

# Stimulus conflation and tuning selectivity in V4 neurons: a model of visual crowding

Brad C. Motter

Veterans Affairs Medical Center, Syracuse, NY, USA  
Department of Neuroscience and Physiology,  
SUNY Upstate Medical University, Syracuse, NY, USA



**Visual crowding is a fundamental constraint on our ability to identify peripheral objects in cluttered environments. This study proposes a descriptive model for understanding crowding based on the tuning selectivity for stimuli within the receptive field (RF) and examines potential neural correlates in cortical area V4. For V4 neurons, optimally sized, letter-like stimuli are much smaller than the RF. This permits stimulus conflation, the fusing of separate objects into a single identity, to occur within the RF of single neurons. Flanking interactions between such stimuli were found to be limited to the RF. The response to an optimal stimulus centered in the neuron's RF, is suppressed by the simultaneous presentation of flanking stimuli within the RF. The degree of suppression is a function of the neuron's stimulus tuning properties and the position of the flanker within the RF. A single neuron may show suppression or facilitation depending on the detailed stimulus conditions and the relationship to tuning selectivity. Loss of activity in the set of neurons tuned to a particular stimulus alters its overall representation and potential identification, thus forming a basis for visual crowding effects. The mechanisms that determine the outcome of conflation are associated with object identification, and are not some other independent visual phenomena.**

## Introduction

Visual crowding is the breakdown of our ability to identify peripheral objects in the presence of other nearby objects. A standard psychophysical paradigm used to study crowding examines the identification of a target letter placed between two flanking letters. As the target-to-flanker separation decreases, letter identification is impaired, suggesting a spacing threshold (for reviews, see: Levi, 2008; Strasburger, Rentschler, & Jüttner, 2011; Whitney & Levi, 2011). Psychophysical studies have proposed a variety of different hypotheses about the underlying mechanisms of visual crowding. One hypothesis states that crowding results from an

inappropriate integration of features within a small integration area necessary for recognition of the target. Flanking features that fall within the integration area mix with target features, resulting in deficits in target identification (Pelli et al., 2007; Pelli & Tillman, 2008).

The current study examines the hypothesis that crowding-like phenomena occur at the level of single cortical neurons. Because crowding occurs when target and flankers are presented dichoptically (Flom, Heath, & Takahashi, 1963; Tripathy & Levi, 1994) a cortical origin is expected, although the site is unclear. Imaging evidence for crowding is as early as V1 (Millin, Arman, Chung, & Tjan, 2014) and as late as temporal lobe (Louie, Bressler, & Whitney, 2007). The range of areas may be associated with the level of complexity of the features being integrated. Visual spatial interactions for letter stimuli occur over relatively large target-flanker separations, as much as 0.5 times the target eccentricity (Bouma, 1970). In the present study, area V4 neurons were investigated because the range of letter interaction (Motter & Simoni, 2007) is roughly the same scale as the classic receptive field (RF) size of cortical area V4 neurons (Motter, 2009).

Historically, crowding has implied an interaction between stimuli. Interactions between stimuli presented within the classic RF of V4 and temporal lobe neurons are the subject of many studies (Chelazzi, Miller, Duncan, & Desimone, 2001; Desimone, 1998; Desimone & Duncan, 1995; Gawne & Martin, 2002; Ghose & Maunsell, 2008; Luck, Chelazzi, Hillyard & Desimone, 1997; Miller, Gochin, & Gross, 1993; Missal, Vogels, Li, & Orban, 1999; Moran & Desimone, 1985; Pollen, Przybyszewski, Rubin, & Foote, 2002; Reynolds & Chelazzi, 2004; Reynolds, Chelazzi, & Desimone, 1999; Sato, 1989; Sripathi & Olson, 2010; Sundberg, Mitchell, & Reynolds, 2009; Zoccolan, Cox, & DiCarlo, 2005). These studies attempted to predict the neuron's response to multiple stimuli as a function of the response to each stimulus taken separately. The observed responses were modeled using either a competition or an input gain model to characterize the

Citation: Motter, B. C. (2018). Stimulus conflation and tuning selectivity in V4 neurons: a model of visual crowding. *Journal of Vision*, 18(1):15, 1–24, <https://doi.org/10.1167/18.1.15>.



spatial summation of the responses to the individual stimuli. Attention appears to affect the input gain but not the combinatorial process of integration itself (Ghose & Maunsell, 2008).

In this study, an alternate approach is used. Instead of interactions between stimuli, we propose that crowding emerges from tuning selectivity, taking the position that a neuron does not individuate between RF stimuli but gives a response proportionate to the degree to which the integration of stimulus information in its RF matches its ideal tuning properties (its “optimal” stimulus). The working hypothesis is that the response to an optimal stimulus in the RF is degraded by conflation with other stimuli because the combined stimuli represent a change away from the optimal stimulus, i.e., an inappropriate integration with respect to the neuron’s tuning. Our objective is to examine whether the spatial RF of a neuron and its tuning properties provide a model for interactions within the analogous crowding integration area. Stimulus positioning within the RF is as important to a neuron’s response as the selectivity tuning to other stimulus features. Thus, the strategy chosen was to place a stimulus that best activates the neuron at the center of the RF and study the modulation of that preferred stimulus response as other stimuli are placed at different flanking positions with respect to the central stimulus. If the central stimulus is truly optimal, the integration of additional stimuli should degrade the neuron’s response. Neither the number of additional stimuli in the RF, nor their individual properties, per se, is the issue; only the neuron’s evaluation of the conflated stimulus condition with respect to tuning is significant. Changes in input gain due to attention were minimized by examining peripheral sites while the subject engages in a foveal discrimination task. Given this perspective, changes in the neuronal response resulting from the conflation of stimuli reflect the normal integrative process of the neuron. The term “crowding” is reserved for the description of the perceptual consequences of the juxtaposing of stimuli. The term “conflation” is used here to describe the spatial combining of stimuli into a new identity within the neural RF. Conflation is only a part of the process underlying crowding, and represents a particular stage of integration.

While acknowledging the range of complexity of visual processing in V4, attested by many studies using large stimulus sets, stimulus warping methods, and natural scene images (Carlson, Rasquinha, Zhang, & Connor, 2011; Connor, Brincat, & Pasupathy, 2007; David, Hayden, & Gallant, 2006; DiCarlo, Zoccolan, & Rust, 2012; Gallant, Connor, Rakshit, Lewis, & Van Essen, 1996; Hegde & Van Essen, 2007; Kobatake & Tanaka, 1994; Nandy, Sharpee, Reynolds, & Mitchell, 2013; Pasupathy & Connor, 1999, 2001; Roe et al., 2012; Rust & DiCarlo, 2010; Zoccolan et al., 2005), we chose to concentrate on a subset of letter-like stimuli

(about 20) for which each subject had very extensive discriminative experience. These stimuli reflect the use of letters and numbers in studies of perceptual crowding, reading, and visual search. In addition, stroke-based structures form the basis of many natural contours as well as letters and symbols and have a prominent role in shape coding (Biederman, 1987; Changizi, Zhang, Ye, & Shimojo, 2006).

Central to our strategy is the identification of an optimal stimulus for a neuron responding to letter-like stimuli. We reasoned that if the stimulus set contains a local optimum (a clear preferred stimulus within the set), then the response to these stimuli under conflating conditions will show a pattern consistent with the integration hypothesis, i.e., a suppression of the response relative to that local optimum stimulus. On the other hand, stimulus sets that yield broad tuning assessments have different potential consequences. Broad tuning typically indicates that the set contains features in the neuron’s tuning domain, but the specific configurations within the stimulus set are not optimally organized or lack important features. If the experimentally determined preferred stimulus is only weakly preferred, i.e., the neuron exhibits broad tuning, then two additional outcomes are probable. The first, but less likely, outcome is that the conflating stimuli just happen to form a stimulus much closer to the optimal stimulus for the neuron as evidenced by a significant increase in the response. The second possible outcome is a response to the conflating stimuli that does not differ greatly from the response to the preferred stimulus because the set of features in the conflating condition remain the same without forming a much better (or worse) approximation of the optimal stimulus. All of these outcomes were obtained.

Most perceptual studies of crowding have an attentive task component based on the task’s report requirement; subjects attend and report what they see. Here each neuron acts as an observer providing a graded response to the report of a stimulus’ presence. In addition, in active vision most peripheral locations are actually not the subject of attentive scrutiny. This report characterizes the response of V4 neurons in peripheral vision when attention is at the fovea. Peripheral attentive conditions will be examined in a separate report.

The results are organized in four primary sections that: (a) establish basic stimulus size versus receptive field size relationships for V4 neurons; (b) establish the integration hypothesis using a single flanker derived from an isolated component of an optimal letter; (c) establish that the general category of dual flanking interactions are correlated with feature tuning properties; and (d) examine the relationship between visual space and character spacing in flanker positioning.

## Methods

### Subjects

Data were obtained from seven rhesus monkeys, trained and prepared for behavioral neurophysiological recording experiments. The subjects also participated in visual search and peripheral attention psychophysical experiments not reported here. Standard electrophysiological techniques were used to obtain recordings from neurons in extrastriate area V4 (Motter, 2006, 2009). The spike activity of single cortical neurons was recorded with glass-coated Elgiloy microelectrodes inserted transdurally into the cortex on each day. Eye position was measured with a scleral search coil system. All experimental protocols were approved by the animal studies committees at the VA Medical Center and at SUNY Upstate Medical University. The study adhered to the ARVO Statement for the Use of Animals in Ophthalmic and Vision Research.

### General procedure

Neurons were approached and their electrical activity isolated while animals worked on a foveal attention task. Short blocks of trials were made to estimate the RF location and preferences for size, color, shape, and orientation of stimuli. A substantial effort was made to find combinations of parameters that elicited robust responses defining an optimal, letter-like stimulus; neurons that were not well or consistently driven were dropped from study attempts. Once an estimate of the optimal stimulus was defined, that stimulus was used to map the RF. V4 RFs were mapped by measuring the response rates to stimuli centered at each position of a  $16 \times 16$  grid as described below. From contour plots of these response rates, the center of the RF was re-estimated and the RF was recentered on the display monitor by appropriate positioning of the fixation target. Then blocks of trials were used to characterize the tuning for size, color or luminance, orientation, and/or shape of stimuli. Reiterations of this entire procedure were made when tuning profiles identified a better estimate of the optimal stimulus.

### Foveal attention task and experimental series

All neurons in this report were examined using a foveal attention task that required fixation of a small ( $<0.4^\circ$ ) target and detection of either an orientation or shape change in that target. All subjects received extensive training with strict reaction time require-

ments. Maintained fixation was required within a  $0.75^\circ$  radius of the target center. The training eliminated observable behavioral responses to peripheral stimuli and yielded consistent performance measures. Detection of a change in the fixation target was signaled by a lever release, eye movement, or button push—depending on the requirements of a second behavior task that was either a visual search or a peripheral attention psychophysical task. These secondary tasks are not included in this report but did require extensive discriminative experience ( $>75,000$  training trials) with the letter-like stimuli. Reaction times were typically within a range of 250 to 400 ms. A liquid reinforcement was used with the drop size correlated to trial difficulty and current performance. Viewing was binocular in all conditions. The spatial and feature properties of each peripheral RF were characterized while the monkey performed the foveal attention task. Once the characterization of the spatial and feature properties was completed, one of four experimental series was initiated.

A single flanker experimental series (SFL) was used to examine suppression or facilitation of the response during the formation or destruction of an “optimal” stimulus. This series can be visualized as studying the response of a neuron as a stimulus is built up from (or pulled apart into) component pieces. As noted by Pasupathy & Connor (2001), boundary conformations at specific positions within the stimulus are important to the shape tuning of V4 neurons. Their observation was extended by isolating a limb of a letter stimulus and using it as a flanker (Figure 1D). Changes in response were examined for different spatial positions of the limb relative to a nonoptimal stimulus that, when joined together, formed an optimal stimulus, or conversely as a limb was placed at different positions relative to an optimal stimulus that then formed a nonoptimal stimulus when joined together.

The remaining three experimental series consisted of measuring the response to the combination of the preferred stimulus placed at the center of the RF and two flanking letter-like stimuli placed equidistant from the RF center along an equal eccentricity axis, or an axis perpendicular to the radial axis, or in some cases along the radial axis, as shown in Figure 1. For the second series the flanking stimuli were chosen to be identical (IDN) to the central stimulus; for the third series the flanking stimuli were chosen to be a different stimulus shape (SHP) that also activated the neuron when presented alone. For the fourth series (CLR) the flanking stimuli were the same shape as the central stimulus but differed in color. The two flanking stimuli were always identical to each other. The flanking separation was varied within the series of presentations made on each trial.

