Mechanisms underlying reshuffling of visual responses by optogenetic stimulation in mice and monkeys

A. Sanzeni[1,2], A. Palmigiano[1], T.H. Nguyen[1], J. Luo[3], J.J. Nassi[3], J.H. Reynolds[3], M.H. Histed[4], K.D. Miller[1], and N. Brunel[2,5]

[1]Center for Theoretical Neuroscience, Columbia University, New York, NY, USA
[2]Department of Neurobiology, Duke University, Durham, NC, USA
[3]Systems Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, USA
[4]National Institute of Mental Health Intramural Program, NIH, Bethesda, MD, USA
[5]Department of Physics, Duke University, Durham, NC, USA

July 14, 2022

Abstract

The ability to observe the response of neural circuits to controlled optogenetic perturbations opens an unprecedented window into the mechanisms governing network dynamics and computations. We combined an analysis of neuronal responses to visual and optogenetic inputs in mice and monkey V1 with theoretical modelling. In both species, we found that optogenetic stimulation of excitatory neurons strongly modulated the activity of single neurons, but had only a weak effect on the distribution of firing rates across the population. Key statistics of responses of mice and monkeys lay on a continuum, with mice/monkeys occupying the low/high rate regions, respectively. At high contrast of the visual stimulus, optogenetic inputs did not significantly affect the distribution of firing rates, i.e. they reshuffled firing rates across the network. We show that neuronal reshuffling emerges generically in randomly connected networks of excitatory and inhibitory neurons, as long as the combination of recurrent coupling and feedforward input is sufficiently strong so that a powerful inhibitory feedback cancels the mean increase in firing rate due to optogenetic input.

Introduction

Since their introduction about two decades ago, optogenetic methods have revolutionized neuroscience (Kim et al., 2017). Optogenetic inhibition of specific areas have allowed neuroscientists to identify regions that are necessary for specific behaviors. Optical stimulation of specific cell types, in conjunction with electrophysiological or optical recordings of neuronal activity, can also be used to experimentally test predictions from network models. An example of such a prediction is the ‘paradoxical effect’ - activating all inhibitory neurons in a network of excitatory (E) and inhibitory (I) neurons leads to a paradoxical decrease in activity, in inhibition-stabilized networks (Ozeki et al., 2009; Tsodyks and Markram, 1997). This prediction was recently tested in multiple mouse cortical areas (V1, S1 and M1) (Sanzeni et al., 2020). The results were consistent with an inhibition-stabilized cortex. Optogenetics can also provide information about the structure of synaptic connectivity. For instance, activation of single neurons in V1 was shown to lead to a suppression of other neurons, whose magnitude depends on the distance between their preferred orientations, consistent with a network with strong orientation specificity of E-I connectivity (Chettih and Harvey, 2019; Sadeh and Clopath, 2020).

Understanding the mechanisms that shape the response of neural circuits to their inputs is another question where optogenetic techniques can provide an invaluable tool, since they allow experimentalists to target specific cell types in a controlled fashion. To obtain deeper insights into such mechanisms, we analyzed electrophysiological recordings from both mouse (Histed, 2018) and monkey (Nassi et al., 2015) V1, during presentation of visual stimuli of various contrasts and/or optogenetic stimulation of E neurons. Optogenetic stimulation of E neurons leads to a broad distribution of firing rate changes in single neurons, with some neurons being strongly excited, while others are strongly suppressed. In mice, these changes lead to an increase of the mean response which is approximately independent of contrast. In monkeys, we find that, although at low contrast the network responds similarly as in mice, at high contrast, approximately equal proportions of neurons are excited and suppressed, and the average firing rate is not significantly affected by optogenetic stimulation. In fact, the entire distribution of firing rates is unaffected by optogenetic stimulation, even though the variance of firing rate changes is comparable to the variance of rates prior to optogenetic stimulation. Thus, optogenetic stimulation of E neurons leads to a reshuffling of firing rates across the network, leaving invariant the whole distribution. We also find a continuum from mice at low contrast to mice at high contrast to monkeys at low contrast to monkeys at high contrast, along which firing rates increase and the optogenetically-induced changes in the mean and distribution of firing rates shrink to zero.

We explored the mechanisms of such a reshuffling using a theoretical analysis of randomly connected networks of E and I neurons. We find that such networks reproduce qualitatively all the above mentioned features, provided the strengths of recurrent coupling and feedforward input are strong enough, the network is inhibition dominated, and opsin-induced currents in E neurons are sufficiently heterogeneous and weak. The continuity from mice to monkeys can be understood if mice are in a more weakly-coupled regime than monkeys, which can arise from weaker recurrent weights and/or weaker feedforward inputs. These results provide new insights into how cortex processes inputs, and on similarities and differences of cortical information processing in different mammalian species.

Results

Heterogeneous optogenetic modulation of visual responses in mice and monkeys

We analyzed recordings obtained in the visual cortex (area V1) of mice (Histed, 2018) and monkeys (Nassi et al., 2015) (experimental procedures are summarized in the *Methods*). These experiments shared a number of common features: visual stimuli were gratings of different contrast ($[0-90]\%$ in mice and $[0-99]\%$ in monkeys); virally expressed opsins were used to optogenetically stimulate pyramidal cells. Both datasets consist of neural responses recorded extracellularly in awake animals to combinations of visual and optogenetic stimuli. Neurons in monkeys were studied with stimuli matched to their stimulus preferences, while in mice stimuli were not matched to specific neurons. There are also a few differences between the recordings in (Histed, 2018) and (Nassi et al., 2015). In monkeys, visual stimuli were drifting sinusoidal gratings centered within the receptive field and matched to each recorded unit's preferred size, orientation, and spatial and temporal frequencies. In mice, on the other hand, multiple neurons were recorded simultaneously and there was no systematic relationship between the visual stimulus and preferences of recorded neurons.

Both in mice and monkeys, as previously observed (Histed, 2018; Nassi et al., 2015), single neuron responses to visual stimuli were strongly modulated by optogenetic stimuli, in a highly heterogeneous

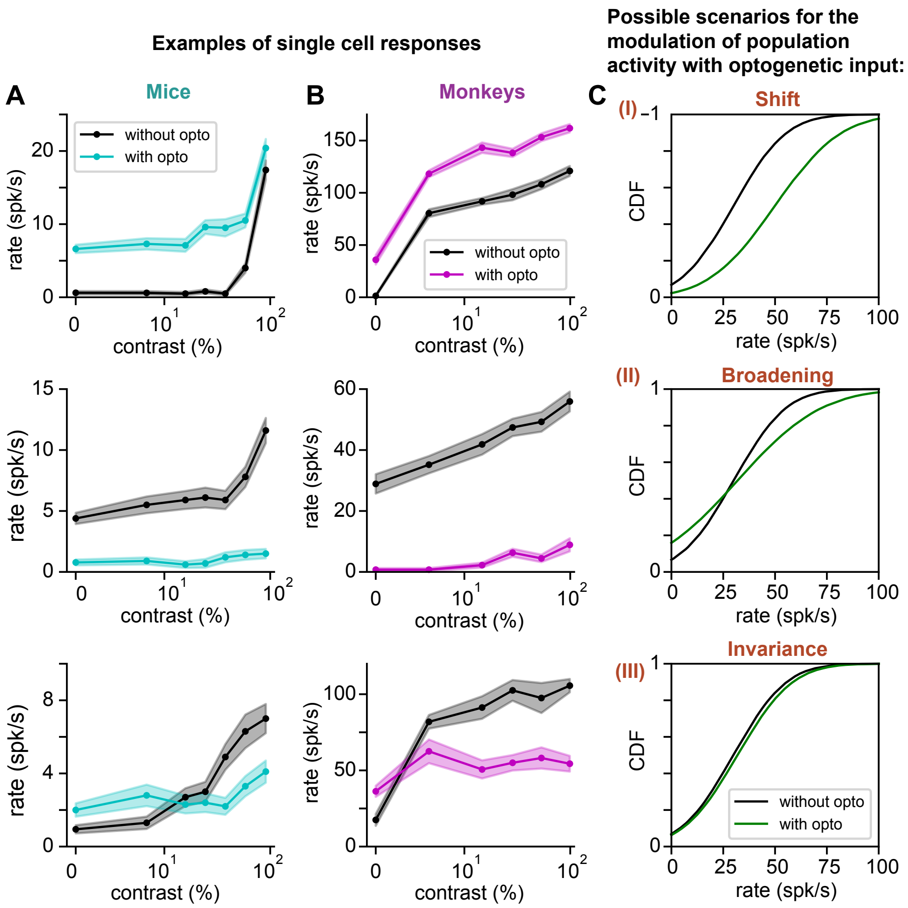

Figure 1: **Optogenetic modulation of visual responses in mice and monkeys: single neuron examples and potential scenarios at the population level.** (**A**) Examples of single neuron responses to visual stimuli in mice, without (black) and with (colored) optogenetic stimulation of E neurons. (**B**) As in **A** but in monkeys. In both species, some neurons strongly increase their rates at all contrasts (top row); others are strongly suppressed (center row); and others exhibit rate increase at low contrast, but suppression at high contrast (bottom row). Note that the same color code is used in all the following figures. (**C**) Possible scenarios of how optogenetic stimulation of E neurons might modify the distribution of firing rates in the network (see text).

fashion (Fig. 1A-B). Specifically, responses to visual stimuli were typically either enhanced by the laser at all contrasts, suppressed by the laser at all contrasts, or enhanced at low contrast and suppressed at high contrast (Fig. 1A-B); the fourth possibility, suppressed at low contrast and enhanced at high contrast, was not observed. These single neuron responses might underlie different scenarios at the population level (Fig. 1C). (i) Optogenetic stimulation could significantly increase the mean activity in the network, while broadening the distribution of single neuron rate. This scenario is what one would expect naively in experiments in which optogenetic stimulation targets E neurons. (ii) Optogenetic stimulation could broaden the distribution of rates, but a large number of suppressed cells could lead to a weak mean increase of the population activity. (iii) Optogenetic stimulation could affect only weakly the distribution of rates. Which of these scenarios better represents the effect of optogenetic stimulation of population activity? Do these effects change with contrast and species?

Optogenetic stimulation reshuffles visual responses in mice and monkeys

To answer the above mentioned questions, we analyzed the statistics of population responses to visual and optogenetic stimuli (Figs. 2). In both mice and monkeys, mean population activity increased monotonically with the contrast of visual stimuli, but firing rates in mice were significantly lower than in monkeys. The mean visual response in mice ranged from 6 ± 0.9 spk/s (mean±sem) at 0% contrast to 16.7 ± 2.0 spk/s at high contrast, while in monkeys it ranged from 20.2 ± 2.0 spk/s to 72.7 ± 6.7 spk/s (Fig. 2A). In both species, visual stimuli of fixed contrast generated a broad distribution of rates (Fig. 2B), and increasing contrast significantly shifted distributions toward higher rates (Fig. 2C–D). Surprisingly, optogenetic stimulation had much weaker effects at the population level (Fig. 2A–D). In mice, the mean and standard deviation of the population activity were significantly modulated by the optogenetic stimulation only for contrasts lower than 55% (Welch's t–test, p–value<0.05) and 90%(f–test, p–value>0.05), respectively. In monkeys, the mean population activity was not significantly modulated at any contrast (Welch's t–test, p–value>0.05) while the modulation of the standard deviation was significant only at contrasts lower than 12% (f–test, p–value>0.05). Even more strikingly, optogenetic stimuli weakly affected the whole distribution of rates in the network (Fig. 2D). In mice, the distribution of rates was significantly affected only at low contrast ($<$ 20%, KS test, p–value<0.05), for which optogenetic stimuli induced a weak but statistically significant shift of the distribution toward higher rates. In monkeys, the distribution of rates was not significantly affected by optogenetic stimulation at any contrast tested (KS test, p–value<0.05). The weak optogenetic modulation observed at the population level was not due to small changes in single neuron firing. In fact, both in mice and monkeys, distributions of optogenetic modulations of single neuron visual responses were broad (Fig. 2E–F). Furthermore, unlike what was found in response to visual stimuli (Fig. 2C–D), the distribution of changes in rate due to optogenetic stimulation were characterized by a mean that was considerably smaller than the standard deviation (Fig. 2A, E) and by a significant fraction of suppressed cells (e.g. at high contrast, 16/55 in mice and 37/50 in monkeys, respectively). Specifically, across contrasts, changes in rates due to optogenetic input in mice and monkeys were 4.2 ± 8.9 spk/s (mean±std) and -0.7 ± 43 spk/s (mean±std), respectively. The distribution of optogenetic responses was contrast invariant in mice (KS test, p-value>0.05). The distribution of optogenetic responses was contrast invariant in mice (KS test, p-value>0.05). In monkeys, on the other hand, optogenetic responses observed at zero contrast followed a distribution that was significantly different from those at higher contrasts (KS test, p-value>0.05). This difference between the two species is consistent with previous analyses (Histed, 2018; Nassi et al., 2015), which have reported that visual and optogenetic stimuli were summed linearly in mice (Histed, 2018) and sublinearly in monkeys (Nassi et al., 2015). The difference in response distribution observed in monkeys at low contrast came from weaker suppression, and appears

