A visual sense of number emerges from divisive normalization in a simple center-surround convolutional network

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 substantially less so to continuous dimensions orthogonal to numerosity, such as size and spacing of the dots. However, the mechanisms that extract numerosity are unknown. 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. Moreover, a model implementation of these principles explained both well-known and relatively novel illusions of numerosity perception across space and time. This supports the conclusion that the neural structures and feedforward processes that encode numerosity naturally produce visual illusions of numerosity. Together, these results identify a set of neurocomputational properties that gives rise to the ubiquity of the number sense in the animal kingdom.

Keywords

numerosity perception; computational modeling; divisive normalization

Introduction

Humans have an intuitive sense of number that allows numerosity estimation without counting (Dehaene, 2011). The prevalence of number sense across phylogeny and ontogeny (Feigenson et al., 2004) suggests common neural mechanisms that allow the extraction of numerosity information from a visual scene. While earlier empirical work highlighted the parietal cortex for numerosity representation (Nieder, 2016), 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 number (abbreviated as *N*) of a dot array in the absence of any behavioral response, but much less so to non-numerical dimensions that are orthogonal to number (i.e., size and spacing, abbreviated as *Sz* and *Sp*, respectively; see **Fig. 1A**) (Park et al., 2016). Subsequent behavioral and neural studies showed that this early cortical sensitivity to numerosity indicates feedforward activity in visual areas V1, V2 and V3 (Fornaciai et al., 2017; Fornaciai and Park, 2021, 2018). 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 (Dakin et al., 2011; Gebuis et al., 2016; Leibovich et al., 2017). In contradiction, however, recent empirical studies 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 (DeWind et al., 2019; Park, 2018; Paul et al., 2022; Van Rinsveld et al., 2020).

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 computational model of Dehaene and Changeux (1993) explains numerosity detection based on several neurocomputational principles. That model (hereafter D&C) assumes a one-dimensional linear retina (each dot is a line segment), and responses are normalized across dot size via a convolution layer that represents combinations of two attributes: 1) dot size, as captured by difference-of-Gaussian contrast filters of different widths; and 2) location, by centering filters at different positions. In the convolution layer, the filter that matches the size of each dot dominates the neuronal activity at the location of the dot owing to a winner-take-all lateral inhibition process. To indicate numerosity, a summation layer pools the total activity over all the units in the convolution layer. While the D&C model provided a proof of concept for numerosity detection, it has several limitations as outlined in the discussion. Of these, the most notable is that strong winner-take-all in the convolution layer discretizes visual information (e.g., discrete locations and discrete sizes yielding a literal count of dots), which is implausible for early vision. As a result, the output of the model is completely insensitive to anything other than number in all situations, which is inconsistent with empirical data (Park et al., 2021).

Recently, several deep-network-based models have been applied to numerosity perception (Creatore et al., 2021; Kim et al., 2021; Nasr et al., 2019; Stoianov and Zorzi, 2012; Testolin et al., 2020). Stoianov and Zorzi (2012) 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 inhibitory connections to units in the first layer that encoded cumulative area. However, the development of center-surround detectors based on unsupervised learning is a common observation (Bell and Sejnowski, 1997), 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 (Kim et al., 2021) found that sensitivity and selectivity to numerosity was well captured in a completely untrained convolutional neural network (AlexNet) (Krizhevsky et al., 2012), 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, these are "black box" models, and it is not always clear *how* these models work; these models contain many mechanisms, and it is not clear which mechanisms are crucial for producing numerosity-sensitive units.

Rather than applying a complex multilayer learning model, we distill the neurocomputational principles that enable the visual system to be sensitive to numerosity while remaining relatively insensitive to non-numerical visual features. These principles are simulated in a single layer model that does not need to be trained. Consistent with prior work, we hypothesize that center-surround contrast filters at different spatial scales play an important role in numerosity perception. In addition to this "convolution" of the input, most prior proposals entail some form of pooling or normalization (e.g., normalization between center-surround units). This can emerge across layers of visual processing, as often assumed in "max pooling" layers of a convolutional neural network (Scherer et al., 2010), or it can occur within a layer, as in the strong winner-take-all lateral inhibition used in the Dehaene and Changeux (1993) model. Furthermore, some models contain both within layer normalization and between layer max pooling (Krizhevsky et al., 2012). Although the functional form of within-layer normalization is similar to between-layer max pooling, it differs anatomically, placing the normalized response earlier in visual processing. In determining the neural mechanisms that are core to numerosity, we note that a moderate level of within-layer normalization is consistent with "divisive normalization" (Carandini and Heeger, 2012), in which the response of each neuron reflects its driving input divided by the summation of responses from anatomically surrounding neurons (i.e., a normalization pool). This normalization is not as extreme as winner-take-all normalization and tends to preserve visual precision through graded activation responses. In the case of early vision, the normalization pool is spatially determined by retinotopic positions. Divisive normalization is known to exist throughout the cortex, reflecting the shunting inhibition of inhibitory interneurons that limit neural activation within a patch of cortex (Carandini and Heeger, 2012). 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 perception would be remiss not to include the effect of within-layer divisive normalization.

To determine the contribution of divisive normalization to numerosity encoding, we implemented an untrained neural network with versus without divisive normalization as applied to center-surround filters at different spatial scales (e.g., as in V1) (**Fig. 1B**). The output simulates the summation of synchronized postsynaptic activity of a large population of neurons at a pre-decisional stage, consistent with previous work (Fornaciai et al., 2017; Park et al., 2016). Our results show that (1) hierarchically organized multiple center-surround filters of varying size make the network insensitive to spacing and that (2) divisive normalization implemented across network units makes the network additionally insensitive to size. Divisive normalization not only occurs over space but also over time (Huber and O'Reilly, 2003). Thus, we additionally implemented temporal divisive normalization to test if it explains contextual effects of numerosity perception (Burr and Ross, 2008; Park et al., 2021).

