Neurocomputational principles underlying the number sense

Joonkoo Park[1,2*], David E. Huber[1]

[1] Department of Psychological and Brain Sciences, University of Massachusetts Amherst, Amherst, MA 01003, U.S.A.

[2] Commonwealth Honors College, University of Massachusetts Amherst, Amherst, MA 01003, U.S.A.

**Corresponding author*

Joonkoo Park, Ph.D.
135 Hicks Way
Amherst MA 01003
Email: joonkoo@umass.edu
Phone: 1-413-545-0051

Abstract

Many species of animals exhibit an intuitive sense of number, suggesting a fundamental neural mechanism for representing numerosity in a visual scene. Recent empirical studies demonstrate that early feedforward visual responses are sensitive to numerosity of a dot array but not to continuous dimensions orthogonal to numerosity, such as size and spacing of the dots. Here we identified the core neurocomputational principles underlying these effects: (1) center-surround contrast filters; (2) at different spatial scales; with (3) divisive normalization across network units. In an untrained computational model, these principles eliminated sensitivity to size and spacing, making numerosity the main determinant of the neuronal response magnitude. In addition to identifying these core principles, we demonstrate that divisive normalization can explain illusions concerning numerosity perception across space and time. These results identify a set of neurocomputational properties that gives rise to the ubiquity of the number sense in the animal kingdom.

Introduction

Humans have an intuitive sense of number that allows numerosity estimation without counting [1]. The prevalence of number sense across phylogeny and ontogeny [2] suggests common neural mechanisms that allow the extraction of numerosity information from a visual scene. Nevertheless, these mechanisms have not yet been identified.

While earlier empirical work highlighted the parietal cortex for numerosity representation [3], growing evidence suggests that numerosity is processed at a much earlier stage. A recent study, using high-temporal resolution electroencephalography together with a novel stimulus design, demonstrated that early visual cortical activity is uniquely sensitive to the numerosity of a dot array in the absence of any behavioral response, but not to non-numerical dimensions that are orthogonal to numerosity (i.e., size and spacing, **Fig. 1A**) [4]. Subsequent behavioral and neural studies showed that this early cortical sensitivity to numerosity indicates feedforward activity in visual areas V1, V2 and V3 [5–7]. These results suggest that numerosity is a basic currency of perceived magnitude early in the visual stream.

Nevertheless, it is unclear how feedforward neural activity creates a representation of numerosity within these brain regions. Specifically, the view of numerosity as a *discrete number* of items seems incompatible with the primary modes of information processing in the brain, such as firing rates and population codes, which are *continuous*. Indeed, some authors assume that continuous non-numerical magnitude information is encoded first and integrated to produce the representation of numerosity [8,9]. In contradiction, however, more empirical studies now demonstrate that the magnitude of visual cortical activity is most sensitive to number and is relatively insensitive to other continuous dimensions such as size and spacing of a dot array [10–12].

What explains this insensitivity to spacing and size effects, despite robust sensitivity to number? Previous computational modeling studies offer some hints to this question. The earliest computational model of numerosity perception by Dehaene and Changeux [13] assumes that the size of each dot in an array becomes normalized within a convolution layer in which the filter that matches the dot size dominates the neuronal activity as a result of a winner-take-all process. At the time this model was proposed, it was not known that numerosity occurs in early visual processing. As applied to early vision, this strong winner-take-all mechanism is implausible, as the model would suggest that visual cortex only knows that dots exist, without knowing the size or the location of the dots. Also, this model does not allow *any* effect of size or spacing, which is inconsistent with empirical data [14].

Recently, several deep-network-based models have been applied to numerosity perception [15–19]. Stoianov and Zorzi [18] developed a hierarchical generative model of the sensory input (images of object arrays) and demonstrated that after learning to generate its own sensory input, some units in the hidden layer were sensitive to numerosity irrespective of total area while other units were sensitive to total area irrespective of numerosity. This suggests an unsupervised learning mechanism for efficient coding of the sensory data that can extract statistical regularities of the input images. The authors provided some suggestions as to the specific neurocomputational principle(s) underlying the success of this model. For example, the first hidden layer developed center-surround representations of different sizes and the second layer developed a pattern of

excitatory and inhibitory connections to the first layer that could be considered a form of normalization. However, the development of center-surround detectors based on unsupervised learning is a common observation [20], indicating that such results are not unique to displays of dot arrays, and are instead a natural byproduct of learning in the visual system. In a more recent study, Kim and colleagues [19] found that sensitivity and selectivity to numerosity was well captured in a completely untrained convolutional neural network (AlexNet) [21], suggesting that a repeated process of convolution and pooling is capable of normalizing continuous dimensions and extracting numerosity information as a statistical regularity of an image. However, why such a process results in numerosity-sensitive units remains unclear.

Rather than applying a complex structure or learning model, we distill the neurocomputational principles that enable the visual system to be sensitive to numerosity while remaining insensitive to non-numerical visual features. Consistent with prior work, we suggest that center-surround contrast filters at different spatial scales play an important role in numerosity perception. In addition, we hypothesize that divisive normalization [22], in which the response of each neuron reflects its driving input divided by the summation of responses from anatomically surrounding neurons, plays a key role in numerosity perception. In the case of early vision, the surround is spatially determined in terms of retinotopic positions. A wealth of evidence indicates that divisive normalization is ubiquitous across species and brain systems and hence thought to be a fundamental computation of many neural circuits. Thus, any theory of numerosity in visual cortex would be remiss, *not* to include the effect of divisive normalization.

To test whether these neurocomputational principles provide a sufficient explanation, we implemented a neural network with or without divisive normalization as applied to center-surround filters at different spatial scales (e.g., as in V1) (**Fig. 1B**). The output is thought to simulate the summation of synchronized postsynaptic activity of a large population of neurons prior to any involvement of behavior, as shown in previous work [4,6]. Our results show that: hierarchically organized multiple center-surround filters of varying size make the network insensitive to spacing; and divisive normalization implemented across network units makes the network insensitive to size. Divisive normalization not only occurs over space but also over time [23]. Thus, we additionally implemented temporal divisive normalization, finding that it explains contextual influences on numerosity, such as illusory numerosity adaptation [24].