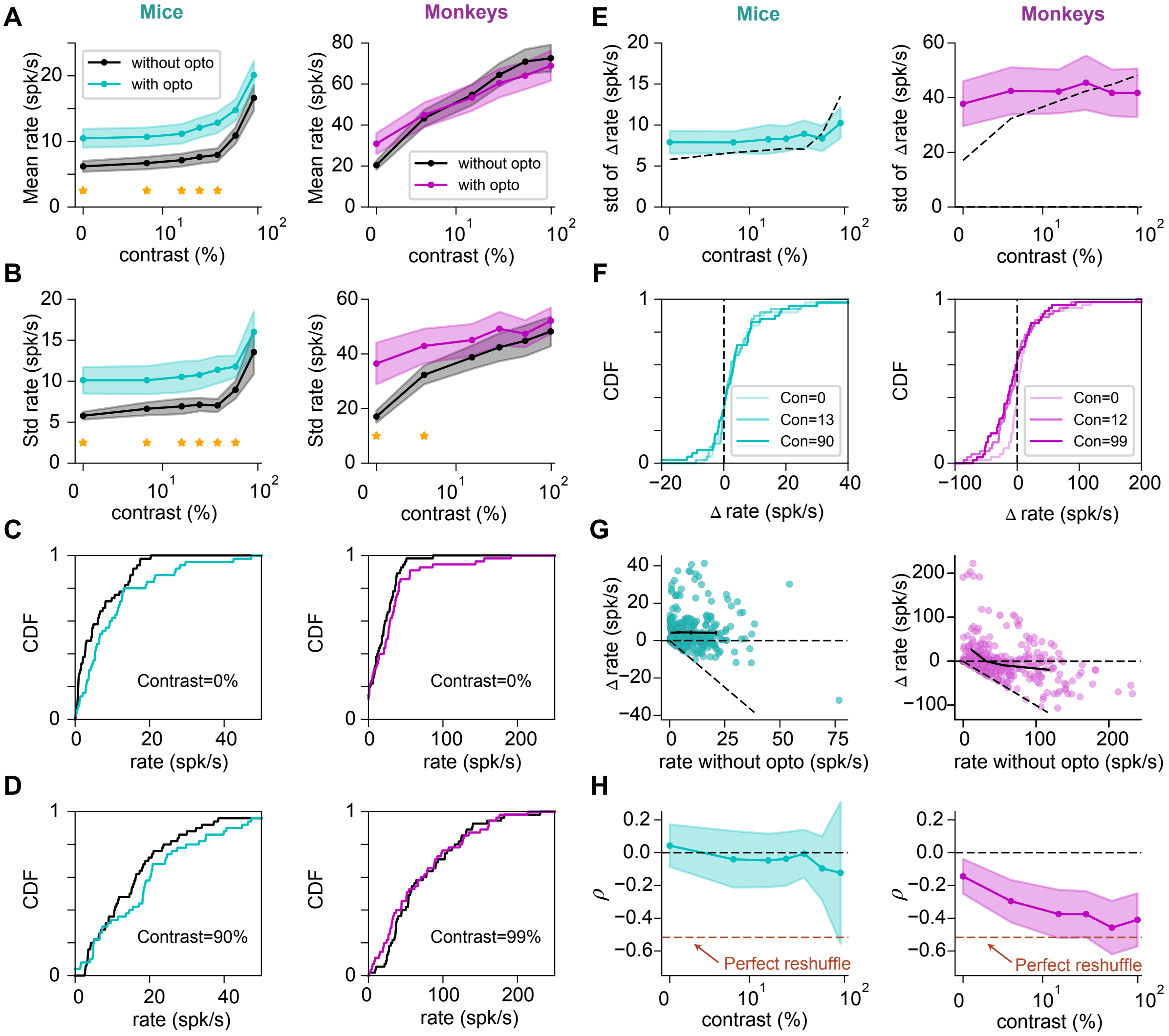

Figure 2: **Effects of optogenetic stimulation on population statistics in mice and monkeys.** (**A**) Population average response to visual and optogenetic stimuli. Optogenetic modulation of mean responses to visual stimuli was weak; it was only significant in mice at low contrast ($< 55\%$, Welch's t–test, p–value>0.05). (**B**) Standard deviation of the distribution of firing rates across the population. In both species, the standard deviation of firing rates is increased significantly by light only at low contrast (in mice, $< 90\%$; In monkeys, $< 12\%$; f–test, p–value>0.05). (**C-D**) Cumulative distribution function (CDF) of firing rates at low **C** and high **D** contrasts of visual stimuli, with (colored) and without (black) optogenetic stimuli. In both species, optogenetic stimuli weakly modulate the CDF. Such modulation was only significant for mice at low contrast (6% and 12%, KS test, p–value<0.05). (**E**) Standard deviation of the change in firing induced by optogenetic stimulation as a function of contrast of the visual stimulus (colored lines). In both species, the standard deviation of the change in firing rate was larger than the mean and of the same order of the standard deviation of rates without the laser (dashed black lines). (**F**) Cumulative distribution function of changes in firing rate (Δr) induced by optogenetic stimulation at three example contrasts (indicated by transparency levels). In both species, and at all contrasts, optogenetic stimuli generated a broad distribution of firing rate changes, with a large fraction of suppressed cells. (**G**) Change in firing rate induced by optogenetic stimulation as a function of baseline activity (i.e. activity without optogenetic stimuli). Each cell is represented once for every value of contrast of the visual stimuli. Continuous black lines represent mean Δr in quartiles of the distribution of rates without optogenetic stimuli. In both species, cells with lower baseline firing tend to be excited by optogenetic stimuli, while cells with higher baseline firing tend to be less excited or, as in the case of monkeys at high contrast, suppressed. (**H**) Normalized covariance of baseline rates and change in rates induced by optogenetic stimuli ($\rho = \mathrm{Cov}(r, \Delta r)/\mathrm{Var}(\Delta r)$). This ratio decreases with contrast in both species and, in monkeys, it approaches the value -0.5. As described in the text, a ratio close to -0.5 is required to obtain invariance of the distribution of rates.

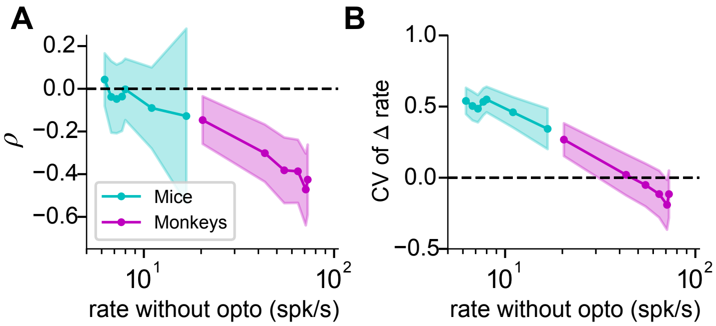

Figure 3: **Optogenetic responses in mice and monkeys as a function of average visual response.** (**A**) Normalized covariance (Eq. (1)) as a function of average visual response (different points correspond to different contrasts). (**B**) Ratio of mean over standard deviation of the change in rate as a function of average visual response. In panels **A** and **B**, data in both animals seem to lie on the same underlying curve, with data in mice/monkeys covering the low/high rate regions, respectively.

to be a consequence of the fact that the decrease in firing of cells cannot exceed their activity in the absence of optogenetic stimuli. The response statistics characterized here suggest that optogenetic stimuli *reshuffled* visual responses in mice and monkeys: they strongly modulated single neuron responses but overall led to a weak modulation of population activity.

How can optogenetic stimulation leave the distribution of rates unchanged, while at the same time leading to large changes in firing rates of individual neurons, whose standard deviation is comparable to the standard deviation of firing rates in the absence of optogenetic inputs (Fig. 2B)? If optogenetic responses were independent from baseline firing, as one might expect intuitively, optogenetic stimuli would significantly broaden the distribution of visual responses. The rate variance with ($\sigma^2_{r+\Delta r}$) and without (σ^2_r) optogenetic stimuli are related by $\sigma^2_{r+\Delta r} = \sigma^2_r + \sigma^2_{\Delta r} + 2\mathrm{Cov}\,(r, \Delta r)$. Thus, independence of r and Δr would imply the rate variance is increased by $\sigma^2_{\Delta r}$, which as we have seen (Fig. 2E) is comparable in size to σ^2_r. If the distribution of rates is not modified by optogenetic stimuli, then $\sigma^2_{r+\Delta r} = \sigma^2_r$, which requires a negative covariance between r and Δr, with value $\mathrm{Cov}\,(r, \Delta r) = -0.5\sigma^2_{\Delta r}$.

The relationship between change in rate induced by laser Δr and rate without laser r is illustrated in Fig. 2G for both mice and monkeys. In mice, the two are weakly anticorrelated (Pearson correlation coefficient was -0.06), consistent with a significant increase in the width of the rate distribution with optogenetic stimulation, while in monkeys, there is a strong anticorrelation between Δr and r (Pearson correlation coefficient was -0.38) – neurons with higher (lower) baseline rates were more likely to be suppressed (excited) by optogenetic stimuli. To quantify the degree to which this anticorrelation leaves the distribution of firing rates invariant, we define a normalized covariance

$$\rho = \frac{\mathrm{cov}\,(r, \Delta r)}{\sigma^2_{\Delta r}}; \tag{1}$$

so that ρ is equal to -0.5 when optogenetic stimuli leave invariant the distribution of rates. ρ is plotted as a function of visual contrast in Fig. 2H. The normalized covariance is close to zero for mice at low contrast, but becomes significantly negative at high contrast. In monkeys, it is negative at all contrasts and gets close to -0.5 at high contrast, consistent with invariance of the distribution of rates and consequent reshuffling of rates by optogenetic stimulation.

The analysis shown so far shows common trends in mice and monkeys but also significant quantitative differences. We wondered whether these quantitative differences might at least partly be explained by the differences between the firing rates observed in the two animals. Therefore, we plotted dimensionless variables characterizing the effect of optogenetic stimuli on the statistics of network activity in the presence of a given visual stimulus as a function of the average firing rate induced by that visual stimulus. Remarkably, Figs. 3A–B show for all considered statistics, curves traced by monkey data seem to be the continuation of those for mouse data, with mice occupying the low rate region of underlying common curves, while monkeys occupy the high rate region. These figures show that in both species, the mean

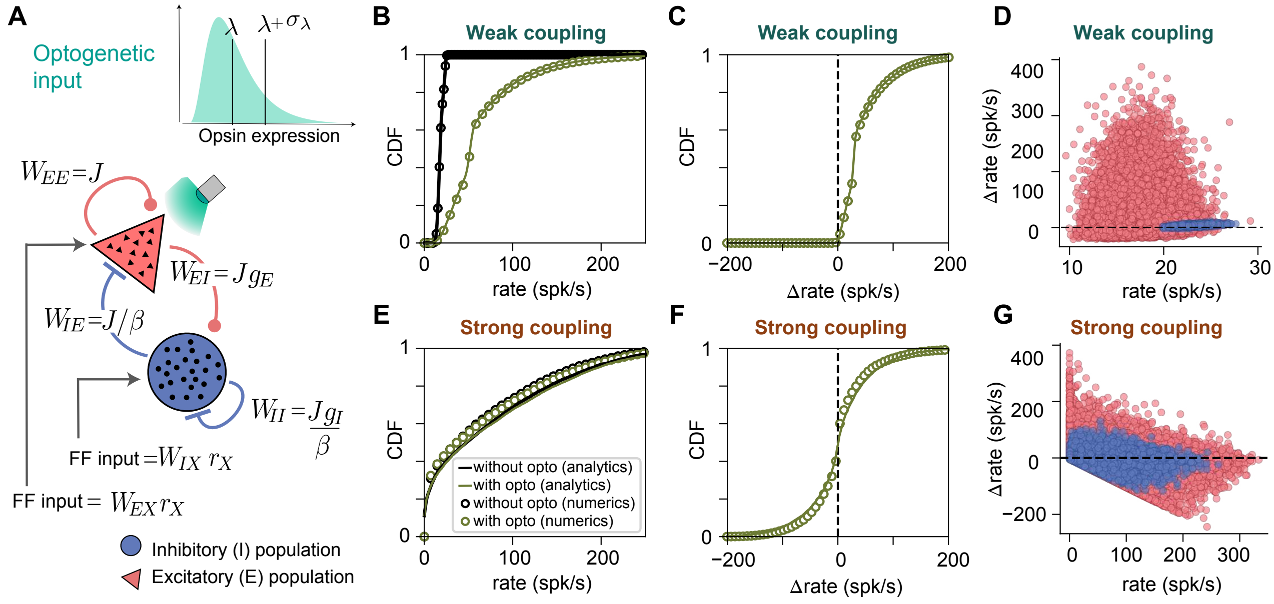

Figure 4: **Strong coupling leads to large fractions of suppressed cells, negative covariance and activity reshuffling in randomly connected E-I network models.** (**A**) Schematic of network model. We investigated networks with randomly connected E and I rate units. The network is driven by a feedforward input, targeting both cell types, and by an optogenetic input targeting only E neurons. In the model, opsin expression in E cells is distributed according to a lognormal distribution.(**B-D**) Responses to optogenetic stimulation of E cells in a weakly coupled network model. In all panels, dots represent numerical simulations while lines are mean field theory predictions. For weak coupling, optogenetic stimulation strongly affects the distribution of firing rates (**B**), with only a weak fraction of suppressed cells (**C**). Changes in firing rates are positively correlated with baseline activity (**D**, Pearson correlation coefficient are 0.1 and 0.6 for E and I cells, respectively). (**E-G**) As in panels **B-D**, but in a strongly coupled network model. For strong enough coupling, optogenetic stimulation does not affect significantly the distribution of rates (**E**). Firing rate changes are broadly distributed, with a mean around zero and a large fraction of suppressed cells (**F**). Firing rate changes are anticorrelated with baseline activity (**G**, Pearson correlation coefficient are -0.3 and -0.2 for E and I cells, respectively). In the simulations, $J = 0.01$mV and 0.3mV for weak and strong coupling, respectively. Other simulation parameters can be found in table 1 in the *Methods*. In all panels, only cells whose response was higher than 1 spk/s in at least one of the two conditions are shown (100% in weak coupling and 43% for strong coupling.)

of Δr relative to its standard deviation decreases as a function of contrast until it becomes negligible at high firing rate. The normalized covariance also decreases with firing rate, getting closer to -0.5 at high rate.

Network reshuffling emerges in strongly coupled network models

To understand the mechanisms underlying reshuffling of visual responses by optogenetic stimuli, as well as the observed differences between mice and monkeys, we investigated the response properties of a mathematical model of cortical circuits. We analyzed a network of randomly connected E and I rate neurons (Harish and Hansel, 2015; Kadmon and Sompolinsky, 2015; Sompolinsky et al., 1988). In the model, the single neuron response function (firing rate vs input current relationship) was taken to be the f-I curve of leaky integrate-and-fire neurons driven by white noise (Amit and Brunel, 1997; Amit and Tsodyks, 1991; Sanzeni et al., 2020; Siegert, 1951). This transfer function shows all the expected features of transfer functions of cortical neurons in vivo: supralinear at low mean input due to fluctuations in these inputs, transitioning to sublinear at high input due to saturation, but the results presented here do not depend on this specific choice (see supplementary information, Fig. S1). Visual stimuli are described by feedforward inputs to E and I units; optogenetic inputs are described by an additional input to E cells, whose strength is drawn from a log-normal distribution.

Through numerical simulations and analytical calculations, we found that the network model reproduces the main features of the experimental data provided that recurrent connectivity and visual input are strong enough, and the distribution of opsin expression is heterogeneous enough. In particular, we first focus on the dependence of optogenetic responses on the network's effective synaptic coupling strength (referred below as coupling strength), which is defined as the synaptic strength times the sensitivity of the postsynaptic cell to synaptic input ((Ahmadian and Miller, 2021)). This sensitivity is the cell's gain, i.e. the slope of its input/output function. In the supralinear portion of the input/output function, this increases with network activation and thus with the strength of visual input.