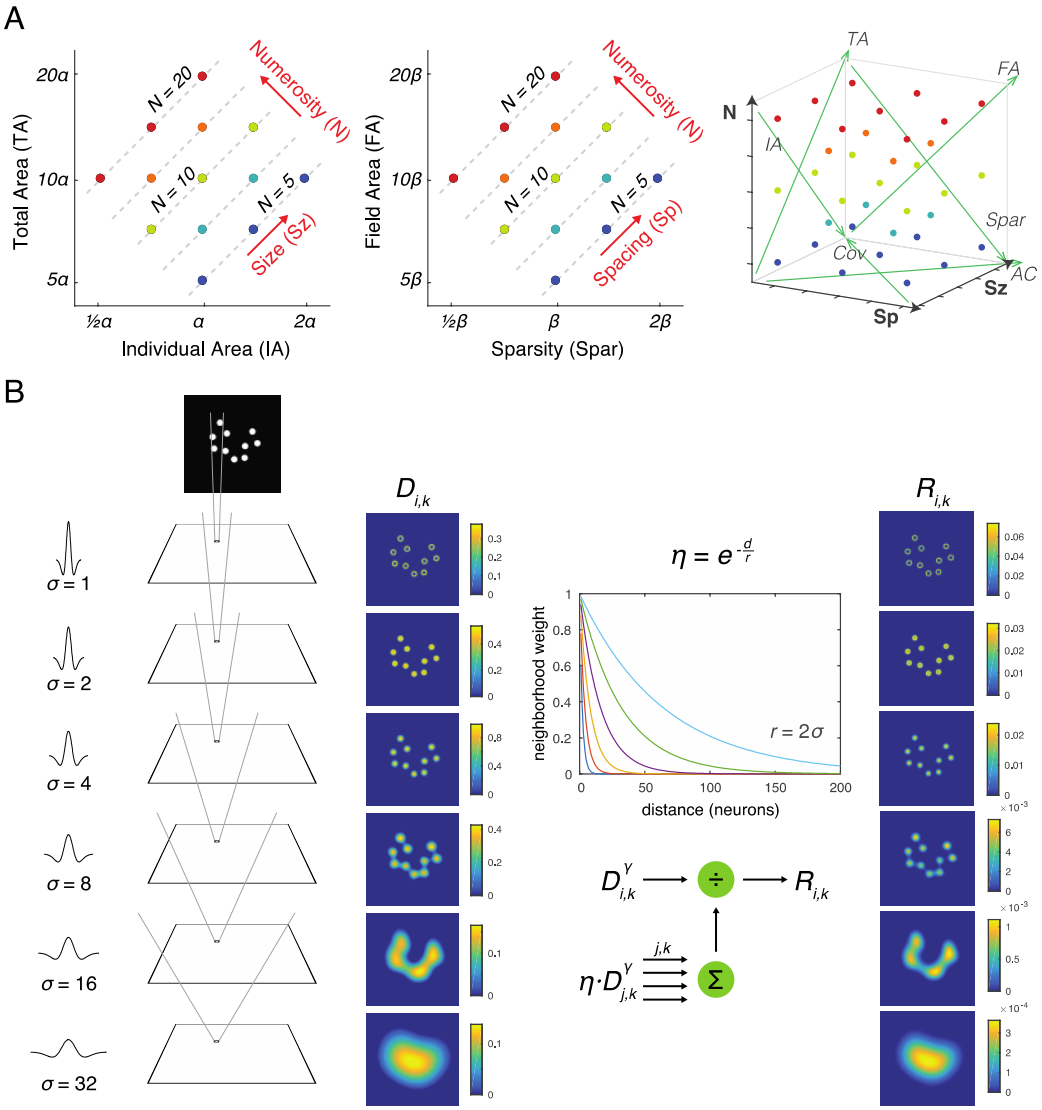

Figure 1. *Stimulus design and computational methods. A. Properties of magnitude dimensions represented in three orthogonal axes defined by log-scaled number (N), size (Sz), and spacing (Sp) (Table 1). B. Schematic illustration of the computational process from a dot-array image to the driving input (i.e., the model without divisive normalization), D, of the simulated neurons, versus the normalized response (i.e., the model with divisive normalization), 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.*

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**). The summed driving input in each filter size showed different effects as a function of number, size, and spacing (**Fig. 2A**), but when the driving input was summed across all filter sizes it was most strongly modulated by both number and size equally but not by spacing (**Fig. 2B**), suggesting that the neural activity tracks total area (*TA*; see **Table 1**; **Fig. 2–figure supplement 1**). The effect of spacing existed in the fourth and sixth largest filter sizes, largely indicating effects of field area and density, respectively (**Fig. 2A**); 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.

