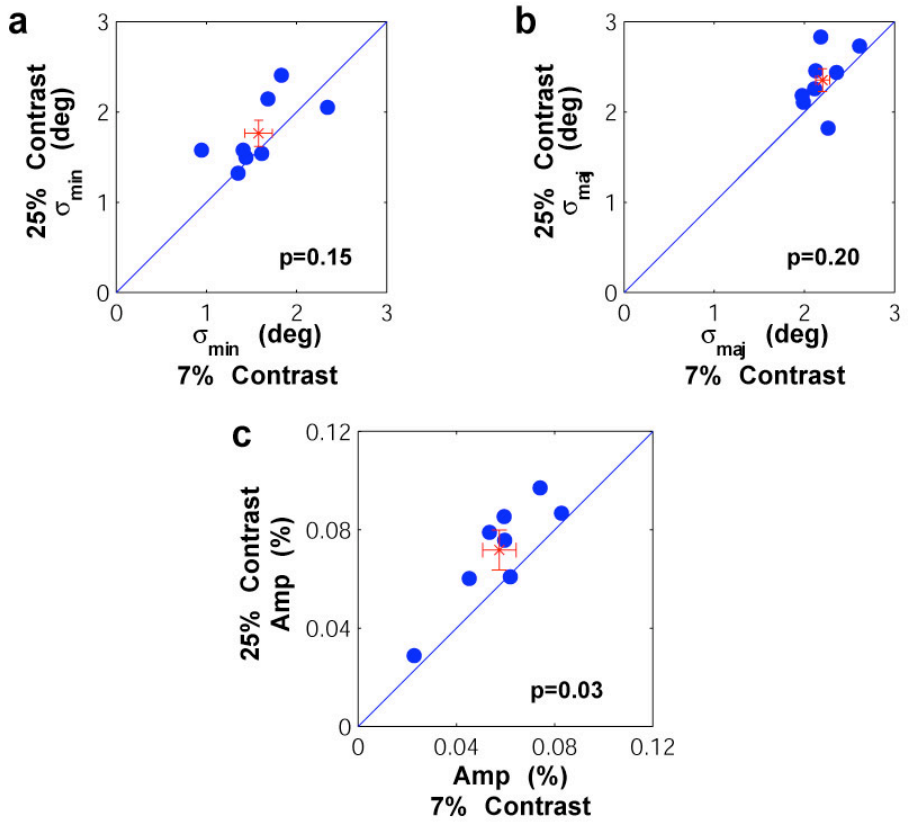
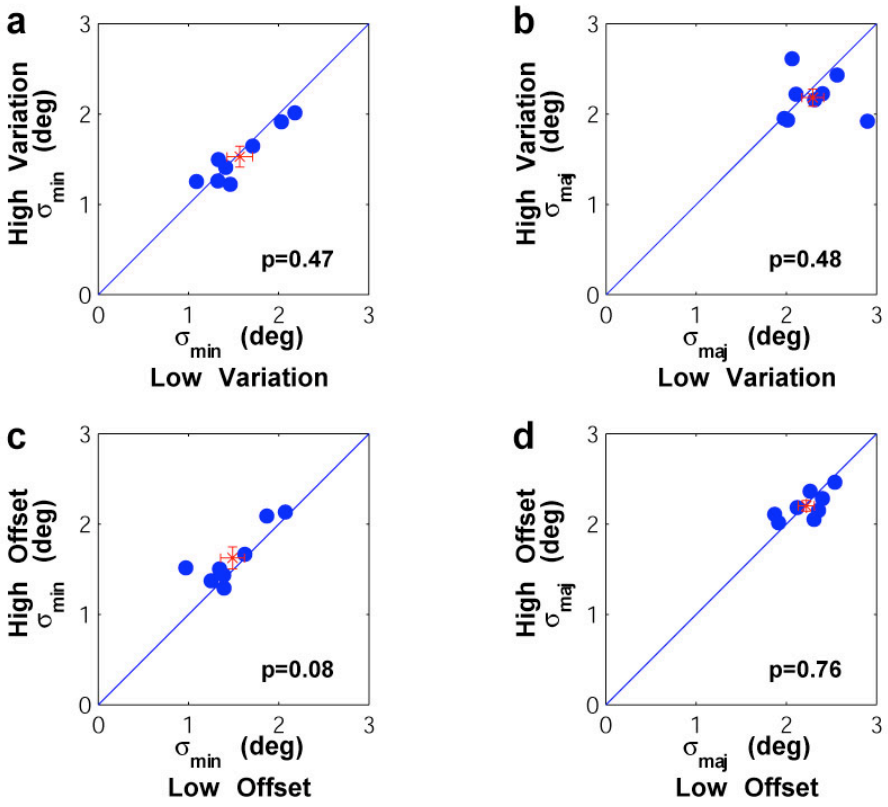