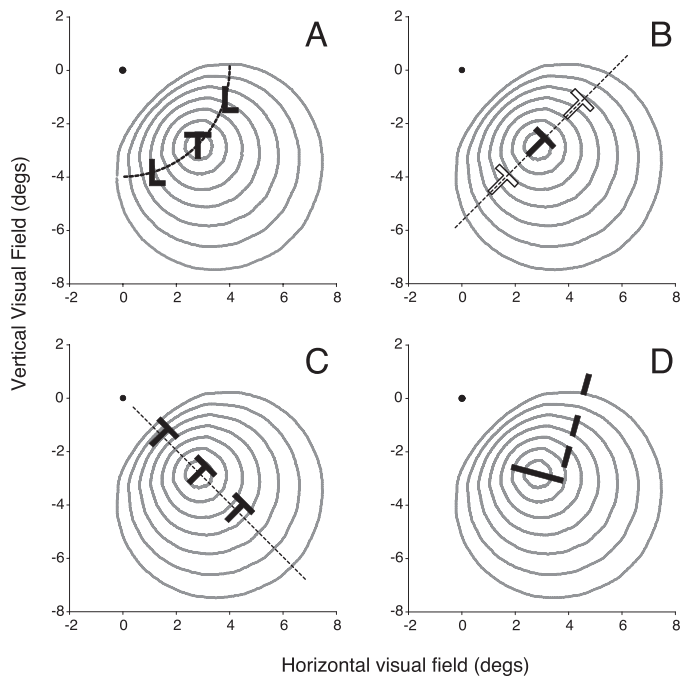


Figure 1. Flanker alignments and positions within receptive field. Nested circular contours depict a model representation of the response sensitivity profile of a V4 receptive field (RF) based on a V4 RF organization that incorporates the cortical magnification factor (Motter, 2009). Flankers in the SHP, IDN and CLR series were aligned (A) along equal eccentricity arcs cutting through the RF, (B) along an axis perpendicular to the radial axis, or (C) along the radial axis cutting through the RF. Flanking separation was defined as center to center distance. Flanking stimuli either (A) differed in shape, (C) were identical to the center stimulus, or (B) were the same shape but different color. The orientation of the individual stimuli was set by each neuron's orientation preference. (D) For the single flanker (SFL) series a small limb flanker was detached from the center stimulus and presented at various separations along an axis perpendicular to the long axis of the center stimulus. These displacements cut across the RF in different directions dependent on the stimulus orientation. Three such separation positions are illustrated in (D) for a stimulus that forms an “L” when the flanker abuts the center bar stimulus. Dot represents fixation target.

## Stimulus presentation

Stimuli were presented on a CRT monitor located at a viewing distance of 57 cm using 22 pixels/°. Stimulus generation and presentation was controlled by custom software using standard graphic routines with stimulus timing synchronized with the vertical refresh of the graphics display system (Motter, 2009). The stimuli consisted of 20 letter-like figures such as T, I, E, L, F, O, or Z and their nonrotationally invariant, mirror images (see Supplementary Figure S1). A preferred orientation was determined using 22.5° increments. The

orientation tuning curve was determined using 11.25°, 22.5°, or 45° increments based on the stimulus's shape symmetry and referenced to the preferred orientation. For each block of about 20 trials, a set of eight stimulus arrangements were used. During each trial up to seven different orientations, shapes, sizes, colors, luminances, or stimulus positions were sequentially delivered to the area of the RF. A stimulus duration of 200 ms and an interstimulus interval of 400 ms were used for presentation timing (Motter, 2006). The number of stimulus presentations delivered per trial as well as the sequence order was pseudorandomized to avoid prediction of both trial duration and stimulus type. This paradigm is used for the characterization studies of all neurons, with the exception of the RF mapping discussed below.

## Receptive field mapping

V4 RFs were mapped using an invisible 16 × 16 grid centered on the estimated RF center. Grid spacing was set at 0.5, 0.75, 1.0, or 1.5 times stimulus length adjusted as necessary to assure that the grid covered the estimated RF size.

The mapping procedure used a moving-spot style stimulus presentation (ON, 100 ms; OFF, 150–200 ms) of the preferred stimulus. The foveal attention paradigm was used. Trials were 500 to 4,500 ms in duration. Response averages for each grid location were obtained by a reverse correlation technique (Motter, 2009). A 2D contour plot of the response rates was constructed. The area bound by the contour line that enclosed response rates above 50% of the maximum response was determined. The radius of a circle with an area equal to this enclosed contour was determined. This measure is similar to the half width at half height standard. The full RF radius was defined as twice that value. This method of estimating the RF radius avoids the measurement uncertainty at the edge of the RF produced by low response rates at the RF edge. This measure of RF size conforms to more standard single flash measures displaced along an axis crossing the RF (see figure 2 of Motter, 2009).

## Spatial comparisons across different RF sizes

Examination of the relation between flanking separation distance and the modulation of the response to the central stimulus requires a consideration of the metric of the separation distance. The size of V4 RFs varies considerably at any one eccentricity and rapidly increases with increasing eccentricity (Figure 2). For each neuron it was necessary to adjust the flanking separation scaling to obtain multiple test

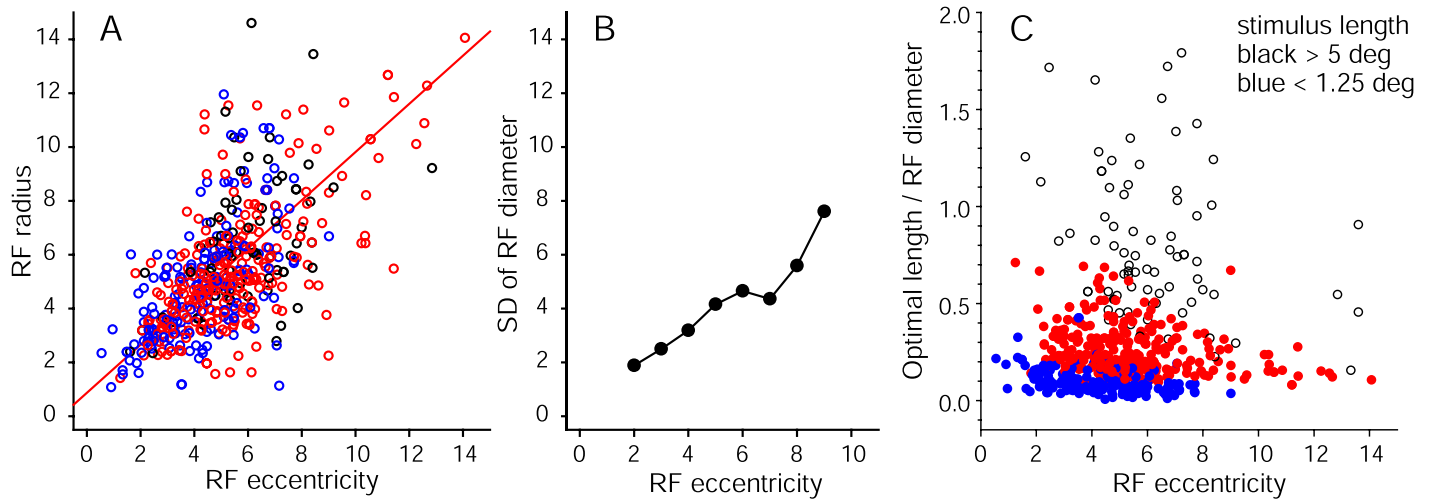


Figure 2. RF size, RF scatter, and stimulus size as function of eccentricity. (A) RF radius plotted as a function of the eccentricity of the RF center. (B) The standard deviation of the scatter of RF diameter increases in proportion to the RF eccentricity. (C) Optimal stimulus length is usually a small fraction of RF size, and when plotted as a function of eccentricity it appears to be independent of eccentricity. For (A) and (C) blue represents stimuli with optimal length less than  $1.25^\circ$ , open black circles represent lengths greater than  $5^\circ$ , and red represents stimuli with intermediate lengths. The blue symbols represent a size range that includes most standard reading text letters.

locations inside the neuron's RF; scaling by eccentricity alone was insufficient. Analysis summaries across neurons that are based on separations measured in degrees of visual angle are not appropriate because a separation of  $X$  degrees will be inside the RF for some neurons and outside the RF for other neurons. Relative position within the RF, rather than absolute separation, was used for comparisons made across neurons. The center-to-center separation between a flanker and the central stimulus (centered in the RF) is expressed as a ratio of the flanker separation to the radius of the RF. When both are expressed in degrees of visual angle, the result is the position of the flanker as a fractional percentage of the RF radius, termed "pRFrad." Some combinations of flanking separation and stimulus size resulted in overlapping or abutting stimuli; these were excluded from quantitative data analysis.

### Spike rate measures and selectivity tuning index

The spike rate activity was measured in defined periods of time around stimulus onset. Individual and population comparisons were based on the neural activity in the interval from 50 to 200 ms after stimulus onset. Stimulus onset time was defined with respect to the midpoint of the first raster scan frame containing the stimulus. For the analysis the response activity of each neuron was normalized to its response to the preferred stimulus in isolation. It is important to establish that the response rates are robust enough

for examining suppression after rate normalization. For the SHP and IDN series the mean response ( $n = 437$ ) to the preferred stimulus was 80 spikes/s with 30 and 136 spikes/s at the 10th and 90th percentiles of response activity. These rates are comparable across all the sets of neurons examined. Population group averaging and comparisons between neurons were based on the normalized data. Population data summaries are depicted using means and standard errors. Nonparametric rank order analyses were used in most comparisons. Multiple linear regressions also were used to identify trends within population activity.

For each neuron assessments of tuning selectivity, positioning, and flanker interactions were made in several different blocks of trials. Significant differences were determined within each block using  $t$ -test comparisons and a distance index method (Sandler, 2008). The latter was chosen because it uses a bootstrap analysis of the probability distribution that is robust for multiple comparisons and because it is sensitive to the temporal structure of the response within the time window of analysis. A significance criterion of 5% was used. These two methods of determining significant response differences were in close agreement within the time window of analysis used. Differences based on the temporal structure of the responses were infrequent and not further reported here.

A tuning index that was independent of the underlying organization of the stimulus dimension was used to define the selectivity of neurons to stimulus shape, orientation, color, luminance, and size. The

index used was introduced by Moody, Wise, di Pellegrino, & Zipser (1998), and defines a selectivity index value  $S$  as:

$$S = \frac{n - \left( \frac{\sum_{i=1,n} R_i}{R_{max}} \right)}{n - 1} \quad (1)$$

Where  $n$  is the number of stimulus-response pairs,  $R_i$  is the response to the  $i$ th stimulus and  $R_{max}$  is the maximum response in the set. This selectivity index is bounded and takes into account changes in activity to each stimulus as opposed to those indices based on maximum and minimum response rates. The index ranges from 0 (identical responses to all stimuli) to 1 (response to only one of the stimuli).

## Results

The activity of neurons with receptive fields (RFs) in the lower contralateral quadrant of the visual field was recorded. Ninety percent (90%) of the RF centers were in the range from 2° to 9° degrees of eccentricity. Postmortem examination of the electrode placements within a grid marked by reference pins indicated that the recordings came from the mid to caudal portion of the crown of the prelunate gyrus, with most locations at or below the level of the tip of the lateral sulcus. This location and the electrophysiological characteristics recorded are consistent with extrastriate area V4 (Nakamura, Gattass, Desimone, & Ungerleider, 1993; Shipp & Zeki, 1985; Stepniewska, Collins, & Kaas, 2005). RF maps and stimulus preference characterizations were obtained in 752 well-isolated neurons; within that group of neurons the SFL (single flanker) studies were completed in 84 neurons, dual flanker studies were completed in 232 neurons in the IDN (identical center and flanker) series, 220 neurons in the SHP (flankers differ from center in shape) series, and 73 neurons in the CLR (flankers differ from center in color) series. Occasionally it was possible to examine neurons in more than one experimental series.

### Receptive field and optimal stimulus size

The description of the results begins with a characterization of the receptive field (RF) size and optimal stimulus size of V4 neurons when using letter-like stimuli. Stressing “letter-like stimuli” may be important because V4 neurons clearly do respond to a wide variety of stimuli, and may change properties based on the stimulus set and task (Cukur, Nishimoto, Huth, & Gallant, 2013). The optimal stimulus was used to map the RF (see Methods). A contoured, response

amplitude profile was constructed for each RF and the area enclosed by the 50% contour was determined. The full RF radius was defined as twice the radius of a circle that matches the area enclosed by the 50% contour boundary. Figure 2A presents the RF radius as a function of the eccentricity of the center of the RF for the subset of neurons with completed studies of stimulus size effects ( $n = 539$ ). The relation between RF eccentricity and RF radius (red regression line) is nearly an identity function. The RF size measures indicate that V4 RFs are much larger than previously reported in anesthetized animals (Desimone & Schein, 1987; Gattass, Sousa, & Gross, 1988). In addition, in the current study the mapping was made with smaller letter-like stimuli, optimally defined for each neuron. There is considerable variation in the RF size at each eccentricity. This variability is depicted in Figure 2B as the standard deviation ( $SD$ ) of RF diameter measured in 1° increment steps of eccentricity. The  $SD$  increases as a function of eccentricity in a manner that nearly compensates for the increase in RF size, so that, at any given eccentricity a 0.2 octave shift in spatial scale to either side of the mean is available from the pool of neurons at that eccentricity. Note that this is the same range as the spatial frequency range over which humans can shift sensitivity when identifying letters in the presence of nearby distracters (Chung & Tjan, 2007). A 0.2 octave shift in behavioral sensitivity therefore does not necessarily require an active process to shift spatial frequency selection, only a utilization of the remaining unaffected resources at a given eccentricity.

