The combination of feedforward and feedback processing accounts for contextual effects in visual cortex

Serena Di Santo[1,*], Mario Dipoppa[3], Andreas Keller[4], Morgane Roth[4], Massimo Scanziani[5], and Kenneth D. Miller[6]

[1]Departamento de Electromagnetismo y Física de la Materia and Instituto Carlos I de Física Teórica y Computacional, Universidad de Granada, Granada, Spain.
[3]Department of Neurobiology, David Geffen School of Medicine, University of California, Los Angeles
[4]Institute of Molecular and Clinical Ophthalmology, Mittlere Strasse 91, CH-4031 Basel, Switzerland
[5]Department of Physiology, University of California, San Francisco, San Francisco, CA 94158-0444, USA
[6]Center for Theoretical Neuroscience and Mortimer B Zuckerman Mind Brain Behavior Institute, Columbia University, NewYorkCity, NY10027, USA
[*]corresponding author: serenadisanto@ugr.es

January 23, 2024

Abstract

Sensory systems must combine local features with context to infer meaning. Accordingly, context profoundly influences neural responses. We developed a unified circuit model demonstrating how feedforward and feedback inputs are integrated to produce three forms of contextual effects in mouse primary visual cortex (V1). First, reanalyzing existing data, we discovered that increasing stimulus size only weakly increases the area of V1 neural response, conflicting with previous models of surround suppression (SS). Second, through modeling, we found that, in Layer 2/3, (1) SS and its contrast dependence are largely inherited from Layer 4; (2) Inverse responses (IR) – size-tuned responses to a

gray "hole" in a full-field grating – are driven by feedback connections provided they are sufficiently wide; (3) Cross-orientation surround facilitation is induced by the summation of feedback input driving IR with the feedforward-driven classical center response. The model accounts for many previous findings and makes multiple testable predictions.

Keywords— contextual modulation, visual cortex, feedback, Higher Visual Areas, inhibitory subclasses

Highlights

- One 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.

1 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 (Fig. 1a-c).

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) (*e.g.*, Angelucci et al., 2017; Dipoppa et al., 2018; Keller et al., 2020a,b; Ziemba et al., 2018; Zipser et al., 1996). V1 neurons are primarily driven by stimuli only within a small region of visual space, known as the neuron's classical receptive field (CRF). Yet 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 (Field et al., 1993; Kirchberger et al., 2021, 2023a; Pak et al., 2020; Roelfsema, 2006).

Here we develop a simple theoretical framework to address the question: how do the properties of feedforward, feedback and lateral input currents contribute to contextual modulation in V1? We arrive at a unified mechanistic picture of how

these different inputs are integrated to produce diverse forms of contextual modulation. This gives a coherent, experimentally testable theory of V1 computation that can serve as a more general framework for studying how cortical circuits integrate context into sensory processing.

Three exemplary paradigms have been widely used to study contextual modulation in V1 (Fig. 1d-f).

- 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* (Allman et al., 1985; Angelucci et al., 2017; Dipoppa et al., 2018; Hubel and Wiesel, 1965; Keller et al., 2020a; Sceniak et al., 2001; Shushruth et al., 2009; Spillmann et al., 2015) 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 (Barlow, 2001, 1972; Coen-Cagli et al., 2015).

- In contrast, addition of a stimulus surround that is very distinct from the CRF stimulus can enhance a neuron's response (Cavanaugh et al., 2002; Field et al., 2010; Keller et al., 2020a; Paffen et al., 2005; Polat et al., 1998; Self et al., 2014; Xing and Heeger, 2001). This *surround facilitation* can enhance the perceptual salience of discontinuities in the visual world, which likely separate different objects (Craft et al., 2007; Lamme, 1995; Rubin, 2001). Given an oriented grating as CRF stimulus, this facilitation can be evoked by an orthogonally oriented stimulus in the surround (*cross-orientation surround facilitation*).

- Surrounding context that indicates occluded structure in the center (Fig. 1b) can evoke illusory perception of that structure (Kanizsa, 1976; Pak et al., 2020). While illusory contour responses in V1 are rare (Peterhans and von der Heydt, 1989; Shin et al., 2023; von der Heydt, 1987; but see Grosof et al., 1993), 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 (Jones et al., 2001; Keller et al., 2020b; Kirchberger et al., 2023b; Rossi et al., 2001). 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 (Keller et al., 2020b; Kirchberger et al., 2023a; Rossi et al., 2001), and are substantially reduced if higher visual areas (HVAs) are optogenetically suppressed (Keller et al., 2020b), suggesting that feedback connections play a key role in inverse responses.

Here we develop a unified model of these three contextual effects in layers

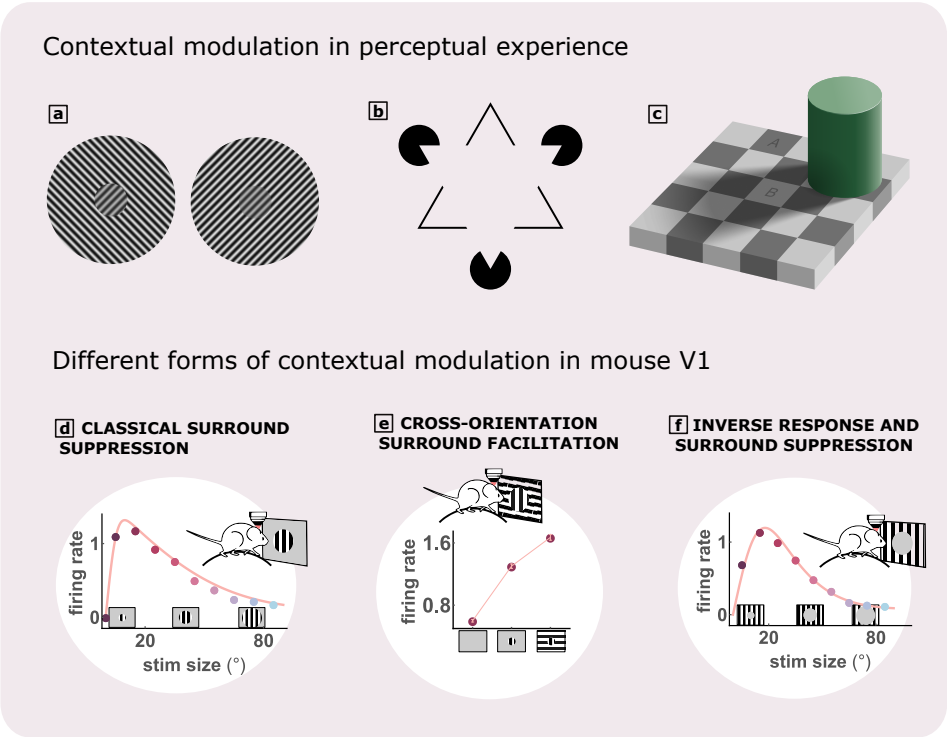

Figure 1: **Different forms of contextual modulation.** **a**. The grating patches in the centers are identical but are perceived as more or less salient due to the distinct surround. **b**. Kanizsa triangle illusion: a nonexistent white triangle in the center appears to occlude the shapes around it **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** Classical size tuning and surround suppression: responses decrease when the stimulus size is increased beyond the CRF size; data (circles) from pyramidal cells studied by Keller et al. (2020b), line is a fit to the data. Here and in what follows, firing rates are in arbitrary units. **e**. Cross-orientation surround facilitation: 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., 2020a). **f**. Inverse response: 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., 2020b) and fit as in (a).

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 (Billeh et al., 2020; Li and Young, 2021; Marques et al., 2018; Rossi et al., 2020) and data from Keller et al. (2020b) on the response profiles across space of each cell type and of inputs from L4 and from an HVA, area LM, to each stimulus type. We focus on L2/3 both because of the availability of this data and because inverse responses, which depend on feedback from HVAs, are found in L2/3 but not in L4 (Keller et al., 2020b), 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 previously to account for various aspects of contextual modulation. Phenomenological models place it within a broad range of nonlinear response properties described as "divisive normalization" (*e.g.*, Carandini and Heeger, 2012; Carandini et al., 2005; Coen-Cagli et al., 2015). Circuit models (*e.g.*, Dipoppa et al., 2018; Keller et al., 2020a; Li and Young, 2021; Mossing et al., 2021; Obeid and Miller, 2021b; Rubin et al., 2015; Schwabe et al., 2006, 2010; Shushruth et al., 2012b) have not previously addressed inverse tuning or the roles of top-down feedback or, in many cases, multiple types of inhibitory cell. In addition, our approach differs in being based on analysis of experimental data on the 2D spatial structure of activity (as also was Dipoppa et al. (2018)) 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, which forces revision of most previous models of contextual modulation.

To isolate basic principles, we focus on an abstract, "minimal" nonlinear model of a single recurrently-connected cell type representing the joint activity of Pyr and PV cells in L2/3, along with external input from SOM cells, that can be fully 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 (Supplementary SS15–SS17). We focus on qualitatively explaining response properties and understanding their underlying mechanisms. This enables multiple insights (summarized in Table 1), including the spatial structure of top-down projections required for inverse responses to arise, how the combination of classical and inverse responses can produce cross-orientation facilitation, and that SOM should make opposite contributions to classical vs. inverse surround suppression.

2 Results

2.1 A spatial schematic description preserving anatomical length-scales

To take into account input from neurons with CRFs distributed across visual space (Dipoppa et al., 2018; Keller et al., 2020b), we require knowledge of two factors: i) 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 ii) how the responses of cells to a stimulus depend on the distance between the cell's CRF center and the stimulus center. For dependence (i), we assume that synaptic strengths decrease with distance between two neurons as a Gaussian function, with length scales taken from experimental measurements (Billeh et al., 2020; Keller et al., 2020b; Li and Young, 2021; Marques et al., 2018; Rossi et al., 2020) (see Supplementary SS1). We take this approach, rather than trying to infer connections that lead to responses that fit the data (*e.g.*, Dipoppa et al., 2018), 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 (Supplementary SS15, SS16). For dependence (ii), we use the dataset of Keller et al. (2020b) and a smoothing method introduced in Dipoppa et al. (2018) 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 (Fig. 2, Supplementary SS2 B). We will also describe these firing-rate fields as Gaussian (or linear combinations of Gaussian) functions centered on the stimulus center.

2.1.1 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 in (Keller et al., 2020b) (Fig. 2; see also Supplementary SS2 B). 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 Section 2.2. 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 (Fig. 2a-d top) (rate fields for all cell types and all stimulus sizes, as computed by two different methods, are shown in Supplementary Fig. SF2). This is quantified by fitting a Gaussian function to each spatial profile (Fig. 2a-d bottom), and comparing (Fig. 2e-h) 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) (shown for all cell types and both methods of calculating spatial rate fields in Supplementary Fig. SF6). 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 (Fig. 2i-l) (shown for all cell types in Supplementary Fig. SF7).

