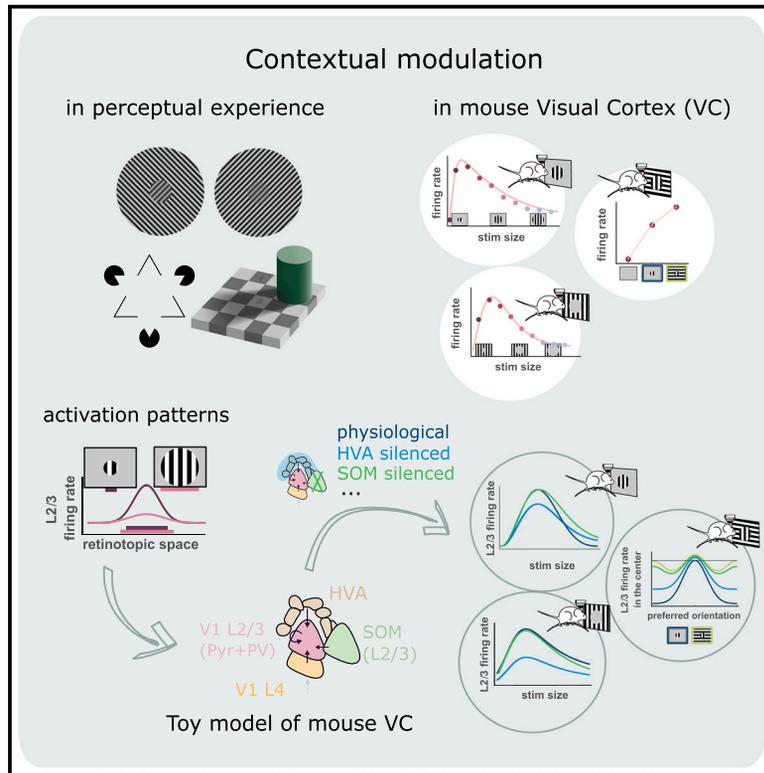


Contextual modulation emerges by integrating feedforward and feedback processing in mouse visual cortex

Graphical abstract



Authors

Serena Di Santo, Mario Dipoppa, Andreas Keller, Morgane Roth, Massimo Scanziani, Kenneth D. Miller

Correspondence

serenadisanto@ugr.es

In brief

Our brain's sensory systems use context to make sense of what we see, but the exact mechanisms behind this are not well understood. Di Santo et al. developed a unifying (toy) model of the mouse visual cortex that shows how brain cells work together to blend sensory input with contextual information.

Highlights

- One “toy” model explains three different types of contextual modulation
- The widths of spatial response patterns grow much more slowly than stimulus size
- Inverse responses depend on the geometry of feedback response fields and projections
- Summation of classical and inverse response accounts for surround facilitation



Article

Contextual modulation emerges by integrating feedforward and feedback processing in mouse visual cortex

Serena Di Santo,^{1,2,8,*} Mario Dipoppa,^{1,3} Andreas Keller,^{4,5,6} Morgane Roth,^{4,5,6} Massimo Scanziani,^{5,6} and Kenneth D. Miller^{1,7}

¹Center for Theoretical Neuroscience and Mortimer B Zuckerman Mind Brain Behavior Institute, Columbia University, New York City, NY 10027, USA

²Departamento de Electromagnetismo y Física de la Materia and Instituto Carlos I de Física Teórica y Computacional, Universidad de Granada, 18071 Granada, Spain

³Department of Neurobiology, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

⁴Department of Biomedicine, University of Basel, 4056 Basel, Switzerland

⁵Department of Physiology, University of California, San Francisco, San Francisco, CA 94143, USA

⁶Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, CA 94143, USA

⁷Department of Neuroscience, Swartz Program in Theoretical Neuroscience, Kavli Institute for Brain Science, College of Physicians and Surgeons and Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York City, NY 10027, USA

⁸Lead contact

*Correspondence: serenadisanto@ugr.es

<https://doi.org/10.1016/j.celrep.2024.115088>

SUMMARY

Sensory systems use context to infer meaning. Accordingly, context profoundly influences neural responses to sensory stimuli. However, a cohesive understanding of the circuit mechanisms governing contextual effects across different stimulus conditions is still lacking. Here we present a unified circuit model of mouse visual cortex that accounts for the main standard forms of contextual modulation. This data-driven and biologically realistic circuit, including three primary inhibitory cell types, sheds light on how bottom-up, top-down, and recurrent inputs are integrated across retinotopic space to generate contextual effects in layer 2/3. We establish causal relationships between neural responses, geometrical features of the inputs, and the connectivity patterns. The model not only reveals how a single canonical cortical circuit differently modulates sensory response depending on context but also generates multiple testable predictions, offering insights that apply to broader neural circuitry.

INTRODUCTION

When an edge or other feature appears in a visual scene, its meaning—is it an object boundary? An element of texture? A shadow?—must be inferred from the larger scene in which it is embedded. More generally, sensory systems must combine local features with context to infer meaning. Accordingly, context profoundly influences our perception, as is made strikingly clear by visual illusions. For example, the perceived contrast or luminance of a patch of the visual scene can be radically altered by whether the context makes the patch appear a small part of a larger object or an independent object, or in light or in shadow (Figures 1A–1C).

Context strongly impacts not only our perception but also neuronal responses. Some of the best-studied examples involve the effects of context on neurons in the primary visual cortex (V1).^{1–6} V1 neurons are primarily driven by stimuli only within a small region of visual space, known as the neuron's classical receptive field (CRF). However, the visual stimulus in the surrounding region (the neuron's "surround") can substantially

influence the cell's response. This contextual influence is likely to be critical to the parsing of objects and formation of a coherent perceptual representation of the visual scene.^{7–11}

Here, we develop a simple theoretical framework to address the question of how the properties of feedforward, feedback, and lateral input currents contribute to contextual modulation in V1. We focus on mouse as a model organism due to the wealth of available information on cell-type-specific physiological and anatomical properties. This allows us to arrive at a unified mechanistic picture of how these different inputs are integrated to produce diverse forms of contextual modulation.

Three exemplary paradigms have been widely used to study contextual modulation in V1 (Figures 1D–1F).

- (1) The response of a cell whose CRF is centered on a stimulus decreases when a similar stimulus is presented in the surround. This "surround suppression"^{5,12–16} may represent a discounting of visual input that can be well predicted from other parts of the visual scene, thus allowing an efficient neural representation.^{17–19}



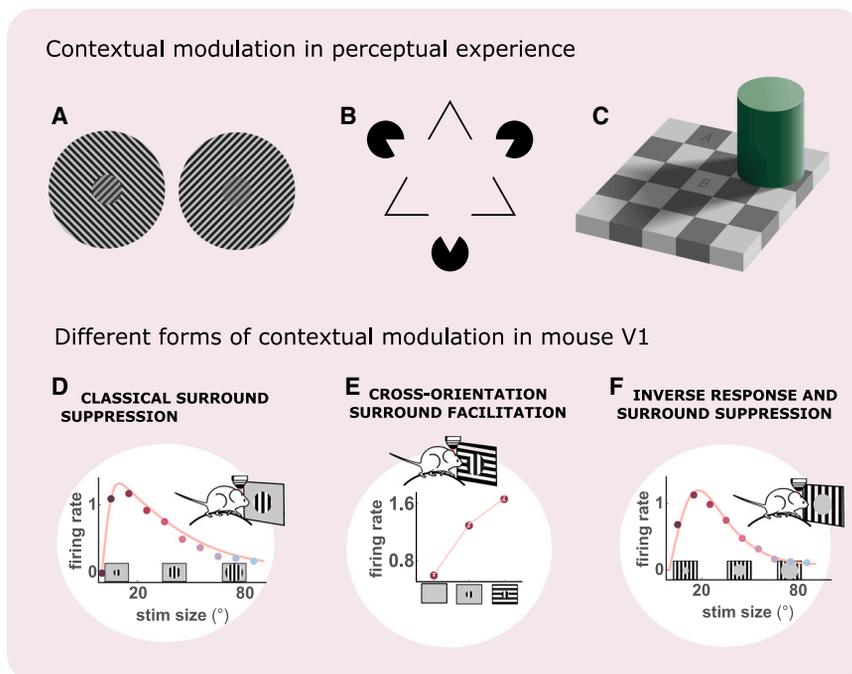


Figure 1. Different forms of contextual modulation

(A) The grating patches in the centers are identical but are perceived as more or less salient due to the distinct surround.

(B) Kanizsa triangle illusion: a nonexistent white triangle in the center appears to occlude the shapes around it.

(C) Checker shadow illusion: the areas labeled A and B have identical brightness but A appears darker than B due to brain's inference that B is in shadow.

(D) Responses decrease when the stimulus size is increased beyond the CRF size; data (circles) from pyramidal cells studied by Keller et al.¹; line is a fit to the data. Here and in what follows, data points represent averages across neurons and firing rates are in arbitrary units.

(E) Responses to a patch of drifting grating are enhanced by addition of an orthogonally oriented surrounding grating. Data replotted with permission from Keller et al.,² represented as mean \pm SEM.

(F) Response to a hole or inverse stimulus in the CRF; i.e., a patch with zero contrast on a full field drifting grating. Note similarity of inverse size tuning to classical size tuning. Data (from Keller et al.¹) and fit as in (A).

- (2) In contrast, addition of a stimulus surround that is very distinct from the CRF stimulus can enhance a neuron's response.^{2,20–24} This “surround facilitation” can enhance the perceptual salience of discontinuities in the visual world, which likely separate different objects.^{25–27} Given an oriented grating as CRF stimulus, this facilitation can be evoked by an orthogonally oriented stimulus in the surround (cross-orientation surround facilitation).
- (3) Surrounding context that indicates occluded structure in the center (Figure 1B) can evoke illusory perception of that structure.^{7,28} While illusory contour responses in V1 are rare^{29–31} (but see Grosf et al.³²), a robust example of such context-induced neural responses in V1 is the “inverse response.” When a large drifting grating is presented, and a uniform patch of mean luminance (a “hole” in the grating) of varying sizes is centered on the CRF, neurons show responses and size tuning to the hole very much like those to drifting gratings of various sizes.^{1,10,33,34} We will refer to the hole stimulus as an “inverse stimulus,” and response suppression for large inverse stimuli as “inverse surround suppression” (although this is something of a misnomer, in that the suppression is induced by expanding the hole, leaving less visual contrast in the surround). In contrast, we will refer to a drifting grating stimulus as a “classical stimulus.” Inverse responses have substantially longer latencies than classical responses^{1,10,33} and are substantially reduced if higher visual areas (HVAs) are optogenetically suppressed,¹ suggesting that feedback connections play a key role in inverse responses.

Here, we develop a unified model of these three contextual effects in layers 2/3 (L2/3) of mouse V1. We consider the four

cortical cell types that have been prominently studied in recent years: excitatory or pyramidal (Pyr) cells and parvalbumin-expressing (PV), somatostatin-expressing (SOM), and vaso-active-intestinal-peptide-expressing (VIP) inhibitory interneurons. We model the cellular response patterns across 2D continuous space, which allows us to determine how the geometry of spatial connections contributes to contextual modulation. We incorporate experimental constraints on the spatial extents of connections^{35–38} and data from Keller et al.¹ on the response profiles across space of each cell type and of inputs from L4 and from an HVA, lateromedial (LM) area, to each stimulus type. We focus on L2/3 both because of the availability of these data and because inverse responses, which depend on feedback from HVAs, are found in L2/3 but not in L4,¹ suggesting that L2/3 is a key site of integration of top-down with bottom-up input.

A number of theoretical works, at different levels of abstraction, have been developed to account for various aspects of contextual modulation. Phenomenological models place it within a broad range of nonlinear response properties described as “divisive normalization.”^{17,39,40} Circuit models^{2,6,38,41–46} have not previously addressed inverse tuning or the roles of top-down feedback or, in many cases, multiple types of inhibitory cells. In addition, our approach differs in being based on analysis of experimental data on the 2D spatial structure of activity and connectivity and developing an analytical framework that gives insight into precisely how interactions across space yield contextual modulation. Our analysis of experimental data produces the surprising finding that, with increasing stimulus size, activity patterns show weak or no increase in spatial width.

To isolate basic principles, we focus on an abstract, “minimal” nonlinear model of a single recurrently connected cell type

Experimental response profiles for classical stimuli

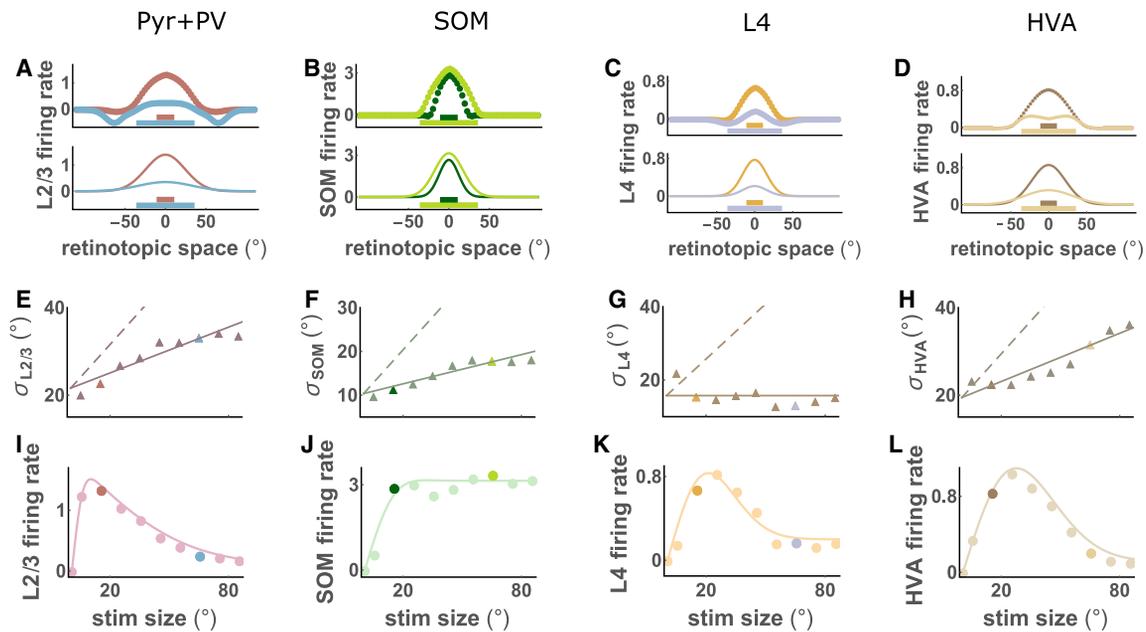


Figure 2. The widths of spatial response patterns grow much more slowly than stimulus size

Pyr + Pv, SOM, L4, and HVA spatial responses, calculated from experimental data.

(A–D) Experimental rate field (top) and best Gaussian fit (bottom) for stimulus sizes $s = \{15^\circ, 65^\circ\}$; x axis represents distance of CRF from stimulus center. Colored horizontal bars represent size of the corresponding stimuli.

(E–H) Full triangles are the widths of the best Gaussian fits for each stimulus size; the full line is a linear fit and the dashed line (slope 1/2) represents the case in which the rate field width grows at the same rate as the stimulus radius (i.e., 1/2 of the stimulus size).

(I–L) Response of cells at stimulus center vs. stimulus size (size-tuning curve). Full circles, data points; full line, fits; dots in different colors highlight responses for the two stimuli in (A)–(D).

representing the joint activity of Pyr and PV cells in L2/3, along with external input from SOM cells, that can be solved analytically. We then show that a full model of the four cell types reproduces all contextual modulation phenomena via the same mechanisms uncovered in this minimal model (Figures S30–S38). Our unifying approach, which focuses on qualitatively explaining response properties and understanding their underlying mechanisms, gives rise to multiple testable predictions (summarized in Table S1).

RESULTS

A spatial schematic description preserving anatomical length scales

To take into account input from neurons with CRFs distributed across visual space, we require knowledge of two factors: (1) how the mean effective strength of synaptic connections depends on the distance in visual space between the CRF centers of the pre- and post-synaptic cells, and (2) how the responses of cells to a stimulus depend on the distance between the cell’s CRF center and the stimulus center. For dependence (1), we assume that synaptic strengths decrease with distance between two neurons as a Gaussian function, with length scales taken from experimental measurements.^{35–38} We take this approach, rather than trying to infer connections that lead to responses

that fit the data, to more tightly constrain our models. We show that our results are robust when considering very large relative errors in the experimental estimates in a biologically realistic model (Figure S33). For dependence (2), we use the dataset of Keller et al.¹ and a smoothing method introduced in⁶ to reconstruct the “firing-rate field” (firing rate vs. CRF spatial position) of Pyr, PV, SOM, and VIP cells in L2/3 of V1, as well as Pyr cells in L4 of V1 and in the higher visual area LM (Figure 2). We will also describe these firing-rate fields as Gaussian (or linear combinations of Gaussian) functions centered on the stimulus center. The transformation from fluorescence signals to firing rates is linear and defined for each cell type using average electrophysiology recordings at baseline and preferred size (maximal response).² In the absence of electrophysiological recordings for LM, we applied the same calibration values used for L2/3 Pyr cells. All the firing rates are expressed in hertz (Hz) and have been baseline subtracted. In the STAR Methods, we provide more detail on the length scales and the smoothing technique employed.

The widths of spatial response patterns grow much more slowly than the stimulus size

We estimated rate fields elicited by classical stimuli of varying sizes from the recordings (Figure 2). We combine Pyr and PV cells into a single joint population, because of the similarity of their responses and for better comparison with the model to be developed in the next section. The rate fields for Pyr + Pv and SOM cells

and for the inputs, L4 and area LM (HVAs), have similar widths as functions of retinotopic space for small and large stimulus size (Figures 2A–2D top). This is quantified by fitting a Gaussian function to each spatial profile (Figures 2A–2D bottom) and comparing (Figures 2E–2H) the width (standard deviation; points and solid lines) to the width expected if the rate field expanded in size equally to the stimulus (dashed lines). The rate fields expand far less than the expected rate and not at all in the case of L4. Finally, the amplitude of the Gaussian fits vs. stimulus size gives the size-tuning curve of the neurons that are centered on the stimulus (Figures 2I–2L). The assumption that rate field width grows to match the stimulus radius is at the core of existing theoretical models of classical surround suppression,^{5,6,38,41,42} in which larger stimuli recruit more wide-ranging lateral activation, and these activated lateral cells have a suppressive influence on the center. The failure of this assumption is also reported (but not discussed) in Dipoppa et al.⁶ and Michaiel et al.⁴⁷ In Figures S1–S4 we show the rate fields for all cell types and all stimuli, as computed by two different methods; in Figure S5 we show that we find no intrinsic qualitative difference between LM cells projecting to a given point in L2/3 from the same retinotopic location and those projecting from peripheral locations (aligned vs. offset cells), contrary to what was suggested by Keller et al.¹ As a result, we treat HVA cells as homogeneous across space. In Figures S6–S9 we show the fit of the spatial profiles as well as the size-tuning curves for all cell types.

In what follows, we first develop a minimal model for ordinary surround suppression. We then apply this model to understand inverse response. Finally, we extend the model to understand cross-orientation surround facilitation.

A minimal model for contextual modulation

We develop a minimal model to address the question of how feedforward, feedback, and lateral input currents contribute to contextual modulation in L2/3. In this model, L2/3 consists of only a single recurrently connected cell type, representing the combination of Pyr and PV cells. This is based on observations that PV responses track those of Pyr cells in multiple contexts and over a wide dynamic range of inputs,^{43,48,49} thereby dynamically balancing (preventing an excess of) excitation and functioning as a stabilizer for the network.^{49–57} In contrast, VIP and particularly SOM cell responses generally differ substantially from those of Pyr and PV cells.^{1,2,6,43,49} We treat SOM cells as a static external input, responding across space as measured experimentally, and ignore VIP cells, which act primarily on SOM cells.^{35,58,59} We also take as external inputs the measured responses across space of L4 and LM area excitatory neurons. We treat LM as a static input, rather than as part of the recurrent circuit, due to the limited information available about non-L2/3 inputs to LM (see discussion). We assume neurons have a supralinear input/output function; for further information, see the STAR Methods.