In Area V4 the size of the optimal letter-like stimulus is typically a small fraction of the RF size. The optimal stimulus size for a neuron was obtained from experimental runs in which the letter-like stimulus size was doubled, and halved, symmetrically around an estimated preferred size. A ratio index was constructed for comparison across neurons by dividing the stimulus length of the optimal stimulus by the RF diameter. The median value of that ratio is 0.2 ( $n = 539$ ). Thus several optimally sized letter stimuli fit within a V4 RF. Figure 2C shows that the ratio index is essentially independent of the RF eccentricity. The blue symbols in Figure 2A and C identify neurons with an optimal stimulus length of 1.25° or less; 1.25° is about twice the size of normal reading text. In Figure 2C these neurons have the smallest index values, emphasizing that several letters in normal reading text fall within the classical RF of V4 neurons.

### The profile of RFs and Bouma’s observations

The V4 RF profile represents a constant-sized, circular sampling of the V1 surface distribution of the

visual field (Motter, 2009). Because the visual field representation across the V1 surface is based on the cortical magnification factor (CMF), there is a radial expansion of the receptive field for V4 (and potentially all cortical areas sampling from V1). Figure 1 illustrates a model V4 RF centered at 4° in the periphery in the lower right visual quadrant; the scaling for the RF is in degrees of visual angle. The RF model was constructed based on a circular sampling of the V1 surface model (Motter, 2009) centered 4° into the periphery and proceeding in 1 mm concentric steps from 1 to 7 mm along the surface. The nested contours represent those steps and generally equate with the sensitivity contours within the RF; sensitivity decreases away from the RF center. The location of highest sensitivity is defined as the RF center even though its location is displaced toward the fovea as a result of the radial asymmetry in sensitivity. The RF maps depicted in the model (Figure 1) are reasonable matches to actual RF measurements as shown in Figure 5 and in Motter (2009). Figure 1C brings a clear perspective to spatial relationships of stimuli placed at varying distances from the fovea. When a target stimulus is placed at the most sensitive location within the RF, there is an inherent asymmetry in sensitivity to any pair of flanking stimuli placed at equal inward-outward distances from the target stimulus along the radial axis. The inward-outward ratio is roughly 1:2. Three hallmarks of crowding, scaling with eccentricity, a radial inward-outward asymmetry, and a radial-tangential anisotropy of the crowding zone (the area within which crowding is observed) are consistent with the 2D RF as mapped in V4 neurons. The scaling hallmark is matched by the increase of the RF size with eccentricity (Figure 2A). The inward-outward asymmetry hallmark derives from the observation (Figure 1C) that for equal target-flanker separations an outward flanker is in a more sensitive part of the RF (and thus more effective) than an inward flanker. The third hallmark of crowding, a radial-tangential anisotropy of the crowding zone, is not a RF property, per se, but can be explained by the inward-outward asymmetry and the common method of plotting the crowding zone. Given that identifying the central target in a dual flanker paradigm requires information from a receptive field centered on the target, then for equally spaced radial flankers the outer flanker will engage the RF at greater distances than the inner flanker (Figure 1C). In fact the outer radial flanker engages the RF at a greater distance than a flanker in any other direction. In a dual flanker task, the radial extent of crowding is therefore determined by the outer flanker, which leads to plotting a radially elongated crowding zone because the inner and outer limits are both plotted at the same distance from the target. These elongated zones have been reported by Toet and Levi (1992) and Pelli et al. (2007) among many others. When a single flanker is used, the

resulting crowding zone should appear more like the RF contours in Figures 1 and 5, and as depicted by Bouma (1978).

## The integration hypothesis tested with single flankers altering letter-like stimuli

Shape coding within V4 neurons emphasizes the representation of curvature or angled line segments (Connor, 2004) as has been shown in studies that extend the surface of a proto-object (Carlson, et al., 2011; Pasupathy & Connor, 1999). The response to multiple stimuli within the RF, in the absence of attention, is generally considered to result from a weighted summation of the response to the stimuli taken individually (Desimone & Duncan, 1995; Ghose & Maunsell, 2008; Reynolds, Chelazzi, & Desimone, 1999). Our integration hypothesis, however, states that the response to multiple stimuli involves a single evaluation of features against the tuning properties of the cell, rather than combining separate estimates. Boundary conformations at specific positions within the stimulus are important to V4 neurons (Pasupathy & Connor, 2001). Expanding on that observation, the single flanker (SFL) experiments examine the response of a neuron as an optimal stimulus is either formed or destroyed when a component piece is added, as illustrated in Figures 1D and 3. The objective was to determine whether the separated stimulus segments portray the facilitative and suppressive conflation actions anticipated by the integration hypothesis.

The experiments examined neurons where a single limb of a letter-like stimulus was the difference between an optimal stimulus and a much less responsive counterpart. Eighty-four (84) neurons from six subjects were examined in this series. The stimulus limb (flanker) was displaced along an axis perpendicular to the long axis of the optimal stimulus. Most of the neurons in this set gave moderate responses to the nonpreferred centered stimulus and low rate responses to the single limb. However, particularly clear examples of optimal stimulus formation can be made when an optimal stimulus is formed from two segments that separately evoke only weak or no response from the neuron. Figure 3A illustrates such an example; another example is provided in the Supplementary Figure S2. Figure 3A and B illustrates a condition where the flanker in each case is ineffective in evoking a response by itself. For the stimulus conditions depicted in Figure 3A, the response spike rasters for the conflating L, beginning at the top, show the response to the long bar presented alone (at most one or two spikes are elicited), followed below by the response rasters to the optimal “L” shape, then followed by the response to the simultaneous presentation of the long bar with the

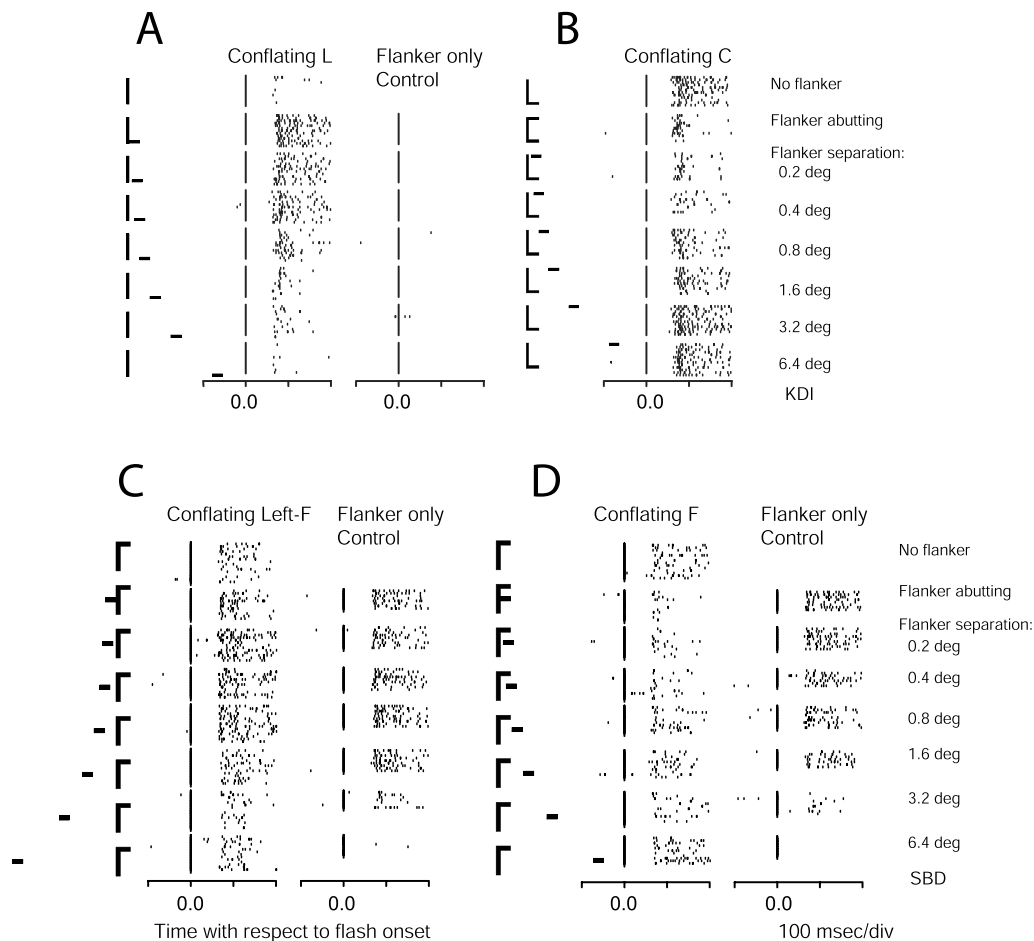


Figure 3. Altering of letter-like stimuli. (A) Conflation of a small flanker with a vertical bar resulting in an optimal “L” shape. Flanker positions depicted on the left and labels on far right, starting at top with only the vertical bar presented. Left rasters show the response to the bar followed beneath by responses to the “L” and then to the various flanker separation combinations. The flanker-only control raster, essentially empty, shows the lack of response to the single flanker presented alone. (B) Conflation that transforms an optimal “L” into a less than optimal “C” for the same neuron. Rasters show response to optimal “L” at top followed beneath by response to “C” and then to the various flanker separation combinations. (C) and (D) Similar to (A) and (B) above, but for a different neuron where the small flanker gave a moderate response (flanker-only control) across the RF from left (C) to right (D) sides of RF. However, the effect of the flanker was to produce facilitation in (C) and suppression in (D). Stimulus size and separations are accurately scaled. Conflation results in either facilitation (A) and (C) or suppression (B) and (D) within the same neuron depending on details of the stimulus position and tuning selectivity of the neuron. The RF radius was  $2.2^\circ$  for (A)/(B) and  $2.6^\circ$  for (C)/(D). Tics in spike rasters represent the time of action potentials synchronized to stimulus onset as marked by the vertical bar at time zero. Different rows in each raster indicate responses on different trials.

lower right limb of the “L” at different positions within the RF. The separation distances (in degrees of visual angle) between stimuli are given by the numbers at the far right. The control rasters in the middle show the lack of response to the stimulus limb presented at the same locations in the absence of the central long bar. This neuron was sharply tuned to both orientation and shape. Note that these were flashed presentations with every combination randomly sequenced. Even when the parts of an optimal stimulus are somewhat separated within the RF, the neuron responds to their presence in a manner consistent with the optimal stimulus. In counterpoint a typical suppressive conflation outcome

is shown in Figure 3B, for the very same neuron. When a second stimulus limb is positioned so that it alters the optimal “L” into a “C,” the neuron’s response is suppressed over the same spatial extent.

Figure 3C and D presents a second example of facilitatory and suppressive conflation in a single neuron; in this case the flanker produces a moderate response when presented alone. Here a difference in the conflation result depends on the position of the flanker on either the left (Figure 3C) or right (Figure 3D), resulting in either facilitation (Figure 3C) or suppression (Figure 3D) relative to the nonoptimal horizontally mirrored “L.” The flanker-only control shows that



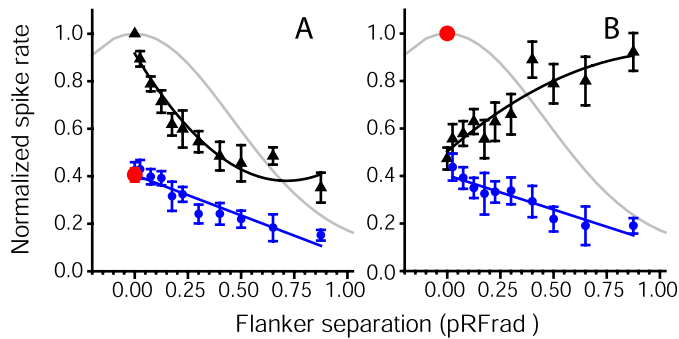


Figure 4. Conflation in the formation or destruction of optimal stimulus. (A) Facilitative conflation. Population average response ( $n = 56$ ) for facilitative conflation of two nonpreferred stimuli resulting in formation of an optimal stimulus, as shown in Figure 3A. Red dot is response to centered nonpreferred stimulus; blue dots are responses to flanker stimulus at different positions within RF. Black triangles are the responses when both are presented simultaneously. As the flanker separation narrows, the response increases to match that of the optimal stimulus. (B) Suppressive conflation. Population average response ( $n = 28$ ). Response to optimal stimulus at RF center (red dot) is reduced when paired with a flanker at sequential positions in the RF (black triangles). The gray curved line shows the RF profile as mapped with the optimal stimulus. The blue and black lines represent simple linear and quadratic fits to the respective data points, provided to emphasize the gradual change across the RF.

while the flanker elicits a similar response on both sides, the conflation result is opposite. The suppression in (Figure 3D) shows a reduction in response consistent with the shape tuning of the neuron. A weighted spatial summation model is hard pressed to explain these two opposed outcomes, whereas the shape sensitivity of the integration hypothesis captures both stimulus conflation scenarios.