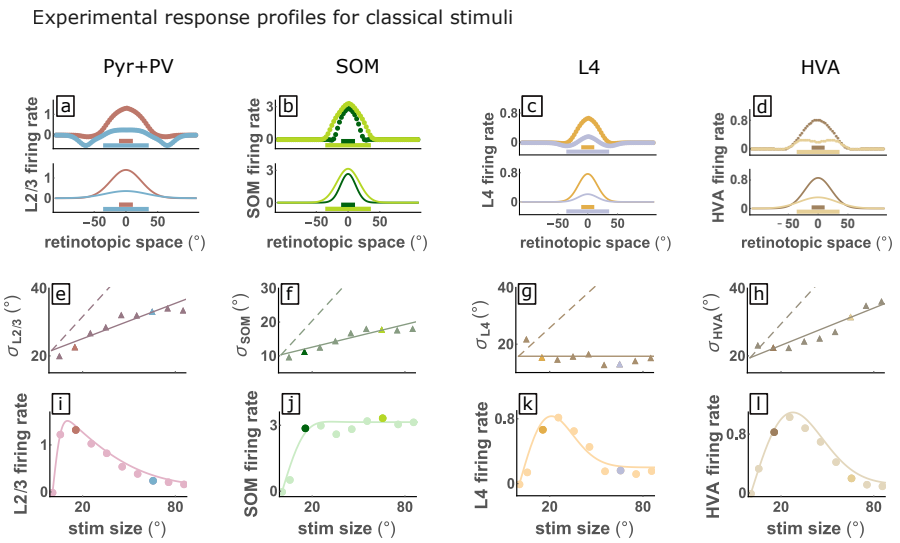

Figure 2: **The widths of spatial response patterns grow much more slowly than stimulus size**. Pyr+Pv, SOM, L4 and HVAs spatial responses, calculated from experimental data (Supplementary SS2 B). **a-d.** experimental rate field (top) and best Gaussian fit (bottom), for stimulus sizes $s = \{15°, 65°\}$; x axis represents distance of RF 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.**)

The assumption that rate field width grows to match the stimulus radius is at the core of existing theoretical models of classical surround suppression (Angelucci et al., 2017; Dipoppa et al., 2018; Li and Young, 2021; Obeid and Miller, 2021a; Rubin et al., 2015), 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, to the best of our knowledge, systematically discussed here for the first time, although it can also be seen in the plots of Dipoppa et al. (2018) (their Supplementary Figure 11).

In what follows, we first develop a minimal model for ordinary surround suppression (Allman et al., 1985; Angelucci et al., 2017; Hubel and Wiesel, 1965; Sceniak et al., 2001; Shushruth et al., 2009; Spillmann et al., 2015). We then apply this model to understand inverse response (Keller et al., 2020b). Finally, we extend the model to understand cross-orientation surround facilitation (Keller et al., 2020a).

2.2 A minimal model for contextual modulation

We develop a minimal model to address the question: how do 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 (Haider et al., 2006; Keller et al., 2020b; Mossing et al., 2021; Palmigiano et al., 2020), thereby dynamically balancing (preventing an excess of) excitation and functioning as a stabilizer for the network (Adesnik, 2017; Ahmadian and Miller, 2021; Bos et al., 2020; Denève and Machens, 2016; Dipoppa et al., 2018; Isaacson and Scanziani, 2011; Kato et al., 2017; Okun and Lampl, 2008; Ozeki et al., 2009; Palmigiano et al., 2020; Renart et al., 2010; Sanzeni et al., 2020). In contrast, VIP and particularly SOM cell responses generally differ substantially from those of Pyr and PV cells (Dipoppa et al., 2018; Keller et al., 2020a,b; Mossing et al., 2021; Palmigiano et al., 2020). 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 (Billeh et al., 2020; Campagnola et al., 2021; Pfeffer et al., 2013). We also take as external inputs the measured responses across space of layer 4 (L4) and Lateromedial area (area LM) 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 (taken to be rectified quadratic, see STAR Methods), describing the steady-state firing rate induced by a given net input to the neuron (Priebe and Ferster, 2008). Such an expansive input/output function is expected when firing activity is driven by input fluctuations rather than by the mean input (Hansel and van Vreeswijk, 2002; Miller and Troyer, 2002), and has been shown (Ahmadian and Miller, 2021; Rubin et al., 2015) to yield network behavior that reproduces a variety of nonlinear visual cortical behaviors, including contrast-dependent surround suppression, that have often been summarized phenomenologically as "divisive normalization" (Carandini and

Heeger, 2012). 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 (Persi et al., 2011), to develop an approximate analytic solution to the one-population model, which yields deeper insight into the mechanisms driving model behavior.

A crucial and novel 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 (Billeh et al., 2020; Keller et al., 2020b; Siegle et al., 2021), but we have recordings of the V1 boutons from only one HVA, area LM (Keller et al., 2020b). Extended Data Fig.9 in Keller et al. (2020b) 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. We infer these constants based on the amplitudes of inverse vs. classical response in L2/3 (see Supplementary SS5). Keller et al. (2020b) suggested that LM cells projecting to a given point in L2/3 from the same retinotopic location or from peripheral locations (aligned vs offset cells) had qualitatively different RF properties, but our analysis finds no such difference (Supplementary SS3), so we treat HVA cells as being homogeneous across space.

2.3 Classical surround suppression in the minimal model

In our minimal model of classical size tuning (Fig. 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 (Fig. 2), with the corresponding Gaussian connectivity (Fig. 3b). The recurrent rate fields resulting from the approximate analytical solution and the simulation of the minimal model agree well (Fig. 3c-d and Supplementary SS6), 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 (Supplementary SS7 B), consistent with effects of optogenetic activation of PV cells (Nienborg et al., 2013b). 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 (Supplementary SS7 C).

2.3.1 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., 2017 and references therein): the minimal 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.

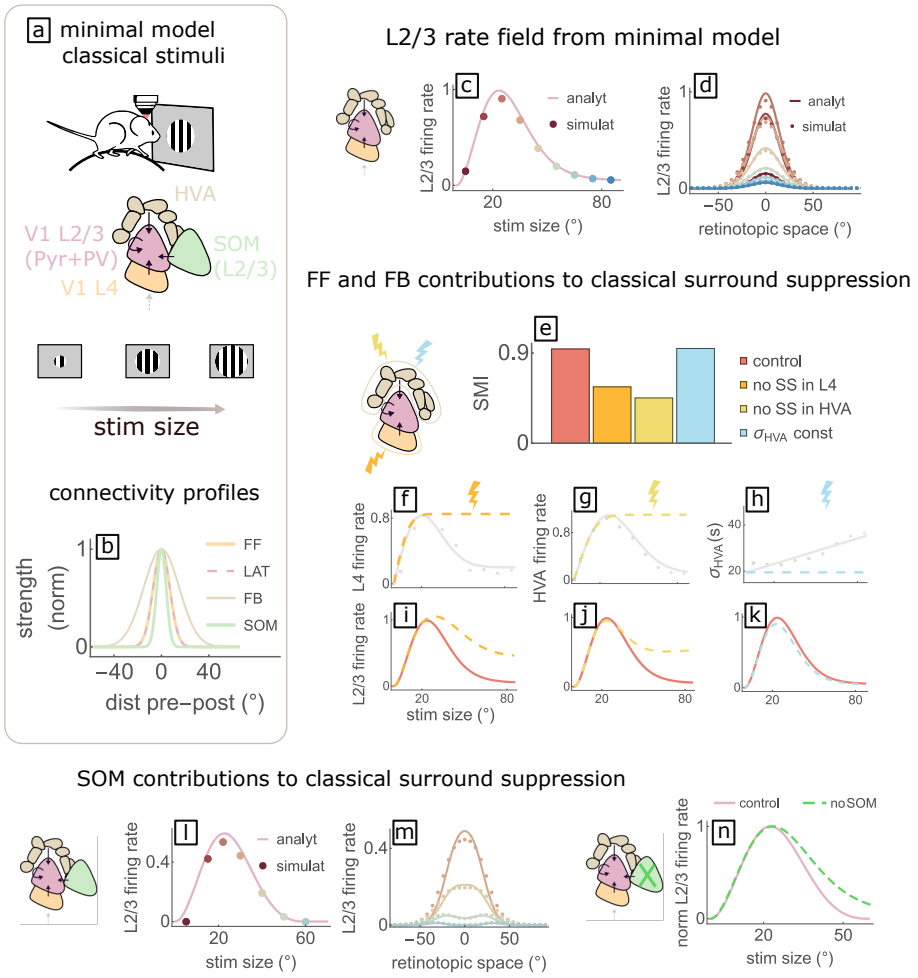

Figure 3: **Feedforward, feedback and lateral SOM inputs shape classical surround suppression. a.** Top to bottom, sketches of: experimental setup with classical stimuli; minimal model with one recurrent cell type in V1 L2/3 receiving feedforward (FF) input from L4, feedback (FB) input from HVAs, lateral input from its own type (LAT) and from SOM cells; classical size tuning stimuli. **b.** Strength of connections vs. CRF distance between pre- and post-synaptic neurons, for the four types of connections.

Figure 3: **c.** Size-tuning curve and **d.** rate field for the recurrent layer of the model. Here and in **l,m,** color code of the full circles indicates stimulus size; full circles represent simulations, full lines represent analytical results. **e-k.** Counterfactual modifications of the inputs. **e.** Bar plot representing the SMI (see main text; = 0, complete suppression; = 1, no suppression) in the control condition, and when we eliminate either: surround suppression (SS) in L4, surround suppression in HVAs, or growth of the width of the HVAs rate field with stimulus size (σ_{HVA} const). **f-h.** Illustration of the counterfactual modifications (dashed, same color code as bar plot) compared to the empirical values (full lines and symbols as in Fig. 2). **i-k.** Effects of the counterfactual modifications on L2/3 size tuning curve (same color code as above). **l-n.** Effects of adding SOM input. **l,m** Same as **c,d** but with SOM input. For large stimuli, the rate field convexity in the center changes. Compare with experimental profiles in Supplementary SS6 A. **n.** Size tuning curve of recurrent layer with (pink; same as **l**) and without (dashed green; same as **c**) SOM input.

In Supplementary SS8, using our analytical solution, we show that surround suppression in the recurrent layer can be produced by either or both of: i) surround suppression in the input population and ii) broadening of the input rate field as stimulus size is increased. In case (i), L2/3 surround suppression is inherited from the input (although it may be enhanced by L2/3 network mechanisms that increase response differences relative to input differences). In case (ii), 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 (Fig. 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 (Figure 3e-g,i-j) and of eliminating the growth of the width of the HVAs rate field with increasing stimulus size (Figure 3h,k; note the width of the L4 rate field does not increase with stimulus size). In each condition, we calculate a *Surround Modulation Index (SMI)*, $SMI = 1 - r_{\text{Large}}/r_{\text{Pref}}$, where r_{Large} and r_{Pref} are the firing rate of the centered cells at the largest stimulus size ($s = 85°$) 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 HVAs rate field width does not alter surround suppression in L2/3.

In Supplementary SS9 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.

