Psychological Review

Complementary Surrounds Explain Diverse Contextual Phenomena Across Visual Modalities

David A. Mély, Drew Linsley, and Thomas Serre

Complementary Surrounds Explain Diverse Contextual Phenomena Across Visual Modalities

David A. Mély, Drew Linsley, and Thomas Serre
Brown University

Context is known to affect how a stimulus is perceived. A variety of illusions have been attributed to contextual processing—from orientation tilt effects to chromatic induction phenomena, but their neural underpinnings remain poorly understood. Here, we present a recurrent network model of classical and extraclassical receptive fields that is constrained by the anatomy and physiology of the visual cortex. A key feature of the model is the postulated existence of near- versus far- extraclassical regions with complementary facilitatory and suppressive contributions to the classical receptive field. The model accounts for a variety of contextual illusions, reveals commonalities between seemingly disparate phenomena, and helps organize them into a novel taxonomy. It explains how center-surround interactions may shift from attraction to repulsion in tilt effects, and from contrast to assimilation in induction phenomena. The model further explains enhanced perceptual shifts generated by a class of patterned background stimuli that activate the two opponent extraclassical regions cooperatively. Overall, the ability of the model to account for the variety and complexity of contextual illusions provides computational evidence for a novel canonical circuit that is shared across visual modalities.

Keywords: visual cortex, illusion, induction, assimilation

Supplemental materials: http://dx.doi.org/10.1037/rev0000109.supp

Spatial context has been known to affect perception since at least Aristotle (Eagleman, 2001). The past several decades of work in visual psychophysics have revealed a plethora of seemingly disparate contextual phenomena (Seriès, Lorenceau, & Frégnac, 2003) whereby subtle differences in experimental conditions yield a wide variety of effects (see Figure 1). In the classical tilt illusion (Goddard, Clifford, & Solomon, 2008; O'Toole and Wenderoth, 1977), the perceived orientation of a center stimulus tilts either toward or away from that of a surround stimulus, depending on their relative orientations. Many variants have been tested with a variety of stimulus parameters including spatial frequency, color, luminance, contrast differences between center and surround stimuli as well as their spatial and temporal separation (see Clifford, 2014, for a review).

Similar effects have been reported in the motion domain—for both direction and speed (Kim & Wilson, 1997; Marshak & Sekuler, 1979; Murakami & Shimojo, 1993, 1996). In color in-duction, both the spatial frequency and phase of the surround controls the direction of the perceived shift in hue of a center stimulus relative to that of the surround (Monnier & Shevell, 2003; Shevell & Monnier, 2005; Smith, Jin, & Pokorny, 2001). In the disparity domain, a center stimulus appears closer or further away from an observer, depending on the relative depth and spacing between center and surround stimuli (Westheimer, 1986; Westheimer & Levi, 1987). While much is known about the psychological basis of these phenomena, our understanding of the underlying neural mechanisms remains, at best, fragmentary.

A widely held assumption is that such contextual phenomena are mediated in the cortex by extraclassical receptive field (eCRF) mechanisms (reviewed in Angelucci & Shushruth, 2013; Seriès et al., 2003): The presentation of a stimulus in the eCRF alone does not typically elicit any response from the neuron but modulates its response to a stimulus presented in the classical receptive field (CRF). Such center-surround interactions have been reported across visual modalities including orientation and spatial frequency (DeAngelis, Freeman, & Ohzawa, 1994), motion (Jones, Grieve, Wang, & Sillito, 2001; Li, Lei, & Yao, 1999), color (Schein & Desimone, 1990; Wachtler, Sejnowski, & Albright, 2003), and disparity (Bradley & Andersen, 1998).
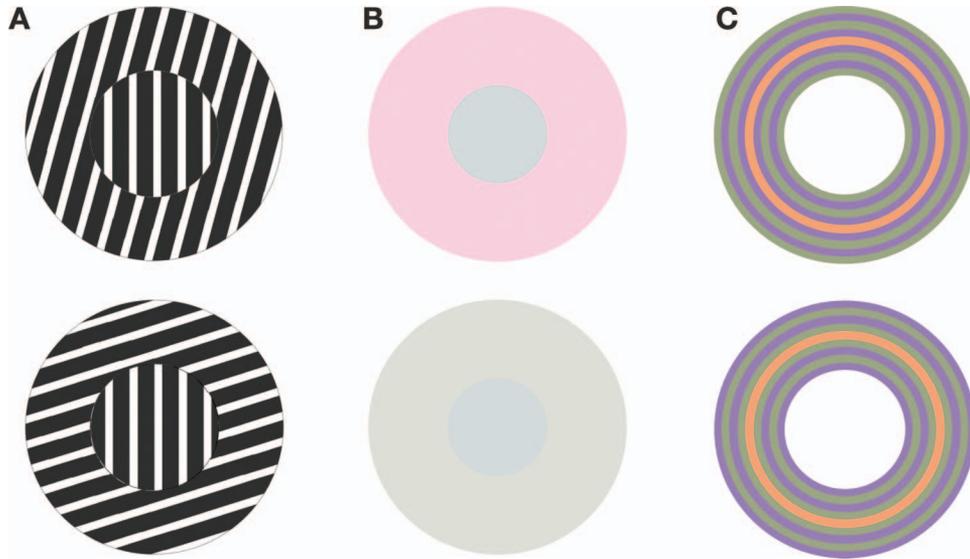
Although several eCRF models have been developed to describe specific phenomena (reviewed in Angelucci & Shushruth, 2013; Schwartz, Hsu, & Dayan, 2007; Seriès et al., 2003; see also Discussion section), a unifying theory, which would integrate disparate aspects of contextual integration and, ultimately, link primate neurophysiology to human behavior, is still lacking. We have thus developed a large-scale recurrent network model of classical and extraclassical receptive fields that distinguishes itself

Figure 1. Representative contextual phenomena explained by the model. (A) Orientation tilt: The perceived center (or test) orientation appears tilted from its true physical orientation, away from the surround (or contextual) orientation when center and surround stimuli are similar (top) and toward the surround orientation when they are dissimilar (bottom). (B) Color induction: A central gray stimulus appears greener when embedded in a pink surround (top) compared to a neutral gray surround (bottom). (C) Enhanced color shifts: The test stimulus is a central, orange ring, embedded in a surround stimulus composed of alternating purple and lime rings. The test ring looks vividly more pink when the adjacent color is purple, followed by lime (top), and looks more yellow when lime is the adjacent color, followed by purple (bottom).

from previous work—allowing us to simulate realistic cortical responses to a variety of full-field, real-world, contextual stimuli defined across visual modalities (we model orientation, color, motion, and binocular disparity). The model is constrained by anatomical data and shown in our experiments to be consistent with primary visual cortex (V1) neurophysiology. A key feature of the model is the postulated existence of near versus far extraclassical eCRF regions with complementary contributions (facilitatory vs. suppressive) to the CRF response. Using an ideal neural observer, we show that the model is consistent with human behavioral responses for a variety of contextual phenomena—revealing commonalities between seemingly disparate phenomena and helping to establish a novel taxonomy of contextual illusions.

Results

The visual cortex is modeled as a dense, regular topographic grid of cortical (hyper)columns which tile the visual field (Figure 2A). Each hypercolumn contains a complete set of units with coinciding CRFs. Their tuning curves are idealized (see Materials and Method section), and centered at regular intervals (e.g., between 0° and 180° for orientation-tuned units). For simplicity, we do not take into account cortical magnification and assume a fixed sampling of the visual field at all eccentricities. The model takes into account connections both within and across hypercolumns in order to explain several CRF and eCRF properties. The resulting circuit motif is replicated for every hypercolumn.

Intracolumnar Recurrent Circuits

Recurrent connections (Figure 2A, red connections) within a column (i.e., originating from within the CRF) include both local

excitatory and inhibitory connections. Inhibitory CRF contributions constitute one of the key mechanisms in an influential model of gain control (divisive normalization, reviewed in Carandini & Heeger, 2012). This model accounts for cross-orientation normalization phenomena (when a grating stimulus is masked by another one at any orientation, see Carandini & Heeger, 1994; Heeger, 1993) and was later extended to capture neural population responses, in order to account for the competitive interactions within a single hypercolumn (Busse, Wade, & Carandini, 2009; Sit, Chen, Geisler, Miikkulainen, & Seidemann, 2009). A recent optogenetic study demonstrated that the underlying circuits are recurrent rather than feedforward (Nassi, Avery, Cetin, Roe, & Reynolds, 2015). We have also confirmed that this form of recurrent intracolumnar inhibition is critical for the model to reproduce these types of competitive interactions within the CRF and for model units to exhibit a realistic contrast response (see supplementary experiments; Figures S1–S5).

Because this form of suppression does not seem to depend on the orientation of the afferent and target cells, it is often called "untuned" inhibition (Rust, Mante, Simoncelli, & Movshon, 2006). In our model, we speculate that such untuned inhibitory recurrent connections within hypercolumns exist for all other visual domains (including color, motion as well as binocular disparity, see Discussion section).

