

Supplementary information

Flexible Gating of Contextual Influences in Natural Vision

R. Coen-Cagli, A. Kohn, O. Schwartz

Supplementary Table 1	2
Supplementary Modeling.....	3
1. Bayesian model and image statistics.....	3
2. Image homogeneity and center-surround similarity	7

Supplementary Table 1

	Population size	Prediction quality (flexible, standard) Fig. 5	<i>p</i>	<i>MR</i> (hom, het) Fig. 6	<i>p</i>	<i>NMR</i> (hom, het) Fig. 7	<i>p</i>
Main text	126	0.69, 0.53	<0.0001	0.47, 0.62	<0.0001	1.32, 1.58	<0.0001
Least phase-sensitive units	63	0.69, 0.51	<0.0001	0.48, 0.64	<0.0001	1.24, 1.58	<0.0001
Most phase-sensitive units	63	0.69, 0.55	<0.0001	0.47, 0.61	<0.0001	1.45, 1.61	0.02
Filter properties from gratings	41	0.7, 0.51	<0.0001	0.41, 0.62	<0.0001	0.93, 1.53	<0.0001
Most well-isolated single units	30	0.74, 0.64	0.004	0.43, 0.57	0.002	1.14, 1.5	0.005
Full population	207	0.58, 0.5	0.0001	0.57, 0.69	<0.0001	1.35, 1.47	0.005

Supplementary Table 1. Results were robust to neuronal exclusion criteria, isolation quality, and method for defining model filter properties. The first row contains the values reported in the main text, as a reference. The second and third rows show results for the median split based on the neurons phase sensitivity index, F1/F0 (**Supplementary Figure 6**). We obtained similar results in the halves of the data, alleviating concern about the use of a quadrature pair to represent the RF. The fourth row shows results when the filters for the RF center and surround were based on measurements of spatial frequency and orientation tuning made with static gratings (**Supplementary Figure 7**). The fifth row shows results for only the most well-isolated single units (waveform SNR>2.75, as defined in Kelly et al. J Neurosci 27, 261-264 (2007)). The bottom row includes neurons that were excluded from the analysis of the main text, because they were not strongly surround-modulated. All reported results were thus robust to any of the selection criteria used.

Supplementary Modeling

1. Bayesian model and image statistics

The Bayesian model we used has been described in detail in previous publications (Coen-Cagli et al., *Advances in NIPS* 369–377, 2009; Coen-Cagli et al., *PLoS Comput Biol* 8, e1002405, 2012). For related work, see (Schwartz and Simoncelli, *Nat Neurosci* 4, 819-825, 2001; Schwartz et al., *Neural Comput* 18, 2680-2718, 2006; Guerrero-Colon et al., *IEEE ICIP* 565-568, 2008; Hammond and Simoncelli, *IEEE Trans Im Proc* 17, 2089-2101, 2008; Karklin and Lewicki, *Nature* 457, 83-86, 2009). Here we provide a brief description of the model, and of the specific implementation used to compare to neuronal responses.

Visual filters and the Bayesian model

Our Bayesian model was a generative model, the Mixture of Gaussian Scale Mixture (MGSM), that captures the characteristic dependencies between the outputs of neighboring filters, when these are applied to natural images (Coen-Cagli et al., *Advances in NIPS* 369 – 377, 2009; Schwartz et al., *Neural Comput* 18, 2680-2718, 2006; Wainwright et al., *Appl Comput Harmon Anal* 11, 89-123, 2001). In our case, the MGSM considers the outputs of two groups of visual filters, representing the RF and surround of a neuron. We derived the V1-like filters from a steerable pyramid (Simoncelli et al., *IEEE Trans Inf Theory* 38, 587-607, 1992; Portilla and Simoncelli, *Int J Comp Vis* 40, 49-70, 2000) and defined the filter outputs— \mathbf{k} and \mathbf{s} , as explained in the **Methods**—as the dot product between an image and the filters. When the filters are applied to homogeneous images, their outputs are dependent, because of global properties of the image. In this case, the filter outputs \mathbf{k} and \mathbf{s} are generated in the MGSM by multiplying local Gaussian variables by the same positive, scalar random variable denoted by \mathcal{V} (termed the mixer). The local Gaussian variables— \mathbf{K} and $\mathbf{\Sigma}$, with the same dimensionality as \mathbf{k} and \mathbf{s} —capture the strength of local image features in the RF and surround, respectively. Thus, the RF and surround filter outputs are given by:

$$\text{(Eq. S1)} \quad \begin{cases} \mathbf{k} = \mathcal{V}\mathbf{K} \\ \mathbf{s} = \mathcal{V}\mathbf{\Sigma} \end{cases}$$

While the dependencies between the elements of \mathbf{K} and $\mathbf{\Sigma}$ are fully described by a covariance matrix $\mathbf{C}_{(\mathcal{K}\mathcal{\Sigma})}$, the joint probability distribution of the elements of \mathbf{k} and \mathbf{s} also have a variance dependence, because the multiplication by the common mixer \mathcal{V} scales both variances similarly. This variance (“bowtie”) dependence is a hallmark feature of applying V1-like filters to natural images (Schwartz and Simoncelli, *Nat Neurosci* 4, 819-825, 2001; Wainwright et al., *Appl Comput Harmon Anal* 11, 89-123,

2001).

When the filters are applied to heterogeneous images, the filters' outputs are independent. In this case, the filter outputs are generated as follows:

$$(Eq. S2) \quad \begin{cases} \mathbf{k} = \mathbf{V}\mathbf{K} \\ \mathbf{s} = \mathbf{U}\mathbf{\sigma} \end{cases}$$

where \mathbf{U} is a mixer variable independent of \mathbf{V} , and \mathbf{K} and $\mathbf{\sigma}$ are described by two separate covariance matrices $\mathbf{C}_{(\mathbf{K})}$ and $\mathbf{C}_{(\mathbf{\sigma})}$.

Inference in the Bayesian model and divisive surround normalization

To specify the relation between the Bayesian model and the response of a neuron to an image, we assumed that V1 neurons aim to compute an optimal estimate of the local content of the image inside the RF, and thus to remove redundancy with the representation of nearby filters. This assumption is steeped in a rich literature that has argued that cortex tries to achieve an efficient representation by reducing redundancy (Attneave, Psychol Rev 61, 183, 1954; Barlow, in Sensory Communication, 217-234, MIT Press, 1961; Dayan and Abbott, Theoretical neuroscience, MIT Press, 1999), and that has shown that many aspects of cortical responses can be explained by this simple computational goal (Schwartz and Simoncelli, Nat Neurosci 4, 819-825, 2001; Simoncelli and Olshausen, Annu Rev Neurosci 24, 1193-1216, 2001; Olshausen and Field, Nature 381, 607-609, 1996; Bell and Sejnowski, Vision Res 37, 3327-3338, 1997; Simoncelli, Curr Opin Neurobiol 13, 144-149, 2003; Zhaoping, Network 17, 301-334, 2006). In the context of the MGSM, the local feature in the target neuron RF (defined, for instance, by a filter of preferred orientation θ and spatial frequency ω) is represented by the value of the underlying Gaussian variable ($\mathcal{K}_{\theta,\omega}$). Intuitively, given the observed sensory inputs (\mathbf{k} and \mathbf{s}), one could compute $\mathcal{K}_{\theta,\omega}$ by inverting (Eq. S1) or (Eq. S2), which amounts to the division of \mathbf{k} by \mathbf{V} . However, whether the outputs are dependent (i.e. (Eq. S1) applies) or independent (i.e. (Eq. S2) applies) for a given image is not known, and \mathbf{V} is also not observed directly. Instead, according to Bayesian inference, an estimate $\overline{\mathcal{K}_{\theta,\omega}}$ can be obtained by integrating out those unobserved variables (for the full analytical derivation see Coen-Cagli et al., Advances in NIPS 369–377, 2009; Schwartz et al., Neural Comput 18, 2680-2718, 2006). This operation, termed marginalization, is a general Bayesian prescription to ignore nuisance variables, and has recently been linked to divisive normalization in other contexts as well (Beck et al. *J Neurosci* **31**, 15310-15319, 2011).