Both the transient and maintained components of the V4 neural response are suppressed by flankers. The effects of suppression or facilitation appear in some cases (Figures 3 and 5) to have a delayed onset correlated with different flanker separations. The different time courses represent the different sensitivities to the shapes, at different stages of conflation. Similar variations can be seen in the responses to different letter-like shapes (see Supplementary Figure S3). Two additional factors contribute to this beyond simple flanker separation: (a) the sensitivity of the RF increases from edge to center; thus, the effectiveness of the flanker in terms of its weighting in the neural integrative process should increase, and (b) the response onset latency decreases from edge to center of the RF on average ( $n = 221$  neurons) about 12 ms. This difference may be sufficient to change the effectiveness of the simultaneous presentation. In addition, Figure 3B and D are examples of response suppression that are

not due to flanker inhibition, per se, but a suppression that results from a change to a less preferred stimulus configuration. Suppression without inhibition interactions appear to be related to tuning properties rather than actual inhibitory interactions between stimuli. This is an important distinction for stimulus interaction modeling.

Conditions where the conflation of two nonpreferred stimuli produced an optimal stimulus were examined in 56 of the 84 neurons. The activity of each neuron was normalized with respect to the response to its optimal stimulus. The separation distances were normalized to each neuron's RF radius. The averaged results are shown in Figure 4A. The response to the RF centered, nonpreferred stimulus when presented by itself is shown as a large red dot. The lower blue circles represent the response to the single flanking segment when presented by itself at different RF locations. The upper black triangles represent the response to the simultaneous presentation of both stimuli. As the gap between the stimuli narrowed, particularly within the inner half of the RF, the neurons' response (triangles) increased, culminating in a response equivalent to the optimal stimulus as the separate stimuli conflated. The light gray curve in Figure 4 represents a Gaussian RF profile fit to the responses to single optimal stimulus presentations presented along an equal eccentricity arc passing through the center of the RF ( $n = 80$  neurons), and serves here as a reference and reminder of the general sensitivity profile of the RF as mapped by an optimal stimulus.

Figure 4B shows the condition where the conflation of the stimuli resulted in suppression. These comparisons (28 different neurons) were made by the addition of a stimulus limb to an optimal stimulus as in Figure 3B. Here the addition of the flanker-limb stimulus produces suppression (black triangles) of the response to the optimal stimulus presented alone (the red dot) at positions within the RF, again particularly within the inner half of the RF. The greatest suppression occurs when the conflation is complete, resulting in a single nonpreferred stimulus at the center of the RF. Note that the stimulus limbs evoke similar, weak responses in the two conditions when presented alone, as seen in comparison of blue circles in left and right sides of Figure 4. Yet when presented as flanker stimuli, these weak stimuli can produce strong facilitatory or suppressive results predicted by the neuron's shape sensitivity. The neuron's response to the flanker when presented alone, even singly at the center of the RF, is not a consistent predictive factor of outcome. A flanker becomes a factor only in the evaluation of the conflating stimulus combination. Thus, response prediction is based on the tuning to the preferred stimulus in the center of the receptive field. Centering the stimulus in the RF is not arbitrary; it is part of the

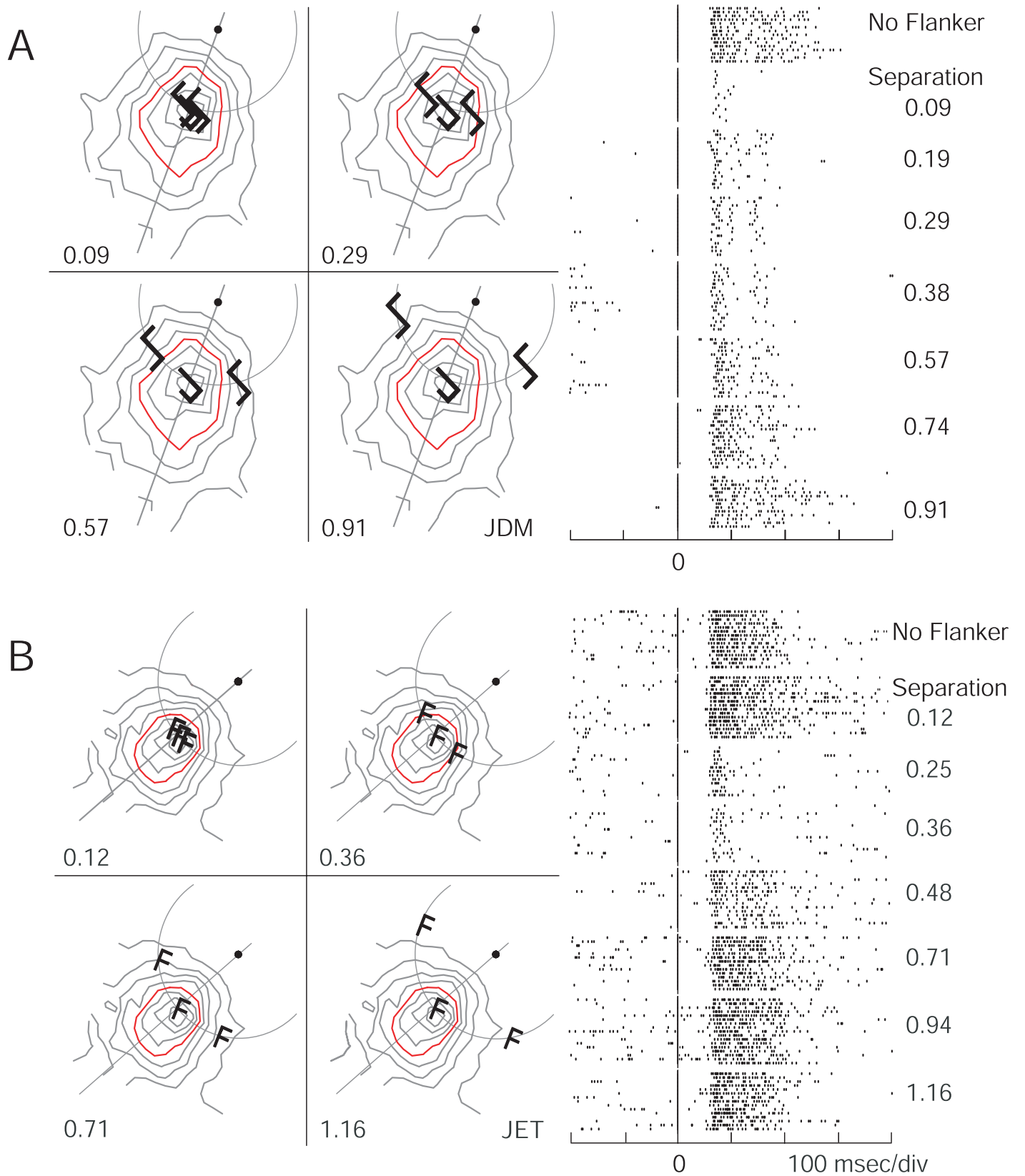


Figure 5. Response to dual flanking stimuli in the SHP and IDN series. (A) SHP series. Left: Stimulus positions superimposed on one V4 neuron for four of the eight stimulus combinations used in its study. Right: Spike rasters for all eight combinations. Topmost raster shows response in the no-flanker condition; remaining groups from top down are in order of increased flanker separation.

←

Suppression occurs as flankers enter the RF, and progressively increases as flanker separation decreases. (B) A different neuron studied in the IDN series. In this case the suppression reverses at the nearest, overlapping separation. Numbers at far right give separations in fractional percent of RF radius and match values in the RF configurations. Black dot in upper right of each contour panel represents the fixation target. Arc lines mark equal eccentricity axes. For comparison purposes red line marks same contour in each panel. Action potentials indicated as in Figure 3.

definition of optimal conditions. Although shape sensitivity is more-or-less invariant across the RF, placement of SFL stimulus combinations off-center within the RF disrupted the profiles show in Figure 4. Beyond overall lower activity levels, off-center placements generated local peaks and troughs in activity. These local perturbations were particularly evident, for example, when the axis of the small flanker displacements crossed the hot spot of the RF center to join an off-center component.

By examining the differences in response between two known stimulus configurations, the single flanker (SFL) series establishes that suppression and facilitation effects of flankers have a consistent spatial correspondence across the receptive field and can be interpreted as expressions of the same mechanism. The response is governed by the neuron's shape-tuning properties, rather than a consistently weighted response to the individual components. Spatial separation appears to be an enabling but secondary issue that sets a spatial limit to the interaction. It is not the physical separation of objects, *per se*, but rather what they form when juxtaposed in the RF as assessed by the shape tuning of the neuron. Overall, the interactions between spatially separate stimuli within the RF appear to be governed by the same processes that are associated with shape sensitivity, suggesting that conflation, and thereby crowding, are manifestations of the neural mechanisms underlying shape processing. The SFL series supports this reasoning and establishes the importance of neuronal shape selectivity and its governance of the spatial interactions in the conflation process. Further support of the interdependence of conflation and shape processing will be addressed below by examining whether a constellation of stimulus selectivity features—shape, orientation, size and color—can predict the outcome of conflating stimulus combinations.

### Response to stimulus conflation in dual flanker conditions

The simplicity of the SFL series and its relatively straightforward relation to shape sensitivity provides a basis for examining more complex interactions arising in a standard dual flanking crowding task where flankers are other letters. The conflation of three letter-like stimuli produces rather complex combinations. Yet

under our working hypothesis, if there is a clear local optimum stimulus, then flanking stimuli will result in suppression. To examine dual flanker conditions, two paradigms were used. In the SHP (shape) series, flankers were different in shape from the central stimulus, *i.e.*, the standard crowding paradigm. In the IDN (identical) series the flankers were identical to the central stimulus. The IDN series controlled for the effectiveness of the stimuli, in that all stimuli were identical. Unlike visual crowding studies where flankers are typically placed along a horizontal or vertical meridian, flankers were positioned with respect to the RF location. The optimal stimulus was centered in the RF (the peak sensitivity location). The flankers were symmetrically positioned along an equal eccentricity arc, or along a line perpendicular to the radial axis passing through the RF center (see Figure 1A and B). In a minority of cases, flankers were symmetrically spaced along the radial axis. The stimulus orientation, color, and luminance of the flankers and central stimuli were the same. The orientation of stimuli was based on the neuron's selectivity, not on the flanker axis passing through the three stimuli. Because of these considerations, stimuli tended to nest together as separation decreased with the IDN series, often resembling a shape variant of the stimulus, whereas the SHP series formed unusual stimulus complexes as their separation decreased.

Figure 5 provides two striking examples of the interaction of dual flankers with a RF-centered, optimal stimulus for two V4 neurons in the SHP and IDN experimental paradigms. Flanker separation conditions are shown that resulted in almost complete suppression of the response to the stimuli. For each neuron, four of the eight stimulus configurations are depicted on the left, accurately scaled and superimposed on the concentric RF sensitivity contours. The response contours are mapped using the central (preferred) stimulus. The neural response spike rasters for all eight conditions are shown on the right. The topmost raster represents the response to the central stimulus presented without any flankers, followed below by the responses to increasing steps of flanker separation. Center to center separations as fractional percentages of the RF radius ( $pRF_{rad}$ ) are shown on the far right. The figures illustrate several commonly observed features: (a) optimal letter-like stimuli were much smaller than their respective V4 RFs, (b) suppression of the response to the central stimulus

occurs when the flankers take positions within the RF, and (c) the suppression becomes strongest when flankers are within the central half of the RF. The red contour line simply marks the same contour within each RF for comparison purposes. There are also striking differences between the illustrated neurons. In Figure 5A the suppression persists as the flanking stimuli touch and merge with the central stimulus. In contrast, in Figure 5B the suppression gives way to facilitation when the central and flanking stimuli merge. In IDN cases facilitation occurred, if at all, when the central and flanking stimuli reached their closest separations and then began to overlap, as in Figure 5B.

For behavioral measures, “critical spacing” defines a spatial threshold for center-to-center separation beyond which crowding flankers have no influence on object identification and below which flankers interfere with object identification (Bouma, 1970; Intriligator, & Cavanagh, 2001; Pelli, Palomares, & Majaj, 2004). An initial hypothesis we entertained was that flanking interactions would be organized based on a center-surround mechanism with a suppressive surround extending beyond the excitable central RF in a fashion similar to the classic Mexican Hat RF organization. However, graphic analyses like those in Figure 5 demonstrated that this was not the case. Letter-like flanking stimuli outside the classic RF had minimal, if any, impact on the response to the central stimulus. Prominent stimulus interactions occurred only within the central areas of the RF. Surround suppression effects (Desimone & Schein, 1987) spanning across the RF boundary were found with large sets of multiple flanking stimuli. Center-surround interactions of V4 neurons may be important for figure-background segmentation (Roe et al., 2012).

## Testing the integration hypothesis

Not all studies resulted in dramatic suppression as in Figure 5. Is the integration hypothesis actually predictive of the dual flanker results? The integration hypothesis is formulated with respect to the optimal stimulus. The better the experimentally determined preferred stimulus is at approximating the optimal stimulus, the more likely the observation of a significant suppression will be. Our experimental estimate of the approximation of the preferred stimulus to the optimal stimulus is the sharpness of the tuning profile. Therefore, the next goal is a quantitative test of the integration hypothesis by examining the correlation between stimulus tuning and flanker suppression. The integration hypothesis suggests that response suppression should deepen as neuronal tuning selectivity increases. Our approach concentrates on the selectivity

of the neuron to the experimentally defined preferred stimulus at the RF center, rather than the flankers because the integration hypothesis is based on the conflated stimulus’s match to the optimal stimulus, not on the effectiveness of the flankers. This result was supported by the single flanker (SFL) series analysis.