In addition to short-range inhibitory connections within hypercolumns, the model also incorporates short-range excitatory connections. In the cortex, such excitatory connections may drive neurons up to 10 times more strongly than their feed-forward inputs (Douglas, Koch, Mahowald, Martin, & Suarez, 1995; Stepanyants et al., 2008). As suggested by Shushruth et al. (2012), we
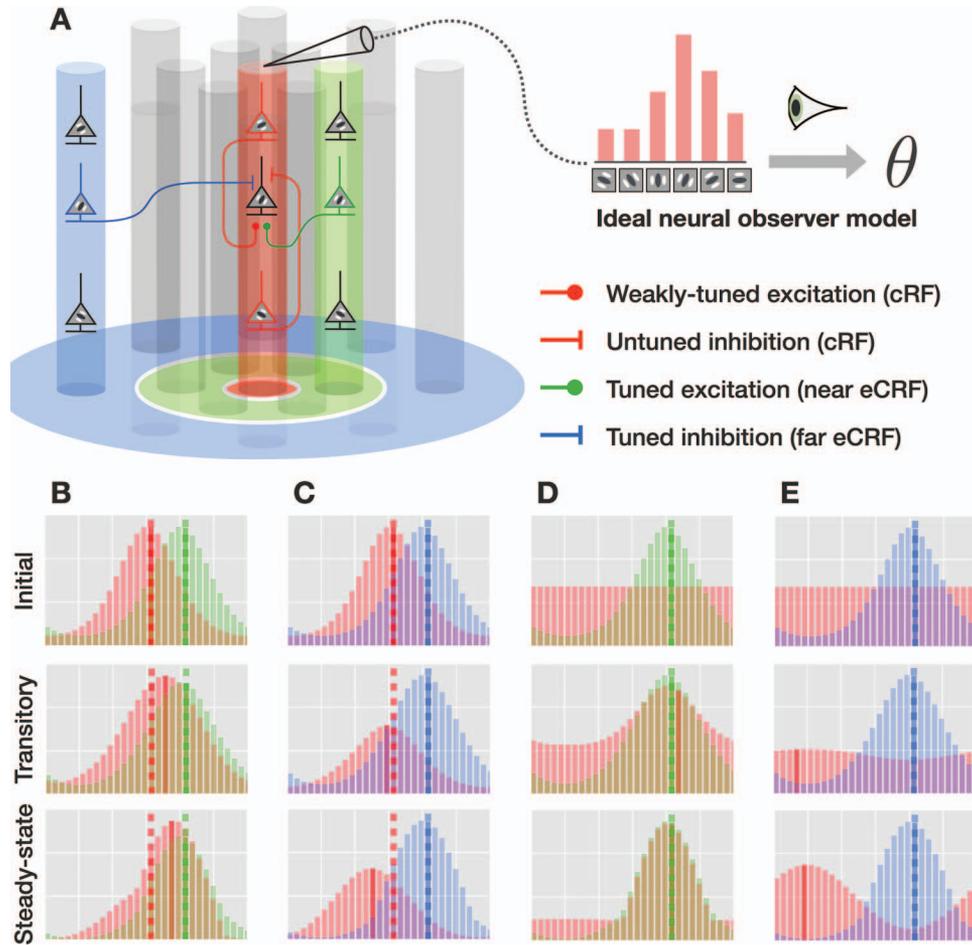
Figure 2. Recurrent network model of center-surround interactions. (A) Connectivity: The model implements excitatory and inhibitory connections both "within and across hypercolumns." (current wording may confuse the reader as green connections are both across hypercolumns and comparatively short-range with respect to the blue connections). The regions shown in red, green, and blue correspond to the CRF, near eCRF, and far eCRF, respectively (defined for the reference column in red). Model inhibitory connections are such that the net inhibition onto a target unit is a function of not just the presynaptic activity, but also the postsynaptic activity (see text for details). Color conventions for the CRF and the near and far eCRFs are used consistently throughout the paper. (B–E) Representative model dynamics: Example population responses (32 direction-tuned model units) following the presentation of a contextual stimulus corresponding to the initial, transitory and steady state (rows). Population responses correspond to locations in the CRF, near eCRF, and far eCRF. Highlighted bars represent directions decoded from the corresponding populations (undefined for flat responses); dashed lines represent values initially decoded at stimulus onset. Each column corresponds to a representative transformation undergone by the center population under the proposed taxonomy of contextual phenomena derived from the model: (B) attractive shift, (C) repulsive shift, (D) bump, (E) notch. (see also Discussion section and Figure 8). Abscissae span the range (−180°, 180°) and ordinates are normalized independently for readability.

have found that recurrent excitation within the CRF is essential to account for some of the more complex aspects of surround suppression (see supplementary experiments, Figures S4 and S5), by placing the column in a regime dominated by recurrent as opposed to feed-forward inputs. However, experimental data on the selectivity of these recurrent excitatory connections are scarce. Here, we assume that the corresponding local excitatory connections within a hypercolumn are only weakly tuned, as perfectly untuned excitation would effectively "flatten out" population response curves.

Intercolumnar Recurrent Circuits

Expanding the optimal stimulus of a cortical neuron immediately beyond its CRF (also commonly referred to as the "minimum response field" or mRF) may facilitate its response (Angelucci, Levitt, & Lund, 2002; Angelucci, Levitt et al., 2002; Briggs & Usrey, 2011; Bringuier, Chavane, Glaeser, & Frégnac, 1999; Sceniak, Ringach, Hawken, & Shapley, 1999; Sengpiel, Sen, & Blakemore, 1997). The area which covers the CRF and its immediate eCRF is sometimes referred to as "peak spatial summation

area." It is considered distinct from the CRF because a direct stimulation of this eCRF region in isolation does not elicit any action potential. Since this region is located immediately beyond the CRF, we deem it the *near eCRF* (or near surround; green annulus and connections in Figure 2A).

A potential neural substrate for the near eCRF includes the short-range, tuned excitatory networks (Lee et al., 2016) which span a spatial extent consistent with that of the eCRF facilitation and amplify co-occurring local inputs at similar orientations (Angelucci et al., 2002; Angelucci, Levitt et al., 2002; Briggs & Usrey, 2011; Sceniak et al., 1999; Sengpiel, 1997). In the model, we assume that all excitatory connections from other hypercolumns centered in the near surround are tuned, irrespective of the visual modality (i.e., the stimulus with the preferred orientation, or direction of motion, etc. in the CRF is also most effective in the near eCRF). Also note that our definition of the near eCRF is purely anatomical and might thus differ from that of others (e.g., Angelucci & Shushruth, 2013), whose definition is functional in nature.

Expanding the optimal stimulus beyond the near eCRF results in neural suppression (first reported by Hubel & Wiesel, 1968, as hypercomplex tuning). Critically, the presentation of the suppressing stimulus in the eCRF alone does not elicit any activity from the recorded cell (see Angelucci & Shushruth, 2013, for review). The tuned nature of these suppressive mechanisms is well documented across visual modalities: from orientation (DeAngelis et al., 1994; Hubel & Wiesel, 1968; Ozeki, Finn, Schaffer, Miller, & Ferster, 2009; Petrov, Carandini, & McKee, 2005; Weliky, Kandler, Fitzpatrick, & Katz, 1995) to color (Schein & Desimone, 1990; Wachtler et al., 2003), spatial frequency (DeAngelis et al., 1994), temporal frequency (Jones et al., 2001; Li et al., 1999), motion direction and speed (Allman, Miezin, & McGuinness, 1985) as well as binocular disparity (Bradley & Andersen, 1998).

Thus, we also define an *inhibitory far eCRF* (or far surround; blue annulus and connections in Figure 2A), located immediately beyond the excitatory near eCRF. In our model, a hypercolumn receives tuned inhibition from hypercolumns centered in its far surround.

To summarize, contributions from the eCRF as a whole arise from surround regions with opposite polarities. We do not assume any gap between the CRF and the near eCRF, nor between the near eCRF and the far eCRF. The model assumes that the near excitatory and far inhibitory eCRFs do not overlap spatially based on partial anatomical evidence (Angelucci, Levitt et al., 2002). In practice, we found that allowing these two eCRF regions to overlap did not affect the model's ability to fit experimental data (see supplementary experiments).

The first key assumption of the model is that, unlike local recurrent interactions within a hypercolumn, interactions across hypercolumns are "tuned" as only units that share the same preferred stimulus are directly connected. The second key assumption of the model is an asymmetry between excitation and inhibition: In the model, excitation only depends on presynaptic activity and is purely additive. Inhibition, on the other hand, from either the CRF or eCRF, depends on both pre- and postsynaptic activity, and ultimately results in a combination of subtractive and divisive effects (Carandini & Heeger, 2012). Similar forms of inhibition have been used in previous recurrent network models to achieve divisive normalization (Grossberg & Todorović, 1988). In prac-

tice, this means that, given a fixed amount of presynaptic inhibition, weakly active units receive less effective inhibition than more active ones. In contrast, any given amount of presynaptic excitation results in the same amount of effective postsynaptic excitation.

Neural Field Model

Upon the presentation of a stimulus, recurrent interactions between units yield complex model dynamics. In particular, population responses at any location are modulated first by their immediate (near and far) eCRFs, then by responses across the visual field as transient activity propagates through the network, until all unit responses settle into a steady-state. Such short- and long-range interactions are modeled using coupled differential equations and the steady-state solution of the resulting neural field model is computed using numerical integration methods (see Materials and Method sections).

Next, we describe experiments conducted *in silico* to compare model responses with published psychophysics data. Psychophysics studies typically record perceptual judgments related to a center stimulus under varying surround conditions. To approximate these judgments, we use an ideal neural observer which maps center population responses to a sensory value. Note that in most cases, several columns may be located within the center stimulus; while any of these columns would be suitable for readout by the ideal observer, we selected the center-most column for simplicity (unless specified otherwise). Surround modulation thus gets translated into measurable perceptual changes in the center that can then be compared with human behavioral data.

We have organized these experiments into three broad categories, each reflecting a key computational mechanism and highlighting commonalities across visual modalities. As we will show, these experiments allow a clear picture to emerge: The diversity of observed contextual phenomena may result from a balance between two opposing "forces" that arise from complementary excitatory-inhibitory eCRF mechanisms. Figure 2 shows examples of CRF and eCRF population responses recorded from the model together with representative transformations they undergo as a result of these two forces (see Discussion section for more details).

All model parameters (Table S1) governing the dynamics and relative strengths of the interactions between the CRF and the eCRF subregions were initially adjusted for the model to reproduce a host of V1 neurophysiology data (see supplementary experiments; Figures S1–S5) including a comparison with data from Busse et al. (2009) and Trott and Born (2015). They were held fixed for all subsequent comparisons with psychophysics data. After scaling the stimulus, the only model parameter that was optimized for individual modalities was the tuning bandwidth of individual model units.

Competitive Activation of the Near Versus far Surrounds

Our comparison between experimental and model data starts with a set of three experiments that span the orientation, motion, and color domains. All experiments involve simple center-surround stimuli, in which the surround stimulus is expected to jointly activate both the excitatory and inhibitory components of

the eCRFs. Thus, these experiments should reveal a fundamental aspect of the model: the outcome of a competition between the near facilitatory and the far suppressive eCRFs when they are simultaneously activated by a surround stimulus.