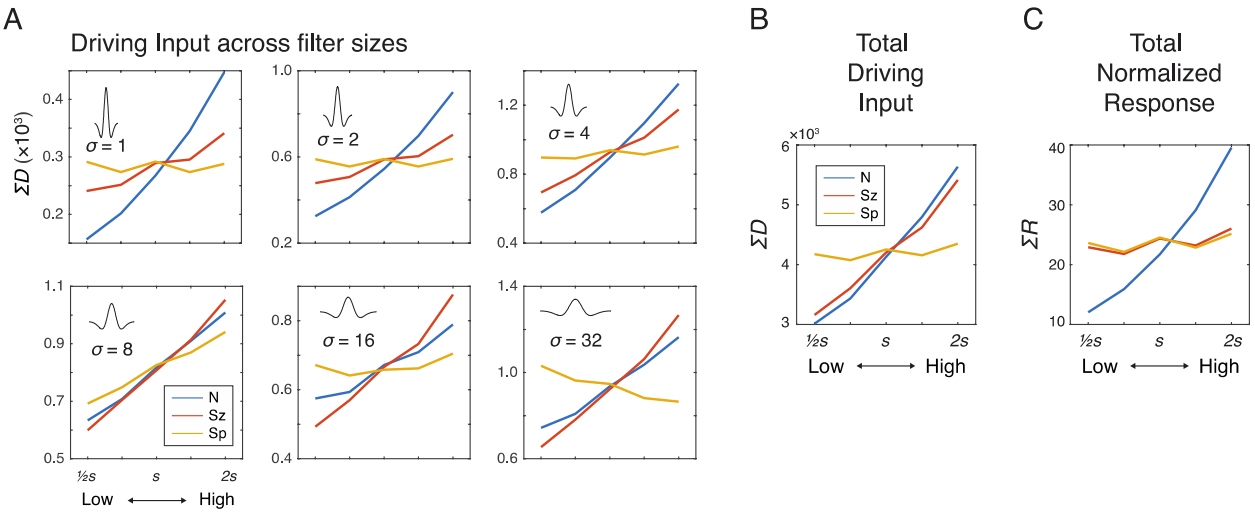

Figure 2. Simulation results showing the effects of number (N), size (Sz), and spacing (Sp) on the driving input and normalized response of the network units. A. Summed driving input (ΣD) separately for each of the six filter sizes as a function of N, Sz, and Sp (see Methods for the specific values of s). B. ΣD across all filters is modulated by both number and size but not by spacing. C. Summed normalized response (ΣR) showed a near elimination of the Sz effect leaving only the effect of N. The results were simulated using r = 2σ and γ = 2, but effects of Sz and Sp were negligible across all the tested model parameters (Fig. 2–figure supplement 2). The value s on the horizontal axis indicates a median value for each dimension (see Materials and Methods).

Divisive normalization nearly eliminates the effect of size

We next added divisive normalization to the center-surround model, 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, Σ*R*, were strongly modulated by number but much less so, if any, by size and spacing (**Fig. 2C**). The pattern of results was largely consistent across different parameter values for neighborhood size (r) and amplification factor (γ) of the normalization model (**Fig. 2–figure supplement 2**); therefore, we chose moderate values of r (=2) and γ (=2) for subsequent simulations. As one way to quantify these modulatory effects, a simple linear regression with Σ*R* as the dependent variable with mean centered values of *N* as the independent variable (as well as *Sz* and *Sp* in separate models) was performed. Then, the slope estimate was divided by the intercept estimate, so that these effects could be easily compared across different sets of images (see **Fig. 2–figure supplement 3**). This baseline-adjusted regression slope for *N*, *Sz*, and *Sp* was 0.5771, 0.0646, and 0.0321, respectively. A multiple 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

Next, we considered if the center-surround model with divisive normalization also explains some of the most well-known visual illusions of numerosity perception. If so, this would support the hypothesis that these visual illusions reflect early visual processing at the level of numerosity encoding, without requiring any downstream processing. In other words, early vision may be the root cause of *both* numerosity encoding and numerosity visual illusions.

Empirical studies have long shown that irregularly spaced arrays (compared with regularly spaced arrays) and arrays with spatially grouped items (compared with ungrouped items) are all underestimated (Frith and Frit, 1972; Ginsburg, 1976; van Oeffelen and Vos, 1982). These illusions were indeed captured by the inclusion of divisive normalization. Irregular arrays yielded a 5.98% reduction (Cohen's d = 4.23) and grouped arrays yielded a 2.99% reduction (d = 10.02) of normalized response (**Fig. 3A-B**). Note that, in the absence of divisive normalization, there was either no effect or an effect in the opposite direction (**Fig. 3–figure supplement 1**).

The underestimation effects in the normalized response 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" (Allik and Tuulmets, 1991), but our results demonstrate that these effects emerge naturally within 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 (Lee et al., 2016). Consistent with this behavioral phenomenon, our simulations revealed that greater heterogeneity leads to greater underestimation (**Fig. 3C**). As compared to the homogeneous array, a moderately heterogeneous array (labeled "less heterogenous") yielded a 1.14% reduction (d = 2.43) and the more heterogeneous array yielded a 5.87% reduction (d = 8.11) in the magnitude of the normalized response. This occurs because the summed normalized response of a single dot saturates as dot area increases (**Fig. 3–figure supplement 2**), which interacts with the heterogeneity of the dot array. As heterogeneity is manipulated by making some dots larger and other dots smaller while keeping total area and numerosity constant, this saturating effect makes the overall normalized response smaller as a greater number of dots deviates from the average size (the gains from making some dots larger is not as great as the losses from making some dots smaller). As in the case of other illusions, the same analysis in the absence of divisive normalization fails to produce this illusion (**Fig. 3–figure supplement 1**).

