Title: Object boundary detection in natural images may depend on 'incitatory' cell-cell interactions

Authors: Gabriel C. Mel[1*], Chaithanya A. Ramachandra[2], and Bartlett W. Mel[1,3]

Affiliations:

[1]Department of Computer Science, University of Southern California, Los Angeles, CA.

[2]Eyenuk, Inc, Los Angeles, CA, USA.

[3]Neuroscience Graduate Program, University of Southern California, Los Angeles, CA.

*Correspondence to: meldefon@gmail.com

Abstract:

Detecting object boundaries is crucial for recognition, but how the process unfolds in visual cortex remains unknown. To study the problem faced by a hypothetical boundary cell, and to predict how cortical circuitry could produce a boundary cell from a population of conventional "simple cells", we labeled 30,000 natural image patches and used Bayes' rule to help determine how a simple cell should influence a nearby boundary cell depending on its relative offset in receptive field position and orientation. We identified three basic types of cell-cell interactions: rising and falling interactions with a range of slopes and saturation rates, as well as non-monotonic (bump-shaped) interactions with varying modes and amplitudes. Using simple models we show that a ubiquitous cortical circuit motif consisting of direct excitation and indirect inhibition – a compound effect we call "incitation" – can produce the entire spectrum of simple cell-boundary cell interactions found in our dataset. Moreover, we show that the synaptic weights that parameterize an incitation circuit can be learned by a single-layer "delta" rule. We conclude that incitatory interconnections are a generally useful computing mechanism that the cortex may exploit to help solve difficult natural classification problems.

Significance statement:

Simple cells in primary visual cortex (V1) respond to oriented edges, and have long been supposed to detect object boundaries, yet the prevailing model of a simple cell – a divisively normalized linear filter – is a surprisingly poor natural boundary detector. To understand why, we analyzed image statistics on and off object boundaries, allowing us to characterize the neural-style computations needed to perform well at this difficult natural classification task. We show that a simple circuit motif known to exist in V1 is capable of extracting high-quality boundary probability signals from local populations of simple cells. Our findings suggest a new, more general way of conceptualizing cell-cell interconnections in the cortex.

Introduction:

The primary visual cortex (area V1) is a complex, poorly understood, multi-purpose image processor optimized to extract information from natural scenes – which are themselves complex, poorly understood signals. Thus, understanding how V1 operates presents a challenging reverse engineering problem. A longstanding hypothesis is that orientation-tuned V1 cells somehow participate in object boundary detection, a core process in biological vision (Biederman, 1987; Gilbert and Wiesel, 1990; Heydt and Peterhans, 1989; Hubel and Wiesel, 1962; Kapadia et al., 1995) that is crucial for the functions of both ventral and dorsal streams (Biederman, 1987; Hoffman, 2000; Rust and Dicarlo, 2010; Theys et al., 2015). However, little progress has been made in refining or testing this hypothesis, in part due to our lack of understanding of the structure of natural object boundaries, and particularly, what a V1 cell needs to do to reliably distinguish boundaries from non-boundaries.

This uncertainty has made it difficult to form specific computational hypotheses as to how V1 circuits perform this behaviorally-relevant classification task. Previous work has analyzed natural image statistics to determine how local boundary segments are arranged in images (Sanguinetti et al., 2010; Sigman et al., 2001), and how these arrangements relate to human contour grouping performance (W. S. Geisler, Perry, Super, & Gallogly, 2001). However, no study has yet attempted to deconstruct the natural boundary *detection* problem in detail, or to link the computations necessary for boundary detection to particular neural mechanisms.

With the goal to better understand the computations underlying object boundary detection in V1 (Figure 1A), we began with a known cell type – orientation-tuned "simple cells" (as defined by Hubel and Wiesel, 1962), and typically modeled as divisively normalized oriented linear filters (Carandini and Heeger, 2012) – and asked how the outputs of a population of simple cells (SCs), whose receptive fields (RFs) densely cover an area of the visual field, should be combined to produce a "boundary cell" (BC) whose firing rate represents the probability that an object boundary is present within its RF (Figure 1B). When framed in this way, Bayes' rule tells us what data to extract from natural images to obtain an answer to the question. In a previous study (Ramachandra and Mel, 2013), we noted that under the simplifying assumption of "class conditional independence" (see methods for a detailed discussion), simple cell-boundary cell interactions are captured by the log-likelihood ratio (LLR) functions embedded in Bayes' rule (colored expressions in Figure 1C), which represent the evidence that a given simple cell provides about the presence of an object boundary within a neighboring boundary cell's receptive field (Figure 1D). We found that SC-BC interactions were diverse, and in some cases involved

compound excitatory and inhibitory effects. However, since only a small number of cells was analyzed in that study, we could not come to general conclusions about the types of cell-cell interactions needed to compute boundary probability, making it difficult to compare and contrast possible neural mechanisms.

In this study, we analyze a much larger dataset, and compute the full set of simple cell-boundary cell interaction functions for a population of 300 simple cells surrounding a "reference location" where a boundary might be detected. We find that the simple cell-boundary cell interactions suggested by the natural image LLR functions follow a predictable pattern that depends on the offset in position and orientation between simple cell and boundary cell receptive fields, and we show that a well-known cortical circuit motif can implement the entire spectrum of individual SC-BC interactions found in our data set. Finally, we demonstrate that a cortically-inspired neural network can produce a boundary-detecting cell from simple cells with a single layer of excitatory synapses and a single inhibitory interneuron. Our findings suggest that cortical sensory computations, including the detection of natural object boundaries, may depend on a specific class of structured excitatory-inhibitory cell-cell interactions.

Materials and methods:

Image preprocessing

As in Ramachandra and Mel (2012), we used a modified version of the COREL database for boundary labeling in natural images. Several image categories, including sunsets and paintings were removed from the full COREL database since their boundary statistics differed markedly from that of typical natural images. Custom code was used to select ~30,000 20x20 pixel image patches for labelling. The "reference location" representing a hypothetical boundary cell's receptive field location was defined as the elongated, horizontal 2x4 pixel region at the center of the patch (dashed box, Figure 1A, B).

Natural image data collection

To collect ground-truth data relating to natural contour statistics, for each image patch to be labeled, a horizontal 2x4 pixel rectangular box was drawn around a centered "reference location" and human labelers were asked to answer the question, "On a scale from 1 to 5, with 1 meaning 'extremely unlikely' and 5 meaning 'extremely likely' – how likely is it that there is an object boundary passing horizontally through the reference box, end to end, without leaving the box?" To qualify as valid, boundary segments also had to be visible and unoccluded within the box. We restricted labelling to horizontal boundaries (i.e. horizontal reference boxes) since pixel lattice discretization made it more difficult to judge oblique orientations, and because we expected cell response statistics in natural images to be roughly orientation invariant. (This expectation was supported by subsequent tests showing that LLR functions obtained for horizontal boundaries also led to high boundary detection performance on oblique boundaries). Labeler responses were recorded, and patches with scores of 1 or 2, were classified as "no" patches, while patches with scores of 4 or 5 were classified as "yes" patches. Agreement between labelers was very high, based on informal observations when two labelers worked together. Rare ambiguous patches that could cause labeler disagreement were often

given scores of 3, so these patches were excluded from our analyses. After labeling, the dataset was doubled by adding left-right flipped versions of each patch, and assigning the same label as the un-flipped counterpart.

Collecting virtual simple cell responses on and off natural boundaries

Given a large set of image patches, some labeled as boundaries and others not, the next step was to collect virtual simple cell responses densely covering boundary vs. non-boundary image patches so that their different statistics could be analyzed. Original color image patches were converted to single-channel (monochrome) intensity images $(0.29\ R + 0.59\ G + 0.11\ B)$. Simple cell-like oriented linear "filters" were created by rotating a 2x4 pixel horizontal filter kernel in 15° increments from 0 to 165° (i.e. 12 orientations). The filter kernel for the horizontal (0°) cell is shown in Figure 2A. Positive filter coefficients represent ON subregions, and negative coefficients represent OFF subregions of the virtual simple cell's receptive field. Blank kernel entries represent zeros. Filter coefficients in rotated kernels were computed by rotating the horizontal kernel using "bilinear interpolation" (https://en.wikipedia.org/wiki/Bilinear_interpolation) (Figure 2B, C). In essence, any entry in an oriented kernel that was affected by the rotated horizontal pattern would be assigned a coefficient interpolated based the nearest 4 coefficients in the original horizontal pattern after back-rotation. (The locations of the faint blue dots in Figure 2A show the locations of the dark blue dots in Figure 2B rotated back by 45° into the horizontal frame of references; the 4 coefficients closest to each faint blue

dot were used for the interpolation). A similar procedure was used to generate all other rotated filter kernels.

A simple cell's "response" value at a particular image location was computed as the dot product between the cell's filter kernel and the underlying image intensity pixel values. A simple cell generated a positive response when its positive kernel coefficients mostly overlapped with bright image pixels and its negative coefficients mostly overlapped with dark image pixels. The largest positive responses occurred on light-dark boundaries of a cell's preferred orientation. A negative response was treated as a positive response of a distinct simple cell with opposite contrast polarity (i.e. rotated by 180°).

Given a labeled image patch, simple cell responses covering that patch were collected on a 5 x 5 grid centered on the patch, at each of 12 orientations (Figure 2D). This led to 5 x 5 x 12 = 300 simple cells responding to any given image patch. Simple cell response data was accumulated separately for "yes" and "no" labeled patches.

Restriction to "normalized" image patches

The prevailing model of a simple cell consists of linear filter whose output is divisively normalized by activity in the cell's surround (Carandini and Heeger, 2012). Normalizing a simple cell's response

involves (1) computing the cell's pre-normalized response (e.g., as we do above using an oriented linear filter); (2) measuring the sum N of other neural activity in the cell's surround including cells of the same or different orientation located in the near or far surround; and (3) inhibiting the simple cell's response as an increasing function of N. Normalization is often called "divisive" as the surround activity term, N, generally appears in the denominator of the overall cell response expression. Response normalization plays a variety of roles in the brain (see Carandini and Heeger, 2012 for an excellent review), though in the visual system normalization is most often discussed as a means of counteracting the effect of multiplicative "nuisance" factors that modulate the responses of entire local population of neurons in a correlated fashion. For example, if the level of illumination varies across the visual field, a common occurrence in natural scenes, the spatial variation in light level tends to drive up response rates of all neurons in brighter areas and drive down response rates of all neurons in darker areas. Normalization circuitry helps to cancel out these correlated, population-level neural response variations, leading to a greater degree of illumination invariance across the visual field, and a greater degree of statistical independence in the neural code (Wainwright & Simoncelli, 2000; Schwartz & Simoncelli, 2001; Liang, Simoncelli, & Lei, 2000; Zetsche & Rohrbein, 2001; Parra, Spence, & Sajda, 2000; Fine et al., 2003; Karklin & Lewicki, 2003, 2005; Zhou and Mel 2008).

The importance of normalization with respect to our analysis is that neural activity in the non-classical surround of a simple cell can powerfully influence its response, and given that our study involves collecting and statistically analyzing simple cell responses in natural images, we made efforts to take account of the effects of normalization circuitry on simple cell responses. As discussed above, the most common assumption about response normalization in the cortex is that it depends on the summed activity of a pool of neurons in the vicinity of a target cell (see Carandini and Heeger 2012 for

177 review), and that the normalization function is "divisive" (meaning the normalization term appears in
178 the denominator). We adopted this assumption, and selected a pool of 100 oriented cells from the
179 300 shown in Figure 2D using an ad-hoc procedure designed to find a subset of the cells that would be
180 minimally correlated in normalized patches. (Limiting the normalizing pool to a subset of the 300 cells
181 turned out not to be unnecessary, as we found that including all 300 cells in the normalizer pool led to
182 functionally equivalent results, but we describe the selection procedure below since the 100-cell
183 normalizer was what was actually used to generate our result figures. The reader not interested in the
184 details of the normalization approach can skip to the next paragraph). We started by defining a "basic
185 normalizer" consisting of the single linear filter value at the reference location. We then culled out a
186 large set of natural image patches whose basic normalizer values fell into a narrow range, that is, all
187 the culled patches had roughly the same filter value at the reference location. The value used for the
188 basic normalizer was 10±2, but the particular value mattered little; what mattered was that the value
189 was fixed across all image patches collected. We then incrementally "grew" a normalization pool as
190 follows. The cell at the reference location was considered to be the first cell in the normalization pool
191 (C_1). A second cell, call it C_2, was added to the pool by choosing that cell (from the 299 remaining) for
192 which the correlation of its absolute value to that of the other cells in the pool (which was only C_1 in
193 this case) was closest to zero. The absolute value was used because negative values were considered
194 to be responses of distinct cells with opposite polarity. A third cell C_3 was then chosen (from the 298
195 remaining) for which the correlation of its absolute value to that of C_1 and C_2 was, on average, closest
196 to zero. This "greedy" ("choose the best each time") procedure continued until 100 cells were chosen,
197 leading to the particular subset of cell RFs shown in Figure 2E. The set of 100 cells is also specified in
198 the Matlab file "normalizer_filter_subs.mat" in the Supplementary Materials. Having chosen the 100-
199 cell normalization pool, we could now cull out a second-generation-normalized set of image patches

from the image database by selecting patches for which the sum of the responses of all 100 cells in the normalization pool (converted to absolute values) fell into a fixed response range N=200±40. The data in our results figures was generated from this set of normalized image patches. However, as mentioned above, random image patches normalized using all 300 filters surrounding the reference location (and N≈600 +/- 40; exact bin center chosen to match mean on our normalized labelled patch database) were indistinguishable as a group (see examples in Figures 2F and 2G), had indistinguishable averages (upper left each image grid), and led to LLR functions for boundary detection that were very similar.

The restriction of our analysis to image patches within a narrow range of normalizer values was functionally similar to carrying out our analysis “under controlled lighting conditions". The benefit of this restriction was that it allowed us to probe the relationships between simple cell responses and boundary probability without having to assume that we know the precise form of the function that normalizes neural responses against changes in lighting or other regional correlating factors. The cost of this restriction is that we obtain information only about the boundary probability computation at a specific normalization level. This would be a serious problem if the boundary probability computation changed significantly at different levels of normalization. To mitigate this risk we verified that our analysis produced similar main results for 2 other value ranges of the 100-cell normalizer (N = 300 ± 40 and N = 400 ± 40). We accomplished this by assuming that the normalizing function divisively scales filter responses, which, given that the filters are linear, is equivalent to divisively scaling the image patches themselves. When we scaled down the image patches from the other two normalizer bins by factors of 300/200 = 1.5 and 400/200 = 2, respectively, and thereafter processed all image patches

equivalently, we obtained results that were functionally indistinguishable from those produced from the original "200" dataset (data not shown). In our main results figures, we include only the data derived from image patches in the original "200" normalizer bin.

Bayesian formalism

We assume that a boundary cell computes $p(\, yes \mid r_1\, r_2\, r_3 \cdots)$, that is, the probability that a boundary is present at the reference location given a set of simple cell responses $r_1\, r_2\, r_3$, etc., whose receptive fields surround and cover the reference location (Fig. 1D). Using Bayes' rule we obtain

$$p(\, yes \mid r_1\, r_2\, r_3 \cdots) = \frac{p(\, r_1\, r_2\, r_3 \cdots \mid yes)\, p(yes)}{p(r_1\, r_2\, r_3 \cdots \mid yes) p(yes) + p(r_1\, r_2\, r_3 \cdots \mid no) p(no)} \qquad \text{(Eq. 1)}$$

Dividing through by the numerator and rearranging, we find

$$p(\, yes \mid f_1\, f_2\, f_3 \cdots) = \frac{1}{1 + \frac{p(no)}{p(yes)} exp\left(-\log \frac{p(f_1\, f_2\, f_3 \cdots \mid yes)}{p(f_1\, f_2\, f_3 \cdots \mid no)}\right)} \qquad \text{(Eq. 2)}$$

239 The log term in the denominator (shaded pink) of Eq. 2 consists of a ratio of two "likelihoods". The
240 numerator and denominator of the likelihood ratio represent, respectively, the probability of
241 measuring a particular combination of simple cell responses ($r_1\ r_2\ r_3 \cdots$) when a boundary is present at
242 the reference location (yes), and the probability of measuring the *same* combination of filter
243 responses when a boundary is not present at the reference location (no). When a simple cell response
244 combination ($r_1\ r_2\ r_3 \cdots$) is measured in the vicinity of a reference location, and based on statistical
245 record keeping we know that that response combination is more likely to be associated with a yes case
246 than a no case, then the likelihood ratio is greater than 1, meaning there is positive evidence for a
247 boundary; the logarithm term is greater than 0, and the exponential term drops below 1 due to the
248 negative sign of the exponent. When the likelihood ratio is *much* greater than 1, the exponential term
249 approaches zero, and the denominator of the full expression approaches 1, indicating that a boundary
250 at the reference location is highly probable. The yellow-shaded term in the denominator of Eq. 2 is the
251 prior "odds" of finding an object boundary at a randomly sampled location, which is the ratio of the
252 prior probabilities of "yes" and "no" cases. Yes cases made up 2.4% of our default dataset (for N=200),
253 meaning the odds of a boundary was 2.4/97.6 = 2.45%. The odds term functions as a sort of evidence
254 threshold: the lower the prior odds of a boundary, the stronger the evidence must be in order to reach
255 a 50% probability that a boundary is present at the reference location. For prior odds of 2.45%, a
256 likelihood ratio of 41 would be needed to reach a 50% boundary probability.
257
258 *Simplifying Bayes rule by assuming class conditional independence (CCI)*
259
260 Using Bayes rule as a tool for interpreting cortical circuit interactions runs into the obstacle that the
261 likelihood expressions contributing to the likelihood ratio involve high-dimensional probability

262 densities, where the dimension corresponds to the number of simple cells that will contribute to the
263 boundary probability calculation (which could number in the 100's). Collecting high-dimensional
264 probability density functions from natural images, and representing them whether mathematically, by
265 rote tabulation, or neurally, is for all intents and purposes intractable. However, if we assume that
266 neighboring simple cells display a certain kind of statistical independence – i.e. "class conditional"
267 independence – this radically simplifies the computational problem by allowing the high-dimensional
268 likelihoods in Bayes rule to be factored and re-expressed as a sum of 1-dimensional log likelihood ratio
269 (LLR) functions $\lambda_i(r_i)$. These 1-dimensional LLR functions can be interpreted as "the varying effect that
270 cell i has on a nearby boundary cell over its range of responses r_i". Fortuitously, the functions λ_i are
271 simple in form (i.e. smooth and unimodal), and can be easily computed by a known cortical circuit
272 motif. We therefore begin our analysis under the CCI assumption, as it provides a stepping stone to
273 understanding the types of cell-cell interactions we should expect to find in a cortical circuit in which
274 boundary probability is computed. We then take steps to bridge the gap between the predictions of
275 the simplified analysis that assumes class conditional independence, and the types of cell-cell
276 interactions that we should expect to find in the more realistic scenario where the simple cells
277 contributing to a boundary probability computation do not satisfy the CCI assumption.
278
279
280 The assumption of class-conditional independence in our scenario implies that simple cell responses
281 are statistically independent both for the set of image patches where a boundary is present at the
282 reference location, and for the set of image patches where a boundary is not present at the reference
283 location. When CCI holds, both the numerator and denominator terms in the likelihood ratio of Eq. 2

can be factored into individual cell response terms, and converted to a sum of cell-specific log likelihood ratio (LLRs) functions, as follows:

$$\log \frac{p(r_1\, r_2\, r_3 \cdots | yes)}{p(r_1\, r_2\, r_3 \cdots | no)} =$$

$$= \log \frac{p(r_1 \mid yes\,) \cdot p(r_2 \mid yes) \cdot p(r_3 \mid yes) \cdots}{p(r_1 \mid no\,) \cdot p(r_2 \mid no) \cdot p(r_3 \mid no) \cdots}$$

$$= \log \frac{p(r_1 \mid yes\,)}{p(r_1 \mid no\,)} + \log \frac{p(r_2 \mid yes\,)}{p(r_2 \mid no\,)} + \log \frac{p(r_3 \mid yes\,)}{p(r_3 \mid no\,)} + \cdots$$

$$= \lambda_1(r_1) + \lambda_2(r_2) + \lambda_3(r_3) + \cdots$$

where $\lambda_i(r_i)$ is the log likelihood ratio function for the ith neighboring cell. In simple terms, the LLR function $\lambda_i(r_i)$ captures how the evidence for a boundary at the reference location depends on the response of the ith neighboring simple cell. We can write the CCI version of Bayes rule as

$$p(\, yes \mid r_1\, r_2\, r_3 \cdots) = \frac{1}{1 + \frac{p(no)}{p(yes)} \cdot exp(-\sum_i \lambda_i(r_i)\,)}$$

which reduces to the remarkably simple formula

$$= \text{sigmoid}\,(\sum_i \lambda_i(r_i)).$$

304

305 In intuitive terms, this equation says that for a boundary cell at the reference location to compute
306 boundary probability within its receptive field, the response of each neighboring simple cell r_i should
307 be passed through a nonlinear function λ_i whose form depends on the neighbor's offset in position
308 and orientation from the reference location; the results from all contributing simple cells should be
309 summed; and the total should be passed through a final sigmoidal nonlinearity to produce the
310 boundary cell's response (Figure 1).