The orientation tilt occurs when the perceived orientation of a center stimulus is biased either toward (Figure 3B) or away (Figure 3A) from the orientation of a surround stimulus, also called the inducing stimulus or inducer (Goddard et al., 2008; O'Toole and
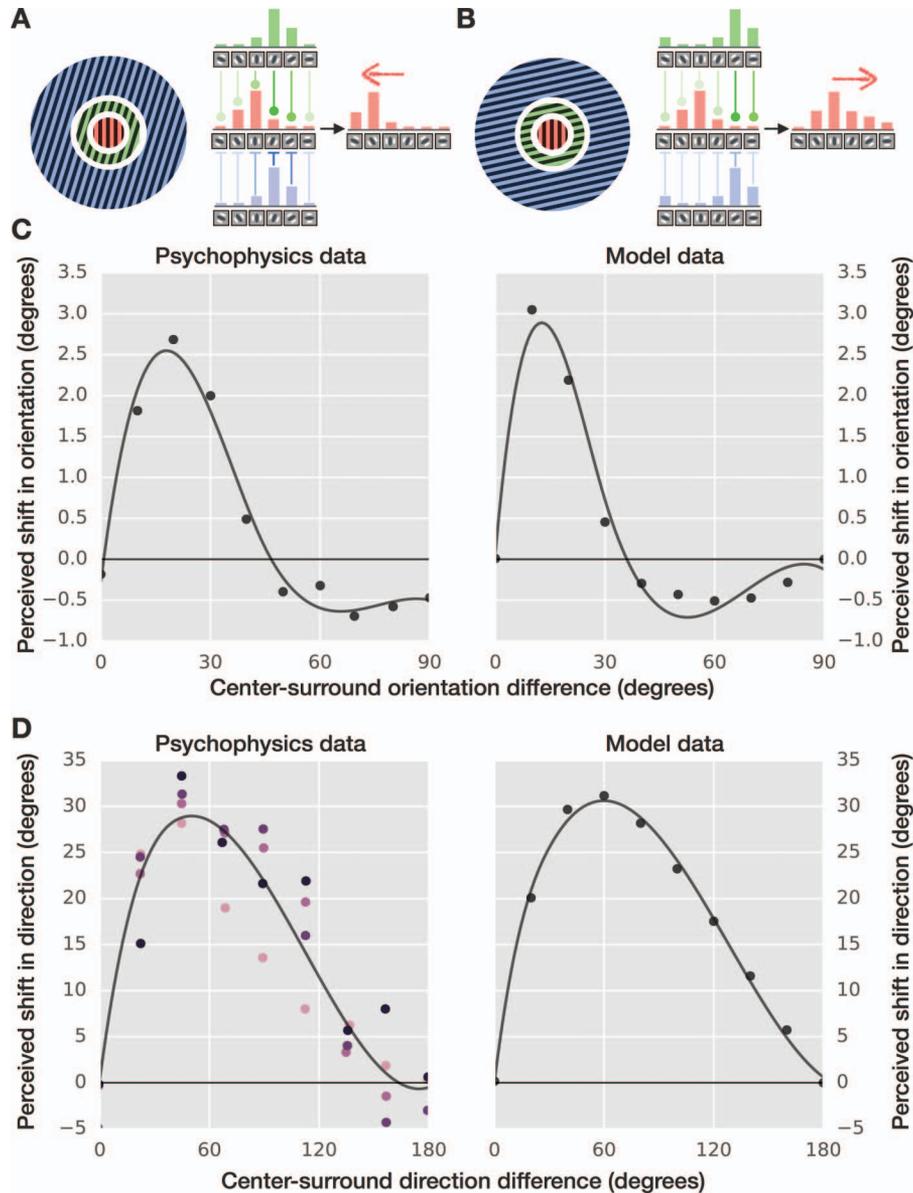
Figure 3. Tilt effects. Competitive activation of the near versus far eCRFs explains the shift in tilt from one direction to the other. (A) Repulsion: For similar center-surround orientations, tuned inhibition from the far eCRF outweighs excitation from the near eCRF, which yields a net repulsive force on the center population responses (away from that the surround orientation). (B) Attraction: For dissimilar center-surround orientations, tuned excitation from the near eCRF prevails, which yields a net attractive force on the center population responses (toward the surround orientation). Note that gaps between the CRF and the near and far eCRFs were added for improved readability only and are not present in the actual model. (C) Orientation tilt: Psychophysics versus model data. Psychophysics data were digitally extracted from Figure 4 in (O'Toole and Wenderoth, 1977) and fitted with splines. The model explains the characteristic shift from perceptual repulsion (positive ordinates) to attraction (negative ordinates). (D) Motion tilt: Psychophysics versus model data. Psychophysics data were digitally extracted from Figure 3 ("periphery" condition) in (Kim & Wilson, 1997) and fitted with splines. Different colors correspond to different subjects. Both psychophysics and model data exhibit a similar dependency on the direction difference between center and surround, as well as a lack of an attractive regime.

Wenderoth, 1977). Figure 3C shows representative psychophysics data (digitally extracted from Figure 4 in O'Toole and Wenderoth, 1977, only data averaged across subjects are available from that study). These data are characterized by two regimes: a *repulsive* regime (i.e., the perceived center orientation shifts away from the surround orientation, corresponding to positive ordinates) when the surround orientation is similar to that of the center and an *attractive* regime (i.e., the perceived center orientation shifts toward the surround orientation, corresponding to negative ordinates) when the surround orientation is different enough from that of the center.

The model successfully reproduces this balance between attraction and repulsion (Figure 3C; a similarly good fit was also obtained using broadband oriented textures as done in Goddard et al., 2008, not shown). The key mechanism which enables the emergence of these two regimes is the postulated asymmetry between facilitatory and suppressive interactions originating from the near and far eCRFs, respectively. The net inhibition in the model, unlike excitation which is only dependent on presynaptic activity, increases monotonically with the level of postsynaptic activity of a target unit. We have confirmed this hypothesis via selective lesioning of the model key components (see Figure S14).

As a result, when neural population responses in the CRF and eCRF overlap significantly (as when center and surround orientations are similar), inhibition predominates and center population

responses get comparatively more suppressed at orientations close to that of the surround. The center of mass of center population response curves shifts away from the surround orientation, biasing the neural decoding accordingly (Figure 3A). The surround thus acts as a repellent in this regime. In contrast, when neural population responses in the CRF and eCRF are far more offset (as when the surround orientation is near orthogonal to that of the center), excitation from the near eCRF predominates, and increases the activity of center units selective for the surround orientation. This results in a force that pushes the center population response toward the surround orientation. This, in turn, biases the decoding of the center orientation in the direction of the surround orientation (Figure 3B). The surround thus acts as an attractor in this regime.

Beyond the orientation domain, tilt effects have also been reported for the perception of motion direction. Figure 3D shows representative psychophysics data (digitally extracted from Figure 3 "periphery" condition in Kim & Wilson, 1997). Unlike in the orientation domain, however, perceptual shifts are always repulsive (the perceived motion direction of the center grating tilts away from that of the surround grating; both gratings have the same contrast and speed). This phenomenon can also be induced using coherently moving random dots (Marshak & Sekuler, 1979); the effect seems to peak for similar center-surround differences in motion direction (between 40° and 60°) for either kind of stimuli.
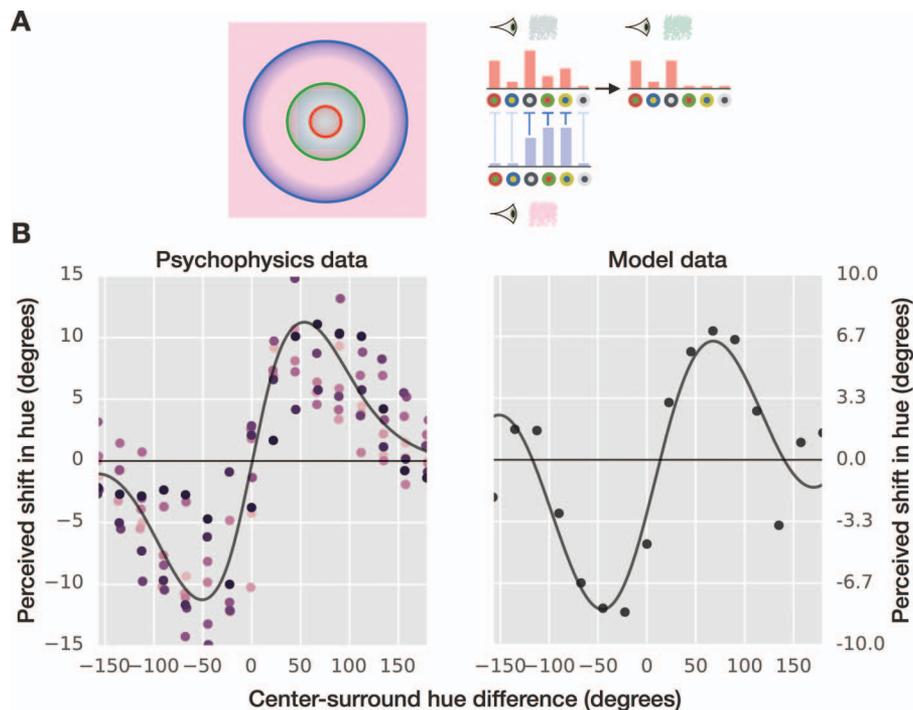
Figure 4. Color induction (or hue tilt effect). This experiment generalizes the tilt effect to opponent population codes. (A) Repulsion: As with the classical tilt effect, the key model mechanism behind perceptual repulsion is the tuned inhibition from the far eCRF. In this example, the pink surround suppresses "red" center neurons, therefore reducing the "redness" of the gray center patch yielding a shift in the perceived center hue toward green. The same explanation also applies to chromatic center stimuli. Colored patches shown next to the eyes correspond to the color decoded under the ideal observer. (B) Psychophysics versus model data: Psychophysics data were digitally extracted from Figure 2 in (Klauke & Wachtler, 2015) and fitted with splines (averaged across eight surround hues). Both model and behavioral data exhibit a characteristic two-lobed shape peaking around ±50°. Please note the difference in ordinate scale between the psychophysics and model data.

We found both a qualitatively and quantitatively good fit between the model and psychophysics data as shown in Figure 3D. In the model, the disappearance of the attractive regime is accounted for by relatively broader model tuning curves (compared to orientation; see supplementary materials and Methods). Interestingly, this seems consistent with neurophysiology data from the primary visual cortex (Albright, Desimone, & Gross, 1984; Ringach, Shapley, & Hawken, 2002), which suggest that tuning for motion direction tends to be broader than for orientation.

In our next experiment, we show that the model is also able to account for tilt effects in the hue domain, more widely known as color induction. The model reproduces the known shifts in human judgment obtained when a center hue is surrounded by an isoluminant background of a different hue (data digitally extracted from Figure 2 in Klauke & Wachtler, 2015, averaged across multiple combinations of center-surround hues sampled uniformly and independently as done in the original experiment).

As with motion induction, only a repulsive tilt effect is observed with hue. The model's ability to account for these data is evident from Figure 4B, which confirms the hypothesis by Klauke and Wachtler (2015) that color induction is in fact just another tilt effect (i.e., a "hue tilt effect"). Furthermore, the same mechanisms that are responsible for the tilt effect in the orientation and motion domains, namely the balance between facilitatory and suppressive forces originating from the eCRF, are also at play in color induction (Figure 4A). However, opponent color coding yields populations from the center and the surround with high overlap, which explains the absence of an attractive regime for this phenomenon.

Exclusive Activation of the Near Versus Far Surrounds

Previous experiments involved stimuli that reflected the outcome of a competition between the near and far eCRF, which were activated jointly. Here, instead, we consider experiments that are based on surround stimuli that activate the near or the far eCRFs separately.

In classical depth induction experiments (Westheimer, 1986; Westheimer & Levi, 1987), human observers are presented binocularly with a center test stimulus (e.g., a thin bar) flanked by two surround stimuli (e.g., parallel thin bars or small squares). The disparities of the flanker stimuli are adjusted so that they appear in the same depth plane, either slightly in front of or behind the center stimulus. The planar separation between the center and flanker stimuli (i.e., their distance in the fronto-parallel plane) is varied systematically. Examples where the flankers appear behind the center stimulus for a shorter and a larger separation are shown in Figure 5A–B.