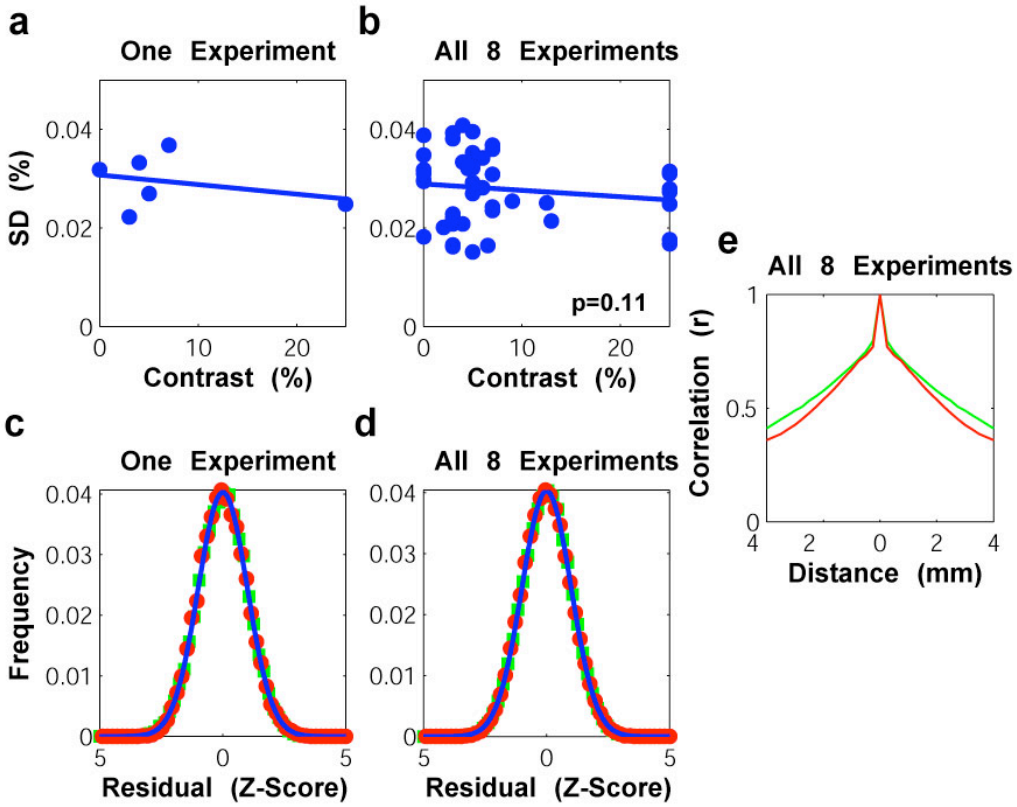
Supplementary Fig. 1 Effect of contrast on response spread. (**a**) and (**b**) scatter plot of $\sigma_{\min}$ and σ_{maj} for two target contrasts (7% and 25%) across all 8 experiments. The red error bars indicate one standard error and are centered on the mean value. The mean of $\sigma_{\min}$ and σ_{maj} are not significantly different for the two contrast ($P > 0.05$, paired *t-test*). (**c**) Scatter plot of response amplitude for the fitted 2D Gaussian in all 8 experiments. Response to 25% contrast target was significantly higher than response to 7% contrast target ($P = 0.03$, paired *t-test*). Despite the differences in response amplitude, response spread is not significantly different for the two target contrasts.



Supplementary Fig. 2 Effect of eye position on response spread. We first calculated the mean offset and standard deviation of eye positions for each target-present trial within the temporal interval over which VSD responses were averaged.