Classical surround suppression in the minimal model

In our minimal model of classical size tuning (Figure 3), inputs to L2/3 are given by the convolution of the rate fields of L4 and HVAs (and SOM when specified), which were derived from the data (Figure 2), with the corresponding Gaussian connectivity (Figure 3B). The recurrent rate fields resulting from the approxi-

mate analytical solution and the simulation of the minimal model agree well (Figures 3C, 3D, S11, and S12) so in what follows we rely on the analytical formulas to efficiently compute model results. The model's free parameters are the amplitudes and widths of the feedforward, feedback, and recurrent connection strengths. Since we consider the recurrent population to be a joint population of Pyr and PV cells, the recurrent connection strength W_0 can be either positive or negative, depending on whether Pyr or PV dominates. We find that L2/3 responses are unstable if W_0 is too strongly positive, while making W_0 too strongly negative decreases surround modulation, consistent with effects of optogenetic activation of PV cells.⁶⁰ We set W_0 to a moderate negative value. The feedforward and feedback connection strengths were then set to reproduce the amplitude of the response in L2/3. In Figures S11 and S12 we show the Pyr + Pv experimental rate fields in two different datasets and for all stimulus sizes; in Figures S13–S15, we provide a validation of the minimal model (see also STAR Methods), along with an analysis of how surround modulation varies with recurrent, feedforward, and feedback strengths.

Feedforward, feedback, and lateral inhibitory input shape classical surround suppression

Many experiments show that feedforward, horizontal, and feedback connections all contribute to V1 classical surround suppression (e.g., Angelucci et al.⁵ and references therein): our model allows us to analyze their interplay. We first analyze the contributions of feedforward (L4) and feedback (HVAs) inputs. In the next section, we consider addition of SOM input.

In STAR Methods we show that surround suppression in the recurrent layer can be produced by either or both of (1) surround suppression in the input population and (2) broadening of the input rate field as stimulus size is increased (Figure S16). In case (1), L2/3 surround suppression is driven by the input (although it may be modulated in L2/3). In case (2), L2/3 surround suppression arises from increased lateral activity driving greater center response suppression. Since the scaling of the widths is weak for LM and absent for L4 (Figure 2), we expect the second mechanism to contribute only marginally to L2/3 surround suppression.

To test this hypothesis, we consider the effects of (counterfactually) eliminating surround suppression in L4 and HVAs (Figures 3F–3G and 3I–3J) and of eliminating the growth of the width of the HVA rate field with increasing stimulus size (Figures 3H and 3K; note the width of the L4 rate field does not increase with stimulus size). In each condition, we calculate a surround modulation index, $SMI = 1 - r_{Large}/r_{Pref}$, where r_{Large} and r_{Pref} are the firing rate of the centered cells at the largest stimulus size ($s = 85^\circ$) and the preferred size, respectively (if $r_{Large} = r_{Pref}$, then $SMI = 0$). Consistent with our hypothesis, eliminating surround suppression in the inputs strongly reduces surround suppression in L2/3, while eliminating growth of the HVA rate field width does not alter surround suppression in L2/3 (Figure 3E).

In Figure S17 we compare the recurrent rate field obtained from a supralinear vs. a linear transfer function and show that the former supports stronger surround suppression, presumably because the nonlinearity increases the gap between responses to stronger vs. weaker stimuli.

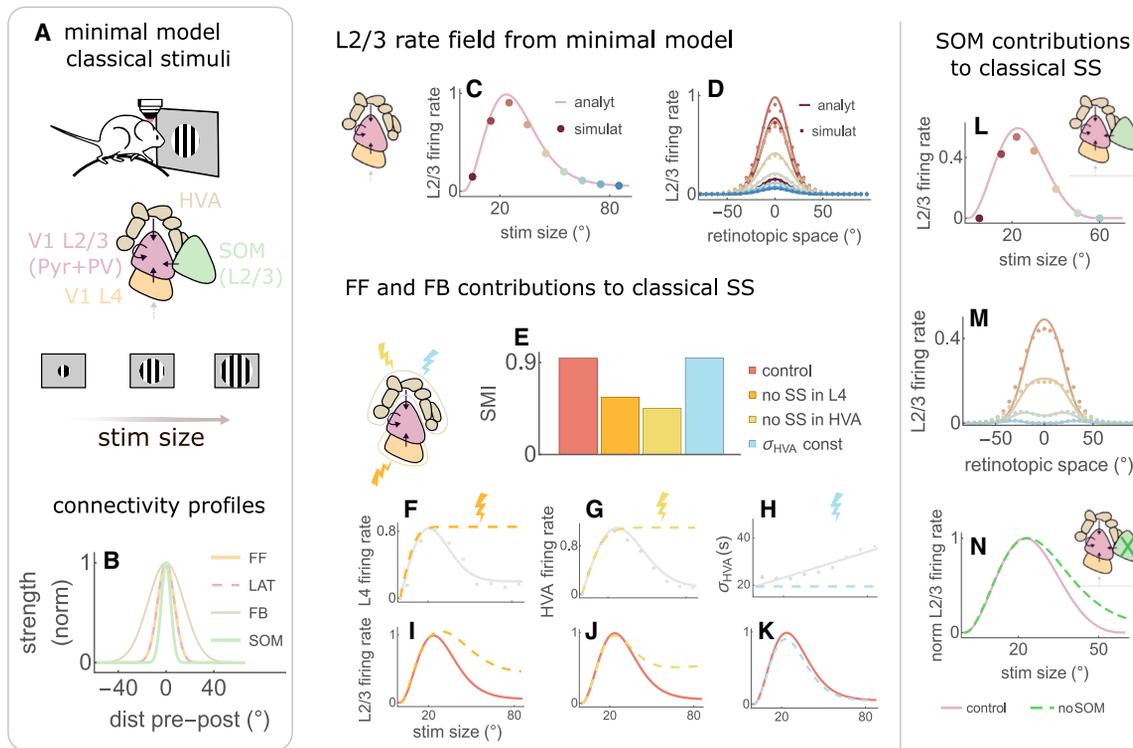


Figure 3. Feedforward, feedback, and lateral SOM inputs shape classical surround suppression

(A) Sketches of experimental setup with classical size tuning stimuli; minimal model with one recurrent cell type in V1 L2/3 receiving feedforward (FF) input from L4, feedback (FB) input from HVAs, and lateral input from its own type (LAT) and from SOM cells.

(B) Strength of the four types of connections vs. CRF distance between pre- and post-synaptic neurons.

(C) L2/3 size-tuning curve and (D) rate field. Here and in (L) and (M), color code of the full circles indicates stimulus size, full circles represent simulations, and full lines represent analytical results.

(E–K) Counterfactual modifications of the inputs. (E) Bar plot representing the *SMI* (see main text) in the control condition and when we eliminate surround suppression (SS) in L4, or HVAs, or growth of the width of the HVA rate field with stimulus size. (F–H) Illustration of the counterfactual modifications (dashed, same color code as E) compared to the empirical values (full lines and symbols as in Figure 2). (I–K) Effects of the counterfactual modifications on L2/3 size-tuning curve.

(L–N) Effects of adding SOM input. (L and M) Same as (C and D) but with SOM input. For large stimuli, the rate field convexity in the center changes. Compare with Figure S5A. (N) Size-tuning curve of recurrent layer with (light red; same as L) and without (dashed green; same as C) SOM input.

Surround suppression is weak for low-contrast stimuli and increases in strength, while optimal size shrinks, with increasing stimulus contrast.⁴³ In Figures S18–S20, we analyze the contrast dependence of size tuning in the model. We find that a supralinear one-population model with Gaussian input whose width increases with stimulus size can generate the observed contrast-dependent size tuning, in line with the results in the two-population E/I model in Rubin et al.⁴¹ However, the contrast dependence in the minimal model is parameter dependent, and in particular does not occur with the connectivity length scales we are using, which are matched with those observed experimentally in mouse L2/3. In Figures S21 and S22, we examine this contrast dependence further and find that additional contributing factors to contrast-dependent size tuning may be (1) scaling of rate field width in the feedback input and (2) large VIP/small SOM response for low contrasts (in agreement with Mossing et al.⁴³). In Figure S23, we show that our results still hold when considering non-Gaussian aspects of the rate fields.

SOM cells enhance classical surround suppression and generate more complex spatial profiles

The rate field of SOM neurons in L2/3 lacks surround suppression (Figure 2J).⁶¹ We compare the size-tuning curve of the model recurrent population in the presence or absence of SOM input, as in Adesnik et al.⁶¹ Since SOM cells prefer larger stimulus sizes, they inhibit L2/3 the most for larger stimuli, thus enhancing L2/3 classical surround suppression (Figures 3L and 3N). This effect is robust against changes in the free parameters (i.e., the connection strengths). However, when the effective strength of SOM projections onto the Pyr + Pv population is large enough, the model Pyr + Pv rate field develops a ring shape for large stimulus sizes (Figure 3M), which is well captured by our analytic solution. This ring shape for large stimulus sizes has been reported in Dipoppa et al.⁶ (replotted in Figure S12). The absence of this feature in our dataset (Figure 2I) could conceivably be explained if those animals had a smaller effective strength of SOM projections than in Dipoppa et al.⁶ From a geometric point of view, the SOM input field is a (relatively narrow)

Experimental response profiles for inverse stimuli

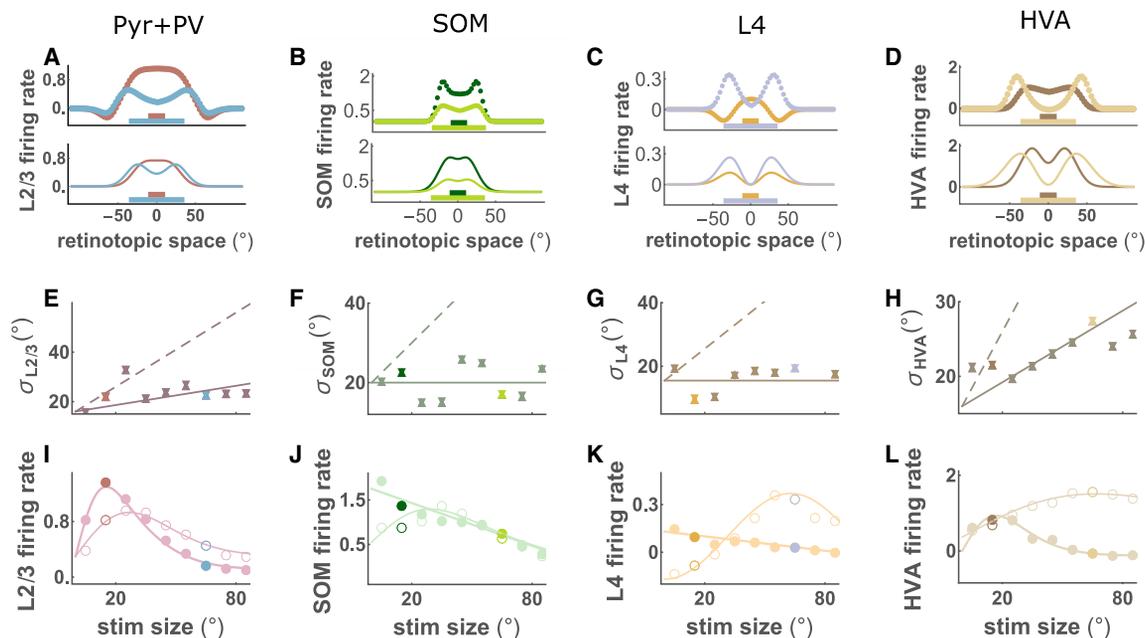


Figure 4. Inverse response and size-tuning properties of V1 and HVAs

Same as Figure 2 but for inverse stimuli.

(A–D) Experimental rate field (top) and best difference-of-Gaussians fit (bottom) for two stimulus sizes (represented at scale by the bottom bars).

(E–H) Up and down triangles (almost overlapping) are the widths of the positive and negative Gaussian functions in the best fits to the rate fields. Differences in widths of the two Gaussians are small. Full line, linear fit; dashed line, fit if width of the positive Gaussian grew as the radius of the stimulus.

(I–L) Inverse size-tuning curves. Full (empty) circles are experimental responses of cells centered on (with an offset of 30° with respect to) the stimulus center, vs. inverse stimulus size. Full lines are fits.

Gaussian-shaped bump (e.g., see SOM rate field in Figure 2), so it most strongly inhibits central Pyr + Pv cells. For large stimuli, its amplitude becomes larger than that of other inputs, generating a concave Pyr + Pv rate field. The excitatory input from HVAs can also contribute to this effect, since for large stimuli it also exhibits a ring profile (Figures 2L and S9).

In conclusion, our model predicts that classical surround suppression in L2/3 and its contrast dependence are largely explained by lateral suppression of its inputs (consistent with Angelucci et al. and Vangeneugden et al.⁶²), marginally enhanced by lateral (PV) inhibition progressively recruited as the stimulus size increases,^{41,42} enhanced by SOM lateral input,⁶¹ and partially supported by the supralinear input/output function^{63,64} (see Table S1).

Inverse response and size-tuning properties of L2/3 are shaped by the feedback input

In the previous sections, we developed the minimal model, with most parameters set from experimental data. Here we use the same model, with no adjustment of parameters except the increased amplitude of HVAs input for inverse vs. classical stimuli (further detailed in Figure S10), to understand responses to inverse stimuli of varying sizes (i.e., hole diameters).

In experiments, letting position 0° be the stimulus center, L2/3 cells with CRF center near zero (“aligned cells”) respond well to inverse stimuli, but the aligned inputs from L4 and HVAs have

very low firing rates (Figures 4 and S1–S4). This led Keller et al.¹ to conclude that the inverse tuning properties of L2/3 are not directly inherited from its retinotopically aligned inputs. Aligned cells in L2/3 show size tuning to inverse stimuli very similar to classical size tuning (Figure 4I, full circles, compare with Figure 2I). However, the spatial representation of inverse stimuli is qualitatively different from the spatial representation of classical ones: for moderate-sized (35°) to large stimuli, the L2/3 rate field for inverse stimuli has a local minimum at 0° , so that the peak response is at a characteristic distance from the center (Figure 4A). This is not true for classical stimuli (Figures S1–S4), although, as we noted previously, it can occur for large enough stimuli both in the model (Figure 3M) and in experiments.⁶ Similarly to the classical stimulus condition, the widths of the inverse rate fields scale weakly with the stimulus size (Figures 4E–4H and S9).

We parameterize the input rate fields with difference-of-Gaussian functions (Figures 4A–4D, bottom) whose parameters vary continuously with stimulus size (see STAR Methods and Figures S6–S9). In doing so, we trade off some details of the spatial profiles of the responses to gain interpretability. In particular, to leverage the minimal model, we adopt a compact analytical expression for the L4 rate field that limits the accuracy of the fit for small stimuli (Figure 4C). In Figure S23 we show that the results presented are qualitatively equivalent with far more accurate fits.

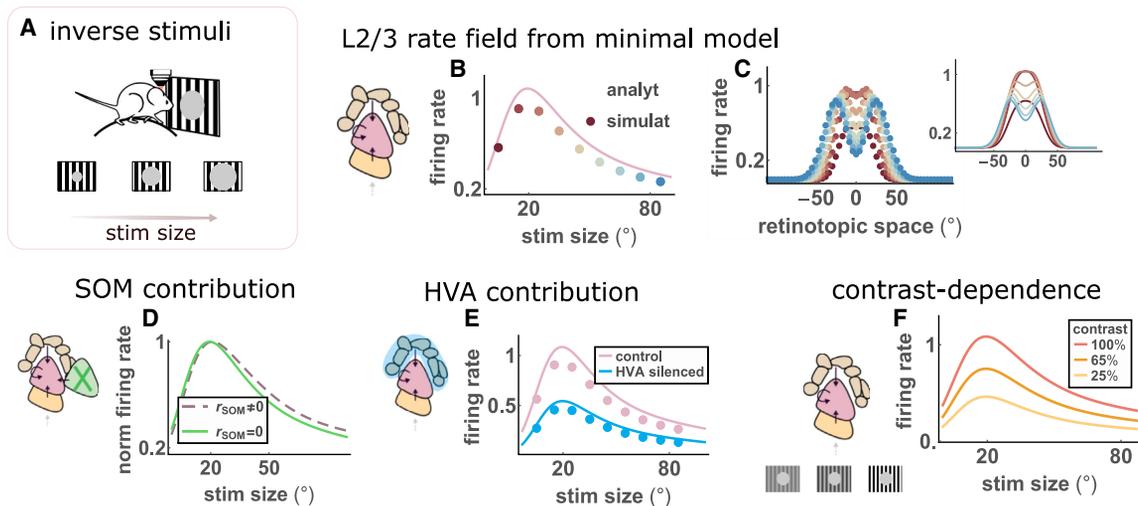


Figure 5. Inverse response and size-tuning properties of L2/3 are shaped by its feedback input

(A) Inverse size-tuning stimuli, (B) size-tuning curve, and (C) rate fields for L2/3 of the model. In (C) results for the simulations (dots) and the analytics (curves) are represented separately for better readability. The model recovers inverse size tuning as well as the change in convexity of the rate field (compare Figures 4A and 4I).

(D) SOM neurons reduce inverse surround suppression (compare Figure 3N for classical stimuli).

(E) Size-tuning curve in the control case (light red) and for silencing of HVAs (rate field of HVAs reduced by a factor 0.65) (compare Figure S12A for classical stimuli).

(F) Size-tuning curves for varying contrast (both L4 and HVA rate fields are reduced by a factor $c = \log_2 2C_0$, C_0 values shown in legend; compare Figure S21 for classical stimuli).

The minimal model reproduces inverse size tuning and the spatial profile of inverse response as a function of size (Figures 5B and 5C): centered cells show high response to small inverse stimuli, while, for large inverse stimuli, offset cells show a large response and centered cells respond less. Adding SOM cell input to the model, using experimentally measured SOM responses, slightly *reduces* the degree of inverse surround suppression, the opposite of the effect seen in the classical case (Figure 5D, and see Figure 3N for comparison). This is because SOM cells respond more strongly to small than to large inverse stimuli (Figures 4B and S9), whereas they respond more strongly to large than to small classical stimuli (Figure 2).

Inverse response is substantially reduced by reduction of HVA input (Figure 5E). For classical stimuli, the same modification affects L2/3 response only marginally. This is consistent with experimental results on optogenetic silencing of HVAs (see Figure 5C in Keller et al.¹). This differential effect of HVA suppression on inverse vs. classical response is consistent across varying levels of HVA suppression (Figure S24).

Finally, the model shows an increase of inverse responses with contrast (Figure 5F, see also Figure S25) similar to that found experimentally.¹

Inverse response requires wide-enough feedback projections and feedback activity profiles scaling with stimulus size

We use our analytical framework to determine the conditions needed for inverse responses to arise. We first investigate the role of the width of the feedback projections $\sigma_{L23-HVA}$. We vary its value around the experimental estimate (Figures 6A and 6B) while the inverse stimulus size is fixed at 15° (Figure 6A). For

this stimulus size, the L2/3 rate field is peaked at its center. For this to occur in the model, $\sigma_{L23-HVA}$ must be sufficiently large, but not too large (Figure 6C), as we now explain.

