J Neurophysiol 121: 115–130, 2019.
First published November 7, 2018; doi:10.1152/jn.00503.2018.

RESEARCH ARTICLE | *Sensory Processing*

Rethinking assumptions about how trial and nuisance variability impact neural task performance in a fast-processing regime

Noam Roth and Nicole C. Rust
Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania

Submitted 25 July 2018; accepted in final form 1 November 2018

Roth N, Rust NC. Rethinking assumptions about how trial and nuisance variability impact neural task performance in a fast-processing regime. *J Neurophysiol* 121: 115–130, 2019. First published November 7, 2018; doi:10.1152/jn.00503.2018.—Task performance is determined not only by the amount of task-relevant signal present in our brains but also by the presence of noise, which can arise from multiple sources. Internal noise, or "trial variability," manifests as trial-by-trial variations in neural responses under seemingly identical conditions. External factors can also translate into noise, particularly when a task requires extraction of a particular type of information from our environment amid changes in other task-irrelevant "nuisance" parameters. To better understand how signal, trial variability, and nuisance variability combine to determine neural task performance, we explored their interactions, both in simulation and when applied to recorded neural data. This exploration revealed that trial variability is typically larger than a neuron's task-relevant signal for tasks with fast reaction times, where spike count integration windows are short. In this low signal-to-trial variability regime, nuisance variability has the counterintuitive property of having a negligible impact on single-neuron task performance, even when it dominates the task-relevant signal. The inconsequential impact of nuisance variability on individual neurons also extends to descriptions of population performance, under the assumption that both trial and nuisance variability are uncorrelated between neurons. These results demonstrate that some basic intuitions about neural coding are misguided in the context of a fast-processing, low-spike-count regime.

NEW & NOTEWORTHY Many everyday tasks require us to extract specific information from our environment while ignoring other things. When the neurons in our brains that carry task-relevant signals are also modulated by task-irrelevant "nuisance" information, nuisance modulation is expected to act as performance-limiting noise. Using both simulated and recorded neural data, we demonstrate that these intuitions are misguided when the brain operates in a fast-processing, low-spike-count regime, where nuisance variability is largely inconsequential for performance.

neural performance; noise; nuisance variability; population coding; trial variability

INTRODUCTION

One overarching goal of systems neuroscience is to establish the relationship between an organism's neural activity and its behavior. One commonly employed approach in this effort is to formulate behavior in terms of a specific task that the organism has to solve and to compare behavioral performance on this task with the neural performance of individual neurons and/or neural populations (e.g., reviewed by Parker and Newsome 1998; Ruff et al. 2018). When viewed from the perspective of individual neurons, neural task performance is proportional to the amount of neural signal reflected in a neuron's responses (Fig. 1*A*) and inversely proportional to the amount of noise, which can arise from multiple sources. Internal noise, or "trial variability," manifests as trial-by-trial variations in neural responses under seemingly identical conditions (Fig. 1*A*). External factors can also translate into noise, particularly when a task requires extraction of a particular type of information from our environment amid changes in other task-irrelevant, nuisance parameters (Fig. 1*B*; Haefner and Bethge 2010; Kim et al. 2016). Stated differently, for any given task neurons in a brain area may be modulated by multiple experimental variables, but when viewed from the perspective of task performance one type of modulation reflects the task-relevant signal, whereas other types of modulation act as noise. Because our intuitions about how neural activity gives rise to behavior rely crucially on our assumptions of how signal, nuisance modulation, and trial variability combine to determine task performance, it is important that those assumptions are valid within the coding regime relevant for any given task.

In the regime where signals are large relative to the size of trial variability, a neuron's response will typically be more modulated by changing an experimental parameter for which it is "tuned" than it will be across trials when that parameter is kept fixed. Tuning functions are often depicted in this way, with large tuning modulations and small error bars to depict trial variability. In this high signal-to-trial variability ratio regime, it naturally follows that when neurons are sensitive to changes in both task-relevant parameters (i.e., signal) as well as task-irrelevant parameters (i.e., nuisance modulation), nuisance modulation will be detrimental to single-neuron task performance (as depicted in Fig. 1*B*). However, as described in more detail below, this intuition is inaccurate when spike counts are small, for example, because of spike count integration windows that are short (as implied for tasks with fast reaction times; Fig. 2). That is, in a low signal-to-trial variability regime, even when nuisance modulation is much larger than the task-relevant signal, it has the nonintuitive property of having a negligible impact on single-neuron task performance.

Address for reprint requests and other correspondence: N. C. Rust, Dept. of Psychology, Univ. of Pennsylvania, Goddard Labs, 428, Philadelphia, PA 19104 (e-mail: nrust@psych.upenn.edu).

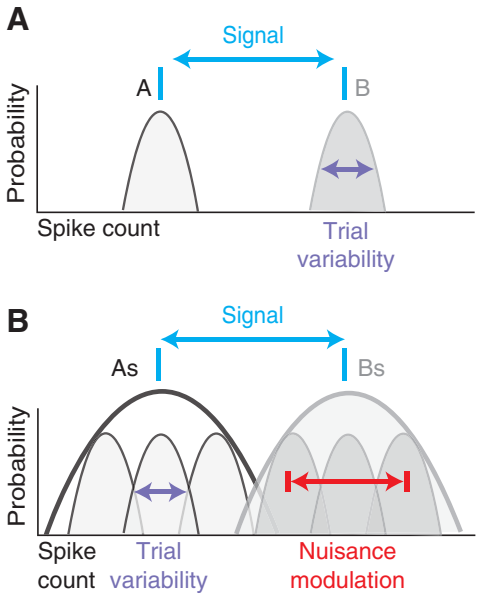

Fig. 1. Classic intuitions about how signal, nuisance modulation, and trial variability impact task performance. *A*: schematic of single-unit task performance (d′) for a classic, 2-way discrimination task in which a subject is asked to label different conditions as "A" or "B" across repeated trials. Shown are hypothetical distributions of spike count responses for the 2 conditions. d′ is measured as the separation of the 2 spike count distributions in units of the number of standard deviations separating their means. d′ is proportional to the amount of signal, which determines the separation between the means of the distributions (cyan), and d′ is inversely proportional to the spread within each distribution, which arises as a result of variability across repeated trials within each condition ("trial variability," purple). *B*: schematic of d′ for the same discrimination task but extended to require grouping multiple conditions into each of 2 sets, "As" and "Bs" (e.g., an object identification task where 2 objects are presented in multiple background contexts). In this case, "nuisance" modulations (e.g., firing modulations by the background context) increase the spread of the responses within each condition and thus lower d′.

Understanding how descriptions of neural task performance formulated for individual neurons extend to the performance of a neural population requires incorporating population considerations such as how task-relevant information is formatted (e.g., linearly or nonlinearly) and whether trial and nuisance variability are correlated between neurons. Investigations focused in part on view-invariant object recognition have demonstrated the means by which individual neurons can multiplex different types of signals such that each type of signal can be extracted from the population with a simple linear decoder (DiCarlo and Cox 2007; Hong et al. 2016; Hung et al. 2005; Li et al. 2009), but little attention has been directed toward understanding how signal and nuisance modulation impact population performance within the context of these linearly separable representations. Some insight into these issues can be gained from work focused on how correlated interactions between neurons impact population performance within a linear decoding scheme (reviewed by Averbeck et al. 2006; Cohen and Kohn 2011; Kohn et al. 2016). However, this work has focused nearly exclusively on correlated trial (as opposed to nuisance) variability (but see Kim et al. 2016). Understanding how nuisance modulation impacts neural task performance thus requires extending these population-based

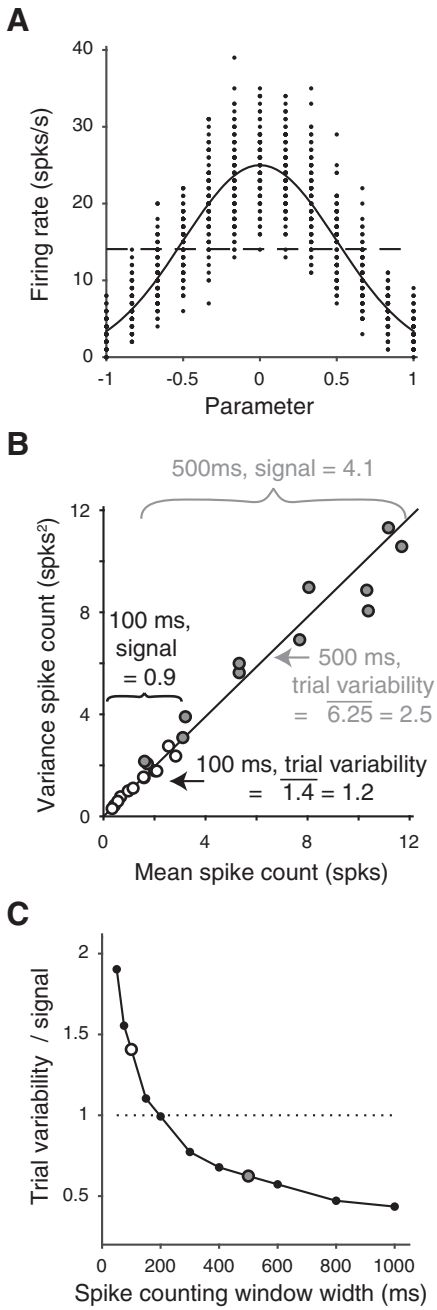

Fig. 2. The relative sizes of signal and trial variability depend on spike count window width. In these simple simulations, a 1-dimensional (1D) Gaussian tuning function with a peak of 25 spikes/s was sampled with 13 points evenly spaced points along the *x*-axis. Firing rate responses were then converted into spike counts for windows of various durations, followed by the introduction of Poisson trial variability. *A*: a 1D Gaussian tuning function along with points that depict the distributions of spike counts computed in a 1-s count window. Dashed line indicates the grand mean firing rate across all conditions. *B*: as imposed by the Poisson simulation, the relationship between the mean spike count within each condition and the variance of spike count across repeated trials for each condition (often called the Fano factor or variance-to-mean ratio) remained linear with a slope of 1, regardless of window width. However, "signal," computed as the standard deviation across the set of mean spike counts for each condition, grew faster than "trial variability," computed as the square root of the average variance across trials within each condition. Values corresponding to signal and trial variability for 100-ms and 500-ms spike count windows are labeled. *C*: plot of the ratio of trial variability relative to signal as a function of spike count window length. Dotted line depicts the benchmark of a ratio equal to 1. White and gray dots correspond to the 100- and 500-ms windows, respectively, from *B*.

approaches to incorporate considerations about nuisance variability.

MATERIALS AND METHODS

Experiments were performed on two adult male rhesus macaque monkeys (*Macaca mulatta)* with implanted head posts and recording chambers. All procedures were reviewed and approved by the University of Pennsylvania Institutional Animal Care and Use Committee. The behavioral and neural data included in this report were also included in a previous report (Roth and Rust 2018). Details of the experimental design and data analysis are described extensively in that report and are summarized here.

Invariant delayed-match-to-sample task. All behavioral training and testing was performed via operant conditioning (juice reward), head stabilization, and high-accuracy infrared video eye tracking. Stimuli were presented with customized software (https://mworks-.github.io) on an LCD monitor with an 85-Hz refresh rate.

As an overview, the monkeys' task required them to make an eye movement response to a specific location when a target object appeared within a sequence of distractor images to receive a reward (Fig. 3*A*). Objects were presented across variation in position, size, and background context (Fig. 3*B*). Stimuli consisted of a fixed set of 20 images that were shown across changes in the identity of four target objects, each presented at five identity-preserving transformations. The task consisted of short blocks (~3 min), each with a fixed target, before another target was pseudorandomly selected.