Example network simulations for weak and strong coupling are shown in Fig. 4B-G. In the absence of optogenetic stimulation, and consistent with previous theoretical studies (Amit and Brunel, 1997; Palmigiano et al., 2020; Roxin et al., 2011; Sanzeni et al., 2022; van Vreeswijk and Sompolinsky, 1996), the distribution of rates is narrow for weak coupling and broad for strong coupling. As predicted from the theoretical analysis described below, responses to optogenetic stimulation strongly depend on the strength of coupling. For weak coupling, optogenetic stimulation generates a significant shift of the distribution of rates (Fig. 4B). This leads to a broad distribution of optogenetic responses, characterized by a positive mean and a lack of significantly suppressed cells (Fig. 4C). Moreover, single neuron responses to optogenetic stimulation are positively correlated with baseline activity (Fig. 4D, Pearson correlation coefficients are 0.1 and 0.6 for E and I cells, respectively). This positive correlation follows from the supralinear shape of the transfer function at low rates. This effect is more pronounced in I cells because, with respect to E cells, they have a shorter membrane time constant and hence a larger supralinear region (Sanzeni et al., 2020). The response structure observed in weakly coupled networks matches naive expectations but strongly differs from experimental observations. When coupling is strong, on the other hand, optogenetic stimulation does not modify significantly the distribution of rates (Fig. 4E). It produces a broad distribution of rate changes, characterized by a mean that is close to zero, and a large fraction of suppressed cells (Fig. 4F). In this regime, single neuron responses to optogenetic stimulation are negatively correlated with baseline activity (Fig. 4G, Pearson correlation coefficient are -0.3 and -0.2 for E and I cells, respectively). These response properties are strikingly similar to experimental observations in monkeys at high contrast.

We now turn to a deeper investigation of the circuit mechanisms underlying these different aspects of the response of the network to optogenetic stimulation, using mean-field analysis.

Mechanisms shaping the response to optogenetic inputs in network models

We performed numerical and mean field analyses of the network (see details in the *Methods*) to elucidate the circuit mechanisms underlying the simulation results highlighted in Fig. 4. These analyses show that strong recurrent connectivity generates a weak change in average firing rate to optogenetic input, provided the network is inhibition-dominated (Fig. 5). In particular, numerical simulations and mean field theory show that the average change in firing rate due to optogenetic stimulation decreases when (Fig. 5A-C, first row) the firing rate r_X of the neurons providing feedforward inputs to the network increases, synaptic efficacies J become stronger, the strength of inhibition onto E neurons g_E becomes stronger, and/or the strength of inhibition onto I neurons g_I becomes weaker. These results can be understood by noticing that, for small enough optogenetic input and neglecting the contributions coming from quenched disorder

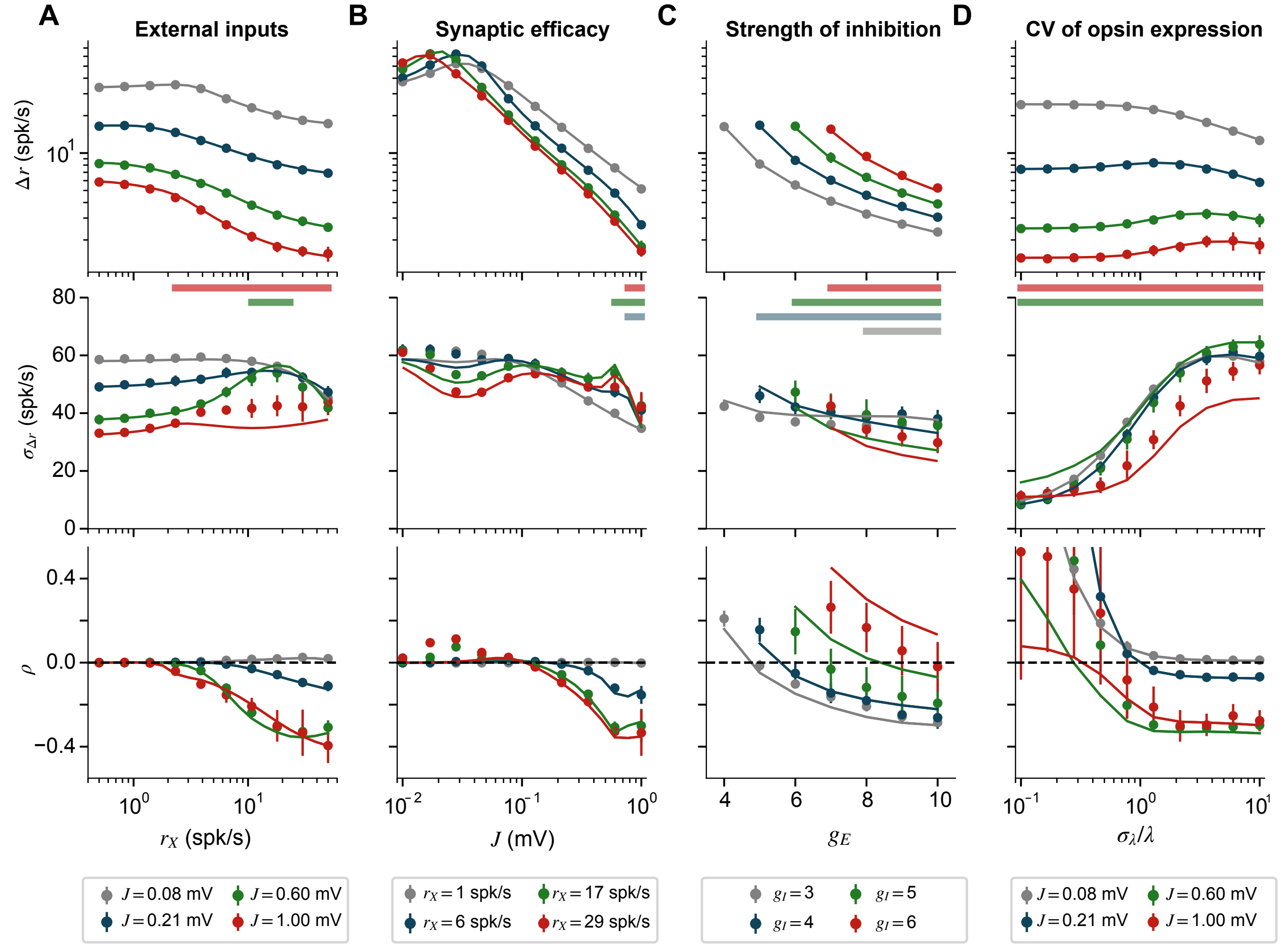

Figure 5: **Dependence of the statistics of network response to optogenetic input on network parameters and optogenetic input statistics.** (**A**) Statistics of network response to optogenetic stimuli as a function of feedforward input r_X, for different values of recurrent synaptic efficacies (J, colors; J is the E-to-E efficacy, but all 4 efficacies ({E,I}-to-{E,I}) are proportional to J). Panels show (from top to bottom): mean change in firing rate; standard deviation of the change in firing rates; normalized covariance between baseline rate and change in rates (Eq. 1). Lines under the second row indicate parameters leading to chaotic dynamics. Dots with errorbars correspond to averages (mean±sem) computed over 20 realizations of the network model. Continuous lines represent mean field theory predictions. Increasing feedforward inputs to the network generates smaller mean rate change, while only mildly affecting the standard deviation of rate changes. Negative covariances are obtained for strong feed-forward inputs, provided J is strong enough. Note that negative covariances can appear both in chaotic networks (red curve in lower panel) and in non-chaotic networks (e.g. green curve in that panel). (**B**) Same as in **A** but as a function of J, for different values of r_X (colors). Increasing J significantly decreases mean rate change and response covariance, while affecting only mildly the standard deviation of rate changes. (**C**) Same as in **A** but as a function of the strength of inhibition to E cells g_E, for different values of the strength of inhibition to I cells g_I. These parameters weakly affect the width of the distribution of responses. Stronger inhibition to E cells, and weaker inhibition to I cells, both lead to lower mean rate changes, and more negative covariance of rate change with baseline rate. Networks with $g_E < g_I$ are not analyzed because they are either unstable or multistable (Sanzeni et al., 2020). (**D**) Same as in **A** but as a function of the CV of the distribution of opsin expression σ_λ/λ, for different values of J. The ratio σ_λ/λ weakly affects the mean rate changes, but strongly affects the width of the distribution of responses, and the response covariance. In particular, heterogeneous opsin expression ($\sigma_\lambda/\lambda > 1$) is required to generate broad distributions of responses and negative covariance. Simulations parameters are: $\sigma_\lambda/\lambda = 2$ in panels **A-C**; $g_E, g_I = 8, 3$ in panels **A,B,D**; $r_X = 10$spk/s and 18spk/s in panel **C** and **D**; $J = 0.6$mV in panel **C**.

in the connectivity and in the opsin expression, the implicit equation defining the mean rates (see Eq. (12) of the *Methods*) can be linearized and solved analytically. Using these approximations, the mean changes in firing induced by optogenetic stimulation in E and I populations are

$$\Delta r_E = \frac{g_I \lambda}{K J \tau_E \left(g_E - g_I\right) \phi'_E}\,, \quad \Delta r_I = \frac{\lambda}{\gamma K J \tau_E \left(g_E - g_I\right) \phi'_E} \tag{2}$$

where λ is the mean optogenetic input, K and γK are the average number of E and I recurrent connections per neuron, while Δr_A, τ_A, and ϕ'_A are the mean change in firing, the membrane time constant, and the single neuron transfer function at the network operation point of population $A \in [E, I]$. Eq. (2) captures qualitatively the dependency on J, r_X and $g_{E,I}$ observed in simulations (Fig. 5A-C, first row). The dependence of the network response on ϕ'_E generates, in models with supralinear single neuron input-output function, an effective strength of interactions between cells that increases with the activity level (Ahmadian et al., 2013). Therefore, Eq. (2) shows that stronger coupling, generated either by larger synaptic efficacy J or through an increased activity level, leads to weaker mean optogenetic response.

In the model, the opsin expression heterogeneity generates a cell-specific input that, unlike the average optogenetic input, cannot be compensated by recurrent interactions and produces highly heterogeneous responses (Figure 5, second row). This effect increases with the variability in opsin expression (σ_λ, Fig. 5D, second row), and weakly depends on coupling strength (Fig. 5A–C, second row). Heterogeneous opsin expression and strong coupling also lead to a large number of suppressed cells (Figs. 4.) In sum, when coupling is strong, cell-to-cell fluctuations (which can be both positive or negative) dominate over the small mean change in input produced by optogenetic stimuli, and therefore determine the changes in firing of neurons.

Another noteworthy result of our analysis is that a combination of strong coupling and a broad distribution of opsin expression naturally leads to a normalized covariance ρ closer to -0.5 (Fig. 5, third row) and to network reshuffling (Fig. 4E–G). The fact that ρ becomes negative is due to an interplay between single neuron nonlinearities and the strong inhibitory feedback canceling the mean network response to optogenetic input. In the model, recurrent inhibition acts to counterbalance the change in rate caused by optogenetic input and thus produce a small change in mean firing rate across the network, while heterogeneities in opsin expression and recurrent interactions generate a broad distribution of firing rate changes, characterized by a large fraction of suppressed cells. Strong couplings leads to a broad distribution of baseline rates and therefore a large fraction of cells with very low baseline rates. Upon heterogeneous optogenetic stimulation, the non-negative nature of firing rates biases the response of cells with very low baseline rates toward excitation (see Fig.4G). To compensate for this bias, and generate a close to zero mean change in rate, the circuit produces, on average, suppression in neurons with higher baseline activity (see Fig.4G). This leads to a negative correlation between baseline activity and response to optogenetic input, i.e. a negative ρ.

Previous theoretical works have shown that the dynamics of rate networks becomes chaotic when recurrent connections are strong (Harish and Hansel, 2015; Kadmon and Sompolinsky, 2015; Sompolinsky et al., 1988). Consistent with these results, we observed the emergence of chaotic dynamics in our simulations as J and/or r_X increase (Fig. 5, lines under the first row). Note that in our simulations chaos is heterogeneous, with a broad distribution of temporally-averaged firing rates that reflects the quenched disorder in the inputs and synaptic connectivity. Since negative covariance and neuronal reshuffling emerge at large contrast, we wondered if the appearance of these features is related to chaotic dynamics. To answer this question, we have systematically compared the regions of parameter space in which negative covariance and chaotic dynamics appear (supplementary information Fig. S2) but found that they only partially overlap. These results show that negative covariance and chaotic dynamics are two distinct phenomena that emerge in networks of strongly coupled neurons.

Network models capture experimental recordings in mice and monkeys

Optogenetic responses in mice and monkeys have qualitatively similar features, which we summarized above with the term "network reshuffling", but also important quantitative differences (Figs. 1 and 2). Compared to monkeys, optogenetic responses in mice had a larger mean over standard deviation ratio and weaker normalized covariance. However, when dimensionless quantities characterizing the statistics of network responses are plotted as a function of population rate in the absence of optogenetic stimuli, these differences seem to largely arise because of different ranges of firing rates, possibly due to a difference in external inputs to both animals (Fig. 3A–B). This observation supports the idea that the strength of feedforward inputs might be the main factor explaining differences between mice and monkeys. To test this hypothesis, we investigated if the same network model can capture responses observed in the two species, provided that the strength of external stimuli to both animals differs. We fitted our network model to recordings in mice and monkeys (see methods section for details of the fitting procedure) assuming common network parameters (J, K, $g_{E,I}$, β, σ_λ/λ, CV_K) across species, but allowing for different feedforward (r_X) and optogenetic (L) inputs. Results are shown in Fig. 6A (dashed lines). The figure shows that our model captures the main features of the recordings: weak mean optogenetic response; large standard deviation of optogenetic response; and increasingly negative normalized covariance with contrast of the visual stimulus. This agreement with experimental data shows that the mechanism we proposed for neuronal reshuffling can emerge with biologically realistic values of model parameters.

Our model, in which the only difference between two species comes from parameters of visual and optogenetic inputs, captures key differences between mice and monkeys: a lower visual response in mice; a larger ratio between mean and standard deviation of the change in rate by optogenetic stimuli in mice; and a more negative normalized covariance in monkeys. These results support the idea that the strength of feedforward inputs is a key factor shaping differences between the two species. At the same time, the model fails to capture quantitatively the mean optogenetic response and the amplitude of the normalized covariance in monkeys. These discrepancies suggest that differences in the network structure and in the distribution of opsin expression might be important to fully capture the observed response properties. In agreement with this observation, we found that repeating the fitting procedure in models with independent structure for mice and monkeys gives a better description of experimental recordings (Fig. 6A, continuous lines). The best fit parameters are characterized by weaker inhibition (bigger g_I produces stronger self-inhibition of I cells and larger network response, e.g. see Eq. (2)) and less heterogeneous optogenetic expression in mice than in monkeys.

Previous theoretical work (Ahmadian and Miller, 2021; Rubin et al., 2015; Sanzeni et al., 2020) has suggested that cortical circuits operate in a loosely balanced regime, characterized by a net input to cells (the sum of excitatory and inhibitory inputs) comparable to its individual (excitatory and inhibitory) components. This property can be quantified by the balance index, defined as the ratio between the net input (excitatory minus inhibitory and the total excitatory inputs to cells (Ahmadian and Miller, 2021). In our model, using the parameters inferred from data, we found a smaller balance index in monkeys than in mice (Fig. 6C). Recordings in mice were consistent with the underlying network being loosely balanced at low and intermediate contrast (balance index greater than 0.1). Somewhat tighter balance was observed at high contrast in mice and for all contrasts in monkeys, for the model in which mice and monkeys were jointly fit, while monkeys showed very tight balance in the model in which they were independently fit. Our theoretical analysis suggested that, in order for the mean optogenetic response to be weak, the average optogenetic current should be small compared to other inputs in the circuit.

