Transformation from local image contrast to location-independent numerosity tuning in human extrastriate cortex

Jacob M. Paul[1*], Martijn van Ackooij[1], Tuomas C. ten Cate[1] & Ben M. Harvey[1]

[1] Experimental Psychology, Helmholtz Institute, Utrecht University, Heidelberglaan 1, Utrecht 3584 CS, Netherlands

* Corresponding author:
Jacob M. Paul
Experimental Psychology, Helmholtz Institute, Utrecht University
Heidelberglaan 1
Utrecht 3584 CS
Netherlands

Keywords: numerosity; primary visual cortex; image contrast; fMRI; Fourier power

Abstract

Many animals use visual numerosity, the number of items in a group, to guide behavior. Neurons in human association cortices show numerosity-tuned responses, decreasing amplitude with distance from a specific numerosity. How are such responses derived from early visual responses? Recent studies show aggregate response amplitudes in human early visual cortex monotonically increase with numerosity, regardless of object size and spacing. This is surprising because numerosity is typically considered a high-level visual or cognitive feature. Here we first use computational modelling of 7T fMRI data to show these monotonic responses originate at the stimulus's retinotopic location in primary visual cortex (V1). Given this location, we then ask whether these monotonic responses can be better described by V1's established response properties. We characterize the Fourier decomposition (into contrast at specific orientations and spatial frequencies) of laboratory numerosity stimuli. This demonstrates that aggregate Fourier power (at all orientations and spatial frequencies) nonlinearly follows numerosity with little effect of item size, spacing or shape: it is proportional to numerosity at a fixed contrast. This nonlinear relationship lets us distinguish predictions of responses to Fourier power and numerosity. Monotonic responses are better predicted by Fourier power, later tuned responses are better predicted by numerosity. Tuned responses emerge after lateral occipital cortex and are independent of retinotopic location. We propose that numerosity's straightforward perception and neural responses reflect its straightforward estimation from early visual spatial frequency domain image representations. Our numerical vision may have built on behaviorally beneficial analysis of spatial frequency in simpler animals.

Significance statement

Quantities like object number (numerosity) guide many animals' behaviors. Numerosity is therefore considered a high-level visual or cognitive feature. Paradoxically, recent studies show aggregate response amplitudes of human early visual areas follow numerosity regardless of object size or spacing. We demonstrate that the aggregate Fourier power of laboratory numerosity displays follows numerosity closely but nonlinearly at a fixed contrast, with little effect of object size or spacing, and predicts early visual responses more closely and parsimoniously than numerosity. Subsequent processing stages overcome this nonlinearity, giving numerosity-selective responses that are separated from position-selective responses. Numerosity's straightforward estimation from early visual spatial frequency domain representations links cognitive quantity processing to early visual contrast representations. This may explain numerosity-guided behaviors of simpler animals.

Humans and many other animals use visual numerosity, the number of items in a set, to guide behavior. Many species have neurons tuned to numerosity, decreasing in response amplitude with distance from a specific preferred numerosity (1–3). Functional magnetic resonance imaging (fMRI) has identified numerosity-tuned neuronal populations in specific areas of human association cortex using both population receptive field modelling approaches (4, 5) and fMRI adaptation (or repetition suppression)(6). Other fMRI studies using multivoxel pattern analyses (7, 8) and representational similarity analyses (9) also support the existence of numerosity-tuned neural populations in the human brain. Response properties of these neurons mirror properties of numerosity perception (3, 6, 10). Numerosity perception is correlated with numerosity tuned responses between trials (3, 10), and repetition suppression (6) and multivoxel pattern discriminability between individuals (11, 12).

There remains considerable debate over how such numerosity-tuned responses are derived from visual inputs. One view proposes that numerosity tuning and perception reflect non-numerical image features that are often correlated with numerosity, like density (13), or contrast energy at high spatial frequencies (14). However, growing convergent evidence from psychophysical, neuroimaging and computational research indicates numerosity itself is represented and perceived (15–19). These views could be reconciled by a non-numerical image feature from which numerosity could be estimated regardless of other image parameters like item size and spacing.

Computational models of numerosity estimation (20, 21) yield numerosity tuned responses via an intermediate stage which monotonic increases in response amplitude with numerosity. Unsupervised machine learning also shows that numerosity-tuned responses can be derived through a monotonic (i.e. summation) coding stage (21). However, none of these models show how the monotonic stage disregards size and spacing, like numerosity perception does.

EEG and fMRI results show that early visual cortex responses to numerosity stimuli appear to monotonically increase with numerosity, regardless of item size or spacing (22, 23). This monotonic response emerges very quickly after stimulus presentation, suggesting it reflects feedforward processing, and is not computed from other parameters like stimulus area and density. This finding has recently been replicated using frequency-tagged EEG (24). This early visual cortex response to numerosity is surprising, as numerosity is generally considered a relatively high-level visual feature. Early visual neurons respond to image contrast in the frequency domain: at specific positions, orientations and spatial frequencies (25). So, it is unclear how early visual neural responses could closely follow numerosity regardless of size and spacing.

Association cortex recording sites with numerosity-tuned responses largely overlap with higher visual field maps (5), but their spatial population receptive fields (pRF) do not necessarily overlap with the numerosity stimulus area (5, 26), and numerosity preferences are unrelated to pRF position or size (26). Numerosity tuning in functionally homologous macaque brain areas locations (27) also does not require the responding single neurons to have spatial receptive fields including the stimulus region, or even have discernable receptive fields at all (28). Indeed, spatial receptive field properties of numerosity-tuned neurons do not influence numerosity preferences, tuning width or firing rate. For early visual monotonic responses, some properties of EEG event related potentials suggest these emerge in V2 or V3 (29), but their precise visual field map and retinotopic location remains unclear.

Here we use computational modelling of human 7T fMRI data to ask precisely where in the early visual cortex monotonic responses emerge, whether they can be explained by the spatial frequency domain image representation in the early visual cortex, and how they are subsequently transformed to give location-independent numerosity tuned responses in the human association cortices.

Results

Numerosity response profiles differ between early visual cortex and association cortices

We presented displays of gradually changing numerosity to our participants while collecting 7T fMRI data (see Methods). We included different stimulus configurations that held total item area, individual item size, or total item perimeter constant across numerosities, or placed items in a dense group (Fig. 1). These configurations varied item size and spacing considerably but produced similar responses.

