# Title: Detecting object boundaries in natural images requires 'incitatory' cell-cell interactions

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

**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: mel@usc.edu

**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 (BC), and to predict how cortical circuitry could produce a BC from a population of conventional "simple cells" (SCs), we labeled 30,000 natural image patches and used Bayes' rule to determine how an SC should affect a BC depending on its offset in receptive field position and orientation. We identified three types of cell-cell interactions: monotonic excitatory and inhibitory, with various thresholds, slopes, and amplitudes, and a spectrum of non-monotonic (U-shaped) interactions. Using simple models we show a common cortical circuit motif consisting of monosynaptic excitation and disynaptic inhibition – an arrangement we call "incitation" – can produce the entire spectrum of SC-BC interactions found in our dataset. Moreover, we show that the synaptic weights that parameterize the incitation circuit can be learned by a simple (1-layer) gradient descent learning rule. We conclude that trainable incitatory interconnections may be a general computational mechanism used by the cortex 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-performing 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 decoding circuit known to exist in V1 is easily capable of computing a high-quality boundary probability signal from a local population of simple cells. Our findings suggest a new and more general way of conceptualizing cell-cell interconnections in the cortex.

**Main text:**

The primary visual cortex (area V1) is a complex, poorly understood, multi-purpose image processing computer optimized to extract information from natural images – which are themselves complex, poorly understood, multi-purpose signals. Thus, understanding how V1 operates presents a challenging reverse engineering problem. A longstanding hypothesis is that V1 cells somehow participate in object boundary detection, a core process in biological vision (Biederman, 1987; Gilbert & Wiesel, 1990; Heydt & Peterhans, 1989; Hubel & Wiesel, 1962a; Kapadia, Ito, Gilbert, & Westheimer, 1995a) that is crucial for the functions of both ventral and dorsal streams (Biederman, 1987; Hoffman, 2000; Rust & Dicarlo, 2010; Theys, Romero, van Loon, & Janssen, 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 distinguishes boundaries from non-

boundaries. This uncertainty has made it difficult to form specific computational hypotheses as to how V1 circuits could perform this particular behaviorally-relevant classification task. Previous work has analyzed natural image statistics to determine how local boundary segments are arranged in images, and how these arrangements relate to human contour grouping performance (W. S. Geisler, Perry, Super, & Gallogly, 2001; Sigman, Cecchi, Gilbert, & Magnasco, 2001). However, no study has yet addressed the issue of how boundary elements should ideally be detected in the first place (independent of the brain), nor what types of neural circuits are capable of doing so. With the goal to better understand the V1 computations underlying object boundary detection (Figure 1A), we began with a known cell type – orientation-tuned "simple cells" (as defined by Hubel & Wiesel, 1962b), and typically modeled as divisively normalized oriented linear filters Carandini & Heeger, 2012) – and asked how the responses of a population of simple cells (SCs) of all orientations, whose receptive fields (RFs) densely cover a region of the image, should be combined to produce a "boundary cell" (BC) whose response signals 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 (Figure 1C). In a previous study (Ramachandra & Mel, 2013), we noted that under an important simplifying assumption (i.e. class conditional independence; see methods for detailed discussion), SC-BC interactions are captured by the log-likelihood ratios (LLRs) 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 the BC's receptive field (Figure 1D). We found that that the SC-BC interactions were diverse, and in some cases involved mixed excitatory and inhibitory effects. However, since only a small number of neighboring cells were analyzed in that study, and the results were heterogeneous, we could not come to general conclusions about types of interactions

between cells needed to compute boundary probability, making it difficult to evaluate or compare candidate neural mechanisms.

To develop a more complete picture of the cell-cell interactions needed for natural boundary detection, in this study we collected and labeled 30,000 natural image patches, with scores ranging from 5, indicating high confidence that a boundary was present at a "reference location" (RL, indicated by a dashed box in Figure 1A), to 1, indicating high confidence that a boundary was *not* present at the RL. From these labeled patches, we histogrammed oriented linear filter values (representing simple cell responses) separately in "yes" (scores of 4-5) and "no" (scores of 1-2) categories (red and blue histograms in Figure 2A, respectively) and computed LLRs for all 300 neighboring simple cells at 12 orientations on a 5x5 pixel lattice centered on the RL. Examples of LLRs are shown in Figure 2B, and the full set is shown in Figure 2C grouped across 5 horizontal shifts at each orientation and vertical position. Given that neurons do not fire at negative rates, the curves to the left and right of the origin in each plot can be considered as the LLRs for two distinct simple cells in an opponent pair, that is, with identical RFs but with their ON and OFF subfields reversed.

For many neighboring cells, the SC-BC interaction was monotonic, that is, as the SC's visually driven response increased, the evidence it provided to the BC, whether positive or negative, increased monotonically in strength. In other cases, however, the LLRs representing SC-BC interactions were non-monotonic "bumps" for one of the two simple cells in each opponent pair. This pattern was observed most often for neighbor cells whose RFs overlapped heavily with the RL (i.e. in the central three columns in Figure 2C). When we generated LLRs for simple cells of different scales and shapes (2x6, 4x6, 4x8, and 6x8 pixel filters) we found a very similar pattern of results, suggesting that the spectrum of SC-BC interactions that includes both monotonic excitatory and inhibitory as well as non-monotonic effects, may be a general feature of the boundary detection problem in natural images (Figure 3).

To gain intuition about the bump-shaped interactions, wherein the evidence that a simple cell provides to a boundary cell grows at first more positive and then less positive (or more negative) as the SC's firing rate increases, we analyzed natural image patches eliciting low, medium, and high responses from a specific neighboring filter, for the same (fixed) filter response at the RL (Figure 4). Patches eliciting low neighbor responses often contained disorganized, smooth, or defocused surfaces with low oriented energy (e.g. sky), and rarely contained a boundary at the RL, leading to low or negative "evidence" from the neighbor cell. Patches eliciting high neighbor responses often contained a boundary through the *neighbor*'s receptive field, rather than at the RL, also leading to low or negative evidence values from the neighbor cell. Only in the medium range, where the neighbor cell's response suggests that the patch contains some structure, but is not so strong as to "steal" the boundary from the RL, does the neighbor cell deliver its maximum evidence value.

We next asked what type of neural interconnection circuit is capable of producing both monotonic and non-monotonic functional connections between nearby cells. Monotonic excitation and inhibition are straightforward, but non-monotonic cell-cell interactions require a compound excitatory-inhibitory (E-I) interconnection scheme. One candidate mechanism is the ubiquitous circuit motif in which a cortical cell both directly excites and disynaptically inhibits other cells in its neighborhood (Buzsáki, 1984; Isaacson & Scanziani, 2011; Klyachko & Stevens, 2006; McBain & Fisahn, 2001; Pfeffer, Xue, He, Huang, & Scanziani, 2013; Pouille & Scanziani, 2001; Swadlow, 2002; Wehr & Zador, 2003) (Figure 5A, rightmost case). If the excitatory effect dominates at low firing rates and the inhibitory effect dominates at high rates, the neighbor cell's net effect on its target can be non-monotonic. When the circuit is simplified to consist of only the direct excitatory or indirect inhibitory pathway, the cell-cell

interaction reduces to conventional monotonic excitation or inhibition (Figure 5A, left and middle cases). To determine whether this circuit motif can produce the full range of LLR interactions found in our data set, we assumed that both the direct excitatory and indirect inhibitory pathways exert a sigmoidal effect on the boundary cell, and fit each LLR function with the difference of an excitatory and an inhibitory sigmoid. Each of the sigmoids was allowed to vary in threshold, gain, and amplitude (Figure 5B). The fits are shown in Figure 5C, confirming that the range of cell-cell interactions needed to calculate boundary probability in natural images, including non-monotonic interactions, can be produced by a simple circuit motif known to be present in V1 (Buzsáki, 1984; George, Lyons-Warren, Ma, & Carlson, 2011; Isaacson & Scanziani, 2011; Klyachko & Stevens, 2006; McBain & Fisahn, 2001; Pfeffer et al., 2013; Pouille & Scanziani, 2001; Swadlow, 2002; Wehr & Zador, 2003). To determine whether the successful fitting of LLRs depended on our particular choice of sigmoidal E and I basis functions, we repeated the fitting procedure using 3 different sets of E and I basis functions and obtained similar results (Figure 6), indicating that the cell-cell interactions needed to detect object boundaries in natural images are easily produced by this general type of compound E-I, or “incitatory” circuit.

We next looked for regularities in the progression of excitatory-inhibitory curve pairs used to fit the LLRs as a function of a neighbor cell's offset in position and orientation from the RL (Figure 5D). We observed the following patterns. First, as the neighbor's orientation offset from the RL increases and approaches 90 degrees (indicated by lightness changes within each plot), excitation becomes weaker, and inhibition becomes both stronger and lower in threshold, resembling cross-orientation suppression (a staple function of V1 (Bishop, Coombs, & Henry, 1973; DeAngelis, Robson, Ohzawa, & Freeman, 1992; Wilson S. Geisler & Albrecht, 1992); though see Priebe & Ferster, 2006). Second, we observed a gradual weakening of both excitation and inhibition as a neighbor cell moves further from

the RL in the direction perpendicular to the boundary orientation (different plot columns), reflecting the expected decline in informativeness as a neighbor cell moves further from the boundary cell in question. To probe this effect further, we characterized each excitatory and inhibitory curve by its gain parameter and plotted the gains separately as a function of a neighbor's orientation difference and spatial offset relative to the RL (Figure 5E). These surfaces confirm that the strength of excitation and inhibition vary systematically for different neighbor cells, but in a non-obvious way that if uncovered experimentally, would be difficult to interpret without an underlying theory of boundary detection and the natural image data to support it.

**Learning the parameters of the incitation circuit**

We showed that the incitatory interconnection circuit of Figure 5A is capable of producing the diverse forms of simple cell-boundary cell interactions needed to compute boundary probability. However, the circuit contains parameters that would need to be set, presumably during development, to allow each simple cell to exert the appropriate effect on every surrounding boundary cell. We asked whether these cell-cell interactions could be learned by a simple, biologically plausible synaptic learning rule operating in the slightly elaborated incitation circuit shown in Figure 7A. In particular, we assumed that each of the 300 oriented receptive fields surrounding a boundary cell is represented by a population of 100 simple cells, all sharing the same underlying linear filter, but each having a different firing threshold and gain reflecting natural variations in neuron size, morphology, firing dynamics, etc. Three examples of filters (pink, blue and green) and their associated simple cell variants are depicted schematically in Figure 7A. Labeled image patches containing boundaries and non-boundaries were presented to the 30,000 simple cells, ground-truth labels from our natural image dataset were presented to the boundary cell (1 for boundary, 0 for no boundary), and the synapses between the

simple cells and the boundary cell and its associated inhibitory neuron were adjusted using a single-layer "delta" rule (Widrow & Hoff, 1960). We then performed virtual neurophysiology to probe the net effect of each underlying oriented filter on the boundary cell's response, induced by that filter's 100 simple cell variants each acting through its learned synaptic weight. These learned "LLR-like functions" (Figure 7B) again included monotonic excitatory and inhibitory as well as non-monotonic mixed E-I effects whose overall pattern was similar to the explicitly calculated LLRs in Figure 2; the averaged LLRs from Figure 2 are included as black dashed curves in Figure 7B. To assess whether the learning of LLR-like functions from natural images depended on having so many (i.e. 100) simple cell variants per receptive field, we repeated the learning experiment with just 10 simple cell variants per RF, in this case differing only in their firing thresholds (rather than both threshold and gain). The resulting LLR-like curves showed signs of the threshold discretization, but were otherwise essentially the same (Figure 8). Thus, the incitatory interconnection scheme depicted in Figure 7A can learn to produce the spectrum of SC-BC interactions needed for boundary detection using a small number of neurons covering each receptive field location.