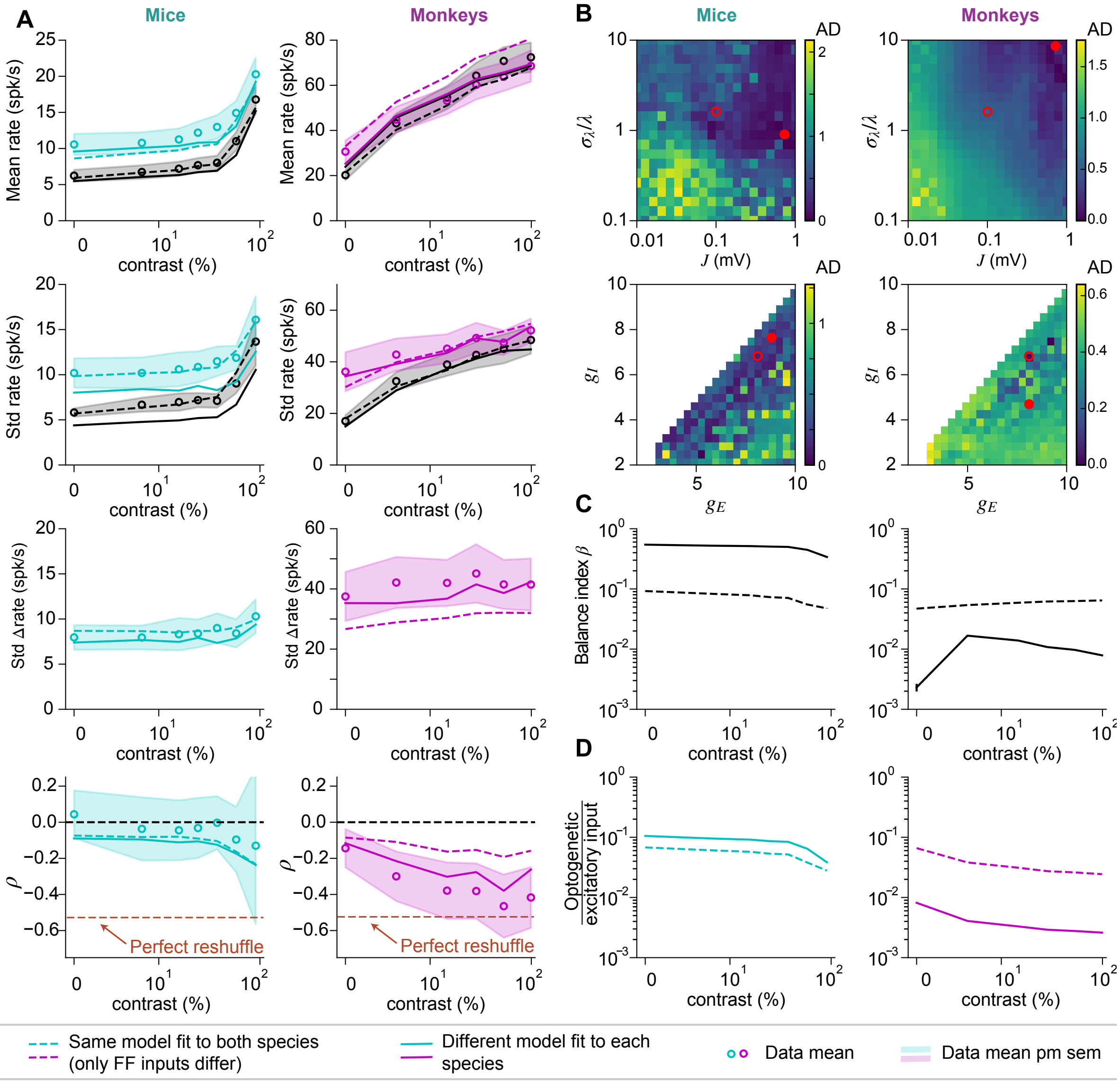

Figure 6: **Fitting mice and monkey data with the network model.** (**A**) Comparison between data (circles represent means, shaded regions are mean±sem) and network model with parameters that best fit data (dashed and continuous lines) in mice (left) and monkeys (right). Dashed lines represent best fits obtained with the same network model in mice and monkeys, but different feedforward and optogenetic inputs. Continuous lines represent best fits obtained with different network models. Model plots are obtained from simulations, averaging over 20 realizations. Network models reproduce key features of experimental data, including: weak mean changes in firing rate; broad distribution of responses; and increasingly negative covariance with contrast of the visual stimulus. (**B**) Average discrepancy (AD, measured as $\log_{10}$ of the ratio between the actual and the minimum error) between best model predictions and experimental data for mice (left) and monkeys (right) obtained with models where two parameters are constrained to the values specified by both axis. Red circles indicate optimal parameters, used to generate panels **A** (open circles: same parameters in mice and monkeys; filled circles: dfferent parameters in mice and monkeys). In both species, consistent with Fig. 5D, better fits are obtained for $\sigma_\lambda/\lambda > 1$, i.e. highly heterogeneous optogenetic expression, and for biologically realistic values of J, in the range of $[0.1-1]$mV. (**C**) Inferred values of the balance index (Ahmadian and Miller, 2021), i.e. the ratio between net input current (excitatory minus inhibitory) and the total excitatory input in response to visual stimuli. The lower index observed in monkeys indicates stronger coupling in this species than in mice. (**D**) Inferred values of the ratio between optogenetic input and the total visually-driven excitatory input. Both in mice and monkeys, the model requires a small mean optogenetic input, relative to visually-driven excitatory input, to capture experimental data.

Consistent with this assumption, we found that the model fitted to data had mean optogenetic inputs significantly smaller than the visually-driven excitatory inputs, with this difference becoming very strong in the model that was fit independently to monkey data (Fig. 6D).

Discussion

Summary of the results

We analyzed electrophysiological recordings of mice and monkey V1 while the animals viewed visual stimuli, with or without concurrent optogenetic stimulation of E neurons. In both species, and for all contrasts of visual stimuli, optogenetic stimuli generate a broad distribution of responses, whose mean is considerably smaller than the standard deviation. This large heterogeneity of responses to optogenetic stimuli results in particular in large fractions of suppressed cells. The main quantitative differences between the two species are that the amplitude of single-cell responses to visual and optogenetic stimulation, and the fraction of suppressed cells, are larger in monkeys, while the mean optogenetic response in monkeys is smaller. The heterogeneity of cell responses underlies a simple pattern at the population level. In monkeys, at all contrasts, optogenetic stimuli strongly modulate the activity of single neurons, while leaving invariant the distribution of rates across the population. This feature is specific to optogenetic stimuli (visual stimuli shift the distribution toward higher rates), and it appears also in mice at high contrast. Intriguingly, statistics of optogenetic responses across the two species form a single, continuous function of baseline (visually-induced) population activity, with mouse and monkey responses corresponding to lower and higher baseline activities, respectively.

To understand the mechanisms that are responsible for this network reshuffling by optogenetic inputs, we analyzed a cortical network model composed of randomly-connected E and I neurons. When recurrent connectivity and/or feedforward input is sufficiently strong, the strong inhibitory feedback balances the mean excitatory optogenetic input, leading to a small change in mean firing rates, while heterogeneities in opsin expression and recurrent interactions lead to a broad distribution of firing rate changes, with a large fraction of suppressed cells. The rectification of the single cell input-output function biases responses of neurons with nearly silent baseline activity towards excitation, while the circuit dynamically generates suppression in neurons with higher baseline activity in order to maintain the small change in mean firing rate. This negative correlation between baseline firing rate and optogenetically induced change in firing rate is consistent with experiment and generates reshuffling of neural activity in the network model like that observed in mice and monkeys.

Theoretical implications: Coupling strength in cortex

Balanced network models (Ahmadian and Miller, 2021; Ahmadian et al., 2013; Amit and Brunel, 1997; Brunel, 2000; Roxin et al., 2011; Rubin et al., 2015; Sanzeni et al., 2020, 2022; van Vreeswijk and Sompolinsky, 1996, 1998) have been successful in reproducing several ubiquitous features of cortical activity: highly irregular firing (Compte et al., 2003; Softky and Koch, 1993), excitation/inhibition balance (Haider et al., 2006), and broad distributions of firing rates (Buzsáki and Mizuseki, 2014; Hromádka et al., 2008; O'Connor et al., 2010). These properties can arise either in networks that are tightly balanced or loosely balanced (Ahmadian and Miller, 2021), meaning that the net input after cancellation of excitation and inhibition is respectively much smaller than, or comparable in size to, the cancelling factors. Here, we showed that such networks can reproduce the surprising effect of an approximate invariance of the distribution of firing rate to optogenetic input, in spite of the fact that this input causes large changes of firing rates of individual neurons - i.e., reshuffling. This provides further evidence that recurrent connectivity in V1 is powerful, and stabilized by strong inhibition. A key property of the loosely balanced regime, as exemplified by the supralinear stabilized network (SSN) model (Ahmadian and Miller, 2021; Ahmadian et al., 2013; Rubin et al., 2015), is that network behavior depends on the network's activation level. In our model, this accounts for the strong dependence of the results of optogenetic stimulation on the state of the network prior to stimulation, which in the experiments described here is controlled by the contrast of the visual stimulus.

The best fits of our network model to experimental data provide evidence that cortical circuits in mice operate in a loosely balanced regime at low contrast, while balance appears to be tighter at high contrast in mice and at all contrasts in monkeys. The tighter balance observed in our analysis is inconsistent with previous estimates that balance in sensory cortex is loose (Ahmadian and Miller, 2021). This inconsistency might be due to the fact that previous estimates were derived assuming average firing rates significantly lower than those reported in Nassi et al. (2015) (however, electrode recording as in Nassi et al. 2015 can produce over-estimates of average firing rates, due to under-representation of weakly active or silent cells, *e.g.*Barth and Poulet 2012). Alternatively, the stronger coupling we have found might be a consequence of the simple structure assumed in our model, which completely neglects the structured connectivity in cortical circuits (further discussed below).

The model investigated here is a randomly connected firing rate model, similar to previously studied models (Harish and Hansel, 2015; Kadmon and Sompolinsky, 2015). Given that the transfer function we used is that derived analytically for leaky integrate-and-fire neurons in the presence of noise, we expect similar results to hold in networks of LIF neurons with similar parameters, and also more generally in networks of more realistic spiking neurons, provided these networks operate in an asynchronous state, which we know can be realized for realistic single neuron and synaptic parameters (Brunel, 2000; Renart et al., 2010). Indeed, the mean-field equations for such networks should be very similar to the equations derived in this paper (SI), with the exception of additional equations relating the variance of the fluctuations in synaptic inputs self-consistently to the statistics of the firing rates of the network. We do not expect this difference to affect our results qualitatively. These results should also hold in networks in weakly synchronized irregular states in which firing rates depend only weakly on the degree of synchronization (Brunel and Hakim, 1999; Brunel and Wang, 2003), as well as in spatially structured networks generating irregular traveling waves (Davis et al., 2020; Muller et al., 2018).

An interesting feature of our analysis is that the coupling strength parameters that best fit the data are close to a transition to chaos in the rate model (Engelken et al., 2020; Harish and Hansel, 2015; Kadmon and Sompolinsky, 2015; Sompolinsky et al., 1988). This makes sense since both reshuffling and chaos require strong coupling. We note however that chaos is not required for our results to hold, since non-chaotic networks reproduce all the main features in the data.

Mice vs monkeys

Our work provides to our knowledge the first systematic comparison of how optogenetic stimuli modulate visual responses in mice and monkeys. It shows that baseline firing plays a key role in determining the statistics of neural responses, and suggests that the different activity levels, due to different strengths of feedforward input, could underlie differences observed in the two species. When key statistics were plotted as a function of baseline rate, curves traced by mice and monkey data seem to lie on the same underlying curve, with mice occupying the low rate region while monkeys occupy the high rate region. Consistent with this observation, fitting data from both species with a network model with the same parameters, but different input strengths, gives a reasonable description of the data, though the quality of the fit improves when network parameters are allowed to differ between the two species.

A recent electron microscopy study performed a detailed comparison of reconstructed neurons and their synaptic connections in mouse and primate V1 (Wildenberg et al., 2021). Surprisingly, synaptic connectivity in primates was found to be much more sparse than in mice, with a larger ratio of excitatory to inhibitory synapses in mice. The larger ratio of excitatory to inhibitory synapses in mice seems consistent with the weaker inhibition we found in mice. However, the weaker coupling in monkeys than in mice reported in (Wildenberg et al., 2021) seems at first sight at odds with our conclusions that monkeys appear to be in a more strongly coupled regime than mice. However, strong coupling can arise from stronger synaptic connections and/or stronger activation levels, and the latter are larger in monkeys. Furthermore, the model reproduces well the data in broad regions of parameter space (see, in Fig. 6B, the broad blue regions which indicate reasonable fits in a broad range of values of J) and thus our results are not incompatible with (Wildenberg et al., 2021). Finally, electron microscopy can tell us about numbers of synapses per neuron, but not about the strength of individual connections, which leaves open the possibility that synaptic coupling strength is comparable in mice and monkey.

The recordings in monkeys were done in cells that optimally responded to the orientation, spatial frequency, and size of the presented stimulus, whereas in the mouse recordings there was no systematic relationship between cell preferences and the stimulus. The use of optimal stimuli in monkeys would bias observed rates to the high end of the rate distribution. It is possible that these higher-firing cells are more likely to be suppressed by optogenetic stimulation, and that this could contribute to the greater suppression seen in optogenetic responses of monkey vs. mouse.

Relations with previous experimental works and computational consequences

The modulation of visual responses by optogenetic stimuli we have described, characterized by large changes in single neuron firing that weakly affect the overall activity of the network, might be relevant to understand responses of neurons observed in other conditions. Large numbers of suppressed cells in response to optogenetic activation of pyramidal cells have been reported in experiments using one-photon stimulation in mouse visual cortex (O'Rawe et al., 2022), ferret visual cortex (Wang et al., 2022), and targeted two-photon stimulation of small ensembles in mouse barrel cortex (Dalgleish et al., 2020).

