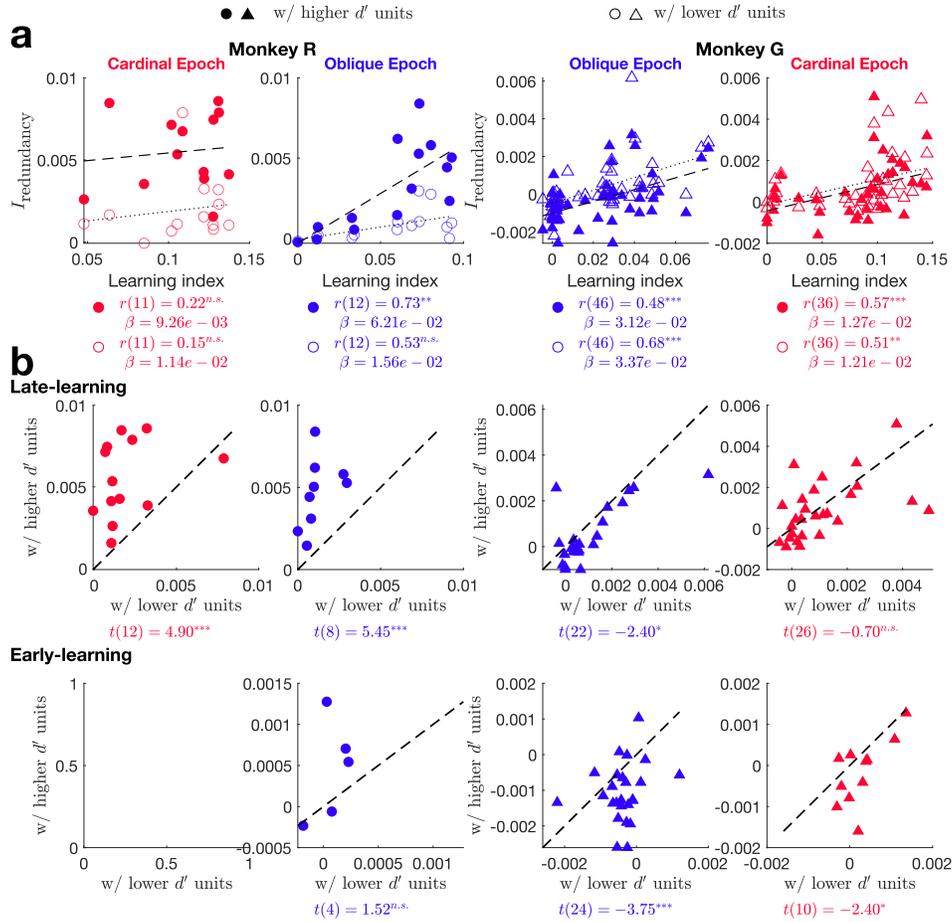.91$; Early-learning: $\beta = -5.9 \times 10^{-5}$, $p = 0.34$; monkey R, oblique task: Late-learning: $\beta = 9.4 \times 10^{-5}$, $p = 0.0023$; Early-learning: $\beta = 1.2 \times 10^{-5}$, $p = 0.027$; monkey G, oblique task: Late-learning: $\beta = 1.6 \times 10^{-4}$, $p = 4.9 \times 10^{-13}$; Early-learning: $\beta = 9.2 \times 10^{-6}$, $p = 0.35$; monkey G, cardinal task: Late-learning: $\beta = 2.3 \times 10^{-4}$, $p = 6.5 \times 10^{-39}$; Early-learning: $\beta = -2.2 \times 10^{-5}$, $p = 0.18$.