In one interesting difference between the learned SC-BC interaction functions compared to the explicitly calculated LLRs, we noted that the learned LLR-like functions also included rightside up U-shaped interactions (e.g. see purple curves in lower right corner of Figure 7B). This and other more subtle differences relative to the literal LLRs can be attributed to the fact that a delta rule attempts to compensate for statistical dependencies between input features, whereas the literal LLRs shown in Figure 2 reflect the simplifying assumption that each oriented filter contributes independently to the response of a boundary cell. The overall similarity of the pattern of learned cell-cell interactions compared to the literal LLRs, however, validates the Bayesian-inspired reverse-engineering approach

to study cell-cell interactions in V1, in that Bayes rule allows us to convert intuitive labels assigned to natural images into predictions as to how cortical neurons should influence each other to solve a particular task (Figure 1).

**Comparing boundary detection performance of five models**

As an additional performance-based approach to evaluating the trained incitation circuit as a model of boundary detection in V1, we compared the Precision-Recall curves of 5 different boundary detectors: (1) the "null hypothesis", consisting of a single conventional simple cell centered at the RL (Figure 7C, blue curve); (2) a straight (unweighted) sum of 7 carefully chosen literal LLRs (Figure 7C, cyan curve) – this is essentially a direct application of Bayes rule as shown in Figure 1D (see Ramachandra and Mel 2012); (3) a *weighted* sum of 300 literal LLRs (again corresponding to the model of Figure 1D, but augmented with weights optimized by the above learning rule), and (4, 5) two learned neuromorphic classifiers as illustrated in Figure 7A with either 100 or 10 simple cell variants per oriented RF (Figure 7C, green and purple curves, respectively). We note that learning is possible even when all modifiable weights are constrained to be positive, so that the model does not require either that weights change sign, or that inhibitory weights are modifiable, both of which are difficult to justify biologically. The results lead us to 4 conclusions: (1) the superior performance of all 4 multi-input classifier variants compared to a single conventional simple cell reinforces the point that individual simple cells are poor quality boundary detectors, but can be significantly improved upon using neurally plausible local circuit computations; (2) the superior performance of all 3 classifier variants that exploit populations of input (300, 3000, or 30000) cells with optimized weights, compared to a classifier with a few carefully chosen unweighted inputs (Figure 7C, cyan curve), points to the value of exploiting information from all around the boundary cell's receptive field – as long as a learning mechanism is available to customize

the interaction between the lower order cells that provide the evidence, and the higher order cell that performs the classification; (3) the similar performance of the 2 learned neuromorphic classifiers compared to a weighted sum of literal LLRs (Figure 7C, red curve) reinforces the point made above regarding the close connection between Bayes rule and the circuit of Figure 7A; and (4) the similar performance of the two learned neuromorphic classifiers to each other, despite one having an order of magnitude more cells than the other, indicates that the incitation circuit of Figure 7A can extract the critical information already from relatively few neurons (Figure 7C, green and purple curves).

In light of these comparisons, we conclude that the requirements for developing a cortical circuit that significantly improves boundary detection performance compared to a lone simple cell are modest, including mainly (1) a compound E-I circuit motif known to exist in V1; (2) natural variation in firing thresholds across the population of simple cells; and (3) a single-layer synaptic learning rule to adjust the circuit parameters. In the cortex, the modifiable parameters could be the strengths of synaptic contacts on the dendrites of different boundary cells and/or interneurons, as shown here, or perhaps the dendritic locations of those synaptic contacts[22–25], and/or a choice among interneurons having different gains and thresholds (Druckmann, Hill, Schürmann, Markram, & Segev, 2013; Markram et al., 2004).

**Experimentally distinguishing boundary cells from conventional simple cells**

Having shown above that V1 circuitry is capable of producing boundary cells from simple cells in principle, we next addressed the question as to how BCs could be detected experimentally, and distinguished from conventional simple cells (or the simple cell-like subunits of complex cells – Hubel & Wiesel, 1962b; Movshon, Thompson, & Tolhurst, 1978; Ohzawa, DeAngelis, & Freeman, 1997).

To determine how BCs would respond to conventional stimuli, we first presented drifting sine wave gratings covering a BC's "classical receptive field", leading to the unremarkable orientation tuning curve shown in Figure 9A. When gratings were added to the surround at varying orientations while the center grating was held at the preferred orientation, the overall effect on the boundary cell's response was suppressive, as indicated by the small circle at the preferred orientation. This suppressive effect arose mainly from the divisive normalization operation applied to every image patch prior to evaluating boundary probability: the presence of the surround grating increased the normalizer value, which, post-normalization, led to a reduced effective contrast at the RF center. This in turn lowered the BC's estimate of boundary probability. Beyond showing this generally suppressive effect of the surround, we were not in a position to make specific predictions regarding the *orientation tuning* of surround suppression in BCs, since any such tuning would flow directly from the orientation tuning of the normalizer we used in our model, which was neutral and built in by assumption. (The mechanisms of divisive normalization were not our focus in the work). We were likewise unable to generate contrast response curves using conventional grating stimuli, or to explore contrast-invariance of orientation tuning, since gratings of any contrast are identical up to a scaling factor, and that factor would have been perfectly cancelled by the simple normalization scheme we used. On the other hand, we were able to use labeled natural edges selected from a fixed normalizer bin (i.e. in which all labeled edge patches had the same normalizer score) to explore the effect of increasing center contrast on orientation tuning curve width. (This was not a perfectly controlled experiment because variations in center contrast in a fixed normalizer bin would have led to *anti*-variations in surround contrast, but given the filter value at the RF center was only one of 100 filters of many orientations used to compute the normalizer value, this effect was likely to be small). Subject to this limitation, as shown in Figure

9B, the boundary cell's tuning width is essentially constant across a roughly 2-fold change in center contrast (the limit of analysis allowed by our labeled database). In an additional experiment using labeled natural edges, this time with center contrast held fixed, but drawing the image patches from two different normalizer bins separated by a factor of 2 in the normalizer value, we saw that the effect of increasing surround contrast was again to suppress boundary probability, just as in Figure 9A. Unlike the case in Figure 9A where sinusoidal gratings were used, the natural image patches evaluated by the boundary cell in this case all had identical center contrast values *after* normalization. Thus, the suppressive effect of the surround in this case was a bona fide probability effect flowing from natural image statistics, rather than resulting from an overt reduction of center contrast by the divisive normalization operator.

To summarize our predictions using oriented edges and gratings with and without surrounds, boundary cells behave similarly to conventional simple cells in that they are (1) orientation tuned, (2) show surround suppression, and (3) have tuning curves whose widths are roughly contrast invariant (Alitto & Usrey, 2004). It is therefore possible that boundary cells have classified as conventional simple cells in previous experiments using simplified stimuli. Among the multiple types of V1 cells that have been previously described, boundary cells share most in common with double opponent cells, which are orientation tuned, have mostly odd-symmetric receptive field profiles as would be expected for boundary detecting cells (Ringach, 2002), and respond to boundaries whether defined by luminance or color (Johnson, Hawken, & Shapley, 2008).

In future neurophysiological studies, an efficient means of dissociating conventional simple cells, which respond to oriented contrast independent of boundary probability, from boundary cells, which

respond to boundary probability independent of oriented contrast, would be to use natural image stimuli drawn from the four corners of the oriented contrast – boundary probability space (Figure 10A). Image patches with low oriented contrast and low boundary probability scores (purple dots) tend to be flat, unstructured image regions; patches with low contrast and high probability (green dots) tend to be well-structured, faint edges; patches with high contrast but low probability (blue dots) tend to consist of contrasty noise or misaligned edges; and regions with high contrast *and* high probability (red dots) are typically well-structured, strong edges (Figure 10B). This factorial stimulus set would make it possible to identify pure simple cells, pure boundary cells, as well as cells of intermediate type.

## DISCUSSION

In the 60 years since Hubel and Wiesel first discovered orientation-tuned simple cells in V1, it has been generally assumed that these cells contribute in some way to the detection of object boundaries (Angelucci et al., 2002; Field, Hayes, & Hess, 1993; Grosof, Shapley, & Hawken, 1993; Kapadia, Ito, Gilbert, & Westheimer, 1995b; Kapadia, Westheimer, & Gilbert, 2000a; Polat, Mizobe, Pettet, Kasamatsu, & Norcia, 1998; Sceniak, Ringach, Hawken, & Shapley, 1999). Consistent with this idea, virtually every modern object recognition system, whether designed by hand or trained from natural image data, includes simple cell-like filtering in its early stages of processing (Fukushima, Miyake, & Ito, 1983; Krizhevsky, Sutskever, & Hinton, 2012; Lades et al., 1993; Lecun, Bottou, Bengio, & Haffner, 1998; Mel, 1997; Riesenhuber & Poggio, 1999). Surprisingly, however, the quantitative relationship between simple cell responses, typically modeled as divisively normalized linear filters (Carandini & Heeger, 2012), and object boundary probability in natural images has been little explored (though see

Ramachandra & Mel, 2013), making it difficult to know whether or how V1 circuits contribute to this behaviorally relevant natural computation. It is important to emphasize that a simple cell on its own is a poor detector of natural object boundaries within its receptive field (see also Arbelaez, Maire, Fowlkes, & Malik, 2011): as shown in Figure 7C (see blue curve), if we use a simple cell's response as an indicator of the presence of an object boundary within its RF, even when the threshold for detection is raised to such a high value (moving upward and leftward along the PR curve) that half of all true boundaries are missed (corresponding to a Recall score of 50%), more than 60% of the "detected" edges at that threshold will be false positives (corresponding to a Precision score of 40%). The reason a simple cell is such an unreliable edge detector is that true object boundaries are rare, and when they do occur, they are mostly of low contrast. Much more common are high contrast non-edge structures (e.g. textures) that contain sufficient oriented energy to strongly drive simple oriented filters.

The poor boundary detection performance of a lone simple cell leads to the hopeful conjecture that V1 also contains "smarter" cells that compute boundary probability by combining the responses of multiple simple cells covering a local neighborhood. In a previous study, we suggested that the appropriate strategy for constructing a boundary cell from a local population of simple cells was to (1) select a small set of simple cells (e.g. 6 cells) that were both individually informative and class-conditionally independent (CCI; see methods for discussion of the CCI assumption); (2) evaluate the log-likelihood ratios for each of the participating simple cells, which describe the functional connections between each simple cell and the boundary cell; and (3) sum the LLRs and pass the total through a sigmoidal nonlinearity to compute boundary probability (Ramachandra & Mel, 2013) (Figure 1B). The present study extends that previous work in four ways: (1) here we collected and analyzed LLRs for *all* of the simple cells at all orientations covering a 5x5 pixel neighborhood in the vicinity of a

boundary cell's RF (300 cells total); (2) we show that the functional connections between SCs and the BC depend systematically on the relative position and orientation of the SC and BC receptive fields, and come in a variety of forms including monotonic excitation and inhibition and various bump-shaped functions; (3) we show that a mixed excitatory-inhibitory, or "incitatory", circuit motif that is known to exist in V1 is capable of producing the entire spectrum of simple cell–boundary cell interactions that we catalogued, and (4) we show that the synaptic weights that parameterize the incitation circuit can be learned by one of the simplest known developmental/plasticity rules: a 1-layer "delta" rule.