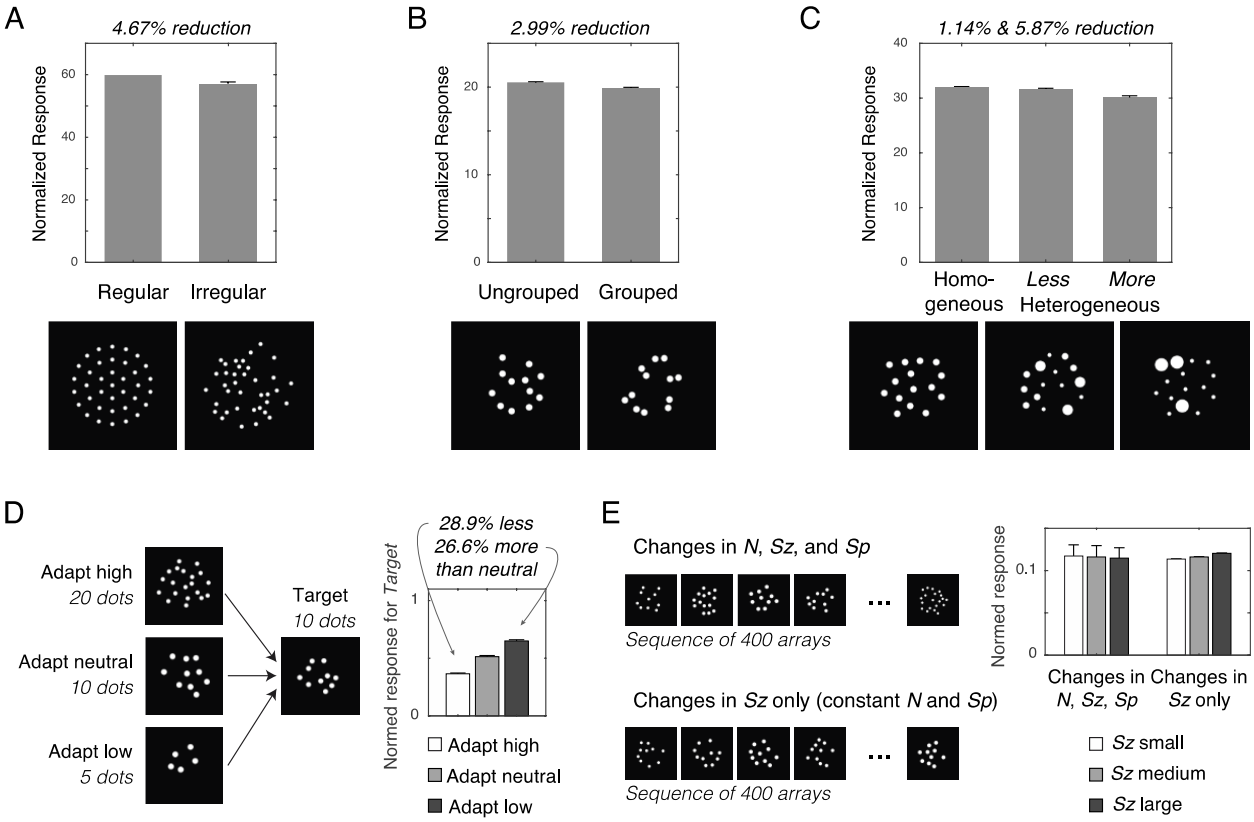

Figure 3. *Simulation of numerosity illusions. Normalized response of the network units influenced by the (A) regularity, (B) grouping, and (C) heterogeneity of dot arrays, as well as by (D) adaptation and (E) context. Error bars represent one standard deviation of the normalized response across simulations; however, the error bars in most cases were too small to be visualized. Spatial normalization effects (A, B, and C) were simulated with $r = 2$ and $\gamma = 2$. Temporal normalization effects (D and E) used these same parameters values in combination with $\omega = 8$ and $\delta = 1$.*

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 (Burr and Ross, 2008). We reasoned that numerosity adaptation might reflect divisive normalization across time, similar to adaptation with light or odor (Carandini and Heeger, 2012), which shifts the response curve and produces a contrast aftereffect. Closely related to temporal adaptation, the recently discovered temporal contextual effect of numerosity perception is an amplified neural response to changes in one dimension (e.g., changes in dot size) when observers experience a trial sequence with only changes in that dimension (Park et al., 2021). Therefore, we also applied the model with temporal normalization to the context effect.

We modeled temporal divisive normalization for a readout neuron that is 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 with the model parameters of $\omega = 8$ and $\delta = 1$ (**Fig. 3D**). Similar to behavioral results (Aagten-Murphy and Burr, 2016), the target of 10 dots was underestimated by 28.9% (d = 18.04) when the adaptor was more numerous than the target and was overestimated by 26.6% (d = 14.06) when the adaptor was less numerous than the target. This pattern held across all tested model parameters (**Fig. 3–figure supplement 3**). It is important to note that the model does not "know" the number of dots in the adaptor image. Instead, temporal divisive normalization compares the spatially normalized response of the current image to that of the adaptor image and because the spatially normalized response is primarily sensitive to variation in number, there is a contrast effect (e.g., "adapt high" reduces the response to the current image). Indeed, because the normalized response is less sensitive to variation in size or spacing, no adaptation effect emerges for those variables (**Fig. 3–figure supplement 4 and Fig. 3–figure supplement 5**). These results confirm that divisive normalization across space and time naturally produces numerosity adaptation.

Using the same model and parameters of temporal normalization (**Eq. 4**), we tested if it can 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 (Park et al., 2016), but that the effect of size becomes apparent when number and spacing are held constant while varying only size (Park et al., 2021). 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. 3E**). In the context where all dimensions varied, the three levels of *Sz* had no linear association with M^*; the 95th percentile confidence interval of the ordinary-least-square linear slope of M^* as a function of *Sz* was [-0.0243, 0.0182], which includes 0. In contrast, in the context where only size varied, M^* was positively correlated with *Sz*; slope confidence interval of [0.00315, 0.00359], which excludes 0. This pattern held across all tested model parameters (**Fig. 3–figure supplement 6**). 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 (Kim et al., 2021; Nasr et al., 2019; Stoianov and Zorzi, 2012), 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. 4A**). 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. 4B**). 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 whereas only smaller filters respond to small dots (**Fig. 2B**). However, normalization dampens this increase. Critically, divisive normalization is a within-layer effect, reflecting recurrent inhibition between center-surround filters owing to inhibitory interneurons. Thus, the effect of dot size is eliminated in early visual responses. 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.