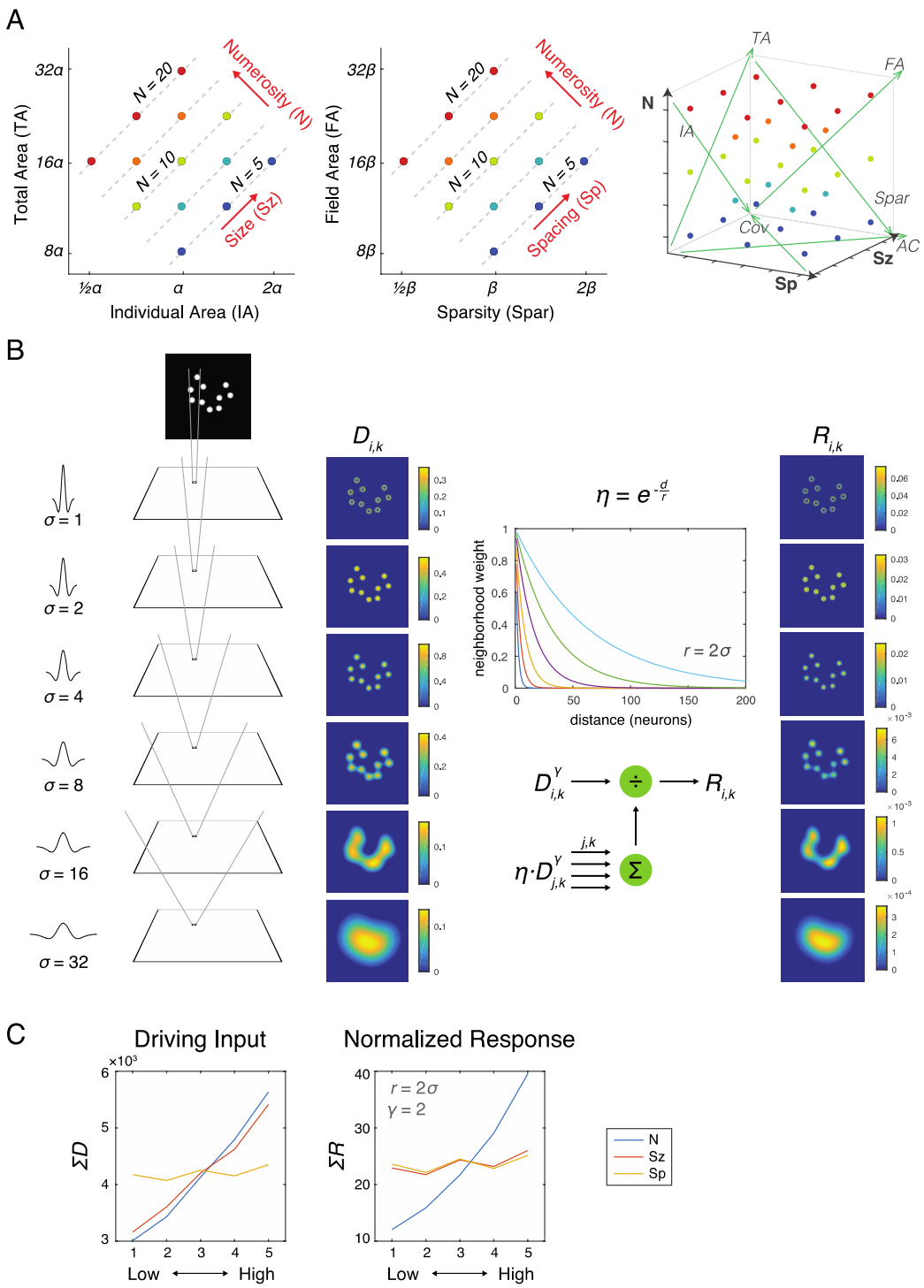

Figure 1. *Stimulus design and computational methods. A. Properties of magnitude dimensions represented in three orthogonal axes defined by log-scaled number (N), size (Sz), and spacing (Sp) (Table 1). B. Schematic illustration of the computational process from a dot-array image to the driving input, D, of the simulated neurons, and then to the normalized response, R. A bitmap image of a dot array was fed into a convolutional layer with DoG filters in six different sizes (Eq. 1). The resulting values, after half wave rectification, represented the driving input. Neighborhood weight, defined by η, was multiplied by*

the driving input across all the neurons across all the filter sizes, the summation of which served as the normalization factor (see Eq. 2 & 3). This illustration of η is showing the case where r is defined by twice the size of the sigma for the DoG kernel. C. Simulation results showing the effects of N, Sz, and Sp on the driving input and normalized response of the network units. The driving input showed a near elimination of the Sp effect. The normalized response additionally showed a near elimination of the Sz effect. The results were simulated using $r = 2\sigma$ and $\gamma = 2$, but effects of Sz and Sp were negligible across all the tested model parameters (Fig. S3).

Results

Center-surround convolution captures total pixel intensities and eliminates the effect of spacing

Images of dot arrays that varied systematically across number, size, and spacing (see Materials and Methods) were fed into a convolutional layer with difference-of-Gaussians (DoG) filters in six different sizes. The driving input, *D*, for each filter was the convolution of a DoG with the display image, or a weighted sum of local pixel intensities (**Fig. 1B**). Prior to divisive normalization, the summed driving input across all spatial positions and filter sizes, ΣD, was most strongly modulated by both number and size equally but not by spacing (**Fig. 1C**), suggesting that the neural activity tracks total area (*TA*; see **Table 1**; **Fig. S1**). The effect of spacing existed in the fourth and sixth largest filter sizes, largely indicating effects of field area and density, respectively (**Fig. S2**); however, the effects in these two filter sizes were in opposite directions, which made the overall effect very small. These results illustrate that having multiple filter sizes is key to normalizing the spacing dimension. In sum, the driving input of the convolutional layer captured total pixel intensity of the image regardless of the number or spatial configuration of dots.

Divisive normalization nearly eliminates the effect of size

Next, we implemented divisive normalization with different parameter values (neighborhood size and amplification factor) to determine the conditions under which divisive normalization might reduce or eliminate the effect of size and whether it might alter the absence of spacing effects in the driving input. Driving input was normalized by the normalization factor defined by a weighted summation of neighboring neurons and filter sizes (**Eq. 2**). The summed normalized responses, ΣR, were strongly modulated by number but much less so, if any, by size and spacing (**Fig. 1C**). The pattern of results was largely consistent across different parameter values for neighborhood size (r) and amplification factor (γ) of the normalization model (**Fig. S3**); therefore, we chose moderate values of r (=2) and γ (=2) for subsequent simulations. A regression model with summed normalized responses as the dependent measure and the three orthogonal dimensions (*N*, *Sz*, *Sp*) as the independent variables revealed a much larger coefficient estimate for *N* (b = 13.68) than for *Sz* (b = 1.541) and for *Sp* (b = 0.7809). In sum, a modest degree of divisive normalization eliminates the effect of size and, at the same time, does not alter the absence of spacing effects.

Divisive normalization across space explains various visual illusions

Numerosity perception has long been demonstrated to be susceptible to illusions. Here, we test if divisive normalization explains some of the most well-known visual illusions for numerosity perception. Empirical studies have long shown that irregularly spaced arrays (compared with regularly spaced arrays), arrays with spatially grouped items (compared with ungrouped items), and arrays with dots that are pairwise connected (compared with unconnected arrays) are all underestimated [25–29]. All of these illusions were captured by the computational model implementing divisive normalization (**Fig. 2A-C**). Critically, there is no learning in the model, and behavior is assumed to reflect the summation of all neural activity across all filter sizes. These effects can be explained by greater normalization when neurons with overlapping normalization neighborhoods are activated, with this greater overlap occurring in subregions of the images for irregular, grouped, or connected (lines) dots. This explanation is functionally similar to one provided by the "occupancy model" [30], but our results demonstrate that these effects emerge naturally from early visual processing.