HVAs activity peaks in a ring of cells with CRFs corresponding approximately to the hole edge, although the ring size grows somewhat more slowly than the hole size (Figures 4 and S26). We posit that each HVA cell on the ring projects a Gaussian-shaped bump of projections back to V1 L2/3, centered on its position on the ring. If these feedback projections are wide enough, the bumps will all overlap in the middle, giving the strongest input, and thus the strongest response, at the stimulus center (shown schematically in Figure 6A, discussed in Figures S27–S29).

To quantify the range of values of $\sigma_{L23-HVA}$ compatible with the inverse response, we compute the second derivative of the recurrent rate field at its center as a function of $\sigma_{L23-HVA}$ for the 15° stimulus (Figures 6C and 6D). A negative second derivative denotes a maximum in the center, as observed experimentally. For this to occur, $\sigma_{L23-HVA}$ needs to fall within the range (12°, 35°) (Figure 6D). The experimental estimate in Marques et al.³⁷ is indicated by the shaded area in Figure 6D, which falls within this range. The same results hold in a model of L2/3 with four recurrent cell types (see Figures S30–S32 and S34–S38).

We can similarly understand why inverse response is size tuned. As the stimulus size increases and the ring-shaped HVA rate field becomes wider, the input to L2/3 cells centered on the stimulus center decreases (the overlap of the bumps in the middle decreases; see Figure 6E). At some point, the input peak moves away from the center, and an ever-deepening trough forms at the stimulus center (Figure 6I),

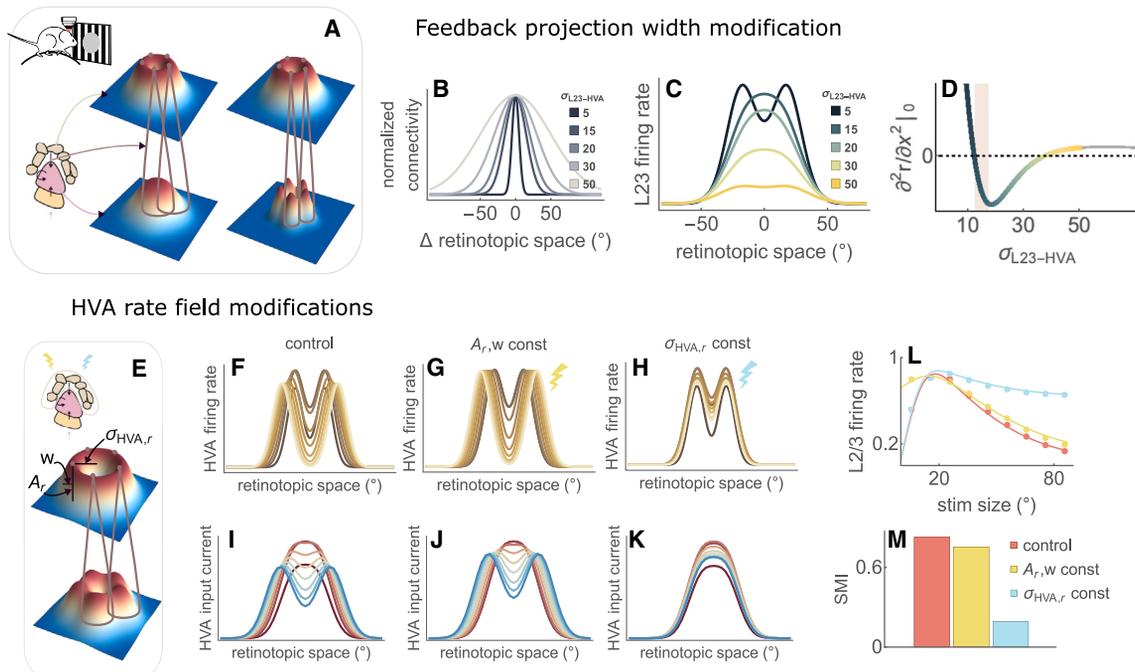


Figure 6. Inverse response requires wide-enough feedback projections and feedback activity profiles that scale with stimulus size

(A) Rate fields in HVAs for a small stimulus (top 3D plots, ring profile). The feedback projections are represented by the empty cones (narrower projection width on the right). Bottom 3D plots: L2/3 field of input current arising from the feedback projections from five example points; the L2/3 input current would be the sum of similar projections from all points in the HVA rate field. Narrower feedback projections yield a trough in the center (right), while broader projections yield a peak (left).

(B) Counterfactual modifications of the width of projections from HVAs to L2/3. Amplitudes are normalized to 1.

(C) L2/3 rate fields for the different values of $\sigma_{L23-HVA}$ and fixed stimulus width 15° .

(D) Second derivative of L2/3 rate field evaluated at the origin, vs. $\sigma_{L23-HVA}$. Color code as in (C); shaded bar represents the experimental estimate in Marques et al.³⁷

(E) Top: parametrization of HVA rate field used for (F)–(M). Bottom: for fixed $\sigma_{L23-HVA} = 20^\circ$, a wide-enough $\sigma_{HVA,r}$ results in a trough in the center.

(F–K) HVA rate field (F–H) and input current (I–K) for the control condition (F and I; parameters as in previous figures) and for two counterfactual modifications: either A_r and w (G and J, dark yellow symbol) or $\sigma_{HVA,r}$ (H and K, light blue symbol) are kept constant (fixed at their value for size $s = 15^\circ$) with varying stimulus size.

(L and M) Size-tuning curve (L) and *SMI* for the control condition (M) (red) and the modified conditions (colors as in G and H).

representing the decrease in response with increasing stimulus size (Figure 6F).

While this explanation attributes the decrease in response with increasing hole size to the increasing width of the ring of HVAs activity (which we also show analytically in Figures S27–S29), another contributing factor could be the decrease in amplitude and thickness of the ring. To test this, we fit the HVA rate field with an alternative parameterization, a circle of radius $\sigma_{HVA,r}$ (ring radius) convolved with a Gaussian of amplitude A_r and width w (ring thickness; see Figure 6E). We then apply two counterfactual modifications to the HVA rate field: (1) eliminate the size dependence of A_r and w while preserving the scaling of $\sigma_{HVA,r}$ with stimulus size (dark yellow symbol, Figure 6G), (2) eliminate the scaling of $\sigma_{HVA,r}$ while preserving A_r and w (light blue symbol, Figure 6H). If our explanation is right, (1) should only have a minor effect, while (2) should significantly reduce inverse surround suppression (which we also show analytically in the STAR Methods).

Indeed, in case (1), the field of HVA input to L2/3 develops a trough for large stimuli (Figure 6J), and the inverse size-tuning curve is very similar to that in the control case, with a minor decrease in *SMI* (Figures 6L–6M). The only notable change is that the response to the smallest size hole, 5° , becomes almost

as large as the preferred size response. In contrast, in case (2), the HVA input field does not develop a trough as the stimulus size increases (Figure 6K) and, correspondingly, inverse surround suppression is drastically reduced (Figures 6L–6M). Note that, for a classical stimulus, surround suppression is only slightly decreased by an analogous modification (compare Figures 3E and 6M, light blue bar).

This difference between the classical and inverse cases in the effect of a lack of widening of the input rate field suggests that, despite the similarity of their size-tuning curves, their origin is substantially different: classical surround suppression in L2/3 is mostly driven by the aligned cells of its inputs, while inverse size tuning is mediated by offset cells in HVAs. As the stimulus becomes larger, peak responses in HVAs move to cells at larger offsets from the center, providing less and less excitatory input to the center as the stimulus radius grows beyond the span of the feedback projections.

In Figures S34–S38, we show that the explanations derived using the minimal model also apply to a more biologically realistic model with four cell types.

The arguments presented here provide a purely geometrical explanation of inverse response and inverse size tuning,

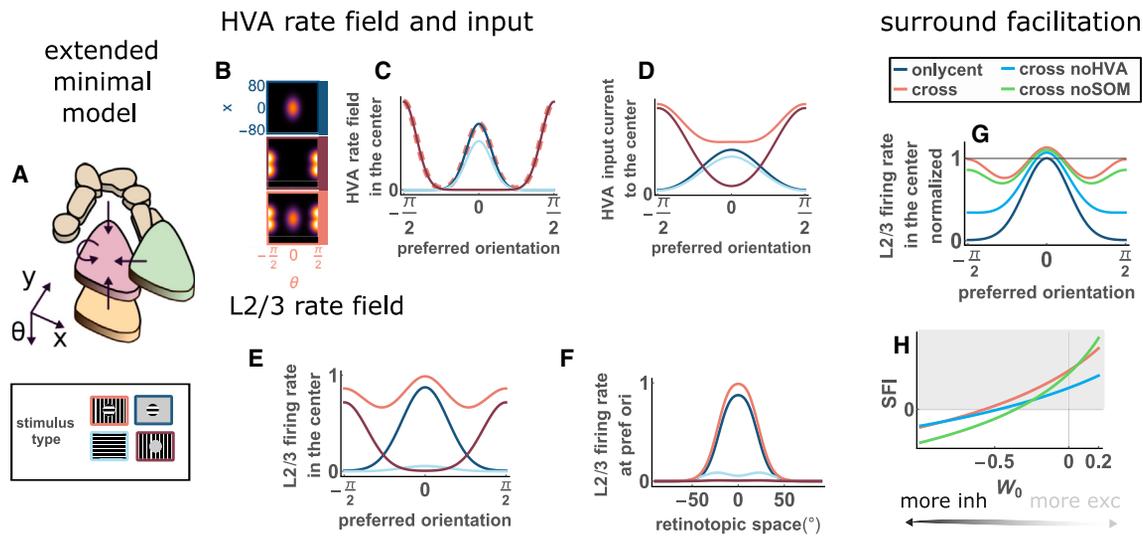


Figure 7. Surround facilitation is generated by inverse response to the orthogonal surround

(A) Sketch of the extended minimal model, accounting for orientation preference as an additional dimension for the rate fields. Bottom: legend of the stimuli analyzed: cross (center at orientation 0, surround at $\pi/2$; light red), classical/only center (dark blue), iso-oriented surround (light blue), and inverse (dark red). (B) Density plot of HVA rate fields in one spatial dimension (vertical axis) and orientation preference (horizontal axis) for classical, inverse, and cross stimuli (color code as in the legend). (C) HVA rate fields of cells whose CRF is centered on the stimulus center and with all possible orientation preferences (horizontal axis). Note, light red line for cross stimulus is shown dashed and thick, so overlapping lines (dark blue, central region; dark red, larger preferred orientations) can be seen; light blue and dark blue (not visible) overlap for larger preferred orientations. (D) Input current from HVAs to the centered cells in L2/3 as a function of their preferred orientation. (E and F) The response of centered L2/3 cells preferring orientation $\theta = 0$ is larger for the cross stimulus than for the classical stimulus alone (surround facilitation). (E) Rate field of L2/3 neurons centered on the stimulus center as a function of their preferred orientation. (F) Rate field of L2/3 cells that prefer orientation $\theta = 0$ as a function of their retinotopic location. (G) We study the cross condition (control) and two manipulations of it (no HVA, no SOM). For each condition, rate fields of L2/3 are normalized to the amplitude of the response to a center-only stimulus. Surround facilitation (responses larger than 1) decreases when HVAs are silenced or SOM cells are hyperpolarized. (H) Larger recurrent excitation (less negative or more positive W_0) contributes to larger SFI.

illustrating the importance of considering a spatial framework with realistic anatomical and physiological length scales to understand contextual effects.

Surround facilitation can be explained by the inverse response to the orthogonal surround

When a stimulus is presented together with an orthogonal surround (“cross” stimulus), the response of L2/3—but not L4—Pyr cells centered on the stimulus is facilitated.^{2,46,65} We propose that this effect may be, at least in part, the counterpart of the inverse response. More specifically, the response of aligned Pyr cells may increase due to the orthogonal surround evoking the same inputs that drive the inverse response. To test this hypothesis, we extend the minimal model to include orientation preference (see Figures 7A and S39) and connectivity that decreases in strength with increasing difference in preferred orientation according to experimental estimates^{36,66,67} (see STAR Methods).

We assume that the rate fields of the input populations for the cross stimulus can be approximated as the sum of the rate fields for a direct stimulus of one orientation in the center plus an inverse stimulus of the orthogonal orientation (Figures 7B and 7C). In general, the response fields are not additive. Nevertheless, for cross stimuli, the classical and inverse stimuli are largely encoded by complementary sets of cells, i.e., cells with ortho-

gonal orientation preferences, which interact minimally³⁶; hence, we make the assumption of additivity.

We consider a classical stimulus with orientation 0, an inverse stimulus with orientation $\pi/2$, and a cross stimulus that combines the two. The extended minimal model shows that L2/3 cells are surround facilitated (Figures 7E and 7F). This effect can be understood as follows: HVA cells at the stimulus center with preferred orientation 0 have zero response to the inverse stimulus (dark red line in Figure 7C). However, L2/3 cells with the same position and preferred orientation receive small but positive HVA input (Figure 7D), which comes from more offset HVA cells via their broad feedback projections to L2/3. Thus the inverse stimulus at orientation $\pi/2$ provides a small HVA input to the L2/3-centered cells that prefer orientation 0. This input generates cross-oriented facilitation (light red curves in Figures 7E and 7F).

Thus, our model suggests that HVAs support L2/3 cross-orientation surround facilitation (Figure 7G, light red curve). If HVAs are blocked (cyan curve), responses to both center-only and cross stimuli decrease; cross is still facilitated relative to center only, but less so than with HVAs intact (the cyan curve in Figure 7G lies below the light red curve). The same is true also when SOM input is silenced (green curve), implying that both HVAs and SOM cells support cross-orientation surround facilitation.

To better quantify surround facilitation, we define the surround facilitation index $SFI = (r_x - r_c)/(r_x + r_c)$, where r_x (r_c) is the firing rate of the L2/3-centered cells in the cross (classical) stimulus condition. We analyze the dependence of SFI on the relative weight of excitation and inhibition in the joint Pyr + Pv population, controlled by the amplitude of the recurrent connections W_0 (Figure 7H). Larger W_0 (i.e., more recurrent excitation) yields larger SFI .

For excitation dominated systems ($W_0 > 0$), SOM input suppresses surround facilitation (green curve above the light red one). On the other hand, for inhibition-dominated systems ($W_0 < 0$), as in Figure 7G, SOM input increases SFI (green curve below the light red). More generally, in Figures S40 and S41, we show that SOM input increases (decreases) SFI if the ratio of SOM input in the inverse vs. center conditions is less (greater) than the ratio of total input in the two conditions.

To summarize, assuming that inputs evoked by center-only and orthogonal surround-only or inverse stimuli add linearly, both HVAs and lateral SOM inputs contribute to cross-orientation surround facilitation, while lateral PV input tends to reduce it. Note that the suppression of SOM was the primary mechanism of enhancement of response to the cross-oriented stimulus proposed in Keller et al.² (in that case due to activation of VIP by the cross stimulus, which inhibited SOM); here, we are proposing that both HVAs and the suppression of SOM contribute to the facilitation.

The assumption of additivity can be experimentally tested by recording the spatial pattern of responses to cross stimuli. In particular, we expect that L2/3 Pyr and PV cells centered on the cross stimulus and with preferred orientation orthogonal to the center stimulus show significant response (Figure 7C). Similar recordings, with a smaller difference in orientation between center and surround, should instead deviate from additivity, perhaps reflecting the neural correlates of the “tilt illusion.”⁶⁸

DISCUSSION

We propose a unifying (toy) model of mouse V1 accounting for three types of contextual modulation and connecting two seemingly unrelated phenomena: inverse response and cross-orientation surround facilitation. To gain robust mechanistic insight, we include in the model as few elements as possible to easily disentangle the effects of each. Instead of quantitatively fine-tuning the model to a particular dataset, we aim to describe results at a qualitative level, at which they are more robust and consistent across different datasets. While the minimal model neglects the full recurrent microcircuitry of L2/3, it yields mechanistic insights into cortical computation, which we show also apply to a fully recurrent four-cell-type circuit (Figures S34–S38).

Despite the very limited number of parameters (as few as three), our minimal approach is highly informative. In addition to the three main types of contextual modulation, it reproduces stabilization of the circuit by PV cells, the modulations of the classical and inverse response when silencing HVAs, and the effects of changing stimulus contrast and hyperpolarizing SOM cells. At the same time, it generates insights that translate into several experimentally testable predictions, as summarized in Table S1.

Previous models of classical surround suppression have assumed it is mediated by lateral inhibition progressively recruited as the stimulus size increases.^{5,6,38,41,42,69} Our analysis suggests that, in mouse L2/3 of V1, this mechanism only marginally contributes to surround suppression, because (1) the feed-forward input to L2/3 is also strongly surround suppressed, and (2) the rate field size increases much more slowly than the stimulus size.

Firstly, given the well-documented existence of surround suppression in earlier stages of the visual pathway, such as the retina^{70,71} and the lateral geniculate nucleus,⁷² we believe that size tuning may be transmitted along the visual pathway and modulated at the different stages, including the primary visual cortex. This observation also applies to other model organisms, such as cats and monkeys,^{73,74} and underscores the importance of considering perception from a more integrated perspective, rather than focusing solely on individual layers or isolated circuits.^{75,76}

Secondly, some of the simplifying assumptions of our framework are based on the large amount of available anatomical and functional information on mouse visual cortex; therefore, it is not straightforward to extend these results to other model organisms. The finding that the growth of the width of input profiles alone does not support surround suppression is parameter dependent and does not hold for length scales that are more representative of cat anatomy (Figures S18–S20).

However, it’s important to note that the observation of a slow rate field size increase in our dataset is purely empirical and, as such, it does not depend on the model. Moreover, this observation is independent on the preprocessing procedure used to define the rate fields (Figures S2–S4) and it is consistent with the recordings in Dipoppa et al.⁶ (their Supplementary Figure 11D) and in Michael et al.⁴⁷ (their Figure 2A), although those authors did not highlight the effect. Pinpointing this counterintuitive feature of neural responses in mouse V1 is one of our major findings. We should, however, caution that all of the recordings in this and those two papers were done with Ca^{++} imaging, and it is conceivable that there is widening of the spiking signal but with intensity below threshold for Ca^{++} recording.

Our finding suggests the interesting psychophysical correlate that the perceived area of stimuli should grow more slowly than the actual area. This is in agreement with the results of Yousif and Keil,^{77,78} who had participants compare total area of multiple patches of color, distributed in random locations within a region, and with different diameter distributions. For samples with equal total area, participants consistently perceived less cumulative area in samples with fewer larger patches than in samples with more smaller patches. This correlate could be further tested with another experiment: alternately present a small circular stimulus centered at position \mathbf{x} , and a larger circular stimulus whose edge intersects \mathbf{x} . We predict that the larger stimulus should be perceived as not extending to the center of the smaller one. We conducted a pilot experiment that appears to support this prediction. These results seem to suggest that recordings of the rate fields as a function of stimulus size for other model organisms may reveal a trend similar to what we observe in mouse.

Limitations of the study

We modeled HVA input based on recordings from area LM.¹ However, in these recordings, LM responses are much weaker in the inverse condition than in the classical condition. Since inverse responses have similar strength to classical responses, we assumed that input from other HVAs must equalize overall HVA responses in the two conditions, and we scaled observed LM inverse responses accordingly. This assumption needs to be tested experimentally.

We treat HVAs as a static input, even though they are part of a coupled system, receiving input from V1 and projecting back to V1. To develop a dynamical model for the coupled population of L4, L2/3, and HVA, we would need to specify the external inputs to this entire system. However, in the dataset considered, L2/3 responses to small-size classical and inverse stimuli are very similar, while LM responds very differently to the two stimuli. This suggests that LM likely receives a significant part of its input from other sources^{79–81} for which we have no recordings. More broadly, being able to determine whether a specific dynamical regime in a circuit arises from the circuit's intrinsic properties (such as its internal connectivity) or is driven by the characteristics of external inputs is at the center of ongoing debate (see e.g., Morales et al.⁸² and Morrell et al.⁸³).