If we introduce a binary variable ξ and denote the inferred probability that the filters are dependent or independent as $p(\xi = \xi_1 | \mathbf{k}, \mathbf{s})$ and $p(\xi = \xi_0 | \mathbf{k}, \mathbf{s})$, respectively, then the estimate of $\overline{\kappa_{\theta, \omega}}$ is given by:

$$(Eq. S3) \quad \overline{\kappa_{\theta, \omega}} \approx p(\xi = \xi_1 | \mathbf{k}, \mathbf{s}) \frac{|k_{\theta, \omega}|}{\sqrt{\lambda_{(\kappa\sigma)}}} + p(\xi = \xi_0 | \mathbf{k}, \mathbf{s}) \frac{|k_{\theta, \omega}|}{\sqrt{\lambda_{(\kappa)}}}$$

where:

$$(Eq. S4) \quad \begin{aligned} \lambda_{(\kappa\sigma)} &= \sqrt{(\mathbf{k}, \mathbf{s}) \mathbf{C}_{(\kappa\sigma)}^{-1} (\mathbf{k}, \mathbf{s})^\top} \\ &= \sqrt{\sum_{i,j=1}^{n_k} (\mathbf{C}_{(\kappa\sigma)}^{-1})_{i,j} \mathbf{k}_i \mathbf{k}_j + \sum_{i,j=n_k+1}^{n_k+n_s} (\mathbf{C}_{(\kappa\sigma)}^{-1})_{i,j} \mathbf{s}_i \mathbf{s}_j + 2 \sum_{i=1}^{n_k} \sum_{j=1+n_k}^{n_k+n_s} (\mathbf{C}_{(\kappa\sigma)}^{-1})_{i,j} \mathbf{k}_i \mathbf{s}_j} \end{aligned}$$

and

$$(Eq. S5) \quad \lambda_{(\kappa)} = \sqrt{\mathbf{k} \mathbf{C}_{(\kappa)}^{-1} \mathbf{k}^\top} = \sqrt{\sum_{i,j=1}^{n_k} (\mathbf{C}_{(\kappa)}^{-1})_{i,j} \mathbf{k}_i \mathbf{k}_j}$$

with n_k indicating the number of center filters, and n_s the number of surround filters.

The term $p(\xi = \xi_1 | \mathbf{k}, \mathbf{s})$ in (Eq. S3) represents the inferred posterior probability that the filter outputs are statistically dependent. This probability can also be computed analytically (Coen-Cagli et al., Advances in NIPS 369–377, 2009; Schwartz et al., Neural Comput 18, 2680-2718, 2006):

$$(Eq. S6) \quad \begin{aligned} p(\xi = \xi_1 | \mathbf{k}, \mathbf{s}) &\propto p(\xi = \xi_1) p(\mathbf{k}, \mathbf{s} | \xi = \xi_1) \\ &= p(\xi = \xi_1) \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C}_{(\kappa\sigma)})}} \cdot \frac{\mathcal{B}\left(1 - \frac{n}{2}; \lambda_{(\kappa\sigma)}\right)}{\lambda_{(\kappa\sigma)}^{\left(\frac{n-1}{2}\right)}} \end{aligned}$$

where $n = n_k + n_s$ is the total number of filters; \mathcal{B} is the type-2 modified Bessel function; and

$p(\xi = \xi_1)$ is the prior probability that the outputs are dependent, which we obtained by training the MGSM on a database of natural images (see **Supplementary Methods 3**). The inferred probability that the inputs are independent $p(\xi = \xi_0 | \mathbf{k}, \mathbf{s})$ has a similar expression (Coen-Cagli et al., PLoS Comput Biol 8, e1002405, 2012), but involving the covariances $\mathbf{C}_{(\kappa)}, \mathbf{C}_{(\sigma)}$.