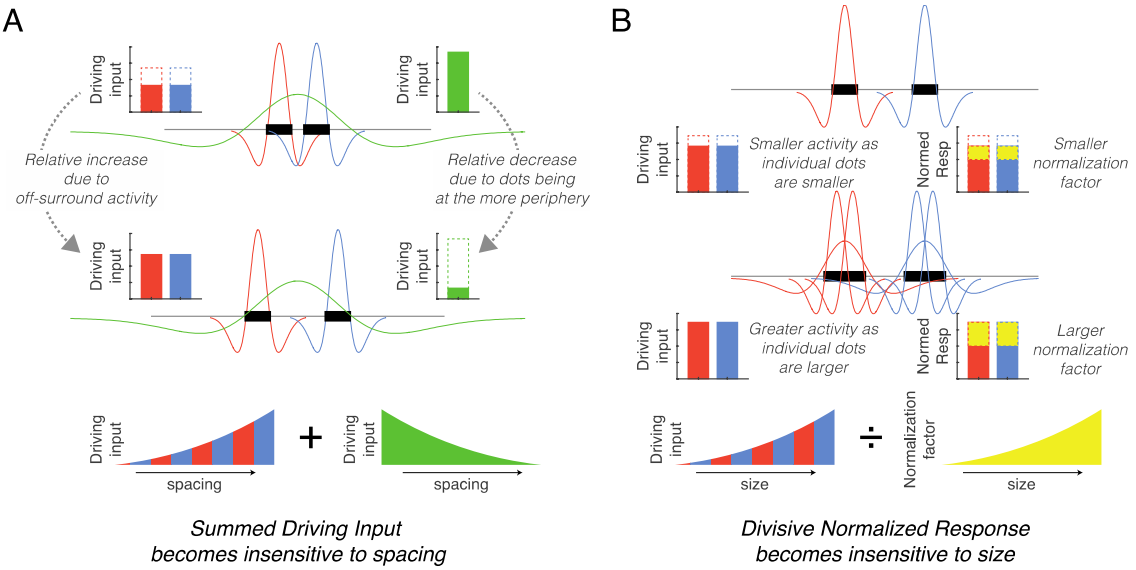

Figure 4. *Simplified schematics explaining the mechanisms underlying the normalization of size and spacing. A. As spacing increases (from top to middle row) the response of small size center-surround filters increases (red and blue) whereas the response of large size center-surround filters decreases (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 normalized output of the neuronal activity is sensitive to numerosity but shows little variation with size and spacing. This pattern is consistent with neural studies finding similar results for the summed response of V1, V2, and V3 in the absence of any behavioral judgment (Fornaciai et al., 2017; Fornaciai and Park, 2018; Paul et al., 2022). 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). Although the few units that were sensitive to numerosity could explain behavior, the abundance of simulated neurons sensitive to non-numerical dimensions is inconsistent with population-level neural activity, which fails to show sensitivity to these non-numerical dimensions in early visual cortex (DeWind et al., 2019; Park, 2018; Van Rinsveld et al., 2020). A key difference between the current model and previous computational models is the inclusion of divisive normalization in the center-surround convolution layer. Unlike prior models, this eliminated the effect of size in the early visual response, without requiring subsequent pooling layers (Creatore et al., 2021; Kim et al., 2021; Nasr et al., 2019; Stoianov and Zorzi, 2012; Testolin et al., 2020) or a decision making process that compares high versus low spatial frequency responses (Dakin et al., 2011).

At first blush, the current model might be considered an extension of Dehaene and Changeux (1993). However, there are four ways in which the current model differs qualitatively from the D&C model. First, the D&C model is one-dimensional, simulating a linear retina, whereas we model a two-dimensional retina feeding into center-surround filters, allowing application to the two-dimensional images used in numerosity experiments (**Fig. 1A**). Second, extreme winner-take-all normalization in the convolution layer of the D&C model implausibly limits visual precision by discretizing the visual response. For example, the convolution layer in the D&C model only knows which of 9 possible sizes and 50 possible locations occurred. In contrast, by using divisive normalization in the current model, each dot produces activity at many locations and many filter sizes despite normalization, and a population could be used to determine exact location and size. Third, extreme winner-take-all normalization also eliminates all information other than dot size and location. By using divisive normalization, the current model represents other attributes such edges and groupings of dots (**Fig. 1B**) and these other attributes provide a different explanation of number sensitivity as compared to D&C. For example, the D&C model as applied to the spacing effect between two small dots (**Fig. 4A**) would represent the dots as existing discretely at two close locations versus two far locations, with the total summed response being two in either case. In contrast, the current model gives the same total response for a different reason. Although the small filters are less active for closely spaced dots, the closely spaced dots *look like a group* as captured by a larger filter, with this addition for the larger filter offsetting the loss for the smaller filter. Similarly, as applied to the dot size effect (**Fig. 4B**), the D&C model would only represent the larger dots using larger filters. In contrast, the current model represents larger dots with larger filters and with *smaller filters that capture the edges of the larger dots*, and yet the summed response remains the same in each case owing to divisive normalization (again, there are offsetting factors across different filter sizes). The final difference

is that the D&C model does not include temporal normalization, which we show to be critical for explaining adaptation and context effects.

Finally, a recent fMRI study reported that neural activity in V1 increases monotonically with numerosity (Paul et al., 2022), which is consistent with the current model at a surface level. The authors, however, concluded that this monotonic increase was better explained by aggregate Fourier power than by numerosity. This explanation is qualitatively different than the center-surround and divisive normalization explanation entailed in the current model. While further investigation may be necessary to distinguish these hypotheses, there are two caveats to consider in relation to the conclusions made by Paul et al. (2022). First, Fourier power uses spatially unbounded sine waves that have little biological plausibility (unlike center-surround or Gabor filters, which are spatially limited). Second, more critically, the aggregate Fourier power metric used by Paul et al. (2022) aggregated only up through the first (or any nth) harmonic, but the value of the harmonic on the frequency spectrum is dictated by dot size and/or dot groupings. In other words, the Fourier metric required a priori knowledge about each image. Including all frequencies, regardless of dot size, would likely produce a different conclusion. It is unclear how the visual system could know in advance an appropriate cutoff for a harmonic, although development of a more biologically plausible Fourier power model might identify testable differences between these accounts.

