Modeling the Diverse Effects of Divisive Normalization on Noise Correlations

Oren Weiss[1], Hayley A. Bounds[2], Hillel Adesnik[3], and Ruben Coen-Cagli[1,4,5]

[1] *Department of Systems and Computational Biology, Albert Einstein College of Medicine*

[2] *Helen Wills Neuroscience Institute, University of California, Berkeley*

[3] *Department of Molecular and Cell Biology, University of California, Berkeley*

[4] *Dominick P. Purpura Department of Neuroscience, Albert Einstein College of Medicine*

[5] *Department of Ophthalmology and Visual Sciences, Albert Einstein College of Medicine*

June 8, 2022

Abstract

Divisive normalization, a prominent model of interactions between neurons, plays a key role in influential theories of neural coding. Testing the mechanistic basis of those theories requires understanding the relationship between normalization and the statistics of neural activity, particularly noise correlations which provide strong constrains on circuit models. Yet the relation between normalization and correlations remains unclear, because models of covariability typically ignore normalization despite empirical evidence suggesting it contributes to stimulus and state modulations of covariability. We propose a pairwise stochastic divisive normalization model that accounts for the effects of normalization and other factors on covariability, and accurately fits Ca^{2+} imaging data from mouse primary visual cortex (V1). We show that normalization modulates noise correlations in qualitatively different ways depending on whether normalization is shared between neurons, and we outline the data regime in which the

corresponding model parameter is identifiable. Our analysis provides the first quantitative evidence that similarly tuned neurons in mouse V1 share their normalization signals, and explains our observation that stronger normalization decorrelates those pairs, particularly at low stimulus contrast. Our model could enable quantifying the role of normalization on information transmission and representation, and mapping the parameters of normalization onto circuit and cellular mechanisms.

1 Introduction

Neurons in the sensory cortices of the brain exhibit substantial response variability across identical experimental trials [Shadlen and Newsome, 1998; Tolhurst, Movshon, and Dean, 1983]. These fluctuations in activity are often shared between pairs of simultaneously recorded neurons, called noise correlations [Cohen and Kohn, 2011]. Because the presence of these correlations can constrain the amount of information encoded by neural populations and impact behavior [Rumyantsev et al., 2020; Kafashan et al., 2021; Bartolo et al., 2020; Abbott and Dayan, 1999; Averbeck and Lee, 2006; Kohn et al., 2016; Hu, Zylberberg, and Shea-Brown, 2014; Moreno-Bote et al., 2014; Zohary, Shadlen, and Newsome, 1994; Kanitscheider, Coen-Cagli, and Pouget, 2015], noise correlations have been widely studied. This work has also revealed that correlations are often modulated by stimulus and state variables [Cohen and Kohn, 2011], and therefore can play an important role in computational theories of sensory coding. For instance, noise correlations could emerge from neurons performing Bayesian inference and reflect the statistics of sensory inputs [Bányai et al., 2019; Berkes et al., 2011; Orbán et al., 2016; Orbán and Bányai, 2019] and prior expectations [Haefner, Berkes, and Fiser, 2016; Lange and Haefner, 2017; 2022]. From a mechanistic point of view, such a statistical structure of noise correlations poses strong constraints on circuit models of cortical activity [Bondy, Haefner, and Cumming, 2018; Doiron et al., 2016; Litwin-Kumar and Doiron, 2012; Hennequin et al., 2018; Heeger and Zemlianova, 2020]. To better understand the functional impact and underlying mechanisms of noise correlations on neural

coding and behavior, we need to be able to quantitatively characterize and interpret how noise correlations in neural populations are affected by experimental variables.

For this reason, successful descriptive models of neural activity have been developed to capture noise correlations [Pillow et al., 2008; Archer et al., 2014; Gardella, Marre, and Mora, 2019; Schneidman et al., 2006; Granot-Atedgi et al., 2013; Zhao and Park, 2017; Sokoloski, Aschner, and Coen-Cagli, 2021; Josić et al., 2009]. However, none of those models considers divisive normalization [Carandini and Heeger, 2012; Heeger, 1992; Albrecht and Geisler, 1991], an operation observed in a wide range of neural systems which has also been implicated in modulating the structure of noise correlations. Experimental phenomena that are accompanied by changes in noise correlations, including contrast saturation [Kohn and Smith, 2005], surround suppression [Liu, Haefner, and Pack, 2016; Snyder et al., 2014; Henry and Kohn, 2020], and attentional modulations of neural activity [Cohen and Maunsell, 2009; Ruff and Cohen, 2014; Mitchell, Sundberg, and Reynolds, 2009] have been successfully modeled using divisive normalization [Cavanaugh, Bair, and Movshon, 2002; Heeger, 1992; Reynolds and Heeger, 2009; Coen-Cagli, Dayan, and Schwartz, 2012], although those models only captured average firing rates of individual neurons. Additionally, some numerical simulation studies have shown how normalization can affect noise correlations [Tripp, 2012; Verhoef and Maunsell, 2017]. These results indicate that it is important to quantify the relative contribution of normalization and other factors to modulation of noise correlations.

We propose a stochastic normalization model, the pairwise Ratio of Gaussians (RoG), to capture the across-trial joint response statistics for pairs of simultaneously recorded neurons. This builds on our previous method, in this paper called the independent RoG, that considered the relationship between normalization and individual neural response variability [Coen-Cagli and Solomon, 2019]. In these RoG models, neural responses are described as the ratio of two random variables: the numerator, which represents excitatory input to the neu-

ron, and the denominator (termed normalization signal), which represents the suppressive effect of summed input from a pool of neurons [Carandini and Heeger, 2012; Sawada and Petrov, 2017]. Our pairwise RoG allows for the numerators and denominators that describe the individual neural responses to be correlated across pairs; these correlations then induce the ratio variables (i.e. the model neurons' activity) to also be correlated. In this paper, we derive and validate a Gaussian approximation to the joint distribution of pairwise responses, which greatly simplifies the problem of fitting the model and interpreting its behavior. We demonstrate that the model fits accurately pairwise neural responses recorded in the mouse primary visual cortex (V1). The model provides a mathematical relationship between noise correlations and normalization, which predicts qualitatively different effects of normalization on noise correlations, depending on the relative strength of the correlation between numerators and between denominators. We show that this captures the varied modulations of noise correlations by stimulus contrast in our data, and could explain the diversity of modulations of noise correlations observed in past work [Ruff and Cohen, 2014; 2016; Verhoef and Maunsell, 2017]. This analysis also provides the first direct quantification that normalization signals are shared between neurons with similar tuning. Lastly, to provide practical guidance for future applications of our model, we outline the regimes of parameter identifiability, and illustrate the conditions under which our pairwise RoG affords better estimates of single-trial normalization signals compared to the independent RoG.

Our results highlight the importance of modeling the relation between normalization and covariability to interpret the rich phenomenology of noise correlations. Our model and code provide a data-analytic tool that allows researchers to further investigate such a relationship, and to quantitatively evaluate predictions made by normative and mechanistic models regarding the role of correlated variability and normalization in neural coding and behavior.

2 Results

2.1 Pairwise Ratio of Gaussians Model Captures Correlated Variability in Mouse V1

We developed the pairwise Ratio of Gaussians model (RoG, Fig 1a) to quantify the relationship between normalization and response covariability (i.e. noise correlations) that has been suggested in some empirical studies of neural activity in visual cortex [Lin et al., 2015; Rikhye and Sur, 2015; Ruff and Cohen, 2016; Verhoef and Maunsell, 2017].

In the standard divisive normalization model (Eq. 2), the mean response is computed as the ratio of the excitatory drive (numerator $\boldsymbol{N}$) to a normalization signal summarizing inhibitory drive (denominator $\boldsymbol{D}$). Our pairwise RoG considers a pair of neurons where each individual neural response is well characterized by the standard normalization equation with corresponding numerators (N_1, N_2) and denominators (D_1, D_2). We then assume that the numerators and denominators are (bivariate) Gaussian random vectors - which allows the possibility for correlations to exist among the numerators (denoted ρ_N) and among the denominators (ρ_D). From this, we derived equations for the mean responses and covariance matrix of the pair as a function of the numerators and denominators (Eqs. 5, 6). These equations depend on the Gaussian approximation to the ratio of two Gaussian random variables. We verified the validity of this approximation, by simulating the activity of pairs of neurons, and comparing the covariance (Fig 1b,c) and correlation (Fig 1d,e) of the true ratio distribution and of the approximate distribution (Eqs. 7, 8). The mean and variance are identical to the independent RoG model (a special case of the pairwise RoG, with numerators and denominators independent between neurons) which we validated previously [Coen-Cagli and Solomon, 2019].

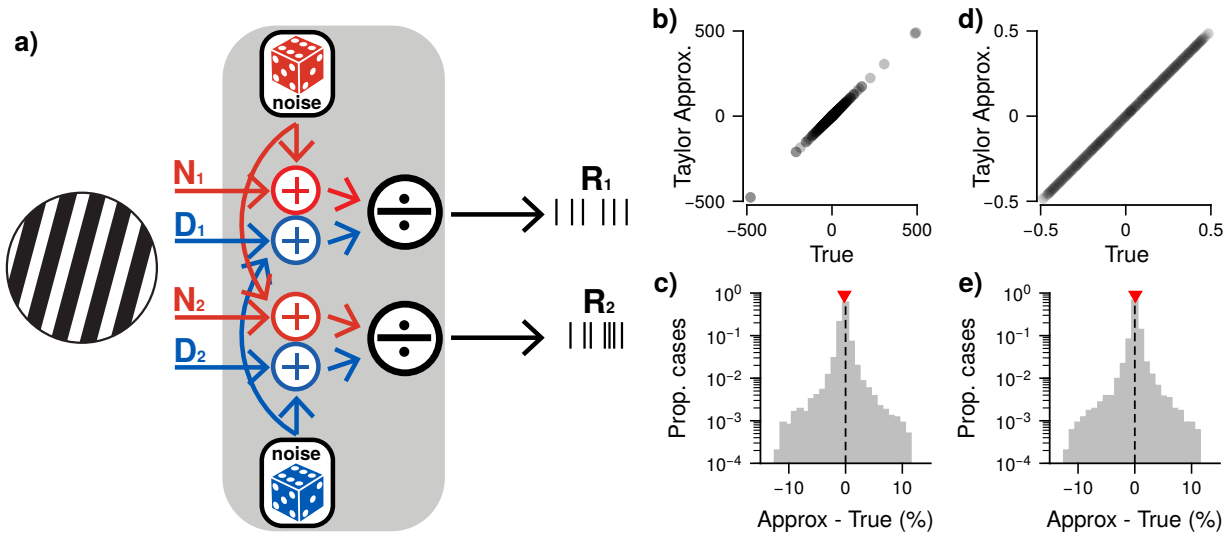


Figure 1: Definition and Validation of the Pairwise Ratio of Gaussians Model
(a) The pairwise RoG model describes pairs of neural responses (R_1, R_2), where each response is computed as the ratio of two stimulus-driven signals: numerators (N_1, N_2) and denominators (D_1, D_2). These variables are corrupted by additive bivariate Gaussian noise with a power-law variance-mean relationship.
(b-e) Comparison of the normal approximation we derived for the pairwise RoG noise covariance (b,c) and noise correlation (d,e) and the true values (estimated across $1e6$ simulated trials) for $1e4$ experiments (i.e. simulated pairs of neural responses). Each experiment used different model parameters and each trial was randomly drawn from the corresponding distribution. (b,d) scatter plot; (c,e) histogram of the percent difference between the Taylor approximation and the true value. The red marker indicates the median percent difference.
Model parameters were drawn uniformly from the following intervals:
$\boldsymbol{\mu_N} \in [0, 100]$, $\boldsymbol{\mu_D} \in [0.5, 1.5]$, $\boldsymbol{\mu_\eta} = 0$, $\boldsymbol{\alpha_N} = 1$, $\boldsymbol{\alpha_D} = 0.001$, $\boldsymbol{\beta_N}, \boldsymbol{\beta_D} \in [1, 1.5]$, $\rho_N, \rho_D \in [-0.5, 0.5]$, $\boldsymbol{\sigma_\eta^2} = 0.1 \cdot \boldsymbol{\sigma_N^2}/\boldsymbol{\sigma_D^2}$, $\rho_\eta = 0$. The ranges of the mean parameters were chosen to reproduce realistic firing rates of V1 cortical neurons, while the $\boldsymbol{\alpha}, \boldsymbol{\beta}$ parameters were chosen such that the variances of the $\boldsymbol{N}$ and $\boldsymbol{D}$ are relatively small and the probability that $\boldsymbol{D} \leq 0$ is negligible [2013]. See Methods for details.

Next, to test how well the model captures experimental data, we applied it to Ca^{2+} imaging data recorded in V1 of mice responding to sinusoidal gratings of varying contrast levels (see Methods 4.5). We analyzed neurons that were strongly driven by the visual stimuli (N=297 neurons, 5597 pairs; see Methods for inclusion criteria). We focused on stimulus contrast tuning (Eq. 9) because the formulation of the corresponding standard normalization model captures firing rate data well [Heeger, 1992], and visual contrast affects both normalization strength and the strength of noise correlations [Kohn and Smith, 2005].

Because the RoG framework has not been validated before on mouse V1 data, we first applied the independent RoG and found that it provided excellent fits (average cross-validated goodness of fit 0.96, 5596/5597 pairs with goodness of fit > 0.5) on par with that found in macaque V1 data recorded with electrode arrays [Coen-Cagli and Solomon, 2019], thus demonstrating that the RoG framework is flexible enough to capture datasets with different statistics. In both cases, the analysis was performed on visually responsive neurons, that therefore exhibited strong contrast tuning of the firing rate, partly explaining the very high performance. However, the independent RoG could not capture correlated variability, which was prominent in the data (median noise correlation across 5597×7 pairs and contrasts 0.116, c.i. [0.114, 0.120]; noise correlation was significantly different from zero in 95% of cases).

Therefore, we tested if the pairwise RoG could capture correlated variability in the data. Figure 2a-f demonstrates that the model is able to capture contrast-dependent noise correlations, both for pairs with positive (example in Fig 2a,c,e; 4371/5597 pairs) and negative median noise correlations (example in Fig 2b,d,f; 1226/5597 pairs). Importantly, even though the ρ_N, ρ_D parameters were stimulus-independent, the pairwise model captured substantial changes in noise correlations with contrast for the majority of pairs analyzed (3076/5597 had greater than 0.5 correlation between the noise correlation values of data and model fit, across contrast levels). Over the entire dataset, the pairwise model slightly outperformed the independent model for the majority of the pairs (mean percent difference of negative log-likelihoods $=1.188, p<10^{-15}$, 3924/5597 pairs with pairwise outperforming independent). Additionally, in line with our predictions from synthetic data (see Appendix 5.4, Fig. S1), the difference between pairwise and independent models increased for pairs of neurons with larger noise correlations (Fig. 2g).