**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 & Sejnowski, 1995; Laughlin, 1989; Olshausen & Field, 1996; Schwartz & Simoncelli, 2001). 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 boundary 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 goal 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 the 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 problem of discriminating object boundaries from non-boundaries based on hundreds of simple cell responses collected from a local neighborhood of 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 data consists of *functions* linking pairs of locations/orientations, which describe how a given simple cell should influence a nearby boundary cell as a part of the 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 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", or represented by scalar 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 7B). 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 (M. P. Jadi & Sejnowski, 2014; Ozeki, Finn, Schaffer, Miller, & Ferster, 2009), which accounts for a number of V1 cell

response properties, "non-binary" interactions between cells would 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, often subject to simple geometric rules. The notion of "surround suppression", for example, reflects both of these tendencies (Adesnik, Bruns, Taniguchi, Huang, & Scanziani, 2012; Cavanaugh, Bair, & Movshon, 2002; Schwabe, Ichida, Shushruth, Mangapathy, & Angelucci, 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 notion that cell-cell interactions have a defined polarity often persists. 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 & Bressloff, 2006); and several studies support the idea that cortical cells affect each other laterally through bowtie-shaped "extension fields" consisting of patterned arrays of positive and negative coefficients (e.g. Bosking, Zhang, Schofield, & Fitzpatrick, 1997; Field et al., 1993; W. S. Geisler et al., 2001; Kapadia, Westheimer, & Gilbert, 2000b; Li, 1999; Sigman et al., 2001). In all of these cases, one neuron's effect on another neuron are described in terms of its scalar connection "strength".

Not all functional interconnections that have been described in the cortex fit such simple descriptions, however. Examples of activity-level-dependent interactions have been reported, where the strength and even polarity of the connection between cells depends on the activity levels of the sending and/or receiving cells. For example, the responses of amplitude-tuned neurons in the auditory cortex grow stronger as the sound pressure level increases up to an optimal intensity level, and then are progressively inhibited as the sound grows louder (Suga & Manabe, 1982); in V1, surround modulation can switch from facilitating to suppressive with increasing center contrast (Ichida, Schwabe, Bressloff,

& Angelucci, 2007; Nauhaus, Busse, Carandini, & Ringach, 2009; Polat et al., 1998; Schwabe, Obermayer, Angelucci, & Bressloff, 2006; Somers et al., 1998); length-tuned neurons respond best to an oriented stimulus up to a certain length, but are then progressively inhibited as the stimulus grows longer (Anderson, Lampl, Gillespie, & Ferster, 2001); and non-monotonic modulatory interactions between a neuron's classical and extra-classical receptive fields have been reported (Polat et al., 1998). These data, though unaccompanied by normative explanations, reinforce the idea that the sign and magnitude one neuron's effect on another can depend not only on the relative position and orientation of their receptive fields (in the case of vision), but also on their relative activity levels.

Our paper represents a fleshing out of this type of effect, and is to our knowledge the first normative theory, parameterized by natural images, that specifies how low-order cells should affect higher-order cells in the cortex to solve a specific, biologically-relevant classification problem. By analyzing natural image data on and off object boundaries, we showed that the local cell-cell interactions needed to solve this classification problem 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. And while such connections cannot (except in special cases) be described by scalar weights, we showed that they are easily produced by a compound E-I circuit motif (see Figure 5) that is known to exist in the cortex (Buzsáki, 1984; Isaacson & Scanziani, 2011; Klyachko & Stevens, 2006; McBain & Fisahn, 2001; Pfeffer et al., 2013; Pouille & Scanziani, 2001; Swadlow, 2002; Wehr & Zador, 2003), and that the synaptic weights that control the net effect of the "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.

**Materials and methods:**

*Data gathering*

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 original 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. To minimize filter pairwise correlations, the selection was restricted to patches with a certain fixed "energy" score, calculated by summing the absolute values of 100 simple cell-like linear filters surrounding the edge location (see below for filter specifications). An elongated box was then drawn on each patch, representing the "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 through the reference box?" 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. Our informal observations, based in part on occasions when two labelers worked together, was that the agreement was very high. Rare ambiguous patches that could cause labeler disagreement were often conservatively labelled as 3 and consequently excluded from later analyses. After labeling, the dataset was doubled by adding left-right flipped versions of each patch, and assigning the same label as the unflipped counterpart.

*Extracting the LLRs*

The original color image patches were first converted to single-channel intensity images $(0.29\ R + 0.59\ G + 0.11\ B)$. Simple cell-like filters at multiple orientations were created by rotating a 2x4 pixel rectangular filter (see $f_1$ in Figure 1B) in 15° increments from 0 to 165°. Interpolation was bilinear. The intensity images were filtered at those 12 orientations over all positions on a 5x5 pixel lattice centered at the reference location. This resulted in 25 positions x 12 orientations = 300 total simple cell responses for each patch. (Given the symmetric form of the filter mask, filter values for orientations from 180° to 345° were the negatives of the first 12 orientations so did not need to be explicitly computed). Patches with a negative reference filter score were vertically flipped so that the reference filter score was positive. Histograms were collected for each of the 300 filters separately for "yes" patches (using 8 to 20 evenly spaced bins depending on the smoothness of the histogram) and "no" patches (using 50 evenly spaced bins). The "yes" and "no" histograms for each filter were normalized to probability distributions. LLRs were then computed as $\log\left(\frac{p_{yes}}{p_{no}}\right)$, where $p_{yes}$ and $p_{no}$ are the boundary and non-boundary pdfs, respectively. To control noise, for each filter, LLR analysis was restricted to a central set of filter values where $p_{yes} > 0.005$ and $p_{no} > 0.002$. Only data inside this region is plotted in Figs. 2-4, Figure 6. The same procedure was repeated with different filters (2x6, 2x8, 4x8, 6x8) to generate the LLR curves shown in Figure 3.

*Class-conditional independence (CCI)*

The assumption of class-conditional independence between nearby filters does not hold in general in natural images, so that the strict application of a naive Bayesian approach that assumes CCI among filters is not expected to perform very well – and in our experiments, if filters are chosen randomly, the

approach performs poorly (often worse than using a single filter). However, either of two simple strategies, both biologically plausible, can mitigate the problems arising from the violation of CCI, allowing a local population of filter outputs mapped through their respective LLRs to be combined linearly to form a much better boundary classifier than a single simple cell. The first strategy is to select small subsets of filters from the neighborhood whose responses are relatively uncorrelated with each other; this was the approach taken in Ramachandra and Mel (2013). In that earlier paper, in addition to selecting de-correlated filters, we used additional ad hoc criteria, such as that the filters should be individually informative about the presence of an object boundary, and that the resulting classifier should be sharply oriented tuned. In principle, a neural learning rule could have been used to select for these criteria, but in that earlier paper a neural model was not defined, a learning rule was not tried, and boundary detection performance was not quantitatively measured. The second strategy, and the one that we pursue in this paper, is to use a learning rule to modify a single layer of excitatory weights in an incitatory circuit, without need for any filter pre-selection (Figure 7A).

*Sigmoid modeling*

Filter LLRs were fit by a difference of 2 sigmoids of the form $s(x) = \frac{A}{1+\exp[-g(x-t)]}$ using a semi-automatic approach. For each LLR, an approximate amplitude $A$, gain $g$, and threshold $t$ for the two sigmoids was chosen automatically, and the parameters were then adjusted by hand so that the difference of the two sigmoids visually matched the LLR as closely as possible. We found visually-guided optimization better captured the essential shape structure of the LLR compared to generic simple error measures such as MSE. A similar fitting procedure was used for the three models in Figure 6 (model details shown in figure). 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 shapes using a difference of two simple functions is mainly of didactic interest; the more practically significant question is whether a weighted sum of simple E and I functions (which will in general involve more than two curves) can produce the LLR-like interactions needed to drive down classification errors during learning (see Figure 7). For the surfaces in Figure 5E, and Figure 6, excitatory gain was computed by measuring the excitatory component's average slope between f = 0 and f = 10. Inhibitory gain in Figure 5 was computed in the same way. In Figure 6, each of the inhibition familes had only a single parameter; this parameter is what is plotted in the inhibitory gain surface plots.

*Learning experiments*

For each patch, each filter value was passed through 100 different sigmoidal nonlinearities (using the sigmoid functional form as above, taking a grid of 10 sigmoid thresholds x 10 sigmoid gains = a total of 100 sigmoids) (or 10; see Figure 8), simulating the responses of 100 simple cells with identical receptive fields but slightly different "F-I curves" (i.e. output nonlinearities). The result was 300 filters x 100 nonlinear variants = 30,000 model simple cell responses per patch. We then used logistic regression to train a linear classifier to distinguish boundary from non-boundary patches using the simple cell responses. A subset of the data (25,000 of the ~30,000 labeled patches) was used for training. During training, data was balanced by duplicating boundary-containing patches such that boundary and non-

boundary exemplars were equal in number. Training was done using batch gradient descent with a learning rate of $\eta = 0.1$, performed for $1{,}000$ iterations. The resulting parameter fits were visualized in Figure 7B by systematically increasing a filter value (x axes) while holding all other filters constant, and plotting the changes in linear classifier score resulting from that filter's 100 nonlinear simple cells combined using the learned weights (y axes). In order to allow shape comparison of the filters' learned interactions with the corresponding LLR function, we scaled the colored interactions within each plot. Each plot has one scaling factor that applies to all 5 colored curves in the plot. The inverse of this scaling factor, which can be thought of equivalently as the weight that the classifier puts on the curves drawn in the subplot, is shown by the grey bars.

*Precision-Recall curves*

Precision-Recall (PR) curves were generated for the learned boundary cell classifier, as well as for the sum of filter LLRs used as a classifier (Figure 1). A classifier consisting of a single linear filter provided the PR baseline (Figure 7). To generate a PR curve, a classifier was applied to each of the 5,000 labeled (untrained) test patches, and the patches were sorted by their boundary probability (BP) scores. A threshold was set at the lowest BP value obtained over the entire test set, and was systematically increased until the highest boundary probability score in the test set was reached. For every possible threshold, above-threshold patches were called putative boundaries and below-threshold patches were called putative non-boundaries, (1) "Precision" was calculated by asking what fraction of patches identified as putative boundaries contained true boundaries (according to their human assigned labels), and (2) "Recall" was calculated by asking what fraction of true boundaries were identified as putative boundaries. As the threshold increased, the P-R values swept out a curve in Precision-Recall space. Perfect performance would be 100% Precision and Recall simultaneously, corresponding to the top right corner of the PR graph. The aqua curve labelled "Sum of 7 random filter LLRs" was generated using the following procedure: 1000 random sets of 7 filters were created. For each set, a single PR curve was generated for a boundary detector consisting of the literal sum of the 7 filter LLR scores on each patch. After all 1000 curves had been generated, the alpha curve was plotted showing the 99$^{th}$ percentile of precision over all PR curves for each recall.

*Stimulus response curves*

Grating stimuli were generated by sampling a sinusoidal grating wave on a 20x20 pixel grid. Frequency was chosen at 0.25 cycles/pixel because it led to relatively artifact-free stimuli and evoked robust boundary cell responses. Gratings were then windowed to a 2 pixel radius around the center of the patch. Gratings were presented to the boundary cell at every 15° of orientation. At each orientation, response was averaged over all phases of the grating. Surround stimuli were created by adding a second grating outside the 2 pixel radius window. Suppression was measured as the average response to the surround stimulus over all phases of the inside and outside gratings, and the orientation of the surround (Figure 9A).

To measure the effects of center and surround contrast on boundary cell response in natural images (Figure 9B), we collected patches with fixed surround contrast and varying center contrast (first plot) and fixed center contrast and varying surround contrast (second plot). Center contrast was taken to be the value of the central reference filter. Surround contrast was taken to be the patch energy used to normalize patches in earlier experiments (see Methods section *Data Gathering*). Stimuli were rotated

in increments of 15° and presented to the boundary cells. Individual tuning curves are plotted, along with the average tuning curve and its standard error.

**References**

Adesnik, Hillel, William Bruns, Hiroki Taniguchi, Z Josh Huang, and Massimo Scanziani. (2012). "A Neural Circuit for Spatial Summation in Visual Cortex." *Nature* 490 (7419): 226–31. https://doi.org/10.1038/nature11526.

Alitto, Henry J., and W. Martin Usrey. (2004). "Influence of Contrast on Orientation and Temporal Frequency Tuning in Ferret Primary Visual Cortex." *Journal of Neurophysiology* 91 (6): 2797–2808. https://doi.org/10.1152/jn.00943.2003.