In the monkey, optogenetic drive to one ocular dominance (OD) domain causes suppression of the activity of cells in the other domain (Chernov et al., 2018). Given that in our monkey data the diameter of the optogenetic perturbation is around 680 μm, it is possible that it targets one OD region more than another and the suppression observed here could also correspond, at least partially, to a mis-match between the OD region that is predominantly stimulated and that of the measured cell. Surprisingly, and at odds with the results presented here (see Fig. 2), recent works in monkeys have reported large changes in mean firing rate with optogenetic stimulation of E cells (Andrei et al., 2019; Chen et al., 2022). In (Andrei et al., 2019), the near absence of suppressed cells could be due to the very weak stimulus contrast utilized (for which both our data and our models indicate weaker suppression), and to the orientation-specific stimulus, which, as visual stimuli, is able to drive the mean activity. In (Chen et al., 2022), optogenetic stimulation to E cells causes a large change in the mean activity from the response to stimuli of contrast up to 60%. This discrepancy is less likely to be explained by the weaker values of contrast utilized. It should be noted, however, that comparisons of results reported in (Nassi et al., 2015) and (Chen et al., 2022) are complicated by the different recording approaches used in these studies: the former used electrophysiology to measure single unit responses to their preferred visual stimuli; the latter used wide-field calcium imaging and measured responses reflected the pooled spiking activity of a large population of neurons with diverse tuning properties. Because averages were performed on the entire population and not only the cells which respond maximally to the visual stimulus (as in our monkey data), the effects of reshuffling could have been partially masked. Finally, because the optogenetic stimulation in (Chen et al., 2022) was delivered in the context of a visual discrimination task, it is possible that the network response in this case differs from the case studied here, were there was no training involved and the optogenetic input was simply a perturbation of activity. In fact, recent work (Akitake et al., 2022) has shown that mice can learn to respond to an optogenetic input, and the response to the stimulation in that case can dramatically increase with learning.

The response properties of cortical networks to optogenetic inputs we have found here might also be relevant to understanding the response of cortex to modulatory inputs (Sherman and Guillery, 1998). In fact, weak changes in mean population activity together with large changes in individual neuron firing rates, appear after learning in experiments involving brain-machine interfaces (Engelhard et al., 2019), and due to behavioral modulations of activity, like head movements (Bouvier et al., 2020), and running (Liska et al., 2022).

Our results also shed light on the mechanisms of normalization, a widespread phenomenon in cortical circuits (Carandini and Heeger, 2012; Reynolds and Heeger, 2009). The normalization model has been shown to be a good phenomenological description of how neurons respond to both visual and optogenetic inputs in primate V1 (Nassi et al., 2015). In our model, as in previous theoretical works (Ahmadian et al., 2013; Rubin et al., 2015), nonlinearities in the mean response of a population of neurons, such as those traditionally described by the normalization model, emerge because of activity-dependent coupling strength. The striking heterogeneity of normalization properties observed across neurons (Nassi et al., 2015), on the other hand, arises from the heterogeneity of weights and inputs, as also observed in previous models (Rubin et al., 2015). A new feature we have found is that the cells whose responses to the optogenetic stimulus are most “normalized” or suppressed tend to be those that have the highest baseline (visually-induced) firing rates.

Limitations and future directions

The network studied here has a purely random connectivity. The mechanism by which the mean response, but not the standard deviation of the response, is suppressed is in large part that of the tightly balanced network (van Vreeswijk and Sompolinsky, 1998). The randomly connected network, when strongly coupled, elicits recurrent inhibition that largely cancels the mean external input to the network, but cannot cancel the differences in input across cells. As described above, analysis of sensory cortical data from mice and cats suggests that balance is loose (Ahmadian and Miller, 2021). We can obtain negative covariances of rate and rate changes comparable to those in the data with balance indices indicating loose balance (not shown), but we did not find parameters to simultaneously also match other observed statistics in the monkey without quite small balance indices, *i.e.*quite tight balance. This suggests a problem with the present scenario although, as noted above, it is possible that the high firing rates reported in monkey V1 might imply a tighter balance than previously estimated.

A related limitation is that, to match the observation that the mean optogenetic response is small relative to the mean visual response, the mean optogenetic input must be small relative to the mean visually-evoked excitatory input to the network. In layer 4 of mouse V1, peak visually-evoked excitatory current has been estimated to be 60-150pA across cells (Li et al., 2013; Lien and Scanziani, 2013). In cultured neurons, optogenetically-evoked currents are 100’s to 1000’s of pA (Mattis et al., 2011), but light is severely attenuated with transmission through cortex, by 50% over about 40 microns (Yona et al., 2016). Thus, it is difficult to assay whether the model’s requirement is met *in vivo* in cortical cells, but this could also be a problem for the present scenario.

Models with structured connectivity might alleviate these possible problems. Both synaptic connectivity and visual input depend on selectivity of neurons for visual stimulus properties such as orientation, spatial phase, and spatial position. In particular, a notable difference between the visual and optogenetic stimuli studied here is that visual stimuli provide input to neurons selectively according to their tuning, while optogenetic stimuli target all excitatory cells indiscriminately. Thus, an alternative explanation of the small mean response might be that neurons inhibit cells that differ from them in such features as preferred orientation (*e.g.* Liu et al., 2011), preferred spatial phase (*e.g.* Kayser and Miller, 2002; Troyer et al., 1998) or preferred direction (Rossi et al., 2020) more strongly than they excite cells with similar tuning, so that a nonspecific stimulus evokes a net inhibition. Such a mechanism might yield small mean response with more substantial optogenetic input, and this might not depend as strongly on the tight cancellation associated with very strong coupling. Given this alternative mechanism for suppressing mean response, the other properties examined here would likely follow: heterogeneity of optogenetic inputs would produce a broad range of optogenetic responses, and a bias of weakly activated cells to show optogenetic excitation would then cause the negative covariance of rate and rate changes and the associated preservation of the overall distribution of firing rates.

In a preliminary investigation of the effects of structure, we implemented a discrete ring model with wider inhibition than excitation, representing inhibition that is wider in preferred orientation than excitation, as observed in mouse simple cells (Liu et al., 2011). We chose ring model parameters that give similar baseline firing rates compared to the structureless model from Figure 5 panel **A**, but with an eight-fold increase in the mean optogenetic input (see Figure S3). We found that when only measuring neurons that match the orientation tuning of the visual stimulus, as is done for the monkey data, the ring model can achieve equal or greater reshuffling compared to the structureless model even with increased optogenetic input. Thus, at least qualitatively, a structured model can alleviate the possible problem with the size of the optogenetic input. In future work we will examine scenarios of structured connectivity to determine whether the data can be quantitatively well fit while relaxing both possible problems discussed above.

Several other factors might also play a role in the differences between monkey and mouse responses. First, as discussed above, the use of optimal stimuli in monkeys might contribute to monkey optogenetic responses showing greater suppression than mouse. Second, the differences in V1 functional architectures of rodents (with 'salt and pepper' orientation selectivity (Ohki et al., 2005a)) and monkeys (with orientation columns (Hubel and Wiesel, 1977)), along with the tendencies of synaptic coupling to decrease in strength both with cortical distance (Rossi et al., 2020) and with difference in feature preferences (Cossell et al., 2015; Ko et al., 2011), might contribute to weaker synaptic coupling in mice as compared to monkeys.

Conclusions

The over 70 million years of evolution separating mice and monkeys (Siepel, 2009) have produced substantial differences in the structure of their cerebral cortices, e.g. in the presence of orientation and ocular dominance columns in primates (Hubel and Wiesel, 1977) but their absence in rodents (Dräger, 1974; Ohki et al., 2005b; Van Hooser et al., 2005) the size of their receptive fields (Van Den Bergh et al., 2010), and their connection statistics (Wildenberg et al., 2021). Our analysis of optogenetic modulation of visual response in V1 shows that, despite these differences, mice and monkeys may share a common remarkable resistance of their cortex to optogenetic inputs at the population level, in spite of large changes in single neuron firing rates. These results might more generally yield insight into circuit mechanisms underlying differences in the effects of sensory or driving vs. modulatory inputs on cortical responses.

Methods

Experimental methods

Mice

A detailed description of the experimental methods is given in (Histed, 2018). In brief, neurophysiological data from Emx1-Cre animals ($n = 4$) were collected. Animals kept on a monitored water schedule were given small drops of water ($\sim$1 μl) every 60–120 s during recording to keep them awake and alert. The visual stimulus, a Gabor patch with spatial frequency 0.1 cycle/deg and sigma 12.5 deg, were presented for 115 ms [full width at half maximum (FWHM) intensity], and successive visual stimuli were presented every 1 s. Optogenetic light pulses were delivered on alternating sets of 10 stimulus presentations (light onset 500 ms before first stimulus; offset 500 ms after end of last stimulus; total light pulse duration 10.2 s). A 1-s delay was added after each set of 10 stimulus presentations. Extracellular probes were 32-site silicon electrodes (Neuronexus, probe model A4x8). Channelrhodpsin-2 (ChR2) was expressed in E neurons (as described in Histed and Maunsell, 2014) using viral (AAV-EF1a-DIO-ChR2-mCherry, serotype 2 or 8; http://openoptogenetics.org) injections into the Emx1-Cre (Gorski et al., 2002; #5628, Jackson Laboratory) line. Virus (0.25–1.0 μl) was injected into a cortical site whose retinotopic location was identified by imaging autofluorescence responses to small visual stimuli. Light powers used for optogenetic stimulation were 500 μW/mm2 on the first recording session; in later sessions, dural thickening was visible and changes in firing rate were smaller, so power was increased (maximum 3 mW/mm2) to give mean spontaneous rate increases of approximately $\sim$5 spk/s in that recording session. Optogenetic light spot diameter was 400–700 μm (FWHM) as measured by imaging the delivered light on the cortical surface.

Spike waveforms were sorted after the experiment using OfflineSorter (Plexon, Inc.). Single units (SU) were identified as waveform clusters that showed clear and stable separation from noise and other clusters, unimodal width distributions, and interspike interval histograms consistent with cortical neuron absolute and relative refractory periods. To compute neurons' visual responses, we counted spikes over a 175-ms period beginning 25 ms after stimulus onset. Zero-contrast responses were computed from the 175 ms period immediately preceding stimulus onset.

Monkeys

A detailed description of the experimental methods is given in (Nassi et al., 2015). In brief, two adult male rhesus macaques (Macaca mulatta) were each implanted with a custom titanium head post and silicone-based artificial dura recording chamber over V1. We injected a VSVg-pseudotyped lentivirus carrying the C1V1-EYFP gene behind the 1.3 kb CaMKIIα promoter (lenti-CaMKIIα-C1V1E162T-ts-EYFP; titer 3 × 1010 TU/ml) into a single location in V1 in each of the two monkeys (monkeys A and M) while they were anesthetized and secured in a stereotactic frame. Injections of the same viral construct were made into V1 of one additional monkey in order to assess specificity of viral expression to E neurons. In a second, distant location in V1 of monkey A, we injected an adeno-associated virus carrying the eArch3.0-EYFP gene behind the 1.3 kb CamKIIα promoter (AAV5-CamKIIα-eArch3.0-ts-EYFP; titer 4 × 1012 VP/ml).

Monkeys were alert and head restrained during all experiments. Single- and multi-unit waveforms were recorded using thin glass-coated tungsten electrodes (Alpha Omega, Engineering Inc., Nazareth, Israel). After isolating a single-unit or multi-unit cluster, we first assessed sensitivity to optogenetic stimulation. We randomly interleaved different stimulation intensities. Stimulation on each trial was continuous and lasted for 200 ms. Each condition was repeated at least five times. For a subset of light-sensitive units, we proceeded to measure responses to simultaneous optogenetic stimulation and visual stimulus presentation.

The presented visual stimuli consisted of circular patches of drifting sinusoidal gratings whose size was matched to each recorded unit's preferred size (median diameter = 0.63deg), of mean luminance matching the surround (42 cd-m2) at the optimal orientation and spatial and temporal frequencies. Each stimulus condition was presented at least five times. For simultaneous optogenetic stimulation and visual stimulus presentation, we varied the contrast of the presented gratings in log steps (0%, 6%, 12%, 25%, 50%, and 99%) and stimulated with four different intensities including zero. Visual contrast and stimulation intensity were randomly interleaved. Here, we show results only for zero and the largest intensity, but similar results hold also for intermediate intensities.

Mathematical methods

Network model. We investigated large randomly connected networks of E and I rate neurons. The single single neuron response function (firing rate vs input current relationship) was taken to be the f-I curve of leaky integrate-and-fire neurons driven by white noise (Amit and Brunel, 1997; Amit and Tsodyks, 1991; Sanzeni et al., 2020; Siegert, 1951). Specifically, the firing of a neuron belonging to population $A \in [E, I]$ in response to an input x, is given by

$$r_A = \phi_A(x) = \left[\tau_{rp} + \tau_A \sqrt{\pi} \int_{u_{min,A}}^{u_{max,A}} e^{u^2} \left(1 + \mathrm{erf}(u)\right) du\right]^{-1}, \tag{3}$$

with

$$u_{max,A} = \frac{\theta - x}{\sigma_A}, \quad u_{min,A} = \frac{V_r - x}{\sigma_A}. \tag{4}$$

In Eq. 3, τ_{rp} indicates the single neuron refractory period so that $1/\tau_{rp}$ is the maximal single neuron firing rate; τ_A is the membrane time constant; σ_A is a parameter controlling the smoothness of the transfer function. Throughout the paper, we assumed the values $\tau_{rp} = 2$ms, $\theta = 20$mV, $V_r = 10$mV, $\tau_{E,I} = 20, 10$ms and $\sigma_{E,I} = 10$mV.

Indicating with r_A^i and μ_A^i the rate and the total input (recurrent + feedforward) of the i-th cell in population $A \in [E, I]$, activity in the network evolves in time according to the equation

$$\tau_A \frac{dr_A^i}{dt} = -r_A^i + \phi_A\left(\mu_A^i\right), \quad \mu_A^i = \sum_{B \in [E,I,X]} \sum_{j=1}^{N_B} c_{AB}^{ij} W_{AB}^{ij} \tau_A r_B^j \tag{5}$$

Here c_{ij} is the adjacency matrix and N_B is the number of neurons in population B where (X) indicates the externally driven input, *i.e.* visual input. In the theoretical analysis of the model (mean field theory described below and simulations shown in Figures 4 and 5 and in supplementary information), the network connectivity was assumed to follow an Erdös–Rényi statistics, i.e. each element of the adjacency matrix c_{ij} was 1 with probability p and 0 otherwise. In fitting the model to data (Figure 6), we considered more general statistics by generating c_{ij} as follows: for each row i, a random number of elements K_i (Gaussian distributed, with mean K and standard deviation $CV_K\, K$) was fixed to 1, the other elements were fixed to 0. Throughout the paper, we assumed uniform Ws (specifically, $W_{AB}^{ij} = W_{AB}$) and $N_X = N_E = 4\,N_I = 10K$. The matrix W_{AB} was parameterized as follows

$$\begin{aligned} W_{EE} &= J, \quad W_{EI} = J g_E, \quad W_{EX} = (G_I g_E N_I/N_E - G_E) W_{EE} \\ W_{IE} &= J/\beta, \quad W_{II} = J g_I/\beta, \quad W_{IX} = (G_I g_I N_I/N_E - G_E) W_{IE}. \end{aligned} \tag{6}$$

These definitions were chosen so that the parameters G_I and G_E give the gain of the E and I population in the strong coupling limit (see (Sanzeni et al., 2020)).

With optogenetic input, the network dynamics becomes

$$\tau_A \frac{d(r_A^i + \Delta r_A^i)}{dt} = -r_A^i - \Delta r_A^i + \phi_A\left(\mu_A^i + \Delta\mu_A^i + \lambda_A^i\right), \quad \Delta\mu_A^i = \sum_{B\in[E,I,X]} \sum_{j=1}^{N_B} c_{AB}^{ij} W_{AB}^{ij} \tau_A \Delta r_B^j, \tag{7}$$

where λ_A^i represents the optogenetic input of the i-th neuron in the network, which varies according to the opsin expression in that neuron, while Δr_A^i and $\Delta\mu_A^i$ indicate the change in rate and in the recurrent inputs produced by optogenetic stimuli.

