Stimulus content shapes cortical response statistics
– Supplemental Experimental Procedures –

Mihály Bányai, Andreea Lazar, Liane Klein, Johanna Klon-Lipok, Wolf Singer, Gergő Orbán

Contents

List of Figures

1 Probabilistic inference in a hierarchical generative model

To obtain an insight into the way bottom-up and top-down information is integrated during inference in a hierarchical generative model, we derive how the posterior distribution of a mid-level latent variable can be obtained. The general approach to specify the internal mental model of an animal is to define a probabilistic generative model, the parameters of which are fitted to a database natural images. Then the statistics learned over latent model variables can be used to predict neuronal responses in populations that are thought to represent the given variables. This approach has proven lucrative to predict mean and variance of responses in V1 (Olshausen & Field, 1996; Schwartz & Simoncelli, 2001; Orbán et al., 2016) with a single layer of latent variables. Introducing stimulus-dependence to V1 correlations necessitate a hierarchical model. The specific form of predicted inter-neuron correlations strongly depends on the structure of the internal model being implemented in the brain. This poses a serious challenge, since our knowledge is limited as to the precise form of such a hierarchical generative model. Therefore we aim to derive predictions that are invariant over the structure of a wide class of hierarchical generative models.

We define a hierarchical generative model in which the stimulus $\mathbf{x}$ is a result of a hierarchical stochastic process (Fig. 1A). High-level complex features, $\mathbf{z}$, define how low-level features, $\mathbf{y}$, are combined (Fig. 1A). These combinations define how likely a particular low-level feature is given a set of high-level latent activations, $\mathrm{P}(\mathbf{y} \mid \mathbf{z})$. Activations of low-level features, $\mathbf{y}$, determine the combinations of pixel intensities (Fig. 1B) and define how likely a particular image is if we assume that a particular setting of low-level features underlies the image, $\mathrm{P}(\mathbf{x} \mid \mathbf{y})$.

Once the prior probability of high-level activations, $\mathrm{P}(\mathbf{z})$, is defined, the model can be specified by formulating the joint distribution of the model variables:

$$\mathrm{P}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathrm{P}(\mathbf{x} \mid \mathbf{y})\, \mathrm{P}(\mathbf{y} \mid \mathbf{z})\, \mathrm{P}(\mathbf{z}) \tag{S1}$$

Inference can be performed in this model by establishing the posterior probability for different variables. We are interested in the posterior of the low-level variable, $\mathbf{y}$, given that we observe the stimulus $\mathbf{x}$. Since we cannot experimentally access the activities of the variables at the topmost layer, the relevant posterior is marginalized over them: $\mathrm{P}(\mathbf{y} \mid \mathbf{x})$. This can be expanded based on the dependencies of the generative model and marginalizing over $\mathbf{z}$:

$$\mathrm{P}(\mathbf{y} \mid \mathbf{x}) = \frac{\mathrm{P}(\mathbf{x} \mid \mathbf{y})\, \mathrm{P}(\mathbf{y})}{\mathrm{P}(\mathbf{x})} = \frac{\int \mathrm{P}(\mathbf{x} \mid \mathbf{y})\, \mathrm{P}(\mathbf{y} \mid \mathbf{z})\, \mathrm{P}(\mathbf{z})\, \mathrm{d}\mathbf{z}}{\mathrm{P}(\mathbf{x})} \tag{S2}$$

This form demonstrates how the prior of $\mathbf{y}$ is dependent on high-level variables $\mathbf{z}$, thus we see that intermediate inferences depend not only on the stimulus $\mathbf{x}$ but also on higher level inferences. However, the integral over the prior of the topmost variable, $\mathbf{z}$, makes this relationship obscure, therefore we intend to obtain a more transparent form.

We aim for an explicit expression on how $\mathbf{y}$ depends on observations, $\mathbf{x}$, and high-level inferences $\mathbf{z}$. In other words, we want to formulate the distribution over $\mathbf{y}$ conditioned on both the parent and children variables:

$$\mathrm{P}(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) = \frac{\mathrm{P}(\mathbf{x} \mid \mathbf{y})\, \mathrm{P}(\mathbf{y} \mid \mathbf{z})\, \mathrm{P}(\mathbf{z})}{\mathrm{P}(\mathbf{x}, \mathbf{z})} = \frac{\mathrm{P}(\mathbf{x} \mid \mathbf{y})\, \mathrm{P}(\mathbf{y} \mid \mathbf{z})}{\mathrm{P}(\mathbf{x} \mid \mathbf{z})} \tag{S3}$$