Anderson, Jeffrey S., Ilan Lampl, Deda C. Gillespie, and David Ferster. (2001). "Membrane Potential and Conductance Changes Underlying Length Tuning of Cells in Cat Primary Visual Cortex." *Journal of Neuroscience* 21 (6): 2104–12.

Angelucci, Alessandra, and Paul C. Bressloff. (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*, 154:93–120. Elsevier.

Angelucci, Alessandra, Jonathan B. Levitt, Emma J. S. Walton, Jean-Michel Hupé, Jean Bullier, and Jennifer S. Lund. (2002). "Circuits for Local and Global Signal Integration in Primary Visual Cortex." *The Journal of Neuroscience* 22 (19): 8633–46.

Arbelaez, P., M. Maire, C. Fowlkes, and J. Malik. (2011). "Contour Detection and Hierarchical Image Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (5): 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." *Proceedings of the Royal Society of London. Series B. Biological Sciences* 212 (1186): 1–34. https://doi.org/10.1098/rspb.1981.0022.

Behabadi, Bardia F, Alon Polsky, Monika Jadi, Jackie Schiller, and Bartlett W Mel. (2012). "Location-Dependent Excitatory Synaptic Interactions in Pyramidal Neuron Dendrites." *PLoS Computational Biology* 8 (7): e1002599. https://doi.org/10.1371/journal.pcbi.1002599.

Bell, A J, and T J Sejnowski. (1995). "An Information-Maximization Approach to Blind Separation and Blind Deconvolution." *Neural Computation* 7 (6): 1129–59.

Biederman, Irving. (1987). "Recognition-by-Components: A Theory of Human Image Understanding." *Psychological Review* 94 (2): 115–147.

Bishop, P. O., J. S. Coombs, and G. H. Henry. (1973). "Receptive Fields of Simple Cells in the Cat Striate Cortex." *The Journal of Physiology* 231 (1): 31–60. https://doi.org/10.1113/jphysiol.1973.sp010218.

Bosking, William H., Ying Zhang, Brett Schofield, and David Fitzpatrick. (1997). "Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex." *Journal of Neuroscience* 17 (6): 2112–27.

Buzsáki, G. (1984). "Feed-Forward Inhibition in the Hippocampal Formation." *Progress in Neurobiology* 22 (2): 131–53.

Carandini, Matteo, and David J. Heeger. (2012). "Normalization as a Canonical Neural Computation." *Nature Reviews Neuroscience* 13 (1): 51–62. https://doi.org/10.1038/nrn3136.

Cavanaugh, James R., Wyeth Bair, and J. Anthony Movshon. (2002). "Selectivity and Spatial Distribution of Signals from the Receptive Field Surround in Macaque V1 Neurons." *Journal of Neurophysiology* 88 (5): 2547–56. https://doi.org/10.1152/jn.00693.2001.

DeAngelis, G. C., J. G. Robson, I. Ohzawa, and R. D. Freeman. (1992). "Organization of Suppression in Receptive Fields of Neurons in Cat Visual Cortex." *Journal of Neurophysiology* 68 (1): 144–63.

Druckmann, Shaul, Sean Hill, Felix Schürmann, Henry Markram, and Idan Segev. (2013). "A Hierarchical Structure of Cortical Interneuron Electrical Diversity Revealed by Automated Statistical Analysis." *Cerebral Cortex* 23 (12): 2994–3006. https://doi.org/10.1093/cercor/bhs290.

Field, David J., Anthony Hayes, and Robert F. Hess. (1993). "Contour Integration by the Human Visual System: Evidence for a Local 'association Field.'" *Vision Research* 33 (2): 173–93. https://doi.org/10.1016/0042-6989(93)90156-Q.

Fukushima, K., S. Miyake, and T. Ito. (1983). "Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition." *IEEE Transactions on Systems, Man and Cybernetics* SMC-13 (5): 826–34. https://doi.org/10.1109/TSMC.1983.6313076.

Geisler, W. S., J. S. Perry, B. J. Super, and D. P. Gallogly. (2001). "Edge Co-Occurrence in Natural Images Predicts Contour Grouping Performance." *Vision Research* 41 (6): 711–24.

Geisler, Wilson S., and Duane G. Albrecht. (1992). "Cortical Neurons: Isolation of Contrast Gain Control." *Vision Research* 32 (8): 1409–10. https://doi.org/10.1016/0042-6989(92)90196-P.

George, Andrew A., Ariel M. Lyons-Warren, Xiaofeng Ma, and Bruce A. Carlson. (2011). "A Diversity of Synaptic Filters Are Created by Temporal Summation of Excitation and Inhibition." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 31 (41): 14721–34. https://doi.org/10.1523/JNEUROSCI.1424-11.2011.

Gilbert, Charles D., and Torsten N. Wiesel. (1990). "The Influence of Contextual Stimuli on the Orientation Selectivity of Cells in Primary Visual Cortex of the Cat." *Vision Research*, Optics Physiology and Vision A Festschrift Honoring Professor Gerald Westheimer on His 65th Birthday, 30 (11): 1689–1701. https://doi.org/10.1016/0042-6989(90)90153-C.

Grosof, D H, R M Shapley, and M J Hawken. (1993). "Macaque V1 Neurons Can Signal 'Illusory' Contours." *Nature* 365 (6446): 550–52. https://doi.org/10.1038/365550a0.

Heydt, R. von der, and E. Peterhans. (1989). "Mechanisms of Contour Perception in Monkey Visual Cortex. I. Lines of Pattern Discontinuity." *The Journal of Neuroscience* 9 (5): 1731–48.

Hoffman, Donald D. (2000). *Visual Intelligence: How We Create What We See*. WW Norton & Company. https://books.google.com/books?hl=en&lr=&id=_2iBn_G3E7MC&oi=fnd&pg=PR9&dq=david+hoffman+visual+intelligence&ots=abym9UY5d5&sig=2qNW0k6I-4bd5iPzlrRGWWOwbto.

Hubel, D H, and T N Wiesel. (1962a). "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex." *The Journal of Physiology* 160 (January): 106–54.
1962b. "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex." *The Journal of Physiology* 160 (January): 106–54.

Ichida, Jennifer M., Lars Schwabe, Paul C. Bressloff, and Alessandra Angelucci. (2007). "Response Facilitation From the 'Suppressive' Receptive Field Surround of Macaque V1 Neurons." *Journal of Neurophysiology* 98 (4): 2168–81. https://doi.org/10.1152/jn.00298.2007.

Isaacson, Jeffry S, and Massimo Scanziani. (2011). "How Inhibition Shapes Cortical Activity." *Neuron* 72 (2): 231–43. https://doi.org/10.1016/j.neuron.2011.09.027.

Jadi, M. P., and T. J. Sejnowski. (2014). "Regulating Cortical Oscillations in an Inhibition-Stabilized Network." *Proceedings of the IEEE* 102 (5): 830–42. https://doi.org/10.1109/JPROC.2014.2313113.

Jadi, Monika, Alon Polsky, Jackie Schiller, and Bartlett W Mel. (2012). "Location-Dependent Effects of Inhibition on Local Spiking in Pyramidal Neuron Dendrites." *PLoS Computational Biology* 8 (6): e1002550. https://doi.org/10.1371/journal.pcbi.1002550.

Johnson, Elizabeth N., Michael J. Hawken, and Robert Shapley. (2008). "The Orientation Selectivity of Color-Responsive Neurons in Macaque V1." *Journal of Neuroscience* 28 (32): 8096–8106. https://doi.org/10.1523/JNEUROSCI.1404-08.2008.

Kapadia, Mitesh K., Minami Ito, Charles D. Gilbert, and Gerald Westheimer. (1995a). "Improvement in Visual Sensitivity by Changes in Local Context: Parallel Studies in Human Observers and in V1 of Alert Monkeys." *Neuron* 15 (4): 843–56. https://doi.org/10.1016/0896-6273(95)90175-2. 1995b. "Improvement in Visual Sensitivity by Changes in Local Context: Parallel Studies in Human Observers and in V1 of Alert Monkeys." *Neuron* 15 (4): 843–56. https://doi.org/10.1016/0896-6273(95)90175-2.

Kapadia, Mitesh K., Gerald Westheimer, and Charles D. Gilbert. (2000a). "Spatial Distribution of Contextual Interactions in Primary Visual Cortex and in Visual Perception." *Journal of Neurophysiology* 84 (4): 2048–62. 2000b. "Spatial Distribution of Contextual Interactions in Primary Visual Cortex and in Visual Perception." *Journal of Neurophysiology* 84 (4): 2048–62.

Klyachko, Vitaly A., and Charles F. Stevens. (2006). "Excitatory and Feed-Forward Inhibitory Hippocampal Synapses Work Synergistically as an Adaptive Filter of Natural Spike Trains." *PLOS Biol* 4 (7): e207. https://doi.org/10.1371/journal.pbio.0040207.

Koch, C, T Poggio, and V Torre. (1983). "Nonlinear Interactions in a Dendritic Tree: Localization, Timing, and Role in Information Processing." *Proceedings of the National Academy of Sciences of the United States of America* 80 (9): 2799–2802.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. (2012). "ImageNet Classification with Deep Convolutional Neural Networks." In *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, 1097–1105. Curran Associates, Inc. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

Lades, M., J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen. (1993). "Distortion Invariant Object Recognition in the Dynamic Link Architecture." *IEEE Transactions on Computers* 42 (3): 300–311. https://doi.org/10.1109/12.210173.

Laughlin, S. B. (1989). "The Role of Sensory Adaptation in the Retina." *Journal of Experimental Biology* 146 (1): 39–62.

Lecun, Y., L. Bottou, Y. Bengio, and P. Haffner. (1998). "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE* 86 (11): 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 (Bristol, England)* 10 (2): 187–212.

Major, Guy, Alon Polsky, Winfried Denk, Jackie Schiller, and David W. Tank. (2008). "Spatiotemporally Graded NMDA Spike/Plateau Potentials in Basal Dendrites of Neocortical Pyramidal Neurons." *Journal of Neurophysiology* 99 (5): 2584–2601. https://doi.org/10.1152/jn.00011.2008.

Markram, Henry, Maria Toledo-Rodriguez, Yun Wang, Anirudh Gupta, Gilad Silberberg, and Caizhi Wu. (2004). "Interneurons of the Neocortical Inhibitory System." *Nature Reviews Neuroscience* 5 (10): 793–807. https://doi.org/10.1038/nrn1519.

McBain, C. J., and A. Fisahn. (2001). "Interneurons Unbound." *Nature Reviews. Neuroscience* 2 (1): 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 Computation* 9 (4): 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., I. D. Thompson, and D. J. Tolhurst. (1978). "Spatial and Temporal Contrast Sensitivity of Neurones in Areas 17 and 18 of the Cat's Visual Cortex." *The Journal of Physiology* 283 (October): 101–20.

Nauhaus, Ian, Laura Busse, Matteo Carandini, and Dario L. Ringach. (2009). "Stimulus Contrast Modulates Functional Connectivity in Visual Cortex." *Nature Neuroscience* 12 (1): 70–76. https://doi.org/10.1038/nn.2232.

Ohzawa, I, G C DeAngelis, and R D Freeman. (1997). "Encoding of Binocular Disparity by Complex Cells in the Cat's Visual Cortex." *Journal of Neurophysiology* 77 (6): 2879–2909.

Olshausen, Bruno A., and David J. Field. (1996). "Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images." *Nature* 381 (6583): 607. https://doi.org/10.1038/381607a0.

Ozeki, Hirofumi, Ian M. Finn, Evan S. Schaffer, Kenneth D. Miller, and David Ferster. (2009). "Inhibitory Stabilization of the Cortical Network Underlies Visual Surround Suppression." *Neuron* 62 (4): 578–92. https://doi.org/10.1016/j.neuron.2009.03.028.