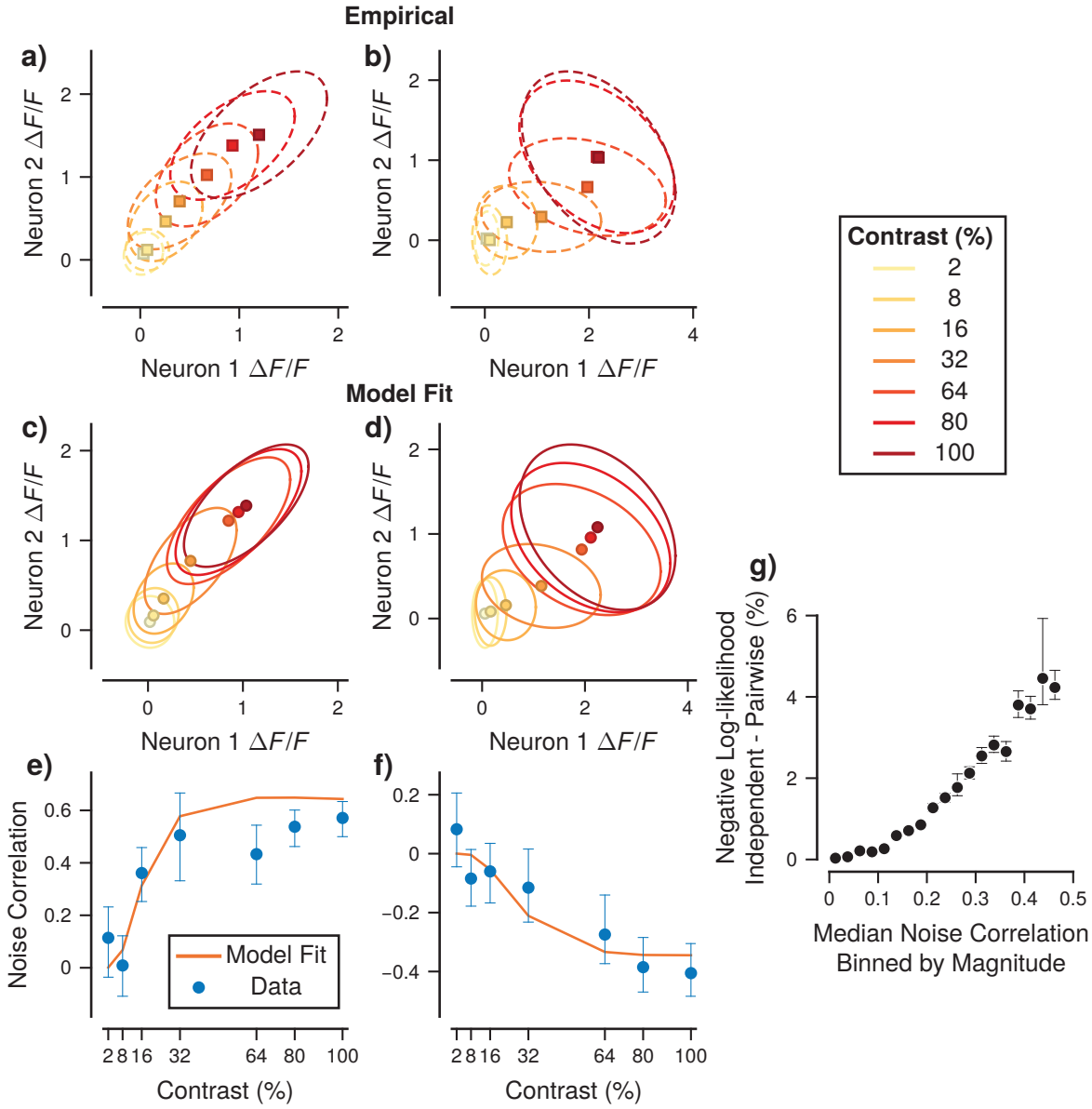


Figure 2: Pairwise RoG Captures Contrast-Dependent Noise Correlations in Mouse V1
(a-f) Pairwise neural responses for two example pairs of neurons in mouse V1 with positive (left column) and negative (right column) median noise correlations. (a,b) shows the empirical means (squares) and covariance ellipses ($\sim$ 1 standard deviation from the empirical mean) for the pairwise responses. (c,d) shows the model fit means (circles) and covariance ellipses. (e,f) compares the relationship between contrast level and noise correlations for model fit and data (error bars are 68% bootstrapped confidence interval). The pair with positive correlations had 93 repeats of each stimulus contrast, the pair with negative correlations had 68 repeats. (g) Model comparison between the pairwise and independent models (error bars 68% bootstrapped confidence interval). Neural pairs are binned by the magnitude of the median noise correlations (across contrasts). Contrast levels are {2,8,16,32,64,80,100}

These results demonstrate the validity of the Gaussian approximation underlying our model, and that the pairwise RoG captures a range of effects of stimulus contrast on noise correlations observed in experimental data.

2.2 Modulations of Correlated Variability Depend on Sharing of Normalization

The results in the previous section illustrate that the model captures both increasing and decreasing trends of noise correlations with contrast. To better interpret these trends and further understand the effects of normalization, we next studied the role of the different model parameters. Specifically, within the RoG modeling framework, there are two main sources of response variability: the numerator (excitatory drive) and the denominator (normalization). Depending on the value of the corresponding ρ parameters, each of these sources of variability can be independent ($\rho = 0$) or shared ($\rho \neq 0$), and therefore contribute differently to noise correlations. Consequently, understanding modulations of noise correlations in the pairwise RoG requires understanding how normalization and external stimuli affect the relative amplitude these sources, and how the effects depend on whether those sources are shared.

First, we studied the relationship between normalization and noise correlations for the lowest contrast stimuli (Fig 3, yellow symbols). When the normalization signals are positively correlated (shared normalization, $\rho_D = 0.5$) but the excitatory drive is independent ($\rho_N = 0$), increasing normalization strength tends to decrease the magnitude of noise correlations (Fig 3a). Conversely, when the normalization signals are independent ($\rho_D = 0$) but there is shared driving input ($\rho_N = 0.5$), the magnitude of noise correlations tends to increase with increasing normalization (Fig 3b). Intuitively, this reflects that increasing the average normalization strength reduces the contribution of fluctuations in normalization signal to the overall response fluctuations [2019]. Therefore, increasing normalization strength reduces a

source of noise correlations when normalization is shared, whereas it reduces a source of independent noise when normalization is not shared.

We found similar effects of normalization on noise correlations at high contrast levels (Fig 3, orange and red symbols) although the slope of the relationship became shallower as contrast increased. This is due to the saturating effect of the contrast response function (Eq. 9): at high contrast, (co)fluctuations of the normalization signal across trials have relatively little effect on the responses of a neural pair, so the correlation in neural responses will be relatively unaffected by normalization strength. We also observed that the magnitude of noise correlations generally increased with contrast when the denominators were correlated and the numerators independent (Fig 3a and S2), whereas it decreased when the numerators were correlated (Fig 3b,c and S2). Therefore, although increasing stimulus contrast increases normalization strength, noise correlations can be modulated in different ways by contrast and by normalization, because changing stimulus contrast also affects the numerator term.

Indeed, these results can be derived from examining the equation for the correlation in the pairwise RoG (Eq .8) rearranged as follows (see Appendix 5.1 for more details):

$$\operatorname{corr}(R_1, R_2) = \frac{\rho_N}{\sqrt{\left(\frac{\delta_{D_1}^2}{\delta_{N_1}^2}+1\right)\left(\frac{\delta_{D_2}^2}{\delta_{N_2}^2}+1\right)}} + \frac{\rho_D}{\sqrt{\left(\frac{\delta_{N_1}^2}{\delta_{D_1}^2}+1\right)\left(\frac{\delta_{N_2}^2}{\delta_{D_2}^2}+1\right)}} \tag{1}$$

and by recognizing that the coefficient of variation of D_i (δ_{D_i}) is a decreasing function of normalization strength (μ_D), whereas the ratio $\frac{\delta_{D_i}}{\delta_{N_i}}$ is often an increasing function of contrast.

We further analyze this equation, first in the case of changing normalization strength while keeping contrast fixed. The term on the right-hand side proportional to ρ_N is an increasing function of the mean normalization strength since the denominator is a decreasing function of μ_D. Conversely, the term proportional to ρ_D decreases with normalization since

the denominator increases with μ_D. In these two cases, the monotonic dependence of noise correlation on normalization strength is guaranteed by Eq. 1 regardless of the specific parameter values. Similar patterns emerge when $\rho_N, \rho_D < 0$, except the signs of the noise correlations are reversed (Fig S2). When the correlations of input and normalization signals are both different from zero the relationship between noise correlation and normalization strength resembles a combination of the two previously described scenarios, and the specific parameter values determine which of the two terms in Eq. 1 dominates. For instance, in our simulations with $(\rho_N = 0.5, \rho_D = 0.5)$ (Fig 3c), the magnitude of noise correlations increases with normalization strength on average similar to (Fig 3b), indicating that the magnitude of the term proportional to ρ_N is usually larger than the magnitude of the ρ_D term, but there is more variability, as evidenced by the increased spread of the scatter. When the input strength and normalization signal have opposite correlations (e.g. $\rho_N = 0.5, \rho_D = -0.5$), we obtained similar results but the magnitude of noise correlations was on average closer to 0 due to cancellation between the two sources of covariability (Fig S2). A similar analysis of Eq. 1 shows that the effects of stimulus contrast are opposite to those of a pure change of normalization strength, because $\frac{\delta_{D_i}}{\delta_{N_i}}$ is often an increasing function of contrast but a decreasing function of normalization strength. Notice that we assumed there is no residual noise component ($\eta \equiv 0$), but all the analyses above remain valid when the amplitude σ_η is relatively small compared to (δ_N, δ_D) (Appendix, Eq. 19).

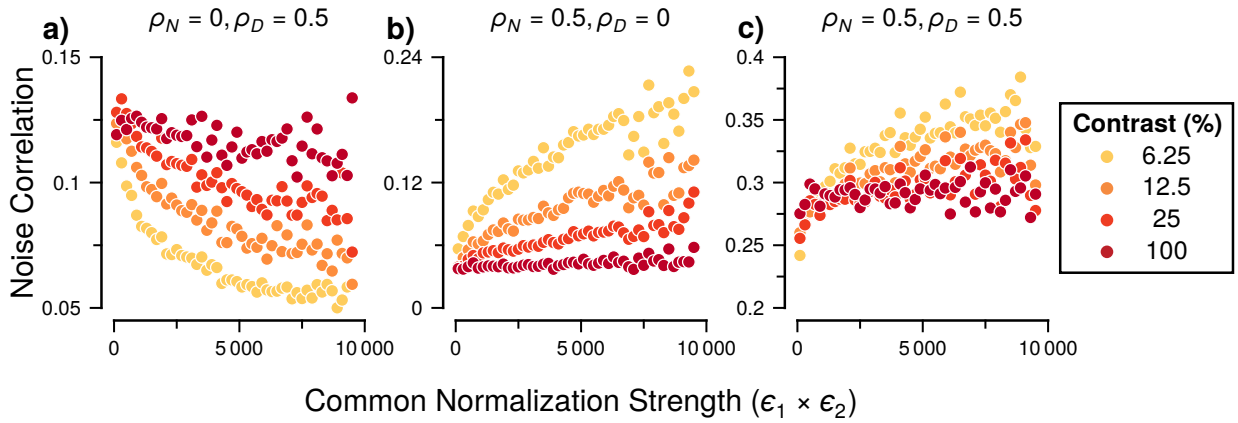


Figure 3: Relationship between Noise Correlations, Normalization Strength and Contrast Depends on the Source of Variability
Each plot in the grid shows, for a given combination of ρ_N, ρ_D, the median noise correlation of all generated neural pairs binned according to $\epsilon_1\epsilon_2$, a contrast independent measure of the common normalization strength. Bins with less than 100 pairs were discarded. Neural responses were generated from the contrast-response parametrization (Eq. 9). Noise correlation strength was computed across 1e3 simulated trials drawn from the pairwise RoG model. For each contrast level and combination of ρ_N, ρ_D, 1e5 simulated experiments were created.
Model parameters were drawn uniformly from the following intervals:
$\boldsymbol{R_{max}} \in [5, 50],\ \boldsymbol{\epsilon} \in [10, 100],\ \boldsymbol{\alpha_N}, \boldsymbol{\alpha_D} \in [0.1, 1],\ \boldsymbol{\beta_N}, \boldsymbol{\beta_D} \in [1, 2],\ \boldsymbol{\eta} \equiv 0$

In summary, our analysis shows that normalization and stimulus contrast can have diverse effects on noise correlations, depending on whether neurons share their normalization signals and on the interplay between multiple sources of variability.

2.3 Identifiability of Correlation Parameters Depends on Noise Correlations

The above analysis demonstrates that the relationship between noise correlations and normalization depends on how this correlated variability arises: either through cofluctuations in the excitatory drive or normalization signal (determined by ρ_N, ρ_D respectively). To employ these insights when fitting to data, we need to know how well we can recover these parameters from data.

To do so, we generated synthetic neural data using realistic values for the single-neuron

parameters and systematically varying the values of the ρ parameters (see Methods 4.4). We then fit the pairwise RoG to these synthetic data and assessed the quality of the maximum likelihood estimators for ρ_N and ρ_D using the mean squared error between the estimates and the ground truth. The model recovered the ρ_N parameter for the majority of the simulations (8490/11628 pairs with lower error than a random estimator, see Fig. 4 caption), but was only able to do so for ρ_D in a smaller fraction of simulations (5429/11628).

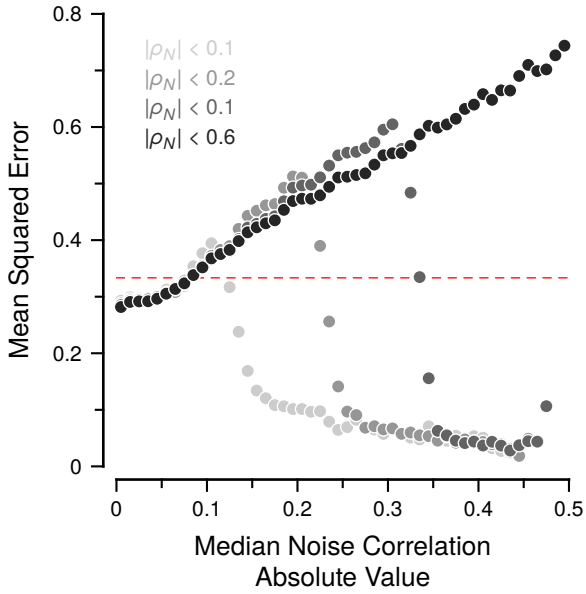


Figure 4: ρ_D is Identifiable when the Magnitude of the Noise Correlations Exceeds ρ_N
Plot generated using 11628 synthetic pairs. The single-neuron parameters were extracted from fits to electrophysiology data [2019] (Methods Section 4.4 with contrast levels {6.25, 12.5, 25, 50, 100}), and we varied systematically the ρ_N, ρ_D parameters between $[-0.5, 0.5]$ with step size of 0.05 (see Methods 4.4). We fit the pairwise RoG to these synthetic pairs and compared the estimated ρ_D value with the ground truth ρ_D value by binning by the magnitude of the noise correlations and computing the mean squared error within that bin. Shading of grays indicates different constraints on the ground truth ρ_N values. As a reference point (red line), we consider the mean squared error between the ground truth and uniformly random ρ_D values in $[-1, 1]$. We used the minimum possible mean squared error (1/3) as a conservative cut-off point.