Trials were initiated by the monkey fixating on a red dot in the center of a gray screen. Cue trials indicating the target object were presented at the beginning of each block. Test trials, which are the focus of this report, always began with a distractor image, and neural responses to this image were discarded to minimize nonstationarities such as stimulus onset effects. In each block, distractors were chosen randomly from a set of 15 possible images without replacement until each distractor was presented once on a correct trial, and the images were then rerandomized. On most trials, a random number of one to six distractors were presented, followed by a target match (Fig. 3*A*). On a small fraction of trials, seven distractors were shown, and the monkey was rewarded for fixating through all distractors. Stimuli were presented for 400 ms (or until the monkeys' eyes left the fixation window), and each stimulus was immediately followed by the presentation of the next stimulus. After the onset of a target match image, monkeys received a juice reward for making a saccade to a response target within 600 ms. If 400 ms had elapsed after target onset and the monkey had not moved its eyes, a distractor stimulus was immediately presented. Within each block, four repeated presentations of the 20 images were collected, and a new target object was then pseudorandomly selected. After the presentation of all four objects as targets, the targets were rerandomized, and at least 20 repeated trials of each condition were collected in total. Overall, monkeys performed this task with high accuracy. Disregarding fixation breaks (*monkey 1*: 8%

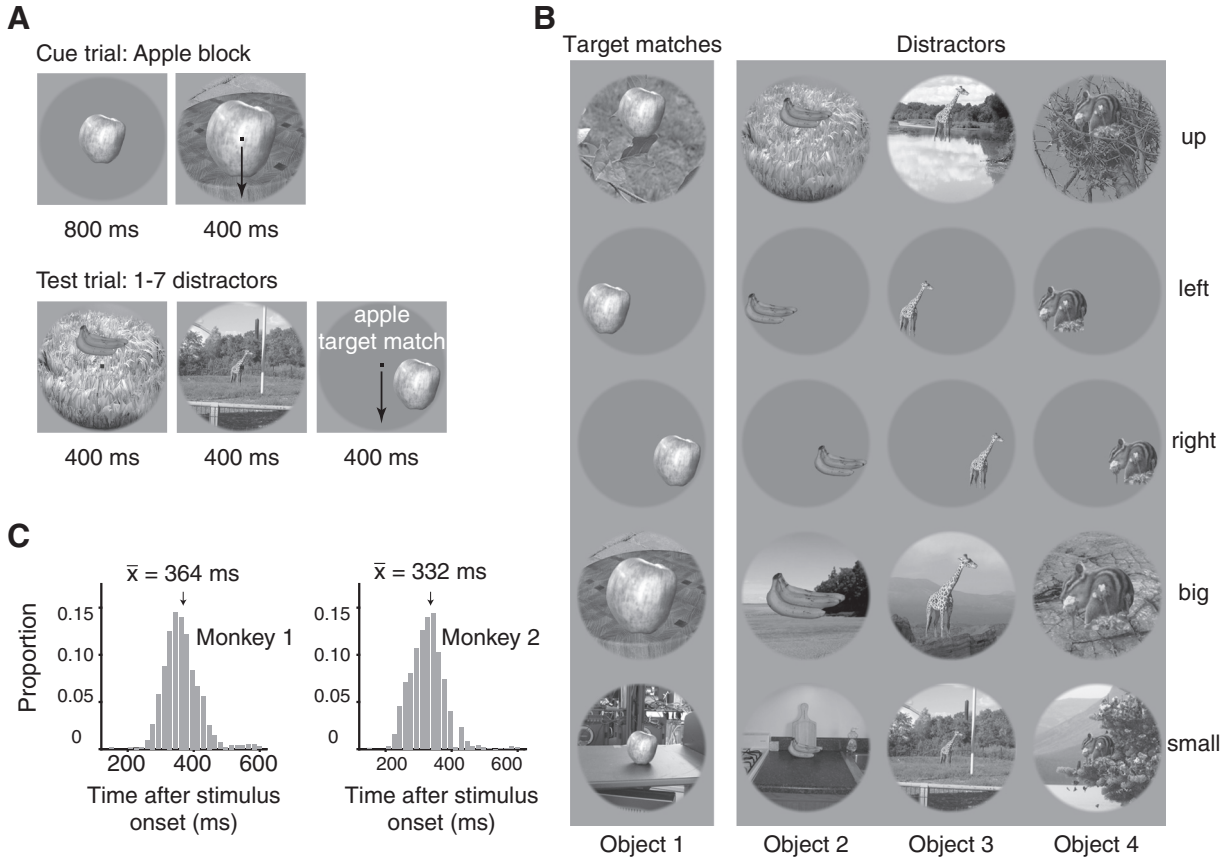

Fig. 3. Invariant delayed-match-to-sample (IDMS) task. *A*: monkeys performed an IDMS task. Each block (~3 min in duration) began with a cue trial indicating the target object for that block. On subsequent trials, monkeys initiated a trial by fixating on a small dot. After a 250-ms delay, a random number (1–7) of distractors were presented, and on most trials this was followed by a target match. Monkeys were required to maintain fixation throughout the distractors and make a saccade to a response dot within a window 75–600 ms after the onset of the target match to receive a reward. In cases where the target match was presented for 400 ms and the monkey had still not broken fixation, a distractor stimulus was immediately presented. *B*: the experiment included 4 objects presented at each of 5 identity-preserving transformations ("up," "left," "right," "big," "small"), for 20 images in total. In any given block, 5 of the images were presented as target matches and 15 were distractors. *C*: histograms of reaction times on correct trials (milliseconds after stimulus onset) during the IDMS task for each monkey, with means indicated by arrows and labeled.

of trials; *monkey 2*: 11% of trials), percent correct on the remaining trials was as follows: *monkey 1*: 96% correct, 1% false alarms, and 3% misses; *monkey 2*: 87% correct, 3% false alarms, and 10% misses.

Neural recording. The activity of inferotemporal cortex (IT) neurons was recorded via a single recording chamber in each monkey. Neural activity was largely recorded with 24-channel U probes (Plexon) with linearly arranged recording sites spaced with 100-μm intervals, with a handful of units recorded with single electrodes (Alpha Omega, glass-coated tungsten). Details of the recording and spike sorting procedure are described at length in Roth and Rust (2018).

We measured neural responses by counting spikes in a window that began 80 ms after stimulus onset. For all the analyses presented in this report except Fig. 6*B*, the spike count window ended at 250 ms. On 1.9% of all correct target match presentations reaction times were faster than 250 ms, and those trials were excluded from analysis such that spikes were only counted during periods of fixation. When combining the units recorded across sessions into a larger pseudopopulation, we screened for units that met three criteria. First, we only included units that were modulated by our task, as quantified by a one-way ANOVA applied to our neural responses (80 conditions $\times$ 20 repeats) with $P < 0.01$. Second, units were required to pass a loose criterion on recording stability, as quantified by computing the variance-to-mean ratio for each unit (calculated by fitting the relationship between the mean and variance of spike count across the 80 conditions) and eliminating units with a variance-to-mean ratio > 5. Finally, we applied a loose criterion on unit recording isolation, quantified by computing the signal-to-noise ratio (SNR) of the waveform (as the difference between the maximum and minimum points of the average waveform divided by twice the standard deviation across the differences between each waveform and the mean waveform) and excluding (multi)units with an SNR < 2. This screen yielded a population of 204 units (of 563 possible units), which included 96 units from *monkey 1* and 108 units from *monkey 2*.

Quantifying single-unit modulation magnitudes. To quantify single-unit modulation magnitudes (Fig. 6*A*, Fig. 7*C*), we applied a bias-corrected, ANOVA-like procedure described in detail by Pagan et al. (2016) and summarized here. This procedure considers the total variance in the spike count responses for each unit across conditions ($n = 80$) and trials ($m = 20$ for each condition) and parses this total variance into the variance that can be attributed to each type of experimental parameter and variance attributed to trial variability. Similar to an ANOVA, the procedure is designed to parse response variance, including the variance that can be attributed to changes in the identity of the visual image, the identity of the target object, and whether each condition was a target match or a distractor. These variances are then converted into measures of spike count modulation (i.e., standard deviation around each unit's grand mean spike count) via a procedure that includes bias correction for overestimates in modulation due to noise.

The procedure began by developing an orthonormal basis of 80 vectors designed to capture all types of modulation with intuitive groupings. The orthonormal basis is most easily understood by envisioning the 80 experimental conditions as a matrix of 20 images (4 objects $\times$ 5 transformations) each viewed in the context of 4 targets (i.e., a 4 $\times$ 20 matrix; Fig. 5*A*). To design the orthonormal basis for this task, we began by constructing a first vector that corresponded to the grand mean spike count response across all conditions; all entries in this vector took on the same, constant value (e.g., 1/80; Fig. 5*B*, "grand mean"). The remaining vectors were designed to capture the types of modulation that neural responses might reflect, which follow from the task design. In the case of our experiment, the first type of modulation differentiated whether a condition was a target match or a distractor, and this corresponds to modulation along the elongated diagonal of the 4 $\times$ 20 matrix (i.e., entries in which the object in view matches the target, e.g., for *object 1*, *images 1–5*; for *object 2*, *images 6–10*, etc.; Fig. 5*B*, "target match"). The second type included 19 vectors to describe the visual modulation, reflected as vertical structure in the orthonormal basis (Fig. 5*B*, "visual"). Notably, although there are 20 different visual images, only 19 were required to capture the visual modulation once the mean firing rate response had also been defined (i.e., degrees of freedom for the visual conditions = 20 – 1). The third type of modulation included three vectors corresponding to response modulations that could be attributed to changing the identity of the target, reflected as horizontal structure in the matrix (Fig. 5*B*, "target identity"). The final type of modulation was that which was required to describe responses that are "peppered" across the matrix, such as differential responses to the same visual image under two different distractor conditions, and we refer to this modulation as "residual" (Fig. 5*B*, "residual"). More technically, residual modulations reflect all nonlinear combinations of visual and target identity signals that do not fall along the elongated diagonal (i.e., are not target match modulation). This final type of modulation included the remaining 64 vectors to form a set of 80 vectors. Once this initial set of vectors was defined, we applied the Gram-Schmidt procedure to convert it into an orthonormal basis. Specifically, we defined each of the n original vectors as $\boldsymbol{v}_i$ and each of the vectors of the resulting orthonormal basis as $\boldsymbol{b}_i$. The Gram-Schmidt process was applied iteratively to each initially defined vector and consisted of two stages: first the vector was orthogonalized relative to all the vectors already incorporated into the final, orthonormal basis, and second, the resulting vector was normalized by its norm. A crucial requirement is that the final vectors $\boldsymbol{b}_1 ... \boldsymbol{b}_n$ spanned the full space. We verified this by checking that the rank of the matrix obtained by juxtaposing the final vectors $[\boldsymbol{b}_1 ... \boldsymbol{b}_n]$ was equal to 80.

The resulting basis $\boldsymbol{b}_i ... \boldsymbol{b}_n$ spanned the space of all possible responses for our task, and each trial-averaged vector of spike count responses to the 80 experimental conditions $\boldsymbol{R}$ can be reexpressed as a weighted sum of these basis vectors. To quantify the amounts of each type of modulation reflected by each unit w_i, we began by computing the squared projection of each basis vector $\boldsymbol{b}_i$ and $\boldsymbol{R}$. An analytical bias correction, described and verified in Pagan and Rust (2014), was then subtracted from this value:

$$w_i^2 = \left(R \cdot b_i^T\right)^2 - \frac{\sigma_t^2 \cdot (b_i^T)^2}{m} \quad (1)$$

where σ_t^2 indicates the trial variance, averaged across conditions ($n = 80$), and m indicates the number of trials ($m = 20$). When more than one dimension existed for a type of modulation (e.g., "visual"), we summed values of the same type (e.g., for total nuisance modulation, we summed over $i = 3{:}80$; see *Eq. 3* below). Next, we applied a normalization factor [$1/(n - 1)$. where $n = 80$] to convert these summed values into variances. Finally, we computed the square root of these quantities to convert them into modulation measures that reflected the number of spike count standard deviations around each unit's grand mean spike count. Target match modulation was computed as

$$\sigma_{TM} = \sqrt{\frac{1}{n-1} \cdot w_2^2} \quad (2)$$

Nuisance modulation was computed as

$$\sigma_{Nui} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=3}^{80} w_i^2} \quad (3)$$

To compute the different subtypes of nuisance modulation, we replaced the weights w_i^2 in *Eq. 2* with the weights that corresponded to the orthonormal basis vectors corresponding to each subtype, including visual modulation ($i = 3$ to 21), target modulation ($i = 22$ to 24), and residual modulation ($i = 25$ to 80), as described above.