Pfeffer, Carsten K., Mingshan Xue, Miao He, Z. Josh Huang, and Massimo Scanziani. (2013). "Inhibition of Inhibition in Visual Cortex: The Logic of Connections between Molecularly Distinct Interneurons." *Nature Neuroscience* 16 (8): 1068–76. https://doi.org/10.1038/nn.3446.

Polat, Uri, Keiko Mizobe, Mark W. Pettet, Takuji Kasamatsu, and Anthony M. Norcia. (1998). "Collinear Stimuli Regulate Visual Responses Depending on Cell's Contrast Threshold." *Nature* 391 (6667): 580–84. https://doi.org/10.1038/35372.

Pouille, F, and M Scanziani. (2001). "Enforcement of Temporal Fidelity in Pyramidal Cells by Somatic Feed-Forward Inhibition." *Science (New York, N.Y.)* 293 (5532): 1159–63. https://doi.org/10.1126/science.1060342.

Priebe, Nicholas J., and David Ferster. (2006). "Mechanisms Underlying Cross-Orientation Suppression in Cat Visual Cortex." *Nature Neuroscience* 9 (4): 552–61. https://doi.org/10.1038/nn1660.

Ramachandra, Chaithanya A, and Bartlett W Mel. (2013). "Computing Local Edge Probability in Natural Scenes from a Population of Oriented Simple Cells." *Journal of Vision* 13 (14). https://doi.org/10.1167/13.14.19.

Riesenhuber, M, and T Poggio. (1999). "Hierarchical Models of Object Recognition in Cortex." *Nature Neuroscience* 2 (11): 1019–25. https://doi.org/10.1038/14819.

Ringach, Dario L. (2002). "Spatial Structure and Symmetry of Simple-Cell Receptive Fields in Macaque Primary Visual Cortex." *Journal of Neurophysiology* 88 (1): 455–63. https://doi.org/10.1152/jn.2002.88.1.455.

Rust, Nicole C, and James J Dicarlo. (2010). "Selectivity and Tolerance ('invariance') Both Increase as Visual Information Propagates from Cortical Area V4 to IT." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30 (39): 12978–95. https://doi.org/10.1523/JNEUROSCI.0179-10.2010.

Sceniak, Michael P., Dario L. Ringach, Michael J. Hawken, and Robert Shapley. (1999). "Contrast's Effect on Spatial Summation by Macaque V1 Neurons." *Nature Neuroscience* 2 (8): 733–39. https://doi.org/10.1038/11197.

Schwabe, Lars, Jennifer M. Ichida, S. Shushruth, Pradeep Mangapathy, and Alessandra Angelucci. (2010). "Contrast-Dependence of Surround Suppression in Macaque V1: Experimental Testing of a Recurrent Network Model." *NeuroImage*, Computational Models of the Brain, 52 (3): 777–92. https://doi.org/10.1016/j.neuroimage.2010.01.032.

Schwabe, Lars, Klaus Obermayer, Alessandra Angelucci, and Paul C. Bressloff. (2006). "The Role of Feedback in Shaping the Extra-Classical Receptive Field of Cortical Neurons: A Recurrent Network Model." *Journal of Neuroscience* 26 (36): 9117–29. https://doi.org/10.1523/JNEUROSCI.1253-06.2006.

Schwartz, Odelia, and Eero P. Simoncelli. (2001). "Natural Signal Statistics and Sensory Gain Control." *Nature Neuroscience* 4 (8): 819. https://doi.org/10.1038/90526.

Sigman, Mariano, Guillermo A. Cecchi, Charles D. Gilbert, and Marcelo O. Magnasco. (2001). "On a Common Circle: Natural Scenes and Gestalt Rules." *Proceedings of the National Academy of Sciences* 98 (4): 1935–40. https://doi.org/10.1073/pnas.98.4.1935.

Somers, D. C., E. V. Todorov, A. G. Siapas, L. J. Toth, D. S. Kim, and M. Sur. (1998). "A Local Circuit Approach to Understanding Integration of Long-Range Inputs in Primary Visual Cortex." *Cerebral Cortex* 8 (3): 204–17. https://doi.org/10.1093/cercor/8.3.204.

Suga, N., and T. Manabe. (1982). "Neural Basis of Amplitude-Spectrum Representation in Auditory Cortex of the Mustached Bat." *Journal of Neurophysiology* 47 (2): 225–55.

Swadlow, Harvey A. (2002). "Thalamocortical Control of Feed–forward Inhibition in Awake Somatosensory 'barrel' Cortex." *Philosophical Transactions of the Royal Society B: Biological Sciences* 357 (1428): 1717–27. https://doi.org/10.1098/rstb.2002.1156.

Theys, Tom, Maria C. Romero, Johannes van Loon, and Peter Janssen. (2015). "Shape Representations in the Primate Dorsal Visual Stream." *Frontiers in Computational Neuroscience* 9. https://doi.org/10.3389/fncom.2015.00043.

Wehr, Michael, and Anthony M. Zador. (2003). "Balanced Inhibition Underlies Tuning and Sharpens Spike Timing in Auditory Cortex." *Nature* 426 (6965): 442–46. https://doi.org/10.1038/nature02116.

Widrow, Bernard, and Marcian E. Hoff. (1960). "Adaptive Switching Circuits." STANFORD UNIV CA STANFORD ELECTRONICS LABS.

**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.

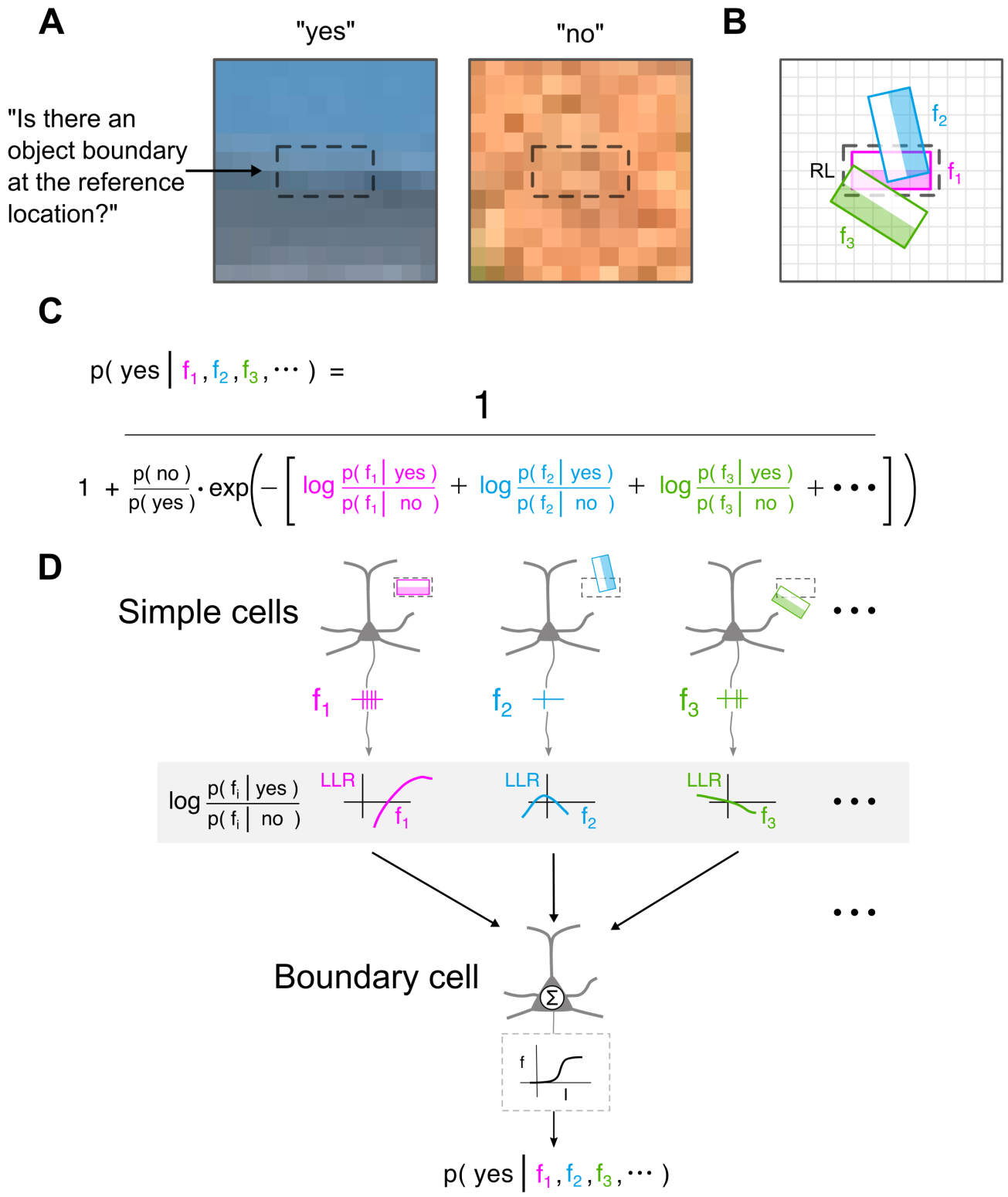


**Figure 1**

**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. (**B**) 3 (of many) oriented linear filters with responses $f_1$, $f_2$, $f_3$ are shown in the vicinity of the RL. (**C**) Under the assumption that filters are class-conditionally independent (see Methods), Bayes' rule gives an expression for boundary probability in terms of individual filter log-likelihood ratios (LLRs) (colored terms in denominator). (**D**) Measured filter values are passed through their respective LLR functions, and the results are summed and passed through a sigmoidal "f-I curve" to yield boundary probability.