Context shapes the readout not only at a neural level but also at a perceptual level. Animals are active interpreters of the visual scene, as opposed to passive decoders.^{7,68,84–91} The study of the mechanisms underlying contextual modulation of responses, including information that higher processing stages project back onto the canvas of visual cortex, opens a window to understand how our brains create our perceptual experiences.

We believe that the simplicity, analytical tractability, and spatial nature of the model presented here, together with the fact that it can be easily extended to account for other feature spaces, will enable further mechanistic insight into visual cortical circuitry and responses in future studies.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to the lead contact, Serena Di Santo (serenadisanto@ugr.es).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- This paper analyzes existing, publicly available data, accessible at <https://doi.org/10.1038/s41586-020-2319-4>.
- All original code has been deposited at Zenodo and is publicly available at <https://doi.org/10.5281/zenodo.13957280> as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

We acknowledge funding sources NIH U01NS108683, NIH R01EY029999, NIH U19NS107613, NSF 1707398, Gatsby Charitable Foundation GAT3708, Kavli Foundation, grant no. IJC2020-044517-I, and grant no. PID2023-149174NB-I00 financed by the Spanish Ministry and Agencia Estatal de Inves-

tigación MICIU/AEI/10.13039/501100011033 and EDRF funds (European Union).

AUTHOR CONTRIBUTIONS

Conceptualization, K.D.M., S.D.S., and M.D.; methodology, S.D.S.; software, S.D.S.; formal analysis, S.D.S.; experimental data, A.K. and M.R.; writing – original draft, S.D.S.; writing – review & editing, K.D.M. and S.D.S.; visualization, S.D.S.; supervision, K.D.M.; funding acquisition, K.D.M. and M.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
- **METHOD DETAILS**
 - Nonlinear transfer function
 - Feedback input
 - The minimal model
 - Validation of the minimal model
 - The minimal model with orientation tuning
 - The full model
 - Inferring parameters of the full model
 - Model selection for the full model
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Data preprocessing
 - Aligned and offset LM neurons are not intrinsically different
 - Functional form of the rate fields
 - Moreover its distribution is very narrow
 - From LM to HVAs
 - The joint population of Pyr and PV
 - Anatomical length scales
 - Input currents
 - Analytical insights on classical size tuning curves with non-normalized input currents
 - Dissecting classical surround suppression
 - Classical surround suppression with non-surround suppressed input rate fields
 - Preferred size depends on the preferred size of the inputs
 - Preferred size decreases with contrast
 - Contrast dependent classical size tuning curves
 - The origin of inverse response and inverse size tuning
 - Analytical results on inverse surround suppression
 - HVAs cells close to the edge of the inverse stimulus respond maximally
 - Lateral input from SOM cells increases surround facilitation and the connection with SOM inverse response

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2024.115088>.

Received: May 19, 2024

Revised: September 27, 2024

Accepted: November 27, 2024

Published: December 21, 2024

REFERENCES

- Keller, A.J., Roth, M.M., and Scanziani, M. (2020). Feedback generates a second receptive field in neurons of the visual cortex. *Nature* 582, 545–549. <https://doi.org/10.1038/s41586-020-2319-4>.
- Keller, A.J., Dipoppa, M., Roth, M.M., Caudill, M.S., Ingrassio, A., Miller, K.D., and Scanziani, M. (2020). A disinhibitory circuit for contextual modulation in primary visual cortex. *Neuron* 108, 1181–1193.e8. <https://doi.org/10.1016/j.neuron.2020.11.013>.
- Zipser, K., Lamme, V.A., and Schiller, P.H. (1996). Contextual modulation in primary visual cortex. *J. Neurosci.* 16, 7376–7389. <https://doi.org/10.1523/JNEUROSCI.16-22-07376.1996>.
- Ziomba, C.M., Freeman, J., Simoncelli, E.P., and Movshon, J.A. (2018). Contextual modulation of sensitivity to naturalistic image structure in macaque v2. *J. Neurophysiol.* 120, 409–420. <https://doi.org/10.1152/jn.00900.2017>.
- Angelucci, A., Bijanzadeh, M., Nurminen, L., Federer, F., Merlin, S., and Bressloff, P.C. (2017). Circuits and mechanisms for surround modulation in visual cortex. *Annu. Rev. Neurosci.* 40, 425–451. <https://doi.org/10.1146/annurev-neuro-072116-031418>.
- Dipoppa, M., Ranson, A., Krumin, M., Pachitariu, M., Carandini, M., and Harris, K.D. (2018). Vision and locomotion shape the interactions between neuron types in mouse visual cortex. *Neuron* 98, 602–615.e8. <https://doi.org/10.1016/j.neuron.2018.03.037>.
- Pak, A., Ryu, E., Li, C., and Chubykin, A.A. (2020). Top-down feedback controls the cortical representation of illusory contours in mouse primary visual cortex. *J. Neurosci.* 40, 648–660. <https://doi.org/10.1523/JNEUROSCI.1998-19.2019>.
- Field, D.J., Hayes, A., and Hess, R.F. (1993). Contour integration by the human visual system: Evidence for a local “association field”. *Vision Res.* 33, 173–193. [https://doi.org/10.1016/0042-6989\(93\)90156-Q](https://doi.org/10.1016/0042-6989(93)90156-Q).
- Roelfsema, P.R. (2006). Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.* 29, 203–227. <https://doi.org/10.1146/annurev.neuro.29.051605.112939>.
- Kirchberger, L., Mukherjee, S., Self, M.W., and Roelfsema, P.R. (2023). Contextual drive of neuronal responses in mouse v1 in the absence of feedforward input. *Sci. Adv.* 9, eadd2498. <https://doi.org/10.1126/sciadv.add2498>.
- Kirchberger, L., Mukherjee, S., Schnabel, U.H., van Beest, E.H., Barseganyan, A., Levelt, C.N., Heimel, J.A., Lorteije, J.A.M., van der Togt, C., Self, M.W., and Roelfsema, P.R. (2021). The essential role of recurrent processing for figure-ground perception in mice. *Sci. Adv.* 7, eabe1833. <https://doi.org/10.1126/sciadv.abe1833>.
- Spillmann, L., Dresch-Langley, B., and Tseng, C. (2015). Beyond the classic receptive field: The effect of contextual stimuli. *J. Vis.* 15, 1–22. <https://doi.org/10.1167/15.9.7>.
- Allman, J., Miezin, F., and McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu. Rev. Neurosci.* 8, 407–430. <https://doi.org/10.1146/annurev.ne.08.030185.002203>.
- Hubel, D.H., and Wiesel, T.N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J. Neurophysiol.* 28, 229–289. <https://doi.org/10.1152/jn.1965.28.2.229>.
- Sceniak, M.P., Hawken, M.J., and Shapley, R. (2001). Visual spatial characterization of macaque v1 neurons. *J. Neurophysiol.* 85, 1873–1887. <https://doi.org/10.1152/jn.2001.85.5.1873>.
- Shushruth, S., Ichida, J.M., Levitt, J.B., and Angelucci, A. (2009). Comparison of spatial summation properties of neurons in macaque v1 and v2. *J. Neurophysiol.* 102, 2069–2083. <https://doi.org/10.1152/jn.00512.2009>.
- Coen-Cagli, R., Kohn, A., and Schwartz, O. (2015). Flexible gating of contextual influences in natural vision. *Nat. Neurosci.* 18, 1648–1655. <https://doi.org/10.1038/nn.4128>.
- Barlow, H.B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1, 371–394. <https://doi.org/10.1068/p01037>.
- Barlow, H. (2001). The exploitation of regularities in the environment by the brain. *Behav. Brain Sci.* 24, 602–671. <https://doi.org/10.1017/S0140525X01000024>.
- Polat, U., Mizobe, K., Pettet, M.W., Kasamatsu, T., and Norcia, A.M. (1998). Collinear stimuli regulate visual responses depending on cell’s contrast threshold. *Nature* 391, 580–584. <https://doi.org/10.1038/35372>.
- Paffen, C.L.E., van der Smagt, M.J., Te Pas, S.F., and Verstraten, F.A.J. (2005). Center-surround inhibition and facilitation as a function of size and contrast at multiple levels of visual motion processing. *J. Vis.* 5, 571–578. <https://doi.org/10.1167/5.6.8>.
- Self, M.W., Lorteije, J.A.M., Vangeneugden, J., van Beest, E.H., Grigore, M.E., Levelt, C.N., Heimel, J.A., and Roelfsema, P.R. (2014). Orientation-tuned surround suppression in mouse visual cortex. *J. Neurosci.* 34, 9290–9304. <https://doi.org/10.1523/JNEUROSCI.5051-13.2014>.
- Cavanaugh, J.R., Bair, W., and Movshon, J.A. (2002). Selectivity and spatial distribution of signals from the receptive field surround in macaque v1 neurons. *J. Neurophysiol.* 88, 2547–2556. <https://doi.org/10.1152/jn.00693.2001>.
- Xing, J., and Heeger, D.J. (2001). Measurement and modeling of center-surround suppression and enhancement. *Vision Res.* 41, 571–583. [https://doi.org/10.1016/S0042-6989\(00\)00270-4](https://doi.org/10.1016/S0042-6989(00)00270-4).
- Rubin, N. (2001). Figure and ground in the brain. *Nat. Neurosci.* 4, 857–858. <https://doi.org/10.1038/nn0901-857>.
- Craft, E., Schütze, H., Niebur, E., and von der Heydt, R. (2007). A neural model of figure-ground organization. *J. Neurophysiol.* 97, 4310–4326. <https://doi.org/10.1152/jn.00203.2007>.
- Lamme, V.A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.* 15, 1605–1615. <https://doi.org/10.1523/JNEUROSCI.15-02-01605.1995>.
- Kanizsa, G. (1976). Subjective contours. *Sci. Am.* 234, 48–52. <https://www.jstor.org/stable/24950327>.
- Shin, H., Ogando, M.B., Abdeladim, L., Durand, S., Belski, H., Cabasco, H., Loeffler, H., Bawany, A., Hardcastle, B., Wilkes, J., et al. (2023). Recurrent pattern completion drives the neocortical representation of sensory inference. *bioRxiv*. <https://doi.org/10.1101/2023.06.05.543698>.
- von der Heydt, R. (1987). Approaches to visual cortical function. *Rev. Physiol. Biochem. Pharmacol.* 108, 69–150. <https://doi.org/10.1007/BFb0034072>.
- Peterhans, E., and von der Heydt, R. (1989). Mechanisms of contour perception in monkey visual cortex. ii. contours bridging gaps. *J. Neurosci.* 9, 1749–1763. <https://doi.org/10.1523/JNEUROSCI.09-05-01749.1989>.
- Grosf, D.H., Shapley, R.M., and Hawken, M.J. (1993). Macaque vi neurons can signal ‘illusory’ contours. *Nature* 365, 550–552. <https://doi.org/10.1038/365550a0>.
- Rossi, A.F., Desimone, R., and Ungerleider, L.G. (2001). Contextual modulation in primary visual cortex of macaques. *J. Neurosci.* 21, 1698–1709. <https://doi.org/10.1523/JNEUROSCI.21-05-01698.2001>.
- Jones, H.E., Grieve, K.L., Wang, W., and Sillito, A.M. (2001). Surround suppression in primate v1. *J. Neurophysiol.* 86, 2011–2028. <https://doi.org/10.1152/jn.2001.86.4.2011>.
- Billeh, Y.N., Cai, B., Gratiy, S.L., Dai, K., Iyer, R., Gouwens, N.W., Abbasi-Asl, R., Jia, X., Siegle, J.H., Olsen, S.R., et al. (2020). Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. *Neuron* 106, 388–403.e18. <https://doi.org/10.1016/j.neuron.2020.01.040>.

36. Rossi, L.F., Harris, K.D., and Carandini, M. (2020). Spatial connectivity matches direction selectivity in visual cortex. *Nature* 588, 648–652. <https://doi.org/10.1038/s41586-020-2894-4>.
37. Marques, T., Nguyen, J., Fioreze, G., and Petreanu, L. (2018). The functional organization of cortical feedback inputs to primary visual cortex. *Nat. Neurosci.* 21, 757–764. <https://doi.org/10.1038/s41593-018-0135-z>.
38. Li, Y., and Young, L.-S. (2021). Unraveling the mechanisms of surround suppression in early visual processing. *PLoS Comput. Biol.* 17, e1008916. <https://doi.org/10.1371/journal.pcbi.1008916>.
39. Carandini, M., Demb, J.B., Mante, V., Tolhurst, D.J., Dan, Y., Olshausen, B.A., Gallant, J.L., and Rust, N.C. (2005). Do we know what the early visual system does? *J. Neurosci.* 25, 10577–10597. <https://doi.org/10.1523/JNEUROSCI.3726-05.2005>.
40. Carandini, M., and Heeger, D.J. (2011). Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62. <https://doi.org/10.1038/nrn3136>.
41. Rubin, D.B., Van Hooser, S.D., and Miller, K.D. (2015). The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron* 85, 402–417. <https://doi.org/10.1016/j.neuron.2014.12.026>.
42. Obeid, D., and Miller, K.D. (2021). Stabilized supralinear network: Model of layer 2/3 of the primary visual cortex. *bioRxiv*. <https://doi.org/10.1101/2020.12.30.424892>.
43. Mossing, D.P., Veit, J., Palmigiano, A., Miller, K.D., and Adesnik, H. (2021). Antagonistic inhibitory subnetworks control cooperation and competition across cortical space. *bioRxiv*. <https://doi.org/10.1101/2021.03.31.437953>.
44. Schwabe, L., Obermayer, K., Angelucci, A., and Bressloff, P.C. (2006). The role of feedback in shaping the extra-classical receptive field of cortical neurons: a recurrent network model. *J. Neurosci.* 26, 9117–9129. <https://doi.org/10.1523/JNEUROSCI.1253-06.2006>.
45. Schwabe, L., Ichida, J.M., Shushruth, S., Mangapathy, P., and Angelucci, A. (2010). Contrast-dependence of surround suppression in Macaque V1: experimental testing of a recurrent network model. *Neuroimage* 52, 777–792. <https://doi.org/10.1016/j.neuroimage.2010.01.032>.
46. Shushruth, S., Mangapathy, P., Ichida, J.M., Bressloff, P.C., Schwabe, L., and Angelucci, A. (2012). Strong recurrent networks compute the orientation tuning of surround modulation in the primate primary visual cortex. *J. Neurosci.* 32, 308–321. <https://doi.org/10.1523/JNEUROSCI.3789-11.2012>.
47. Michael, A.M., Parker, P.R.L., and Niell, C.M. (2019). A Hallucinogenic Serotonin-2A Receptor Agonist Reduces Visual Response Gain and Alters Temporal Dynamics in Mouse V1. *Cell Rep.* 26, 3475–3483.e4. <https://doi.org/10.1016/j.celrep.2019.02.104>.
48. Haider, B., Duque, A., Hasenstaub, A.R., and McCormick, D.A. (2006). Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *J. Neurosci.* 26, 4535–4545. <https://doi.org/10.1523/JNEUROSCI.5297-05.2006>.
49. Palmigiano, A., Fumarola, F., Mossing, D.P., Kraynyukova, N., Adesnik, H., and Miller, K.D. (2020). Common rules underlying optogenetic and behavioral modulation of responses in multi-cell-type v1 circuits. *bioRxiv*. <https://doi.org/10.1101/2020.11.11.378729>.
50. Denève, S., and Machens, C.K. (2016). Efficient codes and balanced networks. *Nat. Neurosci.* 19, 375–382. <https://doi.org/10.1038/nn.4243>.
51. Sanzeni, A., Akitake, B., Goldbach, H.C., Leedy, C.E., Brunel, N., and Histed, M.H. (2020). Inhibition stabilization is a widespread property of cortical networks. *Elife* 9, e54875. <https://doi.org/10.1101/656710>.
52. Ahmadian, Y., and Miller, K.D. (2021). What is the dynamical regime of cerebral cortex? *Neuron* 109, 3373–3391. <https://doi.org/10.1016/j.neuron.2021.07.031>.
53. Isaacson, J.S., and Scanziani, M. (2011). How inhibition shapes cortical activity. *Neuron* 72, 231–243. <https://doi.org/10.1016/j.neuron.2011.09.027>.
54. Okun, M., and Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nat. Neurosci.* 11, 535–537. <https://doi.org/10.1038/nn.2105>.
55. Ozeki, H., Finn, I.M., Schaffer, E.S., Miller, K.D., and Ferster, D. (2009). Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron* 62, 578–592. <https://doi.org/10.1016/j.neuron.2009.03.028>.
56. Renart, A., De La Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K.D. (2010). The asynchronous state in cortical circuits. *Science* 327, 587–590. <https://doi.org/10.1126/science.1179850>.
57. Adesnik, H. (2017). Synaptic Mechanisms of Feature Coding in the Visual Cortex of Awake Mice. *Neuron* 95, 1147–1159.e4. <https://doi.org/10.1016/j.neuron.2017.08.014>.
58. Pfeffer, C.K., Xue, M., He, M., Huang, Z.J., and Scanziani, M. (2013). Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat. Neurosci.* 16, 1068–1076. <https://doi.org/10.1038/nn.3446>.
59. Campagnola, L., Seeman, S.C., Chartrand, T., Kim, L., Hoggarth, A., Gamlin, C., Ito, S., Trinh, J., Davoudian, P., Radaelli, C., et al. (2022). Local connectivity and synaptic dynamics in mouse and human neocortex. *Science* 375, eabj5861. <https://doi.org/10.1126/science.abj5861>.
60. Nienborg, H., Hasenstaub, A., Nauhaus, I., Taniguchi, H., Huang, Z.J., and Callaway, E.M. (2013). Contrast dependence and differential contributions from somatostatin- and parvalbumin-expressing neurons to spatial integration in mouse v1. *J. Neurosci.* 33, 11145–11154. <https://doi.org/10.1523/JNEUROSCI.5320-12.2013>.
61. Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z.J., and Scanziani, M. (2012). A neural circuit for spatial summation in visual cortex. *Nature* 490, 226–231. <https://doi.org/10.1038/nature11526>.
62. Vangeneugden, J., van Beest, E.H., Cohen, M.X., Lorteije, J.A.M., Mukherjee, S., Kirchberger, L., Montijn, J.S., Thamizharasu, P., Camillo, D., Levelt, C.N., et al. (2019). Activity in lateral visual areas contributes to surround suppression in awake mouse v1. *Curr. Biol.* 29, 4268–4275.e7. <https://doi.org/10.1016/j.cub.2019.10.037>.
63. Miller, K.D., and Troyer, T.W. (2002). Neural noise can explain expansive, power-law nonlinearities in neural response functions. *J. Neurophysiol.* 87, 653–659. <https://doi.org/10.1152/jn.00425.2001>.
64. Hansel, D., and van Vreeswijk, C. (2002). How noise contributes to contrast invariance of orientation tuning in cat visual cortex. *J. Neurosci.* 22, 5118–5128. <https://doi.org/10.1523/JNEUROSCI.22-12-05118.2002>.
65. Sillito, A.M., Grieve, K.L., Jones, H.E., Cudeiro, J., and Davis, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* 378, 492–496. <https://doi.org/10.1038/378492a0>.
66. Ko, H., Cossell, L., Baragli, C., Antolik, J., Clopath, C., Hofer, S.B., and Mrsic-Flogel, T.D. (2013). The emergence of functional microcircuits in visual cortex. *Nature* 496, 96–100. <https://doi.org/10.1038/nature12015>.
67. Ko, H., Hofer, S.B., Pichler, B., Buchanan, K.A., Sjöström, P.J., and Mrsic-Flogel, T.D. (2011). Functional specificity of local synaptic connections in neocortical networks. *Nature* 473, 87–91. <https://doi.org/10.1038/nature09880>.
68. Keemink, S.W., Boucsein, C., and van Rossum, M.C.W. (2018). Effects of v1 surround modulation tuning on visual saliency and the tilt illusion. *J. Neurophysiol.* 120, 942–952. <https://doi.org/10.1152/jn.00864.2017>.
69. Angelucci, A., and Bressloff, P.C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate v1 neurons. *Prog. Brain Res.* 154, 93–120. [https://doi.org/10.1016/S0079-6123\(06\)54005-1](https://doi.org/10.1016/S0079-6123(06)54005-1).
70. Schwartz, G.W. (2021). Chapter 8 - surround suppression. In *Retinal Computation*, G.W. Schwartz, ed. (Academic Press), pp. 144–160. <https://doi.org/10.1016/B978-0-12-819896-4.00009-3>.