**Fig. S20: Information redundancy within population of high or low stimulus $d'$ based on passive-viewing data.** Same analysis as in figure S8 but now the units are split by their task sensitivity computed from passive viewing data, as opposed to task data as in figure S8. Size of population was again kept constant within each epoch (monkey R, cardinal: N = 8, monkey R, oblique: N = 9; monkey G, oblique: N = 22; monkey G cardinal: N = 27).
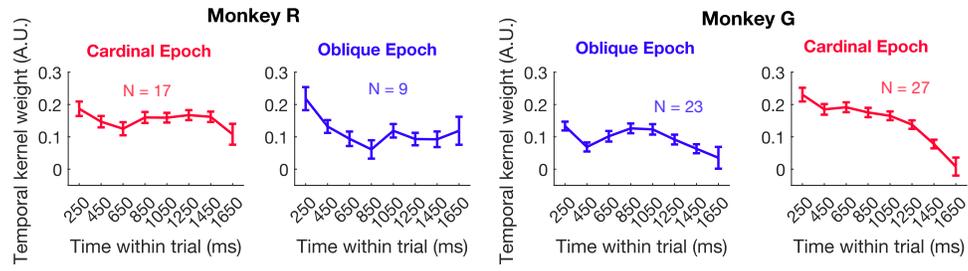
**Fig. S21: Temporal kernels of each epoch.** The lines represent the average across late-learning sessions, while the error bars indicate the standard error of the mean across these sessions.
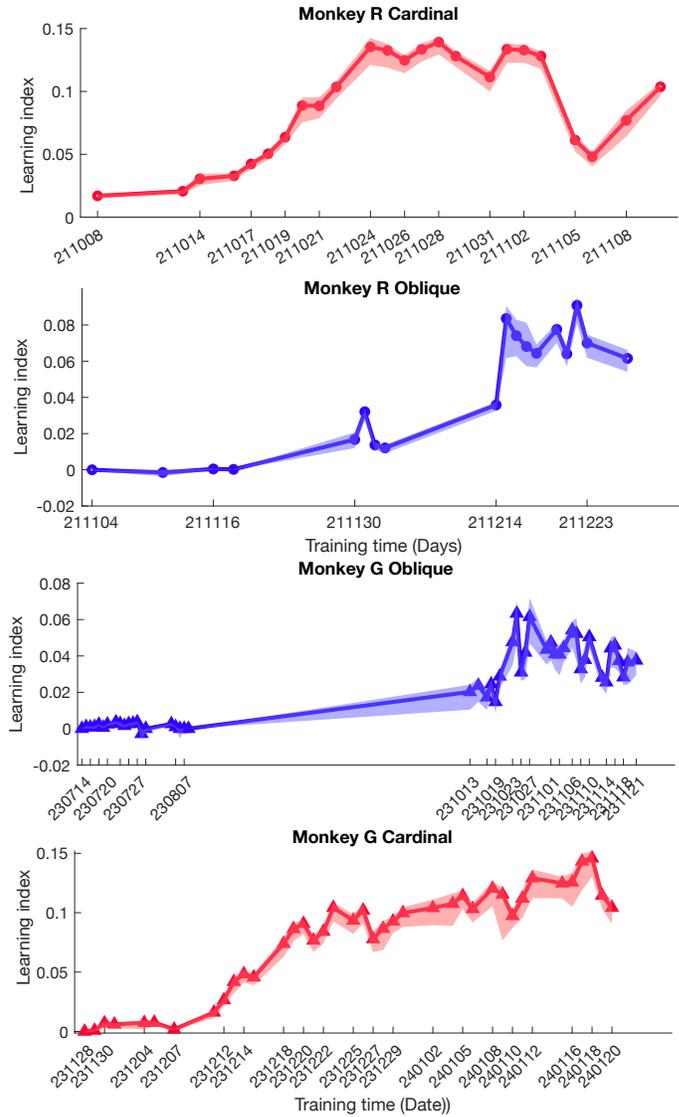
**Fig. S22: Time course of behavioral performance with real calendar dates.** The x-axis represents calendar days, with the spacing reflecting the true time interval (in days) between sessions. For the monkey G oblique epoch, two months of training sessions were excluded from the analysis due to the following reason: the animal was initially trained with the stimulus presented in the left hemifield but struggled to generalize when the stimulus was moved to the right hemifield after implantation surgery to his left V4. To aid in performance recovery, we modified the stimulus size and location during those sessions, leading to inconsistencies in visual stimulus parameters.

**Table S1:** $p-$values from permutation-based test that compares the relationship between $I_{\text{redundancy}}$ and learning index for higher stimulus $d'$ units and that for lower $d'$ units.

|  | Monkey R, cardinal | Monkey R, oblique | Monkey G, oblique | Monkey G, cardinal |
|---|---|---|---|---|
| **$d'$ during task** | | | | |
| Correlation coefficients | $p = 0.032$ | $p = 0.393$ | $p = 0.014$ | $p = 0.098$ |
| Regression coefficients | $p = 0.023$ | $p = 0.007$ | $p < 0.0001$ | $p = 0.021$ |
| **$d'$ during passive viewing** | | | | |
| Correlation coefficients | $p = 0.451$ | $p = 0.260$ | $p = 0.937$ | $p = 0.290$ |
| Regression coefficients | $p = 0.521$ | $p = 0.007$ | $p = 0.660$ | $p = 0.461$ |

**Table S2:** Paired t-test that compares $I_{\text{redundancy}}$ within higher $d'$ units with that within lower $d'$ units

|  | Monkey R, cardinal | Monkey R, oblique | Monkey G, oblique | Monkey G, cardinal |
|---|---|---|---|---|
| **$d'$ during task** | | | | |
| Late-learning | $t(16) = 3.98$ $p = 0.0011$ | $t(8) = 5.90$ $p = 3.6 \times 10^{-4}$ | $t(22) = 3.35$ $p = 0.0029$ | $t(26) = 3.05$ $p = 0.0052$ |
| Early-learning | $t(5) = 0.40$ $p = 0.70$ | $t(4) = 1.73$ $p = 0.16$ | $t(24) = -4.38$ $p = 2.0 \times 10^{-4}$ | $t(10) = 1.10$ $p = 0.30$ |
| **$d'$ during passive viewing** | | | | |
| Late-learning | $t(12) = 4.90$ $p = 3.7 \times 10^{-4}$ | $t(8) = 5.45$ $p = 6.1 \times 10^{-4}$ | $t(22) = -2.40$ $p = 0.025$ | $t(26) = -0.70$ $p = 0.49$ |
| Early-learning | No $d'$ data | $t(4) = 1.52$ $p = 0.20$ | $t(24) = -3.75$ $p = 9.9 \times 10^{-4}$ | $t(10) = -2.40$ $p = 0.037$ |

**Table S3: Spearman correlation between learning index and eye dynamic variables.** Significant correlations ($p < 0.05$) are marked bold.