Trial variability for each unit (σ_{Trial}) was computed in a comparable manner as the square root of the average (across condition) variance across trials:

$$\sigma_{\text{Trial}} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \frac{1}{m-1} \cdot \sum_{t=1}^{m} \left(s_{it} - s_i\right)^2} \quad (4)$$

where the spike count response for a particular trial t of condition i was s_{it} and the mean spike count response across all trials of condition i was s_i.

When estimating modulation population means (Fig. 6*A*, Fig. 7*C*), the bias-corrected squared values were averaged across units before taking the square root. Because these measures were not normally distributed, standard error about the mean was computed via a bootstrap procedure. On each iteration of the bootstrap (across 1,000 iterations), we randomly sampled values from the modulation values for each unit in the population, with replacement. Standard error was computed as the standard deviation across the means of these newly created populations.

Relating modulation magnitudes and single-unit performance. To determine the impact of nuisance modulation on single-unit task performance (d′) (Fig. 7, *A* and *B*) we expressed d′ as a function of the different types of signal modulations described above (*Eqs. 2–4*):

$$\mathrm{d}' = \frac{\left|\mu_{\text{Match}} - \mu_{\text{Distractor}}\right|}{\sigma_{\text{pooled}}} = \sqrt{\frac{a \cdot \sigma_{\text{TM}}^2}{b \cdot \sigma_{\text{Nui}}^2 + \sigma_{\text{Trial}}^2}} \quad (5)$$

where

$$a = \frac{n-1}{3}$$

and

$$b = \frac{n-1}{n}$$

This derivation is described in detail in Pagan and Rust (2014).

To quantify the impact of nuisance modulation on d′, we compared each unit's d′ in the presence of nuisance modulation (*Eq. 5*) vs. d′ when the nuisance modulation term σ_{Nui} was set to zero ($\mathrm{d}'_{\text{NoNui}}$). We then calculated the impact of nuisance modulation as the percent increase in d′ without nuisance:

$$\text{Impact}(\%) = \left(\frac{\mathrm{d}'_{\text{NoNui}}}{\mathrm{d}'} - 1\right) \cdot 100 \quad (6)$$

Population performance. We quantified the ability of the IT population to classify target matches vs. distractors (Figs. 9*C* and Fig. 10), by applying a Fisher linear discriminant (FLD, a linear decoder) using approaches that were described previously in detail (Pagan et al. 2013) and are summarized here.

The general form of a linear decoding axis is

$$f(x) = \boldsymbol{w}^T x + b \quad (7)$$

where $\boldsymbol{w}$ is an N-dimensional vector (where N is the number of units) containing the linear weights applied to each unit and b is a scalar value. We fit these parameters with a FLD, where the vector of linear weights was calculated as

$$\boldsymbol{w} = \Sigma^{-1}\left(\mu_1 - \mu_2\right) \quad (8)$$

and b was calculated as

$$b = \boldsymbol{w} \cdot \frac{1}{2}\left(\mu_1 + \mu_2\right) = \frac{1}{2}\mu_1^T \Sigma^{-1} \mu_1 - \frac{1}{2}\mu_2^T \Sigma^{-1} \mu_2 \quad (9)$$

Here μ_1 and μ_2 are the means of the two classes (target matches and distractors, respectively), and the mean covariance matrix is calculated as

$$\Sigma = \frac{\Sigma_1 + \Sigma_2}{2} \quad (10)$$

where $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are the regularized covariance matrices of the two classes. These covariance matrices were computed with a regularized estimate equal to a linear combination of the sample covariance and the identity matrix $\mathbf{I}$ (Pagan et al. 2016):

$$\Sigma_i = \gamma \Sigma_i + \left(1 - \gamma\right) \cdot \mathbf{I} \quad (11)$$

Regularization with the identity matrix guarantees that the resulting covariance matrix is invertible, in contrast to other approaches, such as regularization with the diagonal matrix, that do not (e.g., when the variance across all of the training trials for a given condition in any split of the data is 0, a matrix regularized with the diagonal of the covariance matrix will not be invertible and the FLD solution will be undefined). To determine γ, we explored a range of values from 0.01 to 0.99, and we selected the value that maximized average performance across all iterations, measured with the cross-validation "regularization" trials set aside for this purpose (see below). We then computed performance for that value of γ with separately measured "test" trials, to ensure a fully cross-validated measure. Because this calculation of the FLD parameters incorporates the off-diagonal terms of the covariance matrix, FLD weights are optimized for both the information conveyed by individual units as well as their pairwise interactions (i.e., correlated trial and nuisance variability; Fig. 8).

We begin by describing our procedures for Fig. 9, followed by the modifications applied to analyze the role of correlations (Fig. 10). For Fig. 9, the decoder was cross-validated via the following resampling procedure. On each resampling iteration, we randomly shuffled the trials for each condition and for each unit and (for numbers of units less than the full population size) randomly selected units. On each iteration, 18 trials from each condition were used for training the decoder, 1 trial was used to determine a value for regularization, and 1 trial from each condition was used for cross-validated measurement of performance.

On each iteration of the resampling, we included 20 target match conditions and 20 (of 60 possible) distractor conditions, to ensure that decoder performance was not biased by unequal numbers of target matches and distractors. To do so, we subselected each set of 20 distractors such that it spanned all possible combinations of mismatched object and target identities (e.g., *objects 1*, *2*, *3*, *4* paired with *targets 4*, *3*, *2*, *1*), of which there are nine possible sets.

We computed population d′ as a measure of population performance. We began by computing the dot product of the test data and the linear weights $\boldsymbol{w}$, adjusted by b (*Eq. 7*). Population d′ was computed for the distributions of these values across the 20 different images presented as target matches versus as distractors:

$$\mathrm{d}' = \frac{\left|\mu_{\text{Match}} - \mu_{\text{Distractor}}\right|}{\sigma_{\text{pooled}}} \quad (12)$$

where μ_{Match} and $\mu_{\text{Distractor}}$ correspond to the mean across the set of matches and distractors, $\sigma_{\text{pooled}} = \sqrt{\left(\sigma_{\text{Match}}^2 + \sigma_{\text{Distractor}}^2\right)/2}$, and σ_{Match} and $\sigma_{\text{Distractor}}$ correspond to the standard deviation across the set of matches and distractors, respectively.

We computed population d′ (*Eq. 12*) on each resampling iteration for the 20 target match conditions and 20 distractor conditions, separately for each set of nine match-distractor combinations, and then averaged across the nine sets. Mean and standard error of population d′ were computed as the mean and standard deviation of d′ across 2,000 resampling iterations. Standard error thus reflected the variability due to the specific trials assigned to training and testing and, for populations smaller than the full size, the specific units chosen.

To determine the impact of correlated trial and nuisance variability on IT population performance (Fig. 10), we began by applying the

FLD described above with the following modifications. First, when available (Fig. 10*A*), the analysis was applied to simultaneously recorded data within each session, and the correlation structure on each trial was kept intact on each resampling iteration (Fig. 10*A*, "Intact"). When >24 units were available within one session, a subset of the 24 units was selected, composed of the units with the most task modulation, quantified via the *P* value of a one-way ANOVA applied to each unit's responses (80 conditions × 20 trials). Second, we randomly shuffled the trials within each condition, for each unit, on each iteration of the bootstrap (Fig. 10*A*, "Shuffle TV"). Third, we randomly shuffled both trial variability as well as the assignment of image identity for each of the 20 distractor conditions and 20 target match conditions on each bootstrap iteration (Fig. 10*A*, "Shuffle TV&NV"). The analysis to determine the impact of correlated trial and nuisance variability on the pseudosimulation (Fig. 10*B*) was performed in the same manner but applied to the pseudosimulated data.

Simulations. To understand the relationship between trial variability and signal in a general case, we simulated a neuron's responses to a hypothetical set of stimulus conditions (Fig. 2). The mean firing rate response for each experimental condition was determined by sampling a Gaussian tuning curve at 13 evenly spaced points between −1 and 1, with a mean of 0 and standard deviation of 0.5. We held the peak firing rate of this simulated neuron at 25 Hz and counted spikes in different durations from 50 to 1,000 ms. Next, we simulated responses across 100 repeated trials with Poisson trial variability. To compute the ratio of trial variability to signal (Fig. 2*B*), we began by creating a modified orthonormal basis similar to the one described above. Specifically, this orthonormal basis $\boldsymbol{b}$ included 13 vectors $\boldsymbol{b}_i$ that reflected *1*) the grand mean spike count across all conditions ($\boldsymbol{b}_1$, 1 dimension) and *2*) the experimental condition the neuron was tuned for ($\boldsymbol{b}_2$–$\boldsymbol{b}_{13}$, 12 dimensions). We next computed signal as

$$\sigma_{\text{Sig}} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=2}^{13} w_i^2} \qquad (13)$$

and trial variability as described in *Eq. 4*, where $n = 13$ and $m = 100$. Finally, we computed the ratio of trial variability to signal for the simulated neuron at each simulated spike count duration.

We also performed a number of simulations based on our recorded data. Each of these simulations began by computing the bias-corrected weights for each recorded unit as described above (*Eq. 1*).

To explore how rescaling the spike counts by different factors of *N* influenced the ratio of trial variability to signal or trial variability to nuisance modulation (Fig. 6*C*), or influenced the impact of nuisance modulation (Fig. 7*B*), we rectified bias-corrected modulations that fell below zero, recomputed the noise-corrected mean spike count responses for each condition, rescaled the mean spike counts by *N*, and generated trial variability with an independent Poisson process.

To estimate the impact of nuisance modulation on population performance, we simulated two versions of each of our recorded units (Fig. 9*C*, "Nuisance-intact" and "Nuisance-removed"). In the "intact" version, we computed each unit's responses as described for the rescaling simulation but with a rescale factor $N = 1$. In the "Nuisance removed" version, we used a similar procedure but set the modulations corresponding to all nuisance dimensions to 0. The responses were thus computed based on the grand mean spike count response as well as the target match modulation. To quantify the impact of nuisance modulation on population performance, we used *Eq. 12* as computed with the population d′ performance values, where d′ refers to the "Nuisance-intact" simulation, and d'_{NoNui} refers to the "Nuisance-removed" simulation.

RESULTS

Impact of spike count window width on relative sizes of signal and trial variability—a simple simulation. To develop intuitions for how we quantify the size of signal modulation relative to the size of trial variability, as well as to establish the relationships between these measures when spikes are counted in windows of different widths, we begin by focusing on a simple simulation of a Gaussian tuning function.

Simulated data were computed as the spike count responses across repeated presentation of *N* experimental conditions, each on *T* repeated trials (Fig. 2*A*). These data thus consist of $N \times T$ responses. For this simulation, the firing rate responses for each experimental condition were determined by sampling a Gaussian tuning curve at $N = 13$ evenly spaced points along the *x*-axis. While the peak firing rate of this simulated neuron was held fixed (25 Hz), spikes were counted in windows of different durations (50 ms to 1 s), followed by the introduction of Poisson trial variability across $T = 100$ repeated trials. The purpose of this simulation is to demonstrate how the relationship between signal and trial variability depends on spike count window width. To prevent confusion, we emphasize that the effect we are highlighting here is not a change in the "variance-to-mean ratio" or "Fano factor" with spike count window width. As imposed by the Poisson process used in this simulation, the relationship between the mean spike count for each condition and the variance in spike counts across repeated trials within each condition is linear with a slope of 1 regardless of whether spikes were counted in windows of 100 ms or 500 ms (Fig. 2*B*).