For a second goal, determination of the population response as a function of flanker separation, a method is needed to avoid the cross cancellation of suppressive and facilitatory responses that occur by averaging. Fortunately the relation between tuning selectivity and response provides a method of sorting the data to minimize the cross cancellation issues. The following sections examine the tuning sensitivity of the neurons, the relationships between tuning and suppression or facilitation, and finally the relationship between the neuronal response and flanker separation.

## Tuning selectivity and receptive field size

Data were collected for shape, orientation, color-luminance, and stimulus size tuning selectivity. For each selectivity index, only the indexed feature (shape in Figure 6A) was changed; the stimuli otherwise maintained the feature preferences of the preferred/optimal stimulus. “Color-luminance” tuning was based on seven photometrically matched color stimuli (RGBCMY & white) and one black stimulus. Tuning selectivity for each of these properties was estimated from the neuron’s responses to stimuli presented one-at-a-time at the center of the RF, and pseudorandomly sequenced within and across trials. A selectivity tuning index was used (see Methods) that is independent of the spacing of the exemplars along any dimension. Figure 6 shows the tuning indexes for a large set of V4 neurons, including neurons for which conflation testing was not completed. Each property was not available for every neuron. An index value near 0.0 indicates a lack of tuning (not a lack of response) where the responses to all stimuli were the same; an index value near 1.0 indicates very sharp tuning (the response to one exemplar completely dominates all the others). The bar histograms for shape, orientation, stimulus size, and color-luminance tuning are similar and show that a large proportion of the recorded neurons were broadly tuned to the exemplars chosen in each of these data sets.

The bias for broadly tuned neurons was expected. A neuron needs to be actively firing to be discovered during recording. To increase the chance of encountering a responsive neuron, multiple stimulus configurations were used during initial search and isolation of a neuron. This choice biases detection and selection toward neurons responding to multiple stimuli within the search set. Consequently, the goal of finding

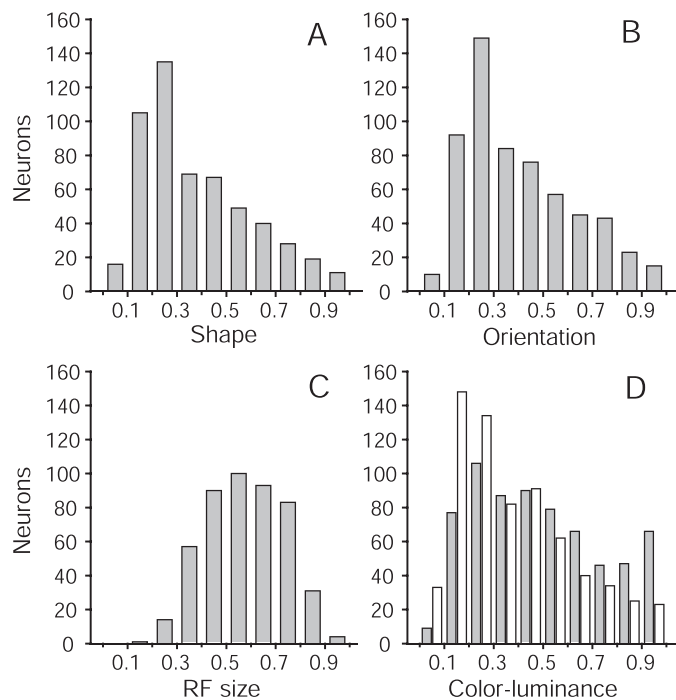


Figure 6. V4 RF stimulus selectivity. Histograms of neuronal selectivity for basic stimulus properties based on the selectivity index. Indexes were computed for each property available for each neuron. The index is based on the response to eight different stimuli, and ranges from no selectivity (index = 0.0) to complete selectivity (index = 1.0). Bar histograms indicate number of neurons in 0.1 index value steps for shape (A), orientation (B), size (C), and color-luminance (D) selectivity. In D the light bars are for the six equiluminant colors (RGBCMY) computed as a separate group, and the dark bars are the index values when white and black stimuli are included.

neurons responsive to letter-like stimuli was met, but at the expense of a bias for finding more broadly tuned neurons.

The tuning results were in general agreement with previous studies, with the exception of stimulus size tuning. Desimone and Schein (1987) and El-Shamayleh and Pasupathy (2016) both reported that increases in stimulus size resulted in increases in neural response up to the edge of the RF for most V4 neurons. Our results indicate that peak size sensitivity occurs for stimuli that are much smaller than the RF diameter. It may be that the initial search for neural activity using smaller letter-like stimuli, with higher spatial complexity (Attneave & Arnoult, 1956, Pelli et al. 2006) engages a different population of V4 neurons, or the differences may result from our larger RF sizes (as noted above) mapped with small letter-like stimuli, or the extensive discrimination training the animals had on letter-like stimuli may play a role in the differences. In any case, given the importance of stimulus size tuning to conflation as

	Orientation	Shape	Size	Color-luminance
RF radius	$r = -0.257$ $p < 0.001$ $n = 542$	$r = -0.258$ $p < 0.001$ $n = 501$	$r = -0.285$ $p < 0.001$ $n = 472$	$r = -0.118$ $p < 0.004$ $n = 613$
Orientation		$r = 0.724$ $p < 0.001$ $n = 429$	$r = 0.580$ $p < 0.001$ $n = 363$	$r = 0.020$ $p = 0.65$ $n = 553$
Shape			$r = 0.650$ $p < 0.001$ $n = 369$	$r = 0.080$ $p = 0.086$ $n = 465$
Size				$r = 0.300$ $p < 0.001$ $n = 417$

Table 1. Correlations among the tuning selectivity indexes and the RF radius in V4 neurons.

established below, its role in shape processing in area V4 needs further attention.

Correlations between stimulus selectivity indexes and between those indexes and the receptive field radius were examined (see Table 1). There are three main observations: First, the indices for shape, orientation, and stimulus size tuning show relatively high and significant correlations, suggesting commonality with an underlying structural dimension for letter-like stimuli. Second, color-luminance selectivity is at best weakly correlated with the others, suggesting an underlying representational dimension that is independent from the others. This difference in form versus color-luminance properties is consistent with the “separate information streams” hypothesis that remains the current model of sensory processing in early visual cortical areas (Roe et al., 2012). Third, as expected, the RF radius is strongly correlated with RF eccentricity ( $r = 0.81$ ,  $n = 691$ ,  $p < 0.001$ ). Interestingly, the correlations between RF radius and stimulus selectivity indexes are all negative and small but significant.

## Relationship between tuning selectivity and flanker modulation

The hypothesis to be tested is whether tuning selectivity can predict the flanker suppression of the response to the preferred stimulus. The outcome measure used for testing the hypothesis was the maximum suppression observed for flanker conditions. Note, only nonabutting and nonoverlapping flanker conditions were used in the following analyses. Flanker suppression is expected when the tuning sensitivity is sharp (high index value), indicating the preferred stimulus is relatively optimal in the set. Likewise, strong tuning makes it less likely that conflation leads

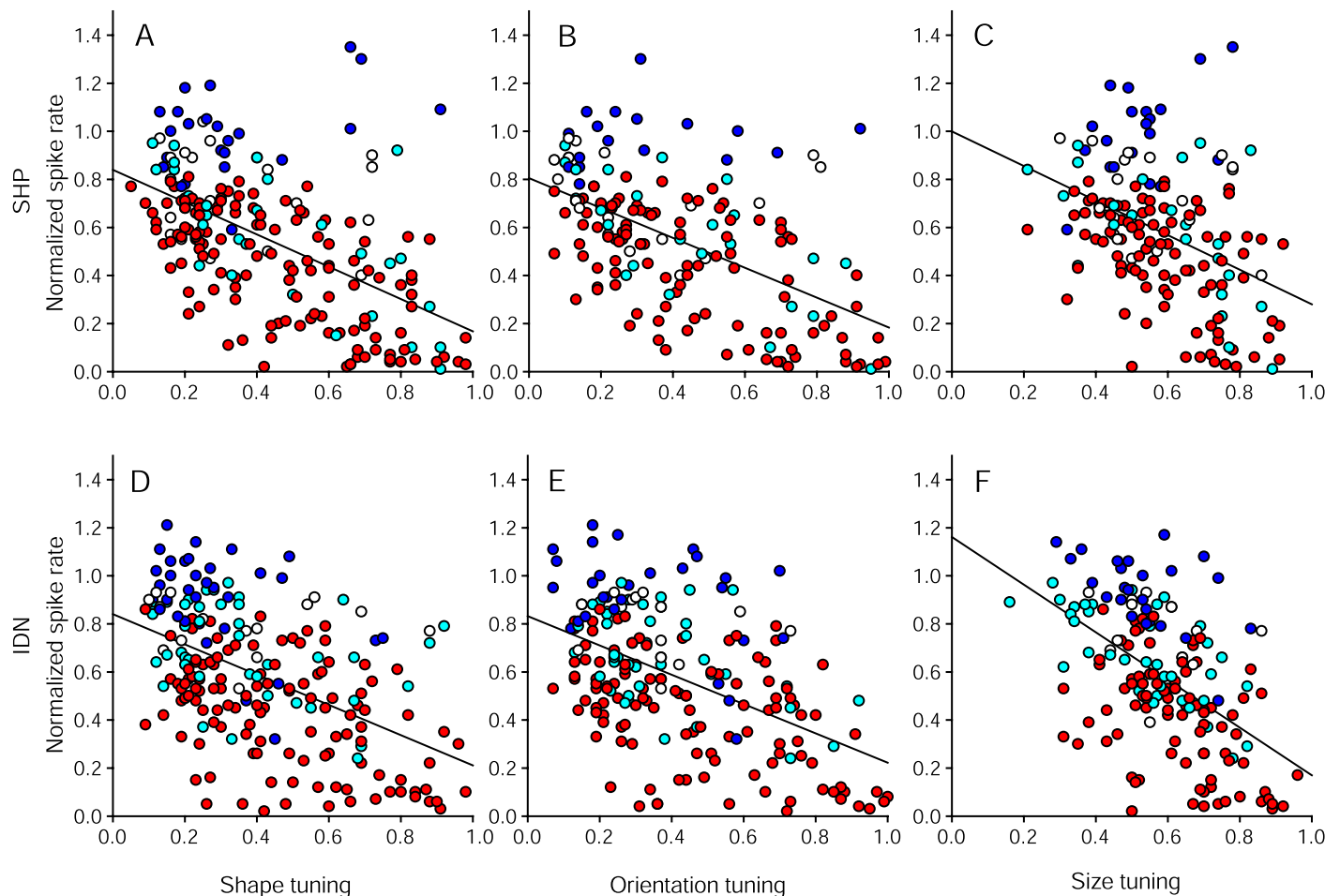


Figure 7. Flanker modulation as a function of tuning selectivity. Each dot represents the peak flanking suppression for each neuron as a function of the tuning selectivity for shape (A) and (D), orientation (B) and (E) and size (C) and (F). Neurons in the SHP experimental series are shown in the top panels and those for the IDN series in the bottom panels. The slanted lines are linear regression fits to the data points. As stimulus-tuning selectivity for each property increased, there was a corresponding increase in the amount of observed flanking suppression. Filled dots represent neurons with significant ( $p < 0.05$ ) modulation relative to the preferred stimulus response. Red dots represent neurons with suppression-only, blue dots with facilitation-only, cyan dots with mixed suppression and facilitation. Open circles represent neurons without any significant modulation.

to facilitation by chance. In contrast, facilitation occurs when the conflating stimulus is closer to the true optimal stimulus. Consequently, if facilitation occurs, it is more likely to occur when tuning is broad, i.e., when the preferred stimulus is only weakly preferred.

The scatterplots in Figure 7 illustrate the maximum response suppression rate during the dual flanking presentations as a function of the tuning selectivity for shape, orientation, and size. The data are shown for both the SHP (shape) and IDN (identical) flanking series. Within each scatterplot, the slanted line represents the linear regression computed for that data set. Qualitatively, the scatterplots and the slopes of the regression lines support the hypothesis. Sharper tuning sensitivities (increasing index values) result in increased likelihood of strong suppression from flankers (decreasing response rates). This result is evident across

the three tuning properties and is essentially the same for both the SHP and IDN series. The color coding in Figure 7 represents a classification of flanker effects made by dividing the neurons into four categories of response type. The color code classifies neurons as showing a suppression-only (red), a facilitation-only (blue), both suppression and facilitation (cyan) and no modulated response (open) with respect to the optimal stimulus presented alone. The classification criterion required significant suppression (or facilitation) in two adjacent spatial flanking positions (using the bootstrapped DI index,  $p < 0.05$ ; see Methods). The suppression-only and mixed suppression groups are consistent with the hypothesis, showing increased suppression by flankers as their respective tuning sensitivities increase. The disproportionate number of neurons at low levels of tuning selectivity matches the

general bias for broadly tuned neurons seen in Figure 6. However, many of these broadly tuned neurons nevertheless show significant levels of suppression. The set of neurons (blue) that had facilitatory-only flanker interactions is predominately clustered in the upper left of each scatterplot instead of spreading evenly across the tuning selectivity range. This result is consistent with the hypothesis that the formation of a new stimulus that better approximates the neuron's optimal stimulus is more likely to occur when tuning is broad and the preferred stimulus is only weakly preferred; the left side of each graph.