To further understand which factors determine the quality of the estimates, we studied how they depended on the noise correlations for a simulated pair. We reasoned that there would be improved estimate quality for the ρ parameters for pairs with larger noise correlations because this would provide more of a constraint to the optimization space. We found that

this intuition held for ρ_N, but for ρ_D the quality of the fit seemed to degrade on average with increasing noise correlations. We hypothesized that this effect was driven by cases in which large values of noise correlations are driven by large values of ρ_N: in those instances, the optimization procedure correctly estimates ρ_N and thus captures noise correlations, but the data do not constrain ρ_D sufficiently. We confirmed this by constraining the magnitude of the ground-truth ρ_N parameter: whenever the magnitude of the noise correlations exceeded the constraint on ρ_N (i.e. noise correlations were in part driven by correlated fluctuations of $\boldsymbol{D}$), the quality of the ρ_D estimator improved dramatically (Fig 4).

This analysis indicates that, while the ρ_N parameter can generally be recovered from the model, the ρ_D parameter can only be inferred when the noise correlations between the pair exceed the magnitude of the (inferred) ρ_N value. This result provides important practical guidance for interpreting model fits to experimental data.

2.4 Shared Normalization Reduces Noise Correlations Between V1 Neurons

Having determined the conditions for identifiability of the correlation parameters ρ_D, we asked how often they are met in our dataset (Fig 5a). We excluded pairs for which either of the ρ values were ± 1 (4712/5597 pairs), because that denotes cases where the data likely did not constrain the pairwise correlation parameters well. Of the remaining pairs (N=885), we found that in a minority (110 pairs) the best fit ρ_N was smaller than median noise correlations, and therefore the estimate of ρ_D was expected to be reliable. Because our analysis focused on neurons that were strongly driven by the stimuli (oriented gratings with fixed orientation), the pairs we considered had similar orientation tuning. Therefore, we expected that they were part of the same normalization circuit. Consistent with this, the estimates of ρ_D were skewed towards positive values (Fig 5b red; median $\rho_D = 0.37$, 95% c.i.[0.28, 0.47]), denoting that normalization signals were often shared. As a control, we

verified that the distribution was symmetric for pairs with unreliable estimates of ρ_D (Fig 5b blue).

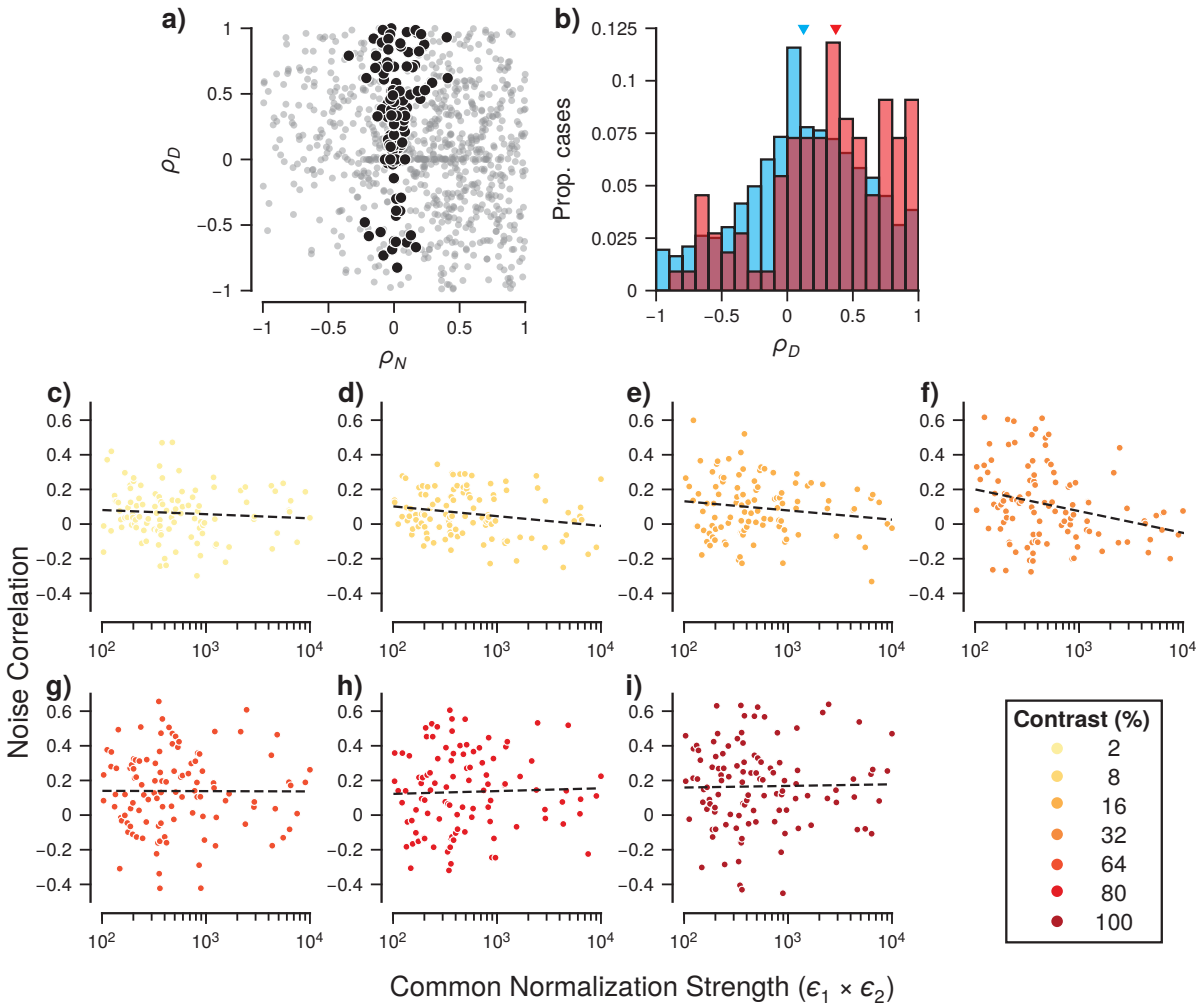


Figure 5: Shared Normalization Signals Reduce Noise Correlations
(a) Scatter of model fit ρ_N vs. ρ_D. Pairs in which the inference of ρ_D is not reliable (i.e. when $|\rho_N|$ exceeds the magnitude of noise correlations, N=775) are shown in gray, while points where the inference of ρ_D is accurate (N=110) are in black. Remaining pairs with poorly constrained ρ_N, ρ_D values not shown. (b) shows the marginal histograms of the fit ρ_D values from (a). The blue indicates all ρ_D values in (a) (median = 0.12 indicated by triangle), while the red indicates those ρ_D values whose fit is accurate (median = 0.37). (c-i) scatter plots of noise correlations as a function of the product of the ϵ parameters (see Methods 4.2) for details), which is the contrast independent component of normalization strength, for neural pairs with accurately estimated ρ_D values. The dotted black line is a linear regression of the noise correlations onto $\log_{10}(\epsilon_1\epsilon_2)$.

We then studied how contrast and normalization strength modulated correlated variability in the pairs with well identified ρ_D. We found that noise correlations increased systematically

with contrast (median Spearman rank correlation = 0.37, 95% c.i. [0.20,0.48]). Furthermore, when we sorted the pairs by the average normalization strength of the two neurons, we observed smaller noise correlations for pairs with stronger normalization (Fig 5c-i). This effect was more pronounced at low contrast (Fig 5c-f; p-values [0.4137, 0.0289, 0.1066, 0.0029]) than at high contrast (Fig 5g-i; p-values [0.9716, 0.7389,0.9449]), as expected from saturation effects of contrast normalization. These findings are consistent with our analysis of Fig. 3, suggesting that stronger shared normalization reduces noise correlations by scaling down a source of shared variability.

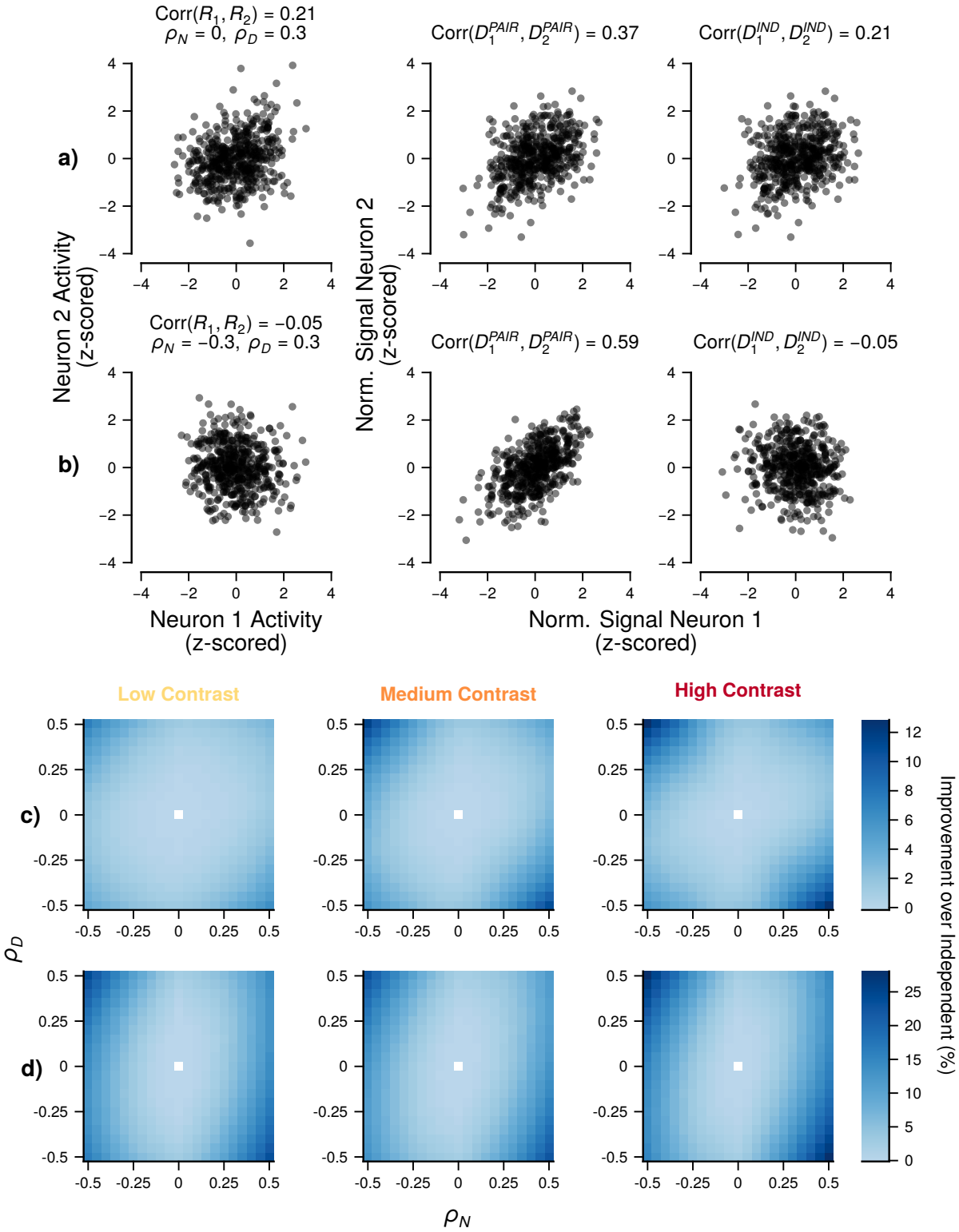
In summary, our result provides the first direct quantification that normalization signals are shared between neurons with similar tuning in V1, as predicted in past work [2019], and an interpretation for why normalization reduces noise correlations between them.

2.5 Pairwise Model Improves Single-Trial Inference of Normalization Strength Even When Noise Correlations are Small

In past work that connected normalization to modulation of noise correlations, stimulus and experimental manipulations (e.g. contrast, attention) are used as proxies for normalization strength [Verhoef and Maunsell, 2017; Tripp, 2012; Ruff and Cohen, 2016] because normalization strength cannot be measured directly. However, these manipulations also affect other factors that drive neural responses, which confounds these as measures of normalization signal. Therefore, quantitatively testing the relationship between noise correlations and normalization requires estimating the single-trial normalization strength for a pair of neurons. One of the advantages of our probabilistic, generative formulation of the pairwise RoG model (Eq. 3) is that it allows us to infer the single-trial normalization strength from measured neural activity (see Methods 4.3). The independent RoG model also provides an estimate for the single-trial normalization, which is known to be a valid estimator for the ground truth normalization strength for data generated from the independent RoG [Coen-

Cagli and Solomon, 2019]. We found similar results for the pairwise RoG (not shown), so we asked how the pairwise estimates for the single-trial normalization strength compares to estimates based on the independent model.

One possibility is that, similar to our results on fit quality, that the estimate derived from the pairwise model would outperform the independent model as noise correlations get larger. However, this is not necessarily the case. Figure 6a,b demonstrates this with two example synthetic neuron pairs. Because the single-trial normalization inference depends on the single-trial neural activity, correlations between neurons will induce correlations between the inferred normalization signals (Fig 6a). However, when noise correlations are small and arise from cancellation between ρ_N and ρ_D, the independent model will infer minimal correlation between normalization signals while the pairwise model will correctly infer that the single-trial normalization is correlated (Fig 6b).



(Caption on next page.)

Figure 6: Inference of Single-Trial Normalization Depends on ρ Parameters and Contrast Level
(a,b) Two simulated experiments drawn from the pairwise RoG with different noise correlations arising from different underlying ρ_N, ρ_D values to compare the pairwise and independent estimators of single-trial normalization with the ground truth normalization signal. (a) has overall noise correlation of 0.21 across contrasts generated with $\rho_N = 0, \rho_D = 0.3$, (b) has overall noise correlation of -0.05 across contrasts generated with $\rho_N = -0.3, \rho_D = 0.3$. Z-scoring performed across trials. Random parameters drawn from $R^{max} \in [10, 100], \boldsymbol{\epsilon} \in [15, 25], \boldsymbol{\alpha_N}, \boldsymbol{\alpha_D} \in [0.1, 1], \boldsymbol{\beta_N}, \boldsymbol{\beta_D} \in [1, 1.5], \boldsymbol{\eta} \equiv 0$. Contrasts levels were {6.25, 12.5, 25, 50, 100}
(c,d) Comparison of the single-trial normalization inference in the pairwise and independent RoG models for two goodness of fit measures, the mean-squared error between the estimate and the true value (c) and the correlation between the estimate and the true value (d), as it depends on ρ_N, ρ_D and the contrast levels. On the left is the lowest contrast level (6.25), in the middle is an intermediate contrast level (25), and at the right displays maximum contrast (100). Each bin corresponds to the median difference between the pairwise and independent models across simulated experiments. (c,d) use 11628 synthetic pairs generated as in Fig. 4.