When the posterior probability that the RF and surround filter outputs are dependent is large (i.e. $p(\xi = \xi_1 | \mathbf{k}, \mathbf{s}) \approx 1$ for a homogeneous image), the estimate of the local component, $\overline{\kappa_{\theta, \omega}}$, is proportional

to the ratio between the observed response of the RF filter ($k_{\theta,\omega}$) and the normalization signal computed from the responses of all other filters ($\lambda_{(k\sigma)}$, a generalized measure of center-surround energy), as shown in (Eq. S4). That is, the estimate involves divisive normalization from the surround. The generalized measure of center and surround contrast energy reduces to a classical description of normalization (Heeger, *Vis Neurosci* 9, 181-197, 1992) when the covariance matrix is diagonal (i.e. when the dependency between k and s is entirely explained by the shared mixer variable, with no linear correlations between their outputs). Critically, when the probability is small (i.e. $p(\xi = \xi_1 | \mathbf{k}, \mathbf{s}) \approx 0$ for a heterogeneous image) the estimate does not involve surround normalization. Hence the prediction is that surround suppression is muted for heterogeneous images.

Training the Bayesian model

In addition to the properties of the center and surround filters, the Bayesian model (MGSM) uses several parameters that are learned from natural images—namely, the prior probability that the outputs are dependent $p(\xi = \xi_1)$, and the covariance matrices for the two mixture components $\mathbf{C}_{(k\sigma)}$, $\mathbf{C}_{(k)}$, $\mathbf{C}_{(\sigma)}$. The values of these parameters determine whether a particular image will be classified as homogeneous or heterogeneous, according to (Eq. S6).

For each set of filters, we trained the Bayesian model on image patches randomly selected from a database commonly used for image compression benchmarks (<http://neuroscience.aecom.yu.edu/labs/schwartzlab/code/standard.zip>). We computed the outputs of the visual filters to the patches, and searched for the parameter values that maximized the likelihood of the filters' outputs, namely $p(\mathbf{k}, \mathbf{s} | \text{parameters})$. To find the maximum likelihood parameters we used an iterative procedure based on the Expectation Maximization (EM) algorithm, as it is standard for mixture models (Coen-Cagli et al., *Advances in NIPS* 369–377, 2009; Meng and Rubin, *Biometrika* 80, 267–278, 1993). The training was unsupervised: we did not pre-specify which patches in the training set were homogeneous or heterogeneous, but rather let the model infer them. Once we had defined the optimal parameters with the standard database, we froze them before classifying the experimental images.

2. Image homogeneity and center-surround similarity

We defined image homogeneity formally, as the result of inference in an MGSM model of natural images (Eq. S6). Intuitively, the inference reflects the statistical similarity of signals in the RF and surround. Here we specify this intuition further, in order to clarify how image homogeneity defines a measure of center-surround similarity and what sort of structure in images leads to their being classified as homogeneous.

We study the properties of the ratio between the probability that an image with filters outputs (\mathbf{k}, \mathbf{s}) (recall that \mathbf{k} represents the outputs of a set of filters in the RF center with different orientations, and \mathbf{s} a set of filters with fixed orientation and different positions in the surround, as explained in the **Methods**) is homogeneous vs. heterogeneous:

$$(Eq. S7) \quad M \doteq \frac{p(\xi = \xi_1 | \mathbf{k}, \mathbf{s})}{p(\xi = \xi_0 | \mathbf{k}, \mathbf{s})} = \frac{p(\xi = \xi_1)}{p(\xi = \xi_0)} \cdot \frac{\sqrt{(2\pi)^{n_k} \det(\mathbf{C}_{(\kappa)})} \sqrt{(2\pi)^{n_s} \det(\mathbf{C}_{(\sigma)})}}{\sqrt{(2\pi)^n \det(\mathbf{C}_{(\kappa\sigma)})}}$$

$$\cdot \frac{\mathcal{B}\left(1 - \frac{n}{2}; \lambda_{(\kappa\sigma)}\right) \lambda_{(\kappa)}^{\left(\frac{n_k-1}{2}\right)} \lambda_{(\sigma)}^{\left(\frac{n_s-1}{2}\right)}}{\mathcal{B}\left(1 - \frac{n_k}{2}; \lambda_{(\kappa)}\right) \mathcal{B}\left(1 - \frac{n_s}{2}; \lambda_{(\sigma)}\right) \lambda_{(\kappa\sigma)}^{\left(\frac{n-1}{2}\right)}}$$