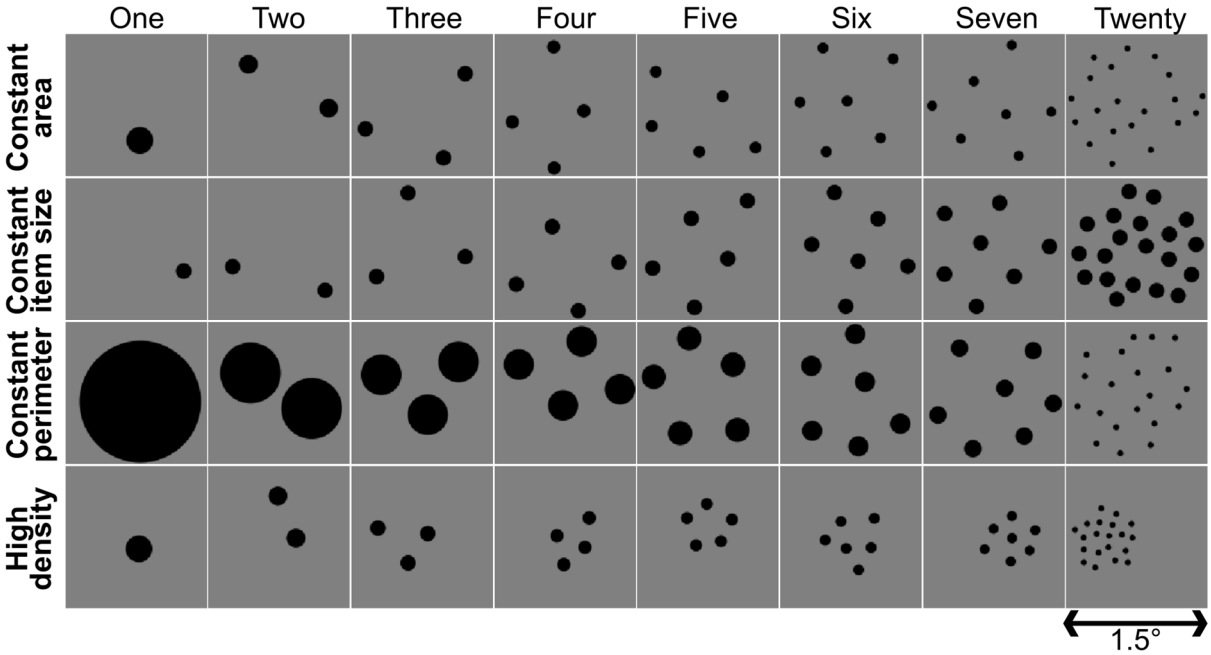

Fig. 1. Example displays from each stimulus configuration and numerosity.

Different cortical locations showed different relationships between numerosity and response amplitude. These were well captured by response models with monotonically increasing, monotonically decreasing, or tuned responses to log(numerosity) at different locations (Fig. S2C-D). We compared the response variance explained by these models in cross-validated data (Fig. 2A and Fig. S4). Separate visual field mapping data demonstrate that monotonically increasing responses were consistently found only in early visual cortex's central visual field representation. Numerosity tuned response were found outside early visual

cortex, in previously described areas of temporal-occipital, parietal-occipital, superior parietal, and frontal cortices containing topographic numerosity maps (5). These largely overlapped with higher extrastriate visual field maps. Monotonically decreasing responses were found next to areas showing tuned responses.