A relatively understudied visual illusion is the effect of heterogeneity of dot size on numerosity perception. A recent behavioral study demonstrated that the point of subjective equality was about 5.5% lower in dot arrays with heterogenous sizes compared with dot arrays with homogeneous sizes [31]. Consistent with this behavioral phenomenon, our simulations revealed that greater heterogeneity leads to greater underestimation (**Fig. 2D**). This occurs because, within a dot array, the effect of each dot's size on the normalized response of the network units with receptive fields targeting the dot saturates as the dot's size increases. As heterogeneity is manipulated by making some dots larger and other dots smaller by the same amount in order to keep the total area and numerosity constant, this saturating effect of individual area on normalized response makes the overall normalized response smaller as a greater number of dots deviates from the average size (see **Fig. S4**).

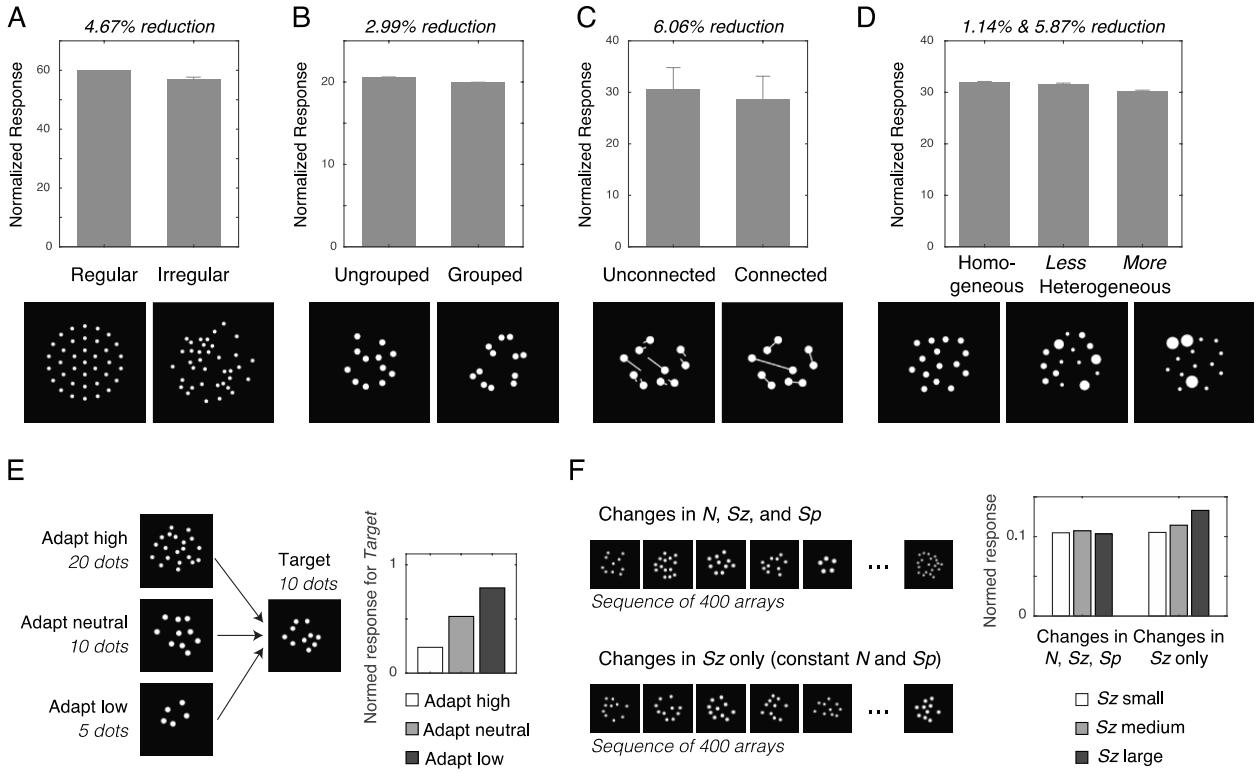

Figure 2. *Simulation of numerosity illusions. Normalized response of the network units influenced by the (A) regularity, (B) grouping, (C), connectedness, and (D) heterogeneity of dot arrays, as well as by (E) adaptation and (F) context. Error bars represent one standard deviation of the normalized response across simulations. In the case of temporal normalization effects (E and F), the results were simulated with* $\omega = 8$ *and* $\delta = 2$*, but the patterns held across all tested model parameter values.*

Divisive normalization across time explains numerosity adaptation and context effects

One of the most well-known visual illusions in numerosity perception is the adaptation effect [24]. We reasoned that numerosity adaptation might be driven by divisive normalization across time, which shifts the response curve and produces temporal contrast effects (i.e., a perceptual aftereffect). We modeled temporal divisive normalization for readout neurons that are driven by the sum of the normalized responses across all units, ΣR. This summed total response (now referred to as M) was temporally normalized (M^*) by the recency weighted average of the driving input (**Eq. 4**). Temporal normalization shifts the sigmoid response curve horizontally along the dimension of M to maximize the sensitivity of M^* based on the recent history of stimulation. Provided that the constant in the denominator is approximately equal to the current trial's response, the results of spatial normalization reported above would not change by also introducing temporal normalization. Temporal normalization was assessed for cases of a target array of 10 dots after observing an array of 5 dots, 10 dots, or 20 dots (**Fig. 2E**). Similar to behavioral results [32], the target of 10 dots was underestimated when the adaptor was more numerous than the target and was overestimated when the adaptor was less numerous than the target.

The most recent stimulus most strongly drives adaptation, but according to the temporal summation in (**Eq. 4**), stimuli prior to the most recent also contribute to temporal normalization, albeit to a lesser extent. Thus, temporal divisive normalization might also explain longer-sequence context effects. Studies show that the effect of size is negligible in the context of a trial sequence that varies size, spacing, and number [4], but that the effect of size becomes apparent when number and spacing are held constant while varying only size [14]. We simulated each of these contexts: The model saw a total of 400 dot arrays that varied across number, size, and spacing or else it saw 400 dot arrays that differed only in size (**Fig. 2F**). In the context where all dimensions varied, the three levels of *Sz* had no linear association with M^*; the 95 percentile confidence interval of the ordinary-least-square linear slope of M^* as a function of *Sz* was [-0.0411, 0.0293], which includes 0. In contrast, in the context where only size varied, M^* was positively correlated with *Sz*; slope confidence interval of [0.00655, 0.00725], which excludes 0. This phenomenon can be explained by the adaptive shifting of the sigmoid response curve across trials. In the former case, because recent trials are often of larger or smaller total response as compared to the current trial, the normalization for the current trial is more often pushed to the nonlinear parts of the normalization curve (e.g., closer to ceiling and floor effects). Thus, the temporally normalized response is relatively insensitive to the small effect of size (keeping in mind that the effect of size is made small by spatial divisive normalization). In contrast, when only size varies across trials, the total response of recent trials is more likely to be well-matched to the total response of the current trial. As a result, the small effect of size is magnified in light of this temporal stability.