To demonstrate this principle, we computed the difference of both the mean squared error (Fig 6c) and correlation (Fig 6d) between the pairwise RoG and independent RoG normalization inference (both measures are relative to the ground truth values) for a large number of simulated pairs (11628) with systematically varying ρ_N, ρ_D values. First, we found that the distinction between the two models depended on the contrast level: at the lowest contrast level, there was no clear difference for either statistic between the pairwise and independent models. As the contrast increased, the quality of the estimators appeared to diverge. Second, for both the mean squared error and correlation between the estimators and the ground truth, the improvement of the pairwise RoG estimate over the independent estimate increased when the magnitude of the ρ_D parameter increased. Consistent with the intuition provided by Fig 6b, the largest improvement occurred when the ρ_N parameter had a large value that was the opposite sign of ρ_D. The dependence on the ρ_D parameter reflects that the estimator for the pairwise model incorporates knowledge about correlation between the normalization signals.

In summary, this analysis shows that the pairwise model produces accurate estimates of the single-trial normalization strength of neural pairs and outlines the conditions in which those estimates are preferable to those obtained with the independent model.

3 Discussion

We introduced a stochastic model of divisive normalization, the pairwise RoG, to characterize the trial-to-trial covariability between cortical neurons (i.e. noise correlations). The model provides excellent fits to Ca^{2+} imaging recordings from mouse V1, capturing diverse effects of stimulus contrast and normalization strength on noise correlations (Figs 2,5). We demonstrated that the effect of normalization on noise correlations differs depending on the sources of the variability, and that the model can accommodate both increases and decreases in noise correlations with normalization (Fig 3) as past experiments had suggested. We then outlined the regimes of identifiability of a key model parameter, which determines how strongly normalization is shared between neurons (Fig 4). This allowed us to obtain the first quantitative evidence that normalization signals are often shared between similarly tuned neurons in mouse V1, also explaining why, for those neuronal pairs, noise correlations were generally weaker when normalization was stronger (Fig 5). Lastly, we derived a Bayesian estimator for the single-trial normalization signals of simultaneously recorded pairs. Surprisingly, this estimator can be more accurate than the estimator based on the model that ignores noise correlations (the independent RoG) even when noise correlations are negligible (Fig 6).

As a descriptive, data analytic tool, our modeling framework complements normative and mechanistic theories of neuronal population variability. For instance, normative probabilistic accounts of sensory processing have suggested that divisive normalization may play a role in the inference of perceptual variables by modulating neural representations of uncertainty [Beck, Latham, and Pouget, 2011; Dehaene, Coen-Cagli, and Pouget, 2021; Hénaff et al., 2020; Bányai et al., 2019; Coen-Cagli, Kohn, and Schwartz, 2015; Festa et al., 2021; Orbán and Bányai, 2019]. Similarly, normalization could play a key role in multisensory cue combination [Hayashi, Kato, and Nozaki, 2020; Ohshiro, Angelaki, and DeAngelis, 2017; 2011]. However, the posited effect of normalization on covariability has not been tested

quantitatively, as normalization signals are not generally measurable. The pairwise RoG will allow researchers to test these hypotheses by providing a means with which to estimate normalization signals from neural data and relate these to measures of neural covariability. In circuit-based models of neural dynamics such as the stabilized supralinear network [Hennequin et al., 2018] and the ORGaNICs architecture [Heeger and Mackey, 2019], the normalization computation emerges naturally from the network dynamics [Heeger and Zemlianova, 2020] and shapes the structure of stimulus-dependent noise correlations [Echeveste et al., 2019]. By quantifying the parametric relation between normalization and covariability, our descriptive tool will enable mapping those parameters onto the different circuit motifs and cell types posited by these network models.

There are numerous descriptive models of correlated neural population activity, among the most popular of these being latent variable models [Sokoloski, Aschner, and Coen-Cagli, 2021; Lin et al., 2015; Goris, Movshon, and Simoncelli, 2014; Whiteway, Averbeck, and Butts, 2020; Whiteway et al., 2019; Archer et al., 2014; Yu et al., 2008; Ecker et al., 2014], in which correlations arise from the mapping of a small number of unobserved variables onto the high dimensional neural response space. This approach effectively partitions noise correlations into underlying causes (i.e. latents) that are responsible for coordinating neural responses, which resembles our attribution of noise correlations to either shared input drive or shared normalization pools. However, because those latent variable approaches do not directly incorporate divisive normalization, they cannot be used to test normative and mechanistic predictions about normalization. An exception to this is the model of Hénaff et. al. that explicitly uses normalization to connect gain variability and perceptual uncertainty [Hénaff et al., 2020]. In this model, which builds upon their previous work [Goris, Movshon, and Simoncelli, 2014], spike counts are modeled as a Poisson process in which the rate parameter is given by the normalization equation (Eq. 2) and the denominator is corrupted by Gaussian noise. They then apply the normal approximation (Eq. 4) to derive an equation for how the gain variability is modulated by the normalization signal. Their model primarily

focuses on individual neural variability and does not explicitly consider pairwise response statistics as we do here.

Three models have directly studied the relationship between normalization and across trial covariability [Ruff and Cohen, 2017; Tripp, 2012; Verhoef and Maunsell, 2017]. Tripp's [Tripp, 2012] simulation work on velocity tuning in the medial temporal cortex (MT) consistently predicted that normalization would decorrelate neural responses. However, we found that noise correlations could also increase with normalization. This is because Tripp modeled correlations to solely arise from tuning similarity between neurons. Conversely, in the RoG framework, noise correlations originate from input correlations ρ_N and correlations between normalization signals ρ_D. Our model then offers more flexibility than Tripp's by allowing relationships between normalization and correlation to depend on the sources of correlations. Verhoef and Maunsell [Verhoef and Maunsell, 2017] investigated the effect of attention on noise correlations by using a recurrent network implementation of the normalization model of attention [Reynolds and Heeger, 2009]. They describe multiple different patterns of the effect of normalization on noise correlations depending on tuning similarity between a pair of neurons and where attention is directed. Our model does not currently account for the effect of attention, but this would be possible by adapting the standard normalization model of attention which would require an additional parameter for the attentional gain. These prior two models are also primarily simulation based, while our model is meant to be data analytic. Lastly, Ruff and Cohen [Ruff and Cohen, 2017] proposed a normalization model to explain how attention increases correlations between V1 and MT neurons [2016]. They modeled the trial-averaged MT neural responses as a function of trial-averaged responses of pools of V1 neurons. After fitting the parameters, single-trial MT responses were predicted by feeding the pooled single-trial V1 responses into the equation. By construction, variability in predicted MT neural responses only arises from variability in the V1 neural responses, which only occur in the numerator of their normalization model. Our model also allows for variability in the denominator of the normalization equation and therefore their model can

be seen as a special case of the pairwise RoG.

An important limitation of our model is that the correlation parameter (ρ_D) is often not identifiable, and in the dataset we analyzed we found a large majority of pairs with best fit (ρ_N, ρ_D) parameters at the ends of their ranges, suggesting they are not well constrained by the data. These issues could be ameliorated by using alternative optimization algorithms, regularizing the model parameters, or optimizing the stimulus design in experiments. Nonetheless, we showed that we were accurately able to recover the correlation between the normalization signals when the median noise correlation exceeded the estimated input correlation (ρ_N) for a pair. In the dataset analyzed, we found that 110/5597 pairs meet this criterion. Although this is a minority of cases, it demonstrates that typical datasets with existing recording technologies could nonetheless provide sufficient power for studies that focus on the ρ parameter values. It will be important in future work to understand which experimental conditions would maximize the yield of pairs with identifiable ρ parameters.

The generality of the modeling framework presented here leaves room for future expansion. One such direction would be to increase the dimensionality to model correlations among a neural population. This would require more correlation parameters, which could make the model more difficult to fit to data. However, reasoning that population variability is low-dimensional [Cunningham and Yu, 2014; Huang et al., 2019; Lin et al., 2015; Rabinowitz et al., 2015; Schölvinck et al., 2015; Umakantha et al., 2021], it is likely this issue could be circumvented by applying dimensionality reduction techniques within the model or by allowing the sharing of correlation parameters across a neural population. Another interesting application of this model would look directly at the effects of normalization on information transmission and representation. The relationship between noise correlations and the amount of information that can be represented by a neural population has been widely discussed [Abbott and Dayan, 1999; Averbeck and Lee, 2006; Averbeck, Latham, and Pouget, 2006; Kohn et al., 2016; Moreno-Bote et al., 2014; Zohary, Shadlen, and

Newsome, 1994]. Moreover, some experimental and theoretical work has connected modulations of information in neural populations with computations that have been modeled with normalization models, such as surround suppression and attention [Henry and Kohn, 2020; Kanashiro et al., 2017; Ringach, 2010; Tripp, 2012]. Our model could be modified to investigate this connection and further illuminate the effects of normalization on information transmission.

4 Methods

4.1 Generative Model – Pairwise Ratio of Gaussians (RoG)

Here we describe in detail the RoG model and derive the Gaussian approximation to the ratio variable. Note that our RoG is entirely different from that used in [Pillow and Simoncelli, 2006], despite the same acronym: [2006] refers to the distribution obtained from the ratio between two Gaussian distributions, whereas we refer to the random variable that results from the ratio of two Gaussian variables.

We build from the standard normalization model [Carandini and Heeger, 2012; Sawada and Petrov, 2017] which computes the across-trial average neural response (e.g. firing rate) as the ratio between a driving input to a neuron (N) and the summed input from a pool of nearby neurons (D):

$$R = f_{\div}(N, D) = \frac{N}{D} \quad \text{where } N \geq 0, D > 0 \tag{2}$$

Where $f_{\div}$ is the division operator; this functional notation is convenient for later derivations in which we consider the derivative of division. Our goal is to model the joint activity of pairs of neurons, so we extend the normalization model by considering two model neurons R_1, R_2. Since we are interested in trial-to-trial variability, we assume that a pair of neural responses $\boldsymbol{R}_t = (R_1, R_2)_t$ on a single trial can be written as the element-wise ratio of two Gaussian

random vectors, $\boldsymbol{N}_t = (N_1, N_2)_t$ and $\boldsymbol{D}_t = (D_1, D_2)_t$, with additive Gaussian random noise $\boldsymbol{\eta}_t = (\eta_1, \eta_2)_t$ to capture the residual (i.e. stimulus-independent) variability.

As detailed further below, the numerators of two neurons can be correlated, and similarly for the denominators. Moreover, in principle there can be correlations between the numerators and denominators (e.g. (N_1, D_2) may be correlated), requiring us to consider the joint, four-dimensional Gaussian distribution for the vector $(\boldsymbol{N}_t, \boldsymbol{D}_t)$. However, in this paper we will consider the simpler model in which $\boldsymbol{N}_t$ and $\boldsymbol{D}_t$ are independent and are each distributed according to their respective two-dimensional Gaussian distributions. This assumption allows for simplified mathematical derivations, and is supported by our previous work which found that including a parameter for the correlation between N and D caused over-fitting to single-neuron data [Coen-Cagli and Solomon, 2019].

We therefore write the generative model for the pairwise RoG as:

$$\begin{aligned} \boldsymbol{N}_t &\sim \mathcal{N}(\boldsymbol{\mu_N}, \boldsymbol{\Sigma_N}) \\ \boldsymbol{D}_t &\sim \mathcal{N}(\boldsymbol{\mu_D}, \boldsymbol{\Sigma_D}) \\ \boldsymbol{\eta}_t &\sim \mathcal{N}(\boldsymbol{\mu_\eta}, \boldsymbol{\Sigma_\eta}) \end{aligned} \tag{3}$$

$$\boldsymbol{R}_t = (R_1, R_2)_t = f_{\div}(\boldsymbol{N}_t, \boldsymbol{D}_t) + \boldsymbol{\eta}_t = \frac{\boldsymbol{N}_t}{\boldsymbol{D}_t} + \boldsymbol{\eta}_t$$

Where $f_{\div}$ is applied elementwise, $\boldsymbol{\mu_N}, \boldsymbol{\mu_D}$ are the 2D vectors of means of the numerator and denominator, respectively, $\boldsymbol{\Sigma_N}, \boldsymbol{\Sigma_D}$ are the respective 2×2 covariance matrices, and $(\boldsymbol{\mu_\eta}, \boldsymbol{\Sigma_\eta})$ is the mean and covariance matrix for the residual component of the model.

For the independent RoG, we used the result that when the denominator has negligible probability mass at values less than or equal to zero, the ratio distribution can be approximated by a Gaussian distribution with mean and variance that can be derived from a Taylor expansion [Díaz-Francés and Rubio, 2013; Hayya, Armstrong, and Gressis, 1975; Marsaglia

et al., 2006; Pham-Gia, Turkkan, and Marchand, 2006]. For our application, this assumption is justified since the denominator is the sum of the non-negative responses from a pool of neurons, and is therefore unlikely to attain values less than or equal to zero. A similar result applies for our pairwise extension, in which we can use the multivariate delta method (an application of a Taylor expansion) to compute the mean and covariance for the joint distribution of ratio variables [Ver Hoef, 2012] under the assumption that $\boldsymbol{\mu_D} > 0$.

To derive equations for the mean and covariance of the pairwise RoG, we use a Taylor expansion around the point $(\boldsymbol{\mu_N}, \boldsymbol{\mu_D})$:

$$\begin{aligned} \boldsymbol{R} &= f_{\div}(\boldsymbol{N}, \boldsymbol{D}) + \boldsymbol{\eta} \\ &= f_{\div}(\boldsymbol{\mu_N}, \boldsymbol{\mu_D}) + \nabla f_{\div}|_{\boldsymbol{\mu_N}, \boldsymbol{\mu_D}} \left(\begin{pmatrix} \boldsymbol{N} \\ \boldsymbol{D} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu_N} \\ \boldsymbol{\mu_D} \end{pmatrix} \right) + \text{higher order terms} + \boldsymbol{\eta} \end{aligned} \tag{4}$$

Using only the first order terms, we can derive expressions for the mean and covariance matrix of the RoG:

$$\boldsymbol{\mu_R} = \mathsf{E}[\boldsymbol{R}] \approx \frac{\boldsymbol{\mu_N}}{\boldsymbol{\mu_D}} + \boldsymbol{\mu_\eta} \tag{5}$$

$$\begin{aligned} \boldsymbol{\Sigma_R} &= \mathsf{E}[(\boldsymbol{R} - \boldsymbol{\mu_R})(\boldsymbol{R} - \boldsymbol{\mu_R})^\top] \\ &\approx \nabla f_{\div}|_{\boldsymbol{\mu_N}, \boldsymbol{\mu_D}} \mathsf{E}\left[\left(\begin{pmatrix} \boldsymbol{N} \\ \boldsymbol{D} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu_N} \\ \boldsymbol{\mu_D} \end{pmatrix} \right) \left(\begin{pmatrix} \boldsymbol{N} \\ \boldsymbol{D} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu_N} \\ \boldsymbol{\mu_D} \end{pmatrix} \right)^\top \right] \nabla f_{\div}|^\top_{\boldsymbol{\mu_N}, \boldsymbol{\mu_D}} + \boldsymbol{\Sigma_\eta} \\ &\approx \nabla f_{\div}|_{\boldsymbol{\mu_N}, \boldsymbol{\mu_D}} \begin{pmatrix} \Sigma_N & 0 \\ 0 & \Sigma_D \end{pmatrix} \nabla f_{\div}|^\top_{\boldsymbol{\mu_N}, \boldsymbol{\mu_D}} + \boldsymbol{\Sigma_\eta} \end{aligned} \tag{6}$$

From 6, we can derive expressions for the variance of each neuron in the pair and their covariance and correlation. First, we adopt the following notation to simplify the equations:

let $\widehat{\sigma_{\eta_i}} = \frac{\mu_{D_i}}{\mu_{N_i}}\sigma_{\eta_i}$ and let $\delta_{X_i} = \frac{\sigma_{X_i}}{\mu_{X_i}}$. Then:

$$\sigma^2_{R_i} \approx \frac{\mu^2_{N_i}}{\mu^2_{D_i}}\left(\delta^2_{N_i} + \delta^2_{D_i} + \widehat{\sigma_{\eta_i}}^2\right) \quad \text{for } i = 1, 2$$

$$\mathrm{Cov}\left(R_1, R_2\right) \approx \frac{\mu_{N_1}\mu_{N_2}}{\mu_{D_1}\mu_{D_2}}\left(\rho_N\delta_{N_1}\delta_{N_2} + \rho_D\delta_{D_1}\delta_{D_2} + \rho_\eta\widehat{\sigma_{\eta_1}}\widehat{\sigma_{\eta_2}}\right) \tag{7}$$

$$\mathrm{Corr}(R_1, R_2) \approx \frac{\rho_N\delta_{N_1}\delta_{N_2} + \rho_D\delta_{D_1}\delta_{D_2} + \rho_\eta\widehat{\sigma_{\eta_1}}\widehat{\sigma_{\eta_2}}}{\sqrt{\delta^2_{N_1} + \delta^2_{D_1} + \widehat{\sigma_{\eta_1}}^2}\sqrt{\delta^2_{N_2} + \delta^2_{D_2} + \widehat{\sigma_{\eta_2}}^2}} \tag{8}$$

Equation 8 generalizes the formula for the spurious correlation of ratios found in population genetics [Pearson, 1897].