|  | Monkey R, cardinal | Monkey R, oblique | Monkey G, oblique | Monkey G, cardinal |
|---|---|---|---|---|
| x-pos-avg | $r(20) = 0.35,$ $p = 0.12$ | $r(12) = 0.09,$ $p = 0.75$ | $r(46) = -0.23,$ $p = 0.11$ | $r(35) = -0.01,$ $p = 0.94$ |
| y-pos-avg | $r(20) = -0.37,$ $p = 0.094$ | $r(12) = 0.11,$ $p = 0.70$ | $r(46) = -0.21,$ $p = 0.16$ | $r(35) = -0.00,$ $p = 0.98$ |
| x-vel-avg | $r(20) = -0.25,$ $p = 0.26$ | $r(12) = -0.20,$ $p = 0.48$ | $r(46) = -0.12,$ $p = 0.43$ | $\mathbf{r(35) = 0.44,}$ $p = 0.0074$ |
| y-vel-avg | $r(20) = 0.16,$ $p = 0.46$ | $r(12) = -0.26,$ $p = 0.37$ | $\mathbf{r(46) = 0.65,}$ $p = 1.0 \times 10^{-6}$ | $r(35) = 0.03,$ $p = 0.87$ |
| pupil-avg | $r(20) = 0.23,$ $p = 0.31$ | $r(12) = 0.23,$ $p = 0.44$ | $\mathbf{r(46) = -0.48,}$ $p = 6.1 \times 10^{-4}$ | $\mathbf{r(35) = -0.62,}$ $p = 6.3 \times 10^{-5}$ |
| x-pos-var | $\mathbf{r(20) = 0.80,}$ $p = 7.5 \times 10^{-6}$ | $r(12) = -0.50,$ $p = 0.069$ | $\mathbf{r(46) = 0.70,}$ $p = 1.2 \times 10^{-7}$ | $\mathbf{r(35) = 0.47,}$ $p = 0.0038$ |
| y-pos-var | $r(20) = -0.17,$ $p = 0.45$ | $r(12) = -0.35,$ $p = 0.23$ | $\mathbf{r(46) = 0.53,}$ $p = 1.5 \times 10^{-4}$ | $\mathbf{r(35) = -0.44,}$ $p = 0.0068$ |
| x-vel-var | $\mathbf{r(20) = 0.45,}$ $p = 0.037$ | $r(12) = 0.21,$ $p = 0.47$ | $\mathbf{r(46) = 0.55,}$ $p = 5.8 \times 10^{-5}$ | $\mathbf{r(35) = -0.40,}$ $p = 0.016$ |
| y-vel-var | $\mathbf{r(20) = 0.56,}$ $p = 0.0073$ | $r(12) = 0.24,$ $p = 0.41$ | $r(46) = -0.16,$ $p = 0.29$ | $\mathbf{r(35) = -0.56,}$ $p = 4.2 \times 10^{-4}$ |
| pupil-var | $r(20) = -0.34,$ $p = 0.12$ | $r(12) = -0.24,$ $p = 0.42$ | $\mathbf{r(46) = 0.53,}$ $p = 1.5 \times 10^{-4}$ | $r(35) = -0.25,$ $p = 0.13$ |

**Table S4: Partial correlation between information redundancy and learning index controlling for eye dynamic variables.** Each item shows the Spearman correlation between $I_{\text{redundancy}}$ and learning index for one epoch (column) with one eye dynamic variable (row) as the control variable. In this table, all correlations remain significantly positive.

| | Monkey R, cardinal | Monkey R, oblique | Monkey G, oblique | Monkey G, cardinal |
|---|---|---|---|---|
| x-pos-avg | $r(19) = 0.70$, $p = 4.1 \times 10^{-4}$ | $r(11) = 0.76$, $p = 0.0028$ | $r(45) = 0.61$, $p = 5.5 \times 10^{-6}$ | $r(34) = 0.66$, $p = 1.2 \times 10^{-5}$ |
| y-pos-avg | $r(19) = 0.69$, $p = 4.8 \times 10^{-4}$ | $r(11) = 0.70$, $p = 0.0077$ | $r(45) = 0.59$, $p = 1.5 \times 10^{-5}$ | $r(34) = 0.67$, $p = 8.8 \times 10^{-6}$ |
| x-vel-avg | $r(19) = 0.73$, $p = 2.0 \times 10^{-4}$ | $r(11) = 0.76$, $p = 0.0024$ | $r(45) = 0.61$, $p = 6.5 \times 10^{-6}$ | $r(34) = 0.59$, $p = 1.4 \times 10^{-4}$ |
| y-vel-avg | $r(19) = 0.73$, $p = 1.7 \times 10^{-4}$ | $r(11) = 0.68$, $p = 0.011$ | $r(45) = 0.46$, $p = 0.0012$ | $r(34) = 0.66$, $p = 1.2 \times 10^{-5}$ |
| pupil-avg | $r(19) = 0.72$, $p = 2.2 \times 10^{-4}$ | $r(11) = 0.68$, $p = 0.010$ | $r(45) = 0.47$, $p = 7.5 \times 10^{-4}$ | $r(34) = 0.48$, $p = 0.0033$ |
| x-pos-var | $r(19) = 0.53$, $p = 0.013$ | $r(11) = 0.73$, $p = 0.0045$ | $r(45) = 0.37$, $p = 0.011$ | $r(34) = 0.56$, $p = 4.0 \times 10^{-4}$ |
| y-pos-var | $r(19) = 0.76$, $p = 6.4 \times 10^{-5}$ | $r(11) = 0.71$, $p = 0.0064$ | $r(45) = 0.51$, $p = 2.8 \times 10^{-4}$ | $r(34) = 0.63$, $p = 3.3 \times 10^{-5}$ |
| x-vel-var | $r(19) = 0.70$, $p = 4.3 \times 10^{-4}$ | $r(11) = 0.69$, $p = 0.0089$ | $r(45) = 0.37$, $p = 0.011$ | $r(34) = 0.63$, $p = 4.5 \times 10^{-5}$ |
| y-vel-var | $r(19) = 0.64$, $p = 0.0018$ | $r(11) = 0.70$, $p = 0.0073$ | $r(45) = 0.61$, $p = 4.9 \times 10^{-6}$ | $r(34) = 0.66$, $p = 1.3 \times 10^{-5}$ |
| pupil-var | $r(19) = 0.73$, $p = 2.0 \times 10^{-4}$ | $r(11) = 0.73$, $p = 0.0046$ | $r(45) = 0.52$, $p = 1.8 \times 10^{-4}$ | $r(34) = 0.66$, $p = 1.1 \times 10^{-5}$ |