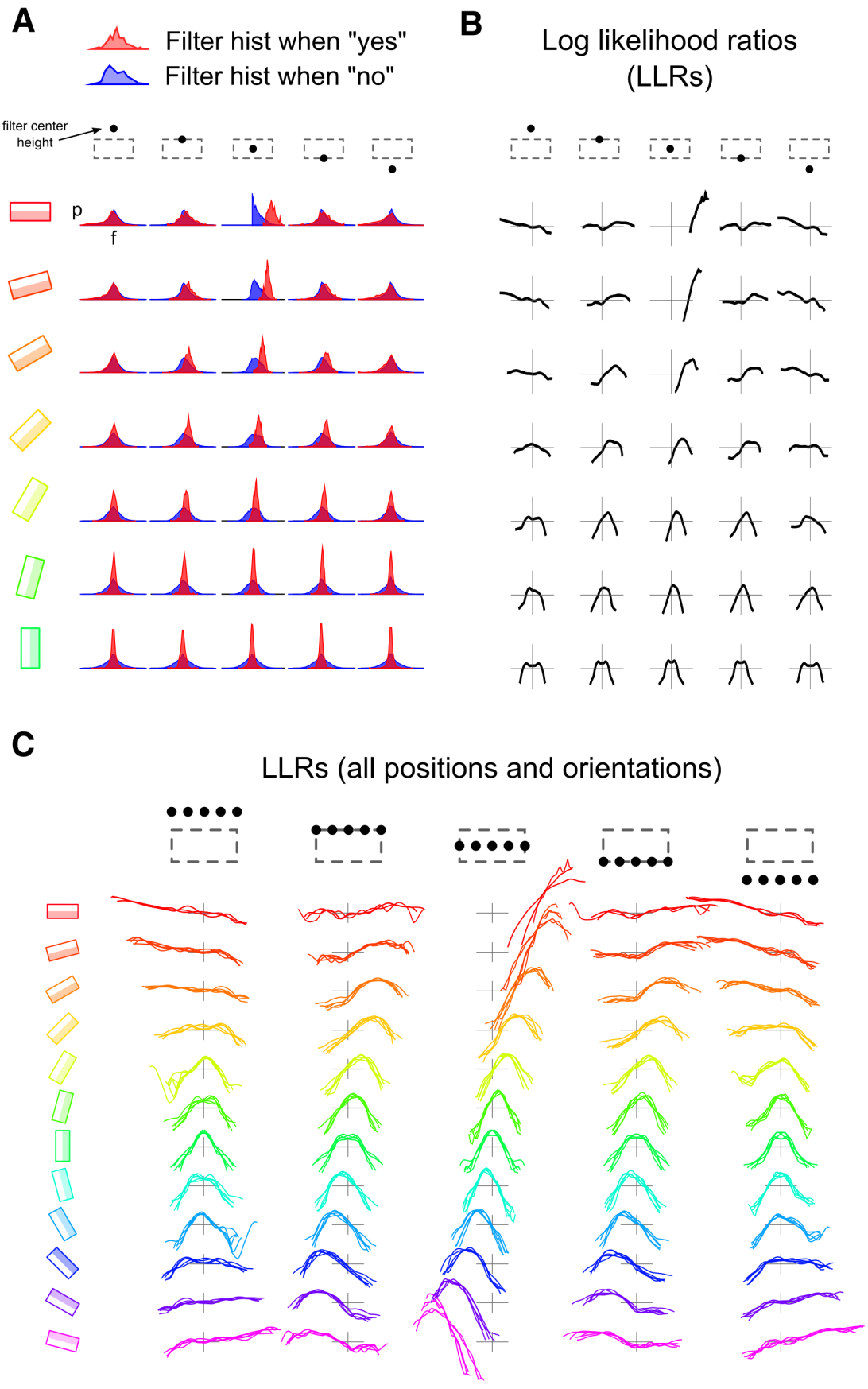


**Figure 2**

**Figure 2. Computing LLRs from natural images.** (**A**) Filter responses from 30,000 labeled image patches potentially containing boundaries at the RL (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 filter histograms is shown for 7 orientations and 5 vertical positions (centered horizontally). (**B**) By dividing the yes and no distributions and taking logs, one obtains the LLRs. (**C**) Full set of 300 LLRs reveals a regular pattern over orientation and location. Cases grouped within each subplot are for 5 horizontal shifts (indicated by black dots at top). Most LLRs are non-monotonic functions of the filter values.

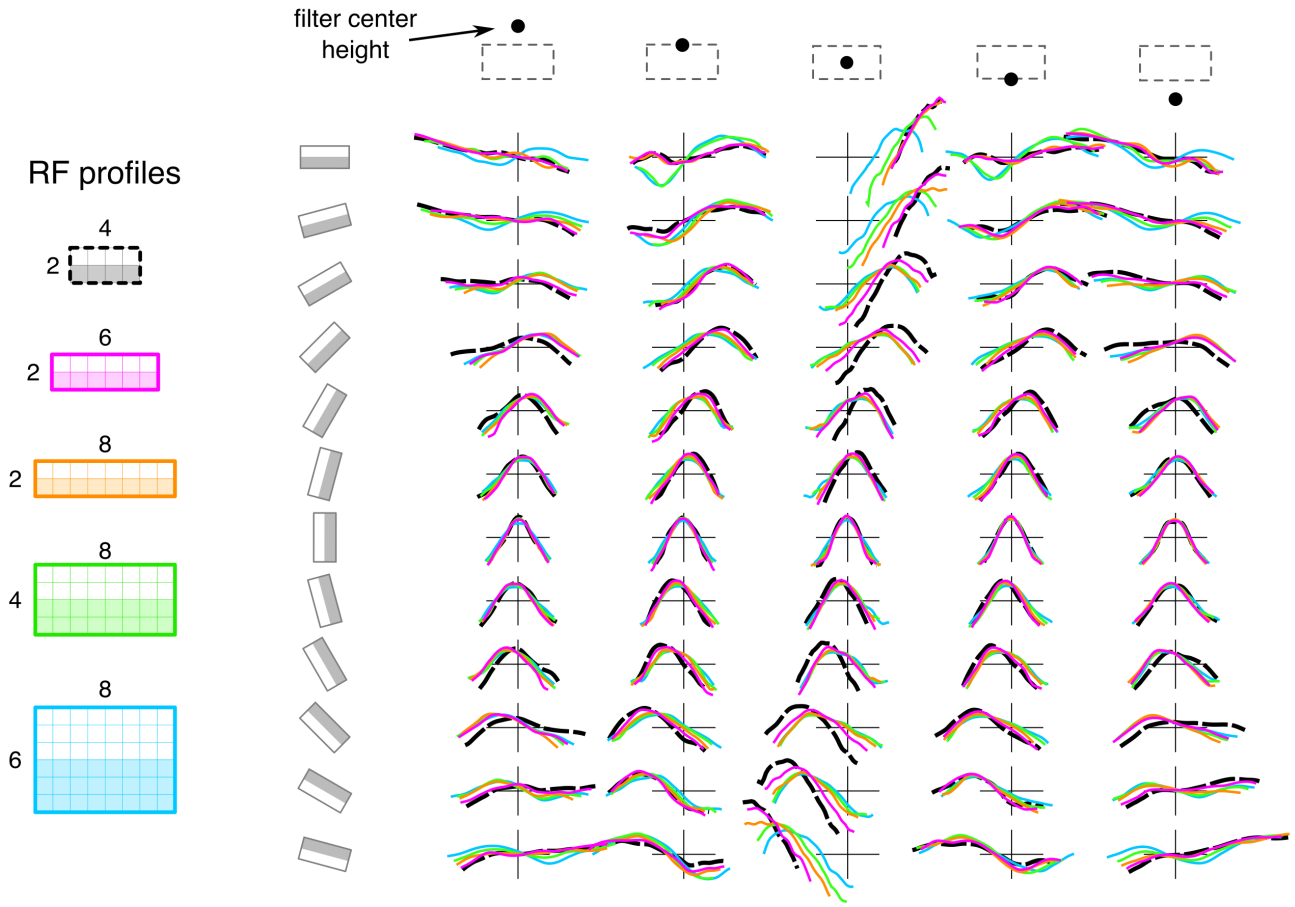


**Figure 3**

**Figure 3. The pattern of LLRs does not depend sensitively on the parameters of the filters.** *Related to Figure 2.* LLRs were generated for each of the filter profiles shown on the left (2x6, 2x8, 4x8, and 6x8 pixels). The LLRs followed the same general pattern for all of the filter profiles and were similarly non-monotonic.

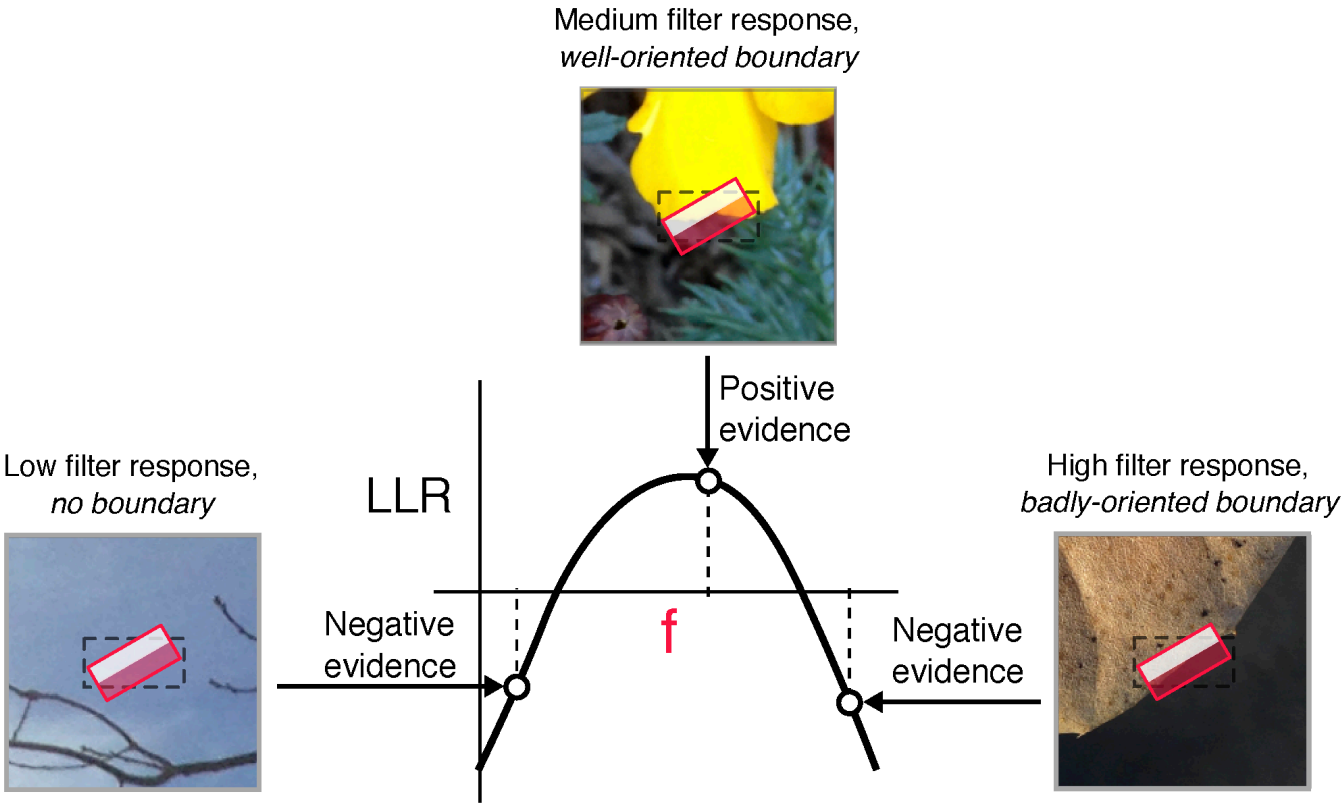


**Figure 4**

**Figure 4. Why a neighbor cell's evidence for a boundary at the RL changes non-monotonically as its response level increases.** *Related to Figure 2.* A low neighbor filter response, as in the patch of sky at left, often signals there is little structure in the image patch, leading to weak or negative evidence for a boundary in the vicinity – including at the RL. A high neighbor filter response, as in the rightmost patch, often signals an object boundary is present *in the neighbor's receptive field*, rather than at the RL, again leading to weak or negative evidence for a boundary at the RL. When the neighbor cell responds strongly enough to indicate the patch contains some structure, but not so strongly as to be competitive with the RL, the neighbor's LLR delivers maximum/positive evidence to the RL, as in the middle patch.