4.2 Parametrization of the Pairwise RoG for Contrast Responses

In the form described so far, the pairwise RoG has 10 stimulus dependent parameters $\left(\mu_{N_1}, \mu_{N_2}, \mu_{D_1}, \mu_{D_2}, \sigma_{N_1}, \sigma_{N_2}, \sigma_{D_1}, \sigma_{D_2}, \rho_N, \rho_D\right)$ and 5 stimulus independent parameters $\left(\mu_{\eta_1}, \mu_{\eta_2}, \sigma_{\eta_1}, \sigma_{\eta_2}\right.$ for the additive noise. For any stimulus condition, there are only 5 relevant measurements that can derived from the neural data (the response means and variances for each neuron in a pair, and their correlation), so the model is over-parametrized. Therefore, to apply the RoG to neural data, we need to reduce the total number of parameters.

The generality of this model provides a procedure for converting a standard normalization model (i.e. a model for the mean response) into a RoG model that specifies both mean and (co)-variance. In this paper, we use the example of contrast-gain control, which has been widely used to model the response of single neurons and neural populations to visual input with varying contrast [Albrecht and Hamilton, 1982; Clatworthy et al., 2003; Geisler and Albrecht, 1992; Heeger, 1992]. By adapting such a model, we can reduce the stimulus dependence of the means of the numerator and denominator $(\mu_{N_1}, \mu_{N_2}, \mu_{D_1}, \mu_{D_2})$. In the contrast-gain control model, the neural response as a function of contrast c $(0 - 100\%)$ is

computed as a "hyperbolic ratio" [Albrecht and Hamilton, 1982; Heeger, 1992]:

$$R(c) = \frac{R^{max} c^2}{\epsilon^2 + c^2} + R_0 \tag{9}$$

Where R^{max} is the maximum response rate, ϵ is the semi-saturation constant (the contrast at which $R(\epsilon) = R^{max}/2$) to prevent division by 0, and R_0 is the spontaneous activity of the neuron (the response at 0% contrast). We can convert this standard model into an RoG by setting the mean of the numerator and denominator in the RoG to the numerator and denominator in this equation:

$$\mu_N = R^{max} c^2, \;\; \mu_D = \epsilon^2 + c^2 \;\; , \mu_\eta = R_0 \tag{10}$$

By using this functional form, we can replace the stimulus dependent parameters of the RoG $(\mu_{N_1}, \mu_{N_2}, \mu_{D_1}, \mu_{D_2})$ with the stimulus independent parameters $R_1^{max}, R_2^{max}, \epsilon_1, \epsilon_2$. Another model simplification is to assume that individual neural variability and mean neural response are related by a power function as has been observed in the visual cortex [Carandini, Heeger, and Movshon, 1997; Goris, Movshon, and Simoncelli, 2014; Gur, Beylin, and Snodderly, 1997]:

$$\sigma^2 = \alpha \cdot \mu^\beta \tag{11}$$

This parametrization allows the Fano Factor (the ratio of the variance to the mean) to vary with stimulus input (as long as $\beta \neq 1$) and for both over-dispersion (Fano factor > 1) and under-dispersion (Fano factor < 1). Moreover, as with the mean, the 4 stimulus-dependent variance parameters of the model $(\sigma_{N_1}, \sigma_{N_2}, \sigma_{D_1}, \sigma_{D_2})$ can be replaced with four pairs of stimulus independent parameters $(\boldsymbol{\alpha_N}, \boldsymbol{\beta_N}, \boldsymbol{\alpha_D}, \boldsymbol{\beta_D})$. Lastly, in principle, the parameters controlling correlation (ρ_N, ρ_D) can vary with stimulus conditions but for computational simplicity we assume that (ρ_N, ρ_D) are stimulus independent. However, even with this

assumption, our model can capture stimulus dependent noise correlations (Fig 2) as have been observed *in vivo* [Kohn and Smith, 2005; Ponce-Alvarez et al., 2013; Sadagopan and Ferster, 2012].

Fitting the RoG to Data

We optimize the values of the parameters, given a dataset, by maximum likelihood estimation. In this paper, we validate various properties of the pairwise RoG using synthetic data produced from the generative model. We will demonstrate the applicability of this model to neural data analysis by fitting the pairwise RoG to Calcium imaging data (see below).

Based on our previous discussion, we have assumed that the model parameters are stimulus independent which we will denote as Θ. We consider our dataset $\mathcal{D} = \{\boldsymbol{R}_t(s)\}$ where s is the stimulus and t indexes the trial. We assume that, for each stimulus, our data is independent and identically distributed according to $\mathcal{N}\left(\boldsymbol{\mu_R}(s), \boldsymbol{\Sigma_R}(s); \Theta\right)$, and that data is independent across stimuli. We can therefore compute the negative log-likelihood of the data using the following equation (see Appendix 5.2 for derivation):

$$\begin{aligned} -\log p(\boldsymbol{R}_t(s);\Theta) \approx \frac{T}{2}\sum_{s\in S}\Big\{ &\log|\boldsymbol{\Sigma_R}(s,\Theta)| + \mathrm{Tr}(\boldsymbol{\Sigma_R}^{-1}(s,\Theta)\widehat{\boldsymbol{\Sigma_R}}(s)) \\ &+ (\widehat{\boldsymbol{\mu_R}}(s) - \boldsymbol{\mu_R}(s,\Theta))^\top \boldsymbol{\Sigma_R}^{-1}(s,\Theta)(\widehat{\boldsymbol{\mu_R}}(s) - \boldsymbol{\mu_R}(s,\Theta))\Big\} \end{aligned} \tag{12}$$

Where $\widehat{\boldsymbol{\mu_R}}(s)$ and $\widehat{\boldsymbol{\Sigma_R}}(s)$ are the empirical mean and covariance across trials computed from the data.

In practice, we have found that it is computationally faster to first optimize the parameters for each neuron in the pair separately (which is equivalent to fitting the independent RoG model), and then optimize the correlation parameters (i.e. the ρ parameters) with the single-neuron model parameters fixed. This two-step optimization process is referred to as the inference functions for marginals method in the copula literature, and is known to be

mathematically equivalent to maximum likelihood estimation for Gaussian copulas [Joe and Xu, 1996], which is the case we consider here. This points to an extension or alternative to the pairwise RoG presented here that considers the bivariate distribution to be some non-Gaussian copula with Gaussian marginals. We assumed that the pairwise distribution is Gaussian for computational simplicity, but others have used non-Gaussian copulas to model neural populations [Berkes, Wood, and Pillow, 2008].

4.3 Inference of Single Trial Normalization from Measured Neural Activity

Because of the probabilistic formulation of the RoG, we can use Bayes theorem to compute the posterior probability of the normalization variable $\boldsymbol{D}_t$ given the observed spike counts $\boldsymbol{R}_t$:

$$\begin{aligned} p(\boldsymbol{D}_t|\boldsymbol{R}_t) &\propto p(\boldsymbol{R}_t|\boldsymbol{D}_t)p(\boldsymbol{D}_t) = p((R_1, R_2), (D_1, D_2)) \\ &\propto p((N_1, N_2), (D_1, D_2))|D_1 D_2| = p(N_1, N_2)p(D_1, D_2)|D_1 D_2| \end{aligned} \tag{13}$$

Where multiplication by $|D_1 D_2|$ occurs due to the change of variables formula for probability density functions. From this distribution, we can find the maximum a posteriori (MAP) estimate of the normalization strength in a single trial by differentiating the posterior distribution with respect the denominator variables and finding the maxima by setting the partial derivatives to 0. For ease of computation, we solve the equivalent problem of finding the minima of the partial derivatives of the negative logarithm of the posterior distribution. Additionally, we assume that there is zero additive noise. In our previous work [Coen-Cagli and Solomon, 2019], we found that, when subtracting the mean additive noise from the simulated spike counts, the MAP estimate remained unbiased. Thus, for simplicity, we assume that we can subtract off the mean spontaneous activity and consider instead the posterior $p\left(\boldsymbol{D}|\boldsymbol{R} - \boldsymbol{\mu}_{\boldsymbol{\eta}}\right)$.

To obtain an estimate for the denominator strength $\boldsymbol{D}_t$ we look at the partial derivatives of the negative log posterior with respect to D_1, D_2 and solve to obtain the MAP estimate. This procedure leads to a two dimensional system of bivariate quadratic equations:

$$\begin{aligned} 2A_1{D_1}^2 + B_1D_1 + CD_1D_2 - 2 = 0 \\ 2A_2{D_2}^2 + B_2D_2 + CD_1D_2 - 2 = 0 \end{aligned} \tag{14}$$

Where the coefficients A_1, A_2, B_1, B_2, C are functions of the parameters of the model (see Appendix S2 for the derivation of (14) and the full expressions of these coefficients).

A basic result from the algebra of polynomial systems (Bézout's theorem) tells us that this system has four pairs of possibly complex valued solutions [Sturmfels, 2002]. In fact, as solving this system amounts to solving a quartic equation in one variable, there exists an algebraic formula for solutions to this system as a function of the coefficients.

Because all of the variables involved are real valued, we are only interested in the existence of real solutions to this two dimensional system. However, there is no theoretical guarantee that there will be any real solutions. In practice we take the real part of the algebraic solution to this system and find which pair of solutions $\left(D_1^{MAP}, D_2^{MAP}\right)$ minimize the negative log posterior. Alternatively, we can also consider finding the MAP by minimizing (Eq. 23), but this is far more time consuming as it would require performing this numerical optimization for each trial.

4.4 Generating Realistic Pairwise Neural Activity from the Model

To constrain our simulations to realistic parameter values for the contrast response function (Eq. 9), we took the single-neuron best fit parameters to macaque V1 data analyzed in our previous work and created parameter pairs by considering all combinations ($N = 11628$ pairs) of these parameters (for details see [Coen-Cagli and Solomon, 2019]). Using the generative model for the pairwise RoG (Eq. 3) and the contrast response parametrization

(Eq. 9), we can simulate single-trial neural activity from these parameter pairs and specific values for (ρ_N, ρ_D).

These synthetic data allow direct comparison of the performance of the pairwise and independent versions of the RoG (see Figure S1). As an assessment of relative model performance, we use the percent difference between the negative log-likelihoods of the pairwise and independent models (relative to the independent model), such that a positive percent difference corresponds to a lower negative log-likelihood for the pairwise model, indicating improved performance over the independent model.

4.5 Data Collection and Processing

Animal Preparation

Data were collected from CaMKII-tTA;tetO-GCaMP6s mice [Wekselblatt et al., 2016], expressing GCaMP6s in cortical excitatory neurons. Mice were implanted with headplates and cranial windows over V1 [Sridharan et al., 2022]. Briefly, mice were anesthetized with 2% isoflurane and administered 2 mg/kg of dexamethasone and 0.5 mg/kg of buprenorphine. Animals were secured in a stereotaxic frame (Kopf) and warmed with a heating pad. The scalp was removed and the skull was lightly etched. A craniotomy was made over V1 using a 3.5 mm skin biopsy bunch. A cranial window, consisting of two 3 mm diameter circular coverslips glued to a 5 mm diameter circular coverslip, was placed onto the craniotomy, and secured into place with Metabond (C&B). Then a custom made titanium headplate was secured via Metabond (C&B) and the animals were allowed to recover in a heated cage.

Behavioral Task

During imaging, mice were head-fixed in a tube, and were performing an operant visual detection task [Bounds et al., 2021]. Briefly, mice were trained to withhold licking when no stimulus was present, and lick within a response window after stimulus presentation. Mice

were water-restricted and given a water reward for correct detection. Visual stimuli were drifting sinusoidal gratings (2 Hz, 0.08 cycles/degree) presented for 500 ms followed by a 1000 ms response window. Stimuli were generated and presented using PsychoPy2 [Peirce et al., 2019]. Contrast of gratings were varied between 7 different levels: {2, 8, 16, 32, 64, 80, 100}, except for 2 recording sessions in which contrast level 80% was omitted. This did not alter any of the analysis, allowing sessions to be combined into a single dataset.

Visual Stimuli and Calcium Imaging

Visual stimuli were presented using a gamma corrected LCD monitor (Podofo, 25 cm, 1024x600 pixels, 60 Hz refresh rate) located 10 cm from the right eye.

Once they learned the task, mice started performing under the 2p microscope, and V1 was imaged via cranial window. Imaging was performed using a 2-photon microscope (Sutter MOM, Sutter inc.), with a 20X magnification (1.0 NA) water-immersion objective (Olympus Corporation). Recordings were done in L2/3 in a 800 x 800 µm field of view, with 75-100 mW of 920 nm laser light (Chameleon; Coherent inc). An electrically-tunable lens (Optotune) was used to acquire 3 plane volumetric images at 6.36 Hz. Planes were 30 µm apart. Acquisition was controlled with ScanImage (Vidrio Technologies).

Calcium imaging data was motion-corrected and ROI extracted using suite2p [Pachitariu et al., 2017], and all data was neuropil subtracted with a coefficient of 0.7.