**Table S5: Partial correlation between $I_{\text{redundancy}}$ and learning index with principal components as control variables.** In each column, we report the largest number of principal components (first row) with which the partial correlation between $I_{\text{redundancy}}$ and learning index (third row) was still significant, and the variance they could explain (second row) for the ten eye dynamic quantities.

|  | Monkey R, cardinal | Monkey R, oblique | Monkey G, oblique | Monkey G, cardinal |
|---|---|---|---|---|
| Largest PC number | 8 | 7 | 7 | 3 |
| Explained variance | 98% | 97% | 92% | 59% |
| Partial correlation | $r(12) = 0.55$ <br> $p = 0.042$ | $\rho(5) = 0.84$ <br> $p = 0.018$ | $r(39) = 0.42$ <br> $p = 0.0069$ | $r(32) = 0.37$ <br> $p = 0.032$ |

**Table S6: Decoding choice with eye dynamic variable.** The first row shows Spearman correlations between the learning index and decoding accuracy of choice from eye dynamic variables. The second row shows the Spearman correlation between $I_{\text{redundancy}}$ and learning index, controlling for the choice decoding accuracy.

| | Monkey R, cardinal | Monkey R, oblique | Monkey G, oblique | Monkey G, cardinal |
|---|---|---|---|---|
| corr (learning index, choice decoding accuracy) | $r(20) = -0.05$, $p = 0.81$ | $r(12) = 0.42$, $p = 0.13$ | $r(46) = 0.68$, $p = 3.7 \times 10^{-7}$ | $r(35) = 0.34$, $p = 0.042$ |
| Partial correlation ($I_{\text{redundancy}}$, learning index \|choice decoding accuracy) | $r(19) = 0.74$, $p = 1.1 \times 10^{-4}$ | $r(11) = 0.78$, $p = 0.0015$ | $r(45) = 0.40$, $p = 0.0048$ | $r(34) = 0.69$, $p = 2.8 \times 10^{-6}$ |

**Table S7: Psychometric and populational neurometric threshold late in learning.** We report the coherence levels for which the animal reached 75% accuracy for each task (Mean ± s.t.d. across late-learning sessions). The third row: the ratios between neurometric and psychometric thresholds.

|  | Monkey R, cardinal | Monkey R, oblique | Monkey G, oblique | Monkey G, cardinal |
|---|---|---|---|---|
| Psychometric threshold | 3.29 ± 1.08% | 5.77 ± 1.90% | 17.49 ± 5.35% | 7.60 ± 1.41% |
| Neurometric threshold | 4.52 ± 1.75% | 5.20 ± 1.16% | 11.78 ± 8.33% | 10.09 ± 2.19% |
| Population N/P Ratio | 1.45 ± 0.45 | 0.90 ± 0.14 | 0.69 ± 0.32 | 1.35 ± 0.32 |

**References and Notes**

1. A. J. Parker, W. T. Newsome, Sense and the single neuron: Probing the physiology of perception. *Annu. Rev. Neurosci.* **21**, 227–277 (1998). doi:10.1146/annurev.neuro.21.1.227 Medline

2. J. J. DiCarlo, D. D. Cox, Untangling invariant object recognition. *Trends Cogn. Sci.* **11**, 333–341 (2007). doi:10.1016/j.tics.2007.06.010 Medline

3. N. C. Rust, J. J. DiCarlo, Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *J. Neurosci.* **30**, 12978–12995 (2010). doi:10.1523/JNEUROSCI.0179-10.2010 Medline

4. F. Attneave, Some informational aspects of visual perception. *Psychol. Rev.* **61**, 183–193 (1954). doi:10.1037/h0054663 Medline

5. H. B. Barlow, in *Sensory Communication*, W. A. Rosenblith, Ed. (1961), pp. 217–233.

6. E. P. Simoncelli, B. A. Olshausen, Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001). doi:10.1146/annurev.neuro.24.1.1193 Medline

7. J. F. Mitchell, K. A. Sundberg, J. H. Reynolds, Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* **63**, 879–888 (2009). doi:10.1016/j.neuron.2009.09.013 Medline

8. M. R. Cohen, J. H. Maunsell, Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* **12**, 1594–1600 (2009). doi:10.1038/nn.2439 Medline

9. R. Moreno-Bote, J. Beck, I. Kanitscheider, X. Pitkow, P. Latham, A. Pouget, Information-limiting correlations. *Nat. Neurosci.* **17**, 1410–1417 (2014). doi:10.1038/nn.3807 Medline

10. A. Kohn, R. Coen-Cagli, I. Kanitscheider, A. Pouget, Correlations and neuronal population information. *Annu. Rev. Neurosci.* **39**, 237–256 (2016). doi:10.1146/annurev-neuro-070815-013851 Medline

11. A. M. Ni, D. A. Ruff, J. J. Alberts, J. Symmonds, M. R. Cohen, Learning and attention reveal a general relationship between population activity and behavior. *Science* **359**, 463–465 (2018). doi:10.1126/science.aao0284 Medline

12. H. von Helmholtz, *Handbuch der Physiologischen Optik* (Leopold Voss, 1867).

13. A. Yuille, D. Kersten, Vision as Bayesian inference: Analysis by synthesis? *Trends Cogn. Sci.* **10**, 301–308 (2006). doi:10.1016/j.tics.2006.05.002 Medline

14. B. Peters, J. J. DiCarlo, T. Gureckis, R. Haefner, L. Isik, J. Tenenbaum, T. Konkle, T. Naselaris, K. Stachenfeld, Z. Tavares, D. Tsao, I. Yildirim, N. Kriegeskorte, How does the primate brain combine generative and discriminative computations in vision? arXiv:2401.06005 [q-bio.NC] (2024).