Discussion

Despite the ubiquity of number sense across animal species, it was previously unclear how unadulterated perceptual responses produce the full variety of numerosity perception effects. Recent empirical studies demonstrate that feedforward neural activity in early visual areas is uniquely sensitive to the numerosity but much less so, if any, to the dimension of size and spacing, which are continuous non-numerical dimensions that are orthogonal to numerosity. Despite recent advances showing that numerosity information *can* be extracted from a deep neural network [15,18,19], precisely *how* early visual areas normalize the effects of size and spacing was unclear.

The current study identified the key neurocomputational principles involved in this process. First, the implementation of hierarchically organized multiple sizes of center-surround filters effectively normalizes spacing owing to offsetting factors (**Fig. 3A**). On the one hand, relatively smaller filters that roughly match or are slightly bigger than each dot produce a greater response when the dots are farther apart because their off-surround receptive fields do not overlap. On the other hand, relatively larger filters that cover most of the array produce a greater response when the dots are closer together because stimulation at the center of the on-surround receptive fields is maximized. When summing these opposing effects, which occur at different center-surround filter sizes, the overall neural activity is relatively invariant to spacing. Second, the implementation of divisive normalization reduces the effect of size by reducing activity at larger filter sizes that have overlapping normalization neighborhoods (**Fig. 3B**). More specifically, increase in size produces greater overall unnormalized activity because more filters (e.g., both larger and smaller) are involved in responding to larger dots while fewer filters (e.g., only

smaller) are involved in responding to smaller dots; remember that driving input is proportional to total area (**Fig. 1C**). However, normalization dampens this increase. In sum, contrast filters at different spatial scales and divisive normalization naturally increases sensitivity to the number of items in a visual scene. Because these neurocomputational principles are commonly found in visual animals, this suggests that numerosity is a natural property of the visual system. Our implementation of divisive normalization across the time domain additionally explains adaptation and context effects of numerosity perception.

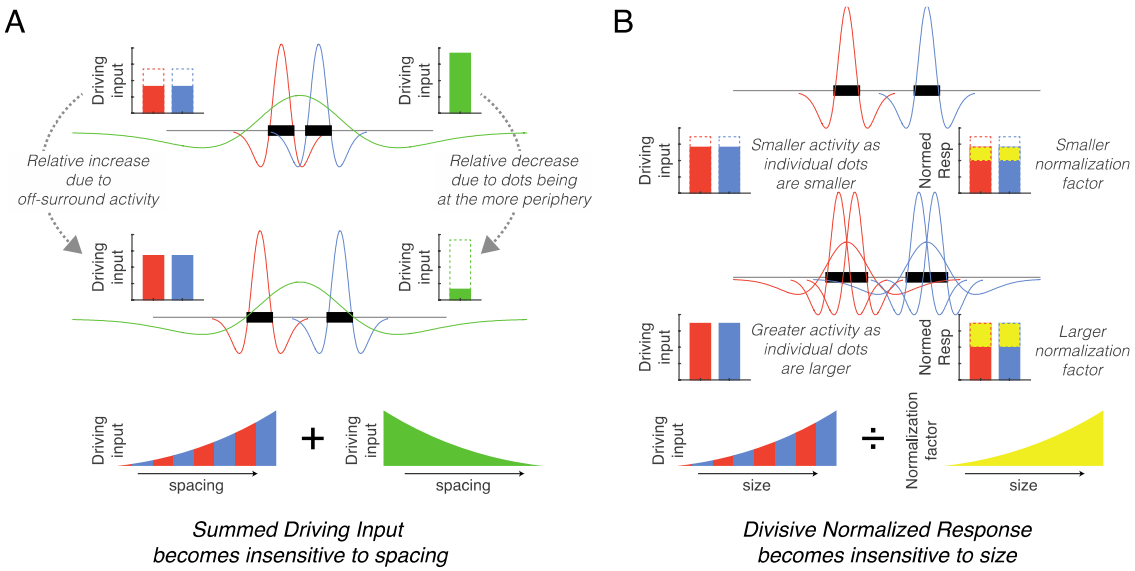

***Figure 3**. Simplified schematics explaining the mechanisms underlying the normalization of size and spacing. A. As spacing increases (from top to middle row) the response of small size center-surround filters increases (red and blue) whereas the response of large size center-surround filters decreases (green), with these effects counteracting each other in the total response. B. As dot size increases (from top to middle row), more filters are involved in responding to the dots thereby increasing the unnormalized response (red and blue), but this results in a greater overlap in the neighborhoods and increases the normalization factor (yellow). These counteracting effects eliminate the size effect.*

A key result from the current model is that the summed output of the neuronal activity is sensitive to numerosity but shows little variation with size and spacing. This pattern is consistent with electroencephalography studies finding similar results for the summed response of V1, V2, and V3 in the absence of any behavioral judgment [6,7]. However, this pattern is different than the behavior of prior deep neural network-based models of numerosity perception, which revealed many units in the deep layers that were sensitive to non-numerical dimensions, along with a few that were numerosity sensitive (or selective). One key difference might be that the current model implements divisive normalization directly in the center-surround responses whereas normalization in deep network models emerges from the connection weights between layers. In summary, our results suggest that (1) center-surround filters, (2) of varying filter sizes, with (3) divisive normalization, naturally reduce the effects of size and spacing while maintaining sensitivity to number without requiring additional layers of processing.

Of note, a primary feature of early visual cortex regions such as V1 and V2 is oriented line filters (i.e., simple cells). The fact that center-surround DoG filters in a feedforward network were sufficient for explaining numerosity perception suggests that neural activity at even earlier stages of the visual pathway, such as the retina or the lateral geniculate nuclei, may be capable of representing numerosity. Consistent with this idea, recent studies show evidence for numerosity processing in the subcortical structure following the monocular visual pathway [33] and in the nervous structure that gives rise to the modulation of the pupillary response [34].

The success of this model does not necessarily imply that neuronal responses in early visual regions directly determine behavioral responses [7]. Prior to behavior, there are many downstream processing steps that incorporate other sources of information, such response bias and decisional uncertainty. Instead, these results, together with previous electrophysiology results, suggest that normalized response magnitude in early visual regions may be the basic currency from which numerosity judgments are made. Future work should explore the link between the neuronal response layer in the current model and various behavioral judgments. This would be a particularly interesting avenue to pursue as task instructions are known to bias decisions on magnitude judgment [35,36], meaning that a top-down attentional system plays a heavy role in that link. Moreover, divisive normalization in higher cortical levels may be crucial for a more complex setting. For instance, while the current normalization model with DoG filters successfully explained visual illusions caused by regularity, grouping, connectedness, and heterogeneity, other numerosity phenomena such as statistical color pairing and topological invariants [28,37] may require the action of neurons with receptive fields that are more complex than center-surround filters. Future work could extend application of divisive normalization to these more complex, further downstream visual representations.