Rearranging Eq. S3 and substituting it into Eq. S2 we obtain

$$\begin{aligned} \mathrm{P}(\mathbf{y}\,|\,\mathbf{x}) &= \frac{\int \mathrm{P}(\mathbf{y}\,|\,\mathbf{x},\mathbf{z})\,\mathrm{P}(\mathbf{x}\,|\,\mathbf{z})\,\mathrm{P}(\mathbf{z})\,\mathrm{d}\mathbf{z}}{\mathrm{P}(\mathbf{x})} = \\ &= \int \mathrm{P}(\mathbf{y}\,|\,\mathbf{x},\mathbf{z})\,\mathrm{P}(\mathbf{z}\,|\,\mathbf{x})\,\mathrm{d}\mathbf{z} \end{aligned} \tag{S4}$$

This is the exactly the form of posterior we used to predict stimulus-dependent correlation patterns.

Further insgiht can be obtained by assuming a specific simple model structure, namely that the high-level variable, $\mathbf{z}$, is discrete and that the distribution of high-level features is described by a categorical distribution. In such a case the integral in Eq. S4 is simplified to a sum:

$$\mathrm{P}(\mathbf{y}\,|\,\mathbf{x}) = \sum_i \mathrm{P}(\mathbf{y}\,|\,\mathbf{x},z_i)\,\mathrm{P}(z_i\,|\,\mathbf{x})\,\mathrm{d}\mathbf{z}, \tag{S5}$$

which is a weighted sum over possible configurations of $\mathbf{z}$ (Fig. 1).

In the visual cortex, activity of neurons at different levels of hierarchy (e.g. V1 and V2) represent different levels of latent variables of an internal model. Features these variables are activated by determine the selectivity of the neurons. During perception inferences are assumed to be made in this hierarchical internal model, which determines how the mean, variance and correlations change when the stimulus changes. Critically, the term $\mathrm{P}(\mathbf{y}\,|\,\mathbf{x},z_i)$ is a product of two terms conveying bottom-up and top-down information-respectively:

$$\mathrm{P}(\mathbf{y}\,|\,\mathbf{x},z_i) = \mathrm{P}(\mathbf{y}\,|\,\mathbf{x})\,\mathrm{P}(\mathbf{y}\,|\,z_i)\,. \tag{S6}$$

Both of the terms contribute to setting the mean activations of the low-level neurons, $\mathbf{y}$. The correlation between a pair of low level neurons, y_1 and y_2 does not depend on the stimulus identitiy in a large class of generative models, thus the first term does not contribute to stimulus-dependent correlations. The second term, however, determines how the two low-level variables covary in the context of one value of the high-level variable or the other. If high-level variables are identified with texture-sensitive neurons in V2 then it becomes clear that different textures imply different covariations between low-level feuatures represented in the lower layer, V1. Thus, such a scheme predicts that correlations will be stimulus-dependent if different stimuli induce different configurations of high-level variables being active.

References

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609.

Orbán, G., Berkes, P., Fiser, J., & Lengyel, M. (2016). Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron*, *92*(2), 530–543.

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience*, *4*(8), 819–825.