15. D. J. Felleman, D. C. Van Essen, Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cereb. Cortex* **1**, 1–47 (1991). doi:10.1093/cercor/1.1.1-a

16. T. Moore, M. Zirnsak, Neural mechanisms of selective visual attention. *Annu. Rev. Psychol.* **68**, 47–72 (2017). doi:10.1146/annurev-psych-122414-033400 Medline

17. M. R. Cohen, J. H. Maunsell, Using neuronal populations to study the mechanisms underlying spatial and feature attention. *Neuron* **70**, 1192–1204 (2011). doi:10.1016/j.neuron.2011.04.029 Medline

18. D. A. Ruff, M. R. Cohen, Attention can either increase or decrease spike count correlations in visual cortex. *Nat. Neurosci.* **17**, 1591–1597 (2014). doi:10.1038/nn.3835 Medline

19. A. C. Snyder, M. J. Morais, M. A. Smith, Dynamics of excitatory and inhibitory networks are differentially altered by selective attention. *J. Neurophysiol.* **116**, 1807–1820 (2016). doi:10.1152/jn.00343.2016 Medline

20. A. S. Nandy, J. J. Nassi, J. H. Reynolds, Laminar organization of attentional modulation in macaque visual area V4. *Neuron* **93**, 235–246 (2017). doi:10.1016/j.neuron.2016.11.029 Medline

21. G. H. Denfield, A. S. Ecker, T. J. Shinn, M. Bethge, A. S. Tolias, Attentional fluctuations induce shared variability in macaque primary visual cortex. *Nat. Commun.* **9**, 2654 (2018). doi:10.1038/s41467-018-05123-6 Medline

22. T. S. Lee, D. Mumford, Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A* **20**, 1434–1448 (2003). doi:10.1364/JOSAA.20.001434 Medline

23. R. M. Haefner, P. Berkes, J. Fiser, Perceptual decision-making as probabilistic inference by neural sampling. *Neuron* **90**, 649–660 (2016). doi:10.1016/j.neuron.2016.03.020 Medline

24. X. Pitkow, D. E. Angelaki, Inference in the brain: Statistics flowing in redundant population codes. *Neuron* **94**, 943–953 (2017). doi:10.1016/j.neuron.2017.05.028 Medline

25. C. Findling, F. Hubert, International Brain Laboratory, L. Acerbi, B. Benson, J. Benson, D. Birman, N. Bonacchi, E. K. Buchanan, S. Bruijns, M. Carandini, J. A. Catarino, G. A. Chapuis, A. K. Churchland, Y. Dan, F. Davatolhagh, E. E. J. DeWitt, T. A. Engel, M. Fabbri, M. A. Faulkner, I. R. Fiete, L. Freitas-Silva, B. Gerçek, K. D. Harris, M. Häusser, S. B. Hofer, F. Hu, J. M. Huntenburg, A. Khanal, C. Krasniak, C. Langdon, C. A. Langfield, P. E. Latham, P. Y. P. Lau, Z. Mainen, G. T. Meijer, N. J. Miska, T. D. Mrsic-Flogel, J.-P. Noel, K. Nylund, A. Pan-Vazquez, L. Paninski, J. Pillow, C. Rossant, N. Roth, R. Schaeffer, M. Schartner, Y. Shi, K. Z. Socha, N. A. Steinmetz, K. Svoboda, C. Tessereau, A. E. Urai, M. J. Wells, S. J. West, M. R. Whiteway, O. Winter, I. B. Witten, A. Zador, Y. Zhang, P. Dayan, A. Pouget, Brain-wide representations of prior information in mouse decision-making. *Nature* **645**, 192–200 (2025). doi:10.1038/s41586-025-09226-1

26. R. D. Lange, R. M. Haefner, Task-induced neural covariability as a signature of approximate Bayesian learning and inference. *PLOS Comput. Biol.* **18**, e1009557 (2022). doi:10.1371/journal.pcbi.1009557 Medline

27. B. B. Averbeck, P. E. Latham, A. Pouget, Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7**, 358–366 (2006). doi:10.1038/nrn1888 Medline

28. H. Nienborg, B. G. Cumming, Decision-related activity in sensory neurons may depend on the columnar architecture of cerebral cortex. *J. Neurosci.* **34**, 3579–3585 (2014). doi:10.1523/JNEUROSCI.2340-13.2014 Medline

29. A. G. Bondy, R. M. Haefner, B. G. Cumming, Feedback determines the structure of correlated variability in primary visual cortex. *Nat. Neurosci.* **21**, 598–606 (2018). doi:10.1038/s41593-018-0089-1 Medline

30. R. D. Lange, C. Gómez-Laberge, V. K. Berezovskii, A. Pletenev, A. Sherdil, T. Hartmann, R. M. Haefner, R. T. Born, Weak evidence for neural correlates of task-switching in macaque V1. *J. Neurophysiol.* **129**, 1021–1044 (2023). doi:10.1152/jn.00085.2022 Medline

31. M. Sanayei, X. Chen, D. Chicharro, C. Distler, S. Panzeri, A. Thiele, Perceptual learning of fine contrast discrimination changes neuronal tuning and population coding in macaque V4. *Nat. Commun.* **9**, 4238 (2018). doi:10.1038/s41467-018-06698-w Medline