Results from the original study (data digitally extracted from Figure 1, upper panels in Westheimer & Levi, 1987) are shown in Figure 5C. When the flankers are close enough to the center stimulus, they seem to attract it in depth (corresponding to a negative shift in perceived disparity for very small flankers/center separations). That is, the center stimulus appears closer to (further away from) the observer when the flankers are in front of (behind) the center stimulus. Instead, when the flankers are moved far enough laterally, they start to repel the center stimulus (corresponding to a positive shift in perceived disparity for larger flankers/center separations).

The observed shifts in depth found in the model (Figure 5C) match qualitatively the human psychophysics data: Flanker stimuli located close enough to the test stimulus activate solely the near eCRF, resulting in a purely facilitatory net eCRF influence. As with the aforementioned tilt effects, net facilitatory eCRF contributions yield attraction of the center toward the surround. Conversely, flankers that are far enough from the test stimulus activate solely the far eCRF. This results in a net suppressive eCRF influence, which translates into repulsion of the center away from the surround.

A classical stimulus used in motion direction induction (Murakami & Shimojo, 1993, 1996) is a center-surround stimulus consisting of randomly moving dots with some coherence in the surround but no coherence in the center. The presentation of coherently moving dots in the surround elicits the illusory perception of coherent motion in the center—either in the same or in the opposite direction to that of the surround (depending on the experimental condition). In Murakami and Shimojo (1993, 1996), the diameter of the center and surround was fixed to w and $2w$, respectively, and the parameter w was varied systematically. This allowed the dimension of the overall stimulus to vary while the relative size of the center and surround regions were maintained.

Figure 6C shows psychophysical data (digitally extracted from Figure 6 and 7 in Murakami & Shimojo, 1996). For small stimulus sizes, the induced center movement is in the same direction as that of the surround. This corresponds to an attractive regime measured as a negative shift in the point of subjective equality (PSE). For larger sizes, the center induced movement reverses direction. This corresponds to a repulsive regime measured as a positive shift in PSE.

We found these results to be consistent with the model (Figure 6C). With a small enough stimulus, the coherently moving surround dots activate the model near eCRF exclusively (Figure 6A). This leads to a perceptual shift in the direction of the surround, consistent with the analogous case discussed in depth induction. At the population level, facilitation from the near eCRF tends to cause a sharpening in the population response around the surround stimulus value in an otherwise flat population response (as all motion directions are present in the center stimulus). As a result, the center stimulus looks "more like" the surround stimulus. As the stimulus size increases, the coherently moving surround dots start to activate an increasingly large proportion of the far surround (Figure 6B), which yields the opposite (repulsive) effect. At the population level, suppression from the far eCRF causes a small notch around the surround stimulus value and the center stimulus appears to look "less like" the surround.

Cooperative Activation of the Near and Far Surrounds

Thus far, we have seen that a variety of contextual phenomena can be explained as resulting from a balance between two opposing forces: an attractive force derived from facilitatory mechanisms originating from the near eCRF versus a repulsive force derived from suppressive mechanisms originating from the far eCRF. This competition can be tipped from attraction to repulsion by increasing the relative contribution of suppressive mechanisms originating from the far eCRF (relative to facilitatory mechanisms
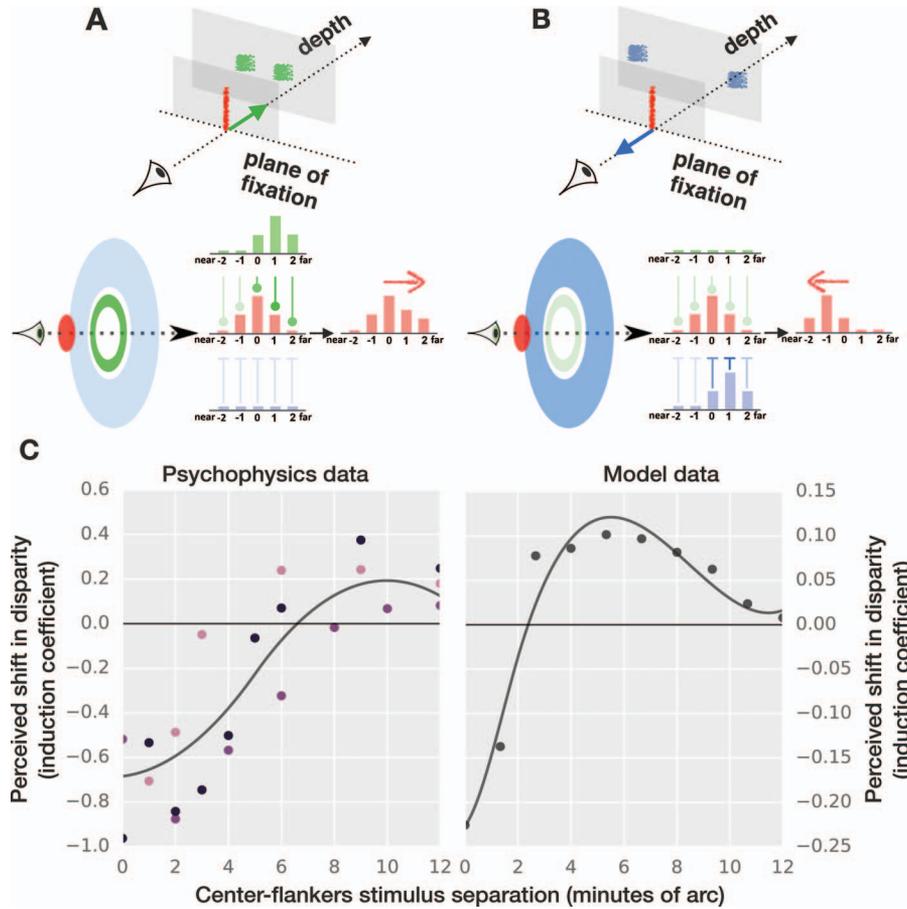
Figure 5. Depth induction. The exclusive activation of either the near or far eCRFs by flankers (as their separation vary) explains the existence of the shift from assimilation to contrast. The perceived depth of a binocular center stimulus (at zero disparity) is affected by binocular flankers located on either sides, presented at either crossed or uncrossed disparities. (A) Attraction: For short separations, flankers activate the near eCRF, which yields a net attractive force on the center population responses (toward the surround disparity) corresponding to a negative perceived shift. (B) Repulsion: For larger separations, flankers activate the far eCRF, which yields a net repulsive force on the center population responses (away from the surround disparity) corresponding to a positive perceived shift. (C) Psychophysics versus model data: Psychophysics data were digitally extracted from Figure 1 (upper panels) in (Westheimer & Levi, 1987) and fitted with splines. Both behavioral and model data capture the balance between stronger attraction toward the flankers at small separations, and weaker repulsion at larger separations. Note that the agreement between the model and human data is only qualitative as the perceived shifts in disparity are on different scales (the model underestimates the strength of the attractive regime in this illusion).

from the near eCRF) either by increasing the spatial extent of the stimulus (so as to activate an increasingly large proportion of the far eCRF) or by increasing the similarity between the center and surround stimulus (so as to increase the overlap between center and near surround population responses). However, we reasoned that if a surround stimulus takes on distinct and appropriate values in the near and far eCRFs (which we deem the near and far values), attraction toward the near value could go in the same direction as repulsion from the far value. Thus, the joint activation of the two eCRF subregions would cooperate rather than compete, resulting in an even larger perceptual shift compared to what would be achieved by presenting either the near or the far stimulus values alone.

The implication for color perception would be that assimilation, the attraction of the perceived center hue toward a neighboring inducing hue (i.e., the near hue), could be amplified by adding an appropriate outer hue (i.e., the far hue). This idea seems consistent with an "enhanced color shift" illusion discovered by Monnier and Shevell (2003), for which we provide a novel explanation. In the classical color assimilation illusion, a colored test ring (e.g., orange) is presented within a narrow uniform surround (e.g., purple or lime), which then attracts the test ring toward its own hue. This effect was found to be greatly amplified when patterned rings (e.g., alternating, thin rings of purple and lime) at an appropriate spatial frequency and phase were used in place of the uniform colored surround (Figure 1C). Such enhancement has also been docu-
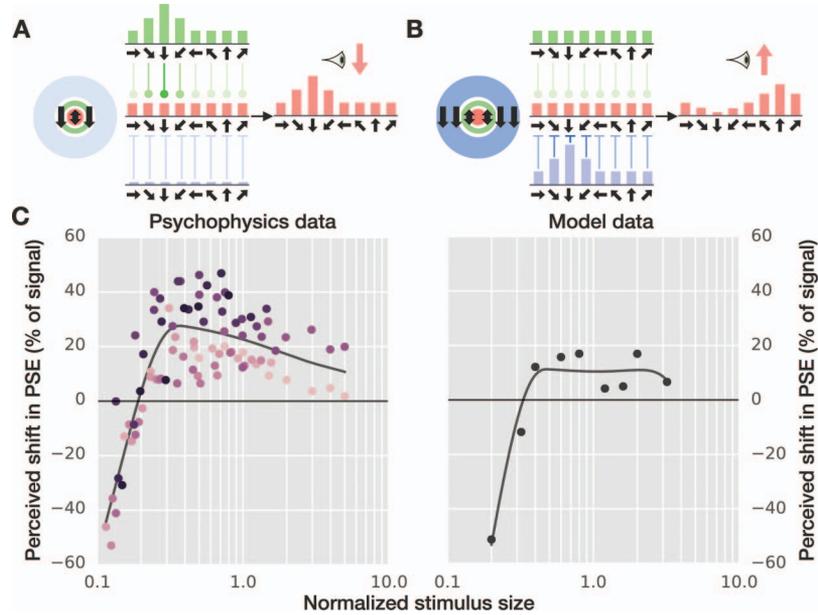
Figure 6. Motion induction. The increasing proportion of the far eCRF activated by larger stimuli explains the shift from assimilation to contrast. (A) Attraction: When the overall stimulus is small enough, the coherently moving surround dots activate the near eCRF exclusively, leading to motion assimilation (i.e., the center dots' direction appears the same as that of the surround dots). (B) Repulsion: Beyond a critical size, activation of the far inhibitory eCRF prevails, which leads to the opposite motion contrast effect (i.e., the center dots' direction appears opposite to that of the surround dots). (C) Psychophysics versus model data: Psychophysics data were digitally extracted from Figure 5 and 6 (Murakami & Shimojo, 1996) and fitted with splines. Both exhibit stronger attraction (negative ordinates) for smaller stimulus sizes, and weaker repulsion (positive ordinates) for larger sizes. Shifts in the point of subjective equality (PSE) were used as a proxy for shifts in perceived motion direction.

mented with achromatic stimuli (Anstis, 2006) and in brightness perception (Anstis, 2006; White, 1979).

Our model provides a simple explanation: As we have established, attraction (i.e., assimilation) toward say purple is caused by the activation of the near surround by a purple stimulus, with respect to a center region coinciding with the test ring. For the appropriate spatial frequency (Figure 7A), the additional lime-colored stimulus activates the far surround, leading to repulsion (i.e., contrast) *away* from lime, thus amplifying the perceptual shift toward purple as purple and lime are roughly perceptual complements. By reversing the phase of the color grating (Figure 7B), the colors stimulating the near and far eCRFs switch, leading to the same effect in the opposite direction.