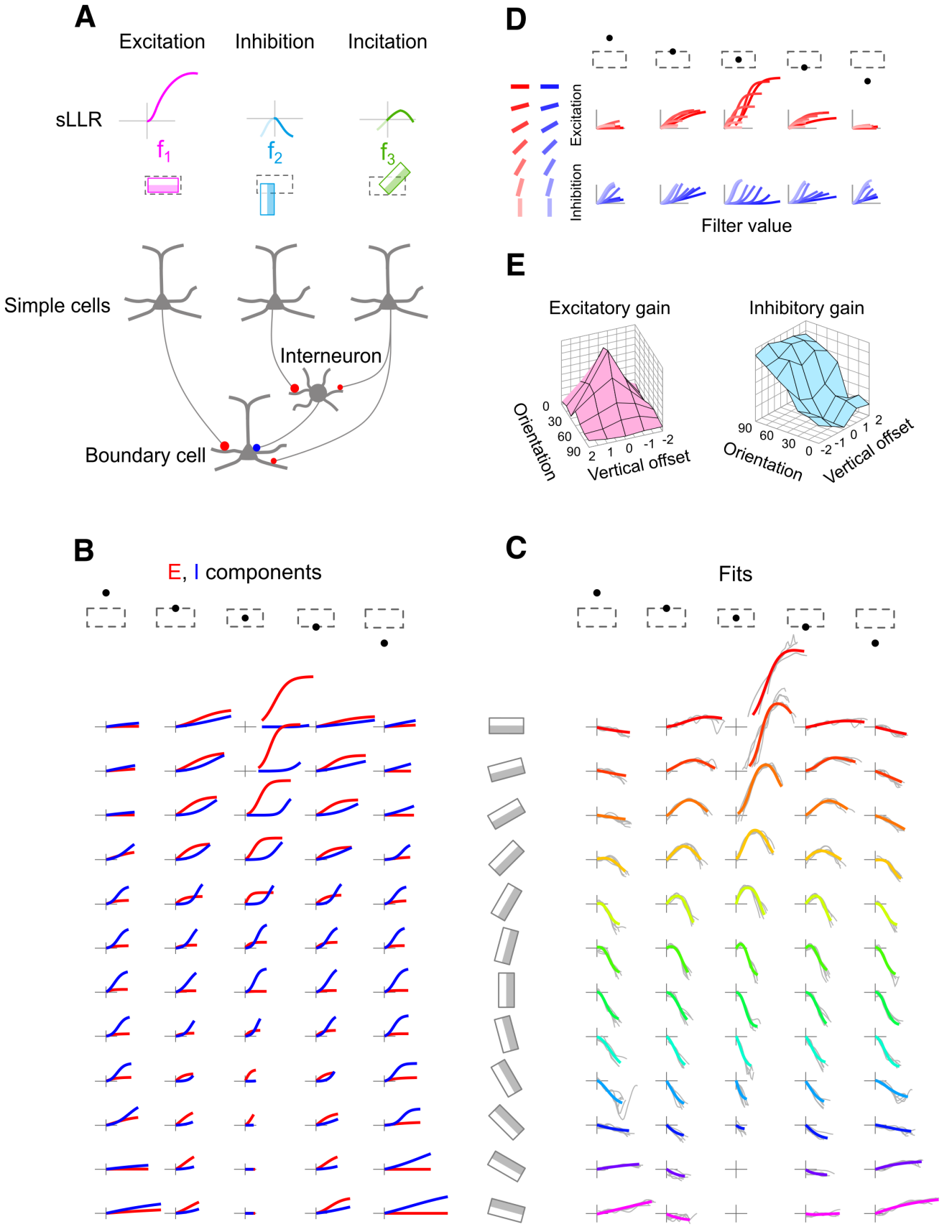


**Figure 5**

**Figure 5. Fitting simple cell–boundary cell interactions (LLRs) with a difference of sigmoids representing separate E and I effects.** (**A**) Circuit motif includes monosynaptic excitation (left stream) and disynaptic inhibition (middle stream) as special cases, and a combined E-I effect capable of generating non-monotonic "incitatory" interactions (right stream). LLRs are shown shifted vertically to emanate from the origin (hence sLLR; combined offsets from all LLRs could be summed to form a single bias term); same in panel (**C**). (**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 (**C**). (**C**) LLR fits are shown in color, on top of the 5-curve groups from Figure 2C 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 inhition grows stronger as the neighbor's orientation deviates from the reference orientation.

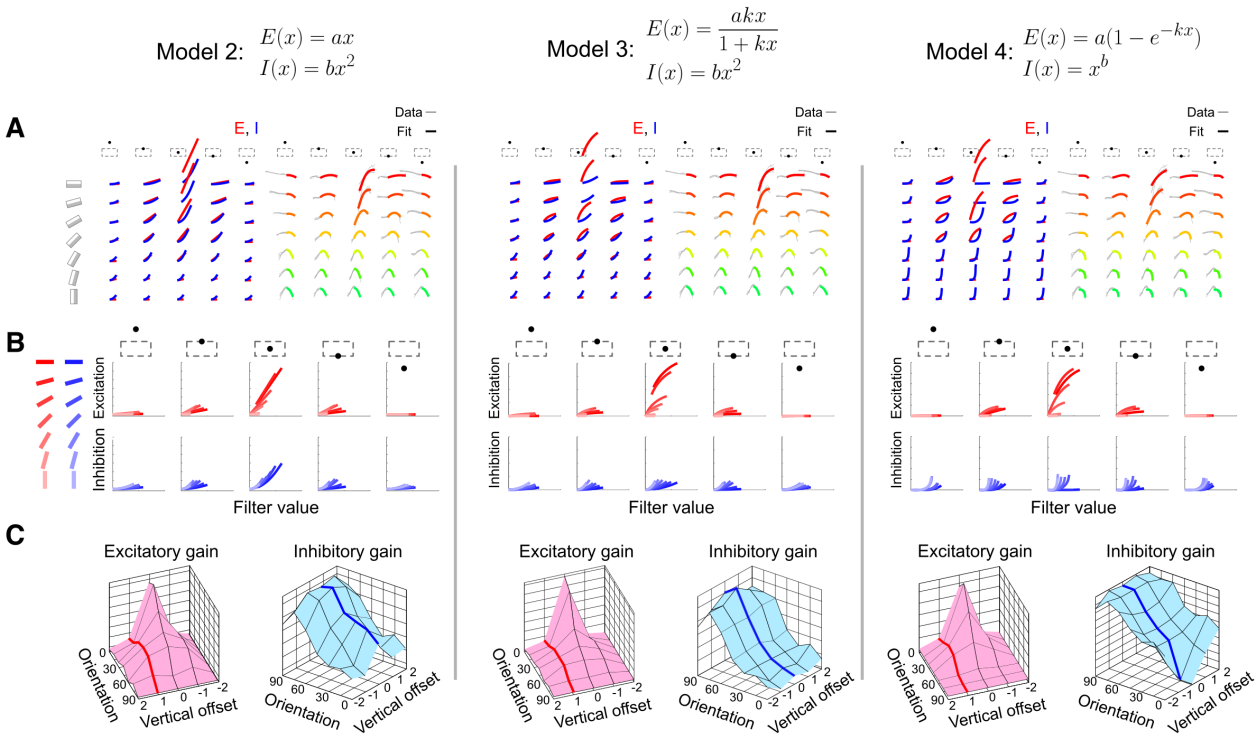


**Figure 6**

**Figure 6. Circuit-level predictions depend only weakly on the choice of parameters representing the excitatory and inhibitory component curves.** *Related to Figure 5.* 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 5c. (**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.

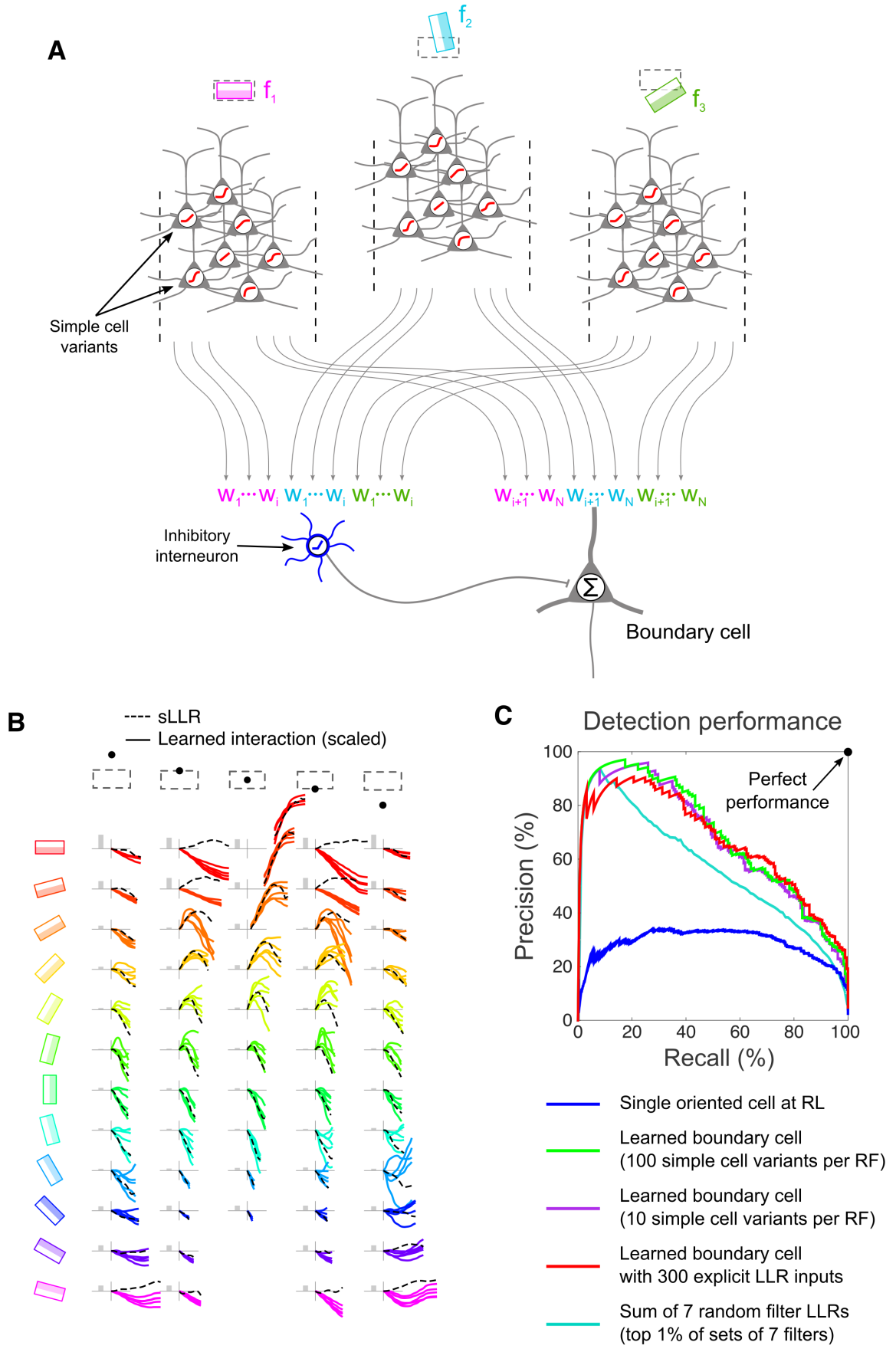


**Figure 7**

**Figure 7. 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 100 simple cells, each with a different fixed i/o nonlinearity. Nonlinearities were sigmoids, $y = \frac{1}{1+e^{-g(x-t)}}$, with gain $g$ between 0.13 and 4.5 and threshold $t$ between -6 and 16. A "delta" rule was used to adjust the weights from each simple cell onto the boundary cell: $\Delta w_i = \eta(t - y)x_i$, 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 $i^{th}$ simple cell, and $\eta$ is the learning rate. In the context of our model, this learning rule is mathematically equivalent (up to a transient initial difference in the learning rate parameter $\eta$) 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 100 associated simple cells was plotted (colored curves). Black dashed curves are averaged LLRs from Figure 2C. The gray bar in each plot represents the weight that the BC puts on that group of 5 colored curves (**C**) Precision-recall curves (on held out data) for the learned boundary cell (green) and weighted sum of LLRs (essentially the explicit Bayesian approach illustrated in Figure 1B and C) (red) are very similar, indicating that the learned neural circuit behaves in accordance with the theoretical prediction. A pure sum of 7 filter LLRs (99$^{th}$ percentile of randomly selected sets of 7 filters) is shown in turqoise. The lower blue PR curve shows that by comparison, a single oriented simple cell at the RL is a poor detector of natural object boundaries.