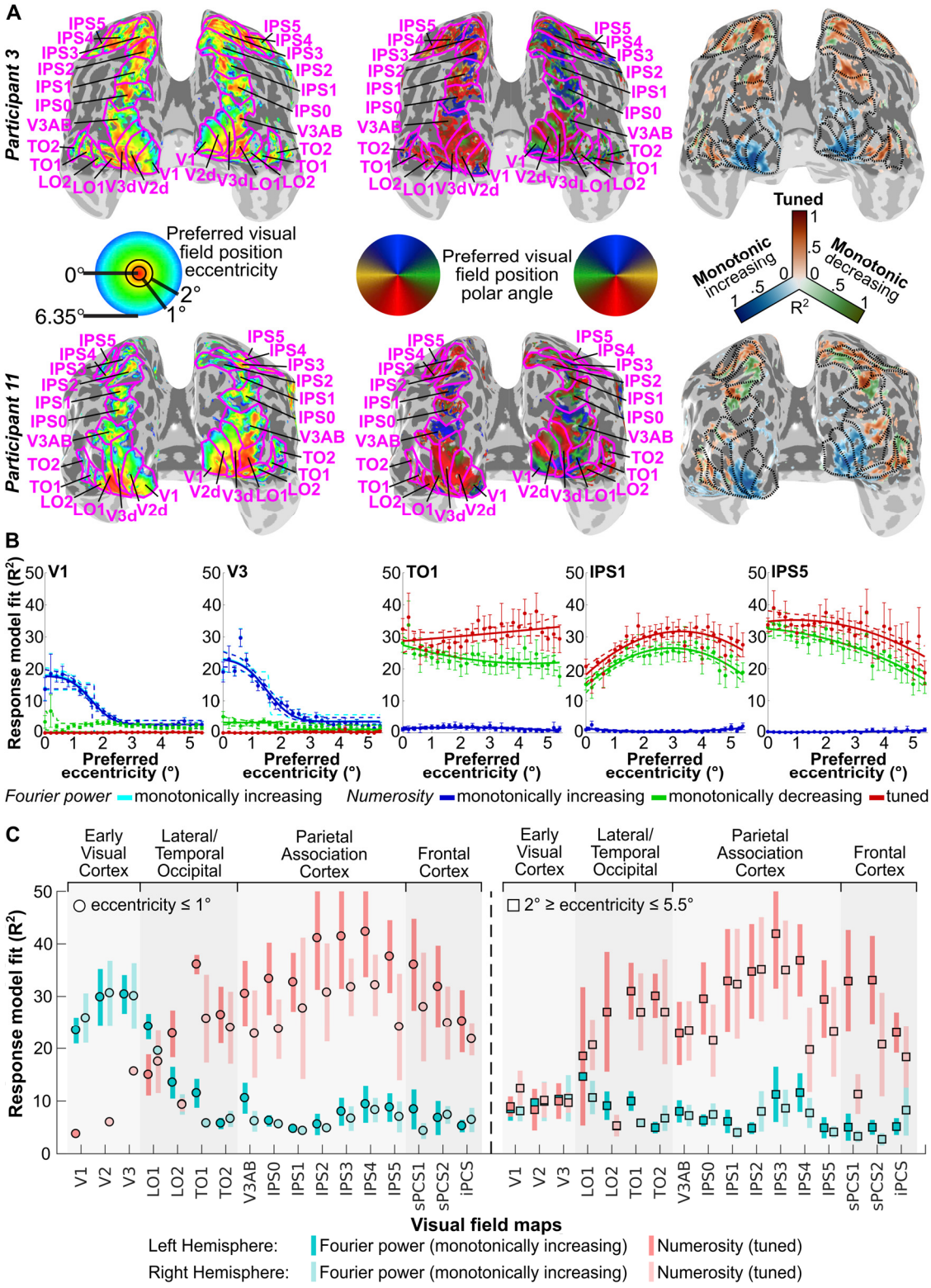

Fig. 2. Relationships between responses to numerosity and visual field position (A) Visual position preferences (eccentricity and polar angle, left and middle) and best fitting numerosity model (right, colors) at each cortical location, for two illustrative participants. Magenta lines and labels show visual field map borders and names respectively. (B) Progression of each numerosity model's fit with preferred visual field eccentricity (colors) in representative visual field maps (grouped across participants). Cyan shows fits of monotonically increasing responses to aggregate Fourier power. Filled circles show mean variance explained per eccentricity bin, error bars show the standard error of the mean. Solid lines show the best fit to changes with eccentricity, dashed lines are bootstrap 95% confidence intervals determined by bootstrapping. (C) Model fit variance explained by numerosity (tuned, reds) and aggregate Fourier power (monotonically increasing, blues) response models for each visual field map hemisphere at eccentricities below 1° (left) and between 2° and 5.5° (right). Points represent the population marginal mean, error bars are 95% confidence intervals; non-overlapping error bars show significant differences at $p < 0.05$.

Early visual monotonic responses

Early visual (V1–V3, LO1) responses were consistently predicted more closely by monotonically increasing rather than tuned responses to numerosity (Fig. 2B). Critically, model fits depended on the recording sites' visual field position preferences: those near fixation (the stimulus location) showed better fits, gradually decreasing to zero into the periphery. These progressions were well captured by cumulative Gaussian sigmoid functions (Fig. S5C). Inflection points of these sigmoid curves fell at eccentricities between one and two degrees of visual angle (Fig. 2B and Fig. S2A).

A location-specific monotonic response to numerosity that has emerged by V1 is perhaps surprising: numerosity is generally considered a complex visual or cognitive feature, and responses of V1 neurons depend on contrast at specific orientations and spatial frequencies within their receptive field. Images are often transformed into this spatial frequency domain using a 2-dimensional Fourier decomposition, which similarly separates changes across an image into the contrast at each orientation and spatial frequency (30). We therefore reasoned that aggregate Fourier power (across all orientations and spatial frequencies) in the spatial frequency domain might closely predict the measured aggregate response of V1 populations. We summed the absolute Fourier decomposition of the displays within the first spatial

frequency harmonic, across all orientations (Fig. 3A-B). This revealed that aggregate Fourier power followed numerosity closely, but nonlinearly, and similarly in all stimulus configurations (Fig. 3C). The aggregate power the second harmonic showed a similar pattern at half the amplitude. Indeed, aggregate Fourier power changed very little over a wide range of item sizes, spacings and shapes that were not tested in our experiments: it followed numerosity closely except in extreme cases and a slight increase with spacing (Fig. 3D-G and Figs. S7-S10).

Connecting dot pairs with bars reduces perceived numerosity. This perception was originally compared to numerosity perception when bars were placed among to dots but fully unconnected (14, 31)(Fig. 3H, black inset). The Fourier power of a connected pair (red) is lower than fully unconnected dots and bars (black), superficially predicting the connectedness illusion. However, rotating the dots and bars to break the connection (magenta) has the same lower Fourier power while it is perceived as more numerous. Furthermore, the Fourier power of the connected pair is greater than the dots alone (grey), while connected pairs are perceived as less numerous than dots alone. Connecting illusory contour inducers (blue) also reduce perceived numerosity and have lower Fourier power than dots alone, but these inducers again have the same effect when rotated to break the connection (cyan). Therefore, while bars and illusory contour affect Fourier power, these effects are not consistent with perceptual effects: the connectedness illusion likely results from higher-order segmentation processes (32).