311

312 *Extracting the log likelihood ratio functions $\lambda_i(r_i)$ from natural images*

313 Histograms of the responses of the 300 simple cells surrounding the reference location were collected
314 separately for "yes" and "no" image patches (total of 600 histograms). Histograms for no patches
315 contained 50 evenly spaced bins. Yes histograms were binned more coarsely because our dataset had
316 many fewer yes patches than no patches; between 8 and 20 evenly spaced bins were used to ensure
317 smoothness of the response histograms for all cells. The yes and no histograms for each simple cell
318 were converted to yes and no probability density functions (pdfs) by dividing the sample count in each
319 bin by two factors: (1) the total sample count in the respective histogram, and (2) the bin width. From
320 these pdfs, LLR functions $\lambda(r)$ were constructed for each simple cell as $\lambda(r) = \log\frac{\mathrm{p}(r|yes)}{\mathrm{p}(r|no)}$, where
321 $p(r|yes)$ and $p(r|no)$ were the cell's yes and no pdf bin values, respectively. The log function was the
322 natural logarithm (base e). To control noise levels, each cell's λ function was considered valid only at
323 non-extreme response levels for which the probability in both the yes and no pdfs exceeded minimum
324 thresholds ($p(r|yes) > 0.005$ and $p(r|no) > 0.002$). Different thresholds were used because

smaller probabilities could be estimated more reliably in the no histograms given the much larger set of no image patches. Only data inside valid response regions is plotted in Figures 3-5 and Figure 7.

The same procedure was repeated using different simple cell profiles (2x6, 2x8, 4x8, and 6x8 pixels) to generate the LLR functions shown in Figure 4.

Modeling log likelihood ratio functions $\lambda(r)$ as differences of sigmoids

As a step in the direction of a circuit level model, we fit the measured LLR functions $\lambda(r)$ for all 300 cells surrounding the reference location with a difference of two sigmoids, $d(r) = \frac{A_1}{1+\exp[-g_1(r-t_1)]} - \frac{A_2}{1+\exp[-g_2(r-t_2)]}$. For each cell, approximate amplitudes A, gains g, and thresholds t for the two sigmoids were chosen automatically by minimizing mean-squared error (MSE). The parameters of the sigmoids were then finely adjusted by hand so that the difference of the two sigmoids visually matched the measured $\lambda(r)$ function as closely as possible. We found visually-guided fine optimization better captured the essential shape structure of the LLR function compared to conventional quantitative measures such as MSE. A similar fitting procedure was used for the three models in Figure 7 (model details can be found in the figure and caption). The risk that human visually-guided optimization of curve shape would alter our conclusions was minimal since (1) human visually-guided optimization is based on a much more sophisticated shape-based metric than, say, MSE, and can therefore be reasonably considered as "ground truth"; (2) our conclusions do not depend on quantitative comparisons of fit quality for different models; and (3) the ability to precisely match individual LLR

347 function shapes using a difference of two simple functions is mainly of didactic interest; the more
348 practically significant question is whether a weighted sum of simple excitatory and inhibitory functions
349 (which will in general involve more than two curves) can produce the LLR-like interactions needed to
350 produce improved boundary detection performance (see Figure 8).
351
352
353 For the surfaces in Figure 6E, and Figure 7, excitatory gain was computed by measuring the excitatory
354 component's average slope between r = 0 and r = 10. Inhibitory gain in Figure 5 was computed in the
355 same way. In Figure 7, each of the inhibition families had only a single parameter; this parameter is
356 what is plotted in the inhibitory gain surface plots.
357

358 *Optimizing cell-cell interactions in a cortically-inspired network using gradient descent learning*

359 In the data analysis portions of the paper, the term "simple cell responses" referred to the value
360 obtained by computing the dot product between the receptive field kernel and the underlying intensity
361 image. This operation yielded the simple cell's linear response value r_i for that receptive field location.
362 In a step towards greater biological realism, for the network-level simulation of Figure 8, each of the
363 300 oriented receptive fields surrounding the reference location was represented a population of 8
364 model simple cells, all sharing the same receptive field and thus the same linear response value r, but
365 having different output nonlinearities (reflecting natural variability within the neuronal population).
366 The differentiation of responses within each 8-cell subpopulation was achieved by modifying the
367 threshold in the sigmoidal output function applied to each cell's linear response, $y_i(r) =$

368 $\frac{A}{1+\exp[-g(r-t)]}$. All cells shared the same amplitude and gain parameters ($A = 1, g$ =0.5), but spanned

369 a range of thresholds from $t = -6$ to 35 in even steps.

370

371

372 The synaptic weights connecting each model simple cell to the boundary cell were obtained as follows.

373 Each image patch created a pattern of activation across the 2,400 model simple cells (300 linear RFs x 8

374 output threshold variants). We used logistic regression to train a linear classifier to distinguish

375 boundary from non-boundary image patches using the 2,400 model simple cells as inputs. A subset of

376 the data (25,000 of the ~30,000 labeled patches) was used for training. During training, data was

377 balanced by duplicating boundary-containing patches such that boundary and non-boundary

378 exemplars were equal in number. Training was done using batch gradient descent with a learning rate

379 of $\eta = 0.1$, performed for 1000 iterations. The net effect on the boundary cell of the 8 model simple

380 cells sharing the same RF location is visualized in Figure 8B. Each graph shows the weighted sum of the

381 outputs of the 8 simple cell variants as a function of r, the cells' shared underlying linear response

382 value. To facilitate comparison of these graphs with the explicit LLR function $\lambda_i(r)$ for that RF

383 location, we scaled the colored interaction functions within each plot. Each plot has one scaling factor

384 that applies to all 5 colored curves in the plot. The inverse of the scaling factor, which can be thought

385 of as the weight that the classifier puts on the curves contained in the subplot, is shown by grey bars.

386

387 *Precision-Recall curves measure boundary classification performance*

388 Precision-Recall (PR) curves were generated for the boundary cell network with optimized weights, as

389 well as for the naïve Bayes classifier (based on a literal sum of all cell LLRs; Figure 1BC) and other

390 classifier variants (Figure 8). The term “classifier” refers to any cell or network or mathematical
391 formula that produces a value in response to an image patch, where larger values are meant to signify
392 higher boundary probability. A classifier consisting of the single simple cell at the reference location
393 provided the PR baseline (Figure 8C, blue curve). To generate a PR curve, a classifier was applied to
394 each of the 5,000 labeled but un-trained “test” image patches, and the patches were sorted by the
395 classifier’s output. A threshold was set at the lowest classifier output obtained over the entire test set,
396 and was systematically increased until the highest output in the test set was reached. For every
397 possible threshold, above-threshold image patches were called putative boundaries and below-
398 threshold patches were called putative non-boundaries. “Precision” was calculated by asking what
399 fraction of patches identified as putative boundaries were true boundaries (according to the human
400 ground truth labels), and “Recall” was calculated by asking what fraction of true boundaries were
401 identified as putative boundaries. As the threshold increased, the Precision and Recall values swept out
402 a curve in PR space. Perfect performance would mean 100% Precision and Recall simultaneously,
403 corresponding to the top right corner of the PR graph. Precision-Recall curves are an alternative to
404 ROC curves, and are preferable in domains where the classes are very unbalanced in terms of prior
405 probability. This is the case in our study: boundary images made up only 2.4% of the overall dataset, vs.
406 97.6% non-boundary cases. See this excellent article for a discussion of this issue
407 (https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-
408 python/).

409 *Boundary cell stimulus responses*

410 The idealized boundary image, analogous to a spike-triggered average stimulus, was computed by
411 averaging all natural image patches weighted by their boundary cell response (Figure 11A). For Figure

412 11B,C, sinusoidal grating stimuli were generated on a 20x20 pixel grid at 24 orientations in 15° steps,
413 and at 60 phase-shifts per cycle. The spatial frequency was chosen to be 0.25 cycles/pixel because it
414 led to relatively artifact-free stimuli at 20x20 pixel resolution, and evoked robust boundary cell
415 responses. For consistency with earlier results, the contrast of each grating image was adjusted to have
416 the same normalizer value (200) as the natural image patches used in the LLR analysis. This was done
417 by generating the grating patch at 100% contrast, computing the normalizer value N on the generated
418 patch using the 100-cell normalization pool (see section above on normalization for details), and
419 scaling down the grating patch by the factor N/200. This procedure reduced the worry that boundary
420 cell responses to the artificially grating stimuli would be distorted by floor or ceiling effects.
421
422
423 Simple cell responses to the grating stimuli were presented to the network of Figure 8 just as was done
424 with natural images patches (see above). For Figure 11C, responses were averaged over all phases of
425 the grating at each orientation. Tuning curves in Figure 11D were obtained by presenting natural image
426 stimuli from the N=200 set of normalized image patches. Red and blue curves are for images with 90th
427 and 10th percentile contrast at the reference location, respectively. These percentiles varied in their
428 contrast by roughly a factor of 2. Contrast was defined as the linear filter response at the reference
429 location divided by the average intensity over the reference filter's 2x4 pixel region of support.
430
431 **Results**
432 To gain insight into the cell-cell interactions needed for natural boundary detection, we collected and
433 labeled 30,000 natural image patches, with scores ranging from 5, indicating high confidence that a

434 boundary was present at a "reference location" (RL, indicated by a dashed box in Figure 1A), down to
435 1, indicating high confidence that a boundary was *not* present at the RL. From these labeled patches,
436 we histogrammed oriented linear filter values (representing simple cell responses) separately for "yes"
437 (scores of 4-5) and "no" (scores of 1-2) cases (red and blue histograms in Figure 3A, respectively).
438 From the responses of 300 neighboring simple cells at 12 orientations on a 5x5 pixel lattice centered
439 on the RL, we computed the likelihoods $p(r_i \mid yes)$ and $p(r_i \mid no)$, meaning the probability of the ith
440 simple cell having a particular response r_i when the patch either does ("yes") or does not ("no")
441 contain a horizontal boundary. We show in the Methods section that, for a boundary cell to compute
442 the probability of a boundary, and contingent on the assumption that the different filter responses are
443 class conditionally independent, the boundary cell should have as its input the sum of the log likelihood
444 ratio (LLR) functions, $\lambda_i(r_i) = \log \frac{p(r_i \mid yes)}{p(r_i \mid no)}$ of all contributing simple cells. Each function $\lambda_i(r_i)$
445 represents the evidence – positive or negative – that simple cell i, when responding at level r_i,
446 provides about the existence of a boundary at the reference location. This evidence should be
447 summed by the boundary cell, and then passed through a sigmoidal output function to yield an actual
448 boundary probability (Figure 1C, D).
449
450 Accordingly, we computed the LLR functions for all of the 300 simple cells surrounding the reference
451 location. Examples of LLR functions are shown in Figure 3B, and the full set is shown in Figure 3C
452 grouped across 5 horizontal shifts at each orientation and vertical position. The LLR functions varied
453 considerably with position and orientation relative to the reference location, but nonetheless
454 conformed to a small number of qualitative shape prototypes (rising, falling, and bump-shaped). When
455 we generated LLR functions for simple cells receptive fields of different sizes and aspect ratios (2x6,
456 4x6, 4x8, and 6x8 pixel RF profiles) we found a qualitatively similar pattern of results, indicating that

457 the basic shapes of the LLR functions do not depend sensitively on the RF profiles of the simple cell
458 receptive fields (Figure 4).
459
460 The LLR functions $\lambda_i(r_i)$ are important functions to understand since they encode the way a simple
461 cell's response r_i individually affects boundary probability at a neighboring receptive field location
462 (positively, negatively, or a combination of both). To gain insight into the forms of the $\lambda_i(r_i)$ functions
463 we observed, we developed a simple mathematical model of the process of LLR function formation. If
464 each cell's yes and no distributions are approximated as gaussian (i.e. unimodal bell-shaped curves)
465 with different means and variances, the resulting LLR functions will be quadratic in form, that is, they
466 will have parabolic shapes. Since the no response distribution is virtually always more variable than the
467 yes distribution, the LLR functions will take the form of downward-pointing parabolas (Figure 5A),
468 qualitatively resembling the LLR functions seen in Figure 3. The particular height and width of each LLR
469 function is determined by the means and variances of the yes and no distributions for that cell (Figure
470 5A, different colored curves). In addition to qualitatively capturing the range of observed LLR function
471 shapes, this model has a simple interpretation in terms of natural image statistics: in image patches
472 that do not contain boundaries at the reference location (which is to say most image patches), the
473 responses of simple cells in the neighborhood tend to vary widely. On the other hand, image patches
474 that do contain boundaries at the reference location are more constrained, and the responses of
475 nearby simple cells tends to be clustered more tightly around characteristic values, r_i^*. The bump-
476 shaped LLR functions, peaked at r_i^*, simply reflect the fact that a simple cell provides the strongest
477 positive evidence for a boundary when its response is more typical of "yes" image patches than "no"
478 patches.
479

480 To facilitate the interpretation of the λ functions as cell-cell interactions, we slightly reformatted them,
481 in two ways. First, the λ functions were shifted vertically in order that they passed through the origin,
482 reflecting the idea that when a simple cell is not firing (corresponding to r_i = 0 on the graph), its
483 influence on the boundary cell (y-value on the graph) should also be zero. This shift was justified given
484 that the outputs of these functions would later be combined additively (Figure 1C, D), and thus the
485 vertical offsets across the entire population of simple cells could be collapsed into a single net offset at
486 the level of the boundary cell (that would likely be small due to cancellation of positive and negative
487 shifts). Second, simple cell firing rates can only be positive, so the left half of each LLR function,
488 corresponding to a negative simple cell firing rate, was "rectified" (i.e. set to zero). Information was
489 not lost since the same or very similar function would be covered by a different simple cell with the
490 same RF but opposite contrast polarity. The right panel of Figure 5A shows the combined effect of the
491 shift and rectify operations. The full set of shifted LLR functions $\lambda_i^s(r_i)$ obtained this way is shown in
492 Figure 5B, with the plots corresponding to the conceptual curves in A marked by asterisks.
493
494 Returning to the interpretation of shifted LLR functions as simple cell-boundary cell interactions, for
495 some simple cells λ^s increased monotonically from the origin, meaning that, as the simple cell's
496 response increased from zero, the evidence it provided to the boundary cell grew steadily more
497 positive. This type of monotonic positive SC-BC interaction was seen for simple cells that were most
498 directly supportive of the hypothesis that a boundary was present at the reference location, such as
499 the simple cell directly overlapping with the RL (middle column in top row of Figure 5B). Referring to
500 the model of Figure 5A, this was a case where the downward-pointing "parabola" peaked far to the
501 right of the origin, so that over the simple cell's entire observed firing range, its effect on the boundary
502 cell remained on the rising limb of the parabola (case 1 in the right panel of Figure 5A). At even higher

503 firing rates than are plotted in Figure 5B, the λ^s function would eventually reach its peak and turn back
504 downward, but such high filter values were so rare in yes patches in our natural image data set that the
505 LLR curves could not be reliably estimated beyond the range shown. Two other cases of pure positive
506 λ^s functions are worth noting: the lower left and right corners of Figure 5B. These cases apply to simple
507 cells whose RFs are nearly "upside down" (i.e. polarity reversed) versions of the reference filter profile,
508 but shifted vertically either 2 pixels above or below the reference location. The fact that these cells are
509 monotonically supportive of the reference hypothesis can be attributed to the existence of many 1-2
510 pixel wide light and dark horizontal bands in our natural image data set.
511
512 For other simple cells, the λ^s functions *decreased* monotonically from the origin, meaning that, as the
513 simple cell's firing rate increased from zero, the evidence it provided to the boundary cell grew
514 increasingly more negative. This type of monotonic negative cell-cell interaction was seen for simple
515 cells whose firing supported a hypothesis *incompatible* with the hypothesis that a boundary was
516 present at the RL. The clearest examples of such cells are those with RFs perpendicular to the RL
517 (middle row, green LLR curves. Referring again to the quadratic LLR model of Figure 5A, these
518 monotonically decreasing λ^s functions arose from cases where the downward-pointing LLR "parabola"
519 was peaked at, or to the left of the origin, so that over the entire response range of the simple cell (to
520 the right of the origin), the λ^s function fell continuously along its descending limb (as in case 2 in Figure
521 5A).
522
523 For the majority of simple cells, however, the λ^s functions were bump-shaped, first rising and then
524 falling as the simple cell's firing rate increased from zero. This type of non-monotonic cell-cell
525 interaction was seen for simple cells whose receptive fields had some degree of overlap in position and