Rather, the effect that we wish to highlight follows from the fact that because the spike count variance across repeated trials grows linearly as mean spike counts are rescaled (Fig. 2*B*), the dispersal (i.e., standard deviation) of spike counts across the different conditions ("signal") grows at a faster rate than the dispersal (i.e., standard deviation) of spike counts across trials ("trial variability"). For example, in this simulation, signal was smaller than trial variability for count windows of 100 ms (0.9 vs. 1.2 spikes for signal and trial variability respectively; Fig. 2*B*) but the relationship was reversed for 500-ms count windows (4.1 vs. 2.5 spikes for signal and trial variability, respectively; Fig. 2*B*). Figure 2*C* plots the ratio of trial variability over signal as a function of spike count window width for a range of window widths—whereas signal was larger than trial variability for long spike count windows, trial variability dominated for spike count windows shorter than 200 ms. This simple simulation thus illustrates that a low signal-to-trial variability regime is not exotic or unexpected but rather follows from counting spikes in short windows, even when neurons are reasonably well tuned and have reasonable peak firing rates. Below we illustrate that one consequence of this relatively simple relationship is a revision to the seemingly straightforward intuitions presented in Fig. 1.

More generally, signal and trial variability can be computed with methods similar to a one-way ANOVA, where the total variance across the $N \times T$ responses is parsed into the variance that can be attributed to changes in the visual image ("signal variance") and the variance across repeated presentations of the same image ("trial variance"). Because signal variance nonintuitively scales as the square of response changes (e.g., when

firing rates double, the signal variance quadruples), we find it useful to consider modulation in units of standard deviation, computed as the square root of these quantities. Conceptually, this simple simulation can also be extended to incorporate nuisance modulation by simulating the responses for a matrix of $N \times M$ experimental conditions, each on T trials, where the first parameter might be considered "signal" and the second parameter "nuisance modulation." In this case, the data would consist of $N \times M \times T$ data points and, similar to a two-way ANOVA, the total variance can be decomposed into signal, nuisance modulation, and trial variability. This is the approach that we apply to recorded neural data, as described below.

IT operates in a regime in which signals are small relative to trial variability during a delayed-match-to-sample object search task. To demonstrate the relevance of the simple simulation presented in Fig. 2 to actual neural data, we quantified signal, nuisance, and trial variability in data recorded from IT as two monkeys performed an invariant delayed-match-to-sample object search (IDMS) task. In this task, monkeys viewed a sequence of images and reported when a sought visual target appeared across variation in the object's position, size, and background context (Fig. 3; Roth and Rust 2018). Specifically, the monkeys' task required them to fixate during the presentation of distractors and make a saccade to a response dot on the screen after target match onset to receive a reward (Fig. 3*A*). This experiment included viewing a fixed set of 20 images, broken down into four objects presented at each of five transformations (Fig. 3*B*). Each of the 20 images was viewed in the context of each of the four objects as a target, resulting in 80 experimental conditions in total. For each condition, we collected at least 20 repeats on correct trials. Monkeys generally performed well on this task (mean percent correct *monkey 1* = 96%, *monkey 2* = 87%). Of particular relevance to this report is their reaction times (computed as the time their eyes left the fixation window relative to the target match stimulus onset), which were fast (means: 364 ms and 332 ms; Fig. 3*C*), implying short spike count integration windows.

As the two monkeys performed this task, we recorded neural activity from IT. Units were screened on the basis of their stability, isolation, and task modulation (see MATERIALS AND METHODS). The data reported here are extracted from trials with correct responses and include the spike count responses of 204 units $\times$ 80 experimental conditions $\times$ 20 repeated trials. Spikes were counted for each unit in windows that always began at 80 ms after stimulus onset with variable widths (detailed below) that terminated no later than 250 ms, and thus always before the monkeys' reaction times on these trials. Distributions of grand mean and peak firing rates computed for the spike count window 80–250 ms are shown in Fig. 4.

The monkeys' task can be envisioned as a two-way classification of the same images presented as target matches (requiring a saccade) versus as distractors (requiring them to maintain fixation). In an earlier report of this data, we found that IT "target match modulation," parameterized in this way, correlated with the monkeys' behavior insofar as the IT population significantly classified correct trials and misclassified trials in which the monkeys made errors (Roth and Rust 2018). The IDMS task also included a number of different nuisance variables, including "visual" (changes in the identity of the visual image, including object identity and object transformation), "target" (changes in the identity of the sought target), and "residual" (nonlinear interactions between visual and target modulation that are not target match modulation, such as a preferential response to a particular image presented as a distractor). We emphasize that our use of the term "residual"—which captures a type of nonlinear tuning—should not be confused with variability across trials (which we term "trial variability"). To quantify the degree to which IT units were modulated by these different variables, we applied an extension of the intuitions presented in Fig. 2, where we parsed total response variance into variance that could be attributed to the target match signal, different types of nuisance modulation, and trial variability. Our procedure included projecting the responses of each unit onto an orthonormal basis designed to group experimental variables into these intuitive sets (Fig. 5; see MATERIALS AND METHODS). Variances were corrected for bias due to limited samples and converted into estimates of modulation (and thus units of standard deviation).

We found that during the IDMS task trial variability was considerably larger than any type of modulation in IT. When applied to spike count windows 80–250 ms after stimulus onset, average trial variability was 5.2$\times$ larger than the target match signal (Fig. 6*A*). Trial variability was also larger than any of the other types of nuisance parameters, including visual

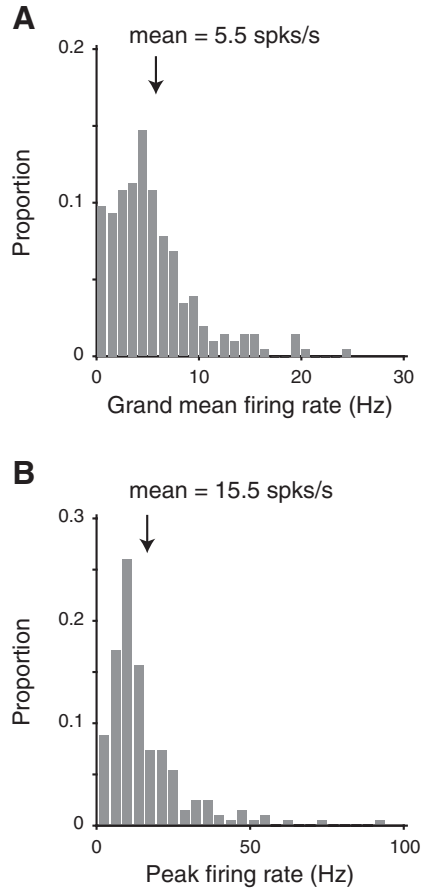

Fig. 4. Firing rate distributions. Firing rate responses were calculated in a window 80–250 ms after stimulus onset and averaged across 20 repeated trials. *A*: distribution of grand mean firing rates, computed across all conditions and trials. *B*: distribution of peak firing rates, computed as the maximum firing rate for each unit across the 80 conditions. Arrows indicate the means across n = 204 units.

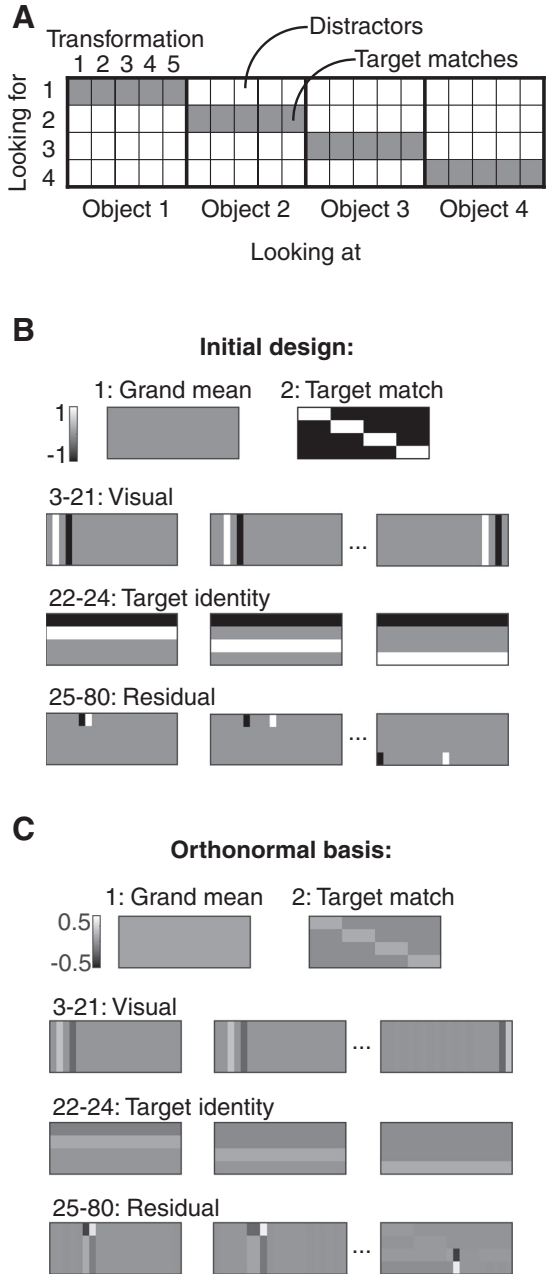

Fig. 5. Constructing an orthonormal basis for the invariant delayed-match-to-sample (IDMS) task. *A*: the complete experimental design of the IDMS task included looking "at" 4 objects, each presented at 5 transformations (for a total of 20 images) viewed in the context of looking "for" each object as a target, and thus $4 \times 20 = 80$ experimental conditions. In this design, target matches fall along the extended diagonal (gray) and distractors are on the off-diagonal (white). *B*: a subset of the 80 initial vectors produced by the first stage of the orthonormal basis design process (see MATERIALS AND METHODS), reconfigured into 4×20 matrices for visualization. *C*: the same basis vectors but after application of the Gram-Schmidt orthogonalization process.

modulation (Fig. 6*A*). This establishes that the average IT unit exists in a regime in which signal and nuisance variability are small compared with trial variability during the IDMS task.

Why was trial variability so much larger than signal and nuisance modulation during the IDMS task? Consistent with the intuitions presented in Fig. 2, this was a result of counting spikes in relatively short spike count windows (170 ms) as a consequence of terminating the count window before the monkeys' reaction times, which were fast (Fig. 3*C*). Within these short spike count windows, the average grand mean spike count was 0.94 spikes per condition per trial and the average peak spike count across the 80 conditions was 2.63 spikes (which translates into mean and peak firing rates of 5.5 spikes/s and 15.5 spikes/s, respectively; Fig. 4). We also found that, consistent with earlier reports, IT trial variability was approximately Poisson (average variance-to-mean ratio across units = 1.20, relative to the Poisson benchmark of 1.0, computed for 170-ms spike count windows positioned 80–250 ms after stimulus onset). Consistent with the simple intuition that trial variability dominates in short spike count windows (Fig. 2), the ratio between trial variability and the target match signal was very large for short integration windows but became smaller as a function of spike count window width (Fig. 6*B*). We also performed simulations to extrapolate our data to larger windows by rescaling the responses by factors of N (Fig. 6*C*). We found that a 16-fold rescaling was required to equate the sizes of the target match signal with trial variability and a 3-fold rescaling was required to equate the sizes of trial variability and nuisance modulation. Under the simplifying assumption of firing rates that remain constant as a function of time, these translate into spike count windows of 2.72 s and 0.5 s for the target match signal and nuisance modulation, respectively.