The original psychophysics data (digitally extracted from the "6-min test" curves of Figure 5 from Shevell & Monnier, 2005) and model data are shown in Figure 7C. The model explains the existence of an optimal spatial frequency value, which maximizes the magnitude of the illusion. The spatial frequency of the stimulus controls the strength of the illusion because it determines how cleanly each of the inducing colors (e.g., lime and purple) activate the near and far eCRFs respectively for a CRF centered on the test ring. The model also postdicts that reversing the phases of the color grating leads to an effect with the same amplitude but opposite direction.

Critically, our explanation only depends on the appropriate hues falling within the near and far eCRFs regions; thus, we predict that the periodicity of the inducing stimulus per se is not important, as long as both regions are correctly stimulated. We show this with our own

versions of the illusion in Figures S6, suggesting that the illusion is just as strong, if not stronger, when the outer rings are replaced with a single uniform region that activates the far surround optimally (which is not the case for the original stimulus by Monnier & Shevell, 2003).

Discussion

We have described a computational neuroscience model of recurrent cortical circuits to account for classical (CRF) and extraclassical receptive field (eCRF) effects. The model was constrained by anatomical data and shown in our experiments to be consistent with V1 neurophysiology. In particular, the model unifies several electrophysiology phenomena such as (cross-orientation) normalization within the CRF (Busse et al., 2009) and modulation by the eCRF (including feature-selective suppression, see Trott & Born, 2015) into a computational neuroscience model of contextual integration.

The model further provides computational evidence for the existence of two eCRF mechanisms with complementary contributions to the CRF (a facilitatory near vs. suppressive far eCRF). In addition, the model predicts that an asymmetry between excitation and inhibition in the eCRF is needed: In our implementation, excitation depends on presynaptic activity only, whereas inhibition depends on both pre- and postsynaptic activities. Another model prediction is that connections within a hypercolumn are weakly
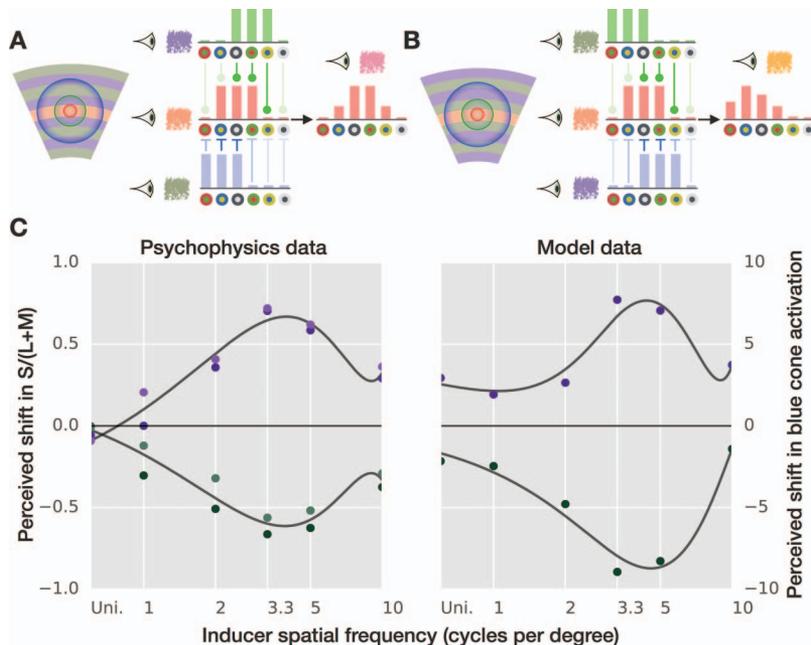
Figure 7. Enhanced color shifts. Cooperative activation of the near and far surrounds explains enhanced perceptual shifts. When distinct and "complementary" hues are used in a patterned surround (or inducer), the resulting shift in color perception of a test hue (here, orange) is amplified relative to a uniform surround of either hue. (A) Shift in one direction: For the optimal spatial frequency, one surround hue (e.g., purple) overlaps optimally with the near eCRF and the other one (e.g., lime) with the far eCRF. For the right color combination (as here with purple and lime which are complementary colors), this results in cooperating perceptual forces: a shift toward purple/away from lime. The colored patches next to the eyes correspond to the color decoded under the ideal observer. (B) Shift in the other direction: when purple and lime are switched. (C) Psychophysics versus model data: Psychophysics data were digitally extracted from Figure 5 (6-min test condition) in (Shevell & Monnier, 2005) and fitted with splines. Purple/green dots correspond to Condition A/B. "Uni'" stands for a uniform inducer composed of a single hue. For both behavioral and model data, there exists an optimal spatial frequency that maximizes the effect in either direction.

tuned or untuned, whereas connections across hypercolumns are tuned. We ran a systematic "lesioning" study on the model, whereby each of the hypothesized mechanisms was removed individually while all remaining parameters were optimized to fit behavioral data across all experiments (see supplementary experiments; Figures S7–S11).

Although our analysis revealed that a model which includes all assumed mechanisms performs best, we also found that one of the mechanisms (subtractive inhibition present in both inter- and intracolumnar connectivity) did not appear to be critical. In general, the model was robust to a range of parameter values (Figure S9–S11) even when relaxing the strict one-to-one mapping for the "tuned" connections from the eCRF onto the CRF (Figure S10). At the same time, while this study has focused on explaining behavioral data for an average observer, the model's variations associated with changes in individual parameter values may help explain intersubject variations observed experimentally (Figure S13).

The model distinguishes itself from previous work in succeeding to account for an array of disparate contextual phenomena spanning experimental conditions. Previous computational models have focused on explaining one or a few eCRF phenomena with an emphasis on surround suppression phenomena (see Angelucci &

Shushruth, 2013; Seriès et al., 2003 for reviews): Phenomenological models of center-surround processing (Cavanaugh, Bair, & Movshon, 2002; Sceniak, Hawken, & Shapley, 2001) and other normative models of visual coding (Coen-Cagli, Dayan, & Schwartz, 2012; Zhu & Rozell, 2013) have been shown to provide a good fit to single-unit contrast and size tuning responses. Recurrent network models have provided a mechanistic account for some of these phenomena (see Angelucci & Shushruth, 2013, for review) and have even led to testable predictions for single-unit electrophysiology (e.g., Rubin, Hooser, & Miller, 2015). But, none of these models have been systematically compared with a broad and diverse set of psychophysical experiments.

Furthermore, our model suggests that several contextual phenomena result from not one, but two opposing forces that yield systematic distortions on center population responses: repulsion from the far suppressive eCRF versus attraction toward the near facilitatory eCRF (see Figure 2B for representative population response dynamics). By revealing commonalities between seemingly disparate perceptual phenomena, the model has helped us establish a novel taxonomy of visual illusions: We have found that the way in which individual stimuli activate these near and far eCRFs (competitively, exclusively or cooperatively; organized by columns in Figure 8) affects the qualitative behavior of the model.
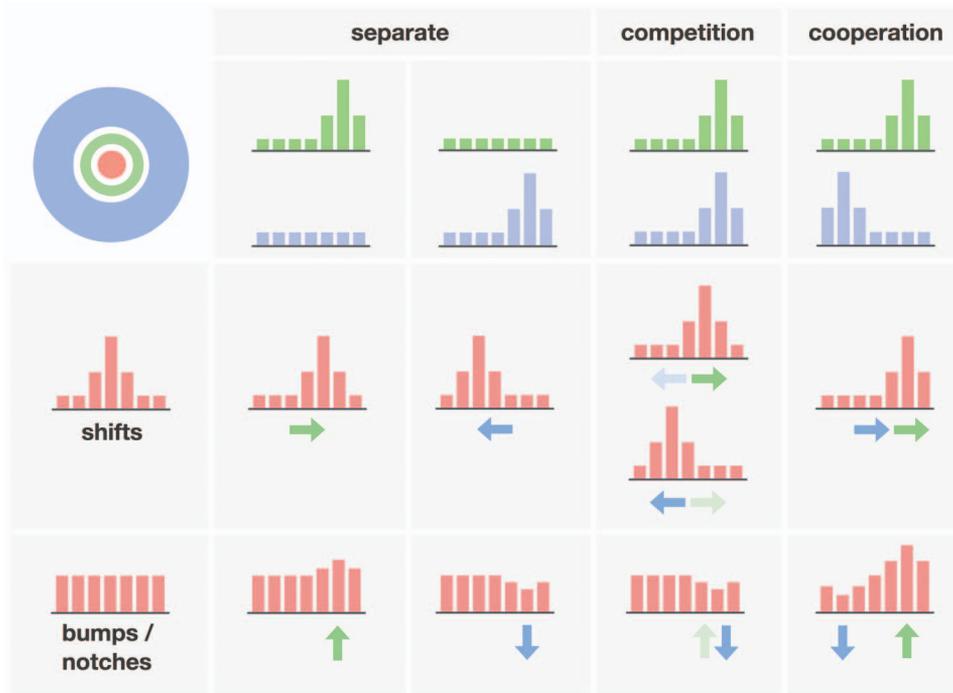
Figure 8. A new taxonomy of contextual phenomena. Rows: Contextual phenomena manifest themselves in the model either as (a) shifts with peaked center population response curves (unambiguous stimuli), or (b) bumps/notches with broad/uniform center population response curves (ambiguous stimuli). Columns: Center-surround stimuli activate the near and far eCRFs in three typical ways: (a) either one separately; (b) both competitively (i.e., near and far eCRFs each induce shifts that tend to stymie each other; green and blue arrows, resp.); and (c) both cooperatively (i.e., the shifts induced by the near and far eCRFs are in the same direction, amplifying the perceptual shift). See Table S2 for a version of this table populated with representative psychophysics studies for each individual case.

A Novel Taxonomy of Contextual Phenomena

Contextual stimuli that yield competitive activation of the near versus far eCRFs were found for a set of tilt illusions including orientation (Goddard et al., 2008; O'Toole and Wenderoth, 1977), motion (Kim & Wilson, 1997), and hue (Klauke & Wachtler, 2015, also known as color induction). In these stimuli, the surround spatially overlaps with both the near facilitatory and far suppressive eCRFs—activating them both competitively. Because of the asymmetry between excitation and inhibition in the model, repulsion from the surround stimulus prevails when the center and surround population responses overlap, that is, when the center and surround stimuli are perceptually similar. Conversely, attraction toward the surround stimulus prevails when such overlap is minimal, such as when the center and surround stimuli are perceptually dissimilar.