For the discrete ring model, the recurrent connection probability and all synaptic weights are scaled by a structured component that depends on the distance between two oriented locations on the ring or between the input orientation and the ring orientation (between $0° - 180°$). This structured factor has the form

$$\frac{\exp\left[\cos(2\Delta\theta)/(2\sigma)^2\right]}{\exp\left[1/(2\sigma)^2\right]}. \tag{8}$$

The synaptic weights between a tuned external population and populations on the ring are shaped by a width σ_X. The recurrent connection probability and weights are shaped by common widths σ_E for excitation and σ_I for inhibition.

Network simulations. Simulations of the network dynamics were performed in python. For each parameter set, we generated one realization of the network structure (adjacency matrix C_{AB}^{ij}, firing of the feedforward population r_X^i (generated from a Gaussian distribution with mean r_X and standard deviation $\sigma_X = 0.2\, r_X$), opsin expression λ_A^i (generated from a lognormal distribution with mean λ and standard deviation σ_λ) and simulated the network dynamics given by Eqs. (5) and (7) using the explicit Runge-Kutta method of order 5 (implemented with the function solve_ivp of the python package scipy.integrate). The network dynamics was run for a total simulated time of $200\tau_E = 4$s and computed rates were stored every $\tau_I/3 = 3.33$ms. The firing rate of a cell was obtained by measuring its average rate over the simulation time (excluding the initial $10\tau_E = 0.2$s.) To measure optogenetic responses, we simulated the same realization of network structure twice, once with $\lambda_A^i \neq 0$ and once with $\lambda_A^i = 0$.

Fitting procedure. Predictions of the model depend on a list of unknown parameters (indicated in what follows as Θ) related to: connectivity (J, K, β, g_E, g_I, CV_K); external inputs (G_E, G_I and one value of r_X per each contrast); and optogenetic inputs (λ, σ_λ). To fit these parameters to experimental data, we assumed $G_I = 2G_E$, consistent with the fact that in vivo, I firing rates are typically about twice E firing rates (e.g. (Sanzeni et al., 2020)). This choice is motivated by the fact that experimental recordings do not distinguish cell types; this limitation makes our model under-constrained by the data (since the same average firing could be obtained by an infinite combination of firing rates of E and I cells). Without loss of generality, we assumed $G_E = 1$ (any other choice can be obtained by rescaling r_X) and $K = 500$ (any other choice can be obtained by rescaling $J \sim 1/\sqrt{K}$). Moreover, we took $\sigma_X = 0$, and controlled variability in the feedforward input by varying CV_K.

With the above assumptions, model parameters were inferred as follows. First, we measured optogenetic responses in a large number (457964) of network simulations; each one obtained assuming model parameters generated randomly and uniformly in the intervals

$$\begin{gathered} g_E \in [3,7]\,, \quad g_I \in [2, g_E - 0.5]\,, \quad \log_{10}(\beta) \in [-1,1]\,, \\ \log_{10}(\sigma_\lambda/\lambda) \in [-1,1]\,, \quad \log_{10}(J) \in [-5,-3]\,, \quad \log_{10}(r_X) \in [-0.3, 1.7]\,, \\ \log_{10}(\lambda) \in [-0.3, 1.7]\,, \quad \log_{10}(CV_K) \in [-3.5, -0.5]. \end{gathered} \tag{9}$$

Starting from the results of these simulations, we trained a multilayer perceptron (MLPRegressor from python package sklearn.neural_network; specifics were activation=ReLU, hidden_layer_sizes=(100, 150, 50),)) to predict the mapping $f(\Theta)$ between model parameters (Θ) and moments of the optogenetic response (r, Δr, σ_r, $\sigma_{\Delta r}$, $\mathrm{Cov}(r, \Delta r)$). The learned multilayer perceptron was able to predict the outcome of simulations with high accuracy (with 20% of held-out simulations, the crossvalidated performance had $R^2 = 0.965$). We then fitted model parameters Θ to data by minimizing the squared error between predictions given by $f(\Theta)$ and moments of the optogenetic responses measured in experiments (least_squares function from python package scipy.optimize).

Parameter	Symbol	Units	Figure 4 (weak, strong)	Figure 5	Figure 6 (comb., mice, monk.)
Mean E connections per neuron	K		2000	500	500
Coeff. of Variation of In-Degree	CV_K		$1/\sqrt{K}$	$1/\sqrt{K}$	(0.0056, 0.0258, 0.0070)
Connection probability	p		0.1	0.1	0.1
Number of excitatory neurons	N_E		K/p	K/p	K/p
Ratio of I to E Population Size	γ		0.25	0.25	0.25
E time constant	τ_E	ms	20	20	20
I time constant	τ_I	ms	10	10	10
Synaptic efficacy	J	mV	(0.01, 0.3)	(**C**) 0.6	(0.099, 0.07, 0.07)
Ratio of I-to-E over E-to-E	g_E		8	(**A,B,D**) 8	(8.10, 8.80, 8.11)
Ratio of I-to-I over E-to-I	g_I		3	(**A,B,D**) 3	(6.82, 7.64, 4.70)
Ratio of E-to-E over E-to-I	β		1	1	(0.177, 0.256, 1.788)
Gain of the E population	G_E		from Eq. 6	1	1
Gain of the I population	G_I		from Eq. 6	2	2
Strength of FF input to E	W_{EX}		3	from Eq. 6	from Eq. 6
Strength of FF input to I	W_{IX}		(0.5,2.5)	from Eq. 6	from Eq. 6
Mean FF input rates	r_X	spk/s	20	(**C**) 10, (**D**) 17.97	([10.41 - 24.9] & [31.96 - 82.21], [3.69 - 9.74], [10.87 - 36.51])
Std FF input rates	σ_X	spk/s	$0.2r_X$	$0.2r_X$	0
Mean optogenetic input strength	λL	mV	20	20	(2.38 & 7.31, 12.69, 3.31)
Std / mean of opsin expression	σ_λ/λ		1	(**A,B**) 2.15, (**C**) 1	(1.62, 0.90, 8.57)

Table 1: List of network parameters.

Mean field theory. For $N_B \gg 1$, μ^i_A and $\Delta\mu^i_A$ in Eqs. (5) and (7) are well approximated by correlated random Gaussian variables with

$$\begin{aligned}
\mu_A &= \tau_A \sum_{B\in[E,I,X]} W_{AB} K_B r_B\,, \quad \sigma^2_{\mu_A} = \tau^2_A \sum_{B\in[E,I,X]} W^2_{AB} K_B \left[\sigma^2_{r_B} + (1-p) r^2_B\right]\,, \\
\Delta\mu_A &= \tau_A \sum_{B\in[E,I]} W_{AB} K_B \Delta r_B\,, \quad \sigma^2_{\Delta\mu_A} = \tau^2_A \sum_{B\in[E,I]} W^2_{AB} K_B \left[\sigma^2_{\Delta r_B} + (1-p)\Delta r^2_B\right]\,, \\
\mathrm{Cov}(\mu_A, \Delta\mu_A) &= \tau^2_A \sum_{B\in[E,I]} W^2_{AB} K_B \left[\mathrm{Cov}(r_B, \Delta r_B) + (1-p) r_B \Delta r_B\right]\,.
\end{aligned} \tag{10}$$

In Eq. (10), μ_A ($\Delta\mu_A$) and σ_{μ_A} ($\sigma_{\Delta\mu_A}$) are the mean and variance of μ^i_A ($\Delta\mu^i_A$). The covariance

$\mathrm{Cov}(\mu_A, \Delta\mu_A)$ is computed starting from

$$\mu_A^i + \Delta\mu_A^i = \sum_{B\in[E,I,X]} \tau_A W_{AB} \sum_{j=1}^{N_B} c_{AB}^{ij} \left(r_B^j + \Delta r_B^j \right) , \tag{11}$$

applying the identity $Var(X+Y) = Var(X) + Var(Y) + 2\mathrm{Cov}(X,Y)$ to both sides and using the central limit theorem. Note that optogenetic input affects firing differently than feed-forward and recurrent inputs. In fact, firing depends on the whole distribution of λ_A^i but only on the first two moments of feed-forward and recurrent inputs.

If the network dynamics settle in a stable fixed point, the statistics of currents and rates in the network are related by

$$r_A^i = \phi_A \left[\mu_A^i \right] , \quad r_A^i + \Delta r_A^i = \phi_A \left[\lambda_A^i + \mu_A^i + \Delta\mu_A^i \right] , \tag{12}$$

and moments of the rates can be computed by self-consistently solving a set of ten equations

$$\begin{cases} r_A = \int_{-\infty}^{\infty} d\mu_A^i P(\mu_A^i)\phi(\mu_A^i) \, , \\ \sigma_{r_A}^2 = \int_{-\infty}^{\infty} d\mu_A^i P(\mu_A^i)\phi(\mu_A^i)^2 - r_A^2 \, , \\ \Delta r_A = \int_{-\infty}^{\infty} d\mu_A^i d\Delta\mu_A^i d\lambda_A^i P(\mu_A^i, \Delta\mu_A^i) P(\lambda_A^i) \left[\Delta\phi_A^i \right] \, , \\ \sigma_{\Delta r_A}^2 = \int_{-\infty}^{\infty} d\mu_A^i d\Delta\mu_A^i d\lambda_A^i P(\mu_A^i, \Delta\mu_A^i) P(\lambda_A^i) \left[\Delta\phi_A^i \right]^2 - \Delta r_A^2 \, , \\ \mathrm{Cov}(r_A, \Delta r_A) = \int_{-\infty}^{\infty} d\mu_A^i d\Delta\mu_A^i d\lambda_A^i P(\mu_A^i, \Delta\mu_A^i) P(\lambda_A^i) \left[\phi(\mu_A^i) - r_A \right] \left[\Delta\phi_A^i - \Delta r_A \right] \, , \end{cases} \tag{13}$$

with $\Delta\phi_A^i = \phi(\mu_A^i + \Delta\mu_A^i + \lambda_A^i) - \phi(\mu_A^i)$.

If the fixed point of the network dynamics is unstable, rates and currents become time-dependent. One can again use a Gaussian ansatz, in which the total inputs at different times ($\mu_A^i(s)$ and $\mu_A^i(t)$) are jointly Gaussian with means $\mu_A(s), \mu_A(t)$ and covariance $\sigma_{\mu_A}^2(s,t)$. Analogously, rates at different times ($r_A^i(s)$ and $r_A^i(t)$) are jointly Gaussian with means $r_A(s), r_A(t)$ and covariance $\sigma_{r_A}^2(s,t)$. At any given pair of time points t and s, the moments of rates and currents are related by Eq. (10). The temporal evolution of the moments can be computed using standard techniques (Sompolinsky et al., 1988). In particular, indicating with $C_{r_A}(s,t) = N_A^{-1} \sum_{i=1}^{N_A} r_A^i(s) r_A^i(t)$ the autocorrelation function of the rates of population A at times s and t, we can define the autocovariance function of the rates as

$$\sigma_{r_A}^2(s,t) = C_{r_A}(s,t) - r_A(s) r_A(t) \, . \tag{14}$$

By multiplying Eq. (7) at times s and t, averaging over cells in population A, and rearranging terms we see that $C_{r_A}(s,t)$ evolves in time according to

$$\left(\tau_A \frac{d}{ds} + 1 \right) \left(\tau_A \frac{d}{dt} + 1 \right) C_{r_A}(s,t) = C_{\phi_A[\mu_A]}(s,t) \, .$$

We can repeat the above steps to derive equations for the evolution in time of all five order parameters per population of neurons (mean, autocovariance, and cross-covariance of rates without and with optogenetic stimuli). Defining

$$M_{\phi_A[\mu_A]}(t) = N_A^{-1} \sum_{i=1}^{N_A} \phi_A \left[\mu_A^i(t) \right]$$

and

$$M_{\phi_A[\tilde{\mu}_A]}(t) = N_A^{-1} \sum_{i=1}^{N_A} \phi_A \left[\mu_A^i(t) + \Delta\mu_A^i(t) + \lambda_A^i \right] \, ,$$

the order parameters evolve in time according to

$$
\begin{aligned}
\left(\tau_A \frac{d}{dt}+1\right) r_A = M_{\phi_A[\mu_A]}(t)\,, \quad \left(\tau_A \frac{d}{dt}+1\right) \Delta r_A = M_{\phi_A[\tilde{\mu}_A]}(t) - M_{\phi_A[\mu_A]}(t)\ , \\
\left(\tau_A \frac{d}{ds}+1\right)\left(\tau_A \frac{d}{dt}+1\right) C_{r_A}(s,t) = C_{\phi_A[\mu_A]}(s,t)\ , \\
\left(\tau_A \frac{d}{ds}+1\right)\left(\tau_A \frac{d}{dt}+1\right) C_{\Delta r_A}(s,t) = C_{\phi_A[\tilde{\mu}_A]}(s,t) - \mathrm{Corr}(\phi_A\left[\mu_A\right], \phi_A\left[\tilde{\mu}_A\right])(s,t) \\
- \mathrm{Corr}(\phi_A\left[\tilde{\mu}_A\right], \phi_A\left[\mu_A\right])(s,t) + C_{\phi_A[\mu_A]}(s,t)\ , \\
\left(\tau_A \frac{d}{ds}+1\right)\left(\tau_A \frac{d}{dt}+1\right) \mathrm{Corr}(r_A, \Delta r_A)(s,t) = \mathrm{Corr}(\phi_A\left[\mu_A\right], \phi_A\left[\tilde{\mu}_A\right])(s,t) - C_{\phi_A[\mu_A]}(s,t)\,.
\end{aligned}
\tag{15}
$$

Numerical solution of Mean field equations We solve for the stationary statistics of the chaotic network by explicitly integrating Eq. (15) in time until the order parameters converge to a steady state. We start with initial values for $r_A(0)$ and $\Delta r_A(0)$ as well as initial functions for $C_{r_A}(s,0) = C_{r_A}(0,t)$, $C_{\Delta r_A}(s,0) = C_{r_A}(0,t)$, and $\mathrm{Corr}(r_A, \Delta r_A)(s,0) = \mathrm{Corr}(r_A, \Delta r_A)(0,t)$. The means $r_A(t)$ and $\Delta r_A(t)$ are integrated with the usual explicit RK4 method, but the autocovariance and cross-covariance functions are integrated on a grid with an explicit Euler-like method derived using forward finite differences. For example, when integrating $C_{r_A}(s,t)$ we can express $C_{r_A}(s+dt, t+dt)$ in terms of $r_A(s)$, $r_A(t)$, $C_{r_A}(s,s)$, $C_{r_A}(t,t)$, $C_{r_A}(s,t)$, $C_{r_A}(s{+}dt,t)$, and $C_{r_A}(s,t{+}dt)$. Assuming that we have solved for the order parameters at all time points $(s \le t_0, t \le t_0)$, we can explicitly solve for the order parameters at time points $(s \le t_0 + dt, t_0 + dt)$ and $(t_0 + dt, t \le t_0 + dt)$. Furthermore, we can ignore portions of the grid for $|s - t| > T$ for some T that is much longer than the width of the steady state autocorrelation function in order to reduce the computation time required for the integration.