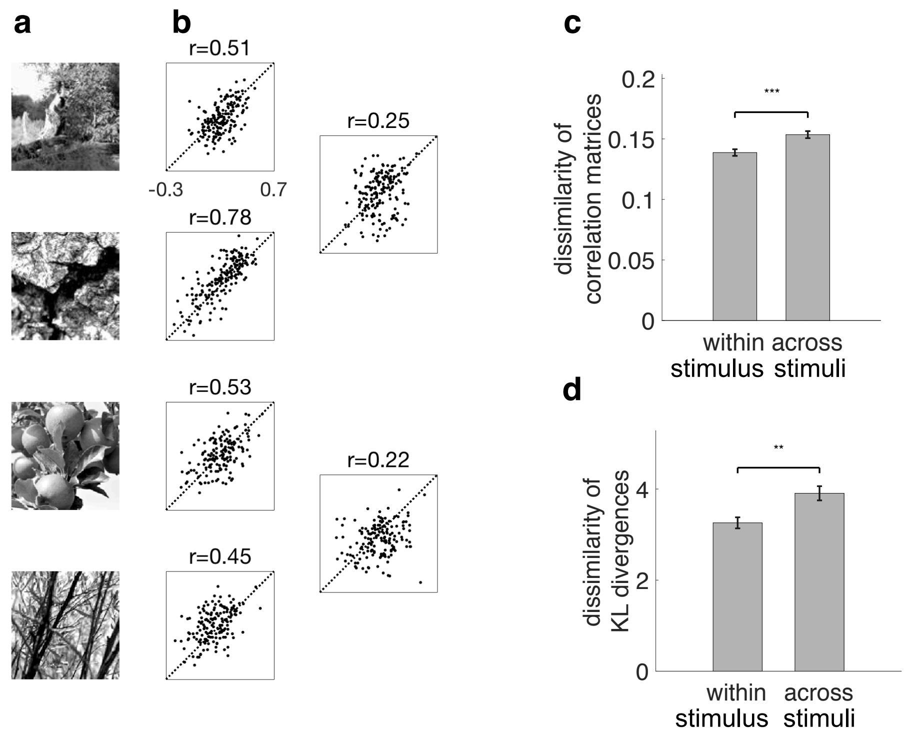

Figure S1. Related to Fig. 3. Stimulus-specificity of response correlation patterns, alternative measures. (**A**) Natural images used in the experiment. (**B**) Pairwise spike count correlation values compared for spiking activity recorded under two conditions. Left: Trials recorded for one particular image (shown on panel **A**) were randomly assigned to two subsets, spike count correlations were calculated for both of the subsets and these values were plotted against each other. Right: Same as left but the subset of trials belonged to responses evoked by different stimuli. Numbers show the Pearson correlation between spike count correlation values obtained for the two conditions. (**C**) Same as Fig. 3F but the comparison is made on a stimulus-by-stimulus basis rather than on a pair-by-pair basis. This comparison yields the same degrees of freedom for the test that we obtain when comparing mean correlations on Fig. 3E. (**D**) Dissimilarity of correlation matrices measured by Kullback-Leibler (KL) divergence.