In sum, analysis of both simulated tuning curves as well as recorded neural data demonstrate that when spikes are counted in short integration windows neurons operate in a regime in which signal and nuisance modulations are small relative to the size of trial variability. This is the regime that much of the brain is expected to exist in for tasks with fast reaction times (e.g., ~250 ms), such as the IDMS task. In the next section, we explore how signal, nuisance, and trial variability combine to determine single-neuron task performance in this regime.

When trial variability is large, nuisance modulation is largely inconsequential for single-unit task performance. As depicted in Fig. 1*B*, the intuition that nuisance modulation should be detrimental to task performance is straightforward. However, as we illustrate here, this intuition only holds in a regime in which signal and nuisance modulations are large compared with trial variability. To illustrate the simple algebraic intuition for why this is the case, we present an extension of our previous formulation for single-unit d′, computed for a two-way classification (Pagan and Rust 2014). To summarize that approach, d′ is a measure of the ratio between signal and noise, where noise is parsed into one component proportional to total nuisance modulation and another component proportional to trial variability (Fig. 1*B*):

$$|d'| = \sqrt{\frac{k_1 \cdot \text{signal modulation}^2}{k_2 \cdot \text{nuisance modulation}^2 + \text{trial variability}^2}} \quad (14)$$

Here, k_1 and k_2 are constants. With this formulation, the impact of nuisance modulation on d′ can be determined by considering the increase in d′ when nuisance modulation is not incorporated into the calculation (i.e., a hypothetical scenario in which nuisance modulation does not exist) compared with when it is (i.e., for the intact data), analogous to the increase in d′ in Fig. 1*A* relative to Fig. 1*B*. Figure 7*A* shows the result of this analysis applied to the IDMS data, which reveals that

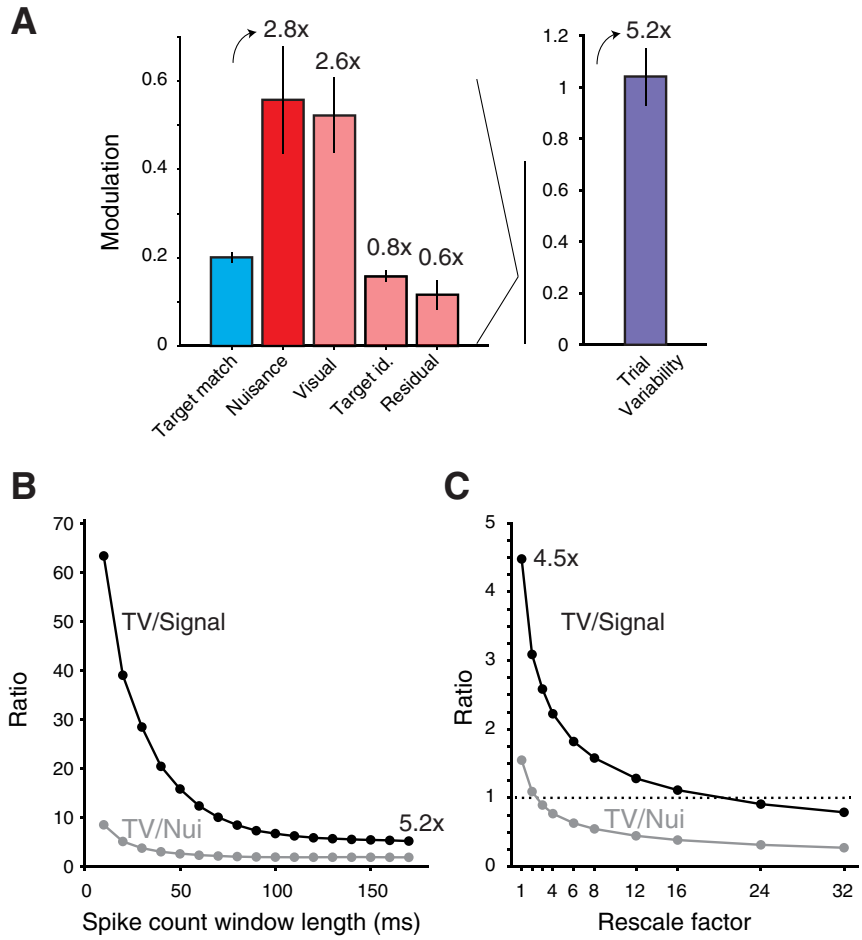

Fig. 6. Trial variability is larger than signal and nuisance modulation in inferotemporal cortex during the invariant delayed-match-to-sample (IDMS) task. Modulations were computed for each type of experimental parameter, in units of the standard deviations around each unit's grand mean spike count (see RESULTS). *A*, *left*: average modulation magnitudes across units, parsed into target match modulation (signal for the IDMS task; cyan), combined nuisance modulation (dark red), and different nuisance modulation subtypes (light red) including visual, target identity, and residual. *Right*: the size of trial variability, computed in a comparable way. Error bars represent standard error across units. Numbers above each type of nuisance modulation indicate its size relative to the target match signal. *B*: the ratios of trial variability (TV) and the target match signal (black) and nuisance modulation (gray) computed as a function of spike count window size. *C*: a simulation extending the plots in *B* to windows of longer duration by rescaling each unit's responses individually (see MATERIALS AND METHODS). Dotted line corresponds to the benchmark of a ratio equal to 1.

removing nuisance only results in a modest increase in d′ across units, with an average increase of 10.1%.

The modest increase in d′ after removal of nuisance modulation is at first glance surprising in light of the fact that nuisance modulations were much larger (2.8×) than the target match signal (Fig. 6*A*), coupled with the intuition that large nuisance modulation should be detrimental to task performance (Fig. 1*B*). However, this result can be understood by examining the trial variability component of the noise, which was 5.2× larger than the target match signal and nearly 2× the size of nuisance modulation (Fig. 6*A*). As a result, trial variability dominated the denominator of the d′ equation. As an illustrative example, compare ratios of the numbers $5/(10 + 100) = 0.045$ versus $5/(0 + 100) = 0.05$: although the first component of the denominator (10) is twofold the size of the numerator (5), including versus excluding it only leads to a change in the total ratio of 10% because the denominator is dominated by the second entry (100). In the case of the data, intact d′ can be approximated from the ratios reported in Fig. 6*A* as $\sqrt{1^2/(2.8^2+5.2^2)}$ (= 0.17), d′ without nuisance can be approximated by $\sqrt{1^2/5.2^2}$ (= 0.19), and the increase is approximated as 14% (where the difference between the 14% approximation and the 10.1% value reported in Fig. 7*A* can be attributed to approximating a ratio after taking the averages vs. computing the ratio in *Eq. 14* for each unit, followed by averaging). Consequently, although the amount of nuisance modulation is large relative to the size of the target match signal, its impact is blunted by the existence of trial variability, which is even larger. In sum, a primary determinant of the impact of nuisance modulation on d′ is the ratio of trial variability over nuisance modulation, and when this ratio is large the existence of nuisance modulation has little consequence for d′. We found this to be the case during the IDMS task.

To illustrate that the amount of nuisance modulation that we observed would have impacted task performance at higher spike counts than we recorded in our data (e.g., if counting windows were longer), we performed a simulation in which we rescaled the responses for each unit in our data (after noise correction; see MATERIALS AND METHODS). Specifically, we kept the proportions and types of signal and nuisance modulation for each unit intact but rescaled the trial-averaged spike count responses for each unit by different factors of *N*, followed by the reintroduction of Poisson trial variability. We then recomputed the impact of nuisance modulation on single-unit d′ as described for Fig. 7*A*. We found that the impact of nuisance on d′ grew substantially with rescaling (Fig. 7*B*). For example, with a sixfold rescaling, which roughly translates into a 1-s counting window (under the assumption that the response properties are constant with time), eliminating nuisance resulted in a 53.0% increase in d′ (compared with the 12.1% increase in simulation with no rescaling; Fig. 7*B*). The increased impact of nuisance with rescaling cannot be attributed to changes in the relative amounts of signal and nuisance modulation, as these remained fixed with rescaling (compare

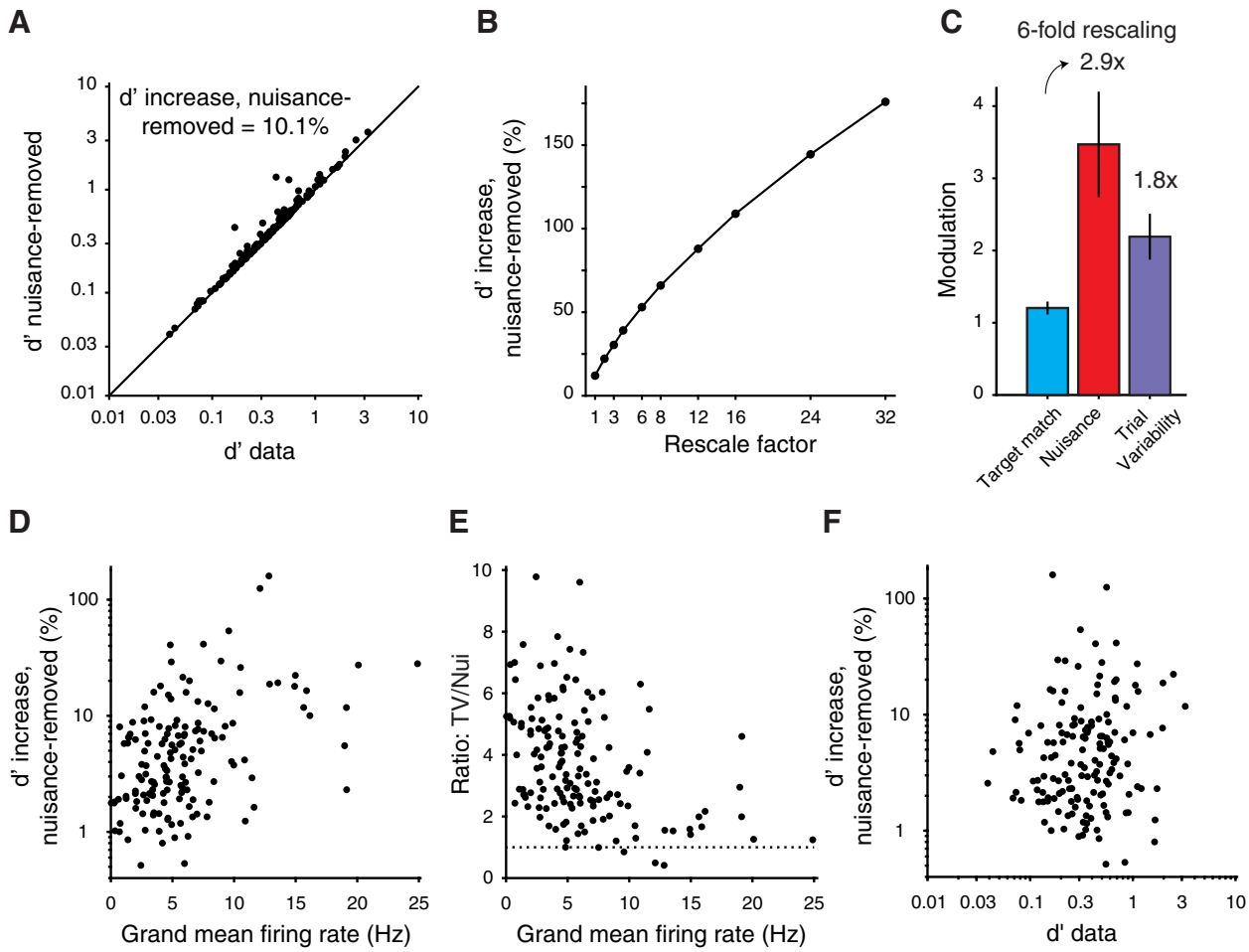

Fig. 7. The impact of nuisance modulation on single-unit task performance (d′). *A*: single-unit d′ computed on the intact data and with the nuisance term set to 0. The impact of nuisance was computed as the average proportional increase in performance when nuisance was removed. *B*: average impact of nuisance modulation on single-unit d′ (computed as described for *A*), applied to data after rescaling firing rates by factors of *N* (see text). *C*: magnitudes of the target match signal, nuisance modulation, and trial variability computed for the 6-fold rescaling to illustrate their ratios. *D*: grand mean firing rate of each unit (computed as described for Fig. 4) plotted against the impact of nuisance modulation on d′; $r = 0.34$; $P = 1.5 \times 10^{-5}$. *E*: grand mean firing rate of each unit plotted against the ratio of trial variability (TV) over nuisance modulation; $r = -0.39$, $P = 5.8 \times 10^{-7}$. *F*: single-unit d′ plotted against the impact of nuisance modulation on d′; $r = 0.04$, $P = 0.64$.