A useful simplification of the equations governing the evolution of the order parameters is to calculate quantities such as $C_{\phi_A[\mu_A]}(s,t)$ by setting the mean and variance of μ_A^i at times s and t equal to the mean and variance at time $\max(s,t)$. While this change increases the required integration time for the order parameters to reach the steady state, it reduces the number of dependent variables needed to calculate quantities like $C_{\phi_A[\mu_A]}(s,t)$ from five to three. With only three dependent variables, it becomes feasible to pre-calculate these autocorrelation functions on a 3D grid and interpolate between the pre-computed points in order to accelerate the integration. In addition, we can precompute quantities such as $M_{\phi_A[\mu_A]}(t)$, which only depends on the mean and variance of the inputs, on a 2D grid and interpolate them as well.

The cross-correlation function $\mathrm{Corr}(\phi_A\left[\mu_A\right], \phi_A\left[\tilde{\mu}_A\right])(s,t)$ cannot be pre-computed and interpolated since the means and variances of the cross-correlated variables are not equal. However, since we have already interpolated $M_{\phi_A[\mu_A]}(t)$ and $M_{\phi_A[\tilde{\mu}_A]}(t)$ we can efficiently calculate the cross-correlation function from a single Gaussian integral.

Indicating with $M_{\phi_A[\mu_A]}\left[\mu_A(t), \sigma^2_{\mu_A}(t,t)\right]$ and $M_{\phi_A[\tilde{\mu}_A]}\left[\mu_A(t) + \Delta\mu_A(t), \sigma^2_{\mu_A+\Delta\mu_A}(t,t)\right]$ the precomputed mean rates as a function of the means and variances of the inputs, we can express the cross-correlation function as

$$
\begin{aligned}
&\mathrm{Corr}(\phi_A\left[\mu_A\right], \phi_A\left[\tilde{\mu}_A\right])(s,t) = \\
&\quad \int_{-\infty}^{\infty} Dx M_{\phi_A[\mu_A]}\left[\mu_A(s) + \mathrm{sign}\,(\eta)\, x\sqrt{|\eta|}\,\sigma_{\mu_A}(s,s), (1-|\eta|)\sigma^2_{\mu_A}(s,s)\right] \\
&\qquad \times M_{\phi_A[\tilde{\mu}_A]}\left[\mu_A(t) + \Delta\mu_A(t) + x\sqrt{|\eta|}\sigma_{\mu_A+\Delta\mu_A}(t,t), (1-|\eta|)\sigma^2_{\mu_A+\Delta\mu_A}(t,t)\right]\ ,
\end{aligned}
\tag{16}
$$

with the Pearson correlation coefficient η defined as

$$
\eta = \frac{\mathrm{Cov}(\mu_A, \mu_A + \Delta\mu_A)(s,t)}{\sigma_{\mu_A}(s,s)\sigma_{\mu_A+\Delta\mu_A}(t,t)}\ .
$$

Acknowledgments

This work was supported by the NIMH Intramural Research Program and by NIH BRAIN U01 NS108683 (to N.B., M.H.H., K.D.M. and J.H.R.). A.S., A.P., T.H.N. and K.D.M. were partially supported by the Gatsby Charitable Foundation GAT3708 and by NSF Grant DBI-1707398. A.P. would like to acknowledge the support of the Swartz Foundation Fellowship for Theory in Neuroscience 2019-4 and 2020-6. T.H.N. would like to acknowledge the support by NIH grant 5R01EY029999. J.J.N was funded, in part, by NIH grant R01-EY021827. J. Luo was supported, in part, by the Fiona and Sanjay Jha Chair in Neuroscience. This work used the computational resources of the Duke Compute Cluster.

References

Ahmadian, Y. and K. D. Miller (2021). What is the dynamical regime of cerebral cortex? *Neuron 109*, 3373–3391.

Ahmadian, Y., D. B. Rubin, and K. D. Miller (2013). Analysis of the stabilized supralinear network. *Neural Computation 25*(8), 1994–2037.

Akitake, B., H. M. Douglas, P. K. LaFosse, C. E. Deveau, A. J. Li, L. N. Ryan, S. P. Duffy, Z. Zhou, Y. Deng, and M. H. Histed (2022). Amplified cortical neural responses as animals learn to use novel activity patterns. *bioRxiv*.

Amit, D. J. and N. Brunel (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex 7*, 237–252.

Amit, D. J. and M. V. Tsodyks (1991). Quantitative study of attractor neural network retrieving at low spike rates I: Substrate – spikes, rates and neuronal gain. *Network 2*, 259–274.

Andrei, A. R., S. Pojoga, R. Janz, and V. Dragoi (2019). Integration of cortical population signals for visual perception. *Nat. Commun. 10*(1), 1–13.

Barth, A. L. and J. F. Poulet (2012). Experimental evidence for sparse firing in the neocortex. *Trends Neurosci. 35*, 345–355.

Bouvier, G., Y. Senzai, and M. Scanziani (2020). Head movements control the activity of primary visual cortex in a luminance-dependent manner. *Neuron 108*, 500–511.

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci. 8*, 183–208.

Brunel, N. and V. Hakim (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comp. 11*, 1621–1671.

Brunel, N. and X.-J. Wang (2003). What determines the frequency of fast network oscillations with irregular neural discharges? *J. Neurophysiol. 90*, 415–430.

Buzsáki, G. and K. Mizuseki (2014, 04). The log-dynamic brain: how skewed distributions affect network operations. *Nature reviews. Neuroscience 15*(4), 264–278.

Carandini, M. and D. J. Heeger (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience 13*(1), 51–62.

Chen, S. C. Y., G. Benvenuti, Y. Chen, S. Kumar, C. Ramakrishnan, K. Deisseroth, W. S. Geisler, and E. Seidemann (2022). Similar neural and perceptual masking effects of low-power optogenetic stimulation in primate V1. *Elife 11*, 1–21.

Chernov, M. M., R. M. Friedman, G. Chen, G. R. Stoner, and A. W. Roe (2018). Functionally specific optogenetic modulation in primate visual cortex. *Proc. Natl. Acad. Sci. U. S. A. 115*(41), 10505–10510.

Chettih, S. N. and C. D. Harvey (2019). Single-neuron perturbations reveal feature-specific competition in V1. *Nature 567*, 334–340.

Compte, A., C. Constantinidis, J. Tegnér, S. Raghavachari, M. Chafee, P. S. Goldman-Rakic, and X.-J. Wang (2003). Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *J. Neurophysiol. 90*, 3441–3454.

Cossell, L., M. F. Iacaruso, D. R. Muir, R. Houlton, E. N. Sader, H. Ko, S. B. Hofer, and T. D. Mrsic-Flogel (2015). Functional organization of excitatory synaptic strength in primary visual cortex. *Nature 518*, 399–403.

Dalgleish, H. W., L. E. Russell, A. M. Packer, A. Roth, O. M. Gauld, F. Greenstreet, E. J. Thompson, and M. Häusser (2020). How many neurons are sufficient for perception of cortical activity? *eLife 9*, e58889.

Davis, Z. W., L. Muller, J. Martinez-Trujillo, T. Sejnowski, and J. H. Reynolds (2020). Spontaneous travelling cortical waves gate perception in behaving primates. *Nature 587*, 432–436.

Dräger, U. C. (1974). Autoradiography of tritiated proline and fucose transported transneuronally from the eye to the visual cortex in pigmented and albino mice. *Brain Res. 82*(2), 284–292.

Engelhard, B., R. Darshan, N. Ozeri-Engelhard, Z. Israel, U. Werner-Reiss, D. Hansel, H. Bergman, and E. Vaadia (2019). Neuronal activity and learning in local cortical networks are modulated by the action-perception state. *bioRxiv*.

Engelken, R., F. Wolf, and L. F. Abbott (2020). Lyapunov spectra of chaotic recurrent neural networks. *arXiv preprint, arXiv:2006.02427*.

Haider, B., A. Duque, A. R. Hasenstaub, and D. A. McCormick (2006). Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *J. Neurosci. 26*, 4535–4545.

Harish, O. and D. Hansel (2015). Asynchronous Rate Chaos in Spiking Neuronal Circuits. *PLoS Comput Biol 11*, e1004266.

Histed, M. H. (2018). Feedforward inhibition allows input summation to vary in recurrent cortical networks. *eNeuro 5*(1).

Hromádka, T., M. R. DeWeese, and A. M. Zador (2008, 01). Sparse representation of sounds in the unanesthetized auditory cortex. *PLOS Biology 6*(1), 1–14.

Hubel, D. and T. Wiesel (1977). Ferrier lecture. functional architecture of macaque monkey visual cortex. *Proc R Soc Lond B Biol Sci. 198*, 1–59.

Kadmon, J. and H. Sompolinsky (2015). Transition to chaos in random neuronal networks. *Phys. Rev. X 5*, 041030.

Kayser, A. S. and K. D. Miller (2002). Opponent inhibition: A developmental model of layer 4 of the neocortical circuit. *Neuron 33*, 131–142.

Kim, C. K., A. Adhikari, and K. Deisseroth (2017). Integration of optogenetics with complementary methodologies in systems neuroscience. *Nat Rev Neurosci 18*, 222–235.

Ko, H., S. B. Hofer, B. Pichler, K. A. Buchanan, P. J. Sjostrom, and T. D. Mrsic-Flogel (2011). Functional specificity of local synaptic connections in neocortical networks. *Nature 473*, 87–91.

Li, Y. T., L. A. Ibrahim, B. H. Liu, L. I. Zhang, and H. W. Tao (2013). Linear transformation of thalamocortical input by intracortical excitation. *Nat. Neurosci. 16*, 1324–1330.

Lien, A. D. and M. Scanziani (2013). Tuned thalamic excitation is amplified by visual cortical circuits. *Nat. Neurosci. 16*, 1315–1323.

Liska, J. P., D. P. Rowley, T. T. K. Nguyen, J.-O. Muthmann, D. A. Butts, J. L. Yates, and A. C. Huk (2022). Running modulates primate and rodent visual cortex via common mechanism but quantitatively distinct implementation. *bioRxiv*.

Liu, B. H., Y. T. Li, W. P. Ma, C. J. Pan, L. I. Zhang, and H. W. Tao (2011). Broad inhibition sharpens orientation selectivity by expanding input dynamic range in mouse simple cells. *Neuron 71*, 542–554.

Mattis, J., K. M. Tye, E. A. Ferenczi, C. Ramakrishnan, D. J. O'Shea, R. Prakash, L. A. Gunaydin, M. Hyun, L. E. Fenno, V. Gradinaru, O. Yizhar, and K. Deisseroth (2011). Principles for applying optogenetic tools derived from direct comparative analysis of microbial opsins. *Nat Methods 9*, 159–172.

Muller, L., F. Chavane, J. Reynolds, and T. J. Sejnowski (2018). Cortical travelling waves: mechanisms and computational principles. *Nat Rev Neurosci 19*, 255–268.

Nassi, J. J., M. C. Avery, A. H. Cetin, A. W. Roe, and J. H. Reynolds (2015). Optogenetic activation of normalization in alert macaque visual cortex. *Neuron 86*(6), 1504–1517.

O'Connor, D. H., S. P. Peron, D. Huber, and K. Svoboda (2010). Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron 67*(6), 1048 – 1061.

Ohki, K., S. Chung, Y. H. Ch'ng, P. Kara, and R. C. Reid (2005a). Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature 433*, 597–603.

Ohki, K., S. Chung, Y. H. Ch'ng, P. Kara, and R. C. Reid (2005b). Functional imaging with cellular resolution reveals precise microarchitecture in visual cortex. *Nature 433*(7026), 597–603.

Ozeki, H., I. M. Finn, E. S. Schaffer, K. D. Miller, and D. Ferster (2009). Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron 62*, 578–592.

O'Rawe, J. F., Z. Zhou, A. J. Li, P. K. LaFosse, H. C. Goldbach, and M. H. Histed (2022). Cortical suppression from excitatory input due to heterogeneous recurrent connectivity. *bioRxiv*.

Palmigiano, A., F. Fumarola, D. P. Mossing, N. Kraynyukova, H. Adesnik, K. D. Miller, and K. Miller (2020). Structure and variability of optogenetic responses identify the operating regime of cortex. *bioRxiv*, 2020.11.11.378729.

Renart, A., J. de la Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K. D. Harris (2010). The asynchronous state in cortical circuits. *Science 327*, 587–590.

Reynolds, J. H. and D. J. Heeger (2009, 2022/07/11). The normalization model of attention. *Neuron 61*(2), 168–185.

Rossi, L. F., K. D. Harris, and M. Carandini (2020). Spatial connectivity matches direction selectivity in visual cortex. *Nature 588*, 648–652.

Roxin, A., N. Brunel, D. Hansel, G. Mongillo, and C. van Vreeswijk (2011). On the distribution of firing rates in networks of cortical neurons. *Journal of Neuroscience 31*(45), 16217–16226.

Rubin, D. B., S. D. Van Hooser, and K. D. Miller (2015). The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron 85*, 402–417.

Sadeh, S. and C. Clopath (2020). Theory of neuronal perturbome in cortical networks. *Proc Natl Acad Sci U S A 117*, 26966–26976.

Sanzeni, A., B. Akitake, H. C. Goldbach, C. E. Leedy, N. Brunel, and M. H. Histed (2020). Inhibition stabilization is a widespread property of cortical networks. *Elife 9*.

Sanzeni, A., M. H. Histed, and N. Brunel (2020). Response nonlinearities in networks of spiking neurons. *PLoS Comput Biol 16*, e1008165.

Sanzeni, A., M. H. Histed, and N. Brunel (2022). Emergence of irregular activity in networks of strongly coupled conductance-based neurons. *Phys. Rev. X 12*, 011044.

Sherman, S. M. and R. W. Guillery (1998). On the actions that one nerve cell can have on another: distinguishing ”drivers” from ”modulators”. *Proc Natl Acad Sci U S A 95*, 7121–7126.

Siegert, A. J. F. (1951). On the first passage time probability problem. *Phys. Rev. 81*, 617–623.

Siepel, A. (2009). Phylogenomics of primates and their ancestral populations. *Genome Res. 19*(11), 1929–1941.

Softky, W. R. and C. Koch (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci. 13*, 334–350.

Sompolinsky, H., A. Crisanti, and H. J. Sommers (1988). Chaos in random neural networks. *Phys. Rev. Lett. 61*, 259–262.

Troyer, T. W., A. E. Krukowski, N. J. Priebe, and K. D. Miller (1998). Contrast-invariant orientation tuning in cat visual cortex: thalamocortical input tuning and correlation-based intracortical connectivity. *J. Neurosci. 18*, 5908–5927.