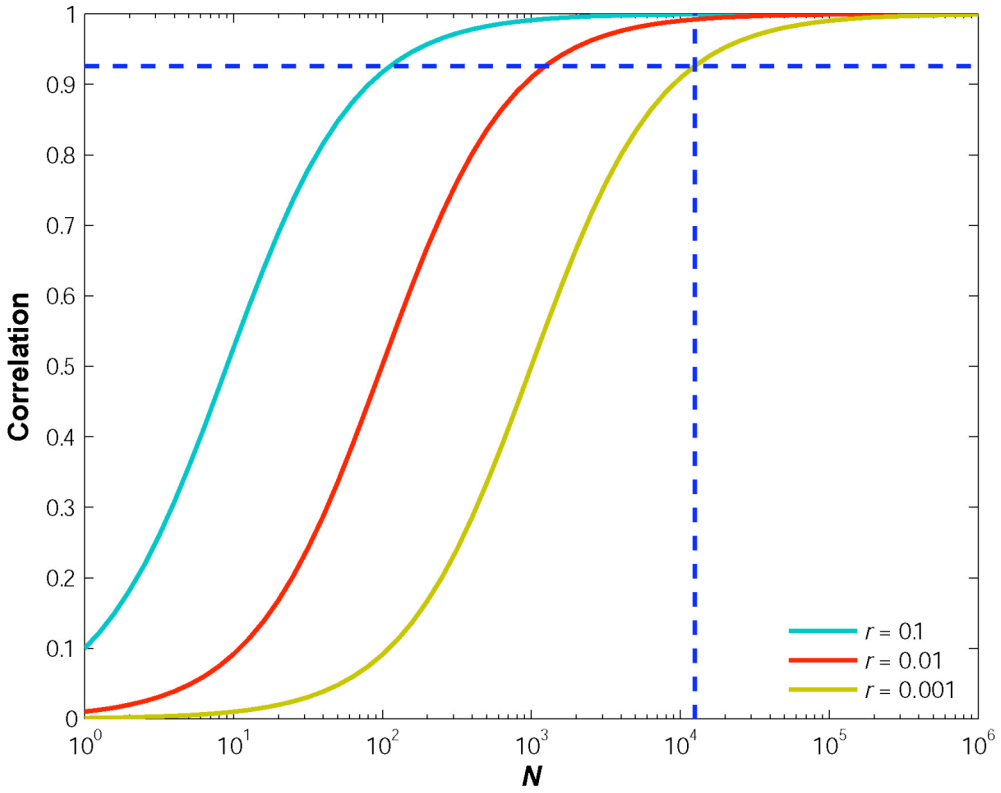
(**a**) and (**b**) scatter plots of $\sigma_{\min}$ and σ_{maj} for trials with values below the median variation (standard deviation) vs. trials with values above the median variation across all 8 experiments. (**c**) and (**d**) scatter plots of $\sigma_{\min}$ and σ_{maj} for trials with values below the median offset vs. trials with values above the median offset across all 8 experiments. Error bars – same convention as in **Supplementary Fig. 1**.

If variability and/or offset of eye positions contribute significantly to neural response spread, we would expect the spread to be significantly lower in the subset of trials with lower variance and/or offset. Instead, the average values fall close to the identity line, indicating that the contribution of eye position to the VSD response spread is minimal. The average value of the median offset was 0.32 deg and the average value of the median standard deviation was 0.003 deg^2 ($N = 8$).