Fig. 6*A* and Fig. 7*C*). Rather, the increased impact of nuisance with rescaling is due to a decrease in magnitude of trial variability relative to the magnitude of nuisance modulation (compare Fig. 6*A* and Fig. 7*C*).

One of the central predictions of the results presented above is that as total spike counts increase, the impact of nuisance modulation should increase as well. Did this predicted relationship hold across the units that we recorded in IT? The grand mean firing rates of individual units were in fact strongly correlated with the impact of nuisance modulation on d′ during the IDMS task (Fig. 7*D*; $r = 0.34$, $P = 1.5 \times 10^{-5}$). Additionally, whereas increases in firing rate corresponded with increases in all three of the variables used to compute d′ (correlations with grand mean firing rate: target match modulation $r = 0.65$, $P = 9.9 \times 10^{-20}$; nuisance modulation $r = 0.73$, $P = 4.6 \times 10^{-27}$; and trial variability $r = 0.9$, $P = 1.7 \times 10^{-51}$), the predicted determinant of the impact of nuisance on d′—the ratio of trial variability over nuisance modulation—was (as expected) negatively correlated with grand mean firing rate (Fig. 7*E*, $r = -0.39$, $P = 5.8 \times 10^{-7}$). These results demonstrate that nuisance modulation is more impactful for units with higher firing rates, because (as predicted) the ratio between trial variability and nuisance modulation decreases as firing rates increase.

Notably, what we did not observe in our data was a significant correlation between the impact of nuisance modulation and d′ (Fig. 7*F*, $r = 0.04$, $P = 0.64$), and we did not observe a correlation between the ratio of trial variability over nuisance and d′ ($r = -0.09$; $P = 0.25$). This can be explained by the fact that the complete transfer of the relationship between firing rate and the impact of nuisance modulation (Fig. 7*D*) to a relationship between d′ and the impact of nuisance modulation (Fig. 7*F*) requires that the variation in d′ across units is determined exclusively by differences in firing rate between units. This is unlikely, as another crucial determinant of d′ is likely to be the type of signals reflected in a unit's responses (e.g., whether or not it carries the target match signal). In other words, units with similar grand mean firing rates are likely to reflect very different types of modulation (e.g., target match, visual, etc.) and thus have very different d′ for the IDMS task, and, similarly, units with very different firing rates could very well have the same d′. Consistent with this interpretation,

although it was the case that d′ and grand mean firing rates were in fact correlated ($r = 0.43$, $P = 4.9 \times 10^{-8}$) and d′ was correlated with all three components required to compute it (correlations with d′: target match modulation $r = 0.90$, $P = 3.3 \times 10^{-57}$; nuisance modulation $r = 0.26$, $P = 0.0012$; trial variability $r = 0.40$; $P = 2.1 \times 10^{-7}$), the amount of target match modulation was an even larger predictor of a unit's d′ (correlation between magnitude of its target match modulation and d′: $r = 0.90$; $P = 3.3 \times 10^{-57}$). After correction for the correlation of target match and nuisance modulation with firing rate, the partial correlation of the target match signal with d′ remained high (partial correlation: $r = 0.91$, $P = 3.97 \times 10^{-60}$), whereas the partial correlation of d′ and nuisance became nonsignificant (partial correlation: $r = -0.08$, $P = 0.30$).

Together, these results indicate that in a fast-processing regime (where spike counts are low), nuisance modulation is largely inconsequential for single-neuron task performance. In contrast, our simulations reveal that mixing signals in the same proportions but in a regime where spike counts are high (e.g., with long integration windows) would be highly detrimental. In the next section, we extend these single-neuron descriptions to describe the impact of trial and nuisance variability on population performance.

When trial variability is large, nuisance modulation remains largely inconsequential for population performance. As we demonstrate in this section, the impact of nuisance modulation on IT performance described above for single units in a low-spike-count regime (Fig. 7) remains modest even when population factors are considered, as long as a basic set of assumptions hold.

When determining population performance, one important factor to consider is the degree to which trial variability and nuisance modulation are correlated between neurons. To summarize the well-established framework for thinking about correlated trial variability (reviewed by Averbeck et al. 2006; Cohen and Kohn 2011; Kohn et al. 2016), when the component of trial variability that falls along a linear decoding axis is uncorrelated between neurons, it will average away as a function of population size. Relative to this benchmark, correlated trial variability has the potential to be either beneficial or detrimental to performance (Fig. 8*A*). Nuisance modulation is similar insofar as the component of nuisance modulation that falls along a linear decoding axis that is uncorrelated between neurons will also average away as a function of population size. Relative to this benchmark, interactions between neurons can configure nuisance modulation to have beneficial or detrimental consequences (Fig. 8*B*).

To illustrate simple benchmarks for thinking about how nuisance variability impacts population performance under the general assumption that trial and nuisance variability are both uncorrelated across units, we begin by considering performance for a simple population comprised of units that all contain the same signal, no nuisance, and uncorrelated trial variability. A well-established behavior of this type of population is that its population d′ will grow with population size as $\sqrt{\sum_1^N \mathrm{d}'^2}$, where N indicates population size (Fig. 9*A*). Conceptually, this performance growth follows from pooling signal across units while averaging away uncorrelated trial variability. Next, we elaborate the scenario by adding nuisance modulation to every unit to decrease its d′ by a constant factor k (where $k < 1$). If the nuisance modulation is independent across units, performance of this population will grow as $\sqrt{\sum_1^N (k\mathrm{d}')^2}$, as both trial and nuisance variability average away at the same rate (Fig. 9*A*). Following on the intuitions presented above for single neurons (Fig. 7), the proportional impact of nuisance can be computed as the proportional in-

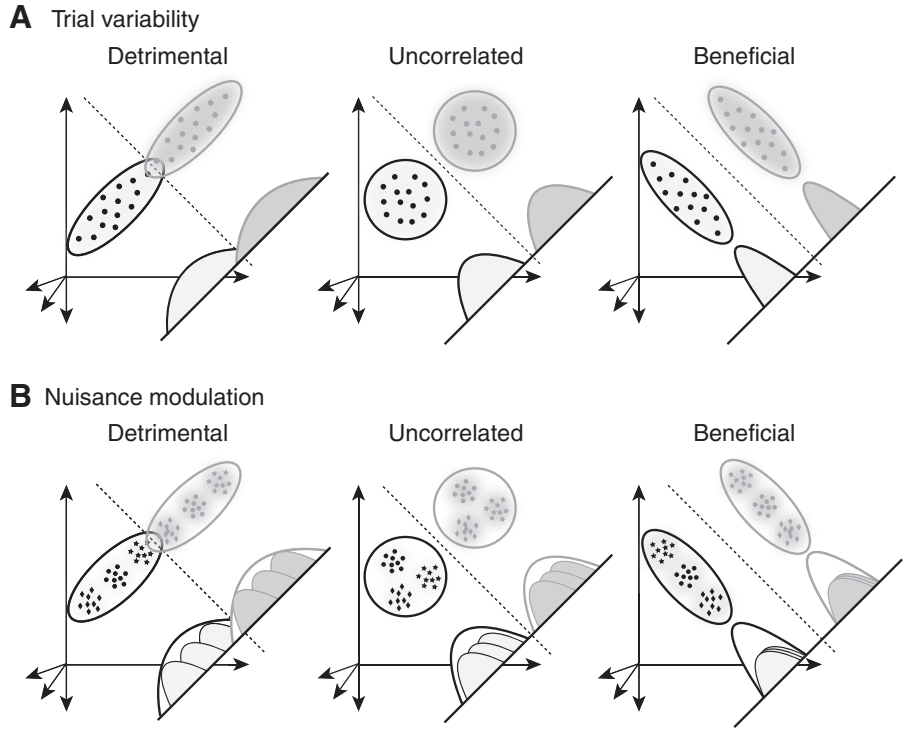

Fig. 8. How correlated trial variability and correlated nuisance modulations can impact task performance. *A*: cartoon depictions of the "beneficial" and "detrimental" impact that correlated trial variability can have on task performance relative to the "uncorrelated" benchmark. Each point depicts a hypothetical population response for a population of 2 neurons on a single trial, and clusters of points depict the dispersion of responses across repeated trials. Dotted lines depict the linear decision boundary optimized for a 2-way classification. Population performance is determined by projecting each class onto an axis perpendicular to the decision boundary. Correlated trial variability between units can be configured to increase or decrease the variance of the projected population response relative to the benchmark of uncorrelated trial variability, and thus have a detrimental or beneficial impact on performance. *B*: same as in *A* but expanded to incorporate correlated nuisance variability. Included are 3 experimental conditions within each set (clusters of points). Like trial variability, correlated nuisance variability between units can be configured to increase or decrease the variance of the projected population response, relative to the benchmark of uncorrelated nuisance variability.

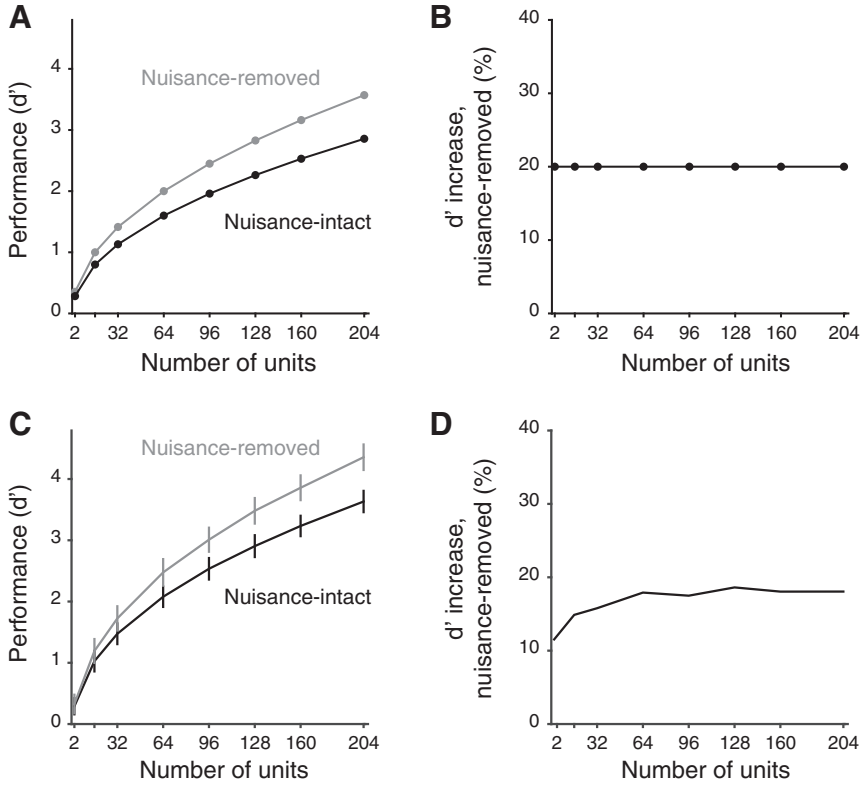

Fig. 9. The proportional impact of nuisance modulation on population performance d′ remains constant as a function of population size. *A*: theoretical benchmarks. Gray, benchmark performance of a population of units with homogeneous signal and independent trial variability computed as $\sqrt{\sum_1^N d'^2}$, where $d' = 0.25$; black, benchmark performance for the same population but with additional independent nuisance variability introduced for each unit that lowers its d′ by k, computed as $\sqrt{\sum_1^N (kd')^2}$, where $k = 0.8$ (corresponding to a 20% reduction in each unit's d′). *B*: proportional impact of nuisance (computed as the proportional increase in performance when nuisance was removed), plotted as a function of population size, computed for the data shown in *A*. *C*: linear decoder performance of the inferotemporal cortex invariant delayed-match-to-sample task data, shown in units of population d′, as a function of population size for 2 simulated populations: "Nuisance-intact," a version of our data in which the responses of each unit are replicated (after noise correction), coupled with independent Poisson trial variability, and "Nuisance-removed," a similar version of our data but with the nuisance modulations for each unit set to 0 (see MATERIALS AND METHODS). Error bars reflect the variability that can be attributed to the random selection of units (for populations smaller than the full data set) and the random assignment of training and testing trials in cross-validation. *D*: proportional impact of nuisance (computed as the proportional increase in performance when nuisance was removed), plotted as a function of population size, computed for the data shown in *C*.