Surround suppression is weak for low-contrast stimuli and increases in strength,

while optimal size shrinks, with increasing stimulus contrast (*e.g.*, Mossing et al., 2021). In Supplementary SS8 D we analyze the contrast dependence of size tuning in the model. We find that a supralinear one-population model with Gaussian input whose rate field width increases with stimulus size can generate the observed contrast-dependent size tuning, i.e. higher contrasts produce larger surround modulation and smaller preferred sizes, in line with the results in the two-population E/I model of Rubin et al. (2015). However, the contrast-dependence in our one-population 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. Instead, and in agreement with (Mossing et al., 2021), our model suggests that in mouse L2/3 contrast-dependence is largely inherited from its feedforward layers. In Supplementary SS10, we examine this contrast dependence further, and find that additional contributing factors may be i) scaling of rate field width in the feedback input and ii) large VIP/small SOM response for low contrasts (in agreement with Mossing et al. (2021)).

In summary, our model suggests that both surround suppression and its contrast dependence in L2/3 are largely inherited from its inputs, although the supralinear input/output function also strengthens surround suppression.

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

The rate field of SOM neurons in L2/3 lacks surround suppression (Fig. 2j), in agreement with previous results (Adesnik et al., 2012a) (see Supplementary SS4 A). We compare the size tuning curve of the model recurrent population in the presence or absence of SOM input, as in Adesnik et al. (2012b). Since SOM cells prefer larger stimulus sizes, they inhibit L2/3 the most for larger stimuli, thus enhancing L2/3 classical surround suppression (Fig. 3l,n). 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 has a "ring" shape for large stimulus sizes (Fig. 3m), which is well captured by the analytic solution of the minimal model. This ring shape for large stimulus sizes has been reported in Dipoppa et al. (2018); their data is replotted for the joint population of Pyr and PV in Supplementary SS6 A).

The absence of this feature in the dataset of (Keller et al., 2020b) (Fig. 2i) could conceivably be explained if those animals had a smaller effective strength of SOM projections than in (Dipoppa et al., 2018). From a geometric point of view, the SOM input field is a (relatively narrow) Gaussian-shaped bump (*e.g.* see SOM rate field in Fig. 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 (Fig. 2l, Supplementary SS4 A).

In conclusion, our model predicts that classical surround suppression in L2/3

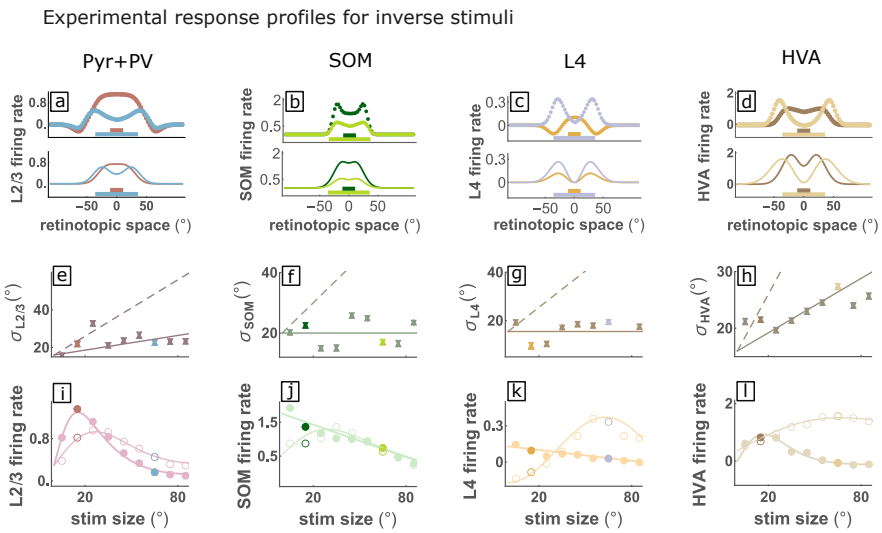

Figure 4: **Inverse response and size tuning properties of V1 and HVAs**. Same as Fig. 2 but for inverse stimuli. **a-d.** Experimental rate field (top) and best difference of Gaussians fit (bottom) for two stimulus sizes. The bottom bars represent the size of the corresponding inverse stimuli. **e-h.** Up and Down full triangles are the widths of the positive and negative Gaussian function in the best difference-of-Gaussian fits to the rate fields. Differences in widths of the two Gaussians are small, *i.e.* up and down triangles are almost overlapping. Full line: linear fit, dashed line: fit if rate field 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.

is largely inherited from surround suppression of its feedforward (L4) and feedback (HVAs) inputs (consistent with the experimental literature, Angelucci et al. 2017; Vangeneugden et al. 2019), marginally enhanced by lateral (PV) inhibition progressively recruited as the stimulus size increases (Obeid and Miller, 2021a; Rubin et al., 2015), enhanced by SOM lateral input (Adesnik et al., 2012a) and partially supported by the supralinear input/output function (Hansel and van Vreeswijk, 2002; Miller and Troyer, 2002).

2.4 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 (described in Section 2.2, further detailed in Supplementary SS5), to understand responses to inverse stimuli of varying sizes (*i.e.*, hole diameters).

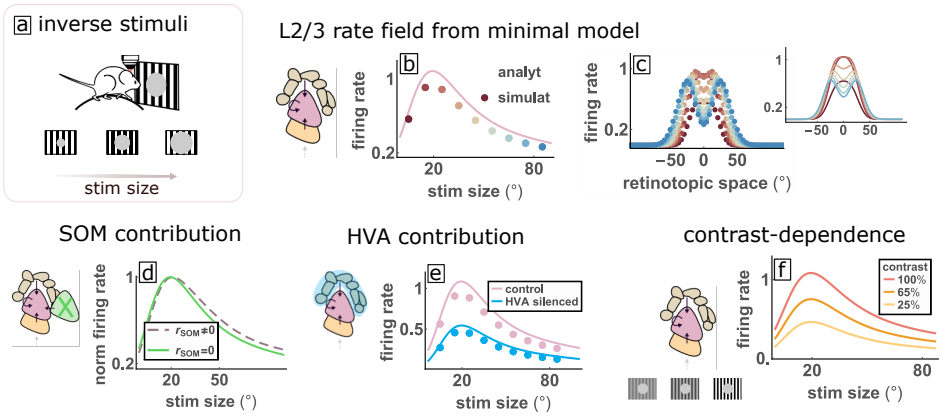

Figure 5: **Inverse response and size tuning properties of L2/3 are shaped by its feedback input**. **a** Inverse size-tuning stimuli, **b** size-tuning curve, and **c** rate fields for L2/3 of the model. In **c,** results for the simulations (dots) and the analytics (curves) are represented separately for better readability. The model recovers inverse size tuning as well as the change in convexity of the rate field (compare Fig. 4a,i). **d.** SOM neurons reduce inverse surround suppression (compare Fig. 3n for classical stimuli). **e.** Size tuning curve in the control case (pink) and for silencing of HVAs (rate field of HVAs reduced by a factor 0.65) (compare Fig. SS13 A for classical stimuli). **f**. Size tuning curves for varying contrast (both L4 and HVAs rate fields are reduced by a factor $c = \log_2 2C_0$, C_0 values shown in legend; compare Fig. SF22 for classical stimuli).

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 (Fig. 4, Supplementary SS4 and SS6). This led the authors of Keller et al. (2020b) 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 (Fig. 4i, full circles, compare with Fig. 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 (Fig. 4a, Supplementary Fig. SF4). This is not true for classical stimuli (Supplementary Fig. SF2), although, as we noted previously, it can occur for large enough stimuli both in the model (Fig. 3m) and in experiments (Dipoppa et al., 2018). Similarly to the classical stimulus condition, the widths of the inverse rate fields scale weakly with the stimulus size (Fig. 4e-h, Supplementary SS4 A). To the best of our knowledge, these latter two features of the inverse rate fields are identified here for the first time.

To leverage our analytical framework, we parameterize the input rate fields with difference-of-Gaussian functions (Fig. 4a-d bottom) whose parameters vary continuously with stimulus size (see STAR Methods and Supplementary SS4, SS4 A). In doing so, we trade off some details of the spatial profiles of the responses to gain interpretability. In particular, to take advantage of the minimal model, we adopt a compact analytical expression for the L4 rate field, that limits the accuracy of the fit for small stimuli (Fig. 4c). However this should be irrelevant for recovering inverse response, because inverse response is mainly dependent on input from HVAs (Keller et al., 2020b). In Supplementary SS11 we show that the results are qualitatively equivalent with far more accurate fits.

The minimal model reproduces inverse size tuning and the spatial profile of inverse response as a function of size (Fig. 5b-c): 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 (Fig. 5d, and see Fig. 3n for comparison). This is because SOM cells respond more strongly to small than to large inverse stimuli (Fig. 4b and Supplementary SS4 A), whereas they respond more strongly to large than to small classical stimuli (Fig. 2).

Inverse response is substantially reduced by reduction of HVAs input by 35% (Fig. 5e). For classical stimuli, the same modification affects L2/3 response only marginally (Supplementary SS13 A). This is consistent with experimental results on optogenetic silencing of HVAs (see Fig.5c in (Keller et al., 2020b)). This differential effect of HVA suppression on inverse vs. classical response is consistent across varying levels of HVA suppression (Supplementary SS13 A). Finally, the model shows an increase of inverse responses with contrast (Fig. 5f, see also Supplementary SS13 B) similar to that found experimentally (Extended Data Fig. 3h of Keller et al. (2020b)).

2.4.1 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 (Fig. 6a,b,

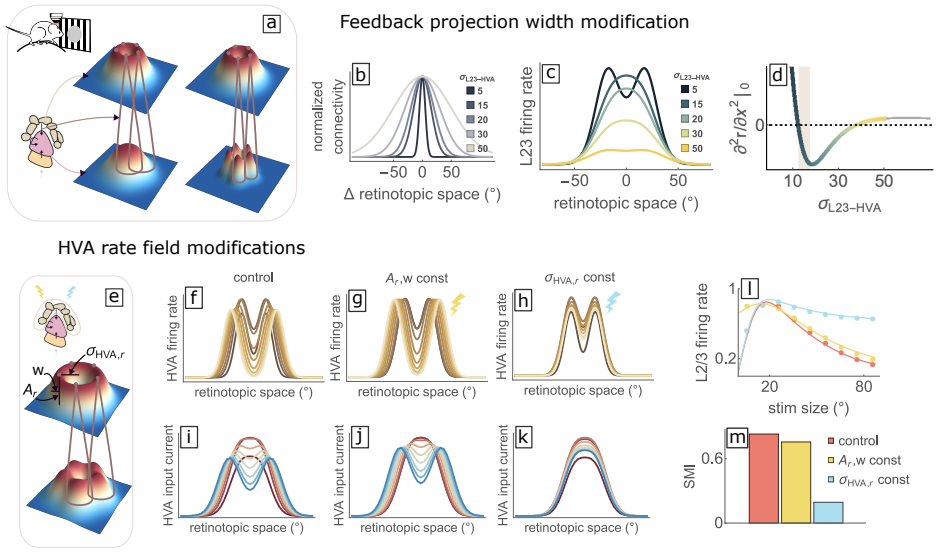

Figure 6: **Inverse response requires wide enough feedback projections and feedback activity profiles that scale with stimulus size. a.** Rate fields in HVAs for a small stimulus (top 3D-plots, ring profile). The feedback projections are represented by the empty cones (narrower projection width on the right). The bottom 3D-plots represent the L2/3 field of input current arising from the feedback projections from 5 example points along the ring of peak firing rate in HVAs; the L2/3 input current would be the sum of similar projections from all points in the HVAs 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, $\sigma_{L23-HVA}$ (see STAR Methods), widths in legend. Amplitudes are normalized to 1. **c.** Rate field of L2/3 for the different values of $\sigma_{L23-HVA}$ for a fixed stimulus width of 15°. Inverse response depends on the feedback projection's width. **d.** Second derivative of L2/3 rate field evaluated at the origin, vs. $\sigma_{L23-HVA}$. The color code indicates the value of $\sigma_{L23-HVA}$ as in **c**, while the shaded bar corresponds to the experimental estimate in (Marques et al., 2018). See text.