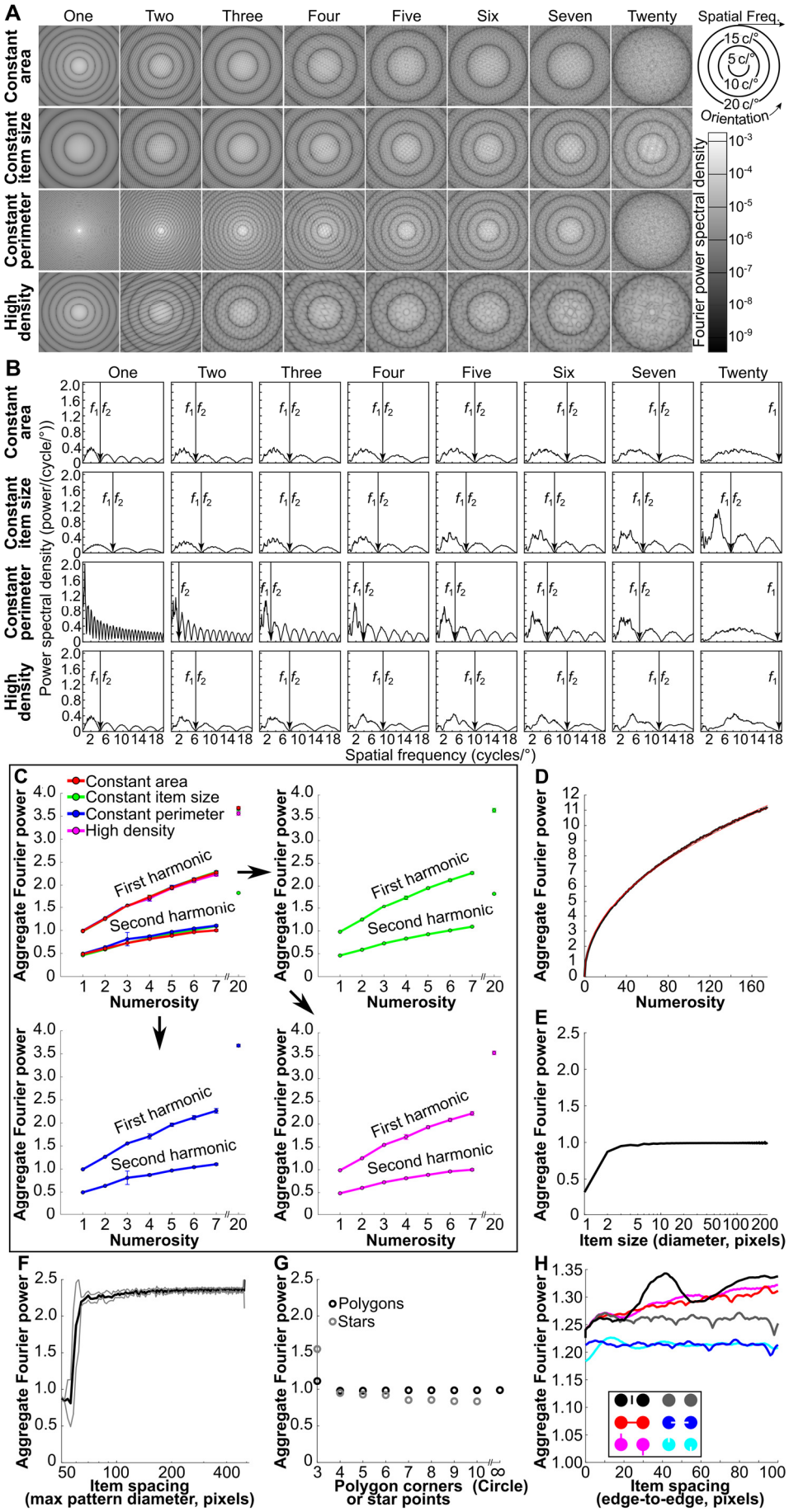
A
One
Two
Three
Four
Five
Six
Seven
Twenty
Spatial Freq.
15 c/°
5 c/°
10 c/°
20 c/°
Orientation
Constant area
Constant item size
Constant perimeter
High density
Fourier power spectral density
B
Power spectral density (power/(cycle/°))
Spatial frequency (cycles/°)
C
Constant area
Constant item size
Constant perimeter
High density
First harmonic
Second harmonic
Aggregate Fourier power
Numerosity
D
Numerosity
E
Item size (diameter, pixels)
F
Item spacing
(max pattern diameter, pixels)
G
Polygons
Stars
Polygon corners (Circle)
or star points
H
Item spacing
(edge-to-edge, pixels)

Fig. 3. Aggregate Fourier power follows numerosity closely with little effect of item size, spacing, shape or connectedness. (A) 2-dimensional Fourier transforms for each image in Fig. 1. (B) Fourier power spectral density at each spatial frequency, collapsed over orientation. The limit of the first harmonic (f_1) is used to determine the aggregate Fourier power of each image. (C) Aggregate Fourier power (B) follows numerosity closely and similarly across stimulus configurations. We divide aggregate Fourier power by the product of number display pixels and the square root of two here, making the power of one circle approximately one for any display resolution. Points show the mean power of all stimuli of the same numerosity in the same stimulus configuration, error bars show the standard deviation. Upper left panel shows all configurations overlaid; other panels show individual configurations. (D) Aggregate Fourier power as a function of numerosity for a fixed item size (black line). Red lines show $y = x^{0.466}$ and $y = x^{0.470}$, which approximate the observed relationship. (E) Aggregate Fourier power of one circle of different diameters is approximately constant for diameters above 3 pixels: smaller circles are inaccurately rendered. (F) Aggregate Fourier power for a group of 7 circles (each 16 pixels diameter) increases slightly with spacing for stimulus area diameters above 70 pixels. Smaller areas have little space between items: at 50 pixels all items touch (Fig. S9). Grey lines show the standard deviation across displays. (G) Aggregate Fourier power is approximately constant for regular polygons above 3 corners. Triangles and 3-pointed stars have greater power than for circles, while other stars have less power. (H) Effects of connections by a bar and an illusory contour on Fourier power, together with control dot pairs with the same change to the dot but no connection and no reduction in perceived numerosity. Compared to dots alone, bars increase Fourier power while illusory contour inducers reduce Fourier power, regardless of connectedness.

Several other non-numerical stimulus features have been proposed to account for numerosity perception because they are correlated with numerosity in some cases, though none shows this generalization across displays. We quantified several such features for each numerosity and stimulus configurations (17, 18) and tested predictions of response models that monotonically follow these features against the recorded early visual (V1-V3) responses. We tested predictions of position selective pRFs responding either to dot bodies (i.e. luminance) or edges (i.e. contrast) (5). All of these models predicted response amplitudes significantly less well than log(numerosity) (all $Z \geq 2.37$, $p < .0176$ in paired Wilcoxon signed rank tests, n=20 hemisphere measurements) (Fig. 4A).

However, early visual responses were significantly better predicted by monotonically increasing responses to log(Fourier power) than log(numerosity) (and all other models), both across all stimulus configurations ($Z = 3.78$, $p = 0.00015$) and within each configuration (all $Z \geq 2.28$, $p < 0.0228$) (Fig. 4C). Conversely, numerosity tuned models predict the tuned response

of six previously identified numerosity maps (5) significantly better than models tuned to aggregate Fourier power across all stimulus configurations (Z = -3.77, p = 0.00016) (Fig. 4B), as previously shown for other non-numerical features (17, 18). Numerosity-tuned models also fit better within each stimulus configuration, though this difference is not significant in the high density configuration (Z = -1.19, p = 0.0731) (Fig. 4D).