Data Processing

Processing of calcium imaging data was performed using custom MATLAB code. Fluorescence traces for individual trials and cells (average of the neuropil subtracted fluorescence across a ROI) consisted of 24 frames: 4 frames of pre-stimulus blank, followed by 3 frames of stimulus presentation and 17 frames of post-stimulus blanks corresponding to the response window for the behavioral task. Baseline fluorescence (F_0) was computed as the median

across pre-stimulus frames (1-5), and the stimulus evoked fluorescence ($\Delta F/F$) was computed as the mean of the normalized fluorescence per frame ($(F(i) - F_0)/F_0$ for $F(i)$ the fluorescence of frame i) across frames corresponding to stimulus response and an additional frame to account for onset delays and calcium dynamics. Spontaneous $\Delta F/F$ was computed as above during blank trials. Cells were included in further analysis if the evoked response at the highest contrast was at least 2 standard deviations above the spontaneous mean fluorescence. Across 9 recording sessions, 297/8810 neurons met this inclusion criterion. This small percentage is due to sessions recorded using gratings with fixed spatial frequency and orientation; thus, included neurons are visually responsive and selective for this combination.

4.6 Code and Data Availability

Code for running model simulations and fitting model to experimental data was written in MATLAB and will be made publicly available on Github (https://github.com/oren-weiss/pairwiseRatioGaussians). Ca^{2+} imaging data will be made publicly available on Zenodo.

Acknowledgement

We thank members of the Kohn and Coen-Cagli laboratories for feedback on the manuscript. We also thank Daniel Quintana for help with animal training.

This work was supported by the National Institutes of Health (NIH) Grant R01- EY030578 (R.C.C.), NIH Grant UF1- NS107574 (H.A.), NIH Grant U19- NS107613 (H.A.), NIH Grant RF1- RF1MH120680 (H.A.), NIH Grant RO1- EY023756 (H.A.), New York Stem Cell Foundation (H.A.), National Science Foundation Graduate Research Fellowship (DGE 1752814 (H.A.B.)).

References

Abbott, LF and P Dayan (1999). "The effect of correlated variability on the accuracy of a population code." In: *Neural Computation* 11.1, pp. 91–101.

Albrecht, DG and WS Geisler (1991). "Motion selectivity and the contrast-response function of simple cells in the visual cortex." In: *Visual neuroscience* 7.6, pp. 531–546.

Albrecht, DG and DB Hamilton (1982). "Striate Cortex of Monkey and Cat: Contrast Response Function." In: *Journal of neurophysiology* 48.1, pp. 217–237.

Archer, EW et al. (2014). "Low-Dimensional Models of Neural Population Activity in Sensory Cortical Circuits." In: vol. 27. Advances in Neural Information Processing Systems. Curran Associates, Inc.

Averbeck, BB, PE Latham, and A Pouget (2006). "Neural Correlations, Population Coding and Computation." In: *Nature Reviews Neuroscience* 7.5, p. 358.

Averbeck, BB and D Lee (2006). "Effects of Noise Correlations on Information Encoding and Decoding." In: *Journal of Neurophysiology* 95.6, pp. 3633–3644.

Bányai, M et al. (2019). "Stimulus complexity shapes response correlations in primary visual cortex." In: *Proceedings of the National Academy of Sciences* 116.7, pp. 2723–2732.

Bartolo, R et al. (2020). "Information-limiting correlations in large neural populations." In: *Journal of Neuroscience* 40.8, pp. 1668–1678.

Beck, JM, PE Latham, and A Pouget (2011). "Marginalization in Neural Circuits with Divisive Normalization." In: *Journal of Neuroscience* 31.43, pp. 15310–15319.

Berkes, P, F Wood, and J Pillow (2008). "Characterizing neural dependencies with copula models." In: *Advances in neural information processing systems* 21.

Berkes, P et al. (2011). "Spontaneous Cortical Activity Reveals Hallmarks of an Optimal Internal Model of the Environment." In: *Science (New York, N.Y.)* 331.6013, pp. 83–87.

Bondy, AG, RM Haefner, and BG Cumming (2018). "Feedback Determines the Structure of Correlated Variability in Primary Visual Cortex." In: *Nature Neuroscience* 21.4, pp. 598–606.

Bounds, HA et al. (2021). “Multifunctional Cre-dependent transgenic mice for high-precision all-optical interrogation of neural circuits.” In: *bioRxiv*.

Carandini, M, DJ Heeger, and JA Movshon (1997). “Linearity and Normalization in Simple Cells of the Macaque Primary Visual Cortex.” In: *Journal of Neuroscience* 17.21, pp. 8621–8644.

Carandini, M and DJ Heeger (2012). “Normalization as a Canonical Neural Computation.” In: *Nature Reviews Neuroscience* 13.1, pp. 51–62.

Cavanaugh, JR, W Bair, and JA Movshon (2002). “Nature and Interaction of Signals from the Receptive Field Center and Surround in Macaque V1 Neurons.” In: *Journal of Neurophysiology* 88.5, pp. 2530–2546.

Clatworthy, PL et al. (2003). “Coding of the Contrasts in Natural Images by Populations of Neurons in Primary Visual Cortex (V1).” In: *Vision research* 43.18, pp. 1983–2001.

Coen-Cagli, R, P Dayan, and O Schwartz (2012). “Cortical surround interactions and perceptual salience via natural scene statistics.” In: *PLoS computational biology* 8.3, e1002405.

Coen-Cagli, R, A Kohn, and O Schwartz (2015). “Flexible Gating of Contextual Influences in Natural Vision.” In: *Nature Neuroscience* 18.11, pp. 1648–1655.

Coen-Cagli, R and SS Solomon (2019). “Relating Divisive Normalization to Neuronal Response Variability.” In: *Journal of Neuroscience* 39.37, pp. 7344–7356.

Cohen, MR and A Kohn (2011). “Measuring and interpreting neuronal correlations.” In: *Nature neuroscience* 14.7, pp. 811–819.

Cohen, MR and JHR Maunsell (2009). “Attention Improves Performance Primarily by Reducing Interneuronal Correlations.” In: *Nature Neuroscience* 12.12, pp. 1594–1600.

Cunningham, JP and BM Yu (2014). “Dimensionality Reduction for Large-Scale Neural Recordings.” In: *Nature Neuroscience* 17.11, pp. 1500–1509.

Dehaene, GP, R Coen-Cagli, and A Pouget (2021). “Investigating the Representation of Uncertainty in Neuronal Circuits.” In: *PLOS Computational Biology* 17.2, e1008138.

Díaz-Francés, E and FJ Rubio (2013). "On the Existence of a Normal Approximation to the Distribution of the Ratio of Two Independent Normal Random Variables." In: *Statistical Papers* 54.2, pp. 309–323.

Doiron, B et al. (2016). "The Mechanics of State-Dependent Neural Correlations." In: *Nature Neuroscience* 19.3, pp. 383–393.

Echeveste, R et al. (2019). *Cortical-like Dynamics in Recurrent Circuits Optimized for Sampling-Based Probabilistic Inference*. 9. Neuroscience, pp. 1138–1149.

Ecker, AS et al. (2014). "State dependence of noise correlations in macaque primary visual cortex." In: *Neuron* 82.1, pp. 235–248.

Festa, D et al. (2021). "Neuronal Variability Reflects Probabilistic Inference Tuned to Natural Image Statistics." In: *Nature Communications* 12.1, p. 3635.

Gardella, C, O Marre, and T Mora (2019). "Modeling the correlated activity of neural populations: A review." In: *Neural Computation* 31.2, pp. 233–269.

Geisler, WS and DG Albrecht (1992). "Cortical Neurons: Isolation of Contrast Gain Control." In: *Vision research* 32.8, pp. 1409–1410.

Goris, RL, JA Movshon, and EP Simoncelli (2014). "Partitioning Neuronal Variability." In: *Nature neuroscience* 17.6, pp. 858–865.

Granot-Atedgi, E et al. (2013). "Stimulus-Dependent Maximum Entropy Models of Neural Population Codes." In: *PLOS Computational Biology* 9.3, e1002922.

Gur, M, A Beylin, and DM Snodderly (1997). "Response Variability of Neurons in Primary Visual Cortex (V1) of Alert Monkeys." In: *Journal of Neuroscience* 17.8, pp. 2914–2920.

Haefner, RM, P Berkes, and J Fiser (2016). "Perceptual decision-making as probabilistic inference by neural sampling." In: *Neuron* 90.3, pp. 649–660.

Hayashi, T, Y Kato, and D Nozaki (2020). "Divisively Normalized Integration of Multisensory Error Information Develops Motor Memories Specific to Vision and Proprioception." In: *Journal of Neuroscience* 40.7, pp. 1560–1570.

Hayya, J, D Armstrong, and N Gressis (1975). "A Note on the Ratio of Two Normally Distributed Variables." In: *Management Science* 21.11, pp. 1338–1341.

Heeger, DJ and KO Zemlianova (2020). "A Recurrent Circuit Implements Normalization, Simulating the Dynamics of V1 Activity." In: *Proceedings of the National Academy of Sciences* 117.36, pp. 22494–22505.

Heeger, DJ (1992). "Normalization of Cell Responses in Cat Striate Cortex." In: *Visual neuroscience* 9.2, pp. 181–197.

Heeger, DJ and WE Mackey (2019). "Oscillatory Recurrent Gated Neural Integrator Circuits (ORGaNICs), a Unifying Theoretical Framework for Neural Dynamics." In: *Proceedings of the National Academy of Sciences* 116.45, pp. 22783–22794.

Hénaff, OJ et al. (2020). "Representation of Visual Uncertainty through Neural Gain Variability." In: *Nature Communications* 11.1, p. 2513.

Hennequin, G et al. (2018). "The Dynamical Regime of Sensory Cortex: Stable Dynamics around a Single Stimulus-Tuned Attractor Account for Patterns of Noise Variability." In: *Neuron* 98.4, 846–860.e5.

Henry, CA and A Kohn (2020). "Spatial Contextual Effects in Primary Visual Cortex Limit Feature Representation under Crowding." In: *Nature Communications* 11.1, p. 1687.

Hu, Y, J Zylberberg, and E Shea-Brown (2014). "The Sign Rule and beyond: Boundary Effects, Flexibility, and Noise Correlations in Neural Population Codes." In: *PLOS Computational Biology* 10.2, e1003469.

Huang, C et al. (2019). "Circuit Models of Low-Dimensional Shared Variability in Cortical Networks." In: *Neuron* 101.2, 337–348.e4.

Joe, H and JJ Xu (1996). "The Estimation Method of Inference Functions for Margins for Multivariate Models." In.

Josić, K et al. (2009). "Stimulus-Dependent Correlations and Population Codes." In: *Neural Computation* 21.10, pp. 2774–2804.

Kafashan, M et al. (2021). “Scaling of sensory information in large neural populations shows signatures of information-limiting correlations.” In: *Nature communications* 12.1, pp. 1–16.

Kanashiro, T et al. (2017). “Attentional Modulation of Neuronal Variability in Circuit Models of Cortex.” In: *eLife* 6. Ed. by P Latham, e23978.

Kanitscheider, I, R Coen-Cagli, and A Pouget (2015). “Origin of Information-Limiting Noise Correlations.” In: *Proceedings of the National Academy of Sciences* 112.50, E6973–E6982.

Kohn, A and MA Smith (2005). “Stimulus Dependence of Neuronal Correlation in Primary Visual Cortex of the Macaque.” In: *Journal of Neuroscience* 25.14, pp. 3661–3673.

Kohn, A et al. (2016). “Correlations and Neuronal Population Information.” In: *Annual Review of Neuroscience* 39.1, pp. 237–256.

Lange, RD and RM Haefner (2017). “Characterizing and interpreting the influence of internal variables on sensory activity.” In: *Current opinion in neurobiology* 46, pp. 84–89.

Lange, RD and RM Haefner (2022). “Task-induced neural covariability as a signature of approximate Bayesian learning and inference.” In: *PLoS computational biology* 18.3, e1009557.

Lin, IC et al. (2015). “The nature of shared cortical variability.” In: *Neuron* 87.3, pp. 644–656.

Litwin-Kumar, A and B Doiron (2012). “Slow Dynamics and High Variability in Balanced Cortical Networks with Clustered Connections.” In: *Nature Neuroscience* 15.11, pp. 1498–1505.

Liu, LD, RM Haefner, and CC Pack (2016). “A Neural Basis for the Spatial Suppression of Visual Motion Perception.” In: *eLife* 5, e16167.

Marsaglia, G et al. (2006). “Ratios of Normal Variables.” In: *Journal of Statistical Software* 16.4, pp. 1–10.

Mitchell, JF, KA Sundberg, and JH Reynolds (2009). “Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4.” In: *Neuron* 63.6, pp. 879–888.

Moreno-Bote, R et al. (2014). “Information-limiting correlations.” In: *Nature Neuroscience* 17.10, pp. 1410–1417.

Ohshiro, T, DE Angelaki, and GC DeAngelis (2017). "A Neural Signature of Divisive Normalization at the Level of Multisensory Integration in Primate Cortex." In: *Neuron* 95.2, 399–411.e8.

Ohshiro, T, DE Angelaki, and GC DeAngelis (2011). "A Normalization Model of Multisensory Integration." In: *Nature Neuroscience* 14.6, pp. 775–782.

Orbán, G and M Bányai (2019). "Noise Correlations and Perceptual Inference." In: *Current Opinion in Neurobiology* 58.

Orbán, G et al. (2016). "Neural Variability and Sampling-Based Probabilistic Representations in the Visual Cortex." In: *Neuron* 92.2, pp. 530–543.

Pachitariu, M et al. (2017). "Suite2p: beyond 10,000 neurons with standard two-photon microscopy." In: *BioRxiv*.

Pearson, K (1897). "Mathematical Contributions to the Theory of Evolution.—On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs." In: *Proceedings of the Royal Society of London* 60.359-367, pp. 489–498.

Peirce, J et al. (2019). "PsychoPy2: Experiments in behavior made easy." In: *Behavior research methods* 51.1, pp. 195–203.

Pham-Gia, T, N Turkkan, and E Marchand (2006). "Density of the Ratio of Two Normal Random Variables and Applications." In: *Communications in Statistics-Theory and Methods* 35.9, pp. 1569–1591.

Pillow, JW and EP Simoncelli (2006). "Dimensionality reduction in neural models: an information-theoretic generalization of spike-triggered average and covariance analysis." In: *Journal of vision* 6.4, pp. 9–9.

Pillow, JW et al. (2008). "Spatio-Temporal Correlations and Visual Signalling in a Complete Neuronal Population." In: *Nature* 454.7207, pp. 995–999.

Ponce-Alvarez, A et al. (2013). "Stimulus-Dependent Variability and Noise Correlations in Cortical MT Neurons." In: *Proceedings of the National Academy of Sciences* 110.32, pp. 13162–13167.

Rabinowitz, NC et al. (2015). “Attention Stabilizes the Shared Gain of V4 Populations.” In: *eLife* 4, e08998.

Reynolds, JH and DJ Heeger (2009). “The normalization model of attention.” In: *Neuron* 61.2, pp. 168–185.

Rikhye, RV and M Sur (2015). “Spatial Correlations in Natural Scenes Modulate Response Reliability in Mouse Visual Cortex.” In: *Journal of Neuroscience* 35.43, pp. 14661–14680.

Ringach, DL (2010). “Population Coding under Normalization.” In: *Vision research* 50.22, pp. 2223–2232.