We would like to understand which patterns of filters outputs lead to higher ratios, corresponding to higher inferred probability that the image is homogeneous.

A heuristic approximation to the inference of homogeneity, for a simplified case

We first consider the following special case, in which there are no correlations between filters (i.e., diagonal covariance matrices) and homogeneous and heterogeneous images are equally probable *a priori*. Specifically:

$$\left\{ \begin{array}{l} p(\xi = \xi_1) = p(\xi = \xi_0) = 0.5 \\ \mathbf{C}_{(\kappa\sigma)} = \mathbf{I}_n \\ \mathbf{C}_{(\kappa)} = \mathbf{I}_{n_k} \\ \mathbf{C}_{(\sigma)} = \mathbf{I}_{n_s} \\ n_k = n_s = 2m \\ n = 4m \end{array} \right.$$

from which it follows (by substitution in (Eq. S4) and (Eq. S5)) that:

$$\left\{ \begin{aligned} \lambda_{(\kappa)} &= \sqrt{\sum_{i=1}^{2m} \mathbf{k}_i^2} \doteq \sqrt{P_k} \\ \lambda_{(\sigma)} &= \sqrt{\sum_{i=1}^{2m} \mathbf{s}_i^2} \doteq \sqrt{P_s} \\ \lambda_{(\kappa\sigma)} &= \sqrt{\sum_{i=1}^{2m} \mathbf{k}_i^2 + \sum_{i=1}^{2m} \mathbf{s}_i^2} \doteq \sqrt{P_k + P_s} \end{aligned} \right.$$

This simplifies (Eq. S7) greatly, to:

$$(Eq. S8) \quad M = \frac{\mathcal{B}(2m-1; \sqrt{P_k + P_s})}{\mathcal{B}(m-1; \sqrt{P_k}) \mathcal{B}(m-1; \sqrt{P_s})} \cdot \frac{(P_k \cdot P_s)^{\frac{m-1}{2}}}{(P_k + P_s)^{\frac{2m-1}{2}}}$$

We can now verify if the intuition is correct that when the drive to the RF and surround are similar, images are inferred homogeneous (i.e. M takes on large values). In this case $P_k = P_s \doteq P$ and therefore

$$(Eq. S9) \quad M = \frac{\mathcal{B}(2m-1; \sqrt{2P})}{\mathcal{B}(m-1; \sqrt{P})^2} \cdot \frac{1}{(2)^{\frac{2m-1}{2}}} \cdot \frac{1}{P^{\frac{1}{2}}}$$

When the inputs are weak, i.e. $P \rightarrow 0$, we can use the known limiting form of the Bessel function for small arguments (Abramowitz and Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables (1972), p.375):

$$(Eq. S10) \quad \mathcal{B}(n; x) \approx \frac{1}{2} \Gamma(n) \left(\frac{2}{x} \right)^n$$

and therefore

$$(Eq. S11) \quad M \approx 2^{\frac{7}{2}} \cdot \frac{\Gamma(2m-1)}{\Gamma(m-1)^2} \cdot \frac{1}{P}$$

thus showing that M diverges for weak and similar RF and surround drive. Similarly, for strong inputs that are similar in the RF and surround, i.e. $P \gg 1$, we can take the first order asymptotic expansion of the Bessel function (Abramowitz and Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables (1972), p.377), and find that:

$$(Eq. S12) \quad M \approx \frac{1}{\pi 2^{m-2}} \cdot e^{\sqrt{2P}}$$

again showing that M diverges for large and similar center and surround energies.