Previous authors (Clifford, 2014; Goddard et al., 2008; Kim & Wilson, 1997; Klauke & Wachtler, 2015) have suggested that surround inhibition may be key to explaining the repulsive regime in tilt effects (see supplementary Discussion for a more in-depth discussion). The proposed mechanisms, which include shifts in neural tuning curves (Klauke & Wachtler, 2015), varying inhibition strength depending on the relative center-surround orientation (Goddard et al., 2008), or recurrent center-surround interactions (Kim & Wilson, 1997) are all consistent with the proposed mechanistic model. In addition, the present study offers a plausible computational explana-

tion for not only the existence of a repulsive regime but also an attractive one for certain classes of stimuli, in agreement with a host of experimental data (Goddard et al., 2008; Kim & Wilson, 1997; O'Toole and Wenderoth, 1977; Westheimer & Levi, 1987).

Another model postdiction is the absence of such attractive regime for contextual stimuli that yield broad-band population responses (arising because of broad neural tuning for the perceptual domain or because the stimulus is inherently ambiguous as in textures with little coherent orientation). For such stimuli, the overlap between center and surround population responses remains large even for maximally dissimilar center and surround stimuli, and the only discernible contextual effect is governed by the repulsive regime. Interestingly, the model achieves its quantitative fit for motion induction experiments via a broadening of neural tuning curves for motion direction compared with orientation, which is consistent with V1 electrophysiology data (Albright et al., 1984; Ringach et al., 2002; see also supplementary Discussion and Figure S12 for a more in-depth discussion).

Stimuli that activate exclusively the near or the far eCRF have been used in classical induction experiments in the domain of depth (Westheimer & Levi, 1987) and motion (Murakami & Shimojo, 1996). In the model, consistent with the proposal by Murakami and Shimojo (1993), rescaling a stimulus display (or similarly, varying the relative spacing between center and surround stimuli) yields a reversal from

attraction to repulsion. A surround stimulus close to the center or presented at a small scale tends to predominantly activate the facilitatory near eCRF, yielding attraction toward the surround. A surround stimulus farther from the center or presented at a larger scale tends to activate the suppressive far eCRF to a greater extent, yielding repulsion away from the surround.

For the last set of illusions called enhanced color shifts (Shevell & Monnier, 2005), the contextual (or surround) stimulus took on "complementary" optimal values in the near and the far eCRFs. As a result, shifts induced by either region of the eCRF tended to cooperate rather than compete with one another. This resulted in a perceptual shift greater than a purely attractive effect involving only the near eCRF or a purely repulsive effect involving only the far eCRF. In addition, the spatial antagonism of the model eCRF captures the existence of an optimal spatial frequency (and phase) such that a cycle of the surround stimulus coincides maximally with the near and far eCRFs. More generally, the model confirms the consensus that assimilation predominates at higher spatial frequencies and finer scales whereas contrast emerges at lower spatial frequencies and coarser scales (Anstis, 2006; Monnier & Shevell, 2003; Murakami & Shimojo, 1993, 1996; Shevell & Monnier, 2005; White, 1979, 1981).

Shevell and Monnier (2005) have previously modeled enhanced color shifts through an S-cone color opponent model (see supplementary Discussion). As in our model, such center-surround spatial antagonism results in the existence of an optimal spatial frequency. By design, their model, however, predicts the existence of enhanced perceptual shifts for S-cone stimuli only. On the other hand, our model predicts that such enhanced color shifts should persist for surround stimuli that do not activate S cones. We have created such stimuli (Figure S6) for the reader to judge for themselves but careful psychophysical work using a properly calibrated monitor will be needed to test this model prediction.

We have found a further subdivision of the above taxonomy (rows in Figure 8) based on a more detailed characterization of the center stimulus and, in particular, whether it is ambiguous (e.g., incoherently moving random dot or achromatic stimuli) or not (e.g., high-contrast gratings and bars, highly coherent moving random dot or saturated chromatic stimuli). Unambiguous center-surround stimuli yield peaked, narrow population responses (simulation results in Figure 2B–C) across the visual field. The effect of the surround on a peaked center population response is to shift its center of mass, biasing the associated decoded value accordingly (see supplementary Discussion for a discussion of the evidence of such shifts in neurophysiology studies). The shift is either toward (attraction) or away (repulsion) from the peak of the surround population response depending on whether the net effect of the eCRF is facilitatory (Figure 2B) or suppressive (Figure 2C). Ambiguous center stimuli yield broad-band (or even flat) center population responses (Figure 2D–E). These can be distorted by a peaked surround population in two ways: a bump centered at the surround stimulus value when tuned facilitation from the eCRF prevails (Figure 2D) or a notch at the surround value when tuned suppression does (Figure 2E).

Table S2 shows how the literature fits in the proposed taxonomy. Note that some table entries are missing for certain visual modalities, which suggests more contextual phenomena remain to be found (e.g., cooperative shifts in orientation, which would result in an "enhanced orientation tilt"). Overall, the present study thus provides a vivid

example of how computational models may help reinterpret results as well as summarize and integrate disparate phenomena.

Open Questions

The neural tuning curves considered in this work (orientation, disparity, motion direction, color opponent) can be found in relatively low-level areas of the visual cortex, such as V1, V2, or MT. Thus, the consistency between model and behavioral data is all the more remarkable as many of the illusions studied here are likely to also involve higher-level visual processes which are known to affect perception including perceptual organization and grouping (e.g., Herzog & Clarke, 2014), attention and other top-down feedback (Gilbert & Li, 2013) including surface-based and other filling-in processes (Grossberg & Todorović, 1988). The model's ability to account for contextual interactions may be limited to the relatively simple stimuli such as the bars and gratings tested here. We expect the model to fail to account for human data for more complex contextual stimuli defined by objects or shapes (e.g., Herzog & Clarke, 2014). At the very least, a more complete model would likely require multiple stages of processing as well as mechanisms of filling-in and contour extraction (Grossberg & Todorović, 1988). Similarly, considering tuning curves found in higher-level areas, such as tuning to hue observed in V4/PIT neurons (as opposed to color-opponent V1 neurons considered here, see Conway, Moeller, & Tsao, 2007) could also improve the fit with experimental data (though hue tuning remains controversial, see Conway, 2009; Mollon, 2009).

More generally, the present model leaves open any role for attention. Indeed, recent work has shown that attention seems to be shifting both the CRF and eCRF independently toward the attended location (Anton-Erxleben, Stephan, & Treue, 2009). It is likely that attention (not accounted for in the present model) may have played a role in shaping the pattern of observed behavioral results. In our simulations, the CRF size was scaled to the center of the stimuli—a role that could possibly be endued to attention (Carandini, 2012). Indeed, one of the main mechanisms in the present model—that of complementary excitatory and inhibitory surround mechanisms—is a key mechanism in one of the leading models of spatial attention (Tsotsos, Culhane, & Cutzu, 2001). In this model, an annular region of inhibition creates a negative attentional field surrounding the region of perceptual facilitation centered on the attended target. In addition, modeling work has also suggested that top-down influences may "gate" the effective contextual interactions mediated by long-range horizontal connections (Setić & Domijan, 2008).

Anatomical data to constrain the patterns of recurrent connectivity (both within and across hypercolumns) in the model are scarce. The near and far eCRFs as modeled are likely to constitute, at best, coarse approximations for more complex patterns of anatomical connections. In particular, both the spatial extent and the relative strength of the near and far eCRFs relative to that of the CRF were held constant across experiments. Given that the experiments considered throughout spanned a range of visual stimuli across modalities and sizes, it is likely that these phenomena recruit neural populations in different cortical areas and visual eccentricities. It is also likely that variations in experimental factors lead to differences in how the center and surround capture attention. We thus expect improvements in the model's quantita-

tive fit by considering additional parameters to control the spatial extent and the relative strength of the near and far eCRFs (e.g., as done in Goddard et al., 2008).

We have also left open the question of whether the connectivity in the near and far eCRFs would draw on slow intra-areal lateral connections or fast intra-areal feedback connections (Angelucci et al., 2002; Angelucci, Levitt et al., 2002; Shushruth & Ichida, 2009). Such lack of refinement, in addition to a lack of realistic modeling of excitatory and inhibitory synapses and their relative timing (Vinck, Womelsdorf, & Fries, 2013), negatively impacts our ability to make predictions about the precise time course of the contextual effects modeled. We also expect that a resolution on the question of feedback versus lateral connectivity will be needed to account for some of the electrophysiology phenomena we left aside in the present study including the known contrast dependence of the eCRF size (see Angelucci & Shushruth, 2013, for review) or cross-orientation enhancements (Levitt & Lund, 1997; Sillito, Grieve, Jones, Cudeiro, & Davis, 1995). Another hypothesis of the model that has yet to be confirmed is the existence of cortical columns for all visual domains beyond orientation (see Sincich & Horton, 2005, for review). At present, the existence of cortical columns for color (Dow, 2002), motion (DeAngelis, Ghose, Ohzawa, & Freeman, 1999), and binocular disparity (DeAngelis & Newsome, 1999) is only partially supported by neurophysiology evidence.

We have assumed for simplicity that the near eCRF is circular (i.e., isotropic with respect to the topography of the visual field). There is, however, evidence for anisotropies in the pattern of horizontal connections between cortical columns (as orientation-tuned cells tend to be more often connected when they share the same selectivity and their CRFs are aligned along an axis parallel to their preferred orientation, see Bosking, Zhang, Schofield, & Fitzpatrick, 1997). There is also more direct evidence for anisotropies in the shape of the eCRF (i.e., various elongations over a wide range of orientations and widths, see Tanaka & Ohzawa, 2009). The function of these anisotropies has been attributed to the computation of higher-order features (including contrast- or texture-defined boundaries) as well as contour integration and pop-out (Hess, Hayes, & Field, 2003; Stemmler, Usher, & Niebur, 1995; Tanaka & Ohzawa, 2009). Future work should test whether these phenomena can be accounted for with a model extension that incorporates such eCRF anisotropies.

More generally our study did not address the role of the perceptual biases and the altered discriminability that arise because of surround mechanisms. It has been suggested that surround mechanisms could constitute one of the primary mechanisms for predictive coding and Bayesian inference type of computations (see Schwartz et al., 2007, for review). We speculate that the computational mechanisms revealed by the contextual illusions studied here play a key role in shaping invariant population codes for object constancy. We have obtained preliminary results suggesting that tuned suppression from the far eCRF may improve the accurate decoding of surface reflectances across changes in illumination (i.e., color constancy; see Mély & Serre, abstract presented at the 2015 Vision Science Society meeting), by helping to discount undesirable variations in center population responses caused by changes in the light source. (This is reminiscent of a color constancy algorithm by Land & McCann, 1971 known as the Retinex.) This raises the intriguing possibility that at least some of the mechanisms unraveled here may support other forms of perceptual

constancy beyond color. Further work will be needed to quantify how object transformations such as changes in illumination or depth affect neural population responses tuned to orientation or binocular disparity and what computational mechanisms are needed to help discount these nuisances. Nonetheless, the ability of the model to account for the variety and complexity of contextual illusions provides computational evidence for a novel canonical cortical circuit shared across visual modalities.

Materials and Method

Additional methods may be found in Supplemental Materials and Methods.