71. Jacoby, J., and Schwartz, G.W. (2017). Three small-receptive-field ganglion cells in the mouse retina are distinctly tuned to size, speed, and object motion. *J. Neurosci.* 37, 610–625. <https://doi.org/10.1523/JNEUROSCI.2804-16.2016>.
72. Bonin, V., Mante, V., and Carandini, M. (2005). The suppressive field of neurons in lateral geniculate nucleus. *J. Neurosci.* 25, 10844–10856. <https://doi.org/10.1523/JNEUROSCI.3562-05.2005>.
73. Alitto, H.J., and Usrey, W.M. (2008). Origin and dynamics of extraclassical suppression in the lateral geniculate nucleus of the macaque monkey. *Neuron* 57, 135–146. <https://doi.org/10.1016/j.neuron.2007.11.019>.
74. Alitto, H.J., and Usrey, W.M. (2015). Surround suppression and temporal processing of visual signals. *J. Neurophysiol.* 113, 2605–2617. <https://doi.org/10.1152/jn.00480.2014>.
75. Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput. Biol.* 4, e1000211. <https://doi.org/10.1371/journal.pcbi.1000211>.
76. Siegle, J.H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T.K., Choi, H., Luviano, J.A., et al. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* 592, 86–92. <https://doi.org/10.1038/s41586-020-03171-x>.
77. Yousif, S.R., and Keil, F.C. (2019). The additive-area heuristic: An efficient but illusory means of visual area approximation. *Psychol. Sci.* 30, 495–503. <https://doi.org/10.1177/0956797619831617>.
78. Yousif, S.R., and Keil, F.C. (2021). How we see area and why it matters. *Trends Cogn. Sci.* 25, 554–557. <https://doi.org/10.1016/j.tics.2021.03.017>.
79. Tohmi, M., Meguro, R., Tsukano, H., Hishida, R., and Shibuki, K. (2014). The extrageniculate visual pathway generates distinct response properties in the higher visual areas of mice. *Curr. Biol.* 24, 587–597. <https://doi.org/10.1016/j.cub.2014.01.061>.
80. Harris, J.A., Mihalas, S., Hirokawa, K.E., Whitesell, J.D., Choi, H., Bernard, A., Bohn, P., Caldejon, S., Casal, L., Cho, A., et al. (2019). Hierarchical organization of cortical and thalamic connectivity. *Nature* 575, 195–202. <https://doi.org/10.1038/s41586-019-1716-z>.
81. Blot, A., Roth, M.M., Gasler, I., Javadzadeh, M., Imhof, F., and Hofer, S.B. (2021). Visual intracortical and transthalamic pathways carry distinct information to cortical areas. *Neuron* 109, 1996–2008.e6. <https://doi.org/10.1016/j.neuron.2021.04.017>.
82. Morales, G.B., Santo, S.D., and Muñoz, M.A. (2021). Quasiuniversal scaling in mouse-brain neuronal activity stems from edge-of-instability critical dynamics. *Proc. Natl. Acad. Sci. USA* 120, e2208998120. <https://doi.org/10.1073/pnas.2208998120>.
83. Morrell, M.C., Sederberg, A.J., and Nemenman, I. (2021). Latent dynamical variables produce signatures of spatiotemporal criticality in large biological systems. *Phys. Rev. Lett.* 126, 118302. <https://doi.org/10.1103/PhysRevLett.126.118302>.
84. Eagleman, D.M. (2001). Visual illusions and neurobiology. *Nat. Rev. Neurosci.* 2, 920–926. <https://doi.org/10.1038/35104092>.
85. Mély, D., Linsley, D., and Serre, T. (2018). Complementary surrounds explain diverse contextual phenomena across visual modalities. *Psychol. Rev.* 125, 769–784. <https://doi.org/10.1037/rev0000109>.
86. Agrochao, M., Tanaka, R., Salazar-Gatzimas, E., and Clark, D.A. (2020). Mechanism for analogous illusory motion perception in flies and humans. *Proc. Natl. Acad. Sci. USA* 117, 23044–23053. <https://doi.org/10.1073/pnas.2002937117>.
87. Agrillo, C., Gori, S., and Beran, M.J. (2015). Do rhesus monkeys (macaca mulatta) perceive illusory motion? *Anim. Cogn.* 18, 895–910. <https://doi.org/10.1007/s10071-015-0860-6>.
88. Bååth, R., Seno, T., and Kitaoka, A. (2014). Cats and illusory motion. *Psychology* 5, 1131–1134. <https://doi.org/10.4236/psych.2014.59125>.
89. Gori, S., Agrillo, C., Dadda, M., and Bisazza, A. (2014). Do fish perceive illusory motion? *Sci. Rep.* 4, 6443. <https://doi.org/10.1038/srep06443>.
90. Endler, J.A., Endler, L.C., and Doerr, N.R. (2010). Great bowerbirds create theaters with forced perspective when seen by their audience. *Curr. Biol.* 20, 1679–1684. <https://doi.org/10.1016/j.cub.2010.08.033>.
91. Santacà, M., Miletto Petrazzini, M.E., Agrillo, C., and Wilkinson, A. (2019). Can reptiles perceive visual illusions? Delboeuf illusion in red-footed tortoise (*Chelonoidis carbonaria*) and bearded dragon (*Pogona vitticeps*). *J. Comp. Psychol.* 133, 419–427. <https://doi.org/10.1037/com0000176>.
92. Priebe, N.J., and Ferster, D. (2008). Inhibition, spike threshold and stimulus selectivity in primary visual cortex. *Neuron* 57, 482–497. <https://doi.org/10.1016/j.neuron.2008.02.005>.
93. Persi, E., Hansel, D., Nowak, L., Barone, P., and van Vreeswijk, C. (2011). Power-law input-output transfer functions explain the contrast-response and tuning properties of neurons in visual cortex. *PLoS Comput. Biol.* 7, e1001078. <https://doi.org/10.1371/journal.pcbi.1001078>.
94. Kaschube, M. (2014). Neural maps versus salt-and-pepper organization in visual cortex. *Curr. Opin. Neurobiol.* 24, 95–102. <https://doi.org/10.1016/j.conb.2013.08.017>.
95. Tremblay, R., Lee, S., and Rudy, B. (2016). Gabaergic interneurons in the neocortex: from cellular properties to circuits. *Neuron* 91, 260–292.
96. Karnani, M.M., Jackson, J., Ayzenshtat, I., Tucciarone, J., Manoocheri, K., Snider, W.G., and Yuste, R. (2016). Cooperative subnetworks of molecularly similar interneurons in mouse neocortex. *Neuron* 90, 86–100. <https://doi.org/10.1016/j.neuron.2016.02.037>.
97. Karnani, M.M., Jackson, J., Ayzenshtat, I., Hamzehei Sichani, A., Manoocheri, K., Kim, S., and Yuste, R. (2016). Opening holes in the blanket of inhibition: localized lateral disinhibition by vip interneurons. *J. Neurosci.* 36, 3471–3480. <https://doi.org/10.1523/JNEUROSCI.3646-15.2016>.
98. Fu, Y., Tucciarone, J.M., Espinosa, J.S., Sheng, N., Darcy, D.P., Nicoll, R.A., Huang, Z.J., and Stryker, M.P. (2014). A cortical circuit for gain control by behavioral state. *Cell* 156, 1139–1152. <https://doi.org/10.1016/j.cell.2014.01.050>.
99. Garcia-Junco-Clemente, P., Ikrar, T., Tring, E., Xu, X., Ringach, D.L., and Trachtenberg, J.T. (2017). An inhibitory pull–push circuit in frontal cortex. *Nat. Neurosci.* 20, 389–392. <https://doi.org/10.1038/nn.4483>.
100. Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J.I., Huang, Z.J., and Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control. *Nature* 503, 521–524. <https://doi.org/10.1038/nature12676>.
101. Jiang, X., Shen, S., Cadwell, C.R., Berens, P., Sinz, F., Ecker, A.S., Patel, S., and Tolias, A.S. (2015). Principles of connectivity among morphologically defined cell types in adult neocortex. *Science* 350, aac9462. <https://doi.org/10.1126/science.aac9462>.
102. Samonds, J.M., Feese, B.D., Lee, T.S., and Kuhlman, S.J. (2017). Nonuniform surround suppression of visual responses in mouse v1. *J. Neurophysiol.* 118, 3282–3292. <https://doi.org/10.1152/jn.00172.2017>.
103. Iacaruso, M.F., Gasler, I.T., and Hofer, S.B. (2017). Synaptic organization of visual space in primary visual cortex. *Nature* 547, 449–452. <https://doi.org/10.1038/nature23019>.
104. Bos, H., Oswald, A.-M., and Doiron, B. (2020). Untangling stability and gain modulation in cortical circuits with multiple interneuron classes. *bioRxiv*. <https://doi.org/10.1101/2020.06.15.148114>.
105. Chettih, S.N., and Harvey, C.D. (2019). Single-neuron perturbations reveal feature-specific competition in V1. *Nature* 567, 334–340. <https://doi.org/10.1038/s41586-019-0997-6>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Keller et al. ¹	Keller et al. ¹	https://doi.org/10.1038/s41586-020-2319-4
Software and algorithms		
GitHub code serenadisanto14/ContextualEffectsVC-v1.0.0	Zenodo	https://doi.org/10.5281/zenodo.13957280

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The responses of thousands of neurons in the primary visual cortex of adult mice of either sex are recorded through 2-photon calcium imaging during visual stimulation. Stimulation time was 2s interleaved by 4s of gray screen. The response amplitude to a stimulus was computed as the average response over the duration of the stimulus presentation. Stimuli were presented at 100% contrast. To estimate the center of the receptive field, the responses to patches of gratings (10°) presented along a grid (15° spacing) were fitted with a two-dimensional Gaussian. The size of the CRF was approximated by the size of the patch of gratings evoking the largest response. The number of recorded cells is 1489 Pyr, 25 PV, 105 SOM, 90 VIP, 38 L4, and 167 LM). The maximum distance between the CRF recorded and the stimulus center is 40°. Further experimental details are available in.¹

METHOD DETAILS

Nonlinear transfer function

We assume neurons have a supralinear input/output function (taken to be rectified quadratic), describing the steady-state firing rate induced by a given net input to the neuron.⁹² Such an expansive input/output function is expected when firing activity is driven by input fluctuations rather than by the mean input,^{63,64} and has been shown to yield network behavior that reproduces a variety of nonlinear visual cortical behaviors.^{41,52} The use of these power-law input/output functions, combined with our assumptions that connectivity and firing-rate fields are described by Gaussian functions, allows us, building on previous theoretical work,⁹³ to develop an approximate analytic solution to the one-population model, which yields deeper insight into the mechanisms driving model behavior.

Feedback input

A crucial aspect of the model is to include inputs from HVAs. A difficulty in doing so is that cells in L2/3 are targeted by all of the HVAs,^{35,76} but we have recordings of the V1 boutons from only one HVA, area LM.¹ Extended Data Fig. 9 in¹ shows that optogenetic silencing of different HVAs generates quantitatively different reductions of classical versus inverse responses. We assume that the HVAs considered altogether have a rate field profile proportional to that recorded in LM, but with proportionality constants that can differ between the classical and inverse stimulus conditions. In Figure S10 we show how we infer these constants based on the amplitudes of inverse vs. classical response in L2/3.

The minimal model

We consider a system with one recurrent population defined on a two-dimensional retinotopic space of coordinates $\{\mathbf{x}\}$, whose rate field is $r(\mathbf{x})$ and which receives a visually-driven input described by a 2D isotropic Gaussian function $I(\mathbf{x}) = I_0 G(\mathbf{x}, v)$ (where $G(\mathbf{x}, v) \equiv \frac{1}{2\pi v} e^{-\frac{\mathbf{x}^2}{2v}}$). Let the recurrent connections be $W(\mathbf{x} - \mathbf{y}) = W_0 G(\mathbf{x} - \mathbf{y}, v_r)$. We consider the SSN dynamical equations for the system, $\tau dr(\mathbf{x})/dt = -r(\mathbf{x}) + [u(\mathbf{x})]_+^2$ where $u(\mathbf{x})$, the total input current, i.e., the sum of the visually driven and recurrent input. We will hereafter focus on the steady state of the system, where $r(\mathbf{x}) = [u(\mathbf{x})]_+^2$, or:

$$r(\mathbf{x}) = \left[I(\mathbf{x}) + \int_{-\infty}^{\infty} d\mathbf{y} W(\mathbf{x} - \mathbf{y}) r(\mathbf{y}) \right]^2 \quad (\text{Equation 1})$$

In Equation 1 we omit the rectification sign, based on the assumption that $u(\mathbf{x})$ is always greater than or equal to 0. We will show that the results we then derive obey this assumption.

Analogously we can write the SSN equations for the total input current at steady state as:

$$u(\mathbf{x}) = I(\mathbf{x}) + \int d\mathbf{y} W(\mathbf{x} - \mathbf{y}) u^2(\mathbf{y}) \quad (\text{Equation 2})$$

To find $u(\mathbf{x})$ we make the ansatz that $u(\mathbf{x})$ has the following form:

$$u(\mathbf{x}) = \alpha(\mathbf{x}, v_u)G(\mathbf{x}, v_u) \quad (\text{Equation 3})$$

where $\alpha(\mathbf{x}, v_u)$ is required to vary slowly as a function of \mathbf{x} . We will henceforth write it simply as $\alpha(\mathbf{x})$, leaving the dependence on v_u implicit. Note that v_u is a free parameter that is constrained only by the fact that when we solve our equation the resulting $\alpha(\mathbf{x})$ must be slowly varying. Then the recurrent term is:

$$\text{Rec}(\mathbf{x}) = \int d\mathbf{y}W(\mathbf{x} - \mathbf{y})u^2(\mathbf{y}) = W_0 \int d\mathbf{y}\alpha^2(\mathbf{y})G(\mathbf{x} - \mathbf{y}, v_r)G^2(\mathbf{y}, v_u) = \frac{W_0}{4\pi v_u} \int d\mathbf{y}\alpha^2(\mathbf{y})G(\mathbf{x} - \mathbf{y}, v_r)G\left(\mathbf{y}, \frac{v_u}{2}\right) \quad (\text{Equation 4})$$

The largest contribution to the integral is given by the maximum of its argument (Laplace approximation), but since $\alpha^2(\mathbf{y})$ varies slowly ($|\nabla\alpha(\mathbf{y})| \ll |\alpha(\mathbf{y})|$), the maximum of the full argument can be approximated with the maximum of $G(\mathbf{x} - \mathbf{y}, v_r)G\left(\mathbf{y}, \frac{v_u}{2}\right)$, which occurs at $\mathbf{y} = \bar{\mathbf{y}} \equiv \frac{v_u\mathbf{x}}{2v_r + v_u}$:

$$\text{Rec}(\mathbf{x}) \approx \frac{W_0\alpha^2(\bar{\mathbf{y}})}{4\pi v_u} \int d\mathbf{y}G(\mathbf{x} - \mathbf{y}, v_r)G\left(\mathbf{y}, \frac{v_u}{2}\right) = \frac{W_0}{4\pi v_u}\alpha^2(\bar{\mathbf{y}})G(\mathbf{x}, v_{ru}) \quad (\text{Equation 5})$$

where we defined $v_{ru} = v_r + \frac{v_u}{2}$. Since $\alpha(\mathbf{y})$ varies slowly with its argument, we take the further approximation that $\alpha(\bar{\mathbf{y}})$ is well approximated by $\alpha(\mathbf{x})$. From Equation 2 we get:

$$\alpha(\mathbf{x})G(\mathbf{x}, v_u) \approx I_0G(\mathbf{x}, v) + \frac{W_0}{4\pi v_u}\alpha^2(\mathbf{x})G(\mathbf{x}, v_{ru}), \quad (\text{Equation 6})$$

We now solve the equation above for $\alpha(\mathbf{x})$ and get:

$$\alpha(\mathbf{x}) \approx \frac{2\pi v_u}{W_0G(v_{ru})} \left(G(v_u) \pm \sqrt{G^2(v_u) - \frac{I_0W_0G(v)G(v_{ru})}{\pi v_u}} \right) \quad (\text{Equation 7})$$

where for better readability we are omitting the spatial dependence of the function $G(\cdot)$. In Figure S13 we show that $|\nabla\alpha(\mathbf{x})| \ll |\alpha(\mathbf{x})|$.

We now choose v_u to satisfy:

$$2v_u^{-1} = v^{-1} + v_{ru}^{-1} \quad (\text{Equation 8})$$

This simplifies the expression under the square root, making it independent on \mathbf{x} :

$$\frac{G(v)G(v_{ru})}{G^2(v_u)} = \frac{v_u^2}{vv_{ru}} \quad (\text{Equation 9})$$

and thus we can write:

$$\alpha(\mathbf{x}) \approx \frac{2\pi v_u}{W_0G(v_{ru})} G(v_u) \left(1 \pm \sqrt{1 - \frac{I_0W_0v_u}{\pi vv_{ru}}} \right) \quad (\text{Equation 10})$$

Finally, using Equation 9 again:

$$u(\mathbf{x}) \approx \frac{2\pi vv_{ru}}{W_0v_u} \left(1 \pm \sqrt{1 - \frac{I_0W_0v_u}{\pi vv_{ru}}} \right) G(v) = \frac{\pi(v+v_{ru})}{W_0} \left(1 \pm \sqrt{1 - \frac{2I_0W_0}{\pi(v+v_{ru})}} \right) G(v) \quad (\text{Equation 11})$$

Note that only one of the two solutions is actually acceptable, depending on whether the recurrent population described is excitatory or inhibitory. The only acceptable solution is the one with $-$ for negative W_0 (because the one with $+$ gives $u(\mathbf{x}) < 0$). For positive W_0 , instead, the solution requires that $2I_0W_0 < \pi(v + v_{ru})$. Violation of this condition means that recurrent excitatory input is very large ($W_0 > \pi(v + v_{ru})/2I_0$), and in this case we find numerically that the system diverges (see Figure S14).

Analogously, if the external input can be described by a sum of Gaussians:

$$I(\mathbf{x}) = \sum_{k=1}^m I_k G(\mathbf{x}, v_k) \quad (\text{Equation 12})$$

making the same ansatz as above, we obtain

$$u(\mathbf{x}) = \frac{\pi(v_1+v_{ru})}{W_0} \left(1 - \sqrt{1 - \frac{2I_1W_0}{\pi(v_1+v_{ru})} - 4v_1^2W_0 \sum_{k=1}^m \frac{I_k G(v_{1k})}{(v_1 - v_k)v_{ru}}} \right) G(v_1), \quad (\text{Equation 13})$$

where we defined

$$\begin{aligned} v_u &= \frac{2v_1 v_{ru}}{v_1 + v_{ru}} \\ v_{1k} &= \frac{v_1 v_k}{v_1 - v_k}. \end{aligned} \tag{Equation 14}$$

Note that, despite the approximations, the analytic solution provides a really good description of the rate field (see e.g. Figures 3C, 3E, 5A, and 5C).