Ruff, DA and MR Cohen (2014). “Attention Can Either Increase or Decrease Spike Count Correlations in Visual Cortex.” In: *Nature Neuroscience* 17.11, pp. 1591–1597.

Ruff, DA and MR Cohen (2017). “A Normalization Model Suggests That Attention Changes the Weighting of Inputs between Visual Areas.” In: *Proceedings of the National Academy of Sciences* 114.20, E4085–E4094.

Ruff, DA and MR Cohen (2016). “Stimulus dependence of correlated variability across cortical areas.” In: *Journal of Neuroscience* 36.28, pp. 7546–7556.

Rumyantsev, OI et al. (2020). “Fundamental bounds on the fidelity of sensory cortical coding.” In: *Nature* 580.7801, pp. 100–105.

Sadagopan, S and D Ferster (2012). “Feedforward Origins of Response Variability Underlying Contrast Invariant Orientation Tuning in Cat Visual Cortex.” In: *Neuron* 74.5, pp. 911–923.

Sawada, T and AA Petrov (2017). “The Divisive Normalization Model of V1 Neurons: A Comprehensive Comparison of Physiological Data and Model Predictions.” In: *Journal of Neurophysiology* 118.6, pp. 3051–3091.

Schneidman, E et al. (2006). “Weak Pairwise Correlations Imply Strongly Correlated Network States in a Neural Population.” In: *Nature* 440.7087, pp. 1007–1012.

Schölvinck, ML et al. (2015). “Cortical State Determines Global Variability and Correlations in Visual Cortex.” In: *Journal of Neuroscience* 35.1, pp. 170–178.

Shadlen, MN and WT Newsome (1998). "The Variable Discharge of Cortical Neurons: Implications for Connectivity, Computation, and Information Coding." In: *Journal of Neuroscience* 18.10, pp. 3870–3896.

Snyder, AC et al. (2014). "Correlations in V1 Are Reduced by Stimulation Outside the Receptive Field." In: *Journal of Neuroscience* 34.34, pp. 11222–11227.

Sokoloski, S, A Aschner, and R Coen-Cagli (2021). "Modelling the neural code in large populations of correlated neurons." In: *eLife* 10, e64615.

Sridharan, S et al. (2022). "High-performance microbial opsins for spatially and temporally precise perturbations of large neuronal networks." In: *Neuron*.

Sturmfels, B (2002). *Solving Systems of Polynomial Equations*. 97. American Mathematical Soc.

Tolhurst, DJ, JA Movshon, and AF Dean (1983). "The Statistical Reliability of Signals in Single Neurons in Cat and Monkey Visual Cortex." In: *Vision research* 23.8, pp. 775–785.

Tripp, BP (2012). "Decorrelation of spiking variability and improved information transfer through feedforward divisive normalization." In: *Neural Computation* 24.4, pp. 867–894.

Umakantha, A et al. (2021). "Bridging neuronal correlations and dimensionality reduction." In: *Neuron* 109.17, pp. 2740–2754.

Ver Hoef, JM (2012). "Who Invented the Delta Method?" In: *The American Statistician* 66.2, pp. 124–127.

Verhoef, BE and JHR Maunsell (2017). "Attention-related changes in correlated neuronal activity arise from normalization mechanisms." In: *Nature Neuroscience* 20.7, pp. 969–977.

Wekselblatt, JB et al. (2016). "Large-scale imaging of cortical dynamics during sensory perception and behavior." In: *Journal of neurophysiology* 115.6, pp. 2852–2866.

Whiteway, MR, B Averbeck, and DA Butts (2020). "A latent variable approach to decoding neural population activity." In: *bioRxiv*.

Whiteway, MR et al. (2019). "Characterizing the nonlinear structure of shared variability in cortical neuron populations using latent variable models." In: *Neurons, behavior, data analysis and theory* 3.1.

Yu, BM et al. (2008). "Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity." In: *Advances in neural information processing systems* 21.

Zhao, Y and IM Park (2017). "Variational Latent Gaussian Process for Recovering Single-Trial Dynamics from Population Spike Trains." In: *Neural Computation* 29.5, pp. 1293–1316.

Zohary, E, MN Shadlen, and WT Newsome (1994). "Correlated neuronal discharge rate and its implications for psychophysical performance." In: *Nature* 370.6485, pp. 140–143.

5 Appendix

5.1 Derivation of Relationship Between Mean Normalization Strength and Noise Correlations

We use the same notation as in the Methods section. The true characterization of how changes in the mean normalization strengths (μ_{D_1}, μ_{D_2}) cause changes in $\rho^{NC} := \mathrm{Corr}(R_1, R_2)$ (8) would be to consider the directional derivative $\nabla \rho^{NC} \cdot \frac{1}{\sqrt{2}}(1,1)$. However, it will suffice to consider the partial derivative with respect to either μ_{D_1} or μ_{D_2}. We note a few identities which are easy to verify:

$$\frac{\partial \delta_D}{\partial \mu_D} = \left(\frac{\beta_D}{2} - 1\right)\frac{\delta_D}{\mu_D} \qquad \widehat{\sigma_R}^2 := \delta_N^2 + \delta_D^2 + \widehat{\sigma_\eta}^2$$

$$\frac{\partial \widehat{\sigma_\eta}}{\partial \mu_D} = \frac{\widehat{\sigma_\eta}}{\mu_D} \qquad \frac{\partial \widehat{\sigma_R}^2}{\partial \mu_D} = (\beta_D - 2)\frac{\delta_D^2}{\mu_D} + 2\frac{\widehat{\sigma_\eta}^2}{\mu_D}$$

Then considering $\frac{\partial \rho^{NC}}{\partial \mu_{D_1}}$, we compute using the chain rule:

$$\begin{aligned}
\frac{\partial \rho^{NC}}{\partial \mu_{D_1}} &= (\rho_N \delta_{N_1}\delta_{N_2} + \rho_D \delta_{D_1}\delta_{D_2} + \rho_\eta \widehat{\sigma_{\eta_1}}\widehat{\sigma_{\eta_2}})\frac{\partial}{\partial \mu_{D_1}}(\widehat{\sigma_{R_1}}^2\widehat{\sigma_{R_2}}^2)^{-1/2} \\
&\quad + \frac{\rho_D \frac{\partial \delta_{D_1}}{\partial \mu_{D_1}}\delta_{D_2} + \rho_\eta \frac{\partial \widehat{\sigma_{\eta_1}}}{\partial \mu_{D_1}}\widehat{\sigma_{\eta_2}}}{(\widehat{\sigma_{R_1}}^2\widehat{\sigma_{R_2}}^2)^{1/2}} \\
&= (\rho_N \delta_{N_1}\delta_{N_2} + \rho_D \delta_{D_1}\delta_{D_2} + \rho_\eta \widehat{\sigma_{\eta_1}}\widehat{\sigma_{\eta_2}})(-\frac{1}{2}(\widehat{\sigma_{R_1}}^2\widehat{\sigma_{R_2}}^2)^{-3/2})\widehat{\sigma_{R_2}}^2\left((\beta_{D_1} - 2)\frac{\delta_{D_1}^2}{\mu_{D_1}} + 2\frac{\widehat{\sigma_{\eta_1}}^2}{\mu_{D_1}}\right) \\
&\quad + \frac{\rho_D \left(\frac{\beta_{D_1}}{2} - 1\right)\frac{\delta_{D_1}}{\mu_{D_1}}\delta_{D_2} + \rho_\eta \frac{\widehat{\sigma_{\eta_1}}}{\mu_{D_1}}\widehat{\sigma_{\eta_2}}}{(\widehat{\sigma_{R_1}}^2\widehat{\sigma_{R_2}}^2)^{1/2}}
\end{aligned}$$

Simplifying, we get the equation:

$$\frac{\partial \rho^{NC}}{\partial \mu_{D_1}} = \frac{1}{\mu_{D_1}\widehat{\sigma_{R_1}}\widehat{\sigma_{R_2}}}\Bigg((\rho_N \delta_{N_1}\delta_{N_2} + \rho_D \delta_{D_1}\delta_{D_2} + \rho_\eta \widehat{\sigma_{\eta_1}}\widehat{\sigma_{\eta_2}})\left(\frac{\left(1-\frac{\beta_{D_1}}{2}\right)\delta_{D_1}^2 - \widehat{\sigma_{\eta_1}}^2}{\widehat{\sigma_{R_1}}^2}\right) \tag{15}$$

$$+ \left(\frac{\beta_D}{2} - 1\right)\rho_D \delta_{D_1}\delta_{D_2} + \rho_\eta \widehat{\sigma_{\eta_1}}\widehat{\sigma_{\eta_2}}\Bigg)$$

To reproduce the patterns noted in Figure 3a, we first set $\eta \equiv 0$ and $\rho_N = 0, \rho_D > 0$. Then

$$\begin{aligned}\frac{\partial \rho^{NC}}{\partial \mu_{D_1}} &= \frac{1}{\mu_{D_1}\widehat{\sigma_{R_1}}\widehat{\sigma_{R_2}}}\Bigg((\rho_D \delta_{D_1}\delta_{D_2})\left(\frac{\left(1-\frac{\beta_{D_1}}{2}\right)\delta_{D_1}^2}{\widehat{\sigma_{R_1}}^2} + \left(\frac{\beta_{D_1}}{2} - 1\right)\rho_D \delta_{D_1}\delta_{D_2}\right)\Bigg) \\ &= \frac{\rho_D \delta_{D_1}\delta_{D_2}}{\mu_{D_1}\widehat{\sigma_{R_1}}\widehat{\sigma_{R_2}}}\left(1-\frac{\beta_{D_1}}{2}\right)\left(\frac{\delta_{D_1}^2}{\widehat{\sigma_{R_1}}^2} - 1\right) < 0 \end{aligned} \tag{16}$$

Because $\beta_{D_1} < 2 \Rightarrow 1 - \frac{\beta_{D_1}}{2} > 0$. Moreover, $\widehat{\sigma_{R_1}}^2 - \delta_{D_1}^2 = \delta_{N_1}^2 > 0 \Rightarrow \frac{\delta_{D_1}^2}{\widehat{\sigma_{R_1}}^2} < 1$. Thus, ρ^{NC} decreases as a function of normalization strength.

Similarly, when $\eta \equiv 0, \rho_D = 0, \rho_N > 0$ (Fig 3b, then

$$\frac{\partial \rho^{NC}}{\partial \mu_{D_1}} = \frac{\rho_N \delta_{N_1}\delta_{N_2}}{\mu_{D_1}\widehat{\sigma_{R_1}}\widehat{\sigma_{R_2}}}\left(1-\frac{\beta_{D_1}}{2}\right)\frac{\delta_{D_1}^2}{\widehat{\sigma_{R_1}}^2} > 0 \tag{17}$$

Demonstrating that ρ^{NC} increases with normalization in this case.

When ρ_N, ρ_D are the opposite signs, the inequalities in (16) and (17) reverse and the relationship between ρ^{NC} and normalization switch to be increasing and decreasing, respectively.

Next, when $\eta \equiv 0$ and $\rho_N = -\rho_D = \rho$, (15) reduces to:

$$\frac{\partial \rho^{NC}}{\partial \mu_{D_1}} = \rho \frac{1}{\mu_{D_1} \widehat{\sigma_{R_1}} \widehat{\sigma_{R_2}}} \left(1 - \frac{\beta_{D_1}}{2}\right) \left(\delta_{D_1} \delta_{D_2} \left(1 - \frac{\delta_{D_1}^2}{\widehat{\sigma_{R_1}}^2}\right) + \delta_{N_1} \delta_{N_2} \frac{\delta_{D_1}^2}{\widehat{\sigma_{R_1}}^2}\right) \tag{18}$$

Thus, when $\rho > 0 \Rightarrow \frac{\partial \rho^{NC}}{\partial \mu_{D_1}} > 0$. In fact, we can use (18) along with (16,17) to generalize to all cases where ρ_N, ρ_D have opposite signs. As an example, suppose $\rho_D < 0$ and $\rho_N = |\rho_D| + C > 0, C > 0$. Then we can decompose $\frac{\partial \rho^{NC}}{\partial \mu_{D_1}}$ as a combination of (18, with $\rho = |\rho_D|$) and (17, with $\rho_N = C > 0$). Since both of these terms are positive, we conclude that ρ^{NC} increases with normalization when $\rho_N > 0, \rho_D < 0, \rho_N > |\rho_D|$.

Lastly, we consider when $\eta \not\equiv 0$. The nature of (15) prevents us from making any conclusions in general, as the sign of $\frac{\partial \rho^{NC}}{\partial \mu_{D_1}}$ will depend on the magnitudes of the various terms, similar to when $\rho_N, \rho_D > 0$. To get some understanding of the behavior of ρ^{NC}, let us assume that $\rho_\eta = 0, \rho_D = 0$, a generalization of (17). For these assumptions, we have

$$\frac{\partial \rho^{NC}}{\partial \mu_{D_1}} = \frac{\rho_N \delta_{N_1} \delta_{N_2}}{\mu_{D_1} \widehat{\sigma_{R_1}} \widehat{\sigma_{R_2}}} \left(\left(1 - \frac{\beta_{D_1}}{2}\right) \frac{\delta_{D_1}^2}{\widehat{\sigma_{R_1}}^2} - \frac{\widehat{\sigma_{\eta_1}}^2}{\widehat{\sigma_{R_1}}^2}\right) \tag{19}$$

Here, if $\frac{\widehat{\sigma_{\eta_1}}^2}{\widehat{\sigma_{R_1}}^2} \approx 0$ (i.e. $\widehat{\sigma_{\eta_1}}^2 << \widehat{\sigma_{R_1}}^2$), then (19) should be approximately equal to (17). This gives the heuristic that if the residual variance is much less than the variance of the neural response, then the presence of this noise should have little effect on these derived relationships.