crease in performance when nuisance is removed, which in this case takes on a constant value of $1 - k$ for any population size (Fig. 9*B*). In this example, nuisance modulation reduces both single-neuron as well as population d′ by 20% regardless of population size (Fig. 9*B*). Consequently, for a population with homogeneous neural signals and uncorrelated trial and nuisance variability, if the impact of nuisance modulation on the performance of individual units is modest, its impact will remain modest for the population.

To test these intuitions on recorded neural data, we applied an approach similar in concept to the single-unit analysis presented in Fig. 7*A*, where we estimated the impact of nuisance by comparing the intact data with a hypothetical version of our data with nuisance removed. However, in the case of the population, we did not have an analytical solution and we thus performed pseudosimulations to determine it. To perform this analysis, we simulated the responses of two versions of each unit: an intact version with the same number and types of signals as well as the same grand mean spike count (after noise correction; see MATERIALS AND METHODS) and a version in which the nuisance modulation was removed. In both cases, we simulated trial variability for each unit with an independent Poisson process. Cross-validated linear decoder performance, measured in units of population d′, grew with increasing population size for the intact and nuisance-removed populations with an approximately fixed ratio (Fig. 9*C*). The proportional impact of nuisance modulation as a function of population size was relatively flat and saturated at ~18% with larger-sized populations (Fig. 9*D*). These results confirm the intuitions established by the benchmark presented in Fig. 9, *A* and *B*: the modest impact of nuisance modulation measured in individual units translates to a modest, proportionally constant impact of performance across the population (under the assumption that trial variability is Poisson and is independent between units).

Our simulation-based approach allowed us to estimate the impact of nuisance modulation on population performance relative to a benchmark of the same population but without nuisance. However, our pseudosimulations also incorporate the assumption that trial variability is independent (i.e., uncorrelated) between units, whereas we do in fact expect it to be weakly correlated (see, e.g., Cohen and Maunsell 2009). To assess whether correlated trial and/or nuisance variability had an impact on IT population performance, we also analyzed the raw, simultaneously recorded data within each session.

When a task does not include nuisance variability (e.g., a 2-way discrimination between exactly 2 conditions), the impact of correlated trial variability on population performance can be measured by comparing performance for the simultaneously recorded, intact data with performance when the trials are independently shuffled for each unit to destroy correlations (Averbeck and Lee 2006). Increases in performance with shuffling indicate that noise correlations are detrimental (Fig. 8*A*, *left*), whereas decreases in performance indicate that noise correlations are beneficial (Fig. 8*A*, *right*). This shuffling procedure can be extended for tasks that incorporate a nuisance component by comparing population performance for the intact data with performance when the experimental conditions are shuffled independently for each unit within each class (i.e., shuffling conditions within the set of target matches and within the set of distractors). Here we present the results of this shuffling procedure for simultaneously recorded populations of size 24. Relative to the intact data, shuffling trial variability resulted in a small increase in performance (Fig. 10*A*, "Intact" vs. "Shuffle TV"; proportional increase with shuffling = 8%), indicating that correlated trial variability was aligned along the

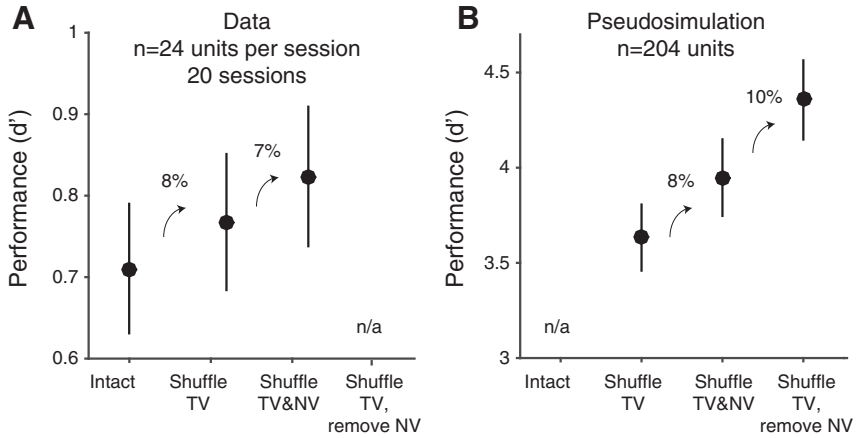

Fig. 10. Understanding how correlated trial variability and correlated nuisance modulations impact task performance. *A*: to assess the impact that correlated trial and nuisance variability between units have on population performance, we applied shuffling procedures to the raw data recorded within each session (across 20 sessions). Shown is linearly decoded population performance (d′) for populations of size 24 for "Intact" (without shuffling), "Shuffle TV" (shuffled trial variability while maintaining nuisance variability correlations intact), and "Shuffle TV&NV" (shuffling both trial and nuisance variability). This analysis cannot be performed in a manner that determines what happens when nuisance variability is removed, indicated by the placeholder "n/a" for comparison with *D*. *B*: The same pseudosimulation data presented in Fig. 9 ($n = 204$). Because those data are simulated as independent between units, the "Intact" condition cannot be computed, as indicated by the placeholder n/a for comparison with *C*. Shown is linearly decoded population performance for: "Shuffle TV"(shuffled trial variability while maintaining nuisance variability correlations intact), "Shuffle TV&NV" (shuffling both trial and nuisance variability), and "Shuffle TV, remove NV" (shuffling trial variability and removing nuisance variability). In both *A* and *B*, numbers above the arrows indicate the proportional increase in d′. Error bars (SE) reflect the variability that can be attributed to the random assignment of training and testing trials in cross-validation.

target match decoding axis in a manner that was weakly detrimental. Next we computed performance when both trial and nuisance variability were shuffled and found that it was slightly higher than shuffling trial variability alone (Fig. 10*A*, "Shuffle TV&NV"; proportional increase = 7%). This suggests that, like trial variability, nuisance variability was correlated in a manner weakly detrimental to performance.

How does the existence of weakly detrimental correlated trial and nuisance variability impact our interpretation of the results presented in Fig. 9? First, we note that the analysis presented in Fig. 9 is not impacted by the existence of correlated trial variability (because any correlations that existed were destroyed in the pseudosimulation process). Second, we note that the ~18% impact of nuisance variability presented in Fig. 9 presents an estimate of the "total" impact arising from what can be considered as two factors: *1*) the impact of nuisance modulations that are uncorrelated and *2*) the additional detrimental impact of nuisance correlations along the decoding axis. To parse their relative contributions, we returned to the pseudosimulation and applied the nuisance shuffling procedure. Shuffling nuisance variability to estimate the second factor led to a small proportional increase (relative to shuffling trial variability alone; Fig. 10*B*, 8%) that was similar to the value measured for the intact data (Fig. 10*A*, 7%, as described above). The remaining proportional impact of (uncorrelated) nuisance modulation, calculated as the increase between shuffling nuisance and removing it altogether, was 10% (Fig. 10*B*; "Shuffle TV&NV" vs. "Shuffle TV, remove NV").

To summarize these results, the impact of nuisance modulation on population performance is predicted to remain proportionally constant as a function of population size under the assumption that signals are homogeneous and that trial and nuisance variability are independent (Fig. 9, *A* and *B*). Consequently, if the impact of nuisance modulation on single-neuron performance is modest (i.e., for a population of size 1) and these assumptions hold, the impact on population performance is expected to remain modest. When measured in IT for the IDMS task under the assumption of uncorrelated, Poisson trial variability, the impact of nuisance modulation was approximately flat as a function of population size and, similar to its impact on single units, remained modest, consistent with these predictions (Fig. 9, *C* and *D*). The impact of nuisance modulation also remained modest when the assumption of uncorrelated trial variability was lifted for measures applied to simultaneously recorded IT population data (Fig. 10). Together, these results suggest that the impact of nuisance modulation on performance during the IDMS task is modest for both IT single units as well as the IT population. We expect these results to generalize to other tasks in which nuisance modulation remains largely uncorrelated along the direction of the linear decoding axis, analogous to results established for correlated trial variability (Fig. 8*B*; Moreno-Bote et al. 2014; see also DISCUSSION).

DISCUSSION

In many everyday situations, we are faced with the challenge of extracting one type of information from our environment while ignoring many other things that are going on around us. This report was inspired by a very simple (albeit ultimately misguided) intuition: when the neurons involved in computing the solutions for these tasks are modulated by both task-relevant signals as well as task-irrelevant nuisance information, nuisance modulation should be a source of noise that limits our ability to perform these tasks (Fig. 1*B*). We found that although this simple intuition is correct when signals are large relative to the size of trial variability, it does not hold when signals are small relative to the size of trial variability—there, nuisance modulation has only a modest impact on single-unit (Fig. 7*A*) as well as population (Fig. 9 and Fig. 10) task performance, even when it is large. This phenomenon can be understood from the perspective of simple algebra in the equation of single-neuron d′ (*Eq. 14*), where the denominator consists of

both nuisance and trial variability terms and large trial variability thus blunts the impact of nuisance variability. Once established for single neurons, the phenomenon extends to population performance, where the proportional impact of nuisance variability is expected to remain constant across all population sizes (including a population size = 1), as long as both signal and trial variability are uncorrelated (Fig. 9). Moreover, a low signal-to-trial variability regime is not exotic or unexpected but follows when spikes are counted in short windows from reasonably well-tuned neurons with reasonable peak firing rates (Fig. 2), as is the case for tasks with fast reaction times. Our results thus reveal that when the brain operates in a fast-processing regime where signals are small relative to the size of trial variability, nuisance modulations are of very little consequence to task performance.

Many of our intuitions about neural coding have been developed within the context of a high-spike-count regime, largely following on foundational work in early and midlevel visual brain areas in primates (e.g., V1, MT) where firing rates are high and/or for tasks in which evidence is integrated over long spike count windows (see, e.g., Roitman and Shadlen 2002). Notably, recent work has called into question whether even in those brain areas high spike counts do in fact translate into a high SNR, due to supra-Poisson trial variability that begins to dominate when spike counts are large (Goris et al. 2014). Moreover, the low-spike-count regime that we present here is likely to be representative of the operating regime in many brain areas during many real-world tasks, particularly for those with small spike count windows, as implied by fast reaction times. As our results demonstrate, within this low-spike-count regime some of the basic intuitions that we have constructed about neural coding in the high-spike-count regime may not hold.