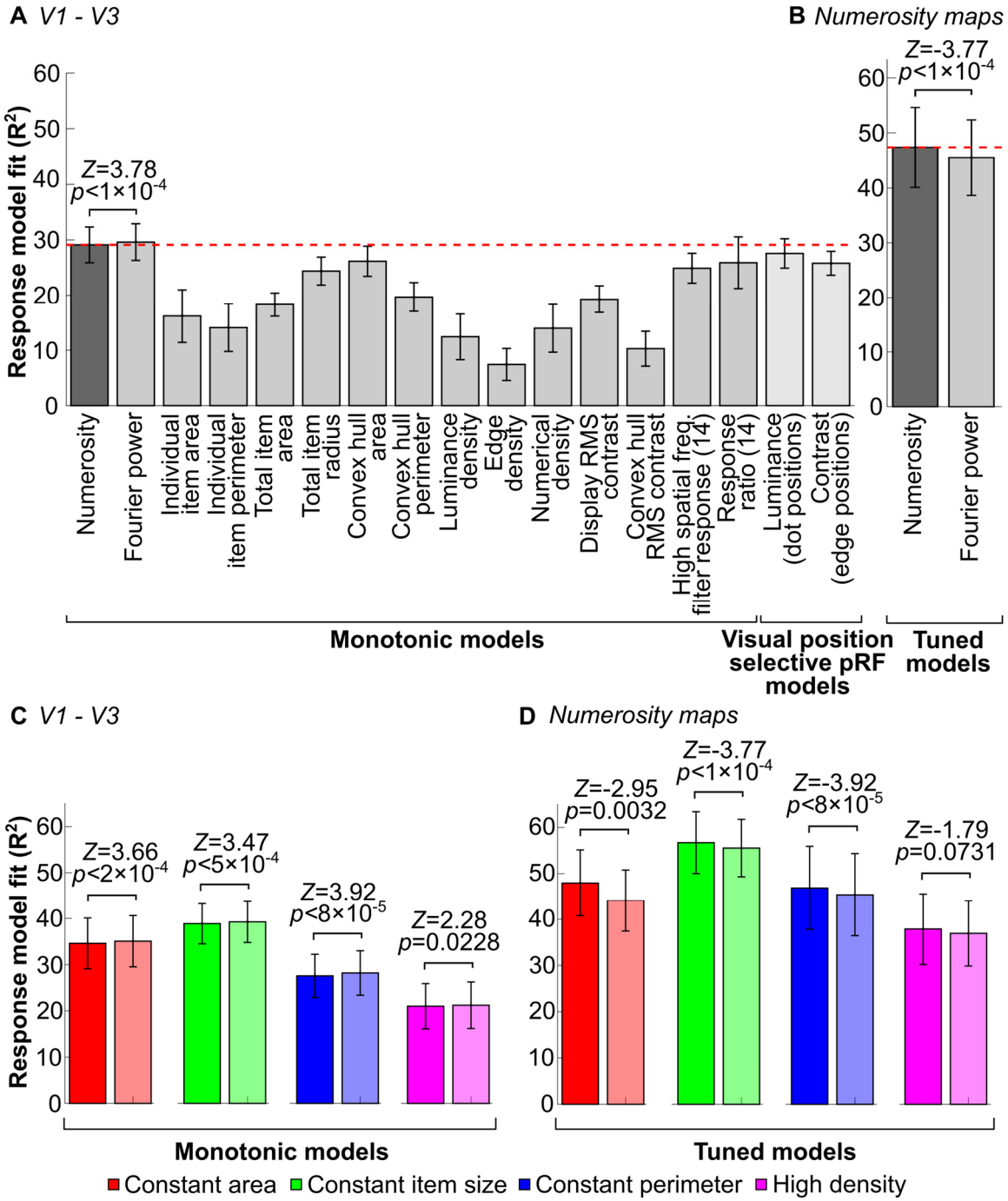

Fig. 4. Responses to aggregate Fourier power predict early visual monotonic responses better than numerosity, but numerosity predicts tuned responses better. (A) Variance explained by monotonic response models for numerosity and non-numerical visual features, and visual position selective pRF models in V1-V3. (B) Variance explained by tuned response models for numerosity and aggregate Fourier power in the association cortex numerosity maps. (C) Variance explained by monotonic response models for numerosity and Fourier power in V1-V3 in each stimulus configuration. (D) Variance explained by tuned response models for numerosity and Fourier power in the association cortex numerosity maps in each stimulus configuration. Bar height is the mean variance explained across all maps. Error bars show 95% confidence intervals, reflecting the range of fits between individual maps. P-values show significance of differences in paired Wilcoxon signed rank tests.

Differences in model fits within and between visual field maps

We separated each visual field map into two eccentricity ranges: near to the stimulus/fixation (below 1° eccentricity) and far from the stimulus/fixation (2–5.5° eccentricity). A three-factor ANOVA (factors: visual field map, eccentricity range, and participant) revealed that monotonically increasing Fourier power response models fit better near fixation ($p < 10^{-10}$, $F(1, 148) = 27.94$), and differed between visual field maps ($p < 10^{-10}$, $F(16, 148) = 10.44$). Post-hoc multiple comparisons revealed better model fits in the early visual field maps (V1, V2, V3, LO1) than elsewhere (Fig. 2C).

Numerosity tuned response model fits also differed between visual field maps ($p < 10^{-10}$, $F(16, 133) = 10.80$) in a similar ANOVA, but here post-hoc multiple comparisons showed poorer model fits in the early visual field maps than elsewhere. Also, unlike monotonically increasing Fourier power responses of early visual field maps, numerosity tuned response model fits did not differ significantly between near and far eccentricities ($p = 0.195$, $F(1, 133) = 1.70$). Therefore, progressions of tuned model fits with eccentricity were captured better by quadratic than sigmoid functions (Fig. S5C).