A step-wise linear regression on the tuning variables' ability to predict the neurons' maximum suppression to flanker stimuli was made using a criteria of  $p = 0.05$ . The results of the step-wise regression indicated that a linear combination of the tuning to orientation and size could predict the maximum suppression, and that shape and color-luminance tuning variables did not significantly add to the ability of the equation to predict maximum suppression for both the SHP and IDN series. A multiple linear regression analysis for the SHP data using orientation and size tuning to predict the neurons' maximum suppression to flanker stimuli indicated that the two predictors explained 30% of the variance:  $R^2 = 0.31$ ,  $F(2, 120) = 27.40$ ,  $p < 0.001$ . It was found that orientation tuning predicted maximum suppression ( $\beta = -0.37$ ,  $p < 0.001$ ) as did size tuning ( $\beta = -0.26$ ,  $p = 0.005$ ). The variance inflation factor test for multicollinearity was  $< 1.5$  for both variables indicating that multicollinearity was not a factor in the fit (Glantz & Slinker, 1990). The regression equation for a least squares fit is SHP response =  $0.994 - (0.426 \times \text{orientation tuning index}) - (0.447 \times \text{size tuning index})$ .

Likewise, a multiple linear regression analysis over the same variables for the IDN series indicated that orientation and size tuning predicted 36% of the variance:  $R^2 = 0.37$ ,  $F(2, 142) = 41.77$ ,  $p < 0.001$ . For the IDN series orientation tuning predicted maximum suppression ( $\beta = -0.35$ ,  $p < 0.001$ ) as did size tuning ( $\beta = -0.33$ ,  $p < 0.001$ ). The variance inflation factors for the IDN variables were  $< 1.5$ . The least squares fitted regression equation was: IDN response =  $1.095 - (0.415 \times \text{orientation tuning index}) - (0.602 \times \text{size tuning index})$ . The variation in the number of neurons results from exclusion of neurons when sufficient tuning data were not available. For the SHP and IDN series the results held for any two of the six contributing animals randomly selected and analyzed separately; the results are not attributable to differences in animals.

Substitution of the shape tuning index for orientation tuning changed the multiple linear regressions only slightly, decreasing the explained variance by an average 5%. The relatively high correlation between shape and orientation tuning suggested a possible collinearity. This possibility was tested by regression of

both factors against maximum suppression. The variance inflation factor for the two variables was  $< 2.4$  in both the SHP and IDN series, indicating that collinearity was actually not an issue. However, combining the shape and orientation indexes by averaging the two did not markedly affect the explained variance. Thus, it appears that the shape and orientation tuning indexes capture independent but nearly equal measures of stimulus properties of the letter-like stimuli. Circular preferred stimuli did have shape but not orientation tuning indexes, but the inclusion or exclusion of these cases did not have an impact on the overall regression on shape tuning selectivity.

While the interchangeability of shape and orientation indexes may be related to their measure of common stimulus form features, the size tuning index quantifies a different stimulus property. It is reasonable that neurons tightly tuned to a preferred size will be less responsive to a conflating stimulus that is larger because of flankers. However, there was no significant correlation between the size tuning index and the optimal stimulus length expressed as a fraction of RF diameter (Pearson correlation,  $r = 0.03$ ,  $p = 0.54$ ,  $n = 473$ ) in the overall population.

### Flankers differing in color or luminance

Color or luminance differences between targets and flankers have been shown to lessen the effects of visual crowding (Kooi, Toet, Tripathy, & Levi, 1994). According to the integration hypothesis, flankers having a different color or luminance will have integration effects if the neurons are tuned to color or luminance. This was tested in the CLR (color) series of experiments ( $n = 73$  neurons), where all stimuli were identical in shape, but the center stimulus had the preferred color-luminance for the neuron and the flanking stimuli had a less preferred color or luminance. The results are shown in Figure 8A. Suppression by flankers increased as a function of the tuning selectivity for color-luminance. The results of a step-wise regression indicated that a linear combination of the tuning to shape and color-luminance tuning predicted the maximum suppression, and that orientation and size tuning variables did not significantly add to the ability of the equation to predict maximum suppression. For the CLR series a multiple linear regression of shape and color-luminance explained 45% of the variance:  $R^2 = 0.46$ ,  $F(2, 66) = 28.61$ ,  $p < 0.001$ . It was found that shape tuning predicted maximum suppression ( $\beta = -0.47$ ,  $p < 0.001$ ) as did color-luminance tuning ( $\beta = -0.41$ ,  $p < 0.001$ ). Variance inflation factors were all less than 2.0.

The predictive power of color-luminance tuning in the CLR series is consistent with the hypothesis that the

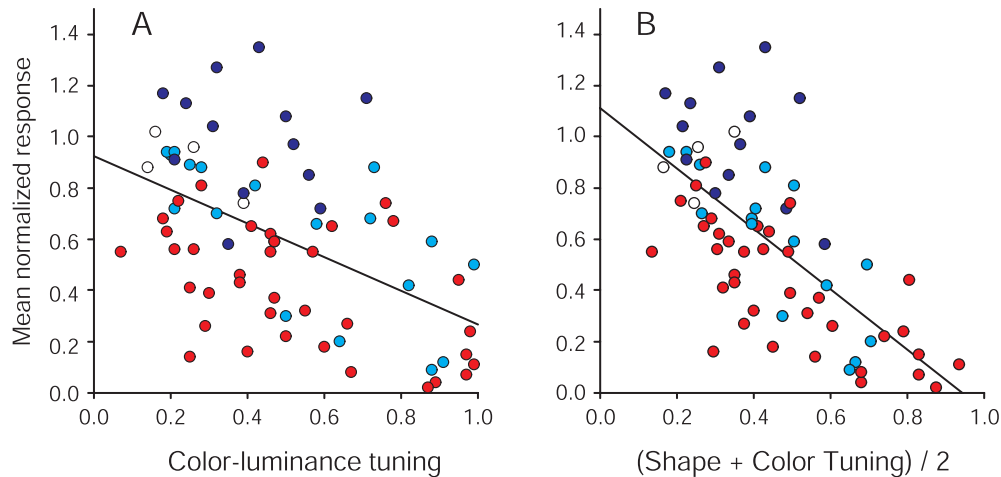


Figure 8. (A) Flanker modulation as a function of color-luminance tuning selectivity. In the CLR (color-luminance) series there is a moderate relationship between the maximum suppression and color-luminance tuning selectivity. (B) Modulation as function of average combined selectivity indexes. Strength of prediction of maximum suppression increases when the simple average of the index values for shape and color-luminance selectivity tuning are used (CLR series). Symbol colors as indicated in Figure 7.

interactions between the central stimulus and flankers are defined by the integration with respect to the optimal stimulus. When the neuron is tuned for color or luminance, then flankers differing in color or luminance result in a suppression of the response to the conflating stimulus. The CLR series differed from the IDN series only in the addition of a color or luminance difference. Interestingly, the regression analysis found that color-luminance tuning essentially replaced the size tuning variable (from the SHP and IDN results) and increased the amount of explained variance. The increase in predictive power likely reflects the fact that the color-luminance tuning is not well correlated to “form” properties (Table 1) and thus adds significant new information. This result suggests that different stimulus property dimensions can contribute independently to the overall evaluation of the conflating stimulus. The CLR results suggest that the behavioral crowding effects when using different flanker color or luminance are not due to the conflation mechanisms within an individual neuron, but may result from feature attentive selection or population considerations (see Discussion). We would expect that conflating conditions that incorporated other stimulus properties, e.g., binocular stereopsis and motion, would increase the predictive power.

The utility of combining the individual tuning measures into a single measure based on simple averaging of the indexes to separate properties identified by the regression analysis was determined. Shape and color-luminance tuning indexes for the CLR series were each averaged together. Figure 8B illustrates the result. An increase in predictive power is evident. A linear regression for the CLR data indicated that the averaged index explained 45% of the variance:  $R^2 =$

0.46,  $F(1, 67) = 57.48$ ,  $p < 0.001$ . As expected the combined index (across uncorrelated features) was a stronger predictor of the maximum suppression response ( $\beta = -0.68$ ,  $p < 0.001$ ) than either index taken separately. These quantitative results strongly support conflation’s integrative hypothesis.

### Flanker modulation as a function of stimulus separation within the RF

Using the above relationships, the response as a function of flanker separation was examined after sorting the population of neurons by tuning selectivity. The sorting diminishes the cross cancellation of facilitatory and suppressive responses that occur in an average. Neurons in the SHP (shape) and IDN (identical) series were each sorted into two groups based on whether their averaged tuning indexes (orientation and size tuning) exceeded the midpoint (0.5) on the tuning selectivity index scale. Responses were pooled across neurons, binned and averaged as a function of the relative RF radius. The responses to the conflating stimuli at different separations within the RF are shown by the bar histograms of Figure 9. The results for neurons with a combined tuning index  $>0.5$  are shown in Figure 9A and D. The dashed horizontal line represents the response to the non-flanked center stimulus. For the SHP condition there is a monotonic increase in suppression (decrease in response) from flanking positions at the edge of the RF to the closest nonoverlapping, nonabutting positions near the center stimulus. The suppression averaged 50% at the smallest flanker separation. In the IDN condition where flankers are identical to the



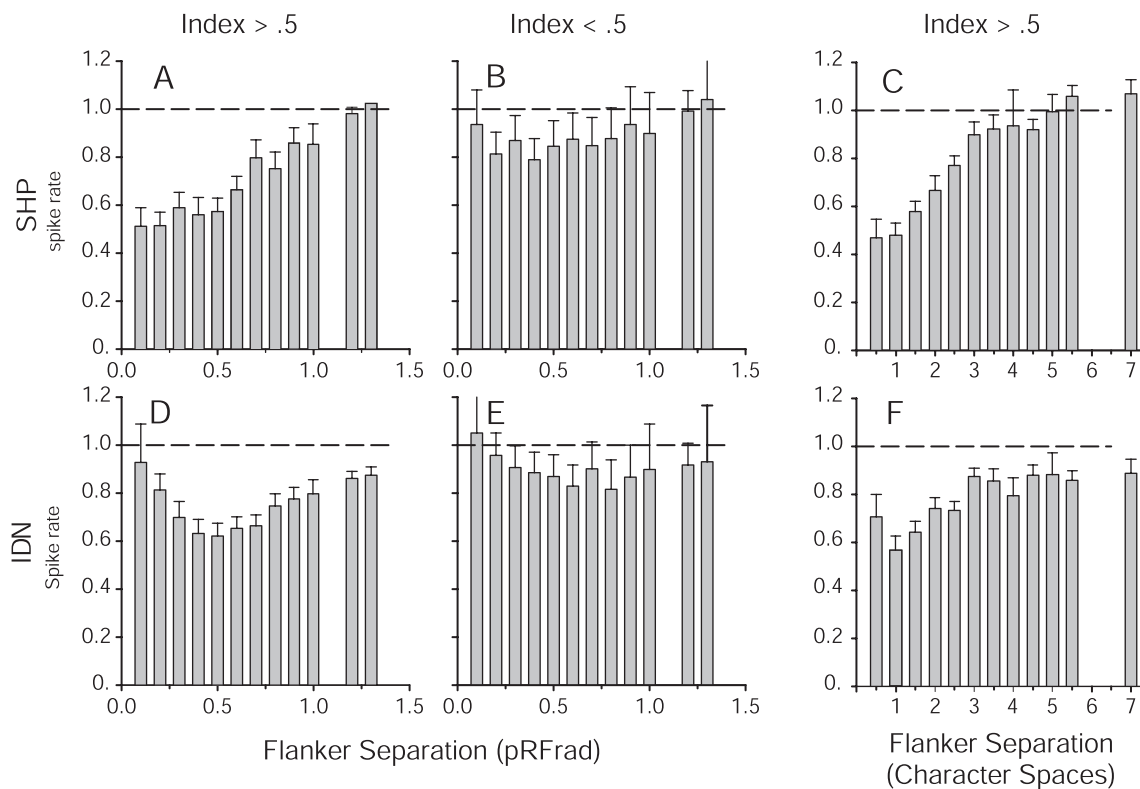


Figure 9. Response modulation as a function of flanker separation within selectivity tuned subgroups. Neurons in the SHP and IDN series were subdivided based on their averaged indexes to stimulus orientation and size selectivity. Average indexes greater than 0.5 (more sharply tuned) are shown in (A) and (D), whereas indexes more broadly tuned (values  $<0.5$ ) are shown in (B) and (E). (A) In the SHP series the suppressive effects progressively increased as flanker separation decreased within the RF. (D) In the IDN series, the suppression at larger separations reversed near the middle of the RF and returned to the preferred stimulus level of response. (B) and (E) The population of more broadly tuned neurons show a hint of the suppression seen in the more sharply tuned subgroups. (C) and (F) When flanker separation is depicted in terms of character spaces, the same progression of increased suppression for smaller separations is obtained, shown here for the more sharply tuned subgroups of both SHP and IDN series.

center stimulus, Figure 9D, there is a somewhat different result. From the RF edge to about halfway to the center, flanker suppression increased to about 60% (similar to the SHP condition), but then reversed coming back to about 90% as the flankers reach the smallest separations. This difference does not result from a sampling bias between the two experimental series as it is present when the analysis was limited to just those neurons ( $n = 30$ ) for which both SHP and IDN results were available, i.e., the same neurons responded differently in the two experimental series. It is likely the response difference results from the difference in end points of the merging stimuli; the ultimate end point of conflation, complete overlap, is for the IDN series the preferred stimulus, whereas for the SHP series it is a comparative jumble. In addition, the effective suppression related to size tuning becomes less effective at small separations. For geometric shape reasons, decreased effectiveness happens primarily with identical stimuli; color differences can disrupt this finding.