IT is a high-level brain area with some degree of tolerance for nuisance variables, even in the case of object identification. At the same time, IT is not completely invariant but rather retains information about nuisance variables (such as changes in an object's position; Hong et al. 2016). Our results shed insight into why the brain might continue to "mix" modulations for different task-relevant parameters in this way even at the highest stages. Specifically, growing evidence suggests that the brain does not seek to produce neurons with increasingly "pure selectivity" at higher stages of processing but rather the brain continues to mix modulations for different task-relevant parameters within individual neurons, both at the locus at which task-relevant solutions are computed as well as downstream (Freedman and Assad 2009; Kobak et al. 2016; Mante et al. 2013; Meister et al. 2013; Raposo et al. 2014; Rigotti et al. 2013; Rishel et al. 2013; Zoccolan et al. 2007). When this is viewed from the perspective that signal mixing introduces noise in the form of nuisance modulation, one might suspect that one or more of these benefits outweigh the performance costs associated with mixed selectivity. However, as we demonstrate here, within the fast-processing, low-spike-count regime that most of these high-level brain areas are likely to operate in, large nuisance modulations are expected to have only a modest impact on task performance, and thus the cost is minimal.

Our experiments employed a population-based approach in which the experimental conditions were fixed despite the units' preferences and our mapping from neural responses onto population performance assumed a particular type of linear decoder (FLD), which we previously reported to significantly classify correct trials and significantly misclassify error trials for this data set (Roth and Rust 2018). At the same time, behavioral performance on the IDMS task was high overall in both monkeys (on average 96% and 87% correct), implying that the monkeys were not working in a behavioral regime in which nuisance modulation was highly detrimental to behavioral performance. Our results provide insight into why this might be the case, insofar as they reconcile how neural task performance could remain high in the face of the considerable nuisance modulation that coexists with target match modulation in IT. However, issues related to the impact of nuisance modulation on task performance can and should be reexamined in future experiments in the context of other tasks in which nuisance modulation has a nonnegligible behavioral impact, as this places a much stronger constraint on the mapping of neural responses onto behavior. For example, these issues could be explored with even more sensitive measures that might involve tailoring the experimental conditions based on units' response selectivity, motivated by reports in early and midlevel visual brain areas that the most sensitive neurons have SNRs that are comparable to behavior (see, e.g., Cohen and Newsome 2009; Goris et al. 2017; Nienborg and Cumming 2006).

The framework with which we explore how interactions between different units impact population performance builds on foundational work focused on correlated trial variability between units, or "noise correlations" (Averbeck et al. 2006; Cohen and Kohn 2011; Kohn et al. 2016). Recent work has emphasized the importance of measuring not just the degree to which neurons are correlated but also how those correlations align with a decoding axis and thus how they impact performance (Moreno-Bote et al. 2014). As an extension of the ideas presented in Fig. 8, these ideas can be envisioned as determining the degree to which the elongations imposed by nuisance variability (i.e., the long axes of the ovals in Fig. 8) are aligned versus misaligned with the decoding axis (i.e., the line). In the delayed-match-to-sample search task that we present here, we found that population performance conformed to the predictions established for populations in which nuisance modulation is uncorrelated. Similarly, one study examined the decoding of orientation information from the V1 population and the impact of another, internal factor that could be considered as a nuisance variable, global fluctuations in firing rate, and determined that this form of nuisance had little impact on population decoding performance (Arandia-Romero et al. 2016). However, in other tasks, nuisance interactions may be inherently more aligned with the task-relevant decoding axis and thus be much more impactful. Such is the case for dissociating self versus object motion, where the brain appears to employ a clever encoding strategy of mixing populations of congruently and incongruently tuned neurons to cancel correlated nuisance variability (Kim et al. 2016).

Our results also speak to general misinterpretations about whether and how a linear population decoder applied to a "tolerant" population can produce an "invariant" response. That is, there is a general sense that even under conditions in which nuisance modulation is detrimental to performance at the level of single neurons, this can be remedied across a population to

produce nuisance modulation "invariance" by combining neural responses via a weighted linear readout, under standard conditions. This type of scenario has been investigated most extensively for the case of invariant object recognition, formulated as the identification of objects across variation in details such as their position, size, and background context (reviewed by DiCarlo et al. 2012). At higher stages of the ventral visual pathway (e.g., IT), information about object identity is more accessible to a weighted linear readout compared with earlier stages (DiCarlo and Cox 2007; Hung et al. 2005; Rust and DiCarlo 2010). Linear separability of the IT population has been demonstrated to be supported not by individual neurons that are themselves invariant or insensitive to identity-preserving transformations but rather neurons that exhibit "tolerance" to nuisance modulation in the form of maintaining their rank-order tuning for object identity in the face of identity-preserving variation (Li et al. 2009). The general intuition behind how the weighted linear readout works is that, by considering the responses across the population, a weighted readout can extract each neuron's signal while averaging away noise, including both trial variability as well as nuisance modulation, when nuisance modulation is formatted in an appropriate manner (Fig. 8). However, as we demonstrate, under a standard set of assumptions, an upper bound exists for the effectiveness of this averaging operation, and the consequence is that the proportional impact on task performance will remain fixed as a function of population size. In other words, a weighted linear readout of "tolerant" units can never extract perfect nuisance modulation "invariance" when the standard assumptions hold.

ACKNOWLEDGMENTS

We thank Marino Pagan for valuable insights and Margot P. Wohl and Krystal Henderson for technical contributions.

GRANTS

This work was supported by the National Eye Institute (Grant R01 EY-020851), the Simons Foundation (through an award from the Simons Collaboration on the Global Brain), and the McKnight Endowment for Neuroscience.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

N.R. and N.C.R. performed experiments; N.R. and N.C.R. analyzed data; N.R. and N.C.R. interpreted results of experiments; N.R. and N.C.R. prepared figures; N.R. and N.C.R. drafted manuscript; N.R. and N.C.R. edited and revised manuscript; N.R. and N.C.R. approved final version of manuscript.

REFERENCES

Arandia-Romero I, Tanabe S, Drugowitsch J, Kohn A, Moreno-Bote R. Multiplicative and additive modulation of neuronal tuning with population activity affects encoded information. *Neuron* 89: 1305–1316, 2016. doi:10.1016/j.neuron.2016.01.044.

Averbeck BB, Latham PE, Pouget A. Neural correlations, population coding and computation. *Nat Rev Neurosci* 7: 358–366, 2006. doi:10.1038/nrn1888.

Averbeck BB, Lee D. Effects of noise correlations on information encoding and decoding. *J Neurophysiol* 95: 3633–3644, 2006. doi:10.1152/jn.00919.2005.

Cohen MR, Kohn A. Measuring and interpreting neuronal correlations. *Nat Neurosci* 14: 811–819, 2011. doi:10.1038/nn.2842.

Cohen MR, Maunsell JH. Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* 12: 1594–1600, 2009. doi:10.1038/nn.2439.

Cohen MR, Newsome WT. Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *J Neurosci* 29: 6635–6648, 2009. doi:10.1523/JNEUROSCI.5179-08.2009.

DiCarlo JJ, Cox DD. Untangling invariant object recognition. *Trends Cogn Sci* 11: 333–341, 2007. doi:10.1016/j.tics.2007.06.010.

DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? *Neuron* 73: 415–434, 2012. doi:10.1016/j.neuron.2012.01.010.

Freedman DJ, Assad JA. Distinct encoding of spatial and nonspatial visual information in parietal cortex. *J Neurosci* 29: 5671–5680, 2009. doi:10.1523/JNEUROSCI.2878-08.2009.

Goris RL, Movshon JA, Simoncelli EP. Partitioning neuronal variability. *Nat Neurosci* 17: 858–865, 2014. doi:10.1038/nn.3711.

Goris RL, Ziemba CM, Stine GM, Simoncelli EP, Movshon JA. Dissociation of choice formation and choice-correlated activity in macaque visual cortex. *J Neurosci* 37: 5195–5203, 2017. doi:10.1523/JNEUROSCI.3331-16.2017.

Haefner RM, Bethge M. Evaluating neural codes for inference using Fisher Information. In: *Advances in Neural Information Processing Systems 23*, edited by Lafferty J. Montreal, Canada: NIPS, 2010, vol. 2, p. 1993–2001.

Hong H, Yamins DL, Majaj NJ, DiCarlo JJ. Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat Neurosci* 19: 613–622, 2016. doi:10.1038/nn.4247.

Hung CP, Kreiman G, Poggio T, DiCarlo JJ. Fast readout of object identity from macaque inferior temporal cortex. *Science* 310: 863–866, 2005. doi:10.1126/science.1117593.

Kim HR, Pitkow X, Angelaki DE, DeAngelis GC. A simple approach to ignoring irrelevant variables by population decoding based on multisensory neurons. *J Neurophysiol* 116: 1449–1467, 2016. doi:10.1152/jn.00005.2016.

Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, Qi XL, Romo R, Uchida N, Machens CK. Demixed principal component analysis of neural population data. *eLife* 5: e10989, 2016. doi:10.7554/eLife.10989.

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

Li N, Cox DD, Zoccolan D, DiCarlo JJ. What response properties do individual neurons need to underlie position and clutter "invariant" object recognition? *J Neurophysiol* 102: 360–376, 2009. doi:10.1152/jn.90745.2008.

Mante V, Sussillo D, Shenoy KV, Newsome WT. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503: 78–84, 2013. doi:10.1038/nature12742.

Meister ML, Hennig JA, Huk AC. Signal multiplexing and single-neuron computations in lateral intraparietal area during decision-making. *J Neurosci* 33: 2254–2267, 2013. doi:10.1523/JNEUROSCI.2984-12.2013.

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

Nienborg H, Cumming BG. Macaque V2 neurons, but not V1 neurons, show choice-related activity. *J Neurosci* 26: 9567–9578, 2006. doi:10.1523/JNEUROSCI.2256-06.2006.

Pagan M, Rust NC. Quantifying the signals contained in heterogeneous neural responses and determining their relationships with task performance. *J Neurophysiol* 112: 1584–1598, 2014. doi:10.1152/jn.00260.2014.

Pagan M, Simoncelli EP, Rust NC. Neural quadratic discriminant analysis: nonlinear decoding with V1-like computation. *Neural Comput* 28: 2291–2319, 2016.

Pagan M, Urban LS, Wohl MP, Rust NC. Signals in inferotemporal and perirhinal cortex suggest an untangling of visual target information. *Nat Neurosci* 16: 1132–1139, 2013. doi:10.1038/nn.3433.

Parker AJ, Newsome WT. 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.

Raposo D, Kaufman MT, Churchland AK. A category-free neural population supports evolving demands during decision-making. *Nat Neurosci* 17: 1784–1792, 2014. doi:10.1038/nn.3865.

Rigotti M, Barak O, Warden MR, Wang XJ, Daw ND, Miller EK, Fusi S. The importance of mixed selectivity in complex cognitive tasks. *Nature* 497: 585–590, 2013. doi:10.1038/nature12160.

Rishel CA, Huang G, Freedman DJ. Independent category and spatial encoding in parietal cortex. *Neuron* 77: 969–979, 2013. doi:10.1016/j.neuron.2013.01.007.

Roitman JD, Shadlen MN. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* 22: 9475–9489, 2002. doi:10.1523/JNEUROSCI.22-21-09475.2002.

Roth N, Rust NC. Inferotemporal cortex multiplexes behaviorally-relevant target match signals and visual representations in a manner that minimizes their interference. *PLoS One* 13: e0200528, 2018. doi:10.1371/journal.pone.0200528.

Ruff DA, Ni AM, Cohen MR. Cognition as a window into neuronal population space. *Annu Rev Neurosci* 41: 77–97, 2018. doi:10.1146/annurev-neuro-080317-061936.

Rust NC, DiCarlo JJ. 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.

Zoccolan D, Kouh M, Poggio T, DiCarlo JJ. Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *J Neurosci* 27: 12292–12307, 2007. doi:10.1523/JNEUROSCI.1897-07.2007.