Figure 6: **e.** Top: parametrization of HVAs rate field used for **f-m**: we fit the rate field as a convolution of a circle of radius $\sigma_{HVA,r}$ with a Gaussian of standard deviation w and amplitude A_r. Bottom: for fixed feedback projection width which would yield a center peak in **c**, a wide enough radius of HVAs activity, $\sigma_{HVA,r}$, results in a trough in the center. **f-k.** HVAs rate field (**f-h**) and input current (**i-k**) for the control condition (**f,i**; parameters as in previous figures), and for two counterfactual modifications: either the amplitude A_r and the width w (**g,j**, dark yellow symbol) or the size of the ring $\sigma_{HVA,r}$ (**h,k**, light blue symbol) are kept constant (fixed at their value for size $s = 15°$) with varying stimulus size. **l.** Size tuning curve and **m.** Surround Modulation Index for the control condition (red) and the modified conditions (colors as in **g,h**)

see also Supplementary SS1) while the inverse stimulus size is fixed at 15° (Fig. 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 (Fig. 6c), as we now explain. HVAs activity peaks in a ring of cells with RFs corresponding approximately to the hole edge, although the ring size grows somewhat more slowly than the hole size (Fig. 4, SS12). 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 Fig. 6a and discussed more precisely in Supplementary SS14).

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 (Fig. 6c-d). 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 $(12°, 35°)$ (Fig. 6d). The experimental estimate in (Marques et al., 2018) is indicated by the shaded area in Fig. 6d, which falls within this range. The same results hold in a model of L2/3 with 4 recurrent cell types (see Supplementary Fig. SF39).

We can similarly understand why inverse response is size tuned. As the stimulus size increases and the ring-shaped HVAs 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 Fig. 6e). At some point, the input peak moves away from the center, and an ever-deepening trough forms at the stimulus center (Fig. 6i), representing the decrease in response with increasing stimulus size (Fig. 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, Supplementary SS17), another contributing factor could be the decrease in amplitude and thickness of the ring. To test this, we fit the HVAs 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 Fig. 6e). We then apply two counterfactual modifications to the HVAs rate field: i) eliminate the size-dependence of A_r and w while preserving the scaling of $\sigma_{HVA,r}$ with stimulus size (dark yellow symbol, Fig. 6g), ii) eliminate the scaling of $\sigma_{HVA,r}$ while preserving A_r and w (light blue symbol, Fig. 6h). If our explanation is right, (i) should only have a minor effect while (ii) should significantly reduce inverse surround suppression.

Indeed, in case (i), the field of HVA input to L2/3 develops a trough for large stimuli (Fig. 6j), and the inverse size tuning curve is very similar to that in the control case, with a minor decrease in SMI (Fig. 6l-m). 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 (ii), the HVA input field does not develop a trough as the stimulus size increases (Fig. 6k) and, correspondingly, inverse surround suppression is drastically reduced (Fig. 6l-m). Note that, for a classical stimulus, surround suppression is only slightly decreased by an analogous modification (compare Fig. 3e and Fig. 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. Our model shows that classical surround suppression in L2/3 is mostly inherited from the aligned cells of its inputs (and thus is not impacted by widening input) and enhanced by both PV inhibitory input and the SOM-VIP subcircuit. On the other hand, 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 Supplementary Fig. SF39 we show that the explanations derived using the minimal model also apply to a more biologically realistic model with 4 cell-types (see in particular SS14).

The arguments presented here provide a purely geometrical explanation of inverse response and inverse size tuning, illustrating the importance of considering a spatial framework with realistic anatomical and physiological length scales to understand contextual effects.

2.5 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 (Keller et al., 2020a; Shushruth et al., 2012a; Sillito et al., 1995). 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

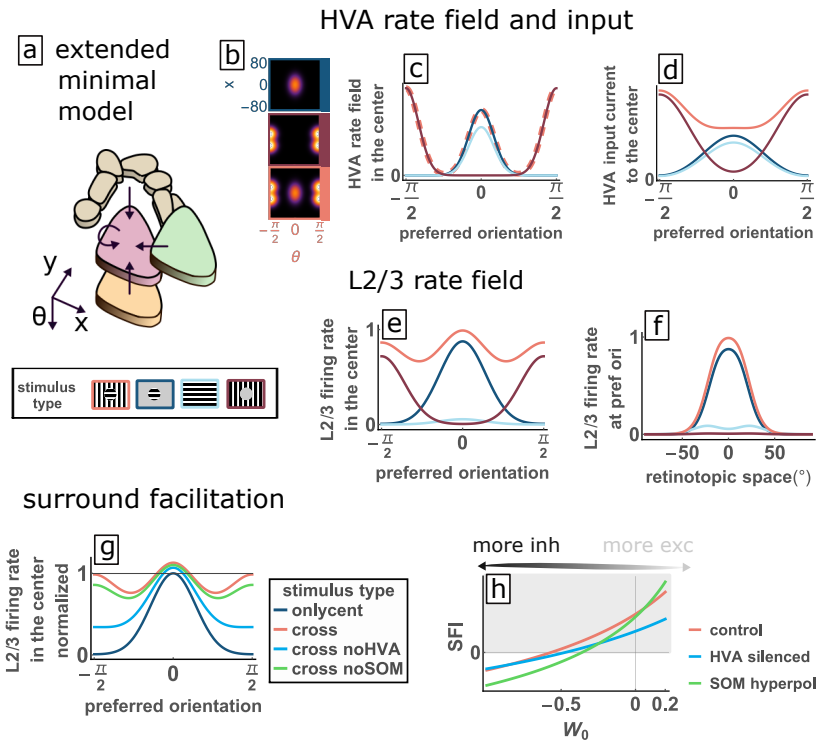

Figure 7: **Surround facilitation is generated by inverse response to the orthogonal surround. a**. Sketch of the extended minimal model, accounting for orientation preference as an additional dimension for the rate fields. Bottom: Legend of the stimuli analysed: cross (center at orientation 0, surround at $\pi/2$; light red), classical/only center (dark blue), iso-oriented surround (light blue), inverse (dark red). **b.** Density plot of HVAs 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). The rate field of the inputs for the cross stimulus (bottom) is the sum of a classical stimulus at orientation $\theta = 0$ (top) plus an inverse stimulus at $\theta = \pi/2$ (middle). **c.** HVAs rate fields of cells whose CRF is centred 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.

Figure 7: **d.** Input current from HVAs to the centered cells in L2/3 as a function of their preferred orientation. Note that the input from L4 is much stronger in the center than in the surround (not shown). **e-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**. Dependence of surround facilitation index (SFI) on the amplitude of the recurrent connections W_0. SFI decreases both when HVAs or SOM cells are silenced. Larger recurrent excitation (less negative or more positive W_0) contributes to larger SFI.

(see Fig. 7a and Supplementary SS18) and connectivity that decreases in strength with increasing difference in preferred orientation according to experimental estimates (Ko et al., 2011, 2013; Rossi et al., 2020) (see Supplementary SS1).

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 (Fig. 7b,c). 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 orthogonal orientation preferences, 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 (Fig. 7e-f). This effect can be understood as follows: HVAs cells at the stimulus center with preferred orientation 0 have zero response to the inverse stimulus (dark red line in Fig. 7c). However, L2/3 cells with the same position and preferred orientation receive small but positive HVA input (Fig. 7d), which comes from more offset HVA cells via their broad feedback projections to L2/3. Thus the inverse stimulus at orientation $\theta = \pi/2$ provides a small HVA input to the L2/3 centered cells that prefer orientation $\theta = 0$. This input generates cross-oriented facilitation (light red curves in Fig. 7e-f).

Thus, our model suggests that HVAs support L2/3 cross-orientation surround facilitation (Fig. 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 Fig. 7g lays 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: $SFI = (r_{\mathrm{x}} - r_{\mathrm{c}})/(r_{\mathrm{x}} + r_{\mathrm{c}})$, where r_{x} (r_{c}) is the firing rate of the L2/3 centered cells in the cross (classical) stimulus condition. We analyse 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 (Fig. 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 Fig.7g, SOM input increases SFI (green curve below the light red). More generally, in Supplementary SS19), we show that SOM input increases (decreases) Pyr+PV 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., 2020a) (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.

3 Discussion

We propose a unifying model accounting for three types of contextual modulation and connecting two previously unrelated phenomena: inverse response and cross-orientation surround facilitation. To our knowledge no previous model has addressed the emergence of inverse responses. Our model suggests that classical surround suppression in L2/3 is largely inherited from surround suppression of its feedforward and feedback inputs (consistent with feedforward and feedback contributions noted in Angelucci et al., 2017; Vangeneugden et al., 2019). Size-tuned inverse response arises from the geometry of feedback projections from HVAs to V1 L2/3, along with a ring-like HVAs spatial pattern of activity to an inverse stimulus that widens with increasing stimulus size. Cross-orientation facilitation can be understood as a combination of classical center response and inverse response.

All previous models of surround suppression have assumed it is mediated by lateral inhibition progressively recruited as the stimulus size increases (Angelucci and Bressloff, 2006; Angelucci et al., 2017; Dipoppa et al., 2018; Li and Young, 2021; Obeid and Miller, 2021a; Rubin et al., 2015). However, we find that, in mouse L2/3 of V1, this mechanism only marginally contributes to surround suppression, because the rate field size increases much more slowly than the stimulus size. Furthermore, the rate field size shows little or no growth with stimulus size in L4, so this mechanism also should not contribute to L4 surround suppression. This holds

true independently of the preprocessing procedure used to define the rate fields (Supplementary SS2 B), and is consistent with the recordings in Dipoppa et al. (2018) (their Supp. Fig. 11D) and in Michaiel et al. (2019) (their Fig.2A), although those authors did not note 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 (2019, 2021), who had participants compare total area of multiple circular patches of color, distributed in random locations within a square 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.

To gain robust mechanistic insight, we include in the minimal model as few elements as possible, to easily disentangle the effects of each. Moreover we keep the parameter space as low-dimensional as possible (as little as 3-dimensional). Instead of quantitatively fine tuning the model to a particular data set, we aim to describe results at a qualitative level, at which they are more robust and consistent across different data sets. 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 4-cell-type circuit (Supplementary SS17).