The results for neurons with tuning indexes  $<0.5$  are shown in Figure 9B and E. As expected, the integration effects of the conflating stimuli are less clear when the neurons are broadly tuned. When the central stimulus is not a dominant preferred (optimal) stimulus, the response averaged across neurons is a mixture of suppression, facilitation, and weak responses at different flanker separations. Because both the flanking separations and RF sizes varied, each neuron does not contribute to every bin; however, in all plots each bar represents averages of  $>60$  neurons. The bin size for the next largest separation was adjusted to meet these minimum neuron counts. The rightmost bar in each plot represents the response average to all flanking separations greater than 1.3 pRFrad.

In summary, the tuning selectivity of neurons for dimensions shared among the conflating stimuli can be used to predict the occurrence of flanking suppression. Sorting by tuning selectivity reveals that suppression begins at the edge of the RF and progressively increases as the flanker separation decreases. Suppression is strong even when the flanking stimuli are identical

optimal stimuli, arguing against a spatial summation model and in favor of the integration model.

## The representation of spatial separation as character spacing

Scaling the spatial separations by the RF radius avoids the problem of averaging visual angle across RFs of different sizes, but it hides any differences related to the size of the individual stimuli. An alternate method is to scale spatial separation as a function of stimulus size, providing a direct visual space metric of separation. This method of spatial scaling is common in psychophysical studies of crowding when using letters and numbers. In those studies the metric of character spacing is often defined by the height of a lower case “x.” In the present study, the letter-like stimuli all have equal height. Character spacing was defined as the center-to-center distance between the central stimulus and a flanker divided by the height of the central stimulus (all in degrees of visual angle). Again, no overlapping or abutting conditions were included.

Recasting the conflation data in terms of character spacing also results in a progressive reduction in the response as the flanking separation decreases, as shown in Figure 9C and F for the SHP and IDN series. Note that a character spacing less than 1.0 can occur without overlap when a flanker is, for example, a bar. The general relationships for character spacing are essentially the same as those for pRFrad scaling; compare for example Figure 9A and 9C, with increased suppression for decreased spatial separation of the stimuli. Character spacing represents a real space metric (not a neural space metric with variability in RF size and cortical magnification issues). Psychophysical studies have shown that the effective crowding distance increases with eccentricity; and that relationship can be normalized by scaling the stimulus size across eccentricity using a cortical magnification function. In this neurophysiological study, a correlative linkage already exists between eccentricity and RF size; therefore, in this data, the character-based scaling incorporates the eccentricity and size factors. This was confirmed by subdividing the data by eccentricity and comparing the resulting groups. They did not differ (sign ranked,  $p = 0.34$ ). These results demonstrate that spatial conflation interactions within RFs can be the basis for psychophysical crowding by flankers.

## Discussion

An integration model was proposed for examining the neural basis of visual crowding. The action of

multiple stimuli presented within a RF was examined from the point of view of a neuron that integrates all information in its RF and responds in proportion to the integrative match to the neuron’s tuning or “optimal” stimulus. Although not entirely a new idea, this is an assertion that multiple stimuli within the RF are neither individuated nor competitive. At the neuronal level the integration hypothesis takes on two forms. When the centered stimulus is the optimal stimulus for that neuron, spatial conflation results in suppression (Figure 4B) of the representation of the identity of that stimulus, a condition analogous to a perceptual incorrect rejection. Within the neural population, that diminished representation occurs in terms of both the level of activation and the number of active neurons that are coding that stimulus. On the other hand, response facilitation (Figure 4A) represents a condition where a conflating pair of nonoptimal stimuli together resemble the optimal stimulus and activate the neuron although the optimal stimulus is not actually present (analogous to a perceptual false positive). Under conflating conditions, the suppression of the activity in neurons tuned to the actual stimulus occurs just as facilitation may increase activity in neurons tuned to other stimuli that are not actually present. In analogy to crowding studies, the central target may be misidentified because there is no clear signal of its presence (suppression) from the tuned mechanisms responsible for its identification, and at the same time other tuned mechanisms at that location responsible for other letters may become activated (facilitation) and incorrectly signal the presence of a different letter or a jumble of letter shapes.

Stimulus conflation leading to suppression is the normal integrative process of the neuron working on an unfavorable mixture of stimuli within its receptive field. This process is very similar in statement to the psychophysical definition of visual crowding given by Pelli et al. (2007), when they state “Crowding is excessive feature integration, inappropriately including extra features that spoil recognition of the target object” (p. 2). Given an optimal target stimulus centered in the RF, conflation with another stimulus results in the loss of the neural representation of that target in the very population of neurons that normally code that target’s presence. Thus, from the perspective of stimulus identification, conflation results in an inappropriate integration.

Under the integration hypothesis, the dominant issue is not the spatial separation between independent objects, but a question of the degree to which the conflating stimulus combination within the RF matches the tuning of the neuron. Within the RF, the flanker position engages both the shape sensitivity gradient of the neuron (alignment and separation) and the spatial sensitivity gradient of the RF. Spatial position within

the RF is a critical tuning property itself, with selectivity depicted by the response amplitude contour gradients. Together the integration along these gradients results in suppression or facilitation, progressing in a monotonic fashion from the edge of the RF to the culminating central configuration (see Figures 4 and 9). Edge to center changes reflect not only the increased mismatch (or match) to shape tuning but also the general increasing sensitivity from edge to center of the RF.

The tuning gradient may be flat for some features while steep for other features (David et al., 2006). The CLR (color) series of experiments demonstrated this point by showing that an additional uncorrelated feature dimension can significantly increase the predictive relationship between tuning and response modulation. Whereas these details may explain the observations in individual neurons, the basis for perception will be dominated by the activity in the population of neurons that are optimally tuned for the stimuli presented.

### Critical spacing

Many visual crowding studies evaluate task performance using a measure termed “critical spacing” which is defined as the largest separation between target and flanker at which any change in target detection performance can be found (Bouma, 1970; Intriligator & Cavanagh, 2001; Pelli et al., 2004). Critical spacing in the periphery has been found to increase proportionately with eccentricity, be as large as one half the target eccentricity, and be relatively insensitive to target size (Levi, 2008; Pelli et al., 2004; Toet & Levi, 1992). The closest analogy to the critical spacing limit at the neuronal level is the radius of the receptive field. Since the RF does not change with changes in stimulus size, the interaction range would be insensitive to changes in stimulus size. However, the V4 radius/eccentricity ratio is nearly 1.0 (Figure 2A) rather than the “half the eccentricity” rule of Bouma (1970) or the 0.5 ratio measures of critical spacing as shown in figure 3 of Pelli et al., (2004). The difference may be in the degree of neural suppression needed to affect performance. Flanker suppression increases from edge to center, so the requirement for a certain level of suppression would reduce the effective RF radius and decrease the V4 ratio. As noted in the results, the most effective suppression occurred inward from about one half the RF radius, quite likely a coincidental observation but intriguing. These issues need to be examined when RF conflation is studied while the subject is making crowding judgments. Until then it seems reasonable to propose that the scaling limitations in peripheral letter crowding are consistent with integration within neural

receptive fields. Area V4 is a candidate but likely not the sole cortical area of conflation interactions.

While the character scaling method fractures the relationship between flanker separation and position within the RF in favor of a relationship between flanker separation and stimulus size, the end result of the two scaling methods in terms of the population response is remarkably similar (See Figure 9). Both measures incorporate the increase in RF size with eccentricity. Both measures give very similar accounts of spatial conflation. Why? Such a result is expected if there is a strong relationship between stimulus height and RF radius, but this relationship is very weak over the range examined (see Figure 2). A more likely possibility combines two factors. The first is that the shape tuning selectivity of a V4 neuron remains invariant within that neuron’s stimulus size tuning range. The second factor is that the response is governed by the shape formed by conflation, not the specific spatial separation. Therefore as the scaling of the conflated stimulus changes (both stimulus size and separation), the shape invariance of the neuron preserves the response to the conflated stimulus across the scale change. If true, then character sized spacing for normal text is a more robust measure of stimulus separation than visual angle. This issue is at the heart of the size versus spacing issue in visual crowding and reading (Chung, 2014; Song, Levi, & Pelli, 2014). We conclude that optimal stimulus size and optimal letter spacing are governed by the same RF-based mechanisms, which in turn are an integral part of the neural basis of object recognition. The RF boundary constraint is important as it sets the maximum extent of any conflation interaction, and is analogous to the integration area in perceptual crowding. The RF boundary constraint will not change as a function of stimulus size; therefore, at any given eccentricity there will be a U-shaped crowding function as stimulus size increases. Hypotheses based on these possibilities need to be tested.

### Target selectivity and flanker similarity

Behavioral studies have found crowding interactions to be strongly dependent on the similarity of target and flanker along various stimulus dimensions (Kooi, et al., 1994; Levi, 2008; Strasburger et al., 2011). Stimulus similarity in conflation and crowding are related but seemingly distinct issues. For neural conflation the critical issue is whether the flanker contains features to which the neuron is sensitive. With respect to conflation, similarity refers to a common afferent access to the neuron. If the flanker does not contain features within the tuning sensitivity of the neuron, it cannot conflate with the target. Feature attentive selection, but perhaps not directed spatial attention, can potentially

alter the strength of afferent information and impact conflation by reducing flanker feature access (David, Hayden, Mazer, & Gallant, 2008; Maunsell & Treue, 2006; Motter, 1994). Generally speaking, across the population of neurons, tuning selectivity implies that flankers that are markedly dissimilar to the target will have fewer features with afferent access to the pool of neurons coding for the target. The IDN series epitomizes similarity, and produced clear neural suppression, but psychophysical measures of crowding, even our own, show less crowding effects for IDN conditions than the neural data suggest should occur. However, we were unsuccessful at eliminating the possibility that flanker information substituted for central stimulus information in behavioral IDN target identification. This is a critical difference that needs resolution.

The pool of neurons most strongly representing a stimulus are those with RFs centered on the stimulus; thus, separate pools of neurons code for target and flankers. Crowding performance encompasses the interpretation of the information from these different pools. Conflation in single neurons and the interpretation of the population pool represent different neural stages in crowding. Physiologically, each successive stage deals with a different level of information integration, potentially at different spatial scales and similarity dimensions. Do the perceptual consequences we term crowding occur at one cortical level of processing? That seems an unnecessary constraint given the broad range of spatial scales in objects that crowd each other. Zhang, Zhang, Xue, Liu, and Yu (2009) have suggested that crowding within characters can explain changes in the acuity threshold slopes as a function of eccentricity as the spatial complexity of Chinese character sets increase. Bernard and Chung (2011) have shown that increases in flanker complexity produce increases in crowding using a series of fonts more complex than the letter-like stimuli used here. A fine detailed analysis of the internal components of a complex font appears to go beyond the spatial limit of V4 neurons. A reasonable hypothesis is that the mechanisms of conflation and spatial detail occur at each recurrent stage in the vertically organized processing unit for each retinotopic position extending from V1 to V4 cortex and beyond. This hypothesis is consistent with imaging results showing crowding at multiple levels.

At some point in conflation there is no longer a threshold signal of the target (or flanker) in the pool of neurons at its spatial location, perhaps even no signal of any reportable feature combination (something is there, but what?). When it comes to a population evaluation of “what is out there,” a difference in the level of activity favoring flanker pools could make the flanker identity stand out. This impacts not only the

identity of what is present in the population of active neurons but also introduces relative positional uncertainty among the stimuli that remain. Directed spatial attention may be able to select which pools of neurons are selected for further processing and even dynamically alter the size of the selected pool (Reynolds, Chelazzi, & Desimone, 1999). Stimulus positional uncertainty is consistent with psychophysical models of crowding that examine attentive misselection as the basis of crowding (He, Cavanagh, & Intriligator, 1996; Strasburger, Harvey, & Rentschler, 1991).

Furthermore, these population issues parallel psychophysical evidence for the distinction between feature confusion and source confusion (Strasburger et al., 2011; Strasburger & Malania, 2013), and within versus between character crowding studies (Zhang et al., 2009). Strong conflation could result in the loss of feature combination identities across the pools of neurons, providing only average feature information, something like texture, to the next processing level or perception in general (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001).

Understanding the functions governing conflation interactions, i.e., the nature of object integration, will require understanding the functions governing the tuning selectivity of the neuron. Stimulus conflation within the receptive field can be a natural ubiquitous spatial bottleneck for visual processing anywhere in the visual system.