526 orientation with the reference location, so that when a boundary was present at the reference
527 location, their characteristic response levels r_i^* fell in the "middling" range. As a result, these cells
528 provided increasing positive evidence for a boundary at the reference location for response levels up
529 to r_i^*, but at even higher response levels, they began to signal stronger evidence for a boundary *at*
530 *their own RF location and orientation* rather than the reference location. This led to a progressive
531 weakening of the evidence signal sent to the reference boundary cell at response levels above r_i^*.
532

533 **A known circuit mechanism can produce the entire observed spectrum of λ^s functions**

534 Given that the input to a boundary cell can be approximate as the sum of the λ^s functions associated
535 with different simple cells (see Figure 1), and that the shifted LLR functions for different simple cells
536 can be either monotonic or nonmonotonic functions of the simple cell's response, we next asked what
537 kind of neural interconnection circuit is capable of producing the types of monotonic and non-
538 monotonic interaction functions that we observed. One candidate mechanism is the ubiquitous circuit
539 motif in which a cortical cell both directly excites and disynaptically inhibits other cells in its
540 neighborhood (Buzsáki, 1984; George et al., 2011; Isaacson and Scanziani, 2011; Klyachko and Stevens,
541 2006; McBain and Fisahn, 2001; Pfeffer et al., 2013; Pouille and Scanziani, 2001; Swadlow, 2002; Wehr
542 and Zador, 2003). For monotonic positive or negative SC-BC interactions, the effect could in principle
543 be mediated by pure excitatory or inhibitory connections, respectively (through an interneuron in the
544 case of a pure inhibitory connection) (left and middle cases in Figure 6A). Non-monotonic cell-cell
545 interactions, however, would seem to require a compound excitatory-inhibitory (E-I) interconnection
546 scheme (Figure 6A, rightmost case), wherein the excitatory effect dominates at low firing rates and the
547 inhibitory effect dominates at high firing rates. To determine whether this circuit motif can produce

548 the full range of cell-cell interactions contained in our data set, we assumed that both the direct
549 excitatory and indirect inhibitory pathways exert a sigmoidal effect on the boundary cell, and therefore
550 fit each LLR function with the difference of two sigmoid functions (i.e. assuming the boundary cell sums
551 the excitatory and inhibitory contributions). Each of the sigmoids was allowed to vary in threshold,
552 gain, and amplitude (Figure 6B). The fits are shown in Figure 6C, confirming that the range of cell-cell
553 interactions needed to calculate boundary probability in natural images, including non-monotonic
554 interactions, can be produced by a simple model of a circuit motif known to be present in V1 (see
555 reference list just above). To determine whether the successful fitting of λ^s functions depended on our
556 particular choice of sigmoidal E and I basis functions, we repeated the fitting procedure using 3
557 alternative sets of E and I sigmoidal basis functions and obtained similar results (Figure 7). This
558 indicates that the basic types of cell-cell interactions needed to detect object boundaries in natural
559 images can be produced easily by this general type of compound E-I, or "incitatory" circuit.
560
561 We next looked for regularities in the progression of excitatory-inhibitory curve pairs used to fit the λ^s
562 functions depending on a neighbor cell's offset in position and orientation from the reference location
563 (Figure 6D). We observed the following patterns. First, as the neighbor's orientation offset from the
564 RL increases and approaches 90 degrees (indicated by lightness changes within each plot), excitation
565 becomes weaker, and inhibition becomes both stronger and lower in threshold, resembling cross-
566 orientation suppression (a staple function of V1 (Bishop et al., 1973; DeAngelis et al., 1992; Geisler and
567 Albrecht, 1992); though see Priebe and Ferster, 2006). Second, we observed a gradual weakening of
568 both excitation and inhibition as a neighbor cell moves further from the RL in the direction
569 perpendicular to the boundary orientation (different plot columns), reflecting the expected decline in
570 informativeness as a neighbor cell moves further from the boundary cell in question. To probe this

571 effect further, we characterized each excitatory and inhibitory curve by its gain parameter and plotted
572 the gains separately as a function of a neighbor's orientation difference and spatial offset relative to
573 the RL (Figure 6E). These surfaces confirm that, under this simple difference-of-sigmoids model, the
574 strength of the excitation and inhibition imparted to a boundary cell by neighboring simple cells varies
575 systematically with offset in RF position and orientation. The pattern is non-obvious, however, so that
576 if measured experimentally, it could be difficult to interpret without the benefit of a normative
577 framework such as the one we have adopted here.
578
579
580 **Optimizing the parameters for a boundary-detecting incitation network**
581
582 In pursuit of our goal to understand how cells in V1 detect natural object boundaries, our approach
583 thus far has been to frame boundary detection as a Bayesian classification problem, where the inputs
584 are simple cells (with the simplifying assumption that all simple cells are CCI) (Figure 1); collect ground
585 truth data from human-labeled natural images (Figures 3, 4); and calculate what the simple cell-
586 boundary cell interactions should look like in situations where the CCI assumption holds true. The
587 approach has led to two main findings. First, simple cell-boundary cell interaction functions λ_i^s are in
588 general nonlinear and often nonmonotonic (Figure 5), and so cannot be captured by a direct
589 connection from an SC to a BC through a conventional positive or negative weight. On the other hand,
590 we showed that the entire family of rising, falling, and bump-shaped SC-BC interaction functions *can* be
591 captured by a 2-stage "incitation" circuit motif that is known to exist in V1 (Figure 6A, 7).
592

593 The precise forms of the SC-BC interaction functions that we might expect to find in V1 remains in
594 question, however: The λ_i^s functions shown in Figures 5B and 6C are only the optimal cell-cell
595 interaction functions if all of the simple cells impinging on the boundary cell are CCI. To reiterate, this
596 rather severe condition means that no simple cell provides any information about any other simple
597 cell's activity level either on or off object boundaries (see Methods and Appendix for details). The
598 condition is, of course, trivially true for a *single* simple cell, and in that case, the simple cell's functional
599 connection to the boundary cell should be exactly its LLR. The assumption may also hold true for a
600 small population of minimally overlapping simple cells (see Ramachandra and Mel, 2013), but it is
601 definitely not true for 300 simple cells with densely overlapping receptive fields, which is the scenario
602 we consider here. Generally speaking, if the number of simple cells exceeds the number of underlying
603 stimulus dimensions, which is at most 64 in our case (i.e. the number of pixels covered by the 300
604 simple cells' RF profiles), then the representation is "overcomplete", and the simple cells are
605 necessarily correlated. For such a correlated cell population, the likelihood terms in Eq. 2 (pink
606 shading) can no longer be factored, and boundary probability can no longer be expressed as a sum of
607 cell-specific LLR terms $\lambda_i(r_i)$.
608
609 This leads to the following interesting question: instead of imposing parameters on the incitation
610 circuit in order to represent cell-specific λ_i^s functions as we did in Figures 6 and 7, suppose the synaptic
611 weights of an incitation circuit are trained to maximize boundary detection performance. What would
612 the individual SC-BC interaction functions look like then? Would they be similar to their cell-specific λ_i^s
613 functions? Or would the SC-BC interaction functions be significantly altered, presumably due to
614 correlations within the simple cell population?
615

616 To answer this question, we set up a slightly augmented incitation network (Figure 8A) whose
617 modifiable parameters included (i) the excitatory weights connecting each simple cell to the boundary
618 cell, and (ii) the excitatory weights connecting each simple cell to the inhibitory interneuron. In
619 addition, to make it possible for the network to learn nonlinear SC-BC interaction functions by
620 modifying only a single layer of excitatory weights, each simple cell was replicated 8 times to form a
621 small population of closely related cells, all sharing the same oriented receptive field, but each having a
622 different firing threshold. The threshold variability can be seen as arising from natural variation in
623 neuron size, morphology, firing dynamics, etc. The functional purpose of this scheme is that each
624 simple cell's activation level (before the threshold is applied) is effectively being recoded through a set
625 of fixed nonlinear basis functions, which facilitates learning. The regularly spaced threshold settings
626 used for the groups of 8 cells are given in the Methods section. Each pre-synaptic simple cell acted on
627 the boundary cell through two adjustable weights, one excitatory weight directly onto the boundary
628 cell, and one excitatory weight onto the boundary cell's inhibitory "partner" cell, which would
629 contribute to disynaptic inhibition of the boundary cell (Figure 8A). The inhibitory neuron was
630 modeled as a linear cell whose firing rate was a weighted sum of its synaptic inputs. Three examples of
631 oriented receptive fields (red, green, and yellow) and their associated simple cell variants are depicted
632 schematically in Figure 8A.
633
634 Training occurred as follows. Labeled image patches containing boundaries and non-boundaries (with
635 equalized probability) were presented to the 2,400 (=300x8) simple cells; ground-truth labels from the
636 natural image dataset were presented to the boundary cell (1 for boundary, 0 for no boundary); and
637 the excitatory synapses between the simple cells and the boundary cell and its associated inhibitory
638 neuron were adjusted using a supervised logistic regression learning rule (Murphy, 2012). We then

639 performed virtual neurophysiology to probe the net effect of each oriented receptive field on the
640 boundary cell's response induced by the 8 simple cell variants (i.e. basis functions) sharing that RF.
641
642 The learned interaction functions again included monotonic rising and falling as well as non-monotonic
643 bump-shaped functions (Figure 8B, colored curves). For some cells the learned SB-BC interaction
644 functions corresponded closely to their respective sLLRs (thin grey lines), most notably the cells
645 centered on the RL at all different orientations (middle column of Figure 8B). For this category of cells,
646 the simplified Bayesian formulation seems to explain their role in the boundary detection computation,
647 but no strong inference can be made along these lines given that we could neither predict, nor
648 retroactively account for, which cells fell into this category. In other cases, one or two of the learned
649 SC-BC interaction functions in each group of 5 overlapped heavily with their corresponding sLLR curves,
650 whereas the other curves in the group were driven apart by the learning rule to cover a much wider
651 spread (vertically or horizontally or both) than the original set of sLLRs. In still other cases, the learned
652 interaction functions were nearly "opposite" to their corresponding sLLR functions (e.g. red curves in
653 columns 2 and 4 of the second row in Figure 8B). In these cases, also, we were unable to explain why
654 the cells' learned interaction functions deviated so extensively from the predictions of the simplified
655 Bayesian model. In the hopes of at least confirming that the deviations were caused by violations in the
656 CCI assumption, we conducted a simple experiment, described next, in which correlations between
657 pairs of simple cell inputs to a boundary cell were systematically manipulated.
658

659 **Probing the relationship between the incitation circuit and Bayes rule**

660 To probe the role of correlations between simple cells in shaping SC-BC interaction functions in a
661 boundary-detecting circuit, we ran a simple experiment in which a boundary cell received input from
662 just two simple cells whose RFs overlapped to varying degrees. For each simple cell pair, we fit the
663 parameters of the incitation circuit either separately (Figure 9A, left) or jointly (Figure 9A, right). We
664 tested pairs of filters ranging from very dependent (Figure 9B, middle columns) to nearly independent
665 (Figure 9B; outer columns). Scatter plots of joint filter responses to boundary (red) and non-boundary
666 (black) patches are shown below each pair. When the SC-BC interaction functions were learned
667 separately, they were nearly identical to the literal LLRs (Figure 9B, first row of blue and orange curves;
668 solid curves show learned interactions, dashed curves show LLRs). On the other hand, when the SC-BC
669 interactions were learned jointly, for SC pairs with heavily overlapping RFs, which led to a breakdown
670 of the CCI assumption, the learned interactions differed significantly from the pure LLRs (Figure 9B,
671 middle columns). Consistent with these observations, we show analytically in the Appendix that an
672 incitation circuit like the one shown in Figure 8A will learn LLRs if the input features are CCI.
673 Consequently, any observable differences between the learned incitation functions and the LLRs must
674 be attributable to a breakdown of class-conditional independence.

675

676 **Comparing boundary detection performance of four models**

677 The optimization of synaptic weights in the incitation circuit of Figure 8A opens up an additional
678 avenue for validation (or refutation) of our overarching boundary cell hypothesis. Our main premise is
679 that a boundary cell in V1 should be able to significantly improve its boundary detection performance
680 (compared to a single simple cell at the reference location) if it can access, through the local cortical
681 circuit, a large and diverse set of simple cells covering the neighborhood. A major additional claim is

682 that the incitation circuit motif, which is known to exist in V1, is well suited to deliver such a
683 performance improvement. To directly test these claims, we compared the Precision-Recall
684 performance of the trained incitation network (Figure 10, red curve) to 3 other boundary detectors: (1)
685 the "null hypothesis", consisting of a single conventional simple cell centered at the reference location
686 (blue curve); (2) an unweighted sum of 300 literal LLRs (orange curve); this is essentially a direct
687 implementation of Bayes rule under the CCI assumption (see Figure 1D); and (3) a *weighted* sum of the
688 same 300 literal LLRs (green curve); this hybrid model honors the basic structure of the Bayesian
689 classifier of Figure 1D, but allows weights on each of the LLR inputs to help compensate for CCI
690 violations in the simple cell population.
691
692 The results shown in Figure 10 support 3 conclusions: (1) the superior performance of all 3 multi-input
693 classifier variants compared to a single conventional simple cell reinforces the point that individual
694 simple cells are seriously underpowered as natural boundary detectors (see Ramachandra and Mel
695 2013); (2) the superior performance of the 2 classifier variants with learned synaptic weights (red and
696 green curves) compared to the simplified Bayesian classifier that receives unweighted LLR inputs
697 (orange curve) attests to the importance of a learning rule that is sensitive to natural image statistics
698 and can help compensate for unwanted input correlations; and (3) the very similar performance of the
699 optimized incitation network (red curve) and the weighted sum of LLRs (green curve), especially in the
700 high recall range, is again suggestive of a non-trivial connection between the simplified Bayesian
701 classifier of Figure 1A and the behavior of the learned boundary detecting incitation circuit of Figure
702 8A.
703

We conclude by noting that the requirements for developing a cortical circuit that produces significantly improved boundary detection performance compared to a conventional simple cell are relatively modest, including (1) a compound E-I circuit motif that we have dubbed an “incitation” circuit, which is known to exist in V1; (2) variability in firing thresholds across the population of simple cells; and (3) a supervised “delta rule” capable of setting the strengths (and/or dendritic locations) of the excitatory contacts from simple cells onto boundary cells and their associated interneurons. Possible sources of the supervisory signal are taken up in the Discussion.

714 **DISCUSSION**

715 In the 60 years since Hubel and Wiesel first discovered orientation-tuned simple cells in V1, it has been
716 generally assumed that these cells contribute in some way to the detection of object boundaries
717 (Angelucci et al., 2002; Field et al., 1993; Grosof et al., 1993; Kapadia et al., 1995, 2000; Polat et al.,
718 1998; Sceniak et al., 1999). Consistent with this idea, virtually every modern object recognition system,
719 whether designed by hand or trained from natural image data, includes simple cell-like filtering in its
720 early stages of processing (Fukushima et al., 1983; Krizhevsky et al., 2012; Lades et al., 1993; Lecun et
721 al., 1998; Mel, 1997; Riesenhuber and Poggio, 1999). Surprisingly, however, the quantitative
722 relationship between simple cell responses, typically modeled as divisively normalized linear filters
723 (Carandini and Heeger, 2012), and object boundary probability in natural images, has been little
724 explored (though see Ramachandra and Mel, 2013), making it difficult to know whether or how V1
725 circuits contribute to this behaviorally relevant natural computation. It is important to emphasize that
726 a simple cell on its own is a poor detector of natural object boundaries within its receptive field (see
727 also Arbelaez et al., 2011): as shown in Figure 10 (blue curve), if we use a simple cell's response as an
728 indicator of the presence of an object boundary within its RF, even when the threshold for detection is
729 raised to such a high value that half of all true boundaries are rejected (corresponding to a Recall score
730 of 50%), almost two thirds of the "detected" edges at that threshold are false positives (corresponding
731 to a Precision score of ~35%). The reason a simple cell is such an unreliable edge detector is that true
732 object boundaries are rare (the overwhelming majority of points in Figure 12A are piled in the lower
733 half of the plot), and when they do occur, they are very often of low contrast. Much more common are
734 high contrast non-edge structures (e.g. textures) that contain sufficient oriented energy to strongly
735 drive simple oriented filters.