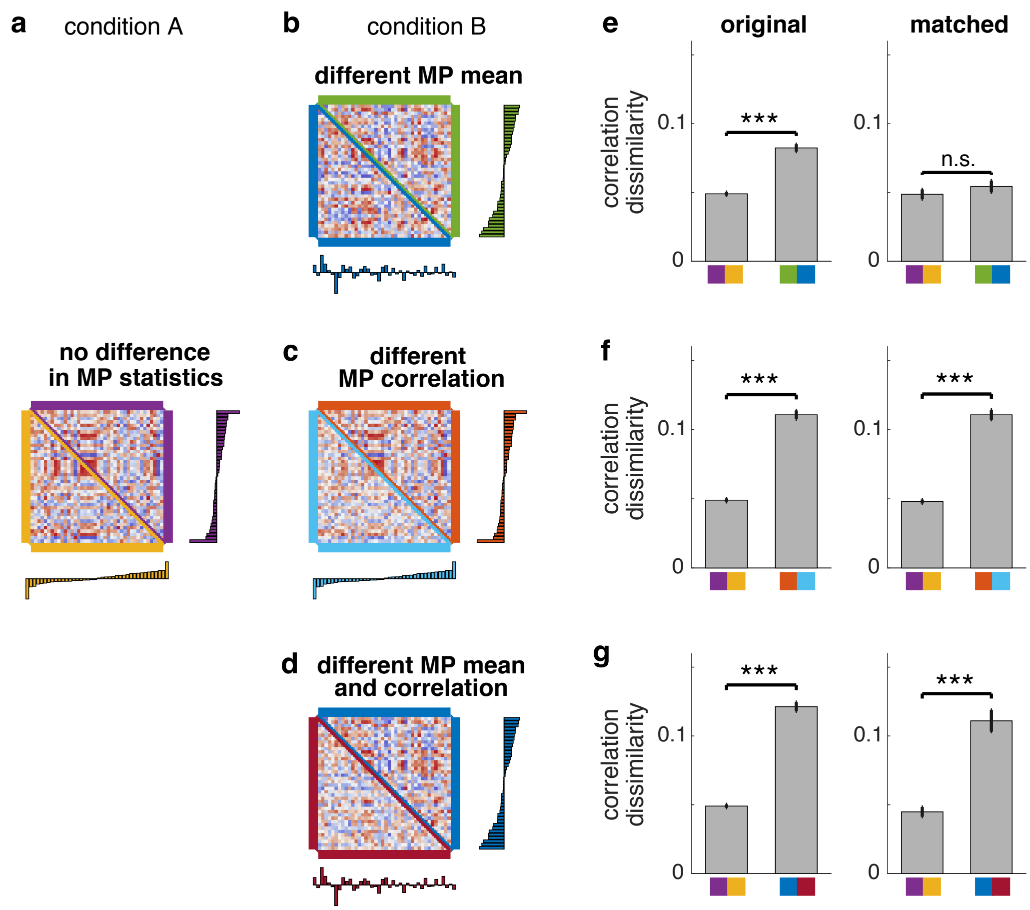

Figure S2 ***(continued on following page)*****.**

Figure S2 ***(preceding page).*** **Related to Fig. 4.** Testing firing rate difference-matching on synthetic data. (**A**) Using a network of neurons in which the activity statistics of membrane potentials was defined through diverse mean membrane potentials and a covariance matrix, we tested dissimilarity of firing rates and correlations under different conditions. Upper and lower triangles of correlation matrices show the correlation structure of membrane potentials in response to different stimuli (identified by different color frames around the triangles). Horizontal and vertical bar graphs show mean activation levels of model neurons. Without loss of generality, the population of neurons is ordered according to the mean activation levels in response to one of the stimuli. In condition A the two stimuli elicit responses with the same mean activity and same correlation structure. (**B–D**, In condition B three scenarios are demonstrated. First, membrane potential means differ while keeping the correlation structure intact (**B**). Second, the membrane potential means are identical but the correlation structure changes (**C**). Third, both mean membrane potential and membrane potential correlations are different across stimuli (**D**). (**E–G**) Correlation dissimilarity of spike count correlation distributions measured in the three conditions of **B–D**, when firing rate changes are not controlled (left column) and when these are matched across conditions (right column). Changes in mean activation levels alone result in apparent differences in the spike count correlation structure (e, left panel) and similar changes in spike count correlation dissimilarity are present in the other two conditions (**F**, **G**, left panels). Matching of firing rate differences eliminates the difference in correlation dissimilarity caused by mean activation differences (**E**, right panel) but not that caused by differences in the correlation structure (**F**, **G**, right panels).

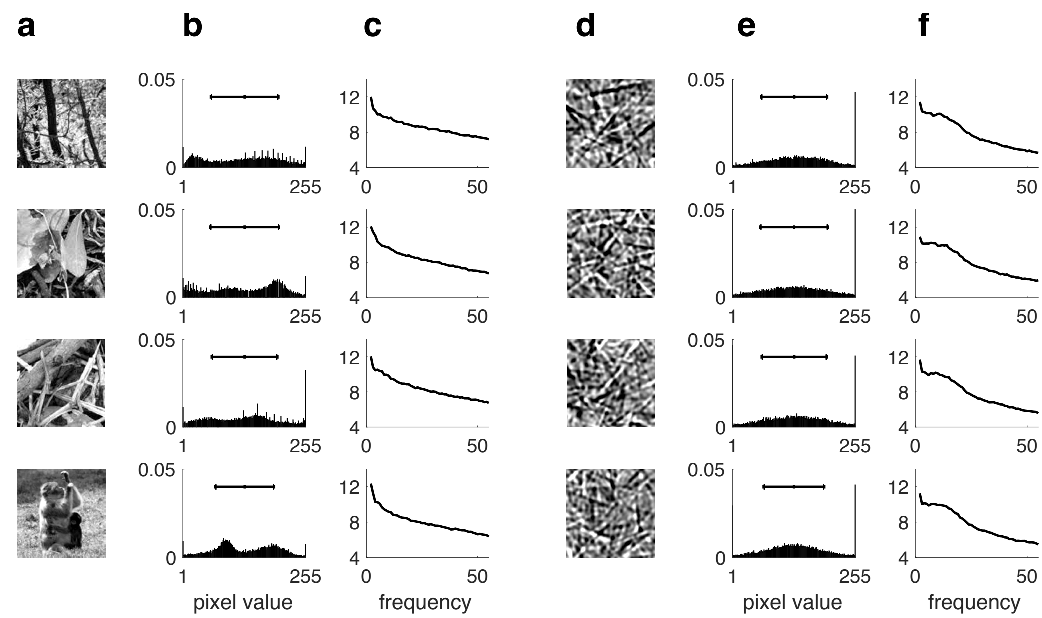

Figure S3. Related to Fig. 5. Stimulus statistics. (**A**) Natural stimuli used in one of the recording sessions. (**B**) Distribution of pixel intensities in the stimuli, error bars indicating mean and standard deviation. (**C**) Spatial frequency spectrum of the stimuli. (**D**) Synthetic stimuli used in the same recording session. (**E–F**), the measures presented in (**B** and **C**) calculated for the synthetic stimuli.