In conclusion, our results indicate that divisive normalization in a convolutional layer with hierarchically organized center-surround filters naturally enhances sensitivity to the discrete number of items in a visual scene by reducing the effects of size and spacing, consistent with recent empirical studies demonstrating direct and rapid encoding of numerosity [4]. Divisive normalization also explains various well-known numerosity illusions across space and time. These results identify the key neurocomputational principles underlying the ubiquity of the number sense in the animal kingdom.

Methods

Stimulus sets

Dot arrays spanning across number, size, and spacing.

Inputs to the neural network were visual stimuli of white dot arrays on a black background (200 × 200 pixels). Dots were homogeneous in size within an array and were drawn within an invisible circular field. Any two dots in an array were at least a diameter apart from edge to edge. The number of dots in an array is referred to as n, the radius of each dot is referred to as r_d, and the radius of the invisible circular field is referred to as r_f. **Table 1** provides mathematical definitions of other non-numerical dimensions based on these terms.

Following the previously developed framework for systematic dot array construction [4,38], stimulus parameters of the dot arrays were distributed systematically within a parameter space defined by three orthogonal dimensions: log-scaled dimensions of number (*N*), size (*Sz*), and spacing (*Sp*) (**Fig. 1A**). *N* simply represents the number of dots. *Sz* is defined as the dimension that varies with individual area (*IA*) while holding *N* constant, hence simultaneously varying in total area (*TA*). *Sp* is defined as the dimension that varies with sparsity (*Spar*) while holding *N* constant, hence simultaneously varying in field area (*FA*). Log-scaling these dimensions allows *N*, *Sz*, and *Sp* to be orthogonal to each other and represent all of the non-numerical dimensions of interest to be represented as a linear combination of those three dimensions (see **Table 1**). Thus, this stimulus construction framework makes is easy to visualize the stimulus parameters and analyze choice behavior or neural data using a linear statistical model. For an implementation of this framework, see the MATLAB code published in the following public repository: https://osf.io/s7xer/.

Table 1. Mathematical relationship between various magnitude dimensions.

Dimension	As a function of n, r_d, r_f	As a function of N, Sz, Sp
Individual area (IA)	πr_d^2	log(IA) = ½log(Sz) – ½log(N)
Total area (TA)	$n \times \pi r_d^2$	log(TA) = ½log(Sz) + ½log(N)
Field area (FA)	πr_f^2	log(FA) = ½log(Sp) + ½log(N)
Sparsity (Spar)	$\pi r_f^2/n$	log(Spar) = ½log(Sp) – ½log(N)
Individual perimeter (IP)	$2\pi r_d$	log(IP) = log(2√π) + ¼log(Sz) – ¼log(N)
Total perimeter (TP)	$n \times 2\pi r_d$	log(TP) = log(2√π) + ¼log(Sz) + ¾log(N)
Coverage (Cov)	$n \times r_d^2 / r_f^2$	log(Cov) = ½log(Sz) – ½log(Sp)
Closeness (Close)	$\pi^2 \times r_d^2 \times r_f^2$	log(Close) = ½log(Sz) + ½log(Sp)

Note: n = number; r_d = radius of individual dot; r_f = radius of the invisible circular field in which the dots are drawn.

Across all the dot arrays, number (*n*) ranged between 5 to 20 dots, dot diameter ($2 \times r_d$) ranged between 9 to 18 pixels, field radius (r_f) ranged between 45 to 90 pixels, all having five levels in logarithmic scale. This approach resulted in 35 unique points in the three-dimensional parameter space (see **Fig. 1B**). For each of the 35 unique points, a total of 100 dot arrays were randomly constructed for the simulation conducted in this study.

Dot arrays for testing regularity effects.

The 'regular' dot array was constructed following the previous study that first demonstrated the regularity effect [25]. This array contained 37 dots with $r_d = 3$ pixels, one of which at the center of the image and the rest distributed in three concentric circles with the radii of 20, 40, and 60 pixels. The 'irregular' arrays were constructed with the same number of and same sized dots randomly placed with $r_f = 72.5$ pixels. This radius for the field area was empirically calculated so that the convex hull of the regular array and the mean convex hull of the irregular arrays were matched. Sixteen irregular arrays were used in the simulation.

Dot arrays for testing grouping effects.

One set of 'ungrouped' dot arrays and another set of 'grouped' dot arrays were constructed. Both ungrouped and grouped arrays contained 12 dots, each of which with $r_d = 4.5$ pixels. However, in

the ungrouped arrays the dots were randomly dispersed, while in the grouped arrays the dots were spatially grouped in pairs. The edge-to-edge distance between the two dots in each pair was approximately equal to r_d. A large number of unique dot arrays were constructed using these criteria for each of the two sets. Then, a subset of unique arrays from each set was chosen so that the convex hull of the arrays between the two sets were numerically matched. A total of 16 grouped and 16 ungrouped arrays entered the simulation.

Dot arrays for testing connectedness effects.

One set of 'connected' dot arrays and another set of 'unconnected' dot arrays were constructed. First, a large number of dot arrays with n = 10, r_d = 6.5 pixels, and r_f = 64 pixels were created. Then, connected dot arrays were constructed by connecting the centers of two closest dots with a thin white line that was 2 pixels in width. The resulting images were visually checked, and all the images in which the lines cross or touch other lines or dots were removed from the set. Then, unconnected dot arrays were constructed from those connected dot arrays by breaking the midpoints of the interconnecting lines and rotating those broken lines about the center of each dot by ±30 degrees in either direction randomly determined. A total of 16 connected and 16 unconnected arrays entered the simulation.

Dot arrays for testing heterogeneity effects.

Three sets of dot arrays equated in the total area (*TA*) were created. The first set of 'homogeneous' (or zero level of heterogeneity) dot arrays contained n = 15 with r_d = 5 pixels within a circular field defined by r_f = 75 pixels. The second set of 'less heterogeneous' dot arrays contained six dots with r_d = 3 pixels, six dots with r_d = 5 pixels, and three dots with r_d = 7.5 pixels. The last set of 'more heterogeneous' dot arrays contained twelve dots with r_d = 2.5 pixels and three dots with r_d = 10 pixels. Hence, the total area (*TA*) of all the arrays were approximately identical to each other while the variability of individual area (*IA*) differed across the sets. Rounding errors due to pixelation and anti-aliasing, however, caused differences the actual cumulative intensity measure of the bitmap images. On average, the cumulative intensity values (0 being black and 1 being white in the bitmap image) were comparable between the three sets of arrays: 1209 in the homogeneous arrays, 1194 in the less heterogeneous arrays, and 1204 in the more heterogeneous arrays. Sixteen arrays in each of the three sets entered the simulation.

Neural network model with divisive normalization

Convolution with DOG filters.