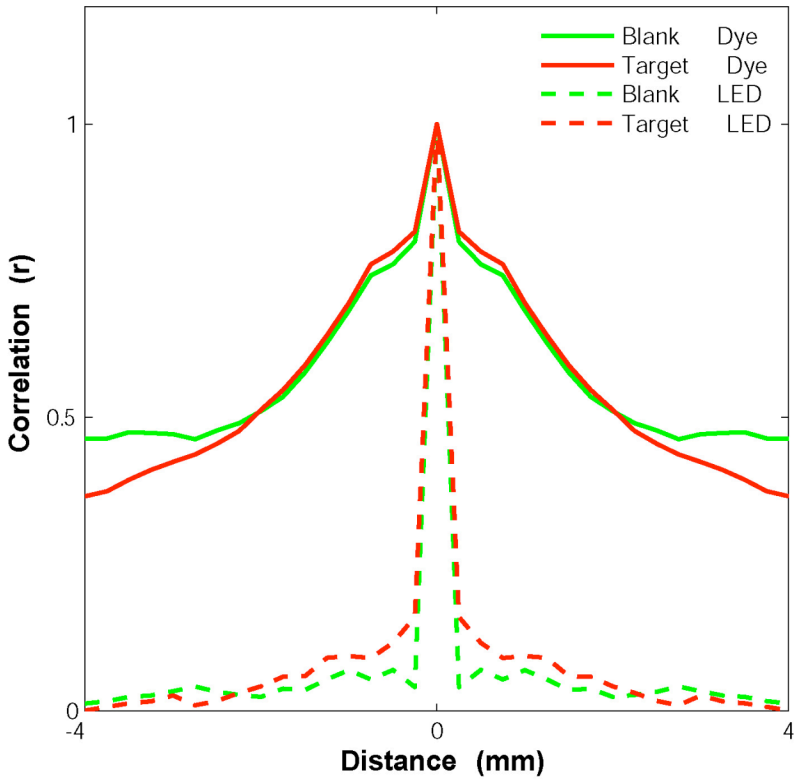


Supplementary Fig. 3 Variability in V1 population responses, as measured by VSD imaging, can be described as an additive (stimulus independent) Gaussian noise with widespread spatial correlations. (**a**) Response variability (over trials) as a function of target contrast at the maximal *d'* site in the example experiment. Blue curve represents linear regression line. (**b**) Same as (a), combined over all 8 experiments. Response variability is not significantly affected by stimulus contrast. The slope of the regression line was not significantly different from zero (*t-test, N = 8*). (**c**) Distribution of z-score values combined across all trials and all sites in an area of 8 x 8 mm in the example experiment. Residual responses at each site were first Z-transformed and then combined across sites. Red symbols – target-present trials; Green symbols – target-absent trials; Blue curve – Standard normal distribution. (**d**) Same as c. but combined across all 8 experiments. The variability in the VSD responses is approximately Gaussian and stimulus independent. (**e**), Average spatial correlations between pairs of sites as a function of their distance (as in **Fig. 2f**) averaged over all 8 experiments. Red – target-present trials; Green – target-absent trials. There is a small tendency for target present-

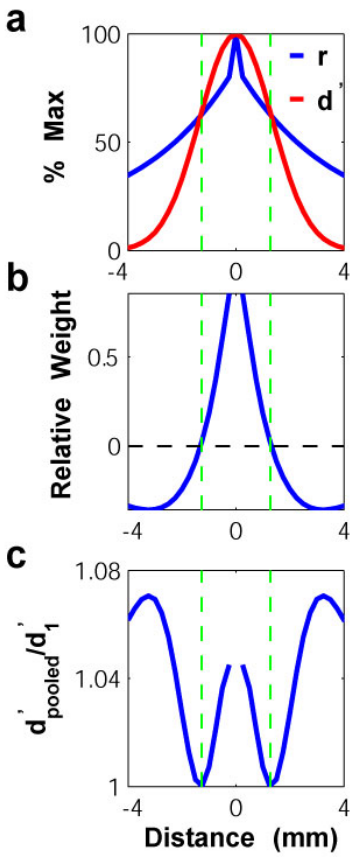
trials to have a slightly lower spatial correlation at long distances but the overall shape of the spatial correlation curves are very similar in target-present and target-absent trials.



Supplementary Fig. 4. Effect of pool size on correlations in population responses. Each curve shows the correlation between pooled responses as a function of pool size for different initial pairwise correlations between individual neurons (see **Supplementary Methods**, Eq. (6)). Dashed blue lines indicate the expected correlation between two neighboring 0.25 x 0.25mm pixels (see text).



Supplementary Fig. 5. Comparison between spatial correlations measured from V1 and spatial correlations measured during control experiment with light emitting diode (LED) (see text). Same conventions as in **Fig. 2f**. In the control experiment, the surface of an LED was illuminated with the same light source used during the VSD experiments. A signal with comparable magnitude as in the VSD experiments was simulated using an LED covered by a -3 log unit neutral density filter.



Supplementary Fig. 6. Optimal two-point pooling in the example experiment. **(a)** Normalized falloff in sensitivity along the minor axis of the average response in **Fig. 2e** (red) and falloff in correlations along the same axis (blue). (**b**) The ratio of the optimal weights (w_2 / w_1) as a function of the distance between the sites (obtained using Eq. 4, **Supplementary Methods**), where the first site, with sensitivity d_1', is the one with the maximal sensitivity (the peak of the red curve in a). (**c**) Expected improvement in the pooled sensitivity relative to d_1' as a function of the distance between the sites, obtained using Eq. 5 , (**Supplementary Methods)**.