32. O. I. Rumyantsev, J. A. Lecoq, O. Hernandez, Y. Zhang, J. Savall, R. Chrapkiewicz, J. Li, H. Zeng, S. Ganguli, M. J. Schnitzer, Fundamental bounds on the fidelity of sensory cortical coding. *Nature* **580**, 100–105 (2020). doi:10.1038/s41586-020-2130-2 Medline

33. M. Kafashan, A. W. Jaffe, S. N. Chettih, R. Nogueira, I. Arandia-Romero, C. D. Harvey, R. Moreno-Bote, J. Drugowitsch, Scaling of sensory information in large neural populations shows signatures of information-limiting correlations. *Nat. Commun.* **12**, 473 (2021). doi:10.1038/s41467-020-20722-y Medline

34. I. Kanitscheider, R. Coen-Cagli, A. Pouget, Origin of information-limiting noise correlations. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6973–E6982 (2015). doi:10.1073/pnas.1508738112 Medline

35. X. Pitkow, S. Liu, D. E. Angelaki, G. C. DeAngelis, A. Pouget, How can single sensory neurons predict behavior? *Neuron* **87**, 411–423 (2015). doi:10.1016/j.neuron.2015.06.033 Medline

36. E. Mach, Über das Sehen von Lagen und Winkeln durch die Bewegung des Auges. *Sitzungsber. Kais. Akad. Wiss.* **43**, 215–224 (1861).

37. S. Appelle, Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychol. Bull.* **78**, 266–278 (1972). doi:10.1037/h0033117 Medline

38. R. D. Lange, A. Chattoraj, J. M. Beck, J. L. Yates, R. M. Haefner, A confirmation bias in perceptual decision-making due to hierarchical approximate inference. *PLOS Comput. Biol.* **17**, e1009517 (2021). doi:10.1371/journal.pcbi.1009517 Medline

39. A. Kohn, Visual adaptation: Physiology, mechanisms, and functional benefits. *J. Neurophysiol.* **97**, 3155–3164 (2007). doi:10.1152/jn.00086.2007 Medline

40. D. H. Ackley, G. E. Hinton, T. J. Sejnowski, A learning algorithm for Boltzmann machines. *Cogn. Sci.* **9**, 147–169 (1985). doi:10.1016/S0364-0213(85)80012-4

41. J. Fiser, P. Berkes, G. Orbán, M. Lengyel, Statistically optimal perception and learning: From behavior to neural representations. *Trends Cogn. Sci.* **14**, 119–130 (2010). doi:10.1016/j.tics.2010.01.003 Medline

42. J. I. Gold, M. N. Shadlen, The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007). [doi:10.1146/annurev.neuro.29.051605.113038](doi:10.1146/annurev.neuro.29.051605.113038) [Medline](Medline)

43. T. D. Hanks, M. E. Mazurek, R. Kiani, E. Hopp, M. N. Shadlen, Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *J. Neurosci.* **31**, 6339–6352 (2011). [doi:10.1523/JNEUROSCI.5613-10.2011](doi:10.1523/JNEUROSCI.5613-10.2011) [Medline](Medline)

44. Y. Niv, Learning task-state representations. *Nat. Neurosci.* **22**, 1544–1553 (2019). [doi:10.1038/s41593-019-0470-8](doi:10.1038/s41593-019-0470-8) [Medline](Medline)

45. V. Rao, G. C. DeAngelis, L. H. Snyder, Neural correlates of prior expectations of motion in the lateral intraparietal and middle temporal areas. *J. Neurosci.* **32**, 10063–10074 (2012). [doi:10.1523/JNEUROSCI.5948-11.2012](doi:10.1523/JNEUROSCI.5948-11.2012) [Medline](Medline)

46. R. Nogueira, J. M. Abolafia, J. Drugowitsch, E. Balaguer-Ballester, M. V. Sanchez-Vives, R. Moreno-Bote, Lateral orbitofrontal cortex anticipates choices and integrates prior with current information. *Nat. Commun.* **8**, 14823 (2017). [doi:10.1038/ncomms14823](doi:10.1038/ncomms14823) [Medline](Medline)

47. P. Berkes, G. Orbán, M. Lengyel, J. Fiser, Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* **331**, 83–87 (2011). [doi:10.1126/science.1195870](doi:10.1126/science.1195870) [Medline](Medline)

48. R. Jardri, S. Duverne, A. S. Litvinova, S. Denève, Experimental evidence for circular inference in schizophrenia. *Nat. Commun.* **8**, 14218 (2017). [doi:10.1038/ncomms14218](doi:10.1038/ncomms14218) [Medline](Medline)

49. J. Park, S. Kim, H. R. Kim, J. Lee, Prior expectation enhances sensorimotor behavior by modulating population tuning and subspace activity in sensory cortex. *Sci. Adv.* **9**, eadg4156 (2023). [doi:10.1126/sciadv.adg4156](doi:10.1126/sciadv.adg4156) [Medline](Medline)

50. E. Zohary, M. N. Shadlen, W. T. Newsome, Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* **370**, 140–143 (1994). [doi:10.1038/370140a0](doi:10.1038/370140a0) [Medline](Medline)

51. L. F. Abbott, P. Dayan, The effect of correlated variability on the accuracy of a population code. *Neural Comput.* **11**, 91–101 (1999). [doi:10.1162/089976699300016827](doi:10.1162/089976699300016827) [Medline](Medline)

52. H. Sompolinsky, H. Yoon, K. Kang, M. Shamir, Population coding in neuronal systems with correlated noise. *Phys. Rev. E* **64**, 051904 (2001). [doi:10.1103/PhysRevE.64.051904](doi:10.1103/PhysRevE.64.051904) [Medline](Medline)