We can also verify the intuition that when the energies in the center and surround are different, images are inferred heterogeneous (i.e. M takes on small values). We consider first the case that $P_s \gg P_k$, or equivalently $P_k \rightarrow 0, P_s \rightarrow \infty$. Under this circumstance we find:

$$(Eq. S13) \quad M \approx \frac{\mathcal{B}(2m-1; \sqrt{P_s})}{\mathcal{B}(m-1; \sqrt{P_k}) \mathcal{B}(m-1; \sqrt{P_s})} \cdot \frac{(P_k \cdot P_s)^{\frac{m-1}{2}}}{(P_s)^{\frac{2m-1}{2}}}$$

and using the limiting form and asymptotic expansion of the Bessels, for $P_k \rightarrow 0, P_s \rightarrow \infty$ respectively, we find:

$$(Eq. S14) \quad M \approx \frac{(P_k)^{m-1}}{(P_s)^{\frac{m}{2}}}$$

This ratio vanishes as the numerator goes to zero and denominator to infinity, thus confirming that images with strong surround and weak center energy are inferred heterogeneous. A similar result holds for strong RF input paired with weak drive to the surround.

The behavior for intermediate values of P_k, P_s cannot be described as easily analytically. However, we can compute M exactly via (Eq. S8) for a range of values of P_k, P_s commonly observed on natural images. In **Supplementary Figure 4a** we assumed $m = 8$ (i.e. 16 filters, or 8 quadrature pairs, in the center and 16 in the surround), and plot the values of $M/(M + 1)$, which is identical to

$p(\xi = \xi_1 | \mathbf{k}, \mathbf{s})$ —the posterior probability that the image is homogeneous. As expected, the largest values are found closer to the diagonal, when P_k, P_s are similar. This is true not only in the asymptotic regimes but also for intermediate energies. However, to achieve a high probability of inferred homogeneity, there is a more stringent requirement that P_k, P_s are similar at low values of P_k, P_s than at high values. Thus, the requirement of similar drive to the center and surround has a strong intensity-dependence. Another way to think about this result is in terms of the so called Bayesian Occam's razor: when the evidence is weak (e.g. a low contrast image, corresponding to low values of P_k, P_s), stick to the simpler model (the center and surround signals are independent) rather than the more complex one (center and surround signals are dependent).

We then searched for a simpler, intuitive similarity measure that could approximate the behavior of the full inference. We found that the measure $S = \frac{(P_k \cdot P_s)^\alpha}{(P_k + P_s)^\beta}$, based on the rightmost term in (Eq. S8), did reasonably well: it increased both when the energies were more similar and when they were larger

(i.e. when the product dominates over the sum). Importantly, simpler measures, based on either the difference between P_k and P_s , or their log-ratio, could not reproduce the characteristic intensity-dependence of the full inference. Thus, simpler measures of similarity—such as those suggested by previous physiological studies—fail to capture the behavior of the inference of homogeneity.

Supplementary Figure 4b illustrates the probability corresponding to the measure S (namely $S/(S+1)$), which is analogous to $M/(M+1)$ plotted in **Supplementary Figure 4a**. We searched manually for values of the exponents in S , and found that the qualitative behavior of $p(\xi = \xi_1 | \mathbf{k}, \mathbf{s})$ could be approximated reasonably well for a large range of values, provided that $\beta \approx 1.7\alpha$.

In summary, this analysis confirms the intuition that MGSM inference about homogeneity defines a measure of center-surround similarity. Such similarity is input strength-dependent, and can be approximated as a product-over-sum ratio with different exponents.