We find that classical surround suppression in L2/3 is enhanced by increasing SOM lateral input (Adesnik et al., 2012a; Nienborg et al., 2013b), reduced by an increase in the firing of PV cells (Nienborg et al., 2013b), and partially supported by the supralinear neuronal input/output function (Hansel and van Vreeswijk, 2002; Miller and Troyer, 2002; Rubin et al., 2015). In addition, we find that the contrast dependence of classical surround suppression is also largely inherited from L4 inputs(Mossing et al., 2021), and show that the spatial response profile of Pyr and PV cells can change from a peak in the center to a trough in the center for increasingly large classical stimuli, as observed in Dipoppa et al. (2018), if SOM input is sufficiently strong.

Keller et al. (Keller et al., 2020b) highlight the importance of input from retinotopically offset neurons in HVAs for the inverse response. Here, we adopt a continuous description of retinotopic space, as in (Dipoppa et al., 2018; Li and Young, 2021), to preserve the information about length scales of the spatial patterns of activity. In addition, we constrain the range of the projections between different populations based on anatomical results (Billeh et al., 2020; Campagnola et al., 2021; Karnani et al., 2016; Li and Young, 2021; Pfeffer et al., 2013; Rossi et al.,

2020). This allows us to characterize the inverse response in terms of the geometry of the feedback input.

Our schematic description of space opens the way to an analytic treatment of the problem, following and generalizing (Persi et al., 2011). This allows us to determine robust determinants of qualitative response properties, such as: the input properties that are necessary to give rise to surround suppression and that determine the preferred size of the response; possible causes for the decrease of the preferred size when the stimulus contrast is increased; the conditions needed for inverse response rate fields to peak at the center; and the conditions required for inverse surround suppression.

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. We could not model HVAs dynamically because, in the data set considered, L2/3 responses to small-size classical and inverse stimuli are very similar, while LM responds very differently to the two stimuli. This implies that LM must receive a significant part of its input from other sources, which we are not modelling (Blot et al., 2021; Harris et al., 2019; Tohmi et al., 2014).

We modeled HVAs input based on recordings from Area LM (Keller et al., 2020b). 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 scaled observed LM inverse responses accordingly. This assumption needs to be tested experimentally.

Despite the very limited number of parameters, our minimal approach is highly informative. In addition to the three overall phenomena of classical and inverse size tuning and surround facilitation, 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 1 and as we now discuss.

3.1 Testable predictions

SOM neurons reduce the degree of inverse surround suppression (Fig. 5d). Thus, we predict that optogenetically suppressing SOM cells should boost the L2/3 response to smaller inverse stimuli more than to larger inverse stimuli.

We reason that the spatial activity patterns evoked by a center stimulus of one orientation and an inverse stimulus of the orthogonal orientation should add up roughly linearly, to give the spatial pattern evoked by a cross-oriented center/surround stimulus. This can be experimentally tested by recording the spatial pattern of responses to center-only and orthogonally-oriented inverse stimuli to see if their sum indeed predicts the spatial response to cross stimuli. The model also predicts that under presentation of the cross stimulus, L2/3 Pyr and PV cells centered on the stimulus and with preferred orientation orthogonal to the center stimulus should show significant response (Fig. 7c).

Observations and predictions	Fig.	Ref
classical and inverse size-dependent spatial responses, cross-orientation surround facilitation	3c-d, 5b-c, 7e-f, SF31	(Keller et al., 2020a,b),..
the rate field width grows much more slowly than the stimulus size.	2e-h, 4e-h	Compatible with (Dipoppa et al., 2018)
L2/3 responses are stabilized by PV inhibition	SF15, SF35	(Denève and Machens, 2016; Sanzeni et al., 2020)..
by activating PV inhibition, classical surround suppression decreases.	SF15	(Nienborg et al., 2013a)
in mouse L2/3 of V1, suppression is largely inherited from the input, rather than due to recruiting more lateral PV inhibitory activity.	3e-k	Compatible with (Mossing et al., 2021)
classical preferred-size-contrast-dependence is largely inherited from feedforward and feedback layers.	SF23b-c	Compatible with (Mossing et al., 2021)
classical preferred-size-contrast-dependence is supported by large VIP/small SOM response for low contrasts.	SF23h	(Mossing et al., 2021)
SOM cells can generate ring profiles in the rate field of Pyr and PV cells.	3m	Compatible with (Dipoppa et al., 2018)
hyperpolarizing SOM reduces classical surround suppression.	3n, SF36	(Adesnik et al., 2012b)
inverse response (but not classical response) is substantially reduced when input from HVAs is reduced.	5e, SF26, SF37	(Keller et al., 2020b; Vangeneugden et al., 2019)
hyperpolarizing SOM increases inverse surround suppression.	5d, SF36	(prediction)
inverse responses increase with contrast.	5f, SF27, SF38	(Keller et al., 2020b)
inverse response requires wide enough feedback projections.	6a-d, SF39a	Compatible with (Marques et al., 2018)
inverse surround suppression is given by widening of HVAs pattern of activity	6e-m, SF39b-c	Compatible with (Keller et al., 2020b)
when optogenetically silencing a fraction of PV neurons, surround facilitation increases.	7h	(prediction)
cross-orientation surround facilitation should be stronger for cells that have orthogonal classical and inverse preferred orientation, compared to cells with the same preferred orientation for classical and inverse stimuli.	7d-e	(prediction)
Pyr and PV cells in L2/3 centered on the stimulus and with preferred orientation orthogonal to the center stimulus show significant response to the cross stimulus.	7e	(prediction)
optogenetically silencing HVAs during cross stimulus presentation reduces surround facilitation.	7g	(prediction)
hyperpolarizing SOM cells during cross stimulus presentation reduces surround facilitation.	7g	Compatible with (Keller et al., 2020a)

Table 1: All observed phenomena that are included in or generated by the model. Light brown rows indicate known effects reproduced by the model, orange ones indicate phenomena or mechanisms that are highlighted here for the first time but are compatible with work in the literature and light red ones indicates testable predictions. The last four predictions follow from the model prediction that cross-orientation surround facilitation is, at least in part, simply the addition of the inverse response to the center response.

More generally, our model raises many interesting questions in understanding how the orientation tuning of classical and inverse responses are related to one another and to cross responses. In our model, we assume that all neurons have the same OSI and that the orientation preference for classical and inverse stimuli is the same. On the contrary, Extended Data Figure 3 in Keller et al. (2020b) shows that for neurons with high OSI for both classical and inverse stimuli, a significant fraction of neurons have a preferred orientation for classical stimuli that is orthogonal to their preferred orientation for inverse stimuli. Our model predicts that these cells should have the strongest cross-orientation surround facilitation, since they would have the strongest response to the cross surround. They might also have the strongest surround suppression, since they would have the weakest top-down input evoked by the iso surround. More generally, it will be interesting to see if responses to orthogonal center and inverse stimuli add linearly, regardless of a neuron's relative orientation preference for center vs. inverse stimuli, and how these relative preferences impact surround suppression and cross-orientation facilitation.

Our model predicts that optogenetically silencing HVAs during cross stimulus presentation should reduce surround facilitation. Likewise, the model predicts that hyperpolarizing or optogenetically silencing SOM cells during cross stimulus presentation should also reduce surround facilitation. These points are illustrated in Fig. 7g,h.

The fact that the SFI increases as a function of the relative weight of excitation and inhibition (Fig. 7h) predicts that, when optogenetically silencing a fraction of PV neurons, surround facilitation should increase. Fig. 7h also shows that SFI drops when HVAs or SOM inputs are reduced.

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 (Agrillo et al., 2015; Agrochao et al., 2020; Bååth et al., 2014; Bach and Poloschek, 2006; Carbon, 2014; Eagleman, 2001; Endler et al., 2010; Gori et al., 2014; Keemink et al., 2018; Luo et al., 2019; Mély et al., 2018; Pak et al., 2019, 2020; Santacà et al., 2019). 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.

References

H. Adesnik. Synaptic Mechanisms of Feature Coding in the Visual Cortex of Awake Mice. Neuron, 95:1147–1159, 2017.

H. Adesnik, W. Bruns, H. Taniguchi, Z. J. Huang, and M. Scanziani. A neural circuit for spatial summation in visual cortex. Nature, 490(7419):226–231, 2012a.

H. Adesnik, W. Bruns, H. Taniguchi, Z. J. Huang, and M. Scanziani. A neural circuit for spatial summation in visual cortex. Nature, 490:226–231, 2012b.

C. Agrillo, S. Gori, and M. J. Beran. Do rhesus monkeys (macaca mulatta) perceive illusory motion? Animal cognition, 18(4):895–910, 2015.

M. Agrochao, R. Tanaka, E. Salazar-Gatzimas, and D. A. Clark. Mechanism for analogous illusory motion perception in flies and humans. Proceedings of the National Academy of Sciences, 117(37):23044–23053, 2020.

Y. Ahmadian and K. D. Miller. What is the dynamical regime of cerebral cortex? Neuron, 2021.

J. Allman, F. Miezin, and E. McGuinness. Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. Annual Review of Neuroscience, 8(1):407–430, 1985. doi: 10.1146/annurev.ne.08.030185.002203. URL `https://doi.org/10.1146/annurev.ne.08.030185.002203`. PMID: 3885829.

A. Angelucci and P. C. Bressloff. 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, 2006.

A. Angelucci, M. Bijanzadeh, L. Nurminen, F. Federer, S. Merlin, and P. C. Bressloff. Circuits and mechanisms for surround modulation in visual cortex. Annual review of neuroscience, 40:425–451, 2017.

R. Bååth, T. Seno, and A. Kitaoka. Cats and illusory motion. Psychology, 5(9): 1131–1134, 2014.

M. Bach and C. M. Poloschek. Optical illusions. Adv Clin Neurosci Rehabil, 6(2): 20–21, 2006.

H. Barlow. The exploitation of regularities in the environment by the brain. Behavioral and Brain Sciences, 24(4):602–607, 2001.

H. B. Barlow. Single units and sensation: a neuron doctrine for perceptual psychology? Perception, 1(4):371–394, 1972.

Y. N. Billeh, B. Cai, S. L. Gratiy, K. Dai, R. Iyer, N. W. Gouwens, R. Abbasi-Asl, X. Jia, J. H. Siegle, S. R. Olsen, et al. Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. Neuron, 106(3):388–403, 2020.

A. Blot, M. M. Roth, I. Gasler, M. Javadzadeh, F. Imhof, and S. B. Hofer. Visual intracortical and transthalamic pathways carry distinct information to cortical areas. Neuron, 109(12):1996–2008.e6, 2021. ISSN 0896-6273. doi: https://doi.org/10.1016/j.neuron.2021.04.017. URL https://www.sciencedirect.com/science/article/pii/S089662732100283X.

H. Bos, A.-M. Oswald, and B. Doiron. Untangling stability and gain modulation in cortical circuits with multiple interneuron classes. bioRxiv, pages 2020–06, 2020.