The model consisted of a convolutional layer with difference-of-Gaussians (DoG) filters of six different sizes, that convolved input values of the aforementioned bitmap images displaying dot arrays. This architecture hence provided a structure for 200 × 200 × 6 network units (or simulated neurons) activated by images of dot arrays (**Fig. 2**). The DoG filters are formally defined as:

$$\Gamma(x,y) = I \cdot \left(\frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} - \frac{1}{2\pi K^2\sigma^2} e^{-\frac{x^2+y^2}{2K^2\sigma^2}} \right), \qquad (1)$$

Where I is the input image, σ^2 is the spatial variance of the narrower Gaussian, and K is the scaling factor between the two variances. As recommended by Marr and Hildreth [39], $K = 1.6$ was used to achieve balanced bandwidth and sensitivity of the filters. Considering that the input values range [0 1], the DoG filters were reweighted so that the sum of the positive portion equals to 1 and the sum of the negative portion equals to -1, making the summation across all domains 0. This reweighting ensured that the response is maximized when the input matches the DoG filter regardless of filter size and that the filter produces a response of value 0 if the input is constant across a region regardless of filter size. Finally, the output of this convolution process was followed by half-wave rectification at each simulated neuron [40], where negative responses were replaced by zero. This stipulation sets the 'firing threshold' of the network such that the simulated neurons would not fire if the input does not match its DoG filter.

Six different σ values were used (σ_k = 1, 2, 4, 8, 16, 32 for filter size k, respectively) which together were sensitive enough to represent various visual features of the input images, from the edge of the smallest dots to the overall landscape of the entire array. The activity of each stimulated neuron, i, in filter size k following this convolution procedure is referred to as $D_{i,k}$.

Divisive normalization.

Following Carandini & Heeger [22], the normalization model was defined as:

$$R_{i,k} = \frac{D_{i,k}^{\gamma}}{c+\sum_{j,k}\eta_{(i,j,k)}D_{j,k}^{\gamma}}, \qquad (2)$$

where distance similarity $\eta_{(i,j)}$ is defined as:

$$\eta_{(i,j,k)} = e^{-\frac{d(i,j)}{r_k}}. \qquad (3)$$

D_i is the driving input of neuron i (i.e., the output of the convolution procedure described above), $d_{(i,j)}$ is the Euclidean distance between neuron i and neuron j in any filter size, c is a constant that prevents division by zero. The denominator minus this constant, which was set to 1, is referred to as the normalization factor. The parameter r_k, defined for each filter size, serves to scale between local and global normalization. As r_k gets larger, activities from broader set of neurons constitute the normalization factor. In our model, r_k was defined as a scaling factor of σ_k (e.g., $r_k = \sigma_k$, $r_k = 2\sigma_k$, or $r_k = 4\sigma_k$), so that neurons with larger filter sizes have their normalization factor computed from broader pool of neighboring neurons. The parameter γ determines the degree of amplification of individual inputs and serves to scale between winner take all and linear normalization. $R_{i,k}$ represents the normalized response of neuron i in filter size k.

Modelling temporal modulation of network units.

Normalized responses of simulated neurons were further modeled to capture temporal modulations, with another normalization process this time working across time. First, a read out neuron was assumed that summed up the normalized responses across all the neurons, $\Sigma R_{i,k}$. This single firing activity, now referred to as M, underwent the following temporal normalization process that resulted in the normalized activity M^*:

$$M_T^* = \frac{M_T^\delta}{c+\sum_{t=1}^{T}\eta_t M_t^\delta}. \qquad (4)$$

The temporal distance η is defined as:

$$\eta_t = e^{-\frac{d}{\omega}}, \qquad (5)$$

where d is the distance between time point t and T. As in Eq. 2 and 3, c is a constant that prevents division by zero, which was set to 1 for convenience. The parameter ω determines the amount of recent history contributing to the normalization factor, and the parameter δ determines the degree of amplification of M_t.

The MATLAB code used to implement the model can be found in the following public repository: https://osf.io/4rwjs/ (will be made public upon acceptance).

Acknowledgements

We thank Dr. Michele Fornaciai for inspiring discussions. This work was supported by the National Science Foundation CAREER Award BCS 1654089 to J.P.

Author contributions

Joonkoo Park: Conceptualization, Methodology, Software, Formal analysis, Writing. David E. Huber: Methodology, Formal analysis, Writing.

Declaration of interests

The authors declare no competing interests.

References

1. Dehaene, S. *The number sense : how the mind creates mathematics*. (Oxford University Press, 2011).

2. Feigenson, L., Dehaene, S. & Spelke, E. Core systems of number. *Trends Cogn. Sci.* **8**, 307–314 (2004).

3. Nieder, A. The neuronal code for number. *Nat. Rev. Neurosci.* **17**, 366–382 (2016).

4. Park, J., DeWind, N. K., Woldorff, M. G. & Brannon, E. M. Rapid and Direct Encoding of Numerosity in the Visual Stream. *Cereb. Cortex* **26**, 748–763 (2016).

5. Fornaciai, M. & Park, J. Disentangling feedforward versus feedback processing in numerosity representation. *Cortex* **135**, 255–267 (2021).

6. Fornaciai, M., Brannon, E. M., Woldorff, M. G. & Park, J. Numerosity processing in early visual cortex. *Neuroimage* **157**, 429–438 (2017).

7. Fornaciai, M. & Park, J. Early numerosity encoding in visual cortex is not sufficient for the representation of numerical magnitude. *J. Cogn. Neurosci.* **30**, 1788–1802 (2018).

8. Gebuis, T., Cohen Kadosh, R. & Gevers, W. Sensory-integration system rather than approximate number system underlies numerosity processing: A critical review. *Acta Psychol. (Amst).* **171**, 17–35 (2016).

9. Leibovich, T., Katzin, N., Harel, M. & Henik, A. From “sense of number” to “sense of magnitude”: The role of continuous magnitudes in numerical cognition. *Behav. Brain Sci.* **40**, e164 (2017).

10. Park, J. A neural basis for the visual sense of number and its development: A steady-state visual evoked potential study in children and adults. *Dev. Cogn. Neurosci.* **30**, 333–343 (2018).

11. Van Rinsveld, A. *et al.* The neural signature of numerosity by separating numerical and continuous magnitude extraction in visual cortex with frequency-tagged EEG. *Proc. Natl. Acad. Sci.* **117**, 5726–5732 (2020).

12. DeWind, N. K., Park, J., Woldorff, M. G. & Brannon, E. M. Numerical encoding in early visual cortex. *Cortex* **114**, 76–89 (2019).

13. Dehaene, S. & Changeux, J.-P. Development of Elementary Numerical Abilities: A Neuronal Model. *J. Cogn. Neurosci.* **5**, 390–407 (1993).

14. Park, J., Godbole, S., Woldorff, M. G. & Brannon, E. M. Context-Dependent Modulation of Early Visual Cortical Responses to Numerical and Nonnumerical Magnitudes. *J. Cogn. Neurosci.* **33**, 2536–2547 (2021).

15. Nasr, K., Viswanathan, P. & Nieder, A. Number detectors spontaneously emerge in a deep neural network designed for visual object recognition. *Sci. Adv.* **5**, (2019).

16. Creatore, C., Sabathiel, S. & Solstad, T. Learning exact enumeration and approximate estimation in deep neural network models. *Cognition* **215**, 104815 (2021).