Our conclusions are primarily in terms of the qualitative effects of center-surround filtering and divisive normalization, which collectively produce sensitivity to numerosity. However, specific quantitative predictions will change depending on specific model assumptions. For instance, our simulations assumed a distribution of filter sizes that ranged from much smaller to much larger than the presented dots. The responses from filters small enough to capture edges of dots tends to offset the responses from filters large enough to capture local groups of dots, producing relative insensitivity to dot spacing and size (see **Fig. 4**). However, there may be extreme cases where this balancing act breaks down. For instance, studies found that when dots are presented in the periphery where receptive field sizes are larger (Li et al., 2021; Valsecchi et al., 2013) or if the dots are crowded and hard to individuate (Anobile et al., 2014), numerosity perception exhibits different behavioral characteristics. We simulated one extreme by submitting to the model images that contained very small dots (too small to allow edge responses) densely packed in a circular aperture. For this extreme, the summation of normalized responses was still primarily sensitive to number, but that sensitivity was smaller compared to our original simulation, and there was also some moderate sensitivity to size and spacing (**Fig. 2–figure supplement 3**). Our simulation also assumed an equal number of small and large center-surround filters although in reality there are likely fewer large filters. This assumption was made out of computational convenience, although we note that similar results would emerge with an unequal distribution of filters if the divisive normalization amplification factor scaled with filter size (e.g., if the larger number of small filters more strongly inhibited each other) or if the neighborhood size of divisive normalization scaled with filter size in a nonlinear manner. By investigating how these assumptions relate to behavior and physiology, future studies may provide additional mechanistic insights into magnitude perception in general.

The success of this model does not necessarily imply that neuronal responses in early visual regions directly determine behavioral responses (Fornaciai and Park, 2018). 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. For instance, if decisional uncertainty is modeled by assuming a constant level of decisional noise, regardless of the visual information, then the model will naturally produce Weber's scaling law of just noticeable differences considering that the normalized response follows a log-linear pattern as a function of numerosity (see **Fig. 2C**). More complex decisional assumptions could be introduced in an attempt to model the effects of task instructions that are known to bias decisions on magnitude judgment (Castaldi et al., 2019; Cicchini et al., 2016). More assumptions about top-down semantic influences may also explain recent coherence illusion results in orientation or color (DeWind et al., 2020; Qu et al., 2022), for instance if observers are drawn to focus on a particular feature of the stimulus when comparing two dot arrays.

Another line of possible future work concerns divisive normalization in higher cortical levels involving neurons with more complex receptive fields. While the current normalization model with center-surround filters successfully explained visual illusions caused by regularity, grouping, and heterogeneity, other numerosity phenomena such as topological invariants and statistical pairing (He et al., 2015; Zhao and Yu, 2016) may require the action of neurons with receptive fields that are more complex than center-surround filters. For example, another well-known visual illusion is the effect of connectedness, whereby an array with dots connected pairwise with thin lines is underestimated (by up to 20%) compared to the same array without the lines connected (Franconeri et al., 2009). This underestimation effect likely arises from barbell-shaped pairwise groupings of dots, rather than the circularly symmetric groupings of dots that are captured with center-surround filters. Nonetheless, a small magnitude (6%) connectedness illusion emerges with center-surround filters (**Fig. 3–figure supplement 7**). Augmenting the current model with a subsequent convolution layer containing oriented line filters and oriented normalization neighborhoods of different sizes might increase the predicted magnitude of the illusion.

In conclusion, our results indicate that divisive normalization in a single 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 (Park et al., 2016). This account predicts that various well-known numerosity illusions across space and time arise naturally within the same neural responses that encode numerosity, rather than reflecting later stage processes. 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 (DeWind et al., 2015; Park et al., 2016), 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. log(*N*) ranged from 2.322 to 4.322 with the median of 3.322; log(*Sz*) ranged from 16.305 to 18.305 with the median of 17.305; log(*Sp*) ranged from 19.646 to 21.646 with the median of 20.646. This approach resulted in 35 unique points in the three-dimensional parameter space (see **Fig. 1A**). 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 (Ginsburg, 1976). 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 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 (Marr and Hildreth, 1980), $K = 1.6$ 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 (Heeger, 1991), 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 (Carandini and Heeger, 2012), 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/.

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.

Declaration of interests

The authors declare no competing interests.

Title and Legend for Figure Supplements

Figure 2 – Figure supplement 1. Additional illustration concerning the driving input. Correlation between summed driving input, Σ*D*, and log-scaled total area (TA), total perimeter (TP), and number (N).

Figure 2 – Figure supplement 2. Simulation results showing the effects of number (N), size (Sz), and spacing (Sp) on the normalized response (i.e., the model with divisive normalization) 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 2 – Figure supplement 3. Simulation results from images of densely packed dot arrays with extremely high numerosity. (A) The dots arrays were systematically constructed ranging equally across the dimensions of *N*, *Sz*, and *Sp*, which was achieved by using the following parameters: number (n) = from 90 to 360, dot radius (r_d) = from 1 to 2 pixels, field radius (r_f) = from 45 to 90 pixels. For each point in the 2×2×2 parameters space, 16 unique arrays were created. (B) Examples of dot array images are shown. These images were submitted to the current computational model with the same parameters used in our original analysis ($r = 2\sigma$ and $\gamma = 2$). (C) Summed driving input (ΣD) was modulated primarily by *N* and *Sz*. Summed normalized response (ΣR) was most modulated by *N* but also by *Sz* and *Sp* to some degree. The slope of the linear fit to *N*, *Sz*, and *Sp* adjusted by the baseline (the slope estimate divided by the intercept estimate in the simple regression) was 0.4086, 0.1958, and 0.1488, respectively. Note that this baseline-adjusted slope allows comparison of relative change in the response driven by *N*, *Sz*, and *Sp*, despite differences in the baseline activity across different sets of images. In our original simulation, the baseline-adjusted slopes for *N*, *Sz*, and *Sp* were 0.5771, 0.0646, and 0.0321, respectively. Thus, the same computational network when representing much more densely packed dot arrays seems to show relatively decreased sensitivity to numerosity. These results indicate that neural sensitivity to various magnitude dimensions and the degree of that sensitivity differ based on the assumptions about the distribution of filters and filter sizes.