Monotonically decreasing responses outside early visual cortex

Some monotonically decreasing responses were seen outside early visual cortex. Average model fits across whole visual field maps did not differ significantly between monotonically decreasing and tuned numerosity models (one-way ANOVA, $p = 0.281$, $F(1,$

283) = 1.17). However, tuned models fit better (p = 0.0003, $F(1, 49)$ = 14.86) within the previously described numerosity maps. Monotonically decreasing models generally fit better just outside the numerosity maps (Fig. 2A). Such responses are predicted by computational models of numerosity-tuned response derivation (21), but at early stages preceding numerosity tuned responses, not alongside them. Alternatively, numerosity-tuned neural populations with preferred numerosities below one (33) would also decrease their responses as numerosity increases. These populations would be expected in the continuous topographic representation of numerosity, near populations with low numerosity preferences. We calculated the cortical distance between each recording site best fit by a numerosity tuned response and the nearest recording site best fit by a monotonically decreasing response. This distance was significantly positively correlated with the recording site's preferred numerosity (Spearman rank order correlation, $\rho = 0.585$, $p < 0.001$; see Fig. S11), suggesting monotonically decreasing sites are tuned, but with preferred numerosities below one.

Discussion

We found monotonic increases in neural population response amplitude with increasing numerosity in the retinotopic locations of our stimuli, beginning in V1. While these monotonic responses follow numerosity closely, they are better predicted by aggregate Fourier power, which follows numerosity closely over a wide range of stimulus parameters for a fixed contrast. Conversely, tuned responses overlapping with visual field maps in association cortices were not limited to the stimulus's retinotopic location, and were better predicted by tuning for numerosity than aggregate Fourier power. We also found monotonically decreasing responses to numerosity near recording sites tuned to low numerosities, likely reflecting tuning for numerosities below one.

Relationships between V1 responses, Fourier power and numerosity

Numerosity is generally seen as a high-level visual feature or cognitive property, while Fourier power is a low-level representation of image contrast. Contrast energy at specific orientations and spatial frequencies drives V1 neurons' responses (25). Therefore, the cortical response to any visual image begins with an approximate Fourier decomposition (25, 30). This transforms the visual image from the spatial domain (the image's projection onto the retina) to the spatial frequency domain (a neural representation of the spatial frequency, orientation and phase of contrast). In the spatial domain, it does not seem possible to estimate numerosity regardless of item size and spacing: the area that must be integrated and the luminance, contrast or edge extent within that area change with size and spacing. However, numerosity is not directly estimated from the visual image, it is estimated from the early visual image representation, which is in the spatial frequency domain. Many functions in the spatial domain correspond to other functions in the frequency domain, and numerosity (spatial domain) corresponds to a nonlinear function of aggregate Fourier power (frequency domain) at a fixed contrast. This makes it straightforward to estimate numerosity from the early visual, frequency domain image representation. Functional neuroimaging measures aggregate responses of large neural populations with a broad range of orientation and spatial frequency preferences (34). Aggregate Fourier power similarly sums contrast energy across orientations and spatial frequencies, and we propose this is why early visual neuroimaging responses follow numerosity with little effect of size or spacing: these early visual responses reflect aggregate Fourier power.

It is possible to generate phase-scrambled images that contain the same Fourier power distributions but with the locations of image contrast (an orthogonal phase component in the frequency domain) randomized. Such images yield strong responses in primary visual cortex, but even V2 responds poorly to such images as they lack the phase (position) structure found

in natural images (35). Hard edges contain contrast at many frequencies with linked phases. Analysis of this phase structure may be important for object individuation. Therefore, we expect phase (position) structure to be required for derivations of numerosity-tuned responses from early visual, frequency domain image representations. As such, derivation of numerosity from frequency domain image representations simplifies object normalization processes that are required to disregard size and spacing, but is still compatible with object individuation processes and may also simplify these.

We sum Fourier power only within the first harmonic. This is a clear local minimum in the Fourier spectrum, but its spatial frequency varies with item size. The visual system seems unlikely to flexibly identify this limit and aggregate responses within it as we do. We use this limit because the aggregate Fourier power in every harmonic (and so the total Fourier power to infinite spatial frequency) is proportional to the power in the first harmonic. For example, we show that the second harmonic's power also follows numerosity closely, at half the amplitude. However, using a discrete Fourier decomposition on a pixelated image we cannot evaluate further harmonics' power because they exceed the Nyquist frequency (a mathematical limit on which frequencies can be evaluated at a given sampling density). We therefore use a metric we can quantify that is equivalent to total Fourier power. The human visual system transforms images into the spatial frequency domain, but it does not use a discrete Fourier decomposition and its input is not pixilated: these are limitations imposed by computer models. Nevertheless, the visual system is still likely to have some spatial frequency limit. The finding that the aggregate response of early visual cortex is not affected by item size (22, 23) suggests it does not follow the Fourier power over a fixed spatial frequency range, but is proportional to aggregate Fourier power in the first harmonic and so to total Fourier power. It is unclear why this is so. It may be that the visual system samples the image very densely (certainly far more densely than 768×768 pixels as we do here) and little power falls beyond the frequencies it can

evaluate. Alternatively, neurons responding to a specific spatial frequency and its harmonics will typically be activated together and are likely to interact. Although the nature of these interactions is unclear, responses to higher harmonics may be suppressed, for example. Finally, there are certainly differences between the visual system's spatial frequency representation and a discrete Fourier decomposition, which is only a mathematical model. But it is clear that numerosity can be straightforwardly estimated from V1's population response, as it is from a Fourier transform. The properties of lab numerosity displays in a Fourier transform and the Fourier decomposition's close relationship to early visual spatial frequency analysis give an insight into how this could be.

Aggregate Fourier power within the first harmonic follows numerosity closely because it is approximately constant with item size and spacing. Why is this constant? Increasing item size moves Fourier power to a narrower range of lower spatial frequencies (Fig. S8). Its extent in the Fourier spectrum (bandwidth) follows 1/period, because spatial frequency is the inverse of spatial period. Conversely, Fourier power spectral density is high at these low frequencies because power spectral density of hard edges follows 1/frequency. Aggregate Fourier power, reflecting the product of bandwidth (1/period, i.e. frequency) and amplitude (1/frequency), is therefore constant. Similarly, a group's aggregate Fourier power is affected only slightly by its spacing. In Fourier decompositions position becomes a phase component that we do not analyze. Like item size, spacing does not affect Fourier power because increasing the distance between items moves power to lower frequencies, with lower bandwidths and higher amplitudes. Nevertheless, the ratio of item size and spacing affects which frequencies fall within the first harmonic (Fig. S9). If the distance between items is smaller than the item size, the between-item component of the Fourier spectrum falls outside the first harmonic, reducing power within this harmonic. Studies of numerosity perception avoid such crowding. If item spacing is very low (almost touching), the local minimum used to identify the first harmonic

reflects group size rather than item size. Aggregate power then reflects one item (the group) rather than the group's numerosity. Neither limitation would affect Fourier power if aggregated by the visual system over all frequencies.