736

737 The poor boundary detection performance of a lone simple cell leads to the conjecture that V1 also
738 contains "smarter" cells that compute boundary probability by combining the responses of multiple
739 simple cells covering a local neighborhood. In a previous study, we suggested that the appropriate
740 strategy for constructing a boundary cell from a local population of simple cells was to (1) select a small
741 set of simple cells (e.g. 6 cells) that were both individually informative and class-conditionally
742 independent (see Methods for discussion of the CCI assumption); (2) evaluate the log-likelihood ratios
743 for each of the participating simple cells, which would tell us the optimal functional interconnections
744 between each simple cell and the boundary cell (according to Bayes rule); and (3) sum the LLRs and
745 apply a fixed sigmoidal nonlinearity to compute boundary probability (Ramachandra and Mel, 2013)
746 (Figure 1C,D). The present study extends that previous work in eight ways: (1) we collected and
747 analyzed individual LLRs for *all* of the simple cells at all orientations covering a 5x5 pixel neighborhood
748 in the vicinity of a boundary cell's RF (300 cells total); (2) we show that the idealized functional
749 interconnections between simple cells and boundary cells depend systematically on the relative
750 positions and orientations of the simple cell and boundary cell RFs (Figure 3) – but are relatively
751 insensitive to the scale or aspect ratio of the simple cell receptive fields (Figure 4); (3) we developed a
752 simple analytical model (i.e. gaussian likelihoods->quadratic LLRs) that shows how the three seemingly
753 different types of SC-BC interaction functions – rising, falling, and bump-shaped functions – represent
754 different ranges of the same underlying (quadratic) function class (Figure 5); (4) we show that a mixed
755 excitatory-inhibitory, or "incitatory", circuit motif that is known to exist in V1 is capable of producing
756 the entire spectrum of natural image-derived SC-BC interaction functions (Figures 6,7); (5) we show
757 that the parameters of a boundary-detecting incitation circuit can be learned by adjusting a single layer
758 of excitatory weights (Figure 8A); (7) we show that a learned incitation circuit can improve the
759 precision of boundary detection in the high-recall range by 43% to 121% compared to a conventional

760 simple cell model (Figure 10); and (8) by “reading out” the weights of the learned incitation circuit, we
761 show that the simple cell-boundary cell interaction functions that we would expect to find in the visual
762 cortex are not likely to be verbatim LLRs, but rather, perturbed versions due to class-conditional
763 dependencies among simple cells whose receptive fields overlap heavily with each other (Figures 8B,
764 9). This could be helpful in interpreting the results of future neurophysiological experiments in V1.
765

766 **Experimental predictions**

767 *Distinguishing boundary cells from conventional simple cells*

768 Having shown that cortical circuitry is capable in principle of producing boundary cells from simple cells
769 using only a single layer of modifiable excitatory weights, it is important to ask how BCs could be
770 detected experimentally, and distinguished from conventional simple cells (or the simple cell-like
771 subunits of complex cells – Hubel and Wiesel, 1962; Movshon et al., 1978; Ohzawa et al., 1997).
772

773 To determine how BCs would respond to various stimuli, stimulus patches were scaled to have the
774 same fixed value of the normalizer used in earlier figures, once again reflecting a simple form of
775 divisive normalization (see Methods section *Boundary cell stimulus responses*). We first constructed a
776 canonical stimulus for a boundary cell akin to a spike triggered average by averaging all image patches
777 weighted by their evoked boundary cell response. As expected, the STA stimulus appears as a localized,
778 polarized, oriented boundary segment reminiscent of a simple cell’s receptive field (Figure 11A). We
779 then presented drifting sine wave gratings covering a boundary cell’s "classical receptive field", leading
780 to the unremarkable phase response and orientation tuning curves shown in Figure 11B, C. Next, we
781 used labeled natural edges with the same normalizer value to explore the effect of increasing center

782 contrast on orientation tuning curve width. (This was not a perfectly controlled experiment because
783 variations in center contrast at a fixed normalizer value would have led to *anti*-variations in surround
784 contrast, but given the filter value at the RF center was only one of 100 filters of many orientations
785 used to compute the normalizer value, this effect was likely small). Subject to this limitation, as shown
786 in Figure 11B, the boundary cell's tuning width is essentially constant across a roughly 2-fold change in
787 center contrast – the limit of analysis allowed by our labeled database (average tuning curve has full
788 width at half height for high contrast stimuli 43.6°; for low contrast stimuli, 39.2°).
789
790 Thus, for oriented edges and gratings presented within the CRF, boundary cells behave similarly to
791 conventional simple cells in that they (1) have phase-dependent responses; (2) are orientation tuned;
792 and (3) have tuning curves whose widths are roughly contrast invariant (Alitto and Usrey, 2004). It is
793 therefore possible that boundary cells exist and have been classified as conventional simple cells in
794 previous experiments using simplified stimuli. Among the multiple types of V1 cells that have been
795 previously described, boundary cells share most in common with double opponent cells, which are
796 orientation tuned, have mostly odd-symmetric receptive field profiles as would be expected for
797 boundary detecting cells (Ringach, 2002), and respond to boundaries whether defined by luminance or
798 color (Johnson et al., 2008).
799
800 In future neurophysiological studies, an efficient means of dissociating conventional simple cells, which
801 respond to oriented contrast independent of boundary probability, from putative boundary cells,
802 which respond to boundary probability independent of oriented contrast, would be to use natural
803 image stimuli drawn from the four corners of the oriented contrast – boundary probability space
804 (Figure 12A). Image patches with low oriented contrast and low boundary probability scores (purple

805 dots) tend to contain flat, unstructured image regions; patches with low contrast and high probability
806 (green dots) tend to contain well-structured, faint edges; patches with high contrast but low
807 probability (blue dots) tend to contain contrasty noise or misaligned edges; and regions with high
808 contrast *and* high probability (red dots) are typically well-structured, strong edges (Figure 12B). This
809 factorial stimulus set would make it possible to identify pure simple cells, pure boundary cells, as well
810 as cells of intermediate type.

811

812 *Diverse inputs to the dendrites of boundary cells?*

813 One of our main findings is that a cell in visual cortex whose job is to detect object boundaries can
814 improve its detection performance if it collects input from many simple cells in its vicinity with a
815 diversity of receptive field positions and orientations. Consistent with this, several recent 2-photon
816 calcium imaging studies have mapped the receptive field properties of individual dendritic spines on V1
817 neurons in mice, ferrets, and monkeys, and have shown that the inputs to a single V1 cell (and often a
818 single dendrite) are quite variable in terms of their receptive field properties, covering a much wider
819 range of preferred orientations and RF positions than might be expected given the target cell's more
820 sharply tuned response preferences (Jia et al, 2010; Wilson et al. 2016; Iacaruso et al. 2016; Scholl et al.
821 2017; Ju et al. 2020). These findings do not prove that many or most cells in V1 are boundary cells,
822 only that most V1 cells appear to receive the requisite diversity of inputs from neighboring cells. What
823 remains to be shown is that the excitatory and inhibitory inputs from surrounding cells are properly
824 weighted and balanced by the local incitation circuit, so as to maximize boundary detection
825 performance. The conceptual experiment described next could help to establish whether the local
826 cortical circuit actually functions in this way.

827

828 *A predictable spectrum of SC-BC interactions?*

829 A key feature of the boundary cell hypothesis is that SC-BC interaction functions take on predictable
830 forms, depending primarily on the offsets in position and orientation between the SC and BC receptive
831 fields (Figure 8b). It may be possible to empirically measure those interaction functions by applying
832 stimuli that distinguish simple cells from boundary cells (Figure 12) while imaging V1 neurons in awake
833 animals (e.g. Tang et al. 2018; Ju et al. 2020). The approach would require the ability (1) to identify
834 pairs of nearby simple and boundary cells and to characterize their RFs, and (2) to transiently activate
835 or inactivate identified cells optogenetically. Then, while presenting carefully curated natural image
836 patches to the boundary cell – that also overlap with the simple cell's RF, the simple cell's activity level
837 could be perturbed, and the response changes in the boundary cell measured. The direction and
838 magnitude of the change in the boundary cell's activity could be compared to the prediction of a
839 trained incitation network (Figure 8B). For a particular simple cell, certain image patches will drive the
840 cell to a level on the rising slope of its SC-BC interaction function, so that a boost in the cell's activity
841 (through optogonetic stimulation) should in turn boost the boundary cell's activity – and conversely for
842 suppression of the simple cell's activity. For other image patches, the simple cell will be firing at or
843 beyond the mode of its SC-BC interaction function with respect to a particular boundary cell, so that a
844 boost in that simple cell's activity will lead to a suppression of the boundary cell's activity (and
845 conversely if the simple cell's activity is optogenetically suppressed).

846

It is worth noting that we cannot assume that the SC-BC interaction functions measured in this way will look exactly like those produced by a particular trained incitation network, since even if the overall idea holds true, the interaction functions depend on the complete set of simple cells providing input to a particular boundary cell, which cannot be known. Nonetheless, by repeating these types of response manipulations for a large number of SC-BC pairs, we can hope to find a basic correspondence between predicted and measured SC-BC interactions, in the sense that the measured interactions should include pure increasing cases, pure decreasing cases and non-monotonic cases, with a systematic dependence on the spatial and orientation offsets of the simple and boundary cells' RFs (Figures 6-8).

It is also worth noting that boundary cells need not reside in, or only in, V1. Nothing precludes that cells that signal boundary probability, rather than boundary contrast, may be found in higher visual areas.

Relationship to previous work on natural image statistics

A number of previous studies have attempted to explain receptive field properties of cells in the retina, LGN and primary visual cortex in terms of natural image statistics and principles such as efficient coding, sparse coding, and independent components analysis (Barlow, 1981; Bell and Sejnowski, 1995; Laughlin, 1989; Olshausen and Field, 1996; Schwartz and Simoncelli, 2001; Zhu and Rozell, 2013). These studies have been mainly concerned with neural *representation*, where the goal is fast/accurate information transmission through a noisy channel, and eventually faithful image reconstruction. In

contrast, our work is primarily concerned with neural *computation*, where the goal is to transform the image into a more abstract shape representation that is more directly useful for guiding behavior.

From a different perspective and with a different goal, Geisler et al. (2001) collected co-occurrence statistics of pre-detected local boundary elements in natural scenes, with the aim to predict human contour grouping performance. Their measurements on natural images included the probability of finding a second boundary element in the vicinity of a first boundary element, depending on the relative offsets in position and orientation of the two elements, or whether two spatially offset boundary elements were more likely to belong to the same or different object. Sigman et al. (2001) also studied co-occurrence statistics of pre-detected boundary elements, coming to the conclusion that boundary elements in natural scenes tend to lie on common circles. The goal to characterize the spatial distribution of pre-detected boundary elements in natural scenes in both of these studies contrasts with our focus here on the detection problem, that is, the problem of discriminating object boundaries from non-boundaries based on populations of simple cell responses collected from a local neighborhood in an image. Furthermore, all of the grouping statistics collected by Geisler et al. and Sigman et al. were represented as scalar values linking pairs of locations/orientations. In contrast, our natural image analysis produces *functions* linking pairs of locations/orientations, which capture how a given simple cell should influence a nearby boundary cell as a part of a boundary detection computation. Also unlike these previous studies, we use our data to constrain and to benchmark cortical circuit models.

Non-monotonic cell-cell interactions have been previously reported

One of our findings is that among the different types of local cell-cell interactions needed for object boundary detection in natural images, many cannot be described as "excitatory" or "inhibitory", nor can they be represented by positive or negative synaptic weights, but are instead U-shaped functions wherein cell 1 might excite cell 2 at low firing rates, reach its peak excitatory effect at intermediate firing rates, and inhibit cell 2 at high firing rates. U-shaped functions of the opposite polarity can also occur (Figure 8B). Should we find the idea surprising that nearby cells in the cortex act on each other non-monotonically?

From one perspective, one might argue that whenever there are excitatory and inhibitory cells wired together in a circuit motif, perhaps we should be surprised if we did *not* find non-monotonic interactions between cells. For example, in the "inhibition-stabilized network" model (Jadi and Sejnowski, 2014; Ozeki et al., 2009), which accounts for a number of V1 cell response properties, "non-binary" interactions between cells would almost certainly be expected to occur. Nevertheless, there has been a historical tendency to think about cell-cell interactions in the cortex as being of a defined polarity, represented by a positive or negative scalar value, and often subject to simple geometric rules. The notion of "surround suppression", for example, reflects both of these tendencies (Adesnik et al., 2012; Cavanaugh et al., 2002; Schwabe et al., 2010). Even as the geometric constraints governing cell-cell interactions become more intricate, such as where interconnection strength and polarity depend on distance or relative orientation, the simplification that cell-cell interactions have a defined polarity is often still relied upon. For example, K.D. Miller's models of map development include short range excitation and medium-range inhibition (Miller, 1994); Angelucci and Bressler's models include near and far suppressive surrounds (Angelucci and Bressloff, 2006); and several studies support the idea that cortical cells affect each other laterally through bowtie-shaped "extension fields" consisting

911 of patterned arrays of positive and negative coefficients (e.g. (Bosking et al., 1997; Field et al., 1993;
912 Geisler et al., 2001; Kapadia et al., 2000; Li, 1999; Sigman et al., 2001). In all of these cases, one
913 neuron's effect on another neuron is described in terms of its scalar connection "strength".
914
915 Not all functional interconnections that have been described in the cortex fit such descriptions,
916 however. Examples of activity-level-dependent interactions have been reported, where the strength
917 and even polarity of the connection between cells depends on the activity levels of the sending and/or
918 receiving cells. For example, the responses of amplitude-tuned neurons in the auditory cortex grow
919 stronger as the sound pressure level increases up to an optimal intensity level, and then are
920 progressively inhibited as the sound grows louder (Suga and Manabe, 1982); in V1, surround
921 modulation can switch from facilitating to suppressive with increasing center contrast (Ichida et al.,
922 2007; Nauhaus et al., 2009; Polat et al., 1998; Schwabe et al., 2006; Somers et al., 1998); length-tuned
923 neurons respond best to an oriented stimulus up to a certain length, but are then progressively
924 inhibited as the stimulus grows longer (Anderson et al., 2001); and non-monotonic modulatory
925 interactions between a neuron’s classical and extra-classical receptive fields have been reported (Polat
926 et al., 1998). These data, though unaccompanied by normative explanations, do support the idea that
927 the sign and magnitude one neuron's effect on another can depend not only on the relative position
928 and orientation of their receptive fields (in the case of vision), but also on their relative activity levels.
929
930 Our paper represents a fleshing out of this type of effect, and is to our knowledge the first normative
931 theory, parameterized by natural images, that specifies how intracolumnar cell-cell interactions may
932 help solve a specific, biologically relevant *classification* problem. By analyzing natural image data on
933 and off object boundaries, we showed that the local cell-cell interactions needed to solve this

classification problem are not capturable by scalar weights, but are in general nonlinear functions that depend on "all of the above" – relative location, relative orientation, and relative activity levels of the sending and receiving cells. We further showed that the SC-BC functional connections needed for boundary detection are easily produced by a compound E-I circuit motif (see Figure 6) that is known to exist in the cortex (Buzsáki, 1984; Isaacson and Scanziani, 2011; Klyachko and Stevens, 2006; McBain and Fisahn, 2001; Pfeffer et al., 2013; Pouille and Scanziani, 2001; Swadlow, 2002; Wehr and Zador, 2003). Finally, we showed that the synaptic weights that control the net effect of an "incitation" motif are easily learned. Future experiments will be needed to establish whether trainable incitation circuits are actually used to help solve the difficult natural classification problems faced by neurons in V1 and other areas of the cortex.

How could a properly parameterized incitation circuit develop?

A possible extension of this work would be to address the limitation that the incitation circuit we show in Figure 8A was trained by a supervised learning rule (logistic regression), but without our providing a biologically based account for the source of the supervision. The original purpose of the exercise was to test whether an incitation circuit with a single layer of modifiable excitatory weights is *capable* of performing object boundary detection at a level comparable to an explicit Bayesian classifier. We found this to be true (Figure 10), suggesting that this particular Bayesian-inspired algorithm lies within the computational scope of cortical tissue. The demonstration leaves open the question, however, as to where a supervisory signal might come from during visual development that alerts a boundary cell and its inhibitory partner when an object boundary is crossing through its receptive field. One possible source of supervision would be a population of neurons located within the same or different area that have access to different visual cues, such as cells sensitive to motion-defined boundaries. Such cells

are found at many levels of the visual system, including the retinas of salamanders (Olveczky et al., 2003); V1, V2, V3 MT and IT in the monkey (Marcar et al., 1995, 2000; Sary et al., 1995; Zeki et al., 2003); and in multiple areas of the human visual cortex (Dupont, 1997; Larsson et al., 2010; Mysore et al., 2006; Zeki et al., 2003). Topographic feedback projections from motion boundary-sensitive cells in these areas to V1 (or locally within V1) could help to instruct boundary cells in V1 so that they may perform based purely on pictorial cues (i.e. when motion signals are unavailable).

Limitations of the model

The boundary detection computation that we have studied was inspired by Bayes rule, and is underlyingly a feedforward computation whose core operation is a sum of LLR terms (Figure 1C). Our attempt to map this computation onto a simple, cortically plausible circuit is shown in Figure 8A, in which a layer of simple cells with varying output nonlinearities activates both (1) a "layer" of boundary cells (though only one BC is shown); and (2) a layer of inhibitory cells, one per BC (though only one inhibitory cell is shown – the one assigned to the one shown boundary cell). Each inhibitory cell, in turn, acts on its associated boundary cells through a fixed connection. Given that the circuit of Figure 8A is purely feedforward, ignoring (1) local or long-range feedback connections that are known to exist in the neocortex (Angelucci et al., 2017); (2) nonlinear dendritic integration effects that could also contribute to boundary detection (Jin et al., 2022); and (3) all dynamics at the synapse, cell, and circuit levels, it falls far short of a fully elaborated cortical circuit model. Rather, the circuit model of Figure 8A should be viewed as a demonstration that a known cortical circuit motif – the incitation motif – is capable of producing cells that superficially resemble simple cells, but are much better at detecting object boundaries in natural scenes than the standard simple cell model (Heeger 1992). A worthy long-range goal would be to fold the boundary-detection capability of a properly parameterized incitation

circuit into a more comprehensive cortical circuit model that addresses a wider range of physiological phenomena (Ozeki et al. 2009; Zhu and Rozell 2013).

Abbreviations

SC: simple cell
BC: boundary cell
RF: receptive field
RL: reference location
LLR: log-likelihood ratio
sLLR: shifted log-likelihood ratio
CCI: class conditional independence
PR: precision-recall

References

Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z.J., and Scanziani, M. (2012). A neural circuit for spatial summation in visual cortex. Nature *490*, 226–231. https://doi.org/10.1038/nature11526.

Alitto, H.J., and Usrey, W.M. (2004). Influence of contrast on orientation and temporal frequency tuning in ferret primary visual cortex. J. Neurophysiol. *91*, 2797–2808. https://doi.org/10.1152/jn.00943.2003.

Anderson, J.S., Lampl, I., Gillespie, D.C., and Ferster, D. (2001). Membrane Potential and Conductance Changes Underlying Length Tuning of Cells in Cat Primary Visual Cortex. J. Neurosci. *21*, 2104–2112. .

Angelucci, A., and Bressloff, P.C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. In Progress in Brain Research, (Elsevier), pp. 93–120.

Angelucci, A., Levitt, J.B., Walton, E.J.S., Hupé, J.-M., Bullier, J., and Lund, J.S. (2002). Circuits for Local and Global Signal Integration in Primary Visual Cortex. J. Neurosci. *22*, 8633–8646. .

Angelucci, A., Bijanzadeh, M., Nurminen, L., Federer, F., Merlin, S., and Bressloff, P.C. (2017). Circuits and Mechanisms for Surround Modulation in Visual Cortex. Annual Review of Neuroscience *40*, 425–451. https://doi.org/10.1146/annurev-neuro-072116-031418.

Arbelaez, P., Maire, M., Fowlkes, C., and Malik, J. (2011). Contour Detection and Hierarchical Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence *33*, 898–916. https://doi.org/10.1109/TPAMI.2010.161.

Barlow, H.B. (1981). The Ferrier Lecture, 1980: Critical Limiting Factors in the Design of the Eye and Visual Cortex. Proc. R. Soc. Lond. B *212*, 1–34. https://doi.org/10.1098/rspb.1981.0022.

Bell, A.J., and Sejnowski, T.J. (1995). An information-maximization approach to blind separation and blind deconvolution. Neural Comput *7*, 1129–1159. .

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. Psychological Review *94*, 115–147. .

Bishop, P.O., Coombs, J.S., and Henry, G.H. (1973). Receptive fields of simple cells in the cat striate cortex. The Journal of Physiology *231*, 31–60. https://doi.org/10.1113/jphysiol.1973.sp010218.

Bosking, W.H., Zhang, Y., Schofield, B., and Fitzpatrick, D. (1997). Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex. J. Neurosci. *17*, 2112–2127. .

Buzsáki, G. (1984). Feed-forward inhibition in the hippocampal formation. Progress in Neurobiology *22*, 131–153. https://doi.org/10.1016/0301-0082(84)90023-6.

Carandini, M., and Heeger, D.J. (2012). Normalization as a canonical neural computation. Nature Reviews Neuroscience *13*, 51–62. https://doi.org/10.1038/nrn3136.

Cavanaugh, J.R., Bair, W., and Movshon, J.A. (2002). Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. J. Neurophysiol. *88*, 2547–2556. https://doi.org/10.1152/jn.00693.2001.