Figure 3 – Figure supplement 1. Simulation of visual illusions considering the driving input (i.e., the model without divisive normalization). No underestimation was observed in any of these cases. If any, irregularly spaced arrays (by 2.19%), grouped arrays (by 1.98%), and more heterogeneous arrays (by 3.06%) were overestimated based on their driving input. In sum, without divisive normalization, the model failed to explain the typically observed visual

illusions. Error bars indicate one standard deviation of the normalized response across simulations.

Figure 3 – Figure supplement 2. 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 occurs because a larger dot involves a greater number of filters overlapping with the dot (i.e., greater driving input), but this greater number of filters leads to a greater normalization factor (increase in the denominator of divisive normalization). 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.

Figure 3 – Figure supplement 3. Adaptation effects as a function of model parameters. In this simulation, the target of 10 dots was preceded by an adaptor of 5, 10, or 20 dots. Temporal normalization could be understood in terms of a sigmoid response curve with the amplification factor (δ) determining the slope of the curve and the recency weighting factor (ω) determining the horizontal position of the curve. Smaller ω values resulted in a relative overestimation of the normalized response to the target, which can be explained by the relative leftward horizontal shift of the sigmoid response curve and hence relative increase in normalized activity (non-linearly as a function of driving input). Larger δ values resulted in greater under- and over-estimation effects, which can be explained by the sharpening of the sigmoid response curve. Error bars indicate one standard deviation of the normalized response across simulations.

Figure 3 – Figure supplement 4. Adaptation effects along the size dimension. The target of medium-sized array was preceded by an adaptor of small-, medium-, or large-sized array. No systematic pattern of adaptation was observed in these simulations. Error bars indicate one standard deviation of the normalized response across simulations.

Figure 3 – Figure supplement 5. Adaptation effects along the spacing dimension. The target of medium-spaced array was preceded by an adaptor of small-, medium-, or large-spaced array. No systematic pattern of adaptation was observed in these simulations. Error bars indicate one standard deviation of the normalized response across simulations.

Figure 3 – Figure supplement 6. Context effects as a function of model parameters. When the model saw 400 dot arrays that varied randomly across number, size, and spacing, the normalized responses to images corresponding to small, medium, and large sizes (*Sz*) showed no association

with size. When then model saw 400 dot arrays that differed only in size, the normalized responses were strongly associated with size. Such a pattern was consistent across all the simulations over various amplification factors (δ) and recency weighting factors (ω) tested. Error bars represent one standard deviation of the normalized response across simulations. Note that the error bars in the exclusive change in size conditions are extremely small.

Figure 3 – Figure supplement 7. Simulation of the connectedness illusion. In order to simulate the connectedness illusion, 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 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. The resulting images were checked again for any cross over of lines, in which case both that image and the corresponding connected image were removed. A total of 16 connected and 16 unconnected arrays entered the simulation. The connected arrays were underestimated by over 6% (Cohen's d = 5.72). Such an underestimation was not observed when considering the driving input (i.e., the model without divisive normalization). Error bars represent one standard deviation of the normalized response across simulations.

References

Aagten-Murphy D, Burr D. 2016. Adaptation to numerosity requires only brief exposures, and is determined by number of events, not exposure duration. *J Vis* **16**:22. doi:10.1167/16.10.22

Allik J, Tuulmets T. 1991. Occupancy model of perceived numerosity. *Percept Psychophys* **49**:303–14.

Anobile G, Cicchini GM, Burr DC. 2014. Separate mechanisms for perception of numerosity and density. *Psychol Sci* **25**:265–70. doi:10.1177/0956797613501520

Bell AJ, Sejnowski TJ. 1997. The "independent components" of natural scenes are edge filters. *Vision Res* **37**:3327–3338. doi:10.1016/S0042-6989(97)00121-1

Burr D, Ross J. 2008. A Visual Sense of Number. *Curr Biol* **18**:425–428. doi:10.1016/j.cub.2008.02.052

Carandini M, Heeger DJ. 2012. Normalization as a canonical neural computation. *Nat Rev Neurosci* **13**:51–62. doi:10.1038/nrn3136

Castaldi E, Piazza M, Dehaene S, Vignaud A, Eger E. 2019. Attentional amplification of neural codes for number independent of other quantities along the dorsal visual stream. *Elife* **8**. doi:10.7554/eLife.45160

Cicchini GM, Anobile G, Burr DC. 2016. Spontaneous perception of numerosity in humans. *Nat Commun* **7**:12536. doi:10.1038/ncomms12536

Creatore C, Sabathiel S, Solstad T. 2021. Learning exact enumeration and approximate estimation in deep neural network models. *Cognition* **215**:104815. doi:10.1016/j.cognition.2021.104815

Dakin SC, Tibber MS, Greenwood JA, Kingdom FAA, Morgan MJ. 2011. A common visual metric for approximate number and density. *Proc Natl Acad Sci* **108**:19552–19557. doi:10.1073/pnas.1113195108

Dehaene S. 2011. The number sense : how the mind creates mathematics. New York: Oxford University Press.

Dehaene S, Changeux J-P. 1993. Development of Elementary Numerical Abilities: A Neuronal Model. *J Cogn Neurosci* **5**:390–407. doi:10.1162/jocn.1993.5.4.390

DeWind NK, Adams GK, Platt ML, Brannon EM. 2015. Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition* **142**:247–265. doi:10.1016/j.cognition.2015.05.016

DeWind NK, Bonner MF, Brannon EM. 2020. Similarly oriented objects appear more numerous. *J Vis* **20**:4. doi:10.1167/jov.20.4.4

DeWind NK, Park J, Woldorff MG, Brannon EM. 2019. Numerical encoding in early visual cortex. *Cortex* **114**:76–89.