L. Campagnola, S. C. Seeman, T. Chartrand, L. Kim, A. Hoggarth, C. Gamlin, S. Ito, J. Trinh, P. Davoudian, C. Radaelli, et al. Connectivity and synaptic physiology in the mouse and human neocortex. bioRxiv, 2021.

M. Carandini and D. J. Heeger. Normalization as a canonical neural computation. Nat. Rev. Neurosci., 13:51–62, 2012.

M. Carandini, J. B. Demb, V. Mante, D. J. Tolhurst, Y. Dan, B. A. Olshausen, J. L. Gallant, and N. C. Rust. Do we know what the early visual system does? Journal of Neuroscience, 25(46):10577–10597, 2005.

C. Carbon. Understanding human perception by human-made illusions. Frontiers in Human Neuroscience, 8, 2014.

J. R. Cavanaugh, W. Bair, and J. A. Movshon. Selectivity and spatial distribution of signals from the receptive field surround in macaque v1 neurons. Journal of neurophysiology, 2002.

R. Coen-Cagli, A. Kohn, and O. Schwartz. Flexible gating of contextual influences in natural vision. Nat Neurosci, 18:1648–1655, 2015.

E. Craft, H. Schutze, E. Niebur, and R. von der Heydt. A neural model of figure–ground organization. Journal of neurophysiology, 97(6):4310–4326, 2007.

S. Denève and C. K. Machens. Efficient codes and balanced networks. Nature Neuroscience, 19(3):375–382, 2016. ISSN 1546-1726. doi: 10.1038/nn.4243. URL https://doi.org/10.1038/nn.4243.

M. Dipoppa, A. Ranson, M. Krumin, M. Pachitariu, M. Carandini, and K. D. Harris. Vision and locomotion shape the interactions between neuron types in mouse visual cortex. Neuron, 98(3):602–615, 2018.

D. M. Eagleman. Visual illusions and neurobiology. Nature Reviews Neuroscience, 2(12):920–926, 2001. ISSN 1471-0048. doi: 10.1038/35104092. URL https://doi.org/10.1038/35104092.

J. A. Endler, L. C. Endler, and N. R. Doerr. Great bowerbirds create theaters with forced perspective when seen by their audience. Current Biology, 20(18):1679–1684, 2010.

D. J. Field, A. Hayes, and R. F. Hess. Contour integration by the human visual system: Evidence for a local "association field". Vision Research, 33(2): 173–193, 1993. ISSN 0042-6989. doi: https://doi.org/10.1016/0042-6989(93)90156-Q. URL https://www.sciencedirect.com/science/article/pii/004269899390156Q.

D. J. Field, J. R. Golden, and A. Hayes. Contour integration and the association field. cortex, 17:2112–2127, 2010.

S. Gori, C. Agrillo, M. Dadda, and A. Bisazza. Do fish perceive illusory motion? Scientific reports, 4(1):1–6, 2014.

D. H. Grosof, R. M. Shapley, and M. J. Hawken. Macaque vi neurons can signal 'illusory'contours. Nature, 365(6446):550–552, 1993.

B. Haider, A. Duque, A. R. Hasenstaub, and D. A. McCormick. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. Journal of Neuroscience, 26(17):4535–4545, 2006.

D. Hansel and C. van Vreeswijk. How noise contributes to contrast invariance of orientation tuning in cat visual cortex. J. Neurosci., 22:5118–5128, 2002.

J. A. Harris, S. Mihalas, K. E. Hirokawa, J. D. Whitesell, H. Choi, A. Bernard, P. Bohn, S. Caldejon, L. Casal, A. Cho, A. Feiner, D. Feng, N. Gaudreault, C. R. Gerfen, N. Graddis, P. A. Groblewski, A. M. Henry, A. Ho, R. Howard, J. E. Knox, L. Kuan, X. Kuang, J. Lecoq, P. Lesnar, Y. Li, J. Luviano, S. McConoughey, M. T. Mortrud, M. Naeemi, L. Ng, S. W. Oh, B. Ouellette, E. Shen, S. A. Sorensen, W. Wakeman, Q. Wang, Y. Wang, A. Williford, J. W. Phillips, A. R. Jones, C. Koch, and H. Zeng. Hierarchical organization of cortical and thalamic connectivity. Nature, 575(7781):195–202, 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1716-z. URL https://doi.org/10.1038/s41586-019-1716-z.

D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. Journal of neurophysiology, 28(2): 229–289, 1965.

J. S. Isaacson and M. Scanziani. How inhibition shapes cortical activity. Neuron, 72(2):231–243, 2011.

H. Jones, K. Grieve, W. Wang, and A. Sillito. Surround suppression in primate v1. Journal of neurophysiology, 86(4):2011–2028, 2001.

G. Kanizsa. Subjective contours. Scientific American, 234(4):48–53, 1976.

M. M. Karnani, J. Jackson, I. Ayzenshtat, J. Tucciarone, K. Manoocheri, W. G. Snider, and R. Yuste. Cooperative subnetworks of molecularly similar interneurons in mouse neocortex. Neuron, 90(1):86–100, 2016.

H. K. Kato, S. K. Asinof, and J. S. Isaacson. Network-Level Control of Frequency Tuning in Auditory Cortex. Neuron, 95(2):412–423, 2017.

S. W. Keemink, C. Boucsein, and M. C. W. van Rossum. Effects of v1 surround modulation tuning on visual saliency and the tilt illusion. Journal of Neurophysiology, 120(3):942–952, 2018. doi: 10.1152/jn.00864.2017. URL https://doi.org/10.1152/jn.00864.2017. PMID: 29847234.

A. J. Keller, M. Dipoppa, M. M. Roth, M. S. Caudill, A. Ingrosso, K. D. Miller, and M. Scanziani. A disinhibitory circuit for contextual modulation in primary visual cortex. Neuron, 108(6):1181–1193, 2020a.

A. J. Keller, M. M. Roth, and M. Scanziani. Feedback generates a second receptive field in neurons of the visual cortex. Nature, 582(7813):545–549, 2020b.

L. Kirchberger, S. Mukherjee, U. H. Schnabel, E. H. van Beest, A. Barsegyan, C. N. Levelt, J. A. Heimel, J. A. Lorteije, C. van der Togt, M. W. Self, et al. The essential role of recurrent processing for figure-ground perception in mice. Science advances, 7(27):eabe1833, 2021.

L. Kirchberger, S. Mukherjee, M. W. Self, and P. R. Roelfsema. Contextual drive of neuronal responses in mouse v1 in the absence of feedforward input. Science Advances, 9:eadd2498, 2023a.

L. Kirchberger, S. Mukherjee, M. W. Self, and P. R. Roelfsema. Contextual drive of neuronal responses in mouse v1 in the absence of feedforward input. Science Advances, 9(3):eadd2498, 2023b.

H. Ko, S. B. Hofer, B. Pichler, K. A. Buchanan, P. J. Sjöström, and T. D. Mrsic-Flogel. Functional specificity of local synaptic connections in neocortical networks. Nature, 473(7345):87–91, 2011.

H. Ko, L. Cossell, C. Baragli, J. Antolik, C. Clopath, S. B. Hofer, and T. D. Mrsic-Flogel. The emergence of functional microcircuits in visual cortex. Nature, 496 (7443):96–100, 2013.

V. A. Lamme. The neurophysiology of figure-ground segregation in primary visual cortex. Journal of neuroscience, 15(2):1605–1615, 1995.

Y. Li and L.-S. Young. Unraveling the mechanisms of surround suppression in early visual processing. PLoS computational biology, 17(4):e1008916, 2021.

J. Luo, K. He, I. M. Andolina, X. Li, J. Yin, Z. Chen, Y. Gu, and W. Wang. Going with the flow: the neural mechanisms underlying illusions of complex-flow motion. Journal of Neuroscience, 39(14):2664–2685, 2019.

T. Marques, J. Nguyen, G. Fioreze, and L. Petreanu. The functional organization of cortical feedback inputs to primary visual cortex. Nature neuroscience, 21(5): 757–764, 2018.

A. M. Michaiel, P. R. L. Parker, and C. M. Niell. A Hallucinogenic Serotonin-2A Receptor Agonist Reduces Visual Response Gain and Alters Temporal Dynamics in Mouse V1. Cell Rep, 26(13):3475–3483, Mar 2019.

K. D. Miller and T. W. Troyer. Neural noise can explain expansive, power-law nonlinearities in neural response functions. Journal of neurophysiology, 87(2): 653–659, 2002.

D. P. Mossing, J. Veit, A. Palmigiano, K. D. Miller, and H. Adesnik. Antagonistic inhibitory subnetworks control cooperation and competition across cortical space. bioRxiv, 2021. doi: 10.1101/2021.03.31.437953. URL https://www.biorxiv.org/content/early/2021/03/31/2021.03.31.437953.

D. Mély, D. Linsley, and T. Serre. Complementary surrounds explain diverse contextual phenomena across visual modalities. Psychological Review, 125, 09 2018. doi: 10.1037/rev0000109.

H. Nienborg, A. Hasenstaub, I. Nauhaus, H. Taniguchi, Z. J. Huang, and E. M. Callaway. Contrast dependence and differential contributions from somatostatin- and parvalbumin-expressing neurons to spatial integration in mouse v1. Journal of Neuroscience, 33(27):11145–11154, 2013a.

H. Nienborg, A. Hasenstaub, I. Nauhaus, H. Taniguchi, Z. J. Huang, and E. M. Callaway. Contrast dependence and differential contributions from somatostatin- and parvalbumin-expressing neurons to spatial integration in mouse V1. J. Neurosci., 33:11145–11154, 2013b.

D. Obeid and K. D. Miller. Stabilized supralinear network: Model of layer 2/3 of the primary visual cortex. bioRxiv, 2021a. doi: 10.1101/2020.12.30.424892. URL https://www.biorxiv.org/content/early/2021/01/02/2020.12.30.424892.

D. Obeid and K. D. Miller. Stabilized supralinear network: Model of layer 2/3 of the primary visual cortex. bioRxiv, https://doi.org/10.1101/2020.12.30.424892, 2021b.

M. Okun and I. Lampl. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. Nature neuroscience, 11(5):535–537, 2008.

H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, and D. Ferster. Inhibitory stabilization of the cortical network underlies visual surround suppression. Neuron, 62(4):578–592, 2009.

C. L. Paffen, M. J. van der Smagt, S. F. Te Pas, and F. A. Verstraten. Center-surround inhibition and facilitation as a function of size and contrast at multiple levels of visual motion processing. Journal of Vision, 5(6):8–8, 2005.

A. Pak, E. Ryu, C. Li, and A. Chubykin. Top-down feedback controls the cortical representation of illusory contours in mouse primary visual cortex. The Journal of Neuroscience, 40:1998–19, 12 2019. doi: 10.1523/JNEUROSCI.1998-19.2019.

A. Pak, E. Ryu, C. Li, and A. A. Chubykin. Top-down feedback controls the cortical representation of illusory contours in mouse primary visual cortex. Journal of Neuroscience, 40(3):648–660, 2020. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.1998-19.2019. URL https://www.jneurosci.org/content/40/3/648.