## Selectivity, clutter and receptive field size

As seen in the single flanker (SFL) series, the influence of a single flanker can modulate activity as effectively as multiple flankers. The issue is not the number of flankers, but the specific factors (spatial position and feature properties) that the flanker brings to the integration that significantly alters the match between the conflated stimulus and the tuning configuration of the neuron. A similar psychophysical result occurs with respect to the specifics of the interacting elements themselves (Rosen, Chakravarthi, & Pelli, 2014). Whereas other studies model the interaction between independent stimuli as a weighted summation or a divisive normalization process (Britten & Heuer, 1999; Desimone & Duncan, 1995; Reynolds et al., 1999; Zoccolan et al., 2005), those approaches often do not work when one of the stimuli is poorly effective in producing a response when presented alone (Britten & Heuer, 1999; Zoccolan et al., 2005). For the integration hypothesis, combining a flanking stimulus with a centered optimal stimulus results in suppression when the conflated stimulus combination evaluates as less optimal, irrespective of the flanker’s individual effectiveness. The center of the pool of neurons whose RFs

overlap the stimulus are the set of neurons with the stimulus in the center of their RFs. Stimulus placements that are off-center in the RF not only produce a less effective response in that neuron, but that neuron in turn occupies an off-center and less effective position in the (topographic) pool of neurons coding the stimulus.

Certainly increasing the number of flankers within the RF increases the probability of a conflated new object, at least up to a point, beyond which individual objects are not recognizable. In area V4 this problem is compounded by the size of the RFs. Conflation changes begin to occur when flanking stimuli are several character spaces away. This would not occur if the RFs were smaller, for example, if they matched their size tuning profiles. Given the powerful detrimental effects, there should be a reason for the existence of RFs much larger than the neuron's optimal stimulus configuration. Large receptive fields were traditionally proposed as the mechanism by which spatial translational invariance of object recognition is achieved. However, this is logically inconsistent with a retinotopically organized area and in fact the actual invariance of stimulus placement within V4 and IT RFs is less than previously imagined (DiCarlo & Maunsell, 2003; Nandy et al., 2013). Our preliminary results suggest that these large RFs have a role in figure-ground segregation under attentive conditions.

The integration hypothesis emphasizes the processing specificity of a neuron that includes not only stimulus properties but centering of the stimulus within the RF. Unfortunately, experimentally the optimal stimulus is a known unknown, the expression of a hypothetical maximum of tuning sensitivities. Fortunately, a strong ranking preference appears to be sufficient to probe the mechanism. Left open is the question of whether all neurons are highly tuned and those found to be less selective simply represent cases at the edge (or outside) of the stimulus domain of the experiment.

## Conclusions

Conflation at a neural level, the fusing of separate objects into a single identity, can be readily observed within cortical area V4 RFs because the optimal stimulus size for letter-like stimuli is markedly smaller than the V4 RF diameter, making it possible to place several items within the RF. Stimulus flanking interactions are wholly contained within the classic RF of the neuron. The size of the RF is analogous to the psychophysically defined crowding integration area. Flanking separations that result in perceptual crowding result in a response suppression to an optimal stimulus centered in V4 RFs. The degree of suppression is

predicted by the tuning selectivity of the neuron and to the flanker separation within the RF. Stimulus conflation, the role of tuning properties, and flanker separation are functions of single neurons. The perceptual assessment of the responses for both target and flankers must occur across the neural population. Attentive strategies potentially affect both individual and population outcomes through control of afferent input. Our results indicate that the neural conflation responses, and ultimately crowding performance, are the result of the neural mechanisms associated with object identification and recognition, and are not some other independent visual phenomena.

*Keywords:* visual crowding, Area V4, receptive fields, tuning selectivity, conflation, object identification

## Acknowledgments

The author thanks Heather Bergsbaken for her continuous, patient, technical support. This work was supported by grants from the National Eye Institute, NIH R01-EY018693, and from the Veterans Affairs Bio-Medical Research Program.

Commercial relationships: none.

Corresponding author: Brad C. Motter.

Email: brad.motter@gmail.com.

Address: Veterans Affairs Medical Center, Syracuse, NY, USA.

## References

- Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin*, *53*, 452–471.
- Bernard, J.-B., & Chung, S. T. L. (2011). The dependence of crowding on flanker complexity and target–flanker similarity. *Journal of Vision*, *11*(8):1, 1–16, doi:10.1167/11.8.1. [PubMed] [Article]
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177–178.
- Bouma, H. (1978). Visual search and reading: eye movements and functional visual field: A tutorial review. In J. Requin (Ed.), *Attention and performance* (7th ed., pp. 115–147). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Britten, K. B., & Heuer, W. H. (1999). Spatial

- summation in the receptive fields of MT neurons. *Journal of Neuroscience*, *19*, 5074–5084.
- Carlson, E. T., Rasquinha, R. J., Zhang, K., & Connor, C. E. (2011). A sparse object coding scheme in area V4. *Current Biology*, *21*, 288–293.
- Changizi, M. A., Zhang, Q., Ye, H., & Shimojo, S. (2006). The structures of letters and symbols throughout human history are selected to match those found in objects in natural scenes. *American Naturalist*, *167*, E117–E139.
- Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. *Cerebral Cortex*, *11*, 761–772.
- Chung, S. T. L. (2014). Size or spacing: Which limits letter recognition in people with age-related macular degeneration? *Vision Research*, *101*, 167–176.
- Chung, S. T. L., & Tjan, B. S. (2007). Shift in spatial scale in identifying crowded letters. *Vision Research*, *47*, 437–451.
- Connor, C. E. (2004). Shape Dimensions and Object Primitives. In: L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (2nd ed., pp. 1080–1089). Cambridge, MA: MIT Press.
- Connor, C. E., Brincat, S. L., & Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Current Opinion in Neurobiology*, *17*, 140–147.
- Cukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*, 767–770.
- David, S. V., Hayden, B. Y., & Gallant, J. L. (2006). Spectral receptive field properties explain shape selectivity in area V4. *Journal of Neurophysiology*, *96*, 3492–3505.
- David, S. V., Hayden, B. Y., Mazer, J. A., & Gallant, J. L. (2008). Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. *Neuron*, *59*, 509–521.
- Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*, 1245–1255.
- Desimone R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.
- Desimone, R., & Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: Sensitivity to stimulus form. *Journal of Neurophysiology*, *57*, 835–868.
- DiCarlo, J. J., & Maunsell, J. H. R. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of Neurophysiology*, *89*, 3264–3278.
- DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*, 425–434.
- El-Shamayleh, Y., & Pasupathy, A. (2016). Contour curvature as an invariant code for objects in visual area V4. *Journal of Neuroscience*, *36*, 5532–5543.
- Flom, M. C., Heath, G. G., & Takahashi, E. (1963, November 15). Contour interaction and visual resolution: Contralateral effects. *Science*, *142*, 979–980.
- Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *Journal of Neurophysiology*, *76*, 2718–2739.
- Gattass, R., Sousa, A. P., & Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *Journal of Neuroscience*, *8*, 1831–1845.
- Gawne, T. J., & Martin, J. M. (2002). Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *Journal of Neurophysiology*, *88*, 1128–1135.
- Ghose, G. M., & Maunsell, J. H. R. (2008). Spatial summation can explain the attentional modulation of neuronal responses to multiple stimuli in area V4. *Journal of Neuroscience*, *28*, 5115–5126.
- Glantz, S. A., & Slinker, B. K. (1990). *Primer of applied regression and analysis of variance* (pp. 191–193). New York, NY: McGraw-Hill.
- He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*, 334–337.
- Hegde, J., & Van Essen, D. C. (2007). A comparative study of shape representation in macaque visual areas V2 and V4. *Cerebral Cortex*, *17*, 1100–1116.
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*, 171–216.
- Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, *71*, 856–867.
- Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision*, *8*, 255–279.
- Levi, D. M. (2008). Crowding—An essential bottleneck

- for object recognition: A minireview. *Vision Research*, 48, 635–654.
- Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding: Selective interference between configural representations of faces in crowded scenes. *Journal of Vision*, 7(2):24, 1–11, doi:10.1167/7.2.24. [PubMed] [Article]
- Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, 77, 24–42.
- Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, 29, 317–322.
- Miller, E. K., Gochin, P. M., & Gross, C. G. (1993). Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Research*, 616, 25–29.
- Millin, R., Arman, A. C., Chung, S. T. L., & Tjan, B. S. (2014). Visual crowding in V1. *Cerebral Cortex*, 24, 3107–3115.
- Missal, M., Vogels, R., Li, C.Y., & Orban, G.A. (1999). Shape interactions in macaque inferior temporal neurons. *Journal of Neurophysiology*, 82, 131–142.
- Moody, S. L., Wise, S. P., di Pellegrino, G., & Zipser, D. (1998). A model that accounts for activity in primate frontal cortex during a delayed matching-to-sample task. *Journal of Neuroscience*, 18, 399–410.
- Moran, J., & Desimone, R. (1985, August 23). Selective attention gates visual processing in the extrastriate cortex. *Science*, 229, 782–784.
- Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *Journal of Neuroscience*, 14, 2178–2189.
- Motter, B. C. (2006). Modulation of transient and sustained response components of V4 neurons by temporal crowding in flashed stimulus sequences. *Journal of Neuroscience*, 26, 9683–9694.
- Motter, B. C. (2009). Central V4 receptive fields are scaled by the V1 cortical magnification and correspond to a constant-sized sampling of the V1 surface. *Journal of Neuroscience*, 29, 5749–5757.
- Motter, B. C., & Simoni, D. A. (2007). The roles of cortical separation and size in active visual search performance. *Journal of Vision*, 7(2):6, 1–15, doi:10.1167/7.2.6. [PubMed] [Article]
- Nakamura, H., Gattass, R., Desimone, R., & Ungerleider, L.G. (1993). The modular organization of projections from areas V1 and V2 to areas V4 and TEO in macaques. *Journal of Neuroscience*, 13, 3681–3691.
- Nandy, A. S., Sharpee, T. O., Reynolds, J. H., & Mitchell, J. F. (2013). The fine structure of shape tuning in area V4. *Neuron*, 78, 1102–1115.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4, 739–744.
- Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology*, 82, 2490–2502.
- Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: Position-specific tuning for boundary conformation. *Journal of Neurophysiology*, 86, 2505–2519.
- Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, 46, 4646–4674.
- Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, 4(12):12, 1136–1169, doi:10.1167/4.12.12. [PubMed] [Article]
- Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, 11, 1129–1135.
- Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, 7(2):20, 1–36, doi:10.1167/7.2.20. [PubMed] [Article]
- Pollen, D. A., Przybyszewski, A. W., Rubin, M. A., & Foote, W. (2002). Spatial receptive field organization of macaque V4 neurons. *Cerebral Cortex*, 12, 601–616.
- Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, 27, 611–647.
- Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, 19, 1736–1753.
- Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., . . . Vanduffel, W. (2012). Toward a unified theory of visual area V4. *Neuron*, 74, 12–29.
- Rosen, S., Chakravarthi, R., & Pelli, D. G. (2014). The Bouma law of crowding, revised: Critical spacing is equal across parts, not objects. *Journal of Vision*, 14(6):10, 1–15, doi:10.1167/14.6.10. [PubMed] [Article]

- Rust, N. C., & DiCarlo, J. J. (2010). Selectivity and tolerance (“invariance”) both increase as visual information propagates from cortical area V4 to IT. *Journal of Neuroscience*, *30*, 12978–12995.
- Sandler, A. J. (2008). Chronic recording during learning. In M. A. L. Nicolelis (Ed.), *Source methods for neural ensemble recordings* (pp. 125–143). Boca Raton, FL: CRC Press.
- Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Experimental Brain Research*, *77*, 23–30.
- Shipp, S., & Zeki, S. (1985). Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex. *Nature*, *315*, 322–325.
- Song, S., Levi, D. M., & Pelli, D. G. (2014). A double dissociation of the acuity and crowding limits to letter identification, and the promise of improved visual screening. *Journal of Vision*, *14*(5):3, 1–37, doi:10.1167/14.5.3. [PubMed] [Article]
- Sripati, A. P., & Olson, C. R. (2010). Responses to compound objects in monkey inferotemporal cortex: the whole is equal to the sum of the discrete parts. *Journal of Neuroscience*, *30*, 7948–7960.
- Stepniewska, I., Collins, C. E., & Kaas, J. H. (2005). Reappraisal of DL/V4 boundaries based on connectivity patterns of dorsolateral visual cortex in macaques. *Cerebral Cortex*, *15*, 809–822.
- Strasburger, H., Harvey, L. O. J., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics*, *49*, 495–508.
- Strasburger, H., & Malania, M. (2013). Source confusion is a major cause of crowding. *Journal of Vision*, *13*(1):24, 1–20, doi:10.1167/13.1.24. [PubMed] [Article]
- Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5):13, 1–82, doi:10.1167/11.5.13. [PubMed] [Article]
- Sundberg, K. A., Mitchell, J. F., & Reynolds, J. H. (2009). Spatial attention modulates center-surround interactions in macaque visual area v4. *Neuron*, *61*, 952–963.
- Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, *32*, 1349–1357.
- Tripathy, S. P., & Levi, D. M. (1994). Long-range dichoptic interactions in the human visual cortex in the region corresponding to the blind spot. *Vision Research*, *34*, 1127–1138.
- Whitney, D., & Levi, D. M. (2011). Visual Crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*, 160–168.
- Zhang, J. Y., Zhang, T., Xue, F., Liu, L., & Yu, C. (2009). Legibility of Chinese characters in peripheral vision and the top-down influences on crowding. *Vision Research*, *49*(1), 44–53.
- Zoccolan, D., Cox, D. D., & DiCarlo, J. J. (2005). Multiple object response normalization in monkey inferotemporal cortex. *Journal of Neuroscience*, *25*, 8150–8164.