Tsodyks, M. V. and H. Markram (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl. Acad. Sci. USA 94*, 719–723.

Van Den Bergh, G., B. Zhang, L. Arckens, and Y. M. Chino (2010). Receptive-field properties of V1 and V2 neurons in mice and macaque monkeys. *J. Comp. Neurol. 518*(11), 2051–2070.

Van Hooser, S. D., J. A. Heimel, S. Chung, S. B. Nelson, and L. J. Toth (2005). Orientation selectivity without orientation maps in visual cortex of a highly visual mammal. *J. Neurosci. 25*, 19–28.

van Vreeswijk, C. and H. Sompolinsky (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science 274*, 1724–1726.

van Vreeswijk, C. and H. Sompolinsky (1998). Chaotic balanced state in a model of cortical circuits. *Neural Computation 10*, 1321–1371.

Wang, S., A. Palmigiano, K. D. Miller, and S. D. Van Hooser (2022). Targeted cortical stimulation reveals principles of cortical contextual interactions. *bioRxiv*.

Wildenberg, G. A., M. R. Rosen, J. Lundell, D. Paukner, D. J. Freedman, and N. Kasthuri (2021). Primate neuronal connections are sparse in cortex as compared to mouse. *Cell Rep 36*, 109709.

Yona, G., N. Meitav, I. Kahn, and S. Shoham (2016). Realistic Numerical and Analytical Modeling of Light Scattering in Brain Tissue for Optogenetic Applications(1,2,3). *eNeuro 3*(1).

Part I

Supplementary information

Theoretical results are preserved with different transfer functions

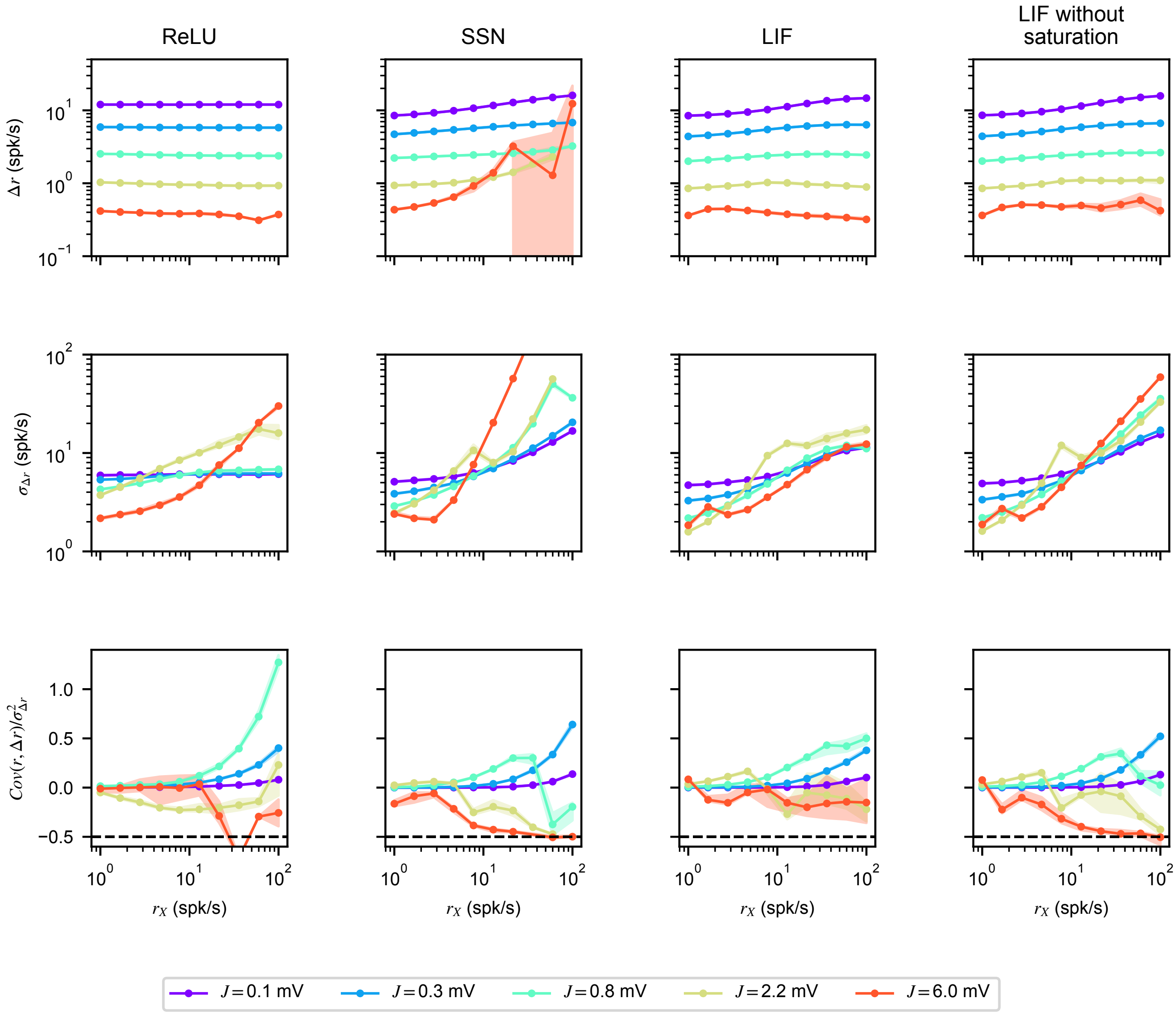

Figure S1: **Neuronal reshuffling emerges in network models with different single transfer functions.** Optogenetic responses obtained in numerical simulations of inhibitory networks (dynamics given determined by Eqs. (5) and (7), without E cells and with optogenetic stimuli targeting I cells) as a function of J (colors) and r_X, for different choices of the single neuron transfer function ϕ_A (titles). Plots are obtained averaging over 20 realizations of the network models (dots represent means, shaded regions are mean±sem). Transfer function used were: $\phi(x) = a\,[x]_+$ (ReLU, $a = 1.5$ mV^{-1}); $\phi(x) = a[x]_+^2$ (SSN, $a = 0.05$ mV^{-2}); $\phi(x)$ as in Eq. (3) with τ_{rp} =2ms (LIF) or $\tau_{rp} = 0$ (LIF without saturation).

Chaos is not required for negative covariance to appear

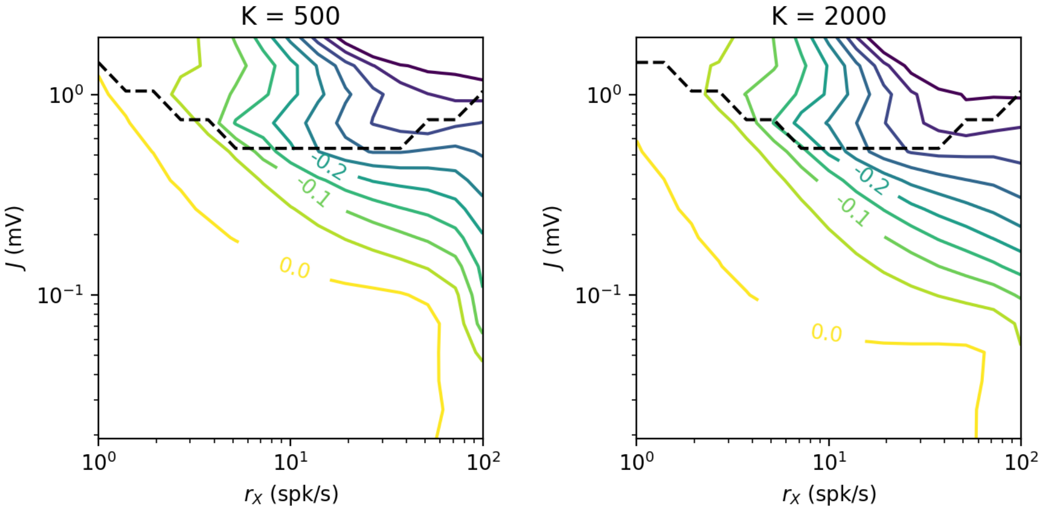

Figure S2: **Phase diagram of chaos and negative covariance.** Colored lines indicate contours of constant normalized covariance between 0 and -0.5 as a function of the feedforward input r_X and recurrent synaptic efficacies J. The dashed line indicates the transition to chaos; Below the dashed line the baseline system is stationary, and above the line the network is chaotic. In order to investigate the effect of finite network size, the left plot is calculated with $K = 500$ while the right plot is calculated with $K = 2000$. Increasing K tends to increase the magnitude of the normalized covariance while decreasing the region of chaos. All values are calculated from mean field theory. Network parameters are identical to those used in Figure 5 panels **A-B**.

Structured Connectivity Reshuffles Tuned Rates Even with Large Optogenetic Input

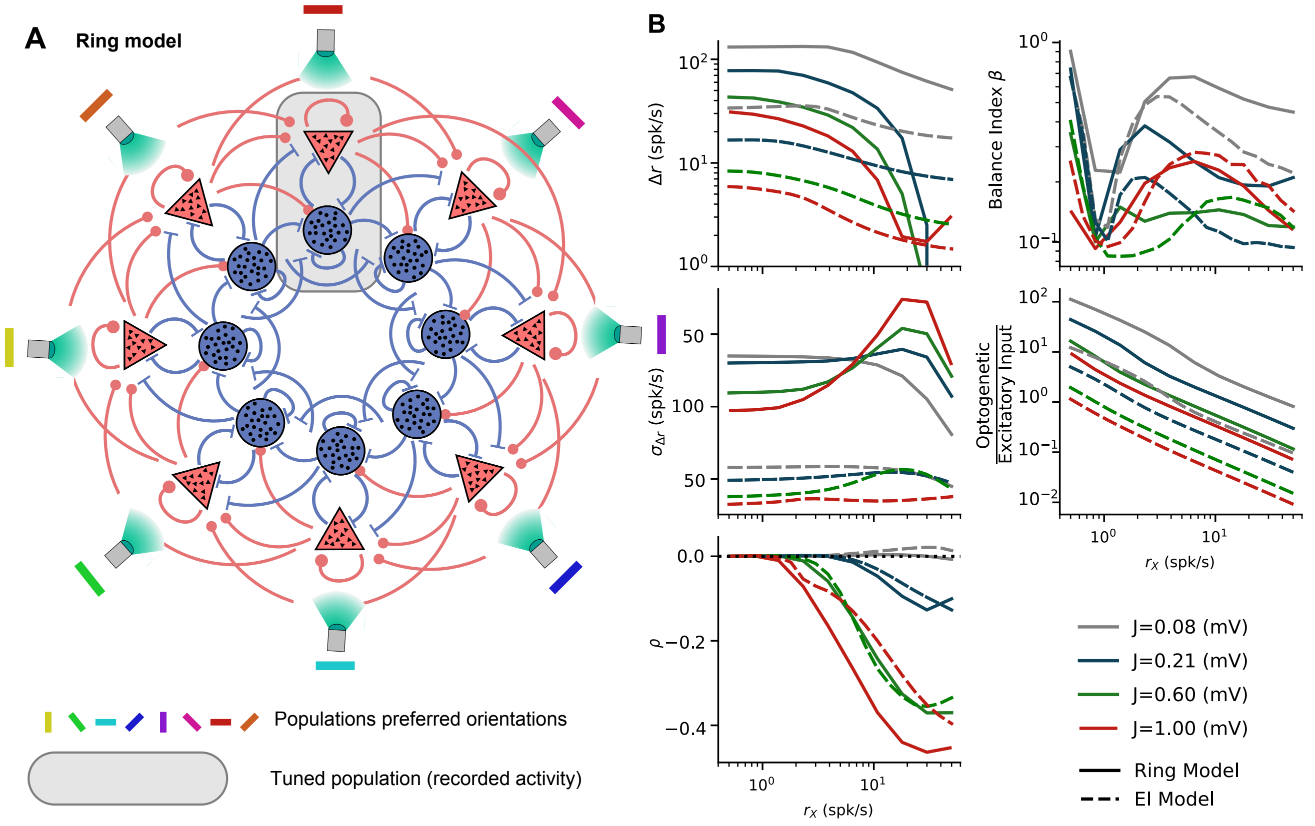

Figure S3: **Dependence of the statistics of tuned population response to optogenetic input on feed-forward input for a discrete ring model.** **A** Structure of the ring model. Each of the 8 blocks has a different preferred orientation (colored bars). All E populations are optogenetically stimulated, but only the population with the center orientation is recorded (gray box). **B** Statistics of tuned population response to optogenetic stimuli for a discrete ring model as a function of feedforward input r_X, for different values of recurrent synaptic efficacies J. Panels in the left column are arranged as in Figure 5. Panels in the right column show (from top to bottom) inferred values of the balance index (Ahmadian and Miller, 2021) and inferred values of the ratio between optogenetic input and the total visually-driven excitatory input. Dashed lines indicate values from Figure 5 panel **A** for the traditional EI model, whereas full lines indicate values for an equivalent ring model. The ring model was not chaotic at any of the plotted parameters. All values are calculated from mean field theory. Most network parameters are identical to those used in Figure 5 panel **A**. Ring model parameters are $N_{loc} = 8$, $K_{rec}(\text{peak}) = 200$, $\sigma_X, \sigma_E, \sigma_I = 20°, 20°, 30°$.

Parameter	Symbol	Units	Figure 5 (**A**)	Figure S3
Mean E connections per neuron	K		500	FF: 500 Rec. (Peak): 200
Coeff. of Variation of In-Degree	CV_K		$1/\sqrt{K}$	$1/\sqrt{K}$
Connection probability	p		0.1	(Peak) 0.1
Number of excitatory neurons	N_E		5000	2000 per Location
Ratio of I to E Population Size	γ		0.25	0.25
E time constant	τ_E	ms	20	20
I time constant	τ_I	ms	10	10
Synaptic efficacy	J	mV	(0.08,0.21,0.60,1.00)	(0.08,0.21,0.60,1.00)
Ratio of I-to-E over E-to-E	g_E		8	8
Ratio of I-to-I over E-to-I	g_I		3	3
Ratio of E-to-E over E-to-I	β		1	1
Gain of the E population	G_E		1	1
Gain of the I population	G_I		2	2
Strength of FF input to E	W_{EX}		from Eq. 6	from Eq. 6
Strength of FF input to I	W_{IX}		from Eq. 6	from Eq. 6
Mean FF input rates	r_X	spk/s	[0.5 - 50]	[0.5 - 50]
Std FF input rates	σ_X	spk/s	$0.2r_X$	$0.2r_X$
Mean optogenetic input strength	λ	mV	20	160
Std / mean of opsin expression	σ_λ/λ		2.15	2.15
Number of locations on ring	N_{loc}		-	8
Width of FF input	σ_X	deg°	-	20
Width of recurrent E	σ_E	deg°	-	20
Width of recurrent I	σ_I	deg°	-	30

Table 2: List of network parameters for Figure 5 Panel **A** and for Figure S3.