Validation of the minimal model

To solve the self-consistency equations for the spatial SSN (Equation 2) we took the ansatz (Equation 3), with the constraint that $\alpha(\mathbf{x})$ is a slowly varying function of its argument. Here we show that the solution that we find for $\alpha(\mathbf{x})$ is actually consistent with the assumption that it varies slowly with \mathbf{x} . Analytically, if we consider only one external input $I(\mathbf{x}) = I_0 G(\mathbf{x}, v)$ and we restrict to one spatial dimension for illustration purposes, we find:

$$\frac{1}{\alpha(x, 0)} \frac{\partial \alpha(x, 0)}{\partial x} = \frac{x \left(-3 + \sqrt{\frac{4v}{v_r} + 1} \right)}{v} = \mathcal{O}\left(\frac{x}{v}\right) \tag{Equation 15}$$

which is very small for the values that are relevant in this context, i.e., $|x| < 50$, $v \sim 20^2$, as shown in Figure S13.

The minimal model with orientation tuning

The minimal model can be extended to account for orientation tuning. Given the salt-and-pepper structure of preferred orientation in mouse V1,⁹⁴ we assume that orientation-preference is independent of the CRF location³⁵: in every retinotopic location all possible preferred orientations are represented. Thus we consider spatial locations and orientation as orthogonal degrees of freedom. Here we reduce the retinotopic space to 1D, for simplicity and without entailing any conceptual change (whereas the results presented in the main text for surround facilitation are obtained for 2 spatial dimensions plus one orientation dimension, for consistency with the other results discussed). The rate field of the recurrent layer is then $r(x, \theta)$. Let us assume that it receives a Gaussian input $I(x, \theta - \psi) = I_0 G(x, v) G_p(\theta - \psi, \lambda)$ at orientation $\psi \in (0, \pi)$, with $G_p(\theta, \lambda) = \frac{1}{\sqrt{2\pi\lambda}} \sum_{k=-\infty}^{\infty} e^{-\frac{(\theta - k\pi)^2}{2\lambda}}$ being a periodic Gaussian⁹³ and $G(x, v) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{x^2}{2v}}$. Let the recurrent connections be $W(x, y, \theta, \phi) = W_0 G(x - y, v_r) G_p(\theta - \phi, \lambda_r)$, and let the response function be rectified quadratic. We set $\sqrt{\lambda_r} = \pi/6$ based on estimates in.³⁶ The SSN equations for the system can be written analogously to Eq1:

$$r(x, \theta) = \left[I(x, \theta) + \int dy d\phi W(x, y, \theta, \phi) r(y, \phi) \right]^2 \tag{Equation 16}$$

and the exact same steps as above can be carried out. Obviously we will take an orientation-tuned ansatz:

$$u(x) = \alpha(x, \theta) G(x, v_u) G_p(\theta - \psi, \lambda_u) \tag{Equation 17}$$

where $\alpha(x, \theta)$ is a slowly varying function of both its arguments. All the calculations then are completely analogous to the purely spatial case. (The convolution of two periodic Gaussians is a periodic Gaussian whose variance is the sum of their variances. Also, note that in the case of periodic Gaussians, the equivalent of Equation 8 holds only approximately, but it is a very good approximation, when λ is sufficiently small compared to π .) We define $\lambda_{ru} = \lambda_r + \frac{\lambda_u}{2}$ and consider $2\lambda_u^{-1} \approx \lambda^{-1} + \lambda_{ru}^{-1}$ then we find:

$$u(x) \approx \frac{\pi}{W_0} \sqrt{\frac{v v_{ru} \lambda \lambda_{ru}}{\lambda_u v_u}} \left(1 \pm \sqrt{1 - 2I_0 W_0 \sqrt{\frac{v_u \lambda_u}{\pi v v_{ru} \lambda \lambda_{ru}}}} \right) G(x, v) G_p(\theta - \psi, \lambda) \tag{Equation 18}$$

The full model

The operating regime of cortical computations emerges from the interplay of Pyramidal cells with multiple interneuron types. In mouse V1, approximately 80% of these interneurons are PV, SOM, or VIP cells.^{58,95} In this Section we show that our results hold in a network model that includes all four of these cell types. We call this the ‘full model’.

Since the synaptic coupling weights are not easily measured in the lab e.g., see varying results in,^{35,58,96} we infer them as in,^{2,6,43,49} together with threshold offsets specific to each cell type, by constraining the model to reproduce the recurrent rate fields recorded in the classical stimulus condition. In order to avoid a proliferation of parameters: i) we set to zero the connections that have been shown experimentally to be very small^{35,58,59,96-101} as in^{2,6} and ii) we fix the connection widths based on anatomical measurements^{35,36,96} (see Figure S30).

We develop a semianalytical non-negative-least-squares approach to fit the free parameters of the model. Since all the inputs that SOM cells receive in the classical stimulus condition are surround suppressed, yet SOM cells are not, it is particularly hard to recover SOM size tuning properties. Possible solutions to this problem are to consider divisive inhibition from SOM to VIP⁶ or additional inputs that grow with stimulus size.⁴³ Here for simplicity we adopt the latter solution. A possible explanation is that SOM cells receive input from the surround that is stronger and/or has different orientation preference than the input from the center.^{37,61,102,103} In Figure S31 we show the size of the extra input relative to the net other input.

We fit the model to reproduce the classical response profiles. We find a collection of parameter sets that fit the data well (see STAR Methods on model comparison). In Figure S30 we plot one of the inferred parameter sets, and the resulting match of experimental and model classical size tuning curves. With the inferred parameters, and given the external inputs from HVAs and L4 and the increase in strength of HVAs input for inverse vs. classical stimuli, the model quantitatively reproduces the inverse response and size tuning of all cell types in L2/3 (Figure S30D). This confirms that the inputs (in particular the input from HVAs) play a fundamental role in shaping the inverse patterns of activity, as suggested in¹ and clarified by the minimal model. In Figure S33, we show that the results presented here are robust against reasonable changes in the connection width, accounting for errors in the experimental estimates considered. In Figure S34–S38 we also show that the full model is stabilized by PV cells and that it recovers the modulations induced by changing the stimulus contrast, by silencing HVAs, and by suppressing SOM cells.

The full model can be leveraged to test the robustness of the mechanisms uncovered through the minimal model. In particular Figure S38 confirms that inverse response and size tuning in the full model require: i) broad enough feedback connections, ii) growth of the width of the HVAs rate field. Therefore the principles discussed and analyzed in the main text for the minimal model extend to this more biologically realistic framework.

Inferring parameters of the full model

In the full model the recurrent layer is composed of Pyramidal neurons (Pyr or E) as well as the 3 main interneuron classes, Parvalbumin (PV or P), Somatostatin (SOM or S) and Vaso-intestinal peptide (VIP or V), which together account for about 80 – 90% of the neurons in mouse V1.^{58,95} In addition, as in the minimal model, we consider external feedforward input from L4 (L) and external feedback input from HVAs (H). In order to keep the number of parameters as small as possible, we consider rectified quadratic transfer functions for all celltypes and we fix the projection span of each connection based on anatomical observations (see STAR Methods on Anatomical length scales). Since the firing rates are not 0, we assume that the total input current to each population is positive. Then we can write the following equation for the rate field of population A ($A \in \{E, P, S, V\}$; note that L4 and HVAs are taken as external inputs with rate fields taken from experimental measurement):

$$\sqrt{r_A(\mathbf{x})} = \sum_B \int W_{AB}(\mathbf{x} - \mathbf{y}) r_B(\mathbf{y}) d\mathbf{y} + T_A \quad (\text{Equation 19})$$

where $B = \{E, P, S, V, L, H\}$ and \mathbf{x} denotes 2 dimensional cortical or –equivalently for the model– retinotopic position. The term T_A is a bias that we consider constant (across space and stimulus size), which can be understood as a cell-type-specific firing threshold.

As in the minimal model, we consider Gaussian assumptions and ansatz:

$$W_{AB}(\mathbf{x} - \mathbf{y}) = \frac{\overline{W}_{AB}}{2\pi v_{AB}} e^{-\frac{(\mathbf{x}-\mathbf{y})^2}{2v_{AB}}} \quad (\text{Equation 20})$$

$$r_A(\mathbf{x}) = \bar{r}_A e^{-\frac{x^2}{2v_A}} + b_A,$$

where b is the baseline, i.e., the firing rate in absence of inputs, uniform across space. Here we aim to solve for the \overline{W}_{AB} given the $r_A(\mathbf{x})$ (see Figures S2–S4) and given the values of v_{AB} . In order to regress the connection weights, we set up a non negative least squares (NNLS) approach similarly to.^{2,6} Let us consider first the NNLS procedure when we constrain only on one stimulus size. Then the loss function reads⁴:

$$S_A = \int_{-\infty}^{\infty} d\mathbf{x} \left[\sqrt{r_A(\mathbf{x})} - \sum_B \int W_{AB}(\mathbf{x} - \mathbf{y}) r_B(\mathbf{y}) d\mathbf{y} - T_A \right]^2 \quad (\text{Equation 21})$$

$$= \int_{-\infty}^{\infty} d\mathbf{x} \left[\sqrt{\bar{r}_A e^{-\frac{x^2}{2v_A}} + b_A} - \sum_B \overline{W}_{AB} \left(r_B \frac{v_B}{v_{AB+}} e^{-\frac{x^2}{2v_{AB+}} + b_B} \right) - T_A \right]^2$$

where we defined $v_{AB+} = v_{AB} + v_B$. It is easy to verify numerically that

$$\sqrt{\bar{r}_A e^{-\frac{x^2}{2v_A}} + b_A} \approx q_A e^{-\frac{x^2}{4v_A}} + \sqrt{b_A}, \quad (\text{Equation 22})$$

is a very good approximation, where we defined $q_A = \sqrt{r_A + b_A} - \sqrt{b_A}$. The NNLS equations consist in minimizing the loss function (which codifies the error between the data and the model) with respect to the parameters. Thus when we minimize with respect to \overline{W}_{AB} we obtain:

$$-2 \int_{-\infty}^{\infty} d\mathbf{x} \left[q_A e^{-\frac{x^2}{2v_A}} - \sum_C \overline{W}_{AC} \left(\overline{r}_C \frac{v_C}{v_{AC+}} e^{-\frac{x^2}{2v_{AC+}} + b_C} \right) + \sqrt{b_A} - T_A \right] \left(\overline{r}_B \frac{v_B}{v_{AB+}} e^{-\frac{x^2}{2v_{AB+}} + b_B} \right) = 0$$

$$\frac{\overline{r}_B v_B 2v_A q_A}{v_{AB+} + 2v_A} - \overline{r}_B v_B \sum_C \overline{W}_{AC} \frac{\overline{r}_C v_C}{v_{AB+} + v_{AC+}} - \overline{r}_B v_B \sum_C \overline{W}_{AC} b_C + \overline{r}_B v_B (\sqrt{b_A} - T_A) + b_B q_A 2v_A - b_B \sum_C \overline{W}_{AC} \overline{r}_C v_C \quad (\text{Equation 23})$$

$$- \int_{-\infty}^{\infty} d\mathbf{x} \frac{b_B}{2\pi} \left(\sum_C \overline{W}_{AC} b_C - (\sqrt{b_A} - T_A) \right) = 0$$

where we defined $v_{AB+} = v_{AB} + v_B/2$. Note that the second derivatives are always negative, thus the problem is always convex. Also note that these are 6 equations for each of the \overline{W}_{AB} with $B = \{E, P, S, V, L, H\}$, i.e., one equation for each of the connection weights that enter into the self-consistent equation for the firing rate of population A. Since the space is continuous and infinite, to ensure convergence of the loss function, we have to impose:

$$\sqrt{b_A} - T_A = \sum_C \overline{W}_{AC} b_C \quad (\text{Equation 24})$$

Conceptually this means that if we knew the \overline{W}_{AB} , then the knowledge of the baseline firings (and the transfer functions) would determine the firing thresholds. Moreover when we take the first derivative of the loss function with respect to T_A and we take into account Equation 24 we obtain:

$$q_A 2v_A - \sum_C \overline{W}_{AC} \overline{r}_C v_C = 0 \quad (\text{Equation 25})$$

Note that we make repeated use of the useful integrals:

$$\int e^{-\frac{(\mathbf{x}-\mathbf{y})^2}{2v_1}} e^{-\frac{y^2}{2v_2}} d\mathbf{y} = \frac{2\pi v_1 v_2}{v_1 + v_2} e^{-\frac{x^2}{2(v_1+v_2)}}$$

$$\int e^{-\frac{x^2}{2v_1}} e^{-\frac{x^2}{2v_2}} d\mathbf{x} = \frac{2\pi v_1 v_2}{v_1 + v_2} \quad (\text{Equation 26})$$

Finally, using both Equations 24 and 25 in Equation 23, we obtain

$$\sum_C (\overline{W}_{AC} \overline{r}_C v_C) \left(\frac{1}{v_{AB+} + 2v_A} - \frac{1}{v_{AC+} + v_{AB+}} \right) = 0. \quad (\text{Equation 27})$$

We can solve this linear equation to find \overline{W}_{AC} and then find T_A from Equation 24. If we solve the linear system in Equation 27 without any further caution, we find an error function that is strictly 0, but values of the parameters that are not compatible with the excitatory or inhibitory nature of the different populations. Therefore we devise an algorithm to solve this system of linear equations which constrains the positivity of each element W_{AB} . We start from a random initial set of W_{AB} that satisfies our positivity constraints and solve one equation at a time. We update the initial parameter set if two conditions are verified: i) the solution still satisfies the positivity constraints and ii) the loss function is actually smaller than its value before this update. This simple algorithm allows us to find a family of solutions, compatible with the results from the simulation of Equation 19 where the \overline{W}_{AB} and the T_A are fixed and the \overline{r}_A and b_A are the steady states reached by the system.

A limitation of this approach is that the additional inputs that we are explicitly ignoring (e.g., other layers of V1, synaptic plasticity, other input pathways, etc ...) will be compensated for by our estimates of the recurrent couplings, thus conditioning the meaning of the effective synaptic coupling strengths.

Next, we generalize the procedure when constraining on all stimulus sizes in the classical stimulus condition. The loss function in this case reads:

$$S_A = \int ds \int d\mathbf{x} \left[\sqrt{r_A(\mathbf{x}, s)} - \sum_{B = \{E, P, S, V, L, H\}} \int W_{AB}(\mathbf{x} - \mathbf{y}) r_B(\mathbf{y}, s) d\mathbf{y} - T_A \right]^2 \quad (\text{Equation 28})$$

where s is stimulus size and $r_A(\mathbf{x}, s) = \bar{r}_A(s)e^{-\frac{x^2}{2v_A(s)}} + b_A$. We can follow all the same steps as above, until the very last one, where we cannot simplify due to the $\int ds$. We obtain:

$$\sqrt{b_A} - T_A - \sum_C \bar{W}_{AC} b_C = 0$$

$$2 \int ds \frac{v_A(s)v_B(s)}{2v_A(s)+v_{AB^+}(s)} q_A(s)\bar{r}_B(s) - \sum_C \bar{W}_{AC} \int ds \frac{v_B(s)v_C(s)}{v_{AB^+}(s)+v_{AC^+}(s)} \bar{r}_C(s)\bar{r}_B(s) = 0$$
(Equation 29)

As before, we solve this numerically for \bar{W}_{AC} with sign constraints.

Note that constraining on inverse size tuning curves would not introduce any conceptual complication (just more terms of the same type to calculate) but here we decide to not infer the model parameters to fit direct and inverse response, but instead we fit only the direct response and then benchmark against the ability to generate inverse size-dependent responses.

SOM cells aligned with the stimulus center respond strongly to large classical stimuli, suggesting that they receive a large excitatory current for such stimuli. Nevertheless, both L2/3 Pyramidal neurons and HVAs neurons are very surround suppressed, thus providing a much smaller excitatory current for large stimuli than for small stimuli. This suggests that SOM cells receive extra excitatory input, that increases with stimulus size. Since, to the best of our knowledge, empirical observations on the origin of such external current are still missing, in order to recover the size tuning curve and spatial profiles of SOM neurons we make a minimal assumption that they receive an extra excitatory current whose amplitude grows linearly with stimulus size. We then infer the effective strength of this interaction by adding one external field X in the NNLS equations:

$$S_A = \int ds \int d\mathbf{x} \left[\sqrt{r_A(\mathbf{x}, s)} - \sum_{B \in \{E, P, S, V, L, H, X\}} \int W_{AB}(\mathbf{x} - \mathbf{y}) r_B(\mathbf{y}, s) d\mathbf{y} - T_A \right]^2$$
(Equation 30)

with

$$r_X(\mathbf{y}, s) = \sqrt{s} e^{-\frac{y^2}{2v_X}}$$

$$W_{AX}(\mathbf{x} - \mathbf{y}) = \delta_{AS} \frac{\bar{W}_{AX}}{2\pi v_{AX}} e^{-\frac{(\mathbf{x}-\mathbf{y})^2}{2v_{AX}}}$$
(Equation 31)

The choice of v_{SX} and v_X is arbitrary and is made here following arguments of simplicity and analogy with other excitatory input currents ($v_{SX} = 8^2$, $v_X = 20^2$). We note that a different choice of these values would only weakly affect the geometry of the SOM rate field.

Model selection for the full model

We find 100 parameters sets through the NNLS procedure described above and we quantify the distance between the data and the simulated full system with:

$$\text{absolute error} = \sum_{B \in \{E, P, S, V\}} \sum_{s=5}^{85} (r_B(0, s) - \bar{r}_B(0, s))^2$$

$$\text{normalized error} = \sum_{B \in \{E, P, S, V, L, H\}} \sum_{s=5}^{85} \left(\frac{r_B(0, s)}{\max_s r_B(0, s)} - \frac{\bar{r}_B(0, s)}{\max_s \bar{r}_B(0, s)} \right)^2$$
(Equation 32)

Here, $r_B(0, s)$ is the firing rate of cell type B at position 0 to a stimulus of size s . As expected the two measures are correlated (Figure S32). Moreover, a fraction of the models found through the NNLS procedure diverge, which is not surprising given that the recurrent equations are not solved exactly. In what follows, as well as in the main text we present simulations with the parameter set that minimizes both the metrics above.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data preprocessing

The distance between the center of the stimulus and the center of the CRF of cells is estimated in¹ through CRF mapping experiments, where a stimulus is presented at many positions spanning the area of the visual field accessible for recording and the response of thousands of neurons are recorded through 2-photon calcium imaging.¹ Since the distance between contiguous presentations of the stimulus is 5° , we take this value to be the error on the measure of the CRF location. For each cell we then take a Gaussian in its spatial location with $\sigma_{err} = 5^\circ$ and with amplitude equal to the recorded (trial averaged) response of that cell, i.e.,

$r_i e^{-\frac{(\mathbf{x} - \mathbf{x}_i)^2}{2v_{err}}}$ with $v_{err} = \sigma_{err}^2$. We then sum the Gaussian functions for all recorded cells and divide by the sum of unitary Gaussians, to account for the non-uniform distribution of data-points. The rate field of population A is thus given by:

$$\tilde{r}_A(\mathbf{x}, s) = \frac{\sum_{i=1}^{N_A} r_i(s) e^{-\frac{(\mathbf{x} - \mathbf{x}_i)^2}{2v_{err}}}}{\sum_{i=1}^{N_A} e^{-\frac{(\mathbf{x} - \mathbf{x}_i)^2}{2v_{err}}}} \quad (\text{Equation 33})$$

where N_A is the number of recorded cells of type A and s indicates the stimulus size. In [Figure S1](#) we show an example of the smoothing that this procedure (first presented in⁶) generates. We call this method the *Gaussian kernel method*. Rate fields are plotted in 1 spatial dimension, but they are in fact 2-dimensional and isotropic in retinotopic space. The retinotopic space is considered continuous, representing the limit of very large number of neurons. We test the robustness of this method by comparing it with a different, more rough procedure: we divide the retinotopic space in bins, then we take the median of the response of all the cells that fall within each bin (according to their CRF measure). This alternative procedure allows to consider the error on the estimate of the response (standard error of the mean calculated on the set of cells falling in each bin), but not the error on the estimate of the CRF position. Also we remark that there exists a non negligible dependence of the spatial profiles calculated in this latter way (and consequently their fits) with the number of bins chosen, that can be imputed to the error on the estimate of the CRF position. Note that the baselines of the inverse rate fields are chosen in accordance with those of the classical rate fields, based on continuity between the responses to the largest direct stimulus and the smallest inverse stimulus, for the cells that are furthest away from the stimulus center. Note that the responses to the largest classical stimuli are far from being uniform in space (as they would be if the stimulus was effectively covering the whole visual field of the animal). Overall the corresponding profiles obtained through the two preprocessing procedures are satisfactorily similar for both direct and inverse stimulus condition (see [Figures S2–S4](#)).