At distances where $d_2'/d_1' > r$ (red curve above blue), w_2 is positive. The value of d'_{pooled} is high when sites are nearby and decreases rapidly to d_1' as the distance between the sites increases. At longer distances, where $d_2'/d_1' < r$, d'_{pooled} starts increasing again and w_2 become negative. Because d' falls off more rapidly than *r*, higher sensitivity is

obtained for second sites approximately 3.2 mm away from the d_1' site. Dashed green horizontal lines are the distances where $d_2'/d_1' = r$.

Supplementary Methods

Optimal pooling rules

We note that the *Optimal* pooling rule relies on the covariance matrix of responses at different sites. If we let x_i be the observed amplitude at site i $(i = 1,\ldots,n)$ on trial j $(j = 1,\ldots,m)$, then the covariance matrix $\mathbf{\Sigma}$ is an *n by n* matrix, where $\Sigma(i,i)$ is the variance of x_i across the *m* trials, and $\Sigma(i,k)$ is the covariance of x_i and x_k across the *m* trials.

The optimal pooling rule for two sites is easy to apply and provides useful intuitions about how correlated noisy responses should be combined (see **Fig. 3** and **Supplementary Fig. 6**). Consider responses from a pair of sites with sensitivity $\langle d_1', d_2' \rangle$, standard deviations $\langle \sigma_1, \sigma_2 \rangle$ and correlation *r*. Applying Eq. **Error! Reference source not found.**), we obtain the optimal set of weights:

$$(1) \qquad \mathbf{w} = \langle w_1, w_2 \rangle = \left\langle \frac{d_1' - r d_2'}{\sigma_1 \left(1 - r^2\right)}, \frac{d_2' - r d_1'}{\sigma_2 \left(1 - r^2\right)} \right\rangle$$

Further, applying Eq. (3), the combined sensitivity at the two sites is given by

$$(2) \qquad d'_{pooled} = \sqrt{\left(d_1'\right)^2 + \frac{1}{1 - r^2}\left(d_2' - d_1' r\right)^2}$$

Supplementary Fig. 6 shows w_2 and d'_{pooled} computed using Equations (4) and (5) in the example experiment.

The basic result for optimal pooling over two sites applies in the more general case where responses are pooled from *n* sites. However, there are often practical difficulties in determining the inverse covariance matrices required by equation (2); for

example, it is often impossible to invert very large matrices. A standard way to get around these difficulties is to determine the optimal weights using Fourier methods. Specifically, the Fourier transform of the average 2D spatial correlation function (the radial 2D version of **Fig. 2f**) gives the power spectrum of the Gaussian noise. From the power spectrum of the noise we can compute a *whitening kernel* (a spatial filter), which, when convolved with the spatial response, will decorrelate the noise in the response. **Fig. 3c** shows a cross section of the whitening kernel for the example experiment. It is well known that the inverse covariance matrix $\Sigma^{-1} = \Lambda^T \Lambda$, where Λ is the whitening matrix, and thus, convolving the whitening filter twice with the average response ***s*** is mathematically equivalent to computing the optimal weights in Eq. (2) [38].

The detection sensitivity of the different pooling rules was determined using a jackknife procedure [40]. In this procedure, a separate analysis is performed for each of the *m* trials. For each trial, model parameters and pooled responses are computed for the remaining *m-1* trials, and an optimal criterion is established based on those trials. This criterion is then applied to the pooled response from the unseen trial; the performance of the model is classified as correct, if the pooled response exceeds the criterion on a target-present trial, or remains below the criterion on a target-absent trial. This procedure is repeated for each trial in our data set to obtain the neurometric function. Analysis of simulated data shows that for the number of trials in our experiments (typically 100 trials), the jackknife procedure underestimates the detection sensitivity of the optimal pooling rule. This is an additional reason why performance estimates for the *Optimal* rule (**Fig. 4** and **Fig. 5**) should be viewed as a lower bound on the actual neuronal sensitivity.

Correlations in neural populations

Exceedingly weak pairwise correlations between neurons can lead to high correlations between pools of neurons. Considering the simple case of two populations of neurons with uniform pairwise correlation within and between pools, the correlation R_N between the pooled (summed) responses is given by:

$$R_N = \frac{Nr}{1+(N-1)r} \tag{6}$$

where N is the number of neurons in each pool and r is the pairwise correlation between individual neurons. This equation shows that even for very low pairwise correlations, the correlation between the pooled responses can reach high values for sufficiently large N (**Supplementary Fig. 4**).