DeAngelis, G.C., Robson, J.G., Ohzawa, I., and Freeman, R.D. (1992). Organization of suppression in receptive fields of neurons in cat visual cortex. Journal of Neurophysiology *68*, 144–163. .

Dupont, P. (1997). The kinetic occipital region in human visual cortex. Cerebral Cortex *7*, 283–292. https://doi.org/10.1093/cercor/7.3.283.

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

Fukushima, K., Miyake, S., and Ito, T. (1983). Neocognitron: A neural network model for a mechanism of visual pattern recognition. IEEE Transactions on Systems, Man and Cybernetics *SMC-13*, 826–834. https://doi.org/10.1109/TSMC.1983.6313076.

Geisler, W.S., and Albrecht, D.G. (1992). Cortical neurons: Isolation of contrast gain control. Vision Research *32*, 1409–1410. https://doi.org/10.1016/0042-6989(92)90196-P.

Geisler, W.S., Perry, J.S., Super, B.J., and Gallogly, D.P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. Vision Res. *41*, 711–724. .

George, A.A., Lyons-Warren, A.M., Ma, X., and Carlson, B.A. (2011). A diversity of synaptic filters are created by temporal summation of excitation and inhibition. J Neurosci *31*, 14721–14734. https://doi.org/10.1523/JNEUROSCI.1424-11.2011.

Gilbert, C.D., and Wiesel, T.N. (1990). The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. Vision Research *30*, 1689–1701. https://doi.org/10.1016/0042-6989(90)90153-C.

Grosof, D.H., Shapley, R.M., and Hawken, M.J. (1993). Macaque V1 neurons can signal "illusory" contours. Nature *365*, 550–552. https://doi.org/10.1038/365550a0.

Heydt, R. von der, and Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. J. Neurosci. *9*, 1731–1748. .

Hoffman, D.D. (2000). Visual intelligence: How we create what we see (WW Norton & Company).

Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. (Lond.) *160*, 106–154. .

Ichida, J.M., Schwabe, L., Bressloff, P.C., and Angelucci, A. (2007). Response Facilitation From the "Suppressive" Receptive Field Surround of Macaque V1 Neurons. Journal of Neurophysiology *98*, 2168–2181. https://doi.org/10.1152/jn.00298.2007.

Isaacson, J.S., and Scanziani, M. (2011). How inhibition shapes cortical activity. Neuron *72*, 231–243. https://doi.org/10.1016/j.neuron.2011.09.027.

Jadi, M.P., and Sejnowski, T.J. (2014). Regulating Cortical Oscillations in an Inhibition-Stabilized Network. Proceedings of the IEEE *102*, 830–842. https://doi.org/10.1109/JPROC.2014.2313113.

Jin, L., Behabadi, B.F., Jadi, M.P., Ramachandra, C.A., and Mel, B.W. (2022). Classical-contextual interactions in V1 may rely on dendritic computations. Neuroscience S0306452222001038. https://doi.org/10.1016/j.neuroscience.2022.02.033.

Johnson, E.N., Hawken, M.J., and Shapley, R. (2008). The Orientation Selectivity of Color-Responsive Neurons in Macaque V1. J. Neurosci. *28*, 8096–8106. https://doi.org/10.1523/JNEUROSCI.1404-08.2008.

Kapadia, M.K., Ito, M., Gilbert, C.D., and Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. Neuron *15*, 843–856. https://doi.org/10.1016/0896-6273(95)90175-2.

Kapadia, M.K., Westheimer, G., and Gilbert, C.D. (2000). Spatial Distribution of Contextual Interactions in Primary Visual Cortex and in Visual Perception. Journal of Neurophysiology *84*, 2048–2062. .

Klyachko, V.A., and Stevens, C.F. (2006). Excitatory and Feed-Forward Inhibitory Hippocampal Synapses Work Synergistically as an Adaptive Filter of Natural Spike Trains. PLOS Biol *4*, e207. https://doi.org/10.1371/journal.pbio.0040207.

Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems 25, F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, eds. (Curran Associates, Inc.), pp. 1097–1105.

Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., Malsburg, C. von der, Wurtz, R.P., and Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers *42*, 300–311. https://doi.org/10.1109/12.210173.

Larsson, J., Heeger, D.J., and Landy, M.S. (2010). Orientation Selectivity of Motion-Boundary Responses in Human Visual Cortex. Journal of Neurophysiology *104*, 2940–2950. https://doi.org/10.1152/jn.00400.2010.

Laughlin, S.B. (1989). The role of sensory adaptation in the retina. J Exp Biol *146*, 39–62. .

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE *86*, 2278–2324. https://doi.org/10.1109/5.726791.

Li, Z. (1999). Visual segmentation by contextual influences via intra-cortical interactions in the primary visual cortex. Network *10*, 187–212. .

Marcar, V.L., Xiao, D.K., Raiguel, S.E., Maes, H., and Orban, G.A. (1995). Processing of kinetically defined boundaries in the cortical motion area MT of the macaque monkey. Journal of Neurophysiology *74*, 1258–1270. https://doi.org/10.1152/jn.1995.74.3.1258.

Marcar, V.L., Raiguel, S.E., Xiao, D., and Orban, G.A. (2000). Processing of Kinetically Defined Boundaries in Areas V1 and V2 of the Macaque Monkey. Journal of Neurophysiology *84*, 2786–2798. https://doi.org/10.1152/jn.2000.84.6.2786.

McBain, C.J., and Fisahn, A. (2001). Interneurons unbound. Nat. Rev. Neurosci. *2*, 11–23. https://doi.org/10.1038/35049047.

Mel, B.W. (1997). SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. Neural Comput *9*, 777–804. .

Miller, K.D. (1994). A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between ON-and OFF-center inputs. Journal of Neuroscience *14*, 409–409. .

Movshon, J.A., Thompson, I.D., and Tolhurst, D.J. (1978). Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. J. Physiol. (Lond.) *283*, 101–120. .

Murphy, K.P. (2012). Machine Learning.

Mysore, S.G., Vogels, R., Raiguel, S.E., and Orban, G.A. (2006). Processing of Kinetic Boundaries in Macaque V4. Journal of Neurophysiology *95*, 1864–1880. https://doi.org/10.1152/jn.00627.2005.

Nauhaus, I., Busse, L., Carandini, M., and Ringach, D.L. (2009). Stimulus contrast modulates functional connectivity in visual cortex. Nat Neurosci *12*, 70–76. https://doi.org/10.1038/nn.2232.

Ohzawa, I., DeAngelis, G.C., and Freeman, R.D. (1997). Encoding of binocular disparity by complex cells in the cat's visual cortex. J. Neurophysiol. *77*, 2879–2909. .

Olshausen, B.A., and Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature *381*, 607. https://doi.org/10.1038/381607a0.

Olveczky, B.P., Baccus, S.A., and Meister, M. (2003). Segregation of object and background motion in the retina. Nature *423*, 401–408. https://doi.org/10.1038/nature01652.

Ozeki, H., Finn, I.M., Schaffer, E.S., Miller, K.D., and Ferster, D. (2009). Inhibitory Stabilization of the Cortical Network Underlies Visual Surround Suppression. Neuron *62*, 578–592. https://doi.org/10.1016/j.neuron.2009.03.028.

Pfeffer, C.K., Xue, M., He, M., Huang, Z.J., and Scanziani, M. (2013). Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. Nature Neuroscience *16*, 1068–1076. https://doi.org/10.1038/nn.3446.

Polat, U., Mizobe, K., Pettet, M.W., Kasamatsu, T., and Norcia, A.M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. Nature *391*, 580–584. https://doi.org/10.1038/35372.

Pouille, F., and Scanziani, M. (2001). Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition. Science *293*, 1159–1163. https://doi.org/10.1126/science.1060342.

Priebe, N.J., and Ferster, D. (2006). Mechanisms underlying cross-orientation suppression in cat visual cortex. Nat Neurosci *9*, 552–561. https://doi.org/10.1038/nn1660.

Ramachandra, C.A., and Mel, B.W. (2013). Computing local edge probability in natural scenes from a population of oriented simple cells. J Vis *13*. https://doi.org/10.1167/13.14.19.

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nat. Neurosci. *2*, 1019–1025. https://doi.org/10.1038/14819.

Ringach, D.L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. J. Neurophysiol. *88*, 455–463. https://doi.org/10.1152/jn.2002.88.1.455.

Rust, N.C., and Dicarlo, J.J. (2010). Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. J. Neurosci. *30*, 12978–12995. https://doi.org/10.1523/JNEUROSCI.0179-10.2010.

Sanguinetti, G., Citti, G., and Sarti, A. (2010). A model of natural image edge co-occurrence in the rototranslation group. J Vis *10*. https://doi.org/10.1167/10.14.37.

Sary, G., Vogels, R., Kovacs, G., and Orban, G.A. (1995). Responses of monkey inferior temporal neurons to luminance-, motion-, and texture-defined gratings. Journal of Neurophysiology *73*, 1341–1354. https://doi.org/10.1152/jn.1995.73.4.1341.

Sceniak, M.P., Ringach, D.L., Hawken, M.J., and Shapley, R. (1999). Contrast's effect on spatial summation by macaque V1 neurons. Nat Neurosci *2*, 733–739. https://doi.org/10.1038/11197.

Schwabe, L., Obermayer, K., Angelucci, A., and Bressloff, P.C. (2006). The Role of Feedback in Shaping the Extra-Classical Receptive Field of Cortical Neurons: A Recurrent Network Model. J. Neurosci. *26*, 9117–9129. https://doi.org/10.1523/JNEUROSCI.1253-06.2006.

Schwabe, L., Ichida, J.M., Shushruth, S., Mangapathy, P., and Angelucci, A. (2010). Contrast-dependence of surround suppression in Macaque V1: Experimental testing of a recurrent network model. NeuroImage *52*, 777–792. https://doi.org/10.1016/j.neuroimage.2010.01.032.

Schwartz, O., and Simoncelli, E.P. (2001). Natural signal statistics and sensory gain control. Nature Neuroscience *4*, 819. https://doi.org/10.1038/90526.

Sigman, M., Cecchi, G.A., Gilbert, C.D., and Magnasco, M.O. (2001). On a common circle: Natural scenes and Gestalt rules. PNAS *98*, 1935–1940. https://doi.org/10.1073/pnas.98.4.1935.

Somers, D.C., Todorov, E.V., Siapas, A.G., Toth, L.J., Kim, D.S., and Sur, M. (1998). A local circuit approach to understanding integration of long-range inputs in primary visual cortex. Cereb Cortex *8*, 204–217. https://doi.org/10.1093/cercor/8.3.204.

Suga, N., and Manabe, T. (1982). Neural basis of amplitude-spectrum representation in auditory cortex of the mustached bat. J. Neurophysiol. *47*, 225–255. .

Swadlow, H.A. (2002). Thalamocortical control of feed–forward inhibition in awake somatosensory 'barrel' cortex. Philosophical Transactions of the Royal Society B: Biological Sciences *357*, 1717–1727. https://doi.org/10.1098/rstb.2002.1156.

Theys, T., Romero, M.C., van Loon, J., and Janssen, P. (2015). Shape representations in the primate dorsal visual stream. Front. Comput. Neurosci. *9*. https://doi.org/10.3389/fncom.2015.00043.

Wehr, M., and Zador, A.M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature *426*, 442–446. https://doi.org/10.1038/nature02116.

Zeki, S., Perry, R.J., and Bartels, A. (2003). The Processing of Kinetic Contours in the Brain. Cereb Cortex *13*, 189–202. https://doi.org/10.1093/cercor/13.2.189.

Zhu, M., and Rozell, C.J. (2013). Visual Nonclassical Receptive Field Effects Emerge from Sparse Coding in a Dynamical System. PLOS Computational Biology *9*, e1003191. https://doi.org/10.1371/journal.pcbi.1003191.

Acknowledgments

Funding for this work was provided by NIH/NEI (EY016093).

Author contributions

BM and CR conceived of the original project. GM and BM collected the natural image data, performed the analysis, developed the models, and wrote the paper.

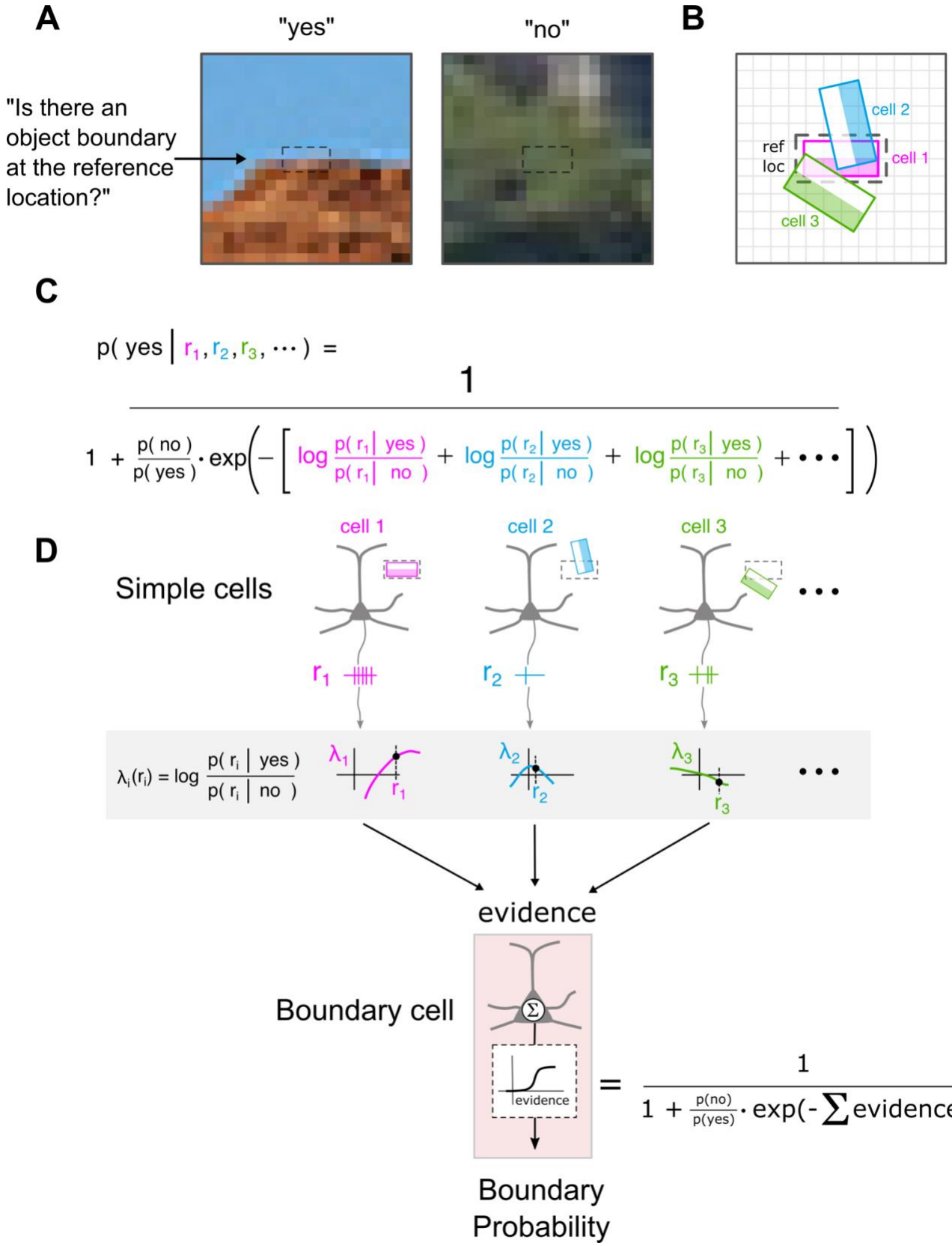
A
"yes"
"no"
"Is there an object boundary at the reference location?"
B
cell 2
ref loc
cell 1
cell 3
C
$p(\text{yes} \mid r_1, r_2, r_3, \cdots) = \dfrac{1}{1 + \frac{p(\text{no})}{p(\text{yes})} \cdot \exp\left(-\left[\log \frac{p(r_1 \mid \text{yes})}{p(r_1 \mid \text{no})} + \log \frac{p(r_2 \mid \text{yes})}{p(r_2 \mid \text{no})} + \log \frac{p(r_3 \mid \text{yes})}{p(r_3 \mid \text{no})} + \cdots\right]\right)}$
D
Simple cells
cell 1
cell 2
cell 3
r_1
r_2
r_3
$\lambda_i(r_i) = \log \frac{p(r_i \mid \text{yes})}{p(r_i \mid \text{no})}$
λ_1
λ_2
λ_3
evidence
Boundary cell
evidence
$= \dfrac{1}{1 + \frac{p(\text{no})}{p(\text{yes})} \cdot \exp(-\sum \text{evidence})}$
Boundary Probability

Figure 1

Figure 1. Calculating boundary probability from natural images using Bayes' rule. **A**, The boundary detection problem can be encapsulated by the question and answers shown; ~30,000 natural image patches were classified in this way. Dashed box indicates a "reference location" where a boundary might appear. Patches shown during labelling were 20x20 pixels. **B**, 3 (of 300) oriented cells with responses r_1, r_2, r_3 are shown in the vicinity of the reference location. **C,** Under the assumption that simple cell responses are class-conditionally independent (see main text), Bayes' rule gives an expression for boundary probability in terms of individual cell log-likelihood ratios (LLRs) (colored terms in denominator). **D**, Cell responses r_i are passed through their respective LLR functions $\lambda_i()$, and the results are summed and passed through a fixed sigmoidal "f-I curve" to yield boundary probability.