A. Palmigiano, F. Fumarola, D. P. Mossing, N. Kraynyukova, H. Adesnik, and K. D. Miller. Structure and variability of optogenetic responses identify the operating regime of cortex. bioRxiv, 2020. doi: 10.1101/2020.11.11.378729. URL https://www.biorxiv.org/content/early/2020/11/11/2020.11.11.378729.

E. Persi, D. Hansel, L. Nowak, P. Barone, and C. van Vreeswijk. Power-law input-output transfer functions explain the contrast-response and tuning properties of neurons in visual cortex. PLoS computational biology, 7(2):e1001078, 2011.

E. Peterhans and R. von der Heydt. Mechanisms of contour perception in monkey visual cortex. ii. contours bridging gaps. Journal of Neuroscience, 9(5):1749–1763, 1989.

C. K. Pfeffer, M. Xue, M. He, Z. J. Huang, and M. Scanziani. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. Nature neuroscience, 16(8):1068–1076, 2013.

U. Polat, K. Mizobe, M. W. Pettet, T. Kasamatsu, and A. M. Norcia. Collinear stimuli regulate visual responses depending on cell's contrast threshold. Nature, 391(6667):580–584, 1998.

N. J. Priebe and D. Ferster. Inhibition, spike threshold and stimulus selectivity in primary visual cortex. Neuron, 57:482–497, 2008.

A. Renart, J. De La Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K. D. Harris. The asynchronous state in cortical circuits. science, 327(5965):587–590, 2010.

P. R. Roelfsema. Cortical algorithms for perceptual grouping. Annual Review of Neuroscience, 29(1):203–227, 2006. doi: 10.1146/annurev.neuro.29.051605.112939. URL https://doi.org/10.1146/annurev.neuro.29.051605.112939. PMID: 16776584.

A. F. Rossi, R. Desimone, and L. G. Ungerleider. Contextual modulation in primary visual cortex of macaques. Journal of Neuroscience, 21(5):1698–1709, 2001.

L. F. Rossi, K. D. Harris, and M. Carandini. Spatial connectivity matches direction selectivity in visual cortex. Nature, 588(7839):648–652, 2020.

D. B. Rubin, S. D. Van Hooser, and K. D. Miller. The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. Neuron, 85(2):402–417, 2015.

N. Rubin. Figure and ground in the brain. Nature neuroscience, 4(9):857–858, 2001.

M. Santacà, M. E. Miletto Petrazzini, C. Agrillo, and A. Wilkinson. Can reptiles perceive visual illusions? delboeuf illusion in red-footed tortoise (chelonoidis carbonaria) and bearded dragon (pogona vitticeps). Journal of Comparative Psychology, 133(4):419, 2019.

A. Sanzeni, B. Akitake, H. C. Goldbach, C. E. Leedy, N. Brunel, and M. H. Histed. Inhibition stabilization is a widespread property of cortical networks. Elife, 9: e54875, 2020.

M. P. Sceniak, M. J. Hawken, and R. Shapley. Visual spatial characterization of macaque v1 neurons. Journal of neurophysiology, 85(5):1873–1887, 2001.

L. Schwabe, K. Obermayer, A. Angelucci, and P. Bressloff. The role of feedback in shaping the extra-classical receptive field of cortical neurons: a recurrent network model. J. Neurosci., 26:9117–9129, 2006.

L. Schwabe, J. M. Ichida, S. Shushruth, P. Mangapathy, and A. Angelucci. Contrast-dependence of surround suppression in Macaque V1: experimental testing of a recurrent network model. Neuroimage, 52:777–792, 2010.

M. W. Self, J. A. Lorteije, J. Vangeneugden, E. H. van Beest, M. E. Grigore, C. N. Levelt, J. A. Heimel, and P. R. Roelfsema. Orientation-tuned surround suppression in mouse visual cortex. Journal of Neuroscience, 34(28):9290–9304, 2014.

H. Shin, M. B. Ogando, L. Abdeladim, S. Durand, H. Belski, H. Cabasco, H. Loefler, A. Bawany, B. Hardcastle, J. Wilkes, K. Nguyen, L. Suarez, T. Johnson, W. Han, B. Ouellette, C. Grasso, J. Swapp, V. Ha, A. Young, S. Caldejon, A. Williford, P. Groblewski, S. Olsen, C. Kiselycznyk, J. Lecoq, and H. Adesnik. Recurrent pattern completion drives the neocortical representation of sensory inference. bioRxiv, 2023. doi: 10.1101/2023.06.05.543698. URL `https://www.biorxiv.org/content/early/2023/06/07/2023.06.05.543698`.

S. Shushruth, J. M. Ichida, J. B. Levitt, and A. Angelucci. Comparison of spatial summation properties of neurons in macaque v1 and v2. Journal of neurophysiology, 102(4):2069–2083, 2009.

S. Shushruth, P. Mangapathy, J. M. Ichida, P. C. Bressloff, L. Schwabe, and A. Angelucci. Strong recurrent networks compute the orientation tuning of surround modulation in the primate primary visual cortex. Journal of Neuroscience, 32 (1):308–321, 2012a.

S. Shushruth, P. Mangapathy, J. M. Ichida, P. C. Bressloff, L. Schwabe, and A. Angelucci. Strong recurrent networks compute the orientation tuning of surround modulation in the primate primary visual cortex. J. Neurosci., 32:308–321, 2012b.

J. H. Siegle, X. Jia, S. Durand, S. Gale, C. Bennett, N. Graddis, G. Heller, T. K. Ramirez, H. Choi, J. A. Luviano, P. A. Groblewski, R. Ahmed, A. Arkhipov, A. Bernard, Y. N. Billeh, D. Brown, M. A. Buice, N. Cain, S. Caldejon, L. Casal, A. Cho, M. Chvilicek, T. C. Cox, K. Dai, D. J. Denman, S. E. J. de Vries, R. Dietzman, L. Esposito, C. Farrell, D. Feng, J. Galbraith, M. Garrett, E. C. Gelfand, N. Hancock, J. A. Harris, R. Howard, B. Hu, R. Hytnen, R. Iyer, E. Jessett, K. Johnson, I. Kato, J. Kiggins, S. Lambert, J. Lecoq, P. Ledochowitsch, J. H. Lee, A. Leon, Y. Li, E. Liang, F. Long, K. Mace, J. Melchior, D. Millman, T. Mollenkopf, C. Nayan, L. Ng, K. Ngo, T. Nguyen, P. R. Nicovich, K. North, G. K. Ocker, D. Ollerenshaw, M. Oliver, M. Pachitariu, J. Perkins, M. Reding, D. Reid, M. Robertson, K. Ronellenfitch, S. Seid, C. Slaughterbeck, M. Stoecklin, D. Sullivan, B. Sutton, J. Swapp, C. Thompson, K. Turner, W. Wakeman, J. D. Whitesell, D. Williams, A. Williford, R. Young, H. Zeng, S. Naylor, J. W. Phillips, R. C. Reid, S. Mihalas, S. R. Olsen, and C. Koch. Survey of spiking in the mouse visual system reveals functional hierarchy. Nature, 592(7852):86–92, 2021. ISSN 1476-4687. doi: 10.1038/s41586-020-03171-x. URL https://doi.org/10.1038/s41586-020-03171-x.

A. M. Sillito, K. L. Grieve, H. E. Jones, J. Cudeiro, and J. Davls. Visual cortical mechanisms detecting focal orientation discontinuities. Nature, 378(6556):492–496, 1995.

L. Spillmann, B. Dresp-Langley, and C. Tseng. Beyond the classic receptive field: The effect of contextual stimuli. Journal of Vision, 15:1–22, 2015.

M. Tohmi, R. Meguro, H. Tsukano, R. Hishida, and K. Shibuki. The extrageniculate visual pathway generates distinct response properties in the higher visual areas of mice. Current Biology, 24:587–597, 2014.

J. Vangeneugden, E. H. van Beest, M. X. Cohen, J. A. Lorteije, S. Mukherjee, L. Kirchberger, J. S. Montijn, P. Thamizharasu, D. Camillo, C. N. Levelt, et al. Activity in lateral visual areas contributes to surround suppression in awake mouse v1. Current Biology, 29(24):4268–4275, 2019.

R. von der Heydt. Approaches to visual cortical function. Rev. Physiol. Biochem. Pharmacol., 108:69–150, 1987.

J. Xing and D. J. Heeger. Measurement and modeling of center-surround suppression and enhancement. Vision research, 41(5):571–583, 2001.

S. R. Yousif and F. C. Keil. The additive-area heuristic: An efficient but illusory means of visual area approximation. Psychological Science, 30:495 – 503, 2019. URL https://api.semanticscholar.org/CorpusID:73461531.

S. R. Yousif and F. C. Keil. How we see area and why it matters. Trends in Cognitive Sciences, 25(7):554–557, 2021.

C. M. Ziemba, J. Freeman, E. P. Simoncelli, and J. A. Movshon. Contextual modulation of sensitivity to naturalistic image structure in macaque v2. Journal of neurophysiology, 120(2):409–420, 2018.

K. Zipser, V. A. Lamme, and P. H. Schiller. Contextual modulation in primary visual cortex. Journal of Neuroscience, 16(22):7376–7389, 1996.

Acknowledgments

We acknowledge funding sources NIH U01NS108683, NIH R01EY029999, NIH U19NS107613, NSF 1707398, Gatsby Charitable Foundation GAT3708.

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., M.R.; Writing-Original Draft, S.D.S.; Writing-Review and Editing, K.D.M., S.D.S. and M.D.; Visualization, S.D.S.; Supervision, K.D.M.; Funding Acquisition,K.D.M., M.S.

Declaration of interests

The authors declare no competing interests.

STAR Methods

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and code should be directed to and will be fulfilled by the lead contact, Serena Di Santo (serenadisanto@ugr.es)

Materials availability

This study did not generate new unique reagents.

Data and code availability

All original code will been deposited and is publicly available as of the date of publication. DOIs are listed in the key resources table.