5.2 Derivation of Negative Log-Likelihood for the Model

The model assumptions that the data is distributed according to a bivariate Gaussian distribution with mean and covariance given by the RoG approximation (5,6). From this, we

calculate the negative log-likelihood:

$$
\begin{aligned}
&-\log p(\{\boldsymbol{R}_t(s)\};\Theta) = -T\sum_{s\in S}\langle \log \mathcal{N}(\boldsymbol{R}_t(s);\boldsymbol{\mu}_{\boldsymbol{R}}(s,\Theta),\boldsymbol{\Sigma}_{\boldsymbol{R}}(s,\Theta))\rangle_T \qquad (20)\\
&\approx \frac{T}{2}\sum_{s\in S}\Big\{\log|\boldsymbol{\Sigma}_{\boldsymbol{R}}(s,\Theta)| + \big\langle(\boldsymbol{R}_t(s)-\boldsymbol{\mu}_{\boldsymbol{R}}(s,\Theta))^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)(\boldsymbol{R}_t(s)-\boldsymbol{\mu}_{\boldsymbol{R}}(s,\Theta))\big\rangle_T\Big\}\\
&\approx \frac{T}{2}\sum_{s\in S}\Big\{\log|\boldsymbol{\Sigma}_{\boldsymbol{R}}(s,\Theta)| + \overbrace{\big\langle\boldsymbol{R}_t(s)^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\boldsymbol{R}_t(s)\big\rangle_T}^{*} - \big\langle\boldsymbol{R}_t(s)^\top\big\rangle_T\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\boldsymbol{\mu}_{\boldsymbol{R}}(s,\Theta)\\
&\qquad - \boldsymbol{\mu}_{\boldsymbol{R}}(s,\Theta)^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\langle\boldsymbol{R}_t(s)\rangle_T + \boldsymbol{\mu}_{\boldsymbol{R}}(s,\Theta)^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\boldsymbol{\mu}_{\boldsymbol{R}}(s,\Theta)\Big\}\\
&\approx \frac{T}{2}\sum_{s\in S}\Big\{\log|\boldsymbol{\Sigma}_{\boldsymbol{R}}(s,\Theta)| + \overbrace{\mathrm{Tr}(\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\widehat{\boldsymbol{\Sigma}_{\boldsymbol{R}}}(s)) + \widehat{\boldsymbol{\mu}_{\boldsymbol{R}}}(s)^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\widehat{\boldsymbol{\mu}_{\boldsymbol{R}}}(s)}^{*} \qquad (21)\\
&\qquad - \boldsymbol{\mu}_{\boldsymbol{R}}(s,\Theta)^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\widehat{\boldsymbol{\mu}_{\boldsymbol{R}}}(s) - \widehat{\boldsymbol{\mu}_{\boldsymbol{R}}}(s)^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\boldsymbol{\mu}_{\boldsymbol{R}}(s,\Theta)\\
&\qquad + \boldsymbol{\mu}_{\boldsymbol{R}}(s,\Theta)^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\boldsymbol{\mu}_{\boldsymbol{R}}(s,\Theta)\Big\}
\end{aligned}
$$

The terms marked by $*$ are equal according to:

$$
\begin{aligned}
\big\langle\boldsymbol{R}_t(s)^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\boldsymbol{R}_t(s)\big\rangle_T &= \mathrm{Tr}\Big(\big\langle\boldsymbol{R}_t(s)^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\boldsymbol{R}_t(s)\big\rangle_T\Big) = \mathrm{Tr}\Big(\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\langle\boldsymbol{R}_t(s)^\top\boldsymbol{R}_t(s)\rangle_T\Big)\\
&= \mathrm{Tr}\Big(\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)(\widehat{\boldsymbol{\Sigma}_{\boldsymbol{R}}}(s) + \widehat{\boldsymbol{\mu}_{\boldsymbol{R}}}(s)\widehat{\boldsymbol{\mu}_{\boldsymbol{R}}}(s)^\top)\Big)\\
&= \mathrm{Tr}(\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\widehat{\boldsymbol{\Sigma}_{\boldsymbol{R}}}(s)) + \mathrm{Tr}\Big(\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\widehat{\boldsymbol{\mu}_{\boldsymbol{R}}}(s)\widehat{\boldsymbol{\mu}_{\boldsymbol{R}}}(s)^\top)\Big)\\
&= \mathrm{Tr}(\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\widehat{\boldsymbol{\Sigma}_{\boldsymbol{R}}}(s)) + \widehat{\boldsymbol{\mu}_{\boldsymbol{R}}}(s)^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(s,\Theta)\widehat{\boldsymbol{\mu}_{\boldsymbol{R}}}(s)
\end{aligned}
$$

Using the fact that $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ and $c = \mathrm{Tr}(c)$ for A, B matrices and c a scalar. Simplifying (21) produces (12).

5.3 Negative Log-Posterior for Inference of Single-Trial Normalization Strength

Using Bayes' theorem and the change of variables formula, we compute the negative log of the probability distribution of the single-trial normalization signals ($\boldsymbol{D}_t$) given the observed

neural activity during that trial ($\boldsymbol{R}_t$) (13) as:

$$-\log p(\boldsymbol{D}_t|\boldsymbol{R}_t = -\log|D_1 D_2| - \log p(N_1, N_2) - \log p(D_1, D_2) \tag{22}$$

Where $\boldsymbol{N}_t = \boldsymbol{R}_t \boldsymbol{D}_t$. We use the generative model (3) and simplify to obtain the following expression:

$$
\begin{aligned}
-\log p(\boldsymbol{D}_t|\boldsymbol{R}_t) = &-\log|D_1 D_2| \\
&+\frac{1}{2}\Bigg[D_1^2\left(\frac{R_1^2}{\sigma_{N_1}^2\left(1-\rho_N^2\right)}+\frac{1}{\sigma_{D_1}^2\left(1-\rho_D^2\right)}\right) && \text{(A1)}\\
&+D_1\left(\frac{-2\mu_{N_1}R_1}{\sigma_{N_1}^2\left(1-\rho_N^2\right)}+\frac{2\rho_N\mu_{N_2}R_1}{\sigma_{N_1}\sigma_{N_2}\left(1-\rho_N^2\right)}-\frac{2\mu_{D_1}}{\sigma_{D_1}^2\left(1-\rho_D^2\right)}+\frac{2\rho_D\mu_{D_2}}{\sigma_{D_1}\sigma_{D_2}\left(1-\rho_D^2\right)}\right) \\
& && \text{(B1)}\\
&+D_2^2\left(\frac{R_2^2}{\sigma_{N_2}^2\left(1-\rho_N^2\right)}+\frac{1}{\sigma_{D_2}^2\left(1-\rho_D^2\right)}\right) && \text{(A2)}\\
&+D_2\left(\frac{-2\mu_{N_2}R_2}{\sigma_{N_2}^2\left(1-\rho_N^2\right)}+\frac{2\rho_N\mu_{N_1}R_2}{\sigma_{N_1}\sigma_{N_2}\left(1-\rho_N^2\right)}-\frac{2\mu_{D_2}}{\sigma_{D_2}^2\left(1-\rho_D^2\right)}+\frac{2\rho_D\mu_{D_1}}{\sigma_{D_1}\sigma_{D_2}\left(1-\rho_D^2\right)}\right) \\
& && \text{(B2)}\\
&+D_1D_2\left(-\frac{2\rho_N R_1R_2}{\sigma_N\sigma_{N_2}\left(1-\rho_N^2\right)}-\frac{2\rho_D}{\sigma_{D_1}\sigma_{D_2}\left(1-\rho_D^2\right)}\right)\Bigg] && \text{(C)}
\end{aligned}
$$

Substituting the coefficients on the right, we can write this equation as:

$$
-\log p\left(\boldsymbol{D}_t|\boldsymbol{R}_t\right) \propto -\log|D_1D_2| + \frac{1}{2}\left(A_1{D_1}^2 + B_1D_1 + A_2{D_2}^2 + B_2D_2 + CD_1D_2\right) \quad (23)
$$

With the assumption that $D_1, D_2 > 0$ (which is expected since these are supposed to the sums of neural activity), we take the partial derivatives of (23) to obtain the polynomials (14).

5.4 Pairwise Model Outperforms the Independent Model in Simulations When Noise Correlations are Large

Our previous work [Coen-Cagli and Solomon, 2019] demonstrated that the independent RoG model provided an excellent fit to single-neuron data. Moreover, by construction, the independent and pairwise model best fit parameters are identical except for the correlation

parameters (the ρ parameters, see Methods). Thus, we expected any differences in the fit quality (as measured by the negative log-likelihood using the best fit parameters) between the independent and pairwise RoG models to be minimal for pairs of neurons with noise correlations close to zero, and that the pairwise RoG would outperform the independent RoG for pairs of neurons with prominent correlated variability.

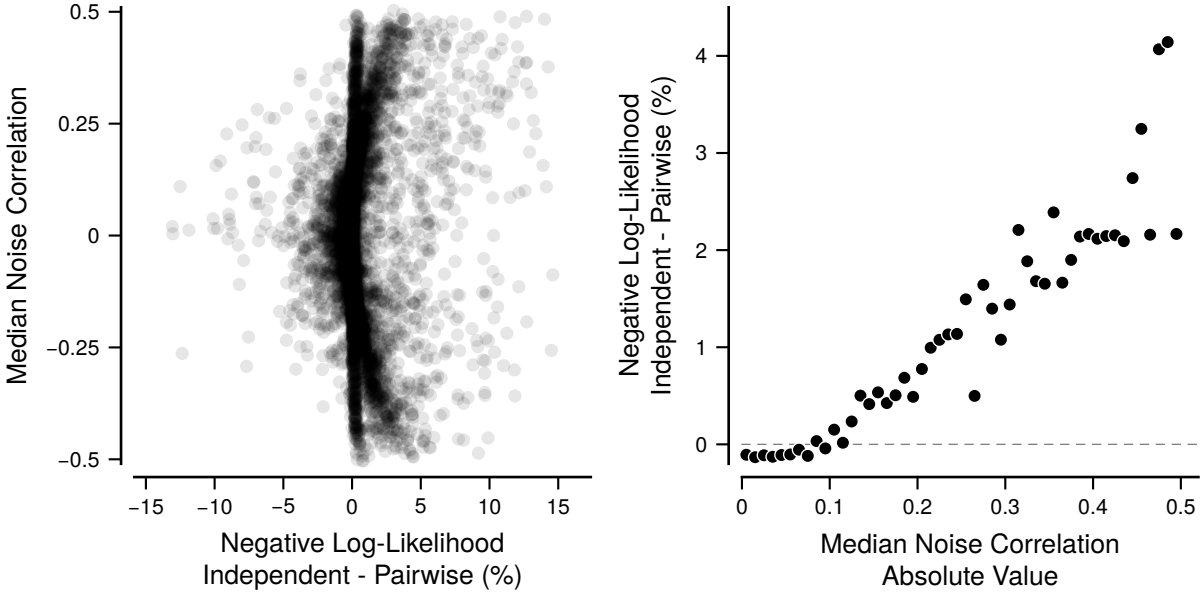


Figure S1: Improvement in Fit of the Pairwise RoG over the Independent RoG Depends on the Magnitude of the Noise Correlations
(a) Scatter plot demonstrating the dependence of the percent difference on the median noise correlation (across contrasts).
(b) Mean percent difference when binning by the magnitude of the median noise correlations.
Data were generated as described in Fig 4.

To test these predictions, we simulated pairs of neurons with parameters selected as described in the previous section (see Methods) and compared the negative log-likelihood values for the pairwise and independent model (a lower negative log-likelihood denotes better goodness of fit). For these simulations, we used 6 stimulus contrasts ($[6.25, 12.5, 25, 50, 100]$) and 1000 trials. We used cross-validation to account for the extra free parameters of the pairwise model (details in Methods). We found that 5443/11628 of the pairs were better captured by the pairwise model, and on average the pairwise model slightly outperformed

the independent model (mean percent difference = 0.964, $p<0.01$). We studied how the difference of likelihoods depends on the noise correlation of each pair (median noise correlation across contrasts; Fig S1a). We found that, for most pairs of neurons with magnitude of noise correlations greater than 0.1, the pairwise model outperformed the independent model (3411/4040 pairs). To quantify the dependence of this model improvement on noise correlation strength, we binned simulated pairs of neurons by the magnitude of noise correlations and computed the average difference of negative log-likelihoods for the pairwise and independent models (Fig S1b). This demonstrates that, as the magnitude of the noise correlations increases, the pairwise model provides an increasingly better fit to the data over the independent model.

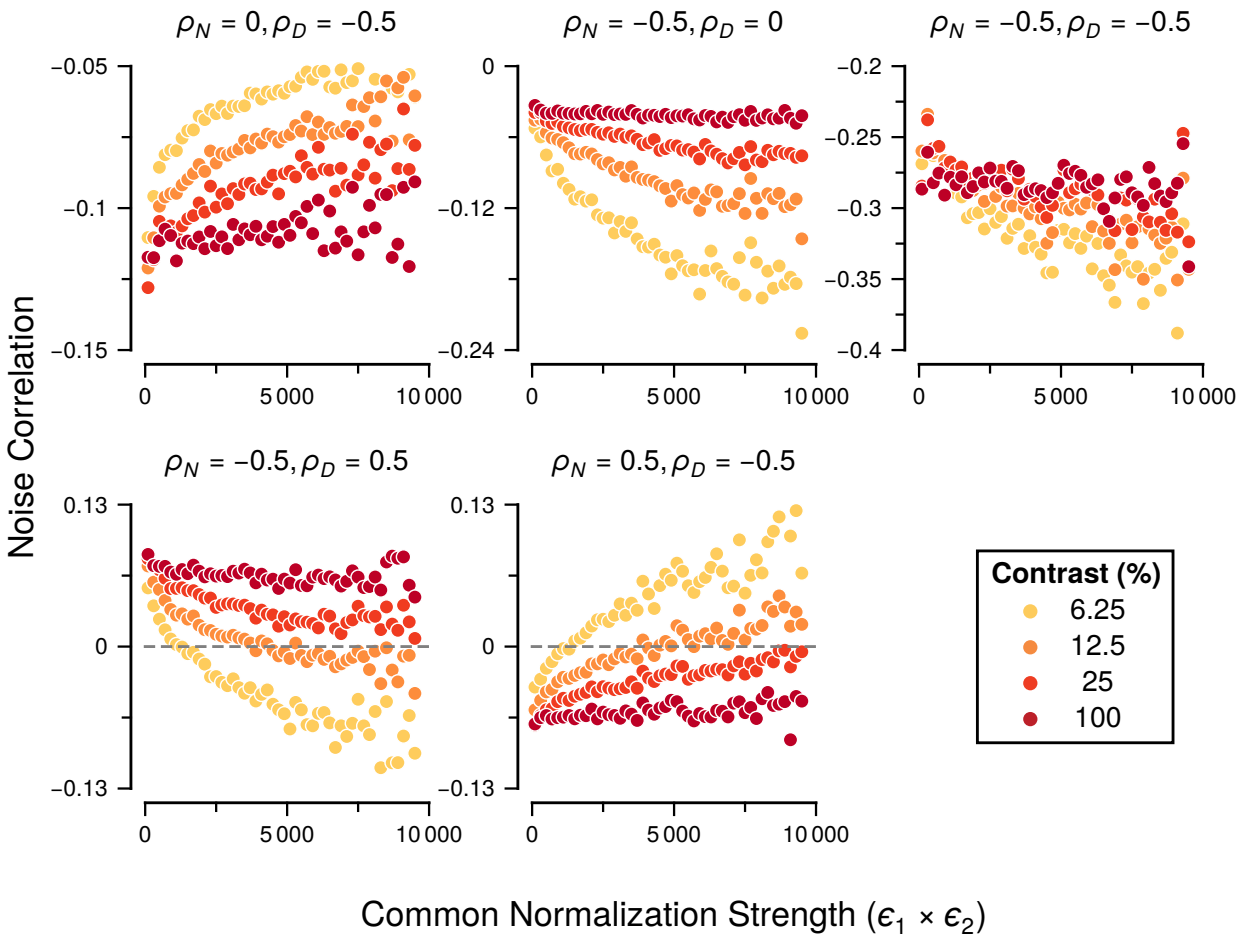


Figure S2: Relationship Between Noise Correlations and Denominator Strength
Expands upon Fig. 3 to include cases where (ρ_N, ρ_D) can be negative or have opposite signs. Figure was created using the exact same method and synthetic dataset as Fig. 3: see the caption in the main text for details.