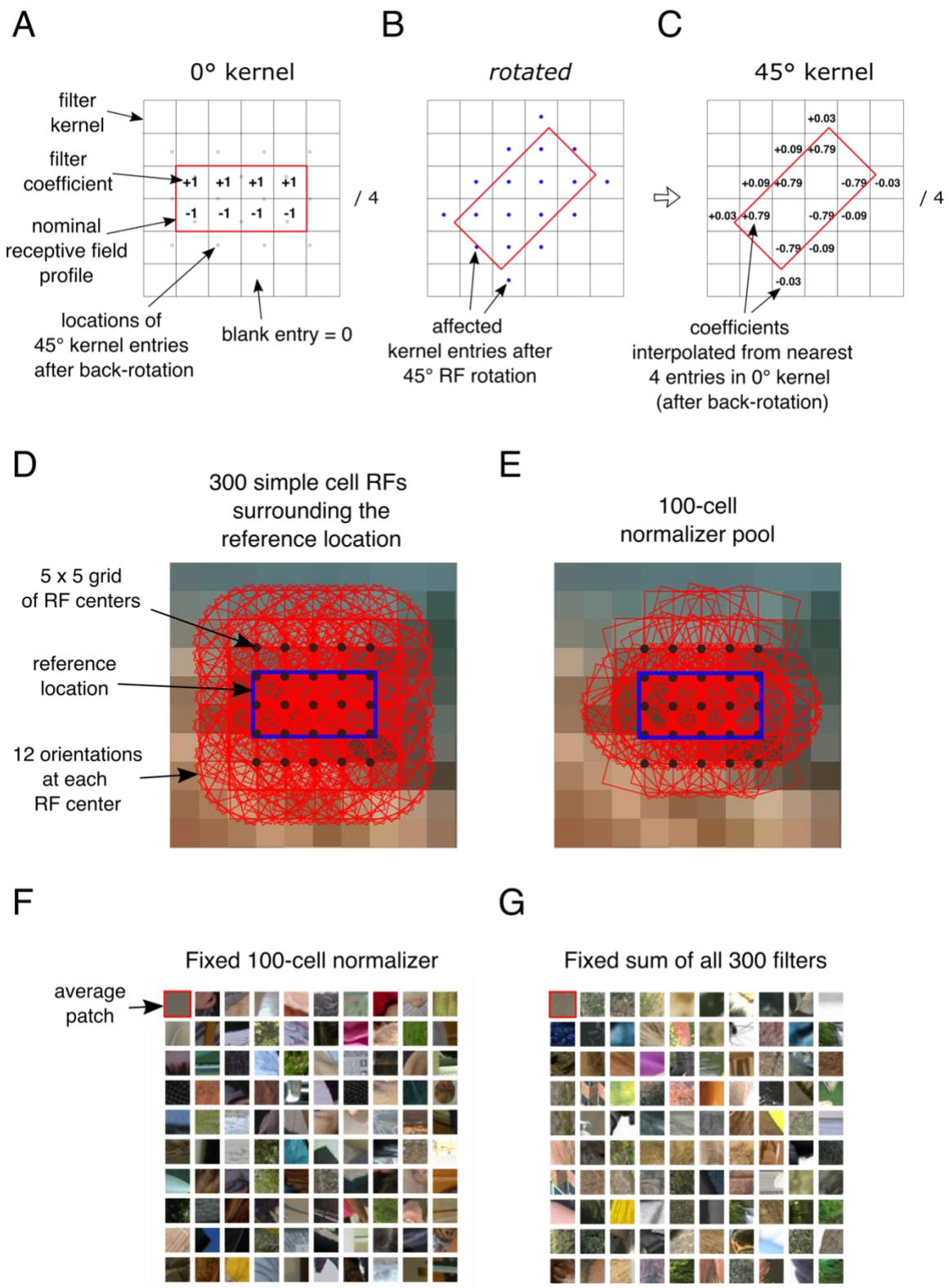
A
0° kernel
filter kernel
filter coefficient
+1 +1 +1 +1
-1 -1 -1 -1
/ 4
nominal receptive field profile
locations of 45° kernel entries after back-rotation
blank entry = 0
B
rotated
affected kernel entries after 45° RF rotation
C
45° kernel
+0.03
+0.09 +0.79
+0.09 +0.79 -0.79 -0.03
+0.03 +0.79 -0.79 -0.09
-0.79 -0.09
-0.03
/ 4
coefficients interpolated from nearest 4 entries in 0° kernel (after back-rotation)
D
300 simple cell RFs surrounding the reference location
5 x 5 grid of RF centers
reference location
12 orientations at each RF center
E
100-cell normalizer pool
F
Fixed 100-cell normalizer
average patch
G
Fixed sum of all 300 filters

Figure 2

Figure 2. Simple cell "filter" construction, layout, and normalization pools. **A,** Diagram shows layout of positive and negative filter coefficients for computing simple cells responses. Positive coefficients represent the ON subregion of the receptive field (RF), and negative coefficients represent the OFF subregion. Faint blue dots indicate off-grid positions (in x-y space) of the 18 pixel locations holding coefficients for the 45° case shown in panel B, back rotated to the horizontal orientation. **B,** Blue dots indicate all pixels that overlap with the red box, representing the original horizontal receptive field rotated by 45°. Each of these 18 pixels will contain a coefficient (see panel C), computed from bilinear interpolation of the coefficients in the original horizontal layout (see panel A). The coefficients "down the middle" of the 45° filter are zero, reflecting an equal contribution from the ON and OFF subregions of the RF. **D,** Layout of 300 filter locations (red) in 5x5 grid surrounding the reference location (blue). **E,** Subset of 100 of the 300 simple cell RFs chosen as the normalizer pool in the following figures. See text for details. **F,** Examples of image patches drawn at random from the image database whose normalizer values (i.e. the sum of the response magnitudes of the 100 cells shown in E) fell into the same bin (200 +/- 40); see text for details. Upper left patch is the average of all. **G,** Same as panel F, but with patch selection based on a different normalizer: the sum of the response magnitudes of all 300 simple cells covering the region (whose RFs are shown in panel D).

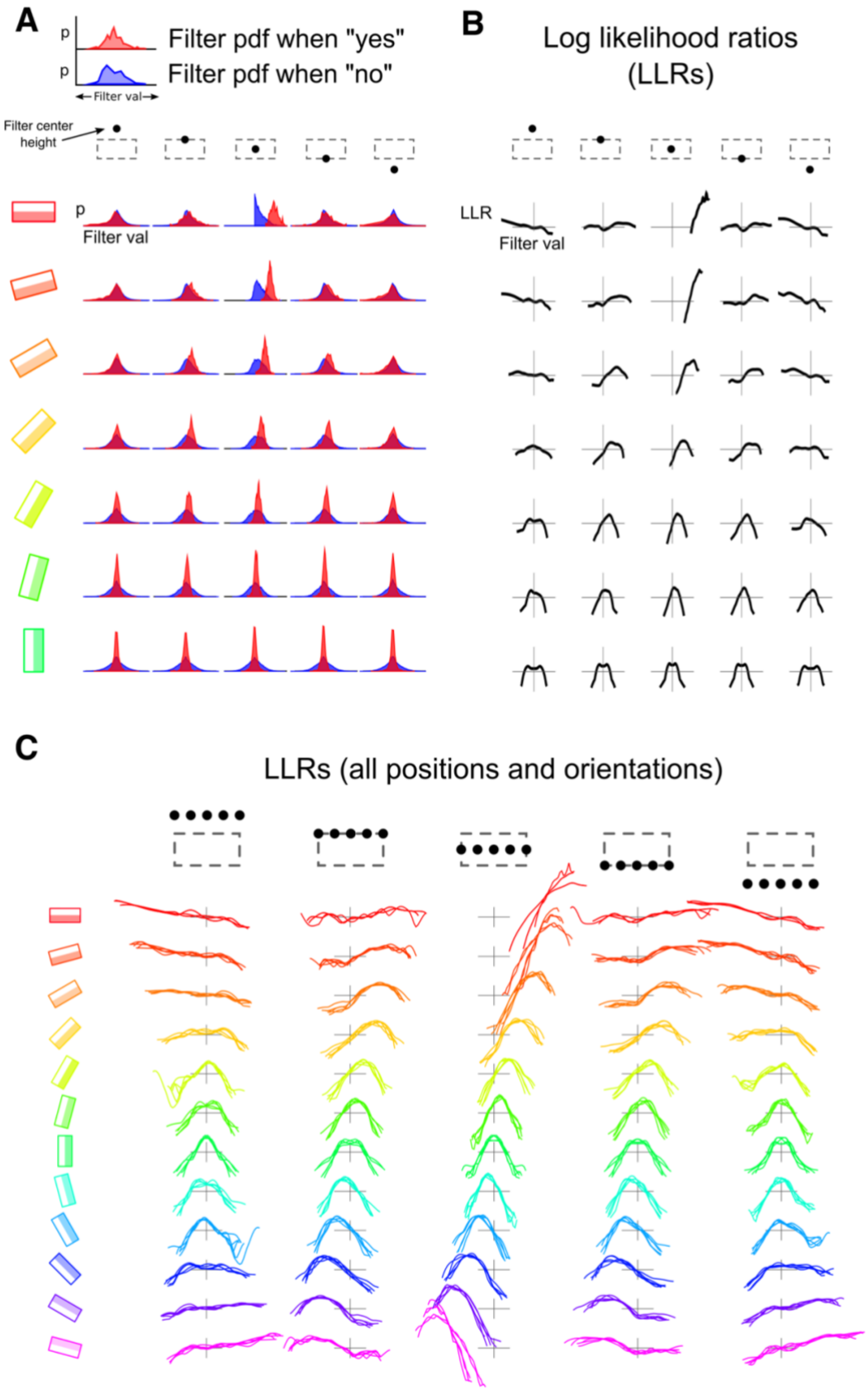
A
p
Filter pdf when "yes"
p
Filter pdf when "no"
Filter val
Filter center height
p
Filter val
B
Log likelihood ratios (LLRs)
LLR
Filter val
C
LLRs (all positions and orientations)

Figure 3

Figure 3. Computing LLR functions from natural images. **A**, Simple cell responses from 30,000 labeled image patches potentially containing boundaries at the reference location (dashed box) were separately histogrammed for "yes" (red) and "no" (blue) cases. Yes (no) cases were those with confidence scores of 4 and 5 (1 and 2). A subset of simple cell response histograms is shown for 7 orientations and 5 vertical positions (centered horizontally). **B,** By dividing the yes and no distributions and taking natural logarithms, one obtains the LLR functions $\lambda_i()$ for each cell which vary as a function of the cell's response r_i. (**C**) The full set of 300 LLR functions reveals a regular pattern over orientation and location. Cases grouped within each subplot are for 5 horizontal shifts (indicated by black dots at top). Many LLR functions are non-monotonic functions of the cell's response.

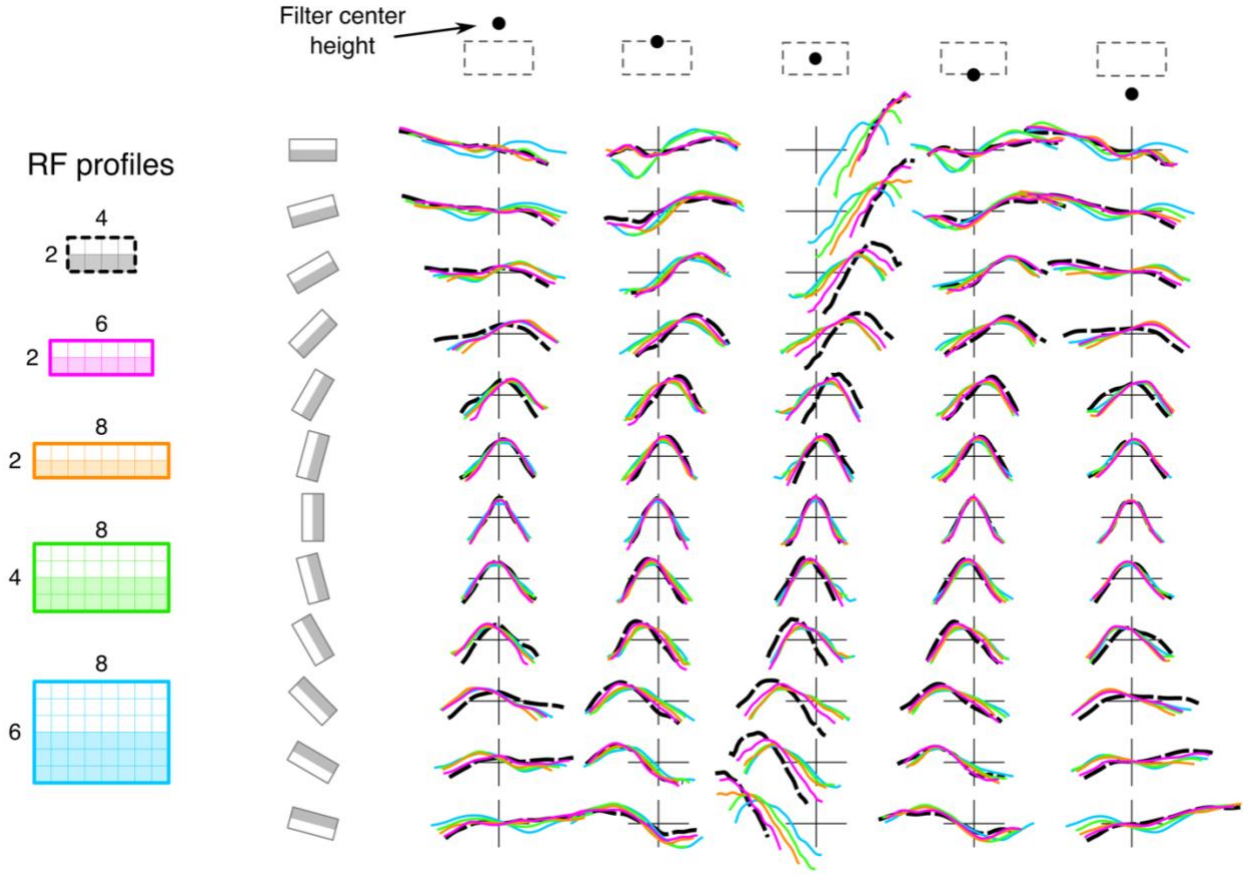

Figure 4

Figure 4. The basic pattern of LLRs forms is conserved across different filter spatial profiles. LLRs were generated for each of the filter profiles shown on the left (2x6, 2x8, 4x8, and 6x8 pixels). The overall spectrum of LLR shapes remains similar for the different cases.

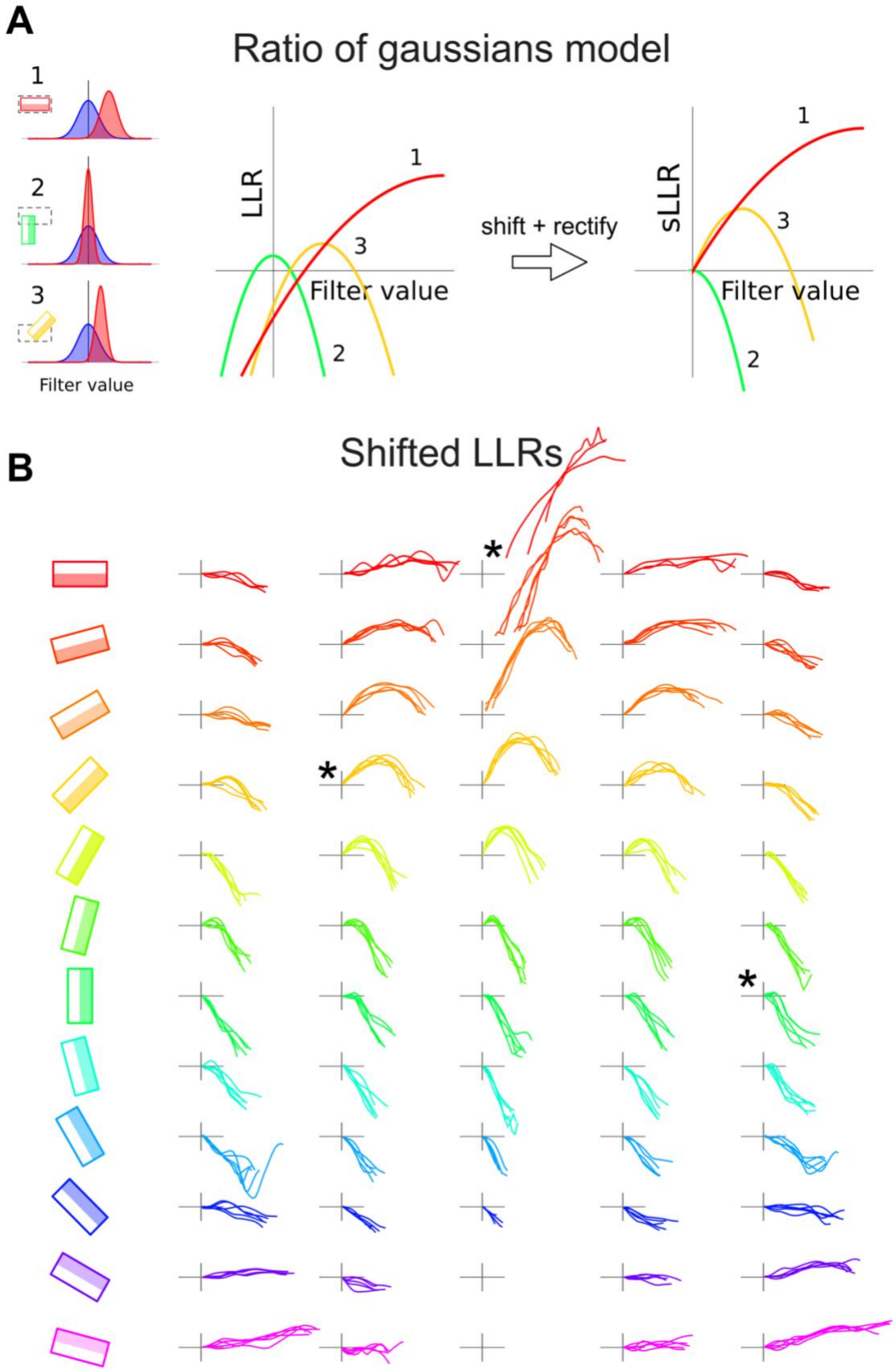
A
Ratio of gaussians model
1
2
3
Filter value
LLR
1
3
Filter value
2
shift + rectify
sLLR
1
3
Filter value
2
B
Shifted LLRs
*
*
*

Figure 5

Figure 5. Interpreting the LLRs as cell-cell interaction functions. **A**, Modelling the "yes" and "no" distributions as gaussians (left panel) leads to parabolic LLRs (middle panel). In order to interpret the LLRs as cell-cell interactions functions, we perform two additional processing steps: (1) When a simple cell is inactive, it should not influence the boundary cell; this is accomplished by shifting the LLR to have zero output (y=0) when the input is zero (x=0); sLLR stands for "shifted LLR". (2) Simple cells cannot have negative firing rates, and so the left halves of the LLRs, corresponding to negative simple cell firing rates, are discarded (these cases are handled by an opponent SC whose RF is identical but with the ON and OFF subfields reversed). This produces the curves in the right panel. **B**. The full set of LLR interactions processed in this way. Many of them are non-monotonic, indicating that that simple cell should have a non-monotonic effect on the boundary cell. The plots corresponding to the three LLRs modelled in panel A are marked with asterisks.