On the other hand, item shape affects aggregate Fourier power considerably. Triangles have a greater aggregate Fourier power than other polygons (which have similar power to circles). Notably, their Fourier spectrum lacks a clear local minimum because a triangle's sides are so far from parallel (Fig. S10). Our procedure may therefore overestimate the first harmonic's extent. Conversely, most stars have less aggregate Fourier power than polygons and circles. Stars contain higher frequency features that fall beyond the first harmonic. Again, neither limitation would affect Fourier power if aggregated by the visual system over all frequencies.

Relationships to numerosity perception

Numerosity estimation from aggregate Fourier power may explain several known effects of stimulus properties on numerosity perception. First, increasing item spacing slightly increases Fourier power and increases perceived numerosity (16). Similarly, tightly crowding items together reduces Fourier power and reduces perceived numerosity (36–38). Second, blurring items decreases their range of spatial frequencies, and reduces perceived numerosity (37). Third, complex shapes disrupt the relationship between numerosity and Fourier power. Such shapes affect both perceived numerosity (31) and early visual response amplitudes (32). So, perceived numerosity, like V1 activation, depends on image properties.

Unlike Fourier decompositions, biological visual systems process different image locations with distinct neural populations. Our stimulus fell entirely in the central visual field. Humans can only integrate a limited spatial extent to estimate numerosity without making eye movements (39, 40). Very high numerosity stimuli must either be too large to see at a glance, be so dense that items are crowded, or use items too small to resolve. Aggregate Fourier power

is unaffected by such limitations, so follows numerosity closely until at least 175. Human vision perceives such high numerosities differently to lower numerosities (41). So, the human visual system may approximate a Fourier decomposition to transform the image into the frequency domain, but has its own limitations.

Previous studies of non-numerical features in numerosity stimuli have focused on total item perimeter, area, density, or pattern extent. These follow numerosity in some stimuli (17, 18), but any single feature can be kept constant across numerosities. Numerosity estimation in the spatial frequency domain moves beyond this approach because item size and spacing have little effect here. Nevertheless, complex shapes, crowding, blurring, phase scrambling and contrast variations disrupt the relationship between numerosity and aggregate Fourier power. These factors therefore allow strong tests that a human or animal is responding to numerosity rather than aggregate Fourier power alone, which would otherwise give the appearance of numerosity-guided behavior and may itself be beneficial.

Some researchers have seen texture or spatial frequency metrics as non-numerical image features that can be used to perform numerosity discrimination tasks without true numerosity perception (13, 14, 42). Numerosity's straightforward estimation from the early visual spatial frequency domain image representation does not imply that humans perceive aggregate Fourier power rather than numerosity, but instead shows how numerosity itself is straightforwardly estimated in the brain.

However, further processes are certainly involved in numerosity perception. There are situations where numerosity perception differs from true image numerosity. First, connecting items with lines or illusory contours (31, 43, 44) reduces perceived numerosity. This is generally thought to reflect higher-level grouping processes rather than image features (42). We show that lines or illusory contour inducers on items affect Fourier power, but it does not matter whether a connection between items is formed while reduced numerosity perception

requires a connection. This is consistent with early visual EEG event related potentials, which initially reflect numerosity with no effect of connectedness, and are only affected by connectedness later (32). So, the connectedness illusion does not affect numerosity perception at the stage of numerosity estimation and is likely to reflect later processes. Similarly, numerosity adaptation affects numerosity perception (45) without affecting image content. This affects numerosity-tuned neural responses (46) but it is unclear whether early visual responses are also affected. Adaptation to the rate of finger tapping also affects visual numerosity perception (47) , and this effect seems very unlikely to arise in early visual cortex. Higher-level effects and contextual effects on numerosity perception are expected considering the extensive network of numerosity-tuned responses in human association cortices, which successively transform the representation of numerosity and include areas involved in attention, multisensory integration and action planning (5, 48).

Our model for estimating numerosity from aggregate Fourier power is conceptually similar to a spatial frequency analysis that Dakin and colleagues proposed to explain why numerosity perception is affected by density in very dense displays (14). That analysis uses the responses of Laplacian Gaussian filters of high spatial frequency as a metric for numerosity and proposes that subjects perceive the response ratio of high and low spatial frequency filters, which is affected by density, rather than perceiving numerosity. Aggregate Fourier power is similarly affected by density. However, Dakin and colleagues' analysis was limited to items of a single size. Any such high spatial frequency filter's response is strongly affected by item size (Fig. S12A-C) while early visual cortex responses (22, 23), numerosity tuned responses (2, 4, 17) and numerosity perception (15, 16) are not. Therefore, this filter response predicts early visual responses poorly, as we have previously shown for numerosity-tuned responses (17, 18). Their response ratio metric (which they do not claim follows numerosity (Fig S12D)) also predicts early visual responses poorly, and humans rapidly and spontaneously perceive

numerosity (15, 16). Early visual responses follow numerosity with very low latency, and these responses are not preceded by responses to image features from which numerosity could be computed (22). This implies a single-stage process like we describe rather than the estimation and subsequent comparison of two measures. Finally, Dakin and colleagues' response ratio metric appears to predict the connectedness illusion, but it also predicts that perceived numerosity would be reduced by lines touching but not connecting items (Fig. S12E), which do not affect perceived numerosity as connecting lines do: the perceptual effects of item connectedness cannot be explained by either image analysis. Therefore, while we were inspired by their insight that numerosity must be estimated from early visual responses, both of Dakin and colleagues' proposed metrics for numerical vision predict neural responses and perception of numerosity poorly, particularly with respect to item size changes.