Inference with realistic covariance structure

So far we have considered the case where the covariances are identity matrices. However, in the MGSM the covariances are optimized to capture structure in natural images, and in practice they are far from diagonal (Coen-Cagli et al. PLoS Comput Biol 8, e1002405 (2012)). How do non-diagonal covariances affect the inference? As shown in (Eq. S4) and (Eq. S5), the entries of the inverse covariances act as weights in the sum of filter energy terms, therefore determining the values of $\lambda_{(\kappa)}, \lambda_{(\sigma)}, \lambda_{(\kappa\sigma)}$. As shown in (Eq. S7), the values of these λ terms directly influence the inference of homogeneity: when $\lambda_{(\kappa\sigma)}$ is small and the product of $\lambda_{(\kappa)}, \lambda_{(\sigma)}$ is large, the image is more likely to be homogeneous; and conversely when $\lambda_{(\kappa\sigma)}$ is large and the product of $\lambda_{(\kappa)}, \lambda_{(\sigma)}$ is small.

To gain some intuition, consider $\lambda_{(\kappa\sigma)}$. Its value is small when the term $(\mathbf{k}, \mathbf{s}) \mathbf{C}_{(\kappa\sigma)} (\mathbf{k}, \mathbf{s})^\top$ is large (and therefore $(\mathbf{k}, \mathbf{s}) \mathbf{C}_{(\kappa\sigma)}^{-1} (\mathbf{k}, \mathbf{s})^\top$ is small). The term $(\mathbf{k}, \mathbf{s}) \mathbf{C}_{(\kappa\sigma)} (\mathbf{k}, \mathbf{s})^\top$ is large when (\mathbf{k}, \mathbf{s}) is a pattern that matches structure in the covariance. For instance, for natural images the variance and covariance of surround filters collinear with the RF (i.e. similarly oriented, and placed at the ends of the RF) are much larger than those of other surround filters (e.g. similarly oriented, but placed at the sides of the RF; Coen-Cagli et al. PLoS Comput Biol 8, e1002405 (2012)), due to the predominance of long lines and edges in scenes. Therefore, the image of a long stripe passing through the RF and collinear surround filters will produce a large value of $(\mathbf{k}, \mathbf{s}) \mathbf{C}_{(\kappa\sigma)} (\mathbf{k}, \mathbf{s})^\top$, and therefore a small value of $\lambda_{(\kappa\sigma)}$, leading to a higher probability of inferred homogeneity. An image containing similarly oriented structure in the RF

but no collinearity, will lead to a smaller value of $(\mathbf{k}, \mathbf{s}) \mathbf{C}_{(\kappa\sigma)} (\mathbf{k}, \mathbf{s})^\top$, and therefore a lower probability of inferred homogeneity.

More precisely, let us consider the eigen-decomposition of the covariance $\mathbf{C}_{(\kappa\sigma)} = \sum_i h_i (\mathbf{v}_i \cdot \mathbf{v}_i^\top)$ and its inverse $\mathbf{C}_{(\kappa\sigma)}^{-1} = \sum_i h_i^{-1} (\mathbf{v}_i \cdot \mathbf{v}_i^\top)^{-1}$, where \mathbf{v}_i, h_i are the orthonormal eigenvectors of the covariance and corresponding eigenvalues, respectively. If the input pattern (\mathbf{k}, \mathbf{s}) is an eigenvector \mathbf{v}_i with large eigenvalue h_i (i.e., it matches the dominant structure in the covariance), then $\lambda_{(\kappa\sigma)}$ amounts to $\sqrt{1/h_i}$, a small number. This leads to a higher probability of inferred homogeneity. Conversely, if the input pattern (\mathbf{k}, \mathbf{s}) is an eigenvector with small eigenvalue, then $\lambda_{(\kappa\sigma)}$ is large, leading to a lower probability of inferred homogeneity.

In summary, visual inputs that resemble the dominant structure (the largest eigenvectors) in $\mathbf{C}_{(\kappa\sigma)}$, and therefore differ from the dominant structure in $\mathbf{C}_{(\kappa)}, \mathbf{C}_{(\sigma)}$, are assigned high probability of being homogeneous. Therefore the full inference about homogeneity in the MGSM amounts to a generalized measure of center-surround similarity, which is dependent on the overall oriented energy in the image, and which uses a metric defined by the structure of the covariance matrices learned from natural images.