Feigenson L, Dehaene S, Spelke E. 2004. Core systems of number. *Trends Cogn Sci* **8**:307–314. doi:10.1016/j.tics.2004.05.002

Fornaciai M, Brannon EM, Woldorff MG, Park J. 2017. Numerosity processing in early visual cortex. *Neuroimage* **157**:429–438.

Fornaciai M, Park J. 2021. Disentangling feedforward versus feedback processing in numerosity representation. *Cortex* **135**:255–267. doi:10.1016/j.cortex.2020.11.013

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

Franconeri SL, Bemis DK, Alvarez GA. 2009. Number estimation relies on a set of segmented objects. *Cognition* **113**:1–13. doi:10.1016/j.cognition.2009.07.002

Frith CD, Frit U. 1972. The solitaire illusion: An illusion of numerosity. *Percept Psychophys* **11**:409–410. doi:10.3758/BF03206279

Gebuis T, Cohen Kadosh R, Gevers W. 2016. Sensory-integration system rather than approximate number system underlies numerosity processing: A critical review. *Acta Psychol (Amst)* **171**:17–35. doi:10.1016/j.actpsy.2016.09.003

Ginsburg N. 1976. Effect of Item Arrangement on Perceived Numerosity: Randomness vs Regularity. *Percept Mot Skills* **43**:663–668. doi:10.2466/pms.1976.43.2.663

He L, Zhou K, Zhou T, He S, Chen L. 2015. Topology-defined units in numerosity perception. *Proc Natl Acad Sci* **112**:E5647–E5655. doi:10.1073/pnas.1512408112

Heeger DJ. 1991. Nonlinear model of neural responses in cat visual cortex. *Comput Model Vis Process* 119–133.

Huber DE, O'Reilly RC. 2003. Persistence and accommodation in short-term priming and other perceptual paradigms: temporal segregation through synaptic depression. *Cogn Sci* **27**:403–430. doi:10.1016/S0364-0213(03)00012-0

Kim G, Jang J, Baek S, Song M, Paik S-B. 2021. Visual number sense in untrained deep neural networks. *Sci Adv* **7**. doi:10.1126/sciadv.abd6127

Krizhevsky A, Sutskever I, Hinton GE. 2012. ImageNet classification with deep convolutional neural networksAdvances in Neural Information Processing Systems. pp. 1106–1114.

Lee H, Baek J, Chong SC. 2016. Perceived magnitude of visual displays: Area, numerosity, and mean size. *J Vis* **16**:12. doi:10.1167/16.3.12

Leibovich T, Katzin N, Harel M, Henik A. 2017. From “sense of number” to “sense of magnitude”: The role of continuous magnitudes in numerical cognition. *Behav Brain Sci* **40**:e164. doi:10.1017/S0140525X16000960

Li MS, Abbatecola C, Petro LS, Muckli L. 2021. Numerosity Perception in Peripheral Vision. *Front Hum Neurosci* **15**. doi:10.3389/fnhum.2021.750417

Marr D, Hildreth E. 1980. Theory of edge detection. *Proc R Soc London Ser B Biol Sci* **207**:187–217. doi:10.1098/rspb.1980.0020

Nasr K, Viswanathan P, Nieder A. 2019. Number detectors spontaneously emerge in a deep neural network designed for visual object recognition. *Sci Adv* **5**. doi:10.1126/sciadv.aav7903

Nieder A. 2016. The neuronal code for number. *Nat Rev Neurosci* **17**:366–382. doi:10.1038/nrn.2016.40

Park J. 2018. 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.

Park J, DeWind NK, Woldorff MG, Brannon EM. 2016. Rapid and Direct Encoding of Numerosity in the Visual Stream. *Cereb Cortex* **26**:748–763. doi:10.1093/cercor/bhv017

Park J, Godbole S, Woldorff MG, Brannon EM. 2021. Context-Dependent Modulation of Early Visual Cortical Responses to Numerical and Nonnumerical Magnitudes. *J Cogn Neurosci* **33**:2536–2547. doi:10.1162/jocn_a_01774

Paul JM, van Ackooij M, ten Cate TC, Harvey BM. 2022. Numerosity tuning in human association cortices and local image contrast representations in early visual cortex. *Nat Commun* **13**:1340. doi:10.1038/s41467-022-29030-z

Qu C, DeWind NK, Brannon EM. 2022. Increasing entropy reduces perceived numerosity throughout the lifespan. *Cognition* **225**:105096. doi:10.1016/j.cognition.2022.105096

Scherer D, Müller A, Behnke S. 2010. Evaluation of Pooling Operations in Convolutional Architectures for Object RecognitionLecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 92–101. doi:10.1007/978-3-642-15825-4_10

Stoianov I, Zorzi M. 2012. Emergence of a “visual number sense” in hierarchical generative models. *Nat Neurosci* **15**:194–6. doi:10.1038/nn.2996

Testolin A, Zou WY, McClelland JL. 2020. Numerosity discrimination in deep neural networks: Initial competence, developmental refinement and experience statistics. *Dev Sci* **23**. doi:10.1111/desc.12940

Valsecchi M, Toscani M, Gegenfurtner KR. 2013. Perceived numerosity is reduced in peripheral

vision. *J Vis* **13**:7–7. doi:10.1167/13.13.7

van Oeffelen MP, Vos PG. 1982. Configurational effects on the enumeration of dots: Counting by groups. *Mem Cognit* **10**:396–404. doi:10.3758/BF03202432

Van Rinsveld A, Guillaume M, Kohler PJ, Schiltz C, Gevers W, Content A. 2020. 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. doi:10.1073/pnas.1917849117

Zhao J, Yu RQ. 2016. Statistical regularities reduce perceived numerosity. *Cognition* **146**:217–222. doi:10.1016/j.cognition.2015.09.018