Aligned and offset LM neurons are not intrinsically different

We ask whether offset LM neurons and aligned (or centered) ones are intrinsically different classes of neurons (as suggested in¹) or they are the same class of neurons. If they were intrinsically different classes of neurons, their properties (e.g., connectivity profiles) could be different and/or the response to a retinotopically matched stimulus could be different. On the contrary, if they were the same population, their response properties to (precisely) retinotopically matched stimuli would be the same, and a difference in response would be imputable to a difference in retinotopic position relative to the stimulus.

In support of the existence of 2 different populations in LM one could argue that the response to the largest stimulus is different ([Figures 5G and 5I](#) in¹). Considering that the largest stimulus is virtually a full field grating, then a difference in the response indicates that they have a different response to the same stimulus.

In other words, if the stimulus was homogeneous in space (e.g., grating covering the full visual field of the animal), the mismatch of responses of aligned and offset neurons would imply that they have intrinsically different response properties, thus they are 2 different populations, potentially with 2 different connectivity profiles.

Nevertheless, the largest stimulus size is a disk of diameter 85° and the offset boutons with the furthest away receptive field considered are at a distance of 42° from the center of the stimulus. This means that at least some of the offset boutons' CRF is close to the border of the largest stimulus. Therefore the premise that the largest stimulus size is a full field grating is not a satisfactory approximation.

In order to address the question more quantitatively, we apply a clustering algorithm (Uniform Manifold Approximation and Projection, UMAP) to the responses. An embedding in 2D is found by searching for a low dimensional projection of the data that has the closest possible equivalent fuzzy topological structure (all the details on the procedure can be found in <https://umap-learn.readthedocs.io/en/latest/faq.html>). Once the clustering is done we plot the scatterplot of all data points in its 2D embedding (see [Figure S5](#)). If there were n different population (with n different response patterns), the plot should show n well-separated sets of points. After the clustering algorithm is applied to the data, only for interpretation purposes, we color each point according to its label.

When we apply the clustering algorithm to the response data of all boutons, the number of clusters found is $n = 1$, i.e., there is only one cloud of points. When we color the points according to their label (aligned, offset or unknown), we notice that the x axis is correlated with the identity of the boutons (aligned boutons—in blue—tend to be on the right). This is consistent with a situation in which the responses change with continuity, the aligned boutons being in one end of the spectrum. This is consistent with the hypothesis that the neurons differ only in their retinotopic location relative to the stimulus, but not in their intrinsic properties. We then add another color to show the boutons that are excluded in [Figure 5I](#) in¹, i.e., the boutons that do not have significant response to the inverse stimulus. We observe that the neurons excluded from [Figure 5I](#) in¹ are the offsets and unknowns that are closest or overlapping with the aligned ones (colored pink in [Figures S11 and S12](#)). This suggests that the selection rule in [Figure 5I](#) is effectively creating two "artificial" clusters, by removing the offset boutons whose response is inbetween the two ends of the spectrum. In fact, if we repeat the clustering analysis with the same selection criteria as [Figure 5I](#), two (non completely separated) clusters arguably appear, corresponding to aligned (blue) and offset+unknown (green+yellow) boutons ([Figures S11 and 12](#)).

Functional form of the rate fields

In order to be able to understand the system analytically we take the approximation:

$$\tilde{r}_A^{(c)}(\mathbf{x}, s) \approx r_A^{(c)}(\mathbf{x}, s) = p_A^c(s) e^{-\frac{\mathbf{x}^2}{2\sigma_A^c(s)}} \quad (\text{Equation 34})$$

where $A = E, P, S, V, L, M$ indicate respectively populations of Pyramidal, PV, SOM, VIP, Layer 4, LM cells and the superscript (c) stands for classical. The prefactor $p_A^c(s)$ represents the size tuning curve, whereas the dependence on retinotopic space is purely Gaussian. Operatively, we fit the experimental rate fields for each stimulus condition with a Gaussian function. We report the goodness of fits in [Figures S6–S9](#).

For inverse rate fields the response profiles are more complex, therefore we schematize them through a difference of Gaussian functions:

$$\tilde{r}_A^{(i)}(\mathbf{x}, s) \approx r_A^{(i)}(\mathbf{x}, s) = p_A^i(s) e^{-\frac{\mathbf{x}^2}{2\sigma_{pA}^i(s)}} - q_A^i(s) e^{-\frac{\mathbf{x}^2}{2\sigma_{qA}^i(s)}}, \quad (\text{Equation 35})$$

with $p_A^i(s) > q_A^i(s) > 0$.

Note that in the inverse stimulus condition we define scaling as the growth of the width of the outer Gaussian of the fit when stimulus size is increased. We do not distinguish between outer and inner scaling because they appear to be very similar. The difference between outer and inner scaling is mostly very small (less than 1%) and:

$$\frac{\bar{v}_{pA}^i - \bar{v}_{qA}^i}{\bar{v}_{pA}^i} < 0.05\%. \quad (\text{Equation 36})$$

Moreover its distribution is very narrow

In order to have a full functional description of the rate fields as a function of the classical stimulus size we need to define the size tuning curve $p_A^c(s)$ for the populations that we are going to use as inputs for the minimal model, i.e., $A = S, L, M$. This will allow to give an analytical account of surround suppression of the inputs by analyzing the interval where $\frac{\partial p_A^c(s)}{\partial s} < 0$ and preferred size (or CRF size) as $s^* : \frac{\partial p_A^c(s)}{\partial s} \Big|_{s=s^*} = 0$ (see [Figure S16](#)). We use the parametrization:

$$p_A^c(s) = \beta_1^A \operatorname{erf} \frac{s}{S_{\beta_1}^A} - \beta_2^A \operatorname{erf} \frac{s}{S_{\beta_2}^A} \quad (\text{Equation 37})$$

$$\bar{v}_A^c(s) = k_{c1}^A + k_{c2}^A s$$

where the first equation encodes the size tuning curve and β_2^A / β_1^A modulates the amount of surround suppression, $S_{\beta_1}^A$ modulates the steepness of the increase of size tuning for small sizes and $S_{\beta_2}^A$ modulates the steepness of the decrease of size tuning for large sizes. Moreover $k_{c2}^A > 0$ indicates the growth of the rate field width with stimulus size.

The size tuning curves described by the functional form in 37, together with the data and the fitted samples of the rate fields are shown in [Figure S9](#).

For the inverse stimuli we need to specify the functions $p_A^i(s)$ and $q_A^i(s)$. We devise such functional forms and show a comparison with the experimental rate fields.

For SOM population we use the parametrization:

$$p_S^i(s) = \alpha_1^S \operatorname{erf} \frac{s}{S_{\alpha_1}^S} - \alpha_2^S \operatorname{erf} \frac{s}{S_{\alpha_2}^S} \quad (\text{Equation 38})$$

$$q_S^i(s) = \rho_1^S \operatorname{erf} \frac{s}{S_{\rho_1}^S} - \rho_2^S \operatorname{erf} \frac{s}{S_{\rho_2}^S}$$

The evidence presented in¹ suggest that HVAs highly influence the inverse size-dependent response profiles, and on the contrary, the input from L4 is very weak in the inverse stimulus condition, due to classical surround suppression, thus while we focus on reproducing accurately the spatial nuances of LM, we allow less accuracy for L4, in order to keep the model simple and the number of fitted parameters as small as possible (for a more nuanced parametrization of the inputs see [Figure S23](#)). Thus for L4 cells we use:

$$p_L^i(s) = q_L^i(s) = \alpha_1^L \operatorname{erf} \frac{s}{S_{\alpha_1}^L} - \alpha_2^L \operatorname{erf} \frac{s}{S_{\alpha_2}^L} \quad (\text{Equation 39})$$

that constrains the center to have zero response, consistently with the very low response of aligned cells in L4 shown in¹. Finally for LM we use:

$$p_M^i(s) = \alpha_1^M + \alpha_2^M s$$

$$q_M^i(s) = p_M^i(s) - \rho_1^M \left(\operatorname{erf} \frac{s}{S_{\rho_1}^M} - \operatorname{erf} \frac{s}{S_{\rho_2}^M} \right) \quad (\text{Equation 40})$$

Note that the size tuning curve of the aligned cells can then be described as $p_A^i(s) - q_A^i(s)$. The size tuning curve of the offset cells is here defined as: $p_A^i(s)e^{-\frac{x^2}{2\sigma_{pA}^2(s)}} - q_A^i(s)e^{-\frac{x^2}{2\sigma_{qA}^2(s)}}$ with $\tilde{\mathbf{x}} = (30^\circ, 30^\circ)$. Aligned and offset size tuning curves pre and post parametrization are shown in Figure S9. All the parameters $\alpha, \beta, S, k, \rho$ are obtained by fitting the recorded rate fields.

Finally the growth of the inner and outer scaling is given by:

$$\begin{aligned} \bar{v}_{pA}^i(s) &= k_{j1}^A + k_{j2}^A s \\ \bar{v}_{qA}^i(s) &= k_{j3}^A + k_{j2}^A s \end{aligned} \quad (\text{Equation 41})$$

The recorded rate field of LM shows a much lower amplitude in the inverse stimulus condition versus the classical one (as evaluated for example computing the integral over space of the rate field). If the input from HVAs consisted solely of LM contribution, we would not be able to recover an inverse response in L2/3 with firing rates as high as or even higher than the classical case, reported in.¹ But we know that V1 receives input from all HVAs. Moreover Extended Data Figure 9 in¹ shows that optogenetic silencing of an individual HVA generates a quantitatively different reduction of classical versus inverse responses for each HVA considered. Thus, to quantitatively recover the relative amplitude of the rate field in L2/3, we take the minimal assumption that HVAs considered all together have the same rate field profiles of LM, but with an amplitude that is inferred according to the amplitude of the inverse response of L2/3 cells. In formulas:

$$r_H^{(i)}(\mathbf{x}, s) = \eta r_M^{(i)}(\mathbf{x}, s) \quad (\text{Equation 42})$$

and from here on we will consider $A = \{E, P, S, V, L, H\}$, where H is a population representing the conjunction of all HVAs (see Figure S10).

From LM to HVAs

The firing rates recorded from LM boutons by Keller et al.,¹ were on average (across space) several times higher in the classical stimulus condition compared to the inverse one. At the same time the firing rates recorded in L4 Pyr cells were also very low for the inverse stimuli, while the firing rates for L2/3 Pyr+PV population were as large as the ones recorded during classical stimuli presentations. Finally VIP firing rates are comparable for the classical vs. inverse stimuli and SOM is somewhat smaller for inverse stimuli, yet its response decreases with inverse size (the opposite of what would be needed to generate inverse surround suppression). If the net incoming current to an L2/3 unit of the Pyr+PV population is taken to be the sum of the lateral, feedforward (from L4) and feedback (from LM) currents, this would be much more excitatory in the classical case compared to the inverse, resulting in a much larger L2/3 firing rate for classical stimuli, in contrast with the experimental observation. Nevertheless L2/3 cells receive feedback input not only from LM, but from all HVAs.^{1,76} Moreover Keller et al.¹ show that optogenetically silencing an individual HVA, suppresses classical and inverse L2/3 response by different amounts (Extended Data Figure 9 *ibidem*). In particular areas M, AL, LM, LI and P all seem to contribute more to inverse than to classical response. This suggests that the total feedback input to L2/3 (proceeding from all HVAs) might be of the same average magnitude for the classical and inverse stimuli. This allows us to recover the observed relative amplitude of L2/3 responses. More precisely, we describe the total (HVAs) feedback current with a rate field that has the same spatial profile as the one recorded from LM, but we introduce a parameter η that rescales the input from LM in the inverse condition. We fit η to recover the amplitude of the inverse responses (see S4).

The joint population of Pyr and PV

In Figures S11 and S12 we show the rate field of the joint population of Pyramidal and PV recorded cells for comparison with the results of the minimal model in Figures 2 and 4. The joint population is built by taking the average rate field of both populations and then weighting each of the two by the appropriate density of cells as reported in.⁹⁵

This joint population is effectively either excitatory or inhibitory, depending on whether Pyramidal or PV cells dominate. Since the cortex has been found to be inhibition stabilized,^{49–51,53,54,104} we choose it to be effectively inhibitory, thereby preventing the system to diverge when external excitatory inputs are injected. The recurrent synaptic weight that we use is then negative, although the results presented in this work still hold with a small positive weight (see Figure S14).

The comparison between the recordings shown here and the results of the minimal model (Figures 2 and 4) should be taken as an estimate of the implications of the simplifying assumptions that the minimal model is adopting. Additional inputs, cell-specific input-output functions, non-Gaussian profiles, systematic errors in the estimate of the firing rates from calcium imaging, synaptic plasticity, etc. are some of the factors plausibly responsible for the mismatch: the goal of the model is to obtain a versatile description of surround suppression, rather than reproducing in detail the experimental size tuning curve (see discussion). In particular we notice that in the data, around the region of positive (larger than baseline) activity there exists a region where cells respond below baseline. This reminds us of the spatial profile of response to optogenetic perturbations (see e.g.,¹⁰⁵). Moreover similar profiles with resonant spatial frequencies have been found for an input roughly equal across the activated region ('pillbox-shaped' input) in an SSN model.⁴¹

Anatomical length scales

We define a two-dimensional Gaussian function $G_{AB}(\mathbf{x}_1, \mathbf{x}_2, v_{AB}) = \frac{1}{2\pi v_{AB}} \exp\left[-\frac{(\mathbf{x}_1 - \mathbf{x}_2)^2}{2v_{AB}}\right]$ to describe the connection strength to a cell of type A at location $\mathbf{x}_1 = (x_1, y_1)$ from a cell of type B at location $\mathbf{x}_2 = (x_2, y_2)$.^{6,36,37,93} We then estimate the length scales of the connectivity $v_{AB} = \sigma_{AB}^2$ based on recent experimental work. We assume isotropy in the azimuth and elevation directions (x and y).

In a recent paper, Rossi et al.³⁶ could trace the excitatory (L4 and L2/3) and inhibitory (L2/3) presynaptic inputs to an L2/3 pyramidal neuron as a function of the horizontal distance. The average resulting curve shows an excellent agreement with a Gaussian function and the measured widths are $\sigma_{EE} \approx 8^\circ$ and $\sigma_{EI} \approx 5^\circ$, where I indicates a generic inhibitory cell. Here and in what follows the magnification factor is taken to be $0.05^\circ/\mu m$ ³⁶.

Distance dependent connections from Pyr to SOM cells in a slice of mouse L2/3 visual cortex were investigated in.⁶¹ Adesnik et al. recorded the spiking activity of an L2/3 SOM in response to blue light spots of increasing diameters to activate progressively wider areas of L2/3. By cutting the slice to transect L2/3 horizontal axons, they show that SOM cells firing is significantly reduced when inputs from distances greater than 8° were cut off. Although this does not exclude the possibility that in the control condition (uncut), nearby ($d < 8^\circ$) Pyr cells are more activated by offset Pyr cells, thus providing more excitation to SOM, the effect is still compelling and we draw from their measure a conservative estimate of $\sigma_{ES} \approx 8^\circ$.

Marques et al.³⁷ measure retinotopic specificity in inputs from the lateromedial (LM) visual area in mouse V1 and quantify the retinotopic span of feedback projections from LM to V1. They report that LM inputs target, on average, retinotopically matched locations in V1, but many of them relay distal visual information. They estimate that about half of the visual coverage relayed by LM varicosities was more than 24° away from their retinotopic position in V1. Assuming Gaussian connectivity, this yields a $\sigma_{EM} \approx 15^\circ$. We assume that this value holds for HVAs in general, i.e., $\sigma_{EH} = \sigma_{EM}$.

In another recent paper, Billeh et al.³⁵ also report the distance-dependent connection probability profiles for different classes of connections. The cortico-cortical connection probabilities for different cell-class pairs were estimated based on a survey of the existing literature, considering valid sources of information (in order): mouse visual cortex, mouse non-visual cortex, rat visual cortex, rat auditory cortex and rat somatosensory cortex. The measures were assumed to be integrated values of (distance-dependent) connections up to a certain average distance ($75\mu m$) between pre- and post-synaptic cells somas and assuming Gaussian probability distribution. This allowed the authors of³⁵ to estimate the widths of the Gaussian connectivity profiles, yielding:

$\sigma_{EE} \approx 6^\circ, \sigma_{PE} = 5^\circ, \sigma_{SE} \approx \sigma_{VE} \approx 5^\circ, \sigma_{EP} \approx 5^\circ, \sigma_{ES} \approx 4^\circ$. Moreover all inhibitory to inhibitory widths were estimated to be the same as PV to PV, which was measured $\sigma_{PP} \approx 6^\circ$.

These values are also largely consistent with more recent measures of simultaneous whole-cell patch-clamp in mouse V1.⁵⁹ The authors fit a Gaussian to the connection probability as a function of lateral intersomatic distance and report: $\sigma_{EE} = 6^\circ, \sigma_{IE} = 5^\circ, \sigma_{EI} = 5^\circ, \sigma_{II} = 6^\circ$, where I is unspecified inhibition.

Figure S30 reports the values of the cortico-cortical widths of the connection probabilities/effective strengths for different cell-class pairs that we draw from this review of the literature: $\sigma_{EE} = \sigma_{EL} = 7^\circ, \sigma_{SE} = 8^\circ, \sigma_{PE} = \sigma_{VE} = \sigma_{P4} = 5^\circ, \sigma_{EP} = 5^\circ, \sigma_{ES} = 4^\circ, \sigma_{II} = 6^\circ$ where I indicates all inhibitory cell types $I = \{P, S, V\}$. Finally, the projection width in orientation space is taken from a fit of the measures in³⁶ and its value is $\lambda = \frac{\pi}{6}$ (same for all cell types).

Input currents

For the classical stimulus condition, the input to retinotopic position \mathbf{x} of the recurrent layer is given by a convolution with the inter-layer connectivity profile $W_{AB}(\mathbf{x} - \mathbf{y}) = \overline{W}_{AB}G(\mathbf{x} - \mathbf{y}, v_{AB})$, where $v_{AB} = \sigma_{AB}^2$:

$$I(\mathbf{x}, s) = \int_{-\infty}^{\infty} \overline{W}_{AB}G(\mathbf{x} - \mathbf{y}, v_{AB})r_B(\mathbf{y}, s)d\mathbf{y} = \overline{W}_{AB}C_0p_B^c(s)2\pi\overline{v}_B^c(s)G(\mathbf{x}, v_{AB} + \overline{v}_B^c(s)) \quad (\text{Equation 43})$$

Thus we can read:

$$\begin{aligned} I_0^c(s) &= \overline{W}_{AB}C_0\overline{r}_B(s)2\pi\overline{v}_B^c(s) \\ v &= v_{AB} + \overline{v}_B^c(s). \end{aligned} \quad (\text{Equation 44})$$

Finally, putting together the functional forms that we derived in the paragraph above, with Equation 44, we can readout the full dependence of the total input current field of the recurrent population on the stimulus size.