53. A. S. Ecker, P. Berens, A. S. Tolias, M. Bethge, The Effect of Noise Correlations in Populations of Diversely Tuned Neurons. *J. Neurosci.* **31**, 14272–14283 (2011). [doi:10.1523/JNEUROSCI.2539-11.2011](doi:10.1523/JNEUROSCI.2539-11.2011) [Medline](Medline)

54. M. R. Cohen, W. T. Newsome, Context-dependent changes in functional circuitry in visual area MT. *Neuron* **60**, 162–173 (2008). [doi:10.1016/j.neuron.2008.08.007](doi:10.1016/j.neuron.2008.08.007) [Medline](Medline)

55. Y. Gu, S. Liu, C. R. Fetsch, Y. Yang, S. Fok, A. Sunkara, G. C. DeAngelis, D. E. Angelaki, Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* **71**, 750–761 (2011). [doi:10.1016/j.neuron.2011.06.015](doi:10.1016/j.neuron.2011.06.015) [Medline](Medline)

56. T. Z. Luo, J. H. Maunsell, Neuronal modulations in visual cortex are associated with only one of multiple components of attention. *Neuron* **86**, 1182–1188 (2015). doi:10.1016/j.neuron.2015.05.007 Medline

57. A. S. Ecker, P. Berens, G. A. Keliris, M. Bethge, N. K. Logothetis, A. S. Tolias, Decorrelated neuronal firing in cortical microcircuits. *Science* **327**, 584–587 (2010). doi:10.1126/science.1179867 Medline

58. A. S. Ecker, P. Berens, R. J. Cotton, M. Subramaniyan, G. H. Denfield, C. R. Cadwell, S. M. Smirnakis, M. Bethge, A. S. Tolias, State dependence of noise correlations in macaque primary visual cortex. *Neuron* **82**, 235–248 (2014). doi:10.1016/j.neuron.2014.02.006 Medline

59. A. C. Snyder, M. J. Morais, A. Kohn, M. A. Smith, Correlations in V1 are reduced by stimulation outside the receptive field. *J. Neurosci.* **34**, 11222–11227 (2014). doi:10.1523/JNEUROSCI.0762-14.2014 Medline

60. A. C. Snyder, M. J. Morais, C. M. Willis, M. A. Smith, Global network influences on local functional connectivity. *Nat. Neurosci.* **18**, 736–743 (2015). doi:10.1038/nn.3979 Medline

61. A. Umakantha, R. Morina, B. R. Cowley, A. C. Snyder, M. A. Smith, B. M. Yu, Bridging neuronal correlations and dimensionality reduction. *Neuron* **109**, 2740–2754.e12 (2021). doi:10.1016/j.neuron.2021.06.028 Medline

62. T. A. Engel, W. Chaisangmongkon, D. J. Freedman, X.-J. Wang, Choice-correlated activity fluctuations underlie learning of neuronal category representation. *Nat. Commun.* **6**, 6454 (2015). doi:10.1038/ncomms7454 Medline

63. M. R. Nassar, D. Scott, A. Bhandari, Noise correlations for faster and more robust learning. *J. Neurosci.* **41**, 6740–6752 (2021). doi:10.1523/JNEUROSCI.3045-20.2021 Medline

64. C. Haimerl, D. A. Ruff, M. R. Cohen, C. Savin, E. P. Simoncelli, Targeted V1 comodulation supports task-adaptive sensory decisions. *Nat. Commun.* **14**, 7879 (2023). doi:10.1038/s41467-023-43432-7 Medline

65. M. Valente, G. Pica, G. Bondanelli, M. Moroni, C. A. Runyan, A. S. Morcos, C. D. Harvey, S. Panzeri, Correlations enhance the behavioral readout of neural population activity in association cortex. *Nat. Neurosci.* **24**, 975–986 (2021). doi:10.1038/s41593-021-00845-1 Medline

66. A. Chadwick, A. G. Khan, J. Poort, A. Blot, S. B. Hofer, T. D. Mrsic-Flogel, M. Sahani, Learning shapes cortical dynamics to enhance integration of relevant sensory input. *Neuron* **111**, 106–120.e10 (2023). doi:10.1016/j.neuron.2022.10.001 Medline

67. R. Bartolo, R. C. Saunders, A. R. Mitz, B. B. Averbeck, Information-limiting correlations in large neural populations. *J. Neurosci.* **40**, 1668–1678 (2020). doi:10.1523/JNEUROSCI.2072-19.2019 Medline

68. N. C. Rabinowitz, R. L. Goris, M. Cohen, E. P. Simoncelli, Attention stabilizes the shared gain of V4 populations. *eLife* **4**, e08998 (2015). doi:10.7554/eLife.08998 Medline

69. S. Ebrahimi, J. Lecoq, O. Rumyantsev, T. Tasci, Y. Zhang, C. Irimia, J. Li, S. Ganguli, M. J. Schnitzer, Emergent reliability in sensory cortical coding and inter-area communication. *Nature* **605**, 713–721 (2022). [doi:10.1038/s41586-022-04724-y](doi:10.1038/s41586-022-04724-y) [Medline](Medline)

70. A. Schoups, R. Vogels, N. Qian, G. Orban, Practising orientation identification improves orientation coding in V1 neurons. *Nature* **412**, 549–553 (2001). [doi:10.1038/35087601](doi:10.1038/35087601) [Medline](Medline)

71. H. Z. Adab, R. Vogels, Practicing coarse orientation discrimination improves orientation signals in macaque cortical area V4. *Curr. Biol.* **21**, 1661–1666 (2011). [doi:10.1016/j.cub.2011.08.037](doi:10.1016/j.cub.2011.08.037) [Medline](Medline)