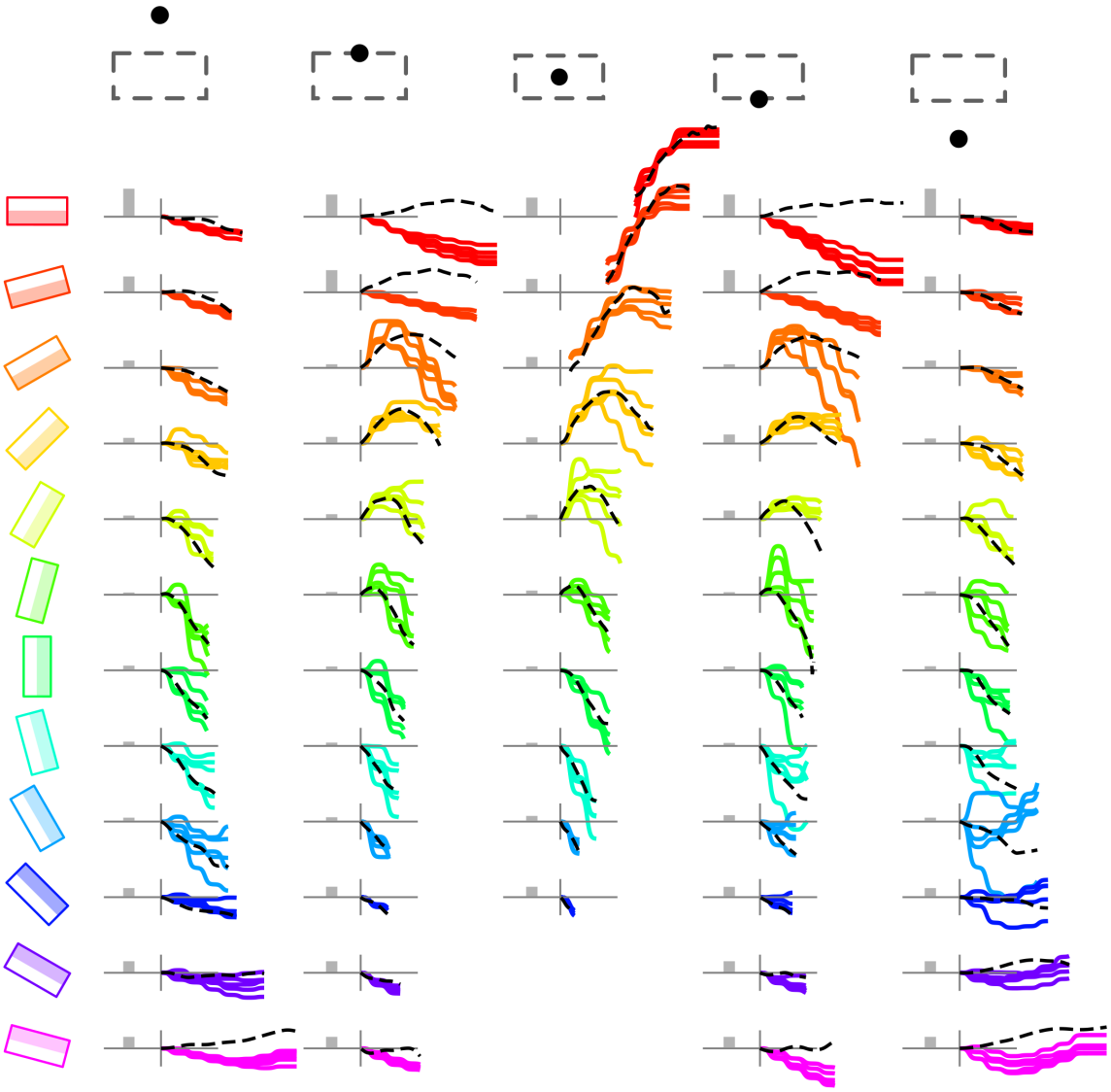

**Figure 8**

**Figure 8. Filter effects on the boundary cell were similar when only 10 nonlinear variants per filter were used.** *Related to Figure 7.* See Figure 7 for explanation.

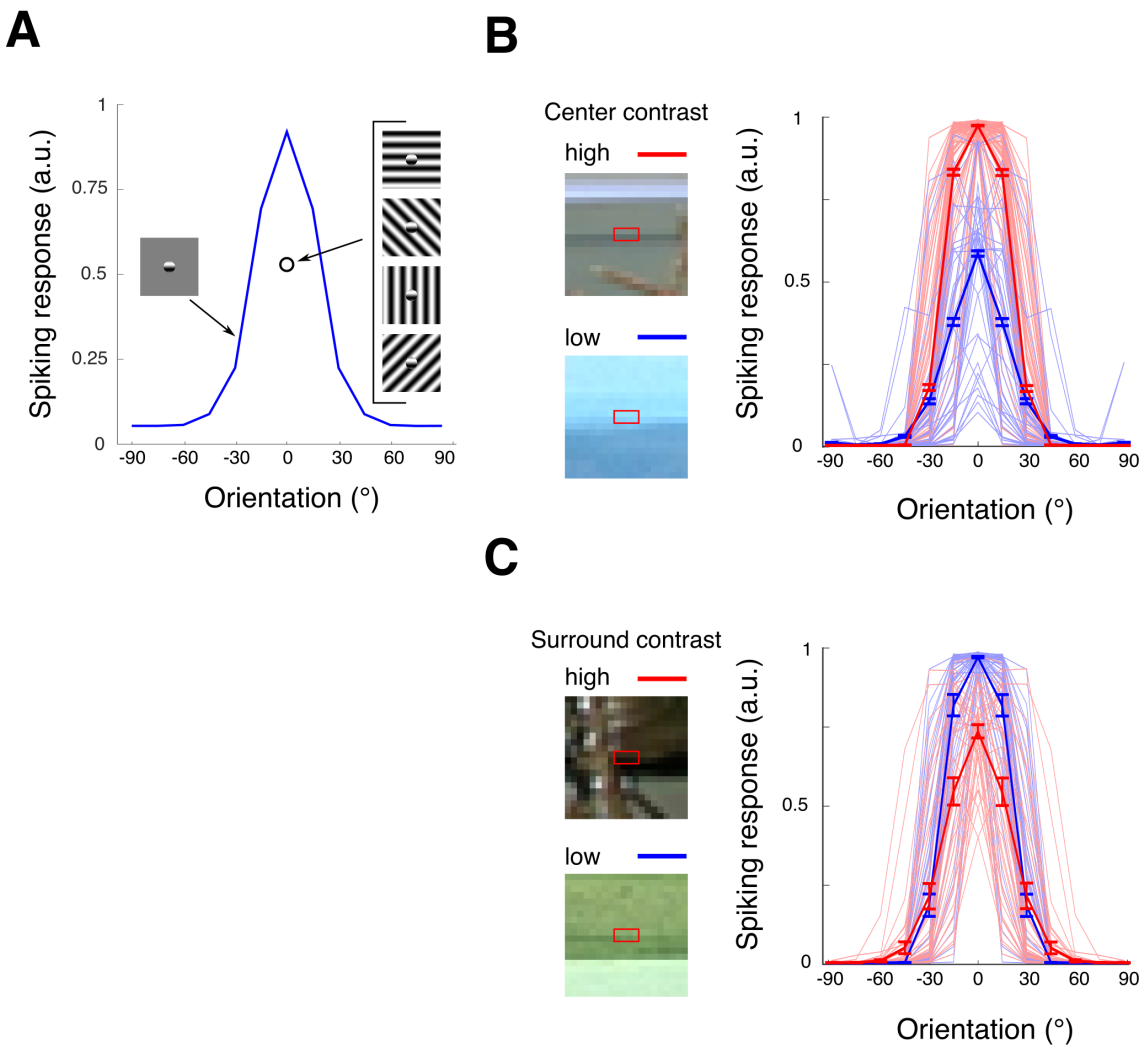


**Figure 9**

**Figure 9. Boundary cell responses to parametric and natural stimuli resemble simple cell responses.** To compute BC responses, the weighted sum of LLRs model (red PR curve in Figure 7C) was used. (**A**) Orientation tuning curve to a sinusoidal grating in a 2 pixel radius window around the receptive field center. At each orientation, responeses were averaged over all phases of the grating. The resulting tuning curve is similar to those obtained for simple cells in V1. Surround suprression was tested by adding a surround grating to a horizontal center grating, and averaging the response over all center and surround phases, and surround orientations. (**B**) Patches with fixed surround contrast and varying center contrast were selected and presented at 15° increments to the boundary cell. For a fixed surround contrat, center contrast increases cell response without increasing tuning width, a hallmark of contrast invariant orientation tuning found in V1 simple cells. (**C**) Holding center contrast fixed while increasing surround contrast has the opposite effect.

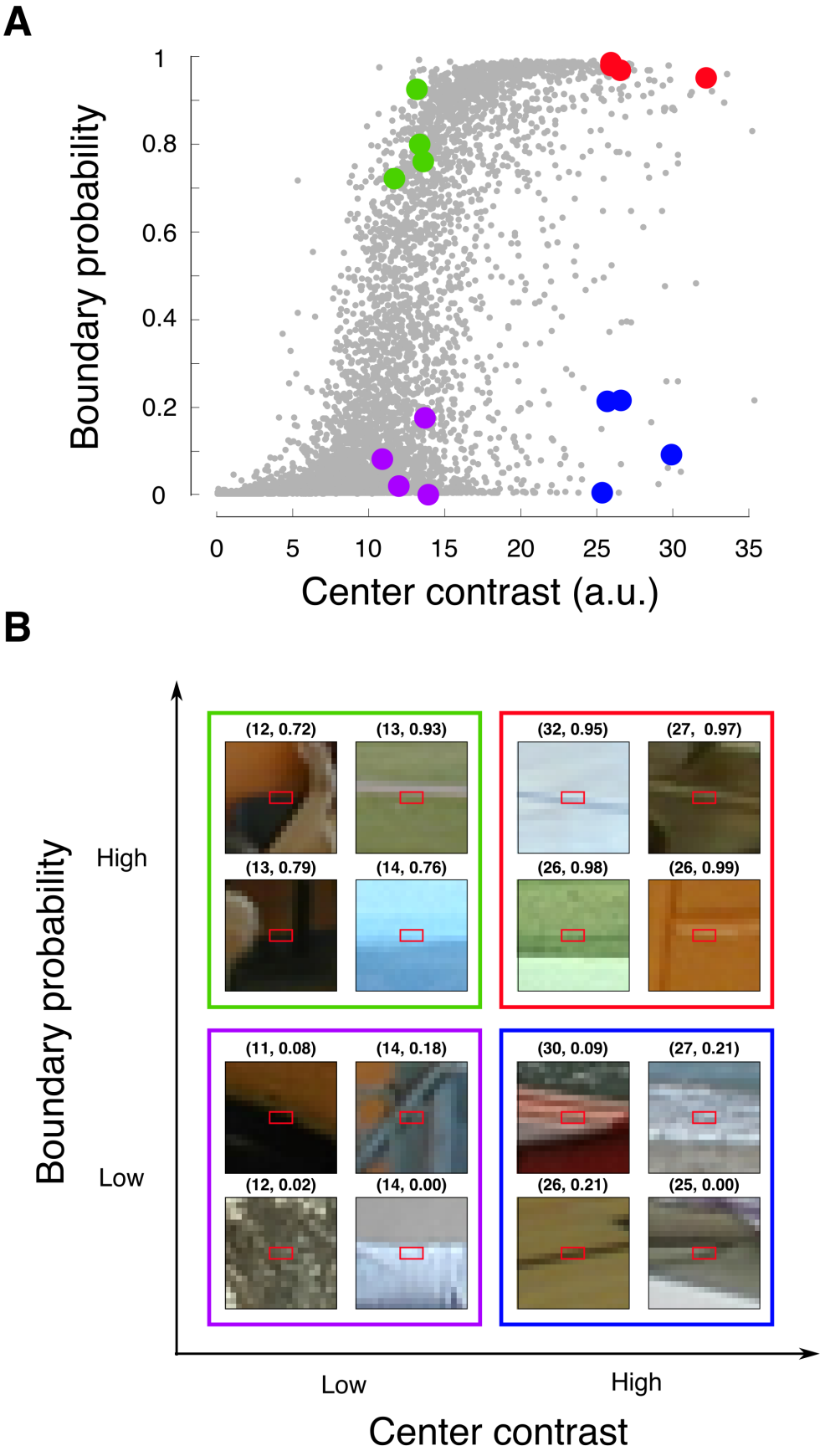


**Figure 10**

**Figure 10. 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 correllated, 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).