17. Testolin, A., Zou, W. Y. & McClelland, J. L. Numerosity discrimination in deep neural networks: Initial competence, developmental refinement and experience statistics. *Dev. Sci.* **23**, (2020).

18. Stoianov, I. & Zorzi, M. Emergence of a 'visual number sense' in hierarchical generative models. *Nat. Neurosci.* **15**, 194–6 (2012).

19. Kim, G., Jang, J., Baek, S., Song, M. & Paik, S.-B. Visual number sense in untrained deep neural networks. *Sci. Adv.* **7**, (2021).

20. Bell, A. J. & Sejnowski, T. J. The "independent components" of natural scenes are edge filters. *Vision Res.* **37**, 3327–3338 (1997).

21. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. in *Advances in Neural Information Processing Systems* 1106–1114 (2012).

22. Carandini, M. & Heeger, D. J. Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**, 51–62 (2012).

23. Huber, D. E. & O'Reilly, R. C. Persistence and accommodation in short-term priming and other perceptual paradigms: temporal segregation through synaptic depression. *Cogn. Sci.* **27**, 403–430 (2003).

24. Burr, D. & Ross, J. A Visual Sense of Number. *Curr. Biol.* **18**, 425–428 (2008).

25. Ginsburg, N. Effect of Item Arrangement on Perceived Numerosity: Randomness vs Regularity. *Percept. Mot. Skills* **43**, 663–668 (1976).

26. Frith, C. D. & Frit, U. The solitaire illusion: An illusion of numerosity. *Percept. Psychophys.* **11**, 409–410 (1972).

27. Franconeri, S. L., Bemis, D. K. & Alvarez, G. A. Number estimation relies on a set of segmented objects. *Cognition* **113**, 1–13 (2009).

28. He, L., Zhou, K., Zhou, T., He, S. & Chen, L. Topology-defined units in numerosity perception. *Proc. Natl. Acad. Sci.* **112**, E5647–E5655 (2015).

29. van Oeffelen, M. P. & Vos, P. G. Configurational effects on the enumeration of dots: Counting by groups. *Mem. Cognit.* **10**, 396–404 (1982).

30. Allik, J. & Tuulmets, T. Occupancy model of perceived numerosity. *Percept. Psychophys.* **49**, 303–14 (1991).

31. Lee, H., Baek, J. & Chong, S. C. Perceived magnitude of visual displays: Area, numerosity, and mean size. *J. Vis.* **16**, 12 (2016).

32. Aagten-Murphy, D. & Burr, D. Adaptation to numerosity requires only brief exposures, and is determined by number of events, not exposure duration. *J. Vis.* **16**, 22 (2016).

33. Collins, E., Park, J. & Behrmann, M. Numerosity representation is encoded in human subcortex. *Proc. Natl. Acad. Sci.* **114**, E2806--E2815 (2017).

34. Castaldi, E., Pomè, A., Cicchini, G. M., Burr, D. & Binda, P. The pupil responds spontaneously to perceived numerosity. *Nat. Commun.* **12**, 5944 (2021).

35. Cicchini, G. M., Anobile, G. & Burr, D. C. Spontaneous perception of numerosity in humans. *Nat. Commun.* **7**, 12536 (2016).

36. Castaldi, E., Piazza, M., Dehaene, S., Vignaud, A. & Eger, E. Attentional amplification of neural codes for number independent of other quantities along the dorsal visual stream. *Elife* **8**, (2019).

37. Zhao, J. & Yu, R. Q. Statistical regularities reduce perceived numerosity. *Cognition* **146**, 217–222 (2016).

38. DeWind, N. K., Adams, G. K., Platt, M. L. & Brannon, E. M. Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition* **142**, 247–265 (2015).

39. Marr, D. & Hildreth, E. Theory of edge detection. *Proc. R. Soc. London. Ser. B. Biol. Sci.* **207**, 187–217 (1980).

40. Heeger, D. J. Nonlinear model of neural responses in cat visual cortex. *Comput. Model. Vis. Process.* 119–133 (1991).

Supplementary Figures

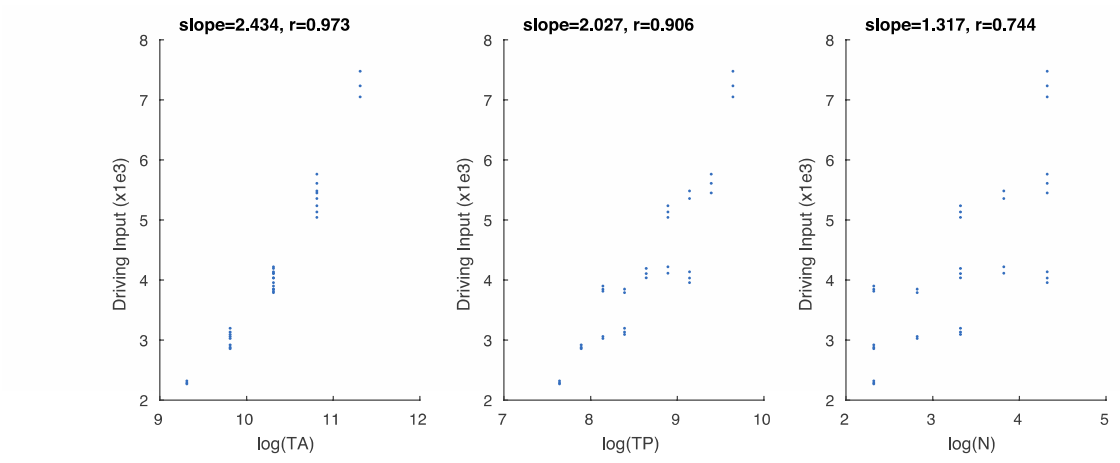

Figure S1. Correlation between summed driving input, Σ*D*, and log-scaled total area (TA), total perimeter (TP), and number (N).