72. Y. Yan, M. J. Rasch, M. Chen, X. Xiang, M. Huang, S. Wu, W. Li, Perceptual training continuously refines neuronal population codes in primary visual cortex. *Nat. Neurosci.* **17**, 1380–1387 (2014). [doi:10.1038/nn.3805](doi:10.1038/nn.3805) [Medline](Medline)

73. A. G. Khan, J. Poort, A. Chadwick, A. Blot, M. Sahani, T. D. Mrsic-Flogel, S. B. Hofer, Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. *Nat. Neurosci.* **21**, 851–859 (2018). [doi:10.1038/s41593-018-0143-z](doi:10.1038/s41593-018-0143-z) [Medline](Medline)

74. K. H. Britten, M. N. Shadlen, W. T. Newsome, J. A. Movshon, The analysis of visual motion: A comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**, 4745–4765 (1992). [doi:10.1523/JNEUROSCI.12-12-04745.1992](doi:10.1523/JNEUROSCI.12-12-04745.1992) [Medline](Medline)

75. S. J. Prince, A. D. Pointon, B. G. Cumming, A. J. Parker, The precision of single neuron responses in cortical area V1 during stereoscopic depth judgments. *J. Neurosci.* **20**, 3387–3400 (2000). [doi:10.1523/JNEUROSCI.20-09-03387.2000](doi:10.1523/JNEUROSCI.20-09-03387.2000) [Medline](Medline)

76. S. R. Allred, B. Jagadeesh, Quantitative comparison between neural response in macaque inferotemporal cortex and behavioral discrimination of photographic images. *J. Neurophysiol.* **98**, 1263–1277 (2007). [doi:10.1152/jn.00016.2007](doi:10.1152/jn.00016.2007) [Medline](Medline)

77. N. T. Markov, J. Vezoli, P. Chameau, A. Falchier, R. Quilodran, C. Huissoud, C. Lamy, P. Misery, P. Giroud, S. Ullman, P. Barone, C. Dehay, K. Knoblauch, H. Kennedy, Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *J. Comp. Neurol.* **522**, 225–259 (2014). [doi:10.1002/cne.23458](doi:10.1002/cne.23458) [Medline](Medline)

78. D. D. Stettler, A. Das, J. Bennett, C. D. Gilbert, Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron* **36**, 739–750 (2002). [doi:10.1016/S0896-6273(02)01029-2](doi:10.1016/S0896-6273(02)01029-2) [Medline](Medline)

79. S. Liu, A. Pletenev, R. M. Haefner, A. Snyder, Chronic Utah array recordings from macaque V4 during learning of orientation discrimination tasks, Zenodo (2025). [doi:10.5281/zenodo.17460460](doi:10.5281/zenodo.17460460)

80. National Research Council, *Guide for the Care and Use of Laboratory Animals* (National Academies Press, ed. 8, 2011).

81. K. A. Phillips, K. L. Bales, J. P. Capitanio, A. Conley, P. W. Czoty, B. A. 't Hart, W. D. Hopkins, S.-L. Hu, L. A. Miller, M. A. Nader, P. W. Nathanielsz, J. Rogers, C. A. Shively, M. L. Voytko, Why primate models matter. *Am. J. Primatol.* **76**, 801–827 (2014). [doi:10.1002/ajp.22281](doi:10.1002/ajp.22281) [Medline](Medline)

82. A. C. Snyder, B. M. Yu, M. A. Smith, Distinct population codes for attention in the absence and presence of visual stimulation. *Nat. Commun.* **9**, 4382 (2018). doi:10.1038/s41467-018-06754-5 Medline

83. A. C. Snyder, B. M. Yu, M. A. Smith, A Stable Population Code for Attention in Prefrontal Cortex Leads a Dynamic Attention Code in Visual Cortex. *J. Neurosci.* **41**, 9163–9176 (2021). doi:10.1523/JNEUROSCI.0608-21.2021 Medline

84. B. R. Cowley, M. A. Smith, A. Kohn, B. M. Yu, Stimulus-Driven Population Activity Patterns in Macaque Primary Visual Cortex. *PLOS Comput. Biol.* **12**, e1005185 (2016). doi:10.1371/journal.pcbi.1005185 Medline

85. P. Gao, S. Ganguli, On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.* **32**, 148–155 (2015). doi:10.1016/j.conb.2015.04.003

86. Smith Laboratory, Spikesort (A MATLAB-based spike sorter for NEV files), Github (2024); https://github.com/SmithLabNeuro/spikesort.

87. H. H. Schütt, S. Harmeling, J. H. Macke, F. A. Wichmann, Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Res.* **122**, 105–123 (2016). doi:10.1016/j.visres.2016.02.002 Medline

88. I. Kanitscheider, R. Coen-Cagli, A. Kohn, A. Pouget, Measuring Fisher information accurately in correlated neural populations. *PLOS Comput. Biol.* **11**, e1004218 (2015). doi:10.1371/journal.pcbi.1004218 Medline

89. P. Schober, C. Boer, L. A. Schwarte, Correlation coefficients: Appropriate use and interpretation. *Anesth. Analg.* **126**, 1763–1768 (2018). doi:10.1213/ANE.0000000000002864 Medline

90. B. A. Olshausen, D. J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996). doi:10.1038/381607a0 Medline

91. R. C. Kelly, M. A. Smith, J. M. Samonds, A. Kohn, A. B. Bonds, J. A. Movshon, T. S. Lee, Comparison of recordings from microelectrode arrays and single electrodes in the visual cortex. *J. Neurosci.* **27**, 261–264 (2007). doi:10.1523/JNEUROSCI.4906-06.2007 Medline