Model Connectivity

A column centered at location (x, y) contains a complete set of N units with CRFs centered at (x, y) and tuning values covering the full range $\theta_{k=1...N}$ (e.g., orientation tuning curves are regularly centered at values $\theta_k \in [0, 180°]$). Tuning curves are idealized—either bell-shaped for disparity (Cumming & Parker, 1997), motion direction (Albright et al., 1984), and orientation (Ringach, Hawken, & Shapley, 1997) or monotonic for color opponency (Johnson, Hawken, & Shapley, 2001).

Each unit (x, y, k) receives excitation Q_k^{xy}, assumed to be weakly tuned and originating from within the same hypercolumn:

$$Q_k^{xy} = \sum_{j=1...N} w_{jk} X_j^{xy} \text{ s.t. } w_{jk} \sim \mathcal{N}(\theta_k, \zeta), \quad (1)$$

where w_{jk} corresponds to excitatory weights between units k and j (with selectivities θ_k and θ_j, respectively) and X_j^{xy} to input activity at location (x, y, j). We assume these weights to be normally distributed, centered at a target unit tuning preference θ_k with standard deviation ζ. Some tuning (albeit weak) is necessary in order to prevent intracolumnar excitation from flattening the population responses to well-defined stimuli. In the color domain, we consider color-opponent model units with monotonic tuning curves. Instead of drawing weights from a normal distribution, which only makes sense for bell-shaped tuning curves, we set $w_{kk} = (\zeta\sqrt{2\pi})^{-1}$ and $w_{jk} = $ const. (when $j \neq k$; under the constraint that the weights sum up to 1).

Each unit (x, y, k) also receives some inhibition U^{xy}, assumed to be untuned and originating from within the same hypercolumn:

$$U^{xy} = \frac{1}{N} \sum_{j=1...N} Y_j^{xy}, \quad (2)$$

where Y_j^{xy} is the output activity of unit j at location (x, y). Unlike the excitation which is linear, inhibition is nonlinear, of the shunting kind (Grossberg & Todorović, 1988), and acts on the output of the presynaptic units (Equation 5). Combined with broad tuning, this allows populations of cells with coinciding CRFs to be significantly driven by a common input. Such a mechanism was used in the model by Shushruth et al. (2012) and found experimentally to be critical to reproduce nonlinear neural effects such as stimulus-matched surround suppression (Trott & Born, 2015). Because the local inhibition is untuned, its strength is independent of a unit selectivity θ_k, and we drop the subscript k for simplicity.

Furthermore, unit (x, y, k) also receives tuned excitation P_k^{xy} from other units with the same selectivity θ_k that are located within its near eCRF $\mathbb{N}^{xy}$, defined relatively to position (x, y):

$$P_k^{xy} = \frac{1}{|\mathbb{N}^{xy}|} \sum_{u,v \in \mathbb{N}^{xy}} X_k^{u,v}. \tag{3}$$

Similarly, unit (x, y, k) also receives tuned inhibition T_k^{xy} from other units with the same selectivity θ_k that are located within its far eCRF $\mathbb{F}^{xy}$, defined relatively to position (x, y):

$$T_k^{xy} = \frac{1}{|\mathbb{F}^{xy}|} \sum_{u,v \in \mathbb{F}^{xy}} Y_k^{u,v}. \tag{4}$$

As in Equation 2, inhibition is nonlinear and acts on the output $Y_k^{u,v}$ of unit k at location (u, v).

Neural Field Model

Neural field dynamics obey the following equations:

$$\eta \partial_t X_k^{xy} + \varepsilon^2 X_k^{xy} = \left[\xi L_k^{xy} - (\alpha X_k^{xy} + \mu) U^{xy} - (\beta X_k^{xy} + \nu) T_k^{xy} \right]_+$$
$$\tau \partial_t Y_k^{xy} + \sigma^2 Y_k^{xy} = \left[\gamma P_k^{xy} + \delta Q_k^{xy} \right]_+, \tag{5}$$

where the feed-forward input L_k^{xy} drives every unit (x, y, k) across the visual field; each is represented by its recurrent input X_k^{xy} and output Y_k^{xy}. The parameters α, β, δ, γ, μ and ξ can be interpreted as synaptic weights (see Table S1 for values used) which control the amount of intra- and intercolumnar excitation and inhibition (Equations 1–4). The steady-state solution is computed using numerical integration (with convergence typically taking ~50 iterations). Population responses at the steady-state $\bar{Y}_k^{xy}$ are a very nonlinear function of the model input L_k^{xy}.

For each unit, the steady-state input and output are given by $\bar{X}_k^{x,y}$ and $\bar{Y}_k^{x,y}$, resp. Due to the rectifying nonlinearity in the dynamics (Equation 5), at steady-state, $\bar{X}_k^{x,y}$ and $\bar{Y}_k^{x,y}$ can either be equal to zero, or to the values below:

$$\bar{X}_k^{x,y} = \frac{\xi L_k^{x,y} - \mu \bar{U}^{x,y} - \nu \bar{T}_k^{x,y}}{\varepsilon^2 + \alpha \bar{U}^{x,y} + \beta \bar{T}_k^{x,y}}$$
$$\bar{Y}_k^{x,y} = \frac{\gamma \bar{P}_k^{x,y} + \delta \bar{Q}_k^{x,y}}{\sigma^2}, \text{ with:} \tag{6}$$

$$\bar{U}^{x,y} = \frac{1}{N} \sum_{j=1...N} \bar{Y}_j^{x,y}$$
$$\bar{T}_k^{x,y} = \frac{1}{|\mathbb{F}^{x,y}|} \sum_{u,v \in \mathbb{F}^{x,y}} \bar{Y}_k^{u,v}$$
$$\bar{P}_k^{x,y} = \frac{1}{|\mathbb{N}^{x,y}|} \sum_{u,v \in \mathbb{N}^{x,y}} \bar{X}_k^{u,v} \tag{7}$$
$$\bar{Q}_k^{x,y} = \sum_{j=1...N} w_{j,k} \bar{X}_j^{x,y}$$

Tuning Curves

The model constitutes an example of tuning curve population model (Rust et al., 2006; Schwartz et al., 2007). We considered two kinds of tuning curves: bell-shaped (orientation, motion direction, binocular disparity) and monotonic, nonsaturating tuning curves (color). All tuning curves were normalized, that is, the maximum unit activity was set to be equal to 1. For nonangular variables (e.g., disparity), bell-shaped tuning curves were parametrized as Gaussian functions:

$$f(\theta | \theta_k, \sigma) = \exp\left(-\frac{(\theta - \theta_k)^2}{2\sigma^2} \right), \tag{8}$$

with preferred stimulus value θ_k and tuning bandwidth σ. When

the variable was circular (e.g., orientation, motion direction), we modeled the tuning curve as a von Mises function instead:

$$f(\theta | \theta_k, \sigma) = \exp\left(\frac{\cos\left((\theta - \theta_k) \frac{2\pi}{I} \right) - 1}{2\sigma^2} \right), \tag{9}$$

where I indicates the length of domain of the tuning curve (e.g., π for orientation vs. 2π for direction). We generally sampled on the order of 30 tuning curve centers regularly spaced in the domain of the considered visual modality. We found that the number of tuning curve centers considered did not impact our results as long as it was large enough.

Monotonic, nonsaturating tuning curves for color were derived by converting stimuli to idealized cone responses first, which were then mapped to opponent color channels similarly to Zhang, Barhomi, and Serre (2012). These included red-on/green-off (R^+G^-), green-on/red-off (G^+R^-), blue-on/yellow-off (B^+Y^-), and yellow-on/blue-off (Y^+B^-), alongside with a pair of luminance-sensitive channels, selective for lighter (Wh^+Bl^-) and darker (Bl^+Wh^-) stimuli.

Model Parameters

All circuit parameters were held constant in all comparisons with psychophysics data. They were determined a priori in order to reproduce key neurophysiology data (see supplementary experiments) and were held constant for all visual modalities except color because of a qualitative difference in tuning curve (see Equation 1). In all subsequent experiments, only two variables were allowed to vary: the stimulus scale and the tuning bandwidth for model units.

The stimuli used in psychophysics studies varied greatly—recruiting neural populations subtending a wide range of CRF (and eCRF) sizes and eccentricities, possibly spanning different visual areas. Rather than adjusting the size of the model CRFs and eCRFs for individual experiments, which would have required structural changes to the model, we instead varied the stimulus scale. Because the connectivity between model hypercolumns was held fixed, this is somewhat akin to varying the magnification factor in the model. Critically, this yielded broad estimates for CRF (and eCRF) sizes within a biologically realistic range (from a fraction of a degree of visual angle to a couple of degrees). The width of the idealized tuning curves which is common to all model units was optimized separately for each experiment (see supplementary materials and Methods for details). We have confirmed that our key model predictions were robust over a range of these parameter values.

Ideal Neural Observer Model

We used an ideal neural observer model to map model population responses to decoded sensory variables, which can then be compared with behavioral judgments collected experimentally. We used a population vector model (Georgopoulos, Schwartz, & Kettner, 1986), in which each unit votes for its preferred sensory value in proportion to its activity (normalized by the summed activities of all units within the same column). This model is not appropriate for color because of the tuning along opponent color pairs rather than a hue angle. Instead, we used cross-validated ridge regression to decode the sine and cosine of hue.

References

Albright, T. D., Desimone, R., & Gross, C. G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *Journal of Neurophysiology, 51,* 16–31.

Allman, J., Miezin, F., & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception, 14,* 105–126.

Angelucci, A., Levitt, J. B., & Lund, J. S. (2002). Anatomical origins of the classical receptive field of single neurons in macaque visual cortical area V1. *Progress in Brain Research, 136,* 373–388.

Angelucci, A., Levitt, J. B., Walton, E. J. S., Hupe, J. M., Bullier, J., & Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *Journal of Neuroscience, 22,* 8633–8646.

Angelucci, A., & Shushruth, S. (2013). Beyond the classical receptive field: Surround modulation in primary visual cortex. In J. S. Werner & L. M. Chalupa (Eds.), *The new visual neurosciences* (pp. 425–444). Cambridge, MA: MIT Press.

Anstis, S. (2006). White's effect in lightness, color, and motion. In M. R. M. Jenkin & L. R. Harris (Eds.), *Seeing spatial form* (pp. 91–100). New York, NY: Oxford University Press.

Anton-Erxleben, K., Stephan, V. M., & Treue, S. (2009). Attention reshapes center-surround receptive field structure in macaque cortical area MT. *Cerebral Cortex, 19,* 2466–2478.

Bosking, W. H., Zhang, Y., Schofield, B., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *The Journal of Neuroscience, 17,* 2112–2127.

Bradley, D. C., & Andersen, R. A. (1998). Center-surround antagonism based on disparity in primate area MT. *Journal of Neuroscience 18,* 7552–7565.

Briggs, F., & Usrey, W. M. (2011). Distinct mechanisms for size tuning in primate visual cortex. *The Journal of Neuroscience, 31,* 12644–12649.