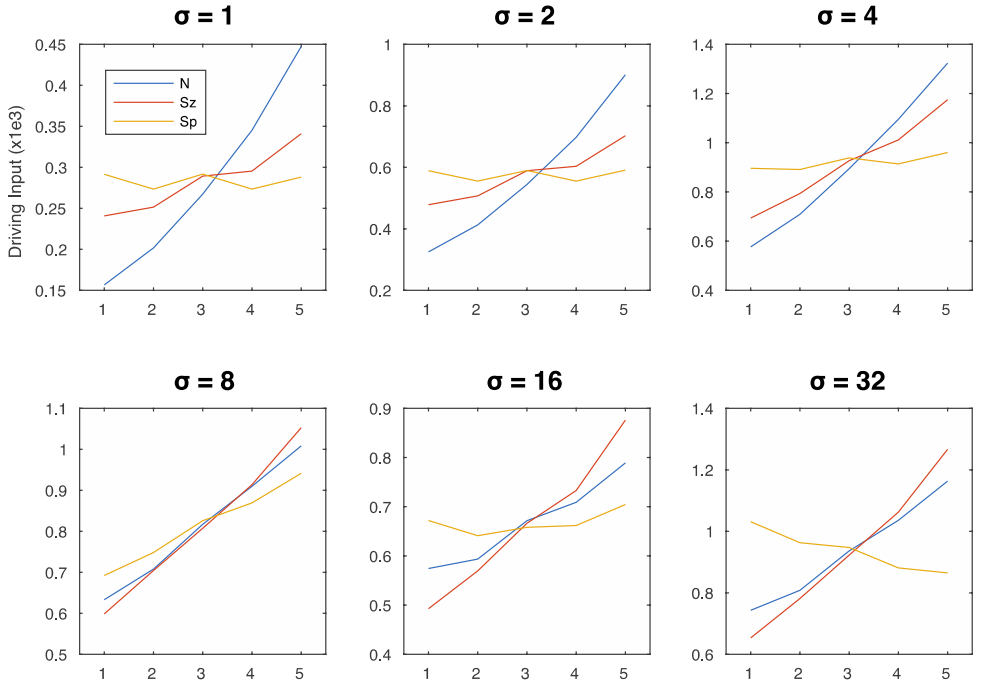

Figure S2. Summed driving input across all the simulated neurons separately in each of the six filters, plotted as a function of number (*N*), size (*Sz*), and spacing (*Sp*). The effect of spacing existed in the fourth and sixth largest filter sizes, indicating effects closely aligning with field

area and density, respectively (see **Table 1**); however, the effects in these two filter sizes were in opposite directions, which made the overall effect very small.

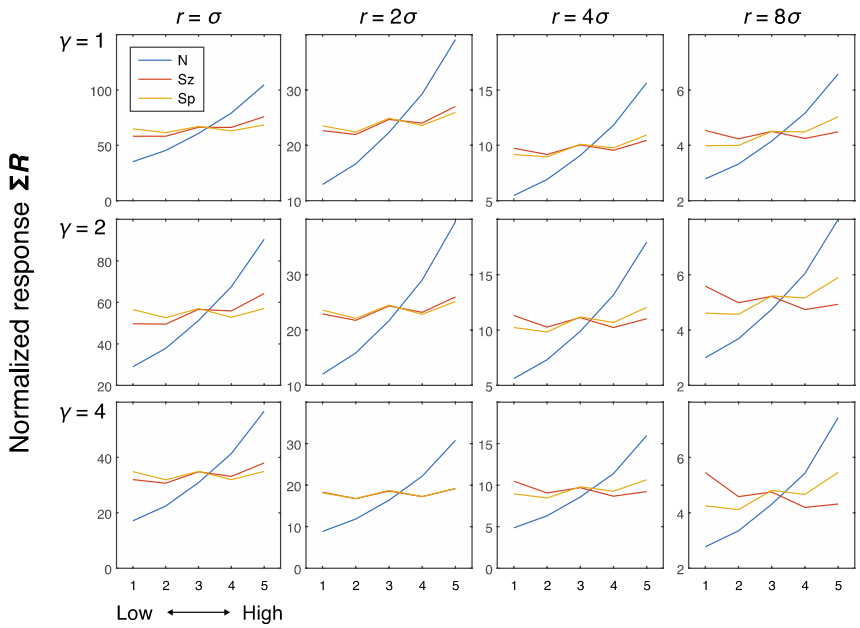

Figure S3. Simulation results showing the effects of number (N), size (Sz), and spacing (Sp) on the normalized response of the network units as a function of neighborhood size (r) and amplification factor (γ). Greater r resulted in a flatter curve for the size effect, and this flattening became more pronounced as γ increased, with the combination of high values for both parameters producing a modest negative effect of size as well as a modest positive effect of spacing. More specifically, the combination of high r and γ values produces a winner-take-all process across large regions of the display. Greater size, in these cases, thus leads to greater normalization factor (denominator) which results in reduced normalization activity, although the extent of this normalization depends on how far away the other dots are located (e.g., less normalization with spacing). Although this is an interesting phenomenon, empirical neural and behavioral studies show a positive effect of size, if any. Hence, larger values of r and γ in this model do not seem to be plausible in the case of numerosity perception. Therefore, we chose moderate values of r (=2) and γ (=2) for subsequent simulations.

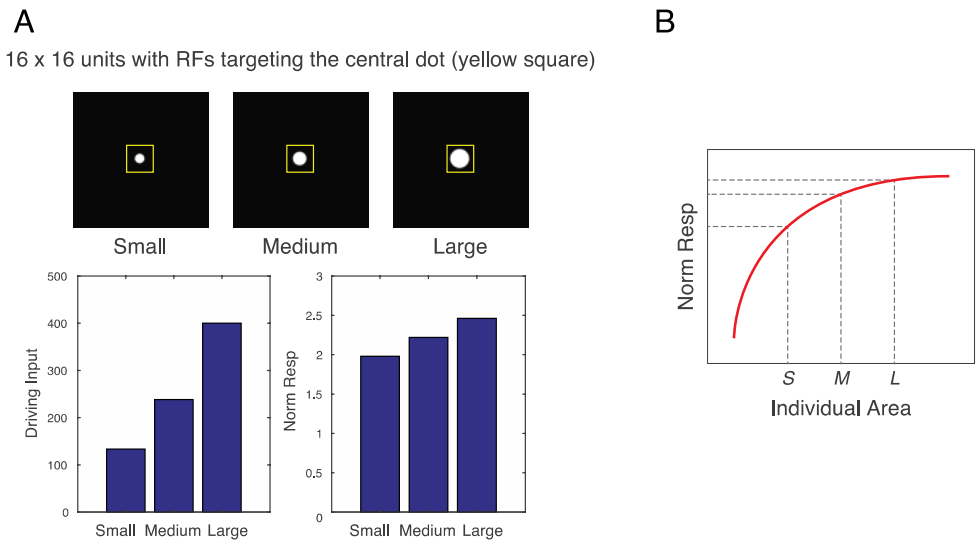

Figure S4. Effects of single dots. A. Images of small (radius = 3.5), medium (radius = 5), and large (radius = 7) singly presented dots were fed into the computational model, and the driving input and the normalized response of the units with the receptive fields (RF) targeting the dots were computed. As expected, driving input was nearly perfectly correlated with area of the dots (r = 0.9983). In contrast, normalized response showed a linear relationship with the radius, which meant a logarithmic relationship with area. This is exactly what the normalization model would predict because a larger dot requires more filters to cover the dot, but more filters increase the normalization factor (denominator) at a faster rate than how the driving input becomes summed up (numerator). In other words, the normalized response becomes tempered in a non-linear way, producing a saturating normalized response as a function of increasing dot area. B. Schematic illustration of the saturating effect of normalized response for a single dot (within a hypothetical dot array) as a function of the area of the dot. Heterogeneous arrays are created by holding the total area and numerosity constant while changing individual dot size. Therefore, when medium-sized dots (M) are replaced with large dots (L), the same number of replacements must be done to go from medium-size dots (M) to small dots (S). However, because of the saturating effect, there is a greater decrease in normalized response than an increase in normalized response. Thus, the overall normalized response becomes necessarily smaller in a heterogeneous array compared to a homogeneous array.