For the inverse stimulus condition, the input to the recurrent layer can be computed through a convolution with the inter-layer connectivity

$$\begin{aligned} I^i(\mathbf{x}, s) &= \int_{-\infty}^{\infty} \overline{W}_{AB}G(\mathbf{x} - \mathbf{y}, v_{AB})r_B(\mathbf{y}, s)d\mathbf{y} \\ &= \overline{W}_{AB}C_0 \left[p_B^i(s)2\pi\overline{v}_{pB}^i(s)G(v_{AB} + \overline{v}_{pB}^i(s)) - q_B^i(s)2\pi\overline{v}_{qB}^i(s)G(v_{AB} + \overline{v}_{pB}^i(s)) \right] \end{aligned} \quad (\text{Equation 45})$$

thus more compactly

$$\begin{aligned}
 I^j(\mathbf{x}, s) &= I_{01}^j(s)G(v_{B1}) + I_{02}^j(s)G(v_{B2}) \\
 I_{01}^j(s) &= \overline{W}_{AB}C_0\rho_B^j(s)2\pi\overline{v}_{\rho B}^j(s) \\
 I_{02}^j(s) &= -\overline{W}_{AB}C_0q_B^j(s)2\pi\overline{v}_{qB}^j(s) \\
 v_{B1} &= v_{AB} + \overline{v}_{\rho B}^j(s) \\
 v_{B2} &= v_{AB} + \overline{v}_{qB}^j(s)
 \end{aligned}
 \tag{Equation 46}$$

Analytical insights on classical size tuning curves with non-normalized input currents

Equation 44 shows that when the inputs are given by the convolution of the firing rates of the input layers with the inter-layer connectivity, the amplitude of the input is proportional to the width of the generating rate field $I^c(\mathbf{x}, s) = I_0^c(s)G(\mathbf{x}, v(s)) \propto 2\pi v_r(s)r(s)G(\mathbf{x}, v(s))$ and $v_r(s)$ and $v(s)$ are related. This is a difference with respect to reference models of classical surround suppression in the literature (e.g., ^{41,42}), where the input currents are not derived from rate fields.

To establish a comparison with existing models, and to simplify the analytical calculations, here we first study classical surround suppression in the case in which $I(\mathbf{x}, s) = I_0(s)G(\mathbf{x}, v(s))$ and $\frac{dI_0(s)}{dv(s)} = 0$ and then show that the simpler case studied analytically provides insights that hold in the more involved case of our main study, where the input currents are derived from rate fields.

Note that when performing calculations for the minimal model, the dependence on stimulus size is not explicitly written. Instead, it is encapsulated within the parameter I_0 . Consequently, this distinction does not alter the functional form of the solution; formally it only involves a redefinition of the parameter I_0 . Nonetheless, from a quantitative perspective, when we consider $I_0^c(s) \propto 2\pi v_r(s)$ and given that $v_r(s)$ grows –however weakly– with stimulus size, the input current to the center cells becomes larger for large stimuli because of the contributions of the offset cells of the input layers.

Dissecting classical surround suppression

Here we will show that classical surround suppression is partially inherited from surround suppression of the input layer(s) and partially due to the width of the input rate field growing with stimulus size. Surround suppression corresponds to a size tuning curve that decreases with increasing stimulus size, i.e., $\frac{r(s)}{ds} \equiv r'(s) < 0$. The dependence on the stimulus size enters in the rate field of the recurrent layer through the amplitude of the input $I_0(s)$ and its width, described here by the input variance $v(s)$. In the simplified case considered here ($\frac{dI_0(s)}{dv(s)} = 0$) we can consider one of these contributions at a time and we will show that:

$$\begin{aligned}
 &\text{if } v'(s) = 0, && \text{then } r'(s) < 0 \Leftrightarrow I_0'(s) < 0 \\
 &\text{if } I_0'(s) = 0 \text{ and } \{W_0 < 0, I_0 > 0\}, && \text{then } r'(s) < 0 \Leftrightarrow v'(s) > 0
 \end{aligned}
 \tag{Equation 47}$$

Since the transfer function is non-negative and monotonic, the size tuning properties of the firing rate are the same as the size tuning properties of the associated total current field:

$$u(\mathbf{x} = 0, s) = \frac{\pi(v(s)+v_{ru}(s))}{W_0} \left(1 - \sqrt{1 - \frac{2I_0(s)W_0}{\pi(v(s)+v_{ru}(s))}} \right) \frac{1}{2\pi v(s)}
 \tag{Equation 48}$$

If $v'(s) = 0$, we find

$$u'(\mathbf{x} = 0, s) = \frac{I_0'(s)'}{2\pi v \sqrt{1 - \frac{2W_0 I_0(s)}{v v_{ru}}}}
 \tag{Equation 49}$$

where one easily reads that $u'(s) < 0 \Leftrightarrow I_0'(s) < 0$ which means that the recurrent layer is surround suppressed when the input layer is surround suppressed.

Conversely, if $I_0'(s) = 0$ (i.e., $I_0(s) = I_0$) we have:

$$\begin{aligned}
 u'(\mathbf{x} = 0, s) &= t_1(s) + t_2(s) + t_3(s); \\
 t_1(s) &= v_{ru}(s) \frac{\pi}{W_0} \left(-1 + \sqrt{1 - \frac{2I_0 W_0}{v(s)+v_{ru}(s)}} \right) v'(s) \\
 t_2(s) &= -I_0 \frac{(v'(s)+v'_{ru}(s))v(s)}{(v(s)+v_{ru}(s))\sqrt{1 - \frac{2I_0 W_0}{v+v_{ru}}}} \\
 t_3(s) &= v(s) \frac{\pi}{W_0} \left(-1 + \sqrt{1 - \frac{2I_0 W_0}{v(s)+v_{ru}(s)}} \right) v'_{ru}(s)
 \end{aligned}
 \tag{Equation 50}$$

Moreover taking into account that $2v_u^{-1} = v^{-1} + v_{ru}^{-1}$ (Equation 8) we find:

$$v'_{ru}(s) = \frac{v_r v'(s)}{\sqrt{v_r^2 + 4v_r v(s)}} > 0 \quad (\text{Equation 51})$$

therefore it is easy to see that for $v'(s) > 0$ and $W_0 < 0, l_0 > 0$:

$$t_1(s) < 0, t_2(s) < 0, t_3(s) > 0 \quad (\text{Equation 52})$$

Finally we can show that $|t_3| < |t_1|$. In fact we just have to show that $v'_{ru} < v_{ru} v'$. Using Equation 51 we easily get to the inequality:

$$2v v_r < 4v v_r + v_r \sqrt{v_r^2 + 4v v_r} + v_r^2 \quad (\text{Equation 53})$$

which is always true, given that the r.h.s has 3 positive terms, the first of which is already larger than the l.h.s. (as shown graphically in Figure S33). Moreover it is easy to see that $\lim_{\sigma(s) \rightarrow \infty} u(s) = 0$.

Thus we have shown that in this system, even if the input is not surround suppressed at all, the recurrent layer can develop surround suppression. Previous models for classical surround suppression described stimuli of increasing size with an input current of increasing width. Here we showed in our setup that our solution is consistent with this result. Nonetheless, our data analyses for feed-forward and feedback inputs shows that the input currents width vary only marginally, while their amplitude vary much more significantly, thus we argue that surround suppression in L2/3 is mostly inherited from its input layers.

Classical surround suppression with non-surround suppressed input rate fields

Here we study numerically the case of one non-surround suppressed excitatory population in the feedforward layer ($\frac{dl_0(s)}{dV(s)} \neq 0$). We want to briefly study the parameter regime for which surround suppression can be imputed to the mechanism of recruiting more lateral inhibition alone. We find that a faster increase in the input width with stimulus size supports classical surround suppression. However, this only holds when the width of the connections from the input layer is small (as in, ⁴¹ modeling cat V1), whereas when it is of the same order of the recurrent projections (as in mouse V1³⁶) this mechanism is not sufficient to generate significant surround suppression (see Figure S39).

Preferred size depends on the preferred size of the inputs

From Equation 49 we deduce that when $v'(s) = 0$, the preferred size of the recurrent layer is the same as the preferred size of the input layer. Therefore, deviations from this behavior might be imputed to variations in the width of the input field or to convergence of inputs with different preferred size. Moreover from the solution of the minimal model with two inputs, Equation 13, we can calculate that when the recurrent layer is receiving two inputs with $v'_1(s) = v'_2(s) = 0$, its preferred size s^* will be in-between the preferred sizes of the two inputs. Indeed one has:

$$v_2 l'_1(s^*) = -v_1 l'_2(s^*) \quad (\text{Equation 54})$$

where we can read that $\text{sign}[l'_1(s^*)] = -\text{sign}[l'_2(s^*)]$, which means that one of the two inputs is increasing while the other one is decreasing (shown in Figures S40 and S41).

Preferred size decreases with contrast

If instead we keep the dependence on stimulus size of one of the two input widths, for instance $v'_2(s) \neq 0$, we find the condition:

$$v_2(s^*) l'_1(s^*) - v_1 l_2(s^*) v'_2(s^*) = -v_1 v_2(s^*) l'_2(s^*), \quad (\text{Equation 55})$$

where one can read that the correction introduced by the dependence $v_2(s)$ reduces the preferred size of the recurrent layer. We hypothesize that this could contribute to contrast dependent classical surround suppression.

In Figure S20 we show that when the width of the input projections is small with respect to the width of the recurrent connections and the width of the input field grows as fast as the stimulus radius (as in ⁴¹), then the preferred size decreases with contrasts. This is in agreement with ⁴¹. However the model that we are considering here has less ingredients than the one in ⁴¹ since i) we have only one recurrent population and ii) in ⁴¹ the input for stimulus size s was a Gaussian function of space with standard deviation $\sigma_{input}(s)$, whereas in our case because of the convolution with the input layer connectivity it has standard deviation $\sqrt{\sigma_{FF}^2 + \sigma_{input}(s)^2}$.

Contrast dependent classical size tuning curves

Here we ask how the minimal model with anatomically-realistic length scales and experimental input rate fields from L4 and HVAs, can recover contrast-dependent surround suppression. The dataset at hand does not contain data on stimuli with different contrast. One simple way of simulating contrast-varying stimuli is to scale the inputs by a factor that is small for small contrasts. This situation is exposed in Figure S21. This model easily reproduces the contrast dependence of the amplitude of response and SMI increases with

increasing contrast for a nonlinear model, but not for a linear one (as expected,^{41,52}). However, with this minimal assumption for contrast-dependence of the inputs, restricting the parameter space to biologically plausible values of the length scales, we cannot reproduce the observation that the preferred size of Pyr cells in L2/3 grows when the contrast decreases.^{5,43} Based on the arguments exposed above we could argue that contrast-dependence of the preferred size of L2/3 cells could be due to: i) an increase of the preferred size of L4 rate field for low contrasts (observed in data⁴³), ii) an increase of the preferred size of HVAs rate field for low contrasts (yet not explored experimentally), iii) a decrease in the scaling of (L4 or) HVAs rate field width with stimulus size (yet not explored experimentally), iv) the fact that for low contrasts VIP are more active, thus SOM are less active (shown in experiments and models in⁴³). We apply one at a time these modifiers to the inputs as a function of contrast and we show that in all cases we reproduce contrast dependent size tuning (see Figure S22). This suggests that all these factors may contribute to contrast dependent size tuning.

The origin of inverse response and inverse size tuning

First of all, let us note that the agreement between analytic results and simulations for the inverse stimulus condition is less accurate than in the classical case because of the imprecision introduced by the parametrization of the inputs (see Figures 4B, 4C, and 4E) and also because the ansatz itself, i.e., a Gaussian modulated by a slowly varying function. Nevertheless, in all the analyzed simulations, the agreement is fairly good, therefore for efficiency reason, we study the analytical solution. To tackle the origin of inverse response we change the value of the projection span from HVAs to L2/3 $\sigma_{L23-HVA}$ and we keep the rate field of HVAs fixed, for a fixed stimulus size. In the main text we explained that when this projection span is too small then inverse response is not allowed, meaning that the response of offset cells is larger than the response of aligned cells. This can be easily understood by looking at the input current from HVAs, which is a convolution between the connectivity and the rate field of HVAs. If the width of the connectivity is large enough, then the convolution transforms the ring profile of the rate field of HVAs into a function with a peak in the center thus generating inverse response.

To tackle the origin of inverse size tuning in an analytical way we counterfactually change the profile width of HVAs, so that the width of the input ring (rate field of HVAs) does not grow with stimulus size. In particular here we use the parameterization of HVAs rate field with a difference of Gaussian and fix the growth of the width of both Gaussians to 0. In this case inverse surround suppression disappears. This is consistent with the results shown in Figures 6F–6M, but in this case the modification leads to a complete fail of inverse surround suppression. This confirms that the origin of inverse surround suppression is the fact that for large stimuli the ring becomes too large to generate a convex function in the center after the convolution with the connectivity. Instead if we fix the width of the ring profile, the convexity of the input current does not change.

Analytical results on inverse surround suppression

Here we show that it is not possible to obtain inverse surround suppression if HVAs ring size does not increase with stimulus size. More precisely we show that the feedback current incoming to the cells aligned with the center of the stimulus grows indefinitely with stimulus size if HVAs ring size is fixed ($v_{pH}(s) = 0, v_{qH}(s) = 0$). With the parametrization chosen (see Equations 40 and 42), and given Equation 43 we have:

$$\begin{aligned}
 I_H(\mathbf{x} = 0, s) &= (\alpha_1^H + \alpha_2^H s) \frac{\bar{v}_{pH}^j}{\bar{v}_{pH}^j + v_{EH}} - \left[(\alpha_1^H + \alpha_2^H s) - \rho_1^H \left(\operatorname{erf} \frac{s}{S_{\rho_1}^H} - \operatorname{erf} \frac{s}{S_{\rho_2}^H} \right) \right] \frac{\bar{v}_{qH}^j}{\bar{v}_{qH}^j + v_{EH}}. \\
 \frac{dI_H(\mathbf{x} = 0, s)}{ds} &= I'_H(\mathbf{x} = 0, s) = \alpha_2^H \frac{\bar{v}_{pH}^j}{\bar{v}_{pH}^j + v_{EH}} - \left[\alpha_2^H - \frac{2\rho_1^H}{\sqrt{\pi}} \left(e^{-\frac{s^2}{(S_{\rho_1}^H)^2}} - e^{-\frac{s^2}{(S_{\rho_2}^H)^2}} \right) \right] \frac{\bar{v}_{qH}^j}{\bar{v}_{qH}^j + v_{EH}}. \quad (\text{Equation 56}) \\
 \lim_{s \rightarrow \infty} I'_H(\mathbf{x} = 0, s) &= \alpha_2^H \left(\frac{\bar{v}_{pH}^j}{\bar{v}_{pH}^j + v_{EH}} - \frac{\bar{v}_{pH}^j - 1}{\bar{v}_{pH}^j - 1 + v_{EH}} \right) = \alpha_2^H v_{EH} \geq 0
 \end{aligned}$$

Thus, under the parameterization chosen, the feedback current coming into the center in L2/3 (i.e., the location that codes for the stimulus center) keeps growing for very large sizes, i.e., there is no inverse surround suppression.

HVAs cells close to the edge of the inverse stimulus respond maximally

Comparing the scaling of the width of the outer Gaussian with stimulus size is useful at an analytical level. However, for the inverse stimulus, another spatial scale can be more relevant physiologically: the retinotopic location (relative to the center) where the rate field is maximal. If this quantity scaled with the stimulus radius, we could conclude that HVAs are mostly active close to the edge of the inverse stimulus. Figure S26 shows that the scaling of this quantity is larger than the scaling of σ_{HVA} . Moreover if we consider all the sizes the data indicates that this quantity scales more weakly than the stimulus radius, but if we consider only intermediate sizes the data are consistent with the hypothesis that the HVAs cells that respond maximally are the ones whose CRF is close to the edge of the ‘hole’. It is tempting to speculate that this supports the involvement of HVAs in contour detection.⁵ When the HVAs rate field is parameterized as a circle around the origin (see in Figure 6E), the y axis of Figure S26 represents the radius of that circle, $\sigma_{HVA,r}$.

Lateral input from SOM cells increases surround facilitation and the connection with SOM inverse response

In Figure S40 we show the rate fields of SOM cells for a classical stimulus, an inverse stimulus at an orthogonal orientation and a cross stimulus, under the assumption that two rate fields of simultaneously presented stimuli at orthogonal orientations are additive (given that the populations of cells code for the two stimuli are different, see main text). We keep the stimulus size fixed to $s = 15^\circ$. The Figure shows that SOM input current is larger in absolute value in the cross than in the classical condition. This could suggest that SOM cells suppress Pyr and PV more in the cross condition than in the classical one, therefore reducing surround facilitation. However this is not true, as shown in Figure 7. On the contrary our model predicts that SOM cells enhance surround facilitation (see Figure S20).

This is due to the definition of surround facilitation index $SFI = \frac{r_x - r_c}{r_x + r_c}$ (based on previous choices in the literature -see e.g., *CMI* in²), as opposed to $\frac{r_x}{r_c}$. Here r_x (resp. r_c) is the firing rate of L2/3 cells centered in the stimulus and tuned to its orientation in the cross (resp. classical) stimulus condition.

To better understand the role of SOM in surround facilitation we dissect the contributions to $\widetilde{SFI} = \frac{u_x - u_c}{u_x + u_c}$, where u_c and u_x are the corresponding currents [u_x]₊² = r_x . \widetilde{SFI} has the same qualitative behavior as SFI as shown in Figure S41 (compare with Figure 7H).

We now define some quantities for the cross ($a = x$) or classical ($a = c$) stimulus condition:

- (1) $u_a > 0$ is the total current incoming to Pyr+PV;
- (2) $u_{Sa} < 0$ is the input from SOM (to Pyr+PV);
- (3) $u_i = u_x - u_c > 0$ is the total current incoming to Pyr+PV by the effect of the surround in the cross stimulus condition;
- (4) $u_{Si} = u_{Sx} - u_{Sc} < 0$ is the input current from SOM in the effect of the surround in the cross stimulus condition;
- (5) \widetilde{SFI}_+ the value of the currents surround facilitation index when SOM cells are active;
- (6) \widetilde{SFI}_- the value of the currents surround facilitation index when SOM cells are silenced.

Then we have:

$$\begin{aligned}
 & \widetilde{SFI}_+ > \widetilde{SFI}_- \\
 & \Leftrightarrow \frac{u_x - u_c}{u_x + u_c} > \frac{u_x - u_{Sx} - u_c + u_{Sc}}{u_x - u_{Sx} + u_c - u_{Sc}} \\
 & \Leftrightarrow \frac{u_i}{u_x + u_c} > \frac{u_i - u_{Si}}{u_x - u_{Sx} + u_c - u_{Sc}} \tag{Equation 57} \\
 & \Leftrightarrow u_i(u_{Sx} + u_{Sc}) < u_{Si}(u_x + u_c) \\
 & \Leftrightarrow u_i(u_{Si} + 2u_{Sc}) < u_{Si}(u_i + 2u_c) \\
 & u_i u_{Sc} < u_{Si} u_c
 \end{aligned}$$

Finally, since $u_{Sc} < 0$ and $u_{Si} < 0$, we have

$$\frac{u_i}{u_c} > \frac{|u_{Si}|}{|u_{Sc}|}, \tag{Equation 58}$$

which means that the inverse response of Pyr+PV population, normalized to its classical response, is larger than the inverse response of SOM population normalized to its classical response. In summary the surround facilitation index decreases when SOM cells are silenced because SOM cells respond less than Pyr+PV cells to the inverse stimulus (when responses are normalized to classical responses). In fact Pyr+PV cells respond more to inverse than to classical stimuli of their respective preferred size, whereas SOM cells respond more to classical than to inverse stimuli (see Figures 3N and 5D in the main text).