Bringuier, V., Chavane, F., Glaeser, L., & Frégnac, Y. (1999). Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science, 283,* 695–699.

Busse, L., Wade, A. R., & Carandini, M. (2009). Representation of concurrent stimuli by population activity in visual cortex. *Neuron, 64,* 931–942.

Carandini, M. (2012). From circuits to behavior: A bridge too far? *Nature neuroscience, 15,* 507–509.

Carandini, M., & Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science, 264,* 1333–1336.

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

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002). Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. *Journal of Neurophysiology, 88,* 2547–2556.

Clifford, C. W. G. (2014). The tilt illusion: Phenomenology and functional implications. *Vision Research, 104,* 3–11.

Coen-Cagli, R., Dayan, P., & Schwartz, O. (2012). Cortical surround interactions and perceptual salience via natural scene statistics. *PLoS Computational Biology 8,* e1002405.

Conway, B. R. (2009). Color vision, cones, and color-coding in the cortex. *The Neuroscientist: A review journal bringing neurobiology, neurology and psychiatry, 15,* 274–290.

Conway, B. R., Moeller, S., & Tsao, D. Y. (2007). Specialized color modules in macaque extrastriate cortex. *Neuron, 56,* 560–573.

Cumming, B. G., & Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature, 389,* 280–283.

DeAngelis, G. C., Freeman, R. D., & Ohzawa, I. (1994). Length and width tuning of neurons in the cat's primary visual cortex. *Journal of Neurophysiology, 71,* 347–374.

DeAngelis, G. C., Ghose, G. M., Ohzawa, I., & Freeman, R. D. (1999). Functional micro-organization of primary visual cortex: Receptive field analysis of nearby neurons. *Journal of Neuroscience, 19,* 4046–4064.

DeAngelis, G. C., & Newsome, W. T. (1999). Organization of disparity-selective neurons in macaque area MT. *Journal of Neuroscience, 19,* 1398–1415.

Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A., & Suarez, H. H. (1995). Recurrent excitation in neocortical circuits. *Science, 269,* 981–985.

Dow, B. M. (2002). Orientation and color columns in monkey visual cortex. *Cerebral Cortex, 12,* 1005–1015.

Eagleman, D. M. (2001). Visual illusions and neurobiology. *Nature Reviews Neuroscience, 2,* 920–926.

Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science, 233,* 1416–1419.

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. Nature reviews. *Neuroscience, 14,* 350–363.

Goddard, E., Clifford, C. W. G., & Solomon, S. G. (2008). Centre-surround effects on perceived orientation in complex images. *Vision Research, 48,* 1374–1382.

Grossberg, S., & Todorović, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception and Psychophysics, 43,* 241–277.

Heeger, D. J. (1993). Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *Journal of Neurophysiology, 70,* 1885–1898.

Herzog, M. H., & Clarke, A. M. (2014). Why vision is not both hierarchical and feedforward. *Frontiers in Computer Neuroscience, 8,* 1–5.

Hess, R. F., Hayes, A., & Field, D. J. (2003). Contour integration and cortical processing. *Journal of Physiology Paris, 97,* 105–119.

Hubel, D., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology, 195,* 215–243.

Johnson, E. N., Hawken, M. J., & Shapley, R. (2001). The spatial transformation of color in the primary visual cortex of the macaque monkey. *Nature Neuroscience, 4,* 409–416.

Jones, H. E., Grieve, K. L., Wang, W., & Sillito, A. M. (2001). Surround suppression in primate V1. *Journal of Neurophysiology, 86,* 2011–2028.

Kim, J., & Wilson, H. R. (1997). Motion integration over space: Interaction of the center and surround motion. *Vision Research, 37,* 991–1005.

Klauke, S., & Wachtler, T. (2015). "Tilt" in color space: Hue changes induced by chromatic surrounds. *Journal of Vision, 15,* 17.

Land, E. H., & McCann, J. J. (1971). Lightness and Retinex theory. *Journal of the Optical Society of America, 61,* 1–11.

Lee, W. C. A., Bonin, V., Reed, M., Graham, B. J., & Hood, G. (2016). Anatomy and function of an excitatory network in the visual cortex. *Nature, 532,* 370–374.

Levitt, J. B., & Lund, J. S. (1997). Contrast dependence of contextual effects in primate visual cortex. *Nature, 387,* 73–76.

Li, C. Y., Lei, J. J., & Yao, H. S. (1999). Shift in speed selectivity of visual cortical neurons: A neural basis of perceived motion contrast. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 4052–4056.

Marshak, W., & Sekuler, R. (1979). Mutual repulsion between moving visual targets. *Science, 205,* 1399–1401.

Mollon, J. D. (2009). A neural basis for unique hues? *Current Biology, 19,* 441–442.

Monnier, P., & Shevell, S. K. (2003). Large shifts in color appearance from patterned chromatic backgrounds. *Nature neuroscience, 6,* 801–802.

Murakami, I., & Shimojo, S. (1993). Motion capture changes to induced motion at higher luminance contrasts, smaller eccentricities, and larger inducer sizes. *Vision Research, 33,* 2091–2107.

Murakami, I., & Shimojo, S. (1996). Assimilation-type and contrast-type bias of motion induced by the surround in a random-dot display: Evidence for center-surround antagonism. *Vision Research, 36,* 3629–3639.

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

O'Toole, B., & Wenderoth, P. (1977). The tilt illusion: Repulsion and attraction effects in the oblique meridian. *Vision Research, 17,* 367–374.

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

Petrov, Y., Carandini, M., & McKee, S. (2005). Two distinct mechanisms of suppression in human vision. *The Journal of Neuroscience, 25,* 8704–8707.

Ringach, D. L., Hawken, M. J., & Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. *Nature, 387,* 281–284.

Ringach, D. L., Shapley, R. M., & Hawken, M. J. (2002). Orientation selectivity in macaque V1: Diversity and laminar dependence. *Journal of Neuroscience, 22,* 5639–5651.

Rubin, D. B., Hooser, S. D. V., & Miller, K. D. (2015). The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron, 85,* 1–51.

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience, 9,* 1421–1431.

Sceniak, M. P., Hawken, M. J., & Shapley, R. (2001). Visual spatial characterization of macaque V1 neurons. *Journal of Neurophysiology, 85,* 1873–1887.

Sceniak, M. P., Ringach, D. L., Hawken, M. J., & Shapley, R. (1999). Contrast's effect on spatial summation by macaque V1 neurons. *Nature Neuroscience, 2,* 733–739.

Schein, S. J., & Desimone, R. (1990). Spectral properties of V4 neurons in the macaque. *Journal of Neuroscience, 10,* 3369–3389.

Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. *Nature Reviews*. Neuroscience, 8, 522–535.

Sengpiel, F. (1997). Binocular rivalry: Ambiguities resolved. *Current Biology, 7,* 447–450.

Sengpiel, F., Sen, A., & Blakemore, C. (1997). Characteristics of surround inhibition in cat area 17. *Experimental Brain Research, 116,* 216–228.

Seriès, P., Lorenceau, J., & Frégnac, Y. (2003). The "silent" surround of V1 receptive fields: Theory and experiments. *Journal of Physiology, 97,* 453–474.

Setić, M., & Domijan, D. (2008). Modeling the top-down influences on the lateral interactions in the visual cortex. *Brain Research, 1225,* 86–101.

Shevell, S. K., & Monnier, P. (2005). Color shifts from S-cone patterned backgrounds: Contrast sensitivity and spatial frequency selectivity. *Vision Research, 45,* 1147–1154.

Shushruth, S., & Ichida, J. (2009). Comparison of spatial summation properties of neurons in macaque V1 and V2. *Journal of Neurophysiology, 1,* 2069–2083.

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

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

Sincich, L., & Horton, J. (2005). The circuitry of V1 and V2: Integration of color, form, and motion. *Annual Review of Neuroscience, 28,* 303–326.

Sit, Y. F., Chen, Y., Geisler, W. S., Miikkulainen, R., & Seidemann, E. (2009). Complex Dynamics of V1 Population Responses Explained by a Simple Gain-Control Model. *Neuron, 64,* 943–956.

Smith, V. C., Jin, P. Q., & Pokorny, J. (2001). The role of spatial frequency in color induction. *Vision Research, 41,* 1007–1021.

Stemmler, M., Usher, M., & Niebur, E. (1995). Lateral interactions in primary visual cortex: A model bridging physiology and psychophysics. *Science, 269,* 1877–1880.

Stepanyants, A., Hirsch, J. A., Martinez, L. M., Kisvarday, Z. F., Ferecsko, A. S., & Chklovskii, D. B. (2008). Local potential connectivity in cat primary visual cortex. *Cerebral Cortex, 18,* 13–28.

Tanaka, H., & Ohzawa, I. (2009). Surround suppression of V1 neurons mediates orientation-based representation of high-order visual features. *Journal of Neurophysiology, 101,* 1444–1462.

Trott, A. R., & Born, R. T. (2015). Input-gain control produces feature-specific surround suppression. *Journal of Neuroscience, 35,* 4973–4982.

Tsotsos, J. K., Culhane, S., & Cutzu, F. (2001). Visual attention and cortical circuits. In J. Braun, C. Koch, & J. L. Davis (Eds.), *From theoretical foundations to a hierarchical circuit for selective attention* (pp. 285–306). Cambridge, MA: MIT Press.

Vinck, M., Womelsdorf, T., & Fries, P. (2013). Gamma-band synchronization and information transmission. In R. Q. Quiroga & S. Panzeri (Eds.), *Principles of neural coding* (pp. 449–470). Boca Raton, FL: CRC Press.

Wachtler, T., Sejnowski, T. J., & Albright, T. D. (2003). Representation of color stimuli in awake macaque primary visual cortex. *Neuron, 37,* 681–691.

Weliky, M., Kandler, K., Fitzpatrick, D., & Katz, L. C. (1995). Patterns of excitation and inhibition evoked by horizontal connections in visual cortex share a common relationship to orientation columns. *Neuron, 15,* 541–552.

Westheimer, G. (1986). Spatial interaction in the domain of disparity signals in human stereoscopic vision. *The Journal of Physiology, 370,* 619–629.

Westheimer, G., & Levi, D. M. (1987). Depth attraction and repulsion of disparate foveal stimuli. *Vision Research, 27,* 1361–1368.

White, M. (1979). A new effect of pattern on perceived lightness. *Perception, 8,* 413–416.

White, M. (1981). The effect of the nature of the surround on the perceived lightness of grey bars within square-wave test gratings. *Perception, 10,* 215–230.

Zhang, J., Barhomi, Y., & Serre, T. (2012). A New Biologically Inspired Color Image Descriptor. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, & C. Schmid (Eds.), *Computer Vision – ECCV 2012* (pp. 312–324). Berlin, Germany: Springer.

Zhu, M., & Rozell, C. J. (2013). Visual nonclassical receptive field effects emerge from Sparse coding in a dynamical system. *PLoS Computational Biology, 9,* 1–15.