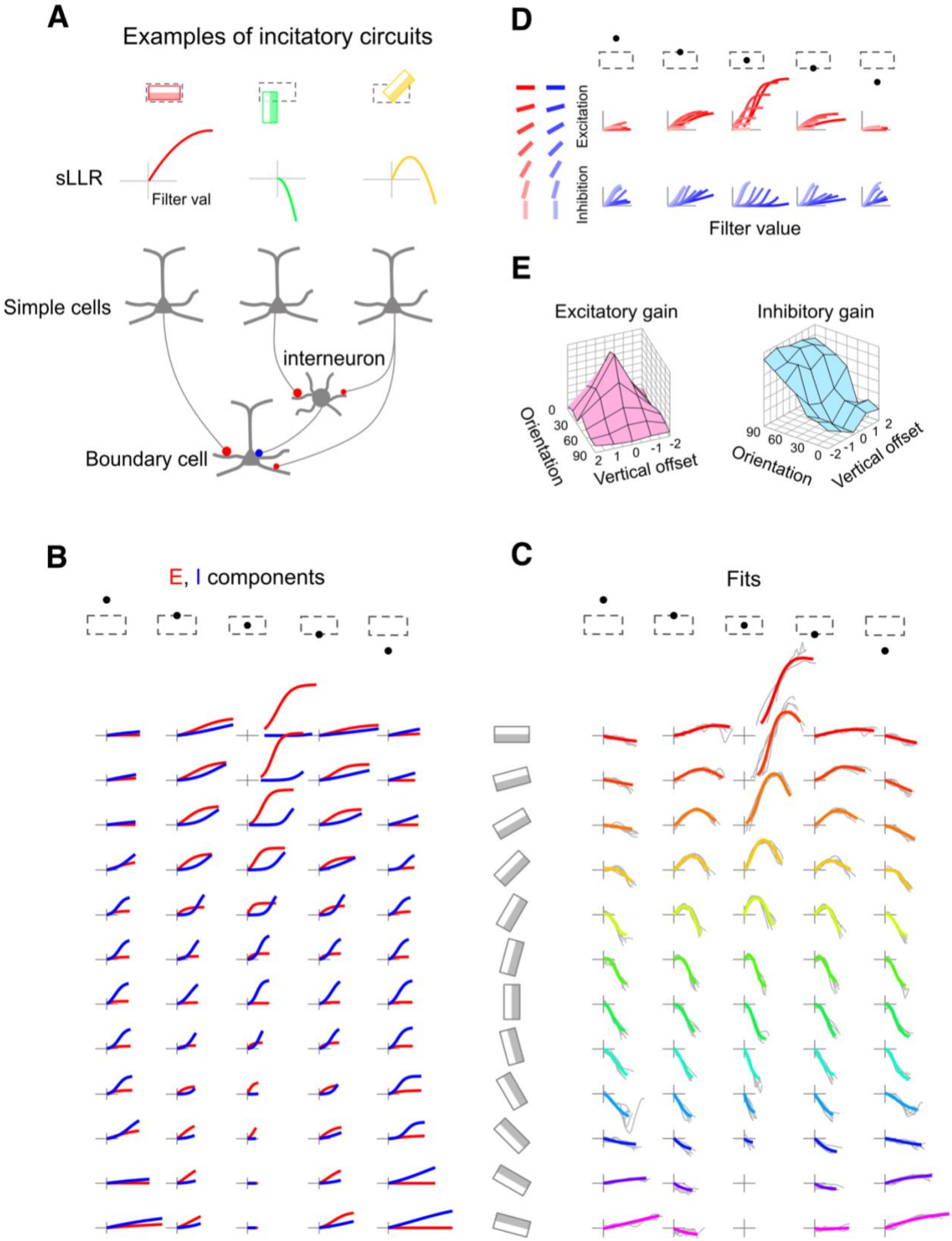
A
Examples of incitatory circuits
sLLR
Filter val
Simple cells
interneuron
Boundary cell
D
Excitation
Inhibition
Filter value
E
Excitatory gain
Inhibitory gain
Orientation
Vertical offset
B
E, I components
C
Fits

Figure 6

Figure 6. Fitting simple cell–boundary cell interactions (LLRs) with a difference of sigmoids representing separate E and I effects. **A**, Each of the three sLLRs shown can be parametrized by an incitatory circuit. The circuits implementing the red and green sLLRs involve pure excitation and inhibition, special cases of incitation, while that of the orange sLLR involves a nontrivial combination of both excitation and inhibition. **B**, E (red) and I (blue) sigmoidal curves were optimized by manipulating their thresholds, slopes and asymptotes so that their difference fit the corresponding LLR shown in panel **C**. **C**. LLR fits are shown in color, on top of the 5-curve groups from Figure 3C shown in light grey. **D**, E and I sigmoids from b are collected across orientations within each subplot, showing smooth progressions of sigmoid parameters. **E**, Plots show gains for the E and I interaction components. For groups of simple cells horizontally centered at the RL, excitation delivered to the boundary cell becomes weaker and inhibition grows stronger as the neighbor's orientation deviates from the reference orientation.

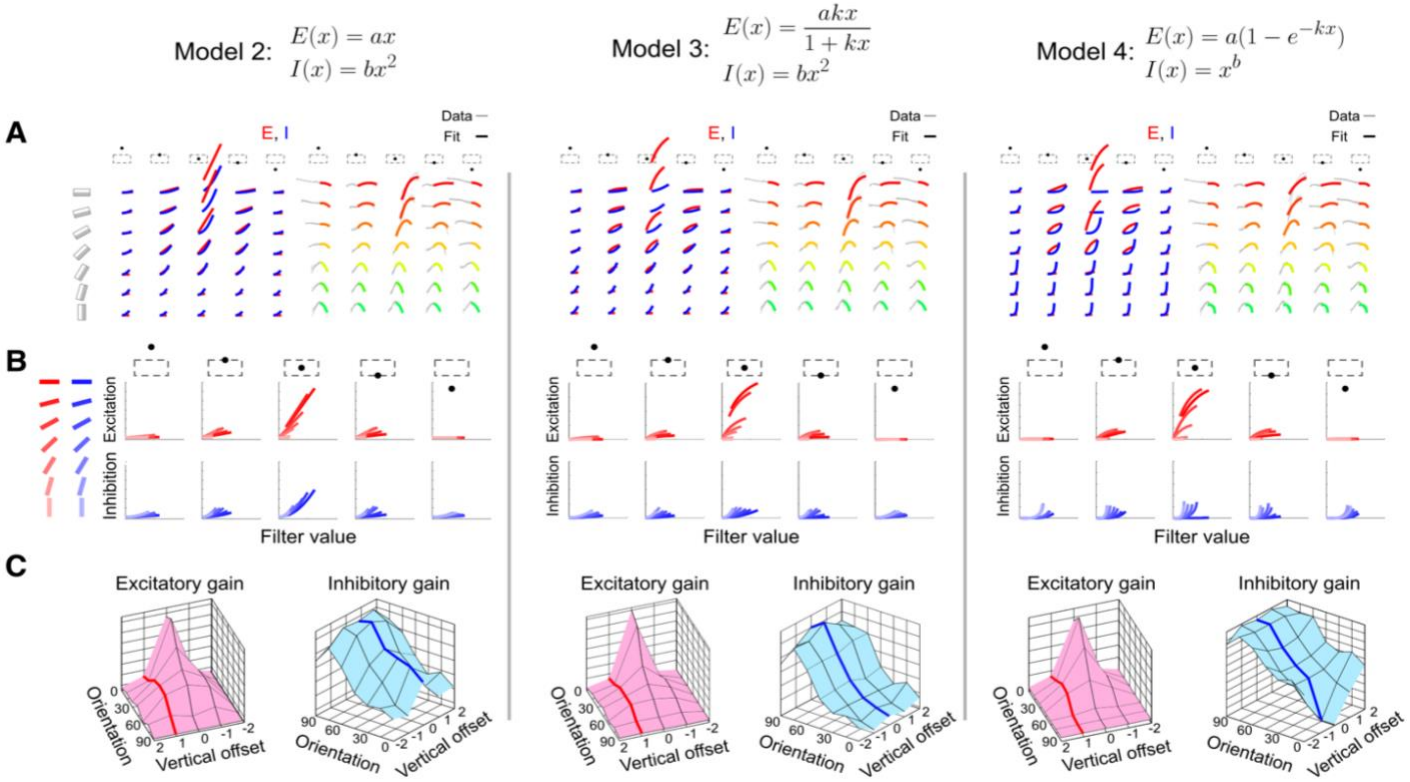
Model 2: $E(x) = ax$ $I(x) = bx^2$
Model 3: $E(x) = \frac{akx}{1+kx}$ $I(x) = bx^2$
Model 4: $E(x) = a(1-e^{-kx})$ $I(x) = x^b$
A
B
C
E, I
Data
Fit
Excitation
Inhibition
Filter value
Excitatory gain
Inhibitory gain
Orientation
Vertical offset

Figure 7

Figure 7. Circuit-level predictions depend only weakly on the choice of parameters representing the excitatory and inhibitory component curves. Three roughly similarly performing models are shown. **A**, Excitatory (red) and inhibitory (blue) curve components (left) and resulting LLR fit (right) are shown for each model. Fit quality is comparable across all three models, and the original model shown in Figure 6C. **B**, Despite having different E-I curve shapes, all three models show the same basic trends in the progression of excitation and inhibition as a function of orientation and vertical offset from the RL. **C**, Summarizing each E and I curve with a single gain parameter shows a similar pattern for the three models.

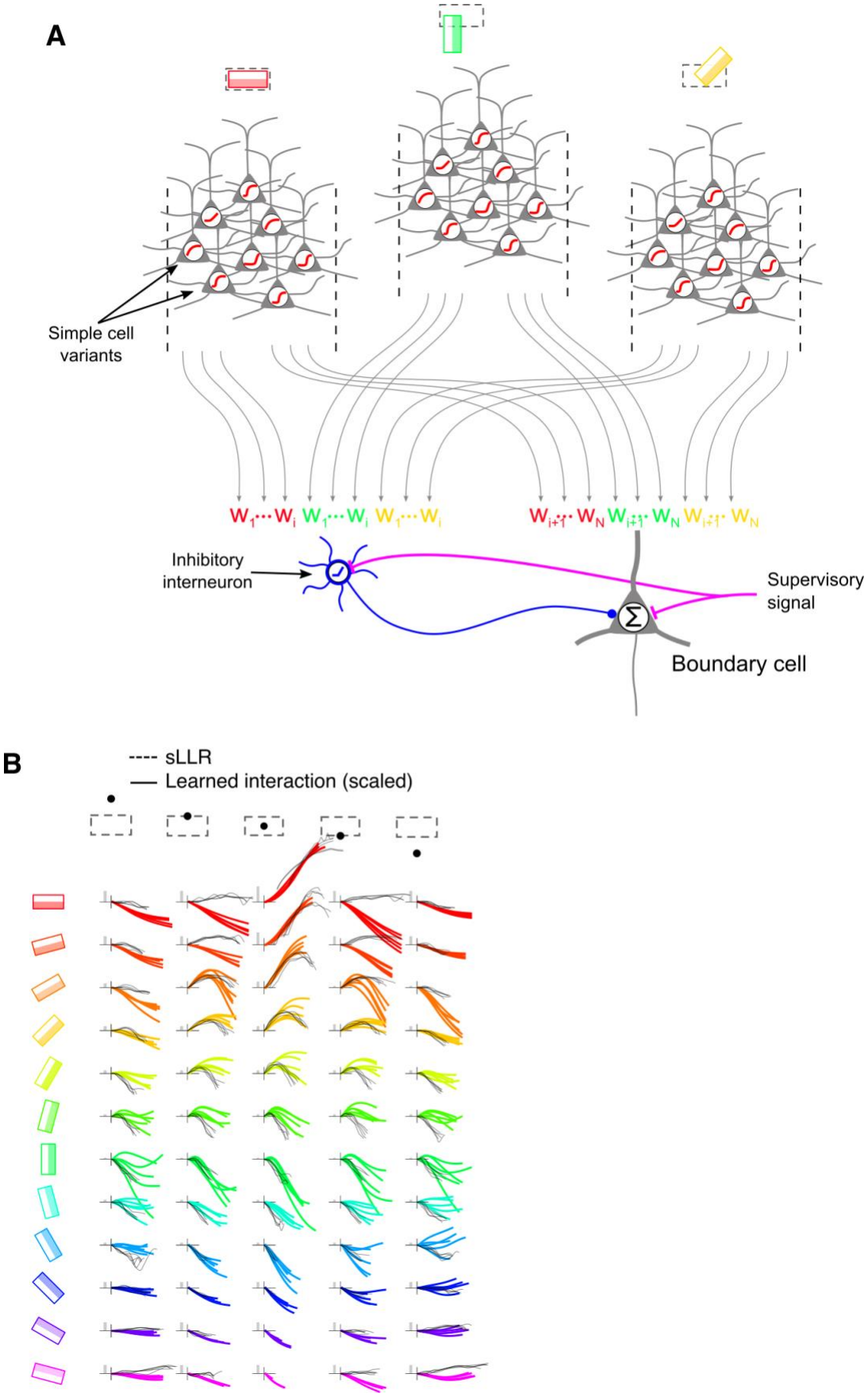
A
Simple cell variants
Inhibitory interneuron
Supervisory signal
Boundary cell
B
---- sLLR
— Learned interaction (scaled)

Figure 8

Figure 8. Simple cell-boundary cell interactions can be learned by a biologically plausible synaptic plasticity rule. **A**, Each oriented filter was represented by a population of 8 simple cells, each with a different fixed i/o nonlinearity. Nonlinearities were sigmoids, $y = \frac{1}{1+e^{-g(\sum_i w_i x_i - t)}}$, with threshold t set at 8 evenly spaced values between -6 and 35. The learning rule used to adjust the weights from each simple cell onto the inhibitory and boundary cell was: $\Delta w_i = \pm\eta(t - y)x_i$, where t is the "training signal" (1 for boundary, 0 for no boundary), y is the response of the boundary cell, x_i is the response of the ith simple cell, η is the learning rate, and the positive (negative) sign was used for the boundary (inhibitory) cell. In the context of our model, this learning rule is mathematically equivalent (up to a transient initial difference in the learning rate parameter η) to a learning rule which constrains all weights to be positive. **B**, To determine the net effect of each filter on the boundary cell (for comparison to the LLRs), the underlying linear filter value was increased from 0 to 1 while holding all other inputs constant, and the weighted sum of the 8 associated simple cells was plotted (colored curves). Black curves are sLLRs from Figure 5B. The gray bar in each plot represents the weight that the BC puts on that group of 5 colored curves.

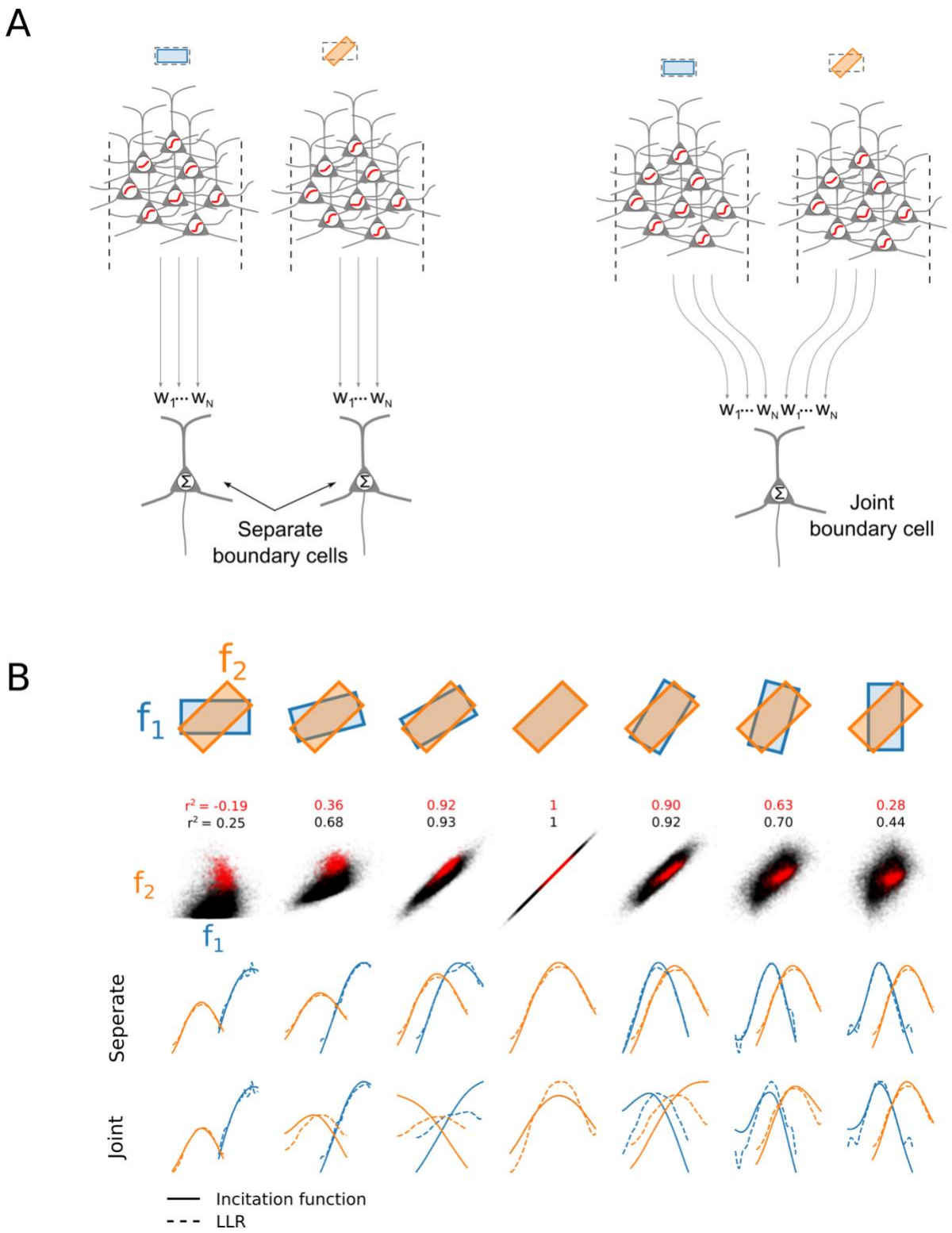
A
$w_1 \cdots w_N$
$w_1 \cdots w_N$
Separate
boundary cells
$w_1 \cdots w_N$ $w_1 \cdots w_N$
Joint
boundary cell
B
f_2
f_1
r^2 = -0.19
r^2 = 0.25
0.36
0.68
0.92
0.93
1
1
0.90
0.92
0.63
0.70
0.28
0.44
f_2
f_1
Seperate
Joint
Incitation function
LLR

Figure 9

Figure 9. The incitation circuit learns literal LLRs when the filters are class-conditionally independent. **A**, We selected several pairs of filters and fitted either 2 separate incitation circuits, one for each filter (left), or one circuit with both inputs (right). **B**, (Top) Filter pairs ranged from very different (left and right) to very similar or identical (middle) filters. (Middle) Scatter plots of joint filter responses for boundary (red) and non-boundary (black) image patches. (Bottom) When filters were fit separately, the learned incitation functions (solid curves) were nearly identical to the filters' LLR curves (dashed). When the filters were fit jointly, pairs with very similar filters no longer learned LLR functions due to a breakdown of CCI.