METHOD DETAILS

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:

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

In Eq.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:

$$u(\mathbf{x}) = I(\mathbf{x}) + \int d\mathbf{y}\, W(\mathbf{x} - \mathbf{y}) u^2(\mathbf{y}) \tag{2}$$

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

$$u(\mathbf{x}) = \alpha(\mathbf{x}, v_u) G(\mathbf{x}, v_u) \tag{3}$$

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:

$$\begin{aligned} \mathrm{Rec}(\mathbf{x}) = \int d\mathbf{y}\, W(\mathbf{x} - \mathbf{y}) u^2(\mathbf{y}) &= W_0 \int d\mathbf{y} \alpha^2(\mathbf{y}) G(\mathbf{x} - \mathbf{y}, v_r) G^2(\mathbf{y}, v_u) \\ &= \frac{W_0}{4\pi v_u} \int d\mathbf{y} \alpha^2(\mathbf{y}) G(\mathbf{x} - \mathbf{y}, v_r) G\left(\mathbf{y}, \frac{v_u}{2}\right). \end{aligned} \tag{4}$$

The largest contribution to the integral is given by the maximum of its argument (Laplace approximation), but since $\alpha^2(\mathbf{y})$ varies slowly ($\alpha'(\mathbf{y}) \ll \alpha(\mathbf{y})$), 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 $\mathbf{y} = \bar{\mathbf{y}} \equiv \frac{v_u \mathbf{x}}{2v_r + v_u}$:

$$\begin{aligned} \mathrm{Rec}(\mathbf{x}) &\simeq \frac{W_0 \alpha^2(\bar{\mathbf{y}})}{4\pi v_u} \int d\mathbf{y} G(\mathbf{x} - \mathbf{y}, v_r) G\left(\mathbf{y}, \frac{v_u}{2}\right) \\ &= \frac{W_0}{4\pi v_u} \alpha^2(\bar{\mathbf{y}}) G(\mathbf{x}, v_{ru}) \end{aligned} \tag{5}$$

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 Eq. 2 we get:

$$\alpha(\mathbf{x})G(\mathbf{x}, v_u) \simeq I_0 G(\mathbf{x}, v) + \frac{W_0}{4\pi v_u}\alpha^2(\mathbf{x})G(\mathbf{x}, v_{ru}), \tag{6}$$

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

$$\alpha(\mathbf{x}) \simeq \frac{2\pi v_u}{W_0 G(v_{ru})}\left(G(v_u) \pm \sqrt{G^2(v_u) - \frac{I_0 W_0 G(v)G(v_{ru})}{\pi v_u}}\right) \tag{7}$$

where for better readability we are omitting the spatial dependence of the function $G(\cdot)$. In Supplementary SS7 A we show that $\alpha(\mathbf{x}) \ll \alpha'(\mathbf{x})$.

We now choose v_u to satisfy:

$$2v_u^{-1} = v^{-1} + v_{ru}^{-1} \tag{8}$$

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

$$\frac{G(v)G(v_{ru})}{G^2(v_u)} = \frac{v_u^2}{v v_{ru}} \tag{9}$$

and thus we can write:

$$\alpha(\mathbf{x}) \simeq \frac{2\pi v_u}{W_0 G(v_{ru})} G(v_u)\left(1 \pm \sqrt{1 - \frac{I_0 W_0 v_u}{\pi v v_{ru}}}\right) \tag{10}$$

Finally, using Eq. 9 again:

$$\begin{aligned} u(\mathbf{x}) &\simeq \frac{2\pi v v_{ru}}{W_0 v_u}\left(1 \pm \sqrt{1 - \frac{I_0 W_0 v_u}{\pi v v_{ru}}}\right) G(v) \\ &= \frac{\pi(v + v_{ru})}{W_0}\left(1 \pm \sqrt{1 - \frac{2 I_0 W_0}{\pi(v + v_{ru})}}\right) G(v) \end{aligned} \tag{11}$$

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 Supplementary SS7 B).

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

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

making the same ansatz as above, we obtain

$$u(\mathbf{x}) = \frac{\pi(v_1 + v_{ru})}{W_0} \left(1 - \sqrt{1 - \frac{2I_1 W_0}{\pi(v_1 + v_{ru})} - 4v_1^2 W_0 \sum_{k=1}^{m} \frac{I_k G(v_{1k})}{(v_1 - v_k) v_{ru}}} \right) G(v_1), \tag{13}$$

where we defined

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

Note that, despite the multiple approximations, the analytic solution provides a really good description of the rate field (see *e.g.* Fig. 3c,e and Fig. 5a,c).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data preprocessing

The distance between the center of the stimulus and the center of the CRF of cells is estimated in (Keller et al., 2020b) 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 (Keller et al., 2020b). 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 $\sigma_{err} = 5°$ 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:

$$\tilde{r}_A(\mathbf{x}, s) = \frac{\sum_{i=1}^{N_A} r_i(s) e^{-\frac{(\mathbf{x}-\mathbf{x}_i)^2}{2v_{err}}}}{\sum_{i=1}^{N_A} e^{-\frac{(\mathbf{x}-\mathbf{x}_i)^2}{2v_{err}}}} \tag{15}$$

where N_A is the number of recorded cells of type A and s indicates the stimulus size. In Supplementary SS2 B we show an example of the smoothening that this procedure (first presented in (Dipoppa et al., 2018)) generates. 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.

Functional form of the rate fields

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

$$\tilde{r}_A^{(c)}(\mathbf{x}, s) \simeq r_A^{(c)}(\mathbf{x}, s) = p_A^c(s) e^{-\frac{\mathbf{x}^2}{2\tilde{v}_A^c(s)}} \tag{16}$$

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 $p^c_A(s)$ 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 Supplementary SS4.

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

$$\tilde{r}^{(i)}_A(\mathbf{x}, s) \simeq r^{(i)}_A(\mathbf{x}, s) = p^i_A(s) e^{-\frac{\mathbf{x}^2}{2\bar{v}^i_{pA}(s)}} - q^i_A(s) e^{-\frac{\mathbf{x}^2}{2\bar{v}^i_{qA}(s)}}, \tag{17}$$

with $p^i_A(s) > q^i_A(s) > 0$.

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 that 1 degree) and:

$$\frac{\bar{v}^i_{pA} - \bar{v}^i_{qA}}{\bar{v}^i_{pA}} < 0.05\%. \tag{18}$$

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 $p^c_A(s)$ 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 $\frac{\partial p^c_A(s)}{\partial s} < 0$ and preferred size (or CRF size) as $s^* : \frac{\partial p^c_A(s)}{\partial s}|_{s=s^*} = 0$ (see Supplementary SS8). We use the parametrization:

$$\begin{aligned} p^c_A(s) &= \beta^A_1 \operatorname{erf} \frac{s}{S^A_{\beta 1}} - \beta^A_2 \operatorname{erf} \frac{s}{S^A_{\beta 2}} \\ \bar{v}^c_A(s) &= k^A_{c1} + k^A_{c2} s \end{aligned} \tag{19}$$

where the first equation encodes the size tuning curve and β^A_2/β^A_1 modulates the amount of surround suppression, $S^A_{\beta 1}$ modulates the steepness of the increase of size tuning for small sizes and $S^A_{\beta 2}$ modulates the steepness of the decrease of size tuning for large sizes. Moreover $k^A_{c2} > 0$ indicates the growth of the rate field width with stimulus size.

The size tuning curves described by the functional form in 19, together with the data and the fitted samples of the rate fields are shown in Supplementary SS4 A.

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

For SOM population we use the parametrization:

$$\begin{aligned} p_S^i(s) &= \alpha_1^S \operatorname{erf} \frac{s}{S_{\alpha 1}^S} - \alpha_2^S \operatorname{erf} \frac{s}{S_{\alpha 2}^S} \\ q_S^i(s) &= \rho_1^S \operatorname{erf} \frac{s}{S_{\rho 1}^S} - \rho_2^S \operatorname{erf} \frac{s}{S_{\rho 2}^S} \end{aligned} \tag{20}$$

The evidence presented in (Keller et al., 2020b) 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 Supplementary SS11). Thus for L4 cells we use:

$$p_L^i(s) = q_L^i(s) = \alpha_1^L \operatorname{erf} \frac{s}{S_{\alpha 1}^L} - \alpha_2^L \operatorname{erf} \frac{s}{S_{\alpha 2}^L} \tag{21}$$

that constrains the center to have zero response, consistently with the very low response of aligned cells in L4 shown in (Keller et al., 2020b). Finally for LM we use:

$$\begin{aligned} p_M^i(s) &= \alpha_1^M + \alpha_2^M s \\ q_M^i(s) &= p_M^i(s) - \rho_1^M \left(\operatorname{erf} \frac{s}{S_{\rho 1}^M} - \operatorname{erf} \frac{s}{S_{\rho 2}^M} \right) \end{aligned} \tag{22}$$

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: $p_A^i(s) e^{-\frac{\tilde{\mathbf{x}}^2}{2\bar{v}_{pA}^i(s)}} - q_A^i(s) e^{-\frac{\tilde{\mathbf{x}}^2}{2\bar{v}_{qA}^i(s)}}$ with $\tilde{\mathbf{x}} = (30°, 30°)$. Aligned and offset size tuning curves pre and post parametrization are shown in Supplementary SS4 A. 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:

$$\begin{aligned} \bar{v}_{pA}^i(s) &= k_{i1}^A + k_{i2}^A s \\ \bar{v}_{qA}^i(s) &= k_{i3}^A + k_{i2}^A s \end{aligned} \tag{23}$$

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 (Keller et al., 2020b). But we know that V1 receives input from all HVAs. Moreover Extended Data Fig. 9 in (Keller et al., 2020b) 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:

$$r_H^{(i)}(\mathbf{x}, s) = \eta r_M^{(i)}(\mathbf{x}, s) \tag{24}$$

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 Supplementary SS5).

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 $W_{AB}(\mathbf{x}-\mathbf{y}) = \bar{W}_{AB}G(\mathbf{x}-\mathbf{y}, v_{AB})$, where $v_{AB} = \sigma_{AB}^2$:

$$\begin{aligned} I(\mathbf{x}, s) &= \int_{-\infty}^{\infty} \bar{W}_{AB}G(\mathbf{x}-\mathbf{y}, v_{AB})r_B(\mathbf{y}, s)d\mathbf{y} \\ &= \bar{W}_{AB}C_0 p_B^c(s)2\pi\bar{v}_B^c(s)G(\mathbf{x}, v_{AB} + \bar{v}_B^c(s)) \end{aligned} \tag{25}$$

Thus we can read:

$$\begin{aligned} I_0^c(s) &= \bar{W}_{AB}C_0\bar{r}_B(s)2\pi\bar{v}_B^c(s) \\ v &= v_{AB} + \bar{v}_B^c(s). \end{aligned} \tag{26}$$

Finally, putting together the functional forms that we derived in the paragraph above, with Eq. 26, 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

$$\begin{aligned} I^i(\mathbf{x}, s) &= \int_{-\infty}^{\infty} \bar{W}_{AB}G(\mathbf{x}-\mathbf{y}, v_{AB})r_B(\mathbf{y}, s)d\mathbf{y} \\ &= \bar{W}_{AB}C_0\left[p_B^i(s)2\pi\bar{v}_{pB}^i(s)G(v_{AB} + \bar{v}_{pB}^i(s)) - q_B^i(s)2\pi\bar{v}_{qB}^i(s)G(v_{AB} + \bar{v}_{pB}^i(s))\right] \end{aligned} \tag{27}$$

thus more compactly

$$\begin{aligned} I^i(\mathbf{x}, s) &= I_{01}^i(s)G(v_{B1}) + I_{02}^i(s)G(v_{B2}) \\ I_{01}^i(s) &= \bar{W}_{AB}C_0 p_B^i(s)2\pi\bar{v}_{pB}^i(s) \\ I_{02}^i(s) &= -\bar{W}_{AB}C_0 q_B^i(s)2\pi\bar{v}_{qB}^i(s) \\ v_{B1} &= v_{AB} + \bar{v}_{pB}^i(s) \\ v_{B2} &= v_{AB} + \bar{v}_{qB}^i(s) \end{aligned} \tag{28}$$