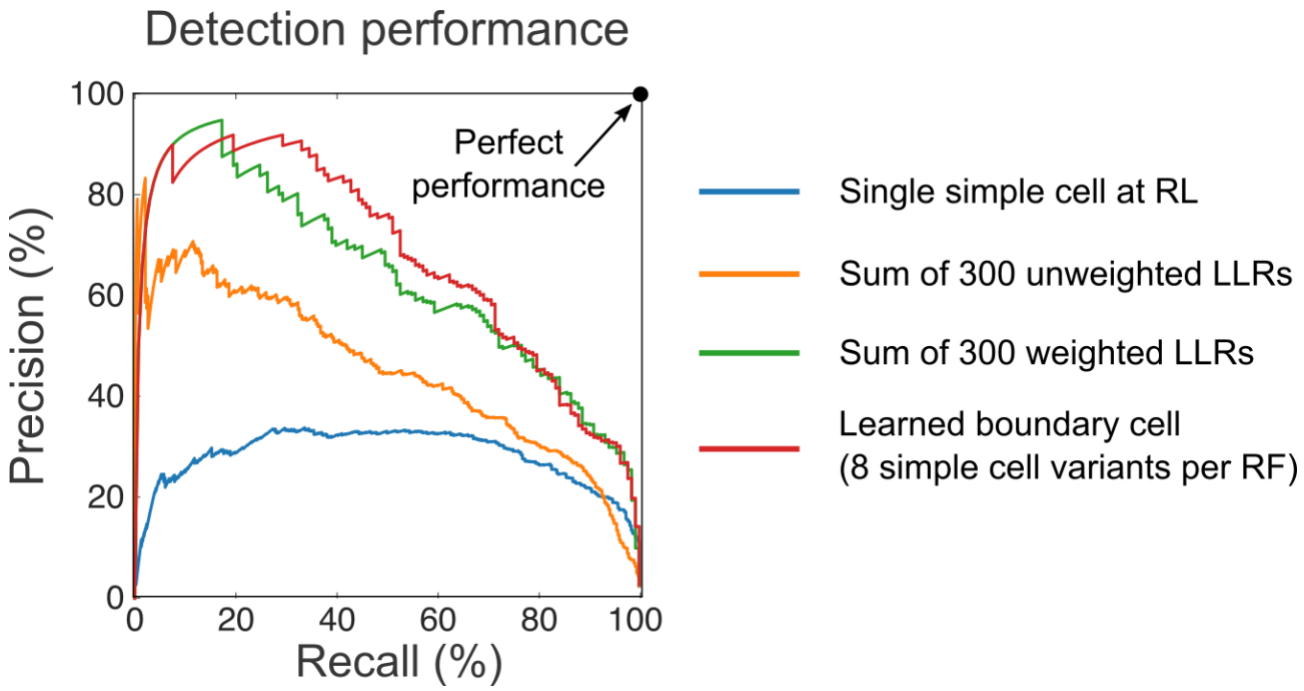
Detection performance
Precision (%)
Recall (%)
0
20
40
60
80
100
Perfect performance
Single simple cell at RL
Sum of 300 unweighted LLRs
Sum of 300 weighted LLRs
Learned boundary cell
(8 simple cell variants per RF)

Figure 10

Figure 10. Precision-recall curves for 4 boundary detecting classifiers. See Methods for definitions of precision and recall. Perfect performance is at upper right-hand corner. A single oriented simple cell at the reference location performs poorly (~20% precision at 90% recall; blue curve). Classifier consisting of a sum of 300 unweighted simple cell LLRs is shown in orange. This case corresponds to the simplified Bayesian classifier depicted in Figure 1C and D. Poor performance was expected given violations of the CCI assumption in the simple cell population. Green curve shows is similar, but with weights on each of the 300 LLRs trained by logistic regression. Performance is dramatically improved, indicating that the learning rule helps to compensate for correlations in the simple cell population. Red curve shows performance of the trained incitation network of Figure 8A. Precision exceeds that of a single simple cell by ~1.5x at 90% recall, and 2.3x at 50% recall.

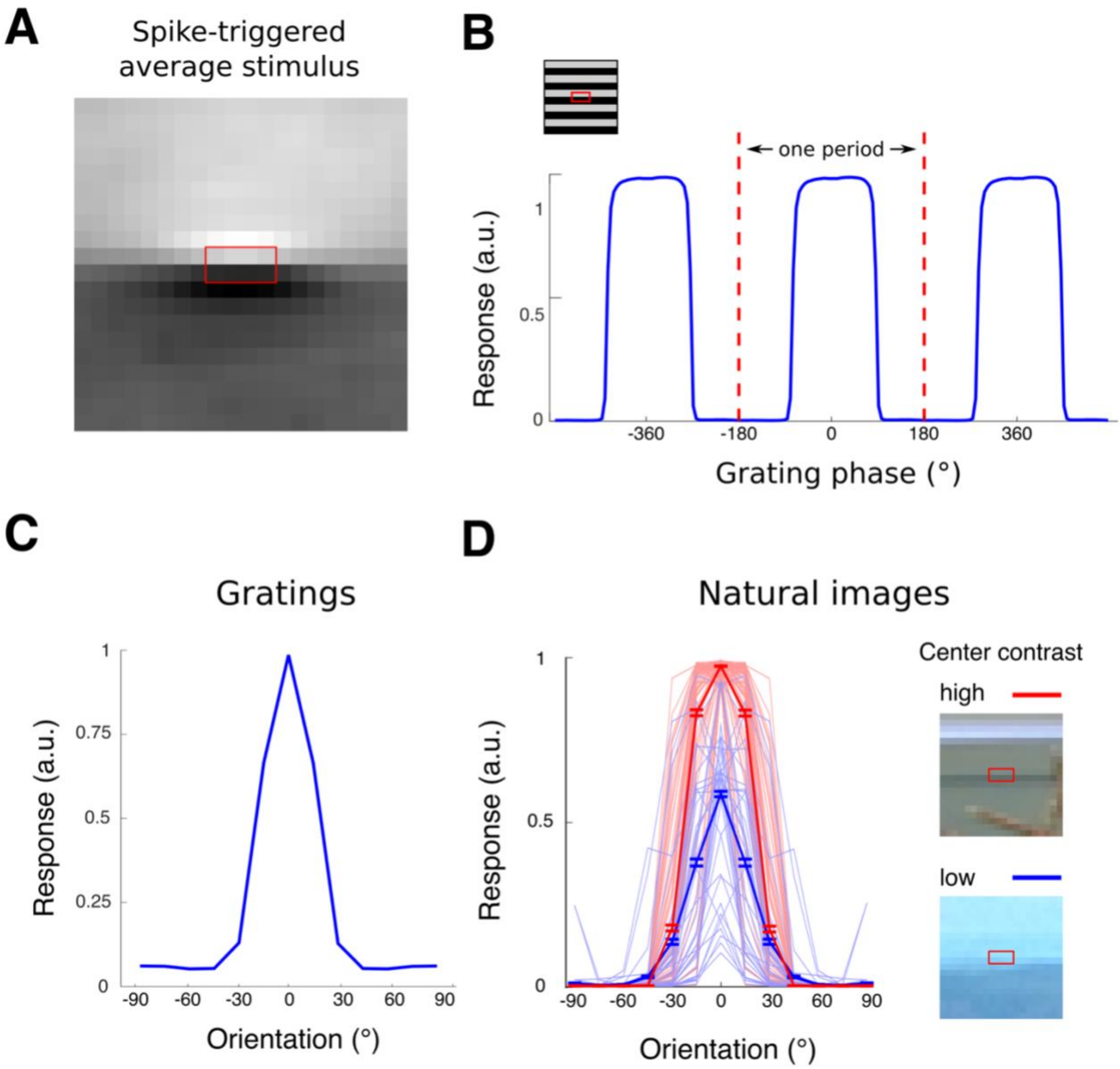
A
Spike-triggered average stimulus
B
one period
Response (a.u.)
1
0.5
0
-360
-180
0
180
360
Grating phase (°)
C
Gratings
Response (a.u.)
1
0.75
0.5
0.25
0
-90
-60
-30
0
30
60
90
Orientation (°)
D
Natural images
Response (a.u.)
1
0.5
0
-90
-60
-30
0
30
60
90
Orientation (°)
Center contrast
high
low

Figure 11

Figure 11. Boundary cell responses to parametric and natural stimuli resemble simple cell responses. To compute BC responses, the weighted sum of LLRs model (orange PR curve in Figure 10) was used. **A**, Spike-triggered average stimulus computed by averaging natural image stimuli weighted by their evoked boundary cell response. **B**, Response of a boundary cell to a grating presented at different phrases. The boundary cell is simple cell-like in that it is sensitive to polarity, responding to only half of all phases. **C**, Orientation tuning curve to the same grating. At each orientation, responses were averaged over all phases of the grating. The resulting tuning curve is similar to those obtained for simple cells in V1. **D**, Patches with fixed surround contrast (normalizer value) and varying center contrast were selected and presented at 15° increments to the boundary cell. For a fixed surround contrast, center contrast increases cell response without increasing tuning width, a hallmark of contrast invariant orientation tuning found in V1 simple cells (full width at half height for high contrast stimuli (red curve) is 43.6°, and for low contrast stimuli, 39.2°).

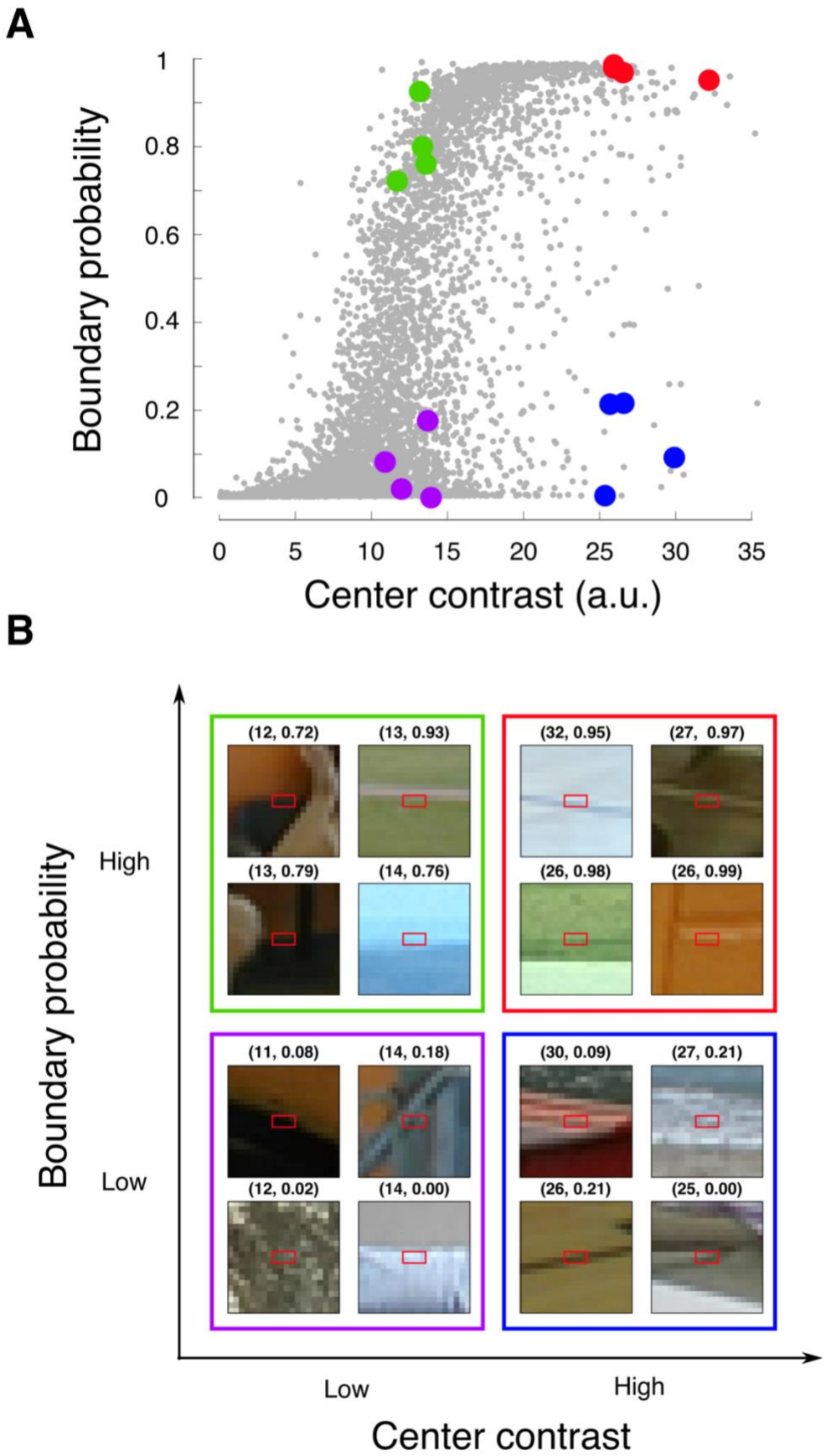
A
Boundary probability
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20
25
30
35
Center contrast (a.u.)
B
Boundary probability
High
Low
(12, 0.72)
(13, 0.93)
(13, 0.79)
(14, 0.76)
(32, 0.95)
(27, 0.97)
(26, 0.98)
(26, 0.99)
(11, 0.08)
(14, 0.18)
(12, 0.02)
(14, 0.00)
(30, 0.09)
(27, 0.21)
(26, 0.21)
(25, 0.00)
Low
High
Center contrast

Figure 12

Figure 12. Distinguishing linear filter responses from boundary probability responses. To determine whether a given cell is computing linear contrast or boundary probability, it is necessary to use a stimulus set which dissociates these two measures. Roughly speaking, what is needed are stimuli whose linear filter and boundary probability scores are "well spread" throughout linear filter-boundary probability space. **A**, Plotting the two scores for all labelled patches shows that they are highly correlated, and that randomly selected patches are likely to lie at the lower left and upper right corners of this space – where linear contrast and boundary probability are either both low or high together. Therefore, if only these stimuli were presented to the cell, it would be difficult to know whether high cell responses were being driven by linear contrast or boundary probability. It would be better to present stimuli that are well spread over the space of the two scores (colored dots) so that cell responses to each variable can be assessed separately. **B**, Examples of these stimuli are shown. They include low contrast non-edges (purple cases), high contrast non-edges (blue cases), low contrast edges (green cases), and high contrast edges (red cases).

Appendix 1: Logistic regression learns LLRs assuming CCI

We are interested in estimating the probability of some event y, in this case whether a patch contains a boundary, from input features $\vec{x}$, in this case the responses of several simple cells. Logistic regression builds a model of $p(y|\vec{x})$ by assuming that the output probability is a sigmoid function of a linear combination of the features:

$$\hat{p}(y|\vec{x}) = \sigma(w^T\vec{x}) = \frac{1}{1+e^{-w^T\vec{x}}}$$

The goal of learning is to pick weights w that minimize the expected cross entropy between the true and model probabilities:

$$w^* = \text{argmin}_w - \sum_{\vec{x}} p(\vec{x}) \sum_y p(y|\vec{x}) \ln \hat{p}\,(y|\vec{x})$$

$$= \text{argmin}_w - \sum_{\vec{x}} p(\vec{x}) \sum_y p(y|\vec{x}) \ln \frac{\hat{p}(y|\vec{x})}{p(y|\vec{x})} - \sum_{\vec{x}} p(\vec{x}) \sum_y p(y|\vec{x}) \ln p\,(y|\vec{x})$$

The left term is the KL divergence between the true and model distributions, and the second term is constant with respect to the weights, and can be ignored. The optimization is then

$$w^* = \text{argmin}_w D_{KL}\big(p(y,\vec{x})||\hat{p}(y,\vec{x})\big)$$

This is minimized when the model distribution $\hat{p}$ matches the true distribution p . To see that under the assumption of class conditional independence, learning the LLR functions achieves this minimum, observe

$$\hat{p}(y,\vec{x}) = p(\vec{x})\hat{p}(y|\vec{x}) = p(\vec{x})\sigma\big(f(\vec{x})\big)$$

Where $f(\vec{x}) = w^T\vec{x}$. Further, class conditional independence implies

$$p(y|\vec{x}) = \frac{p(y)p(\vec{x}|y)}{p(y)p(\vec{x}|y) + p(\bar{y})p(\vec{x}|\bar{y})} = \frac{1}{1+\frac{p(\bar{y})p(\vec{x}|\bar{y})}{p(y)p(\vec{x}|y)}} = \sigma\left(-\log\frac{p(\bar{y})p(\vec{x}|\bar{y})}{p(y)p(\vec{x}|y)}\right)$$

$$= \sigma\left(-\log\frac{p(\bar{y})}{p(y)} + \sum_i \log\frac{p(x_i|y)}{p(x_i|\bar{y})}\right) = \sigma\left(-\log\frac{p(\bar{y})}{p(y)} + \sum_i LLR_i(x_i)\right)$$

so that the objective can be written

$$w^* = \text{argmin}_w D_{KL}\left(p(\vec{x})\sigma\left(-\log\frac{p(\bar{y})}{p(y)} + \sum_i LLR_i(x_i)\right) \;||\; p(\vec{x})\sigma\big(f(\vec{x})\big)\right)$$

One can see by inspection that the two distributions will be equal and the objective will be minimized if and only if

$$f(\vec{x}) = -\log\frac{p(\bar{y})}{p(y)} + \sum_i LLR_i(x_i)$$

that is, when the classifier simply combines the filter values by passing them through their LLR functions and summing the result.

Now consider the problem of learning the optimal weights through gradient descent using a dataset of N input-output pairs. Call the ith input patch $x^{(i)}$ and the ith output label $y^{(i)}$. Label non-boundaries $y_i = -1$ and boundaries $y_i = 1$. The cost function, gradient, and Hessian can be written

$$J = \frac{1}{N}\sum_k \ln\left(1 + e^{-y^{(k)}w^T x^{(k)}}\right)$$

$$\frac{\partial J}{\partial w_i} = \frac{1}{N}\sum_k \frac{-y^{(k)}x_i^{(k)}}{1 + e^{-y^{(k)}w^T x^{(k)}}}$$

$$\frac{\partial^2 J}{\partial w_i\, \partial w_j} = \frac{1}{N}\sum_k \frac{x_i^{(k)}x_j^{(k)}}{\left(1 + e^{-y^{(k)}w^T x^{(k)}}\right)^2}$$

If we collect all of the input patches $x^{(k)}$ into the rows of the matrix X , the Hessian can be written simply as

$$\frac{\partial^2 J}{\partial w\, \partial w^T} = \frac{1}{N}X^T D X$$

Where D is a positive diagonal matrix with entries $D_{kk} = \left(1 + e^{-y^{(k)}w^T x^{(k)}}\right)^{-2}$. First, note that the Hessian is positive semidefinite since $v^T X^T D X v = \left\|\sqrt{D}Xv\right\|^2 \geq 0$ for all v. In the case that the K features making up the columns of the training set X are linearly independent (generically true if the input cells have distinct filter kernels), the Hessian is strictly positive definite since each of the following matrices having rank $\geq K$ implies the same of the next: $X^T X, XX^T, XX^T D, X^T D X$, the last of which is the Hessian. Regardless, by positive semi-definiteness, any local minimum of the objective function J is a global minimum, so gradient descent terminates at the global minimum, which by the argument above must be the LLR solution.