Numerosity-tuned responses

Models for subsequently computing numerosity-tuned responses rely on comparing monotonically increasing and decreasing responses (20, 21), their relative weights determining numerosity preferences. Human early visual population receptive fields have inhibitory surrounds (43), so populations with receptive fields further from stimulus area should monotonically decrease their responses with increasing numerosity. However, we observe no early visual monotonically decreasing responses, perhaps because negative fMRI responses have low amplitudes or fewer neurons with monotonically decreasing responses are needed (21). Using differently weighted excitatory and inhibitory synapses, tuned responses could be computed with monotonically increasing inputs only. So, early visual neural populations alone may provide sufficient inputs to derive numerosity-tuned responses.

This transformation from monotonic to tuned responses also seems to transform Fourier power to numerosity. Which further processes are needed to transform the early visual aggregate Fourier power response into a representation of numerosity that follows perceptual

properties? We did not manipulate image contrast (i.e. dot darkness) in our fMRI stimuli, but Fourier power linearly decreases as contrast is reduced. It is well established that early visual neural response amplitudes depend on image contrast (44), so we expect that reducing image contrast would reduce V1 response amplitudes. Therefore, early visual responses are unlikely to follow numerosity (as others have proposed) as numerosity does not depend on contrast, and more likely to follow aggregate Fourier power (as we propose) as this does depend on contrast. Transforming early visual Fourier power responses into numerosity tuned responses requires normalization for image contrast. Unlike early visual field maps, the responses of the first areas where we find numerosity tuned responses (visual field maps TO1 and TO2, i.e. area MT) are minimally effected by contrast (44). Therefore, contrast normalization at this stage is sufficient to yield contrast-independent numerosity tuned responses. Global image contrast changes do not affect perceived numerosity until items become less visible (45). But contrast variations within a display may disrupt global normalization processes, potentially underlying the reduction in perceived numerosity in contrast-varying displays (37).

After contrast normalization, converting from Fourier power to numerosity is equivalent to including an exponential nonlinearity, compensating for sub-additive accumulation of Fourier power with numerosity. Nonlinear interactions between excitatory and inhibitory inputs to numerosity-tuned populations seem sufficient to implement this. This nonlinearity is the main difference between predictions of monotonically increasing responses to Fourier power and numerosity in our stimulus set. As a compressive nonlinearity, this might reflect fMRI response amplitude saturation with increasing neural activity. Tuned responses would not show such saturation because they don't increase response amplitude with numerosity. We believe this interpretation is unlikely because our numerosity stimuli produce response amplitudes far below the early visual cortex's maximum fMRI response.

Unlike early visual monotonically increasing responses, numerosity tuned responses are not limited to neural populations with spatial receptive fields including the stimulus area. At the macroscopic scale, topographic maps of visual space and numerosity largely overlap, perhaps unsurprising as both are visually driven. This goes against the simplistic view of single brain areas having single functions. Indeed numerosity selectivity is found in many brain areas with diverse functions (5), many with spatial aspects: motion perception, spatial attention and eye movements (46–49). Numerical representations here may facilitate motion tracking, dividing attention and planning eye movements across multiple items respectively (5, 39, 40, 50, 51). But at a finer scale these response preferences are independent (5, 26), perhaps allowing neural responses to numerosity regardless of stimulus position (28). Linking specific numerosities and visual field positions would restrict all of these processes, and there is no link between particular numerosities and visual field positions in our stimuli or natural scenes. Conversely, to begin estimating numerosity from image contrast requires analysis of spatial responses in the stimulus area.

Conclusion

This distinction between the spatial selectivity of early visual monotonically responding populations and tuned populations in association cortices may reflect fundamental differences between the processes that estimate and use numerosity. Simpler animals like bees, zebrafish and newborn chickens display numerosity guided behaviors, but lack the complex numerical cognition of humans. Our findings suggest that apparently numerosity-guided behaviors in these animals may require only selective responses to orientation and spatial frequency in their simpler visual systems (52–54), with or without extracting numerosity from these spatial representations (55). The human brain appears to have built complex cognitive functions with inputs from these simple numerical responses (50).

Methods

Visual stimuli

Following experimental protocols described previously (4, 5, 26), we acquired 7T fMRI data while participants viewed numerosity and visual field mapping stimuli (see SI Methods). Written informed consent was obtained from all participants. All experimental procedures were approved by the ethics committee of University Medical Center Utrecht. Numerosity stimuli consisted of groups of black dots (items) within 0.75° (visual angle, radius) of a fixation cross at the center of the display. We used four stimulus configurations with different progressions of item size and spacing with numerosity (Fig. 1). No numerosity judgments were required. Numerosities progressed gradually in alternating ascending and descending sweeps.

Following previously described methods (17, 18), we quantified several non-numerical features of each display in each numerosity stimulus configuration. Additionally, for each display we determined the aggregate Fourier power within the first harmonic using MATLAB functions (executable code in SI Methods). We first normalized each image to a value of zero on the background and one in the items. We took a two-dimensional Fourier transform of the resulting image (fft2) and shifted the zero-frequency component to the center (fftshift) (Fig. 3A). We then took annular frequency bins (at all orientations) at intervals of 1 cycle/image (excluding zero frequency), in which we summed the absolute power spectral density (PSD) of the Fourier transform. Plotting PSD across frequencies (Fig. 3B) identifies a clear local minimum at the end of the first harmonic. We identified this by finding the lowest frequency above the frequency of the global maximum PSD where: either the first or second derivative of PSD reaches its maximum and the PSD is below 25% of the global maximum PSD. This effectively identifies the sharpest change in PSD, i.e. the end of the first harmonic, avoiding artefacts resulting from the image pixilation or Fourier transform discretization. We summed PSDs of all frequency bins until this frequency to give the aggregate Fourier power in the first

harmonic for each display. We also determined how well the relationship between aggregate Fourier power and numerosity generalizes across a much greater range of item sizes, spacings and shapes that were not included in our fMRI experiments (See SI Methods). Display resolution affects Fourier power. We arbitrarily evaluated our stimuli's power in 768×768-pixel images but divided it by $768 \times 768 \times 2^{0.5}$ to generalize to any resolution. This makes aggregate Fourier power in the first harmonic very nearly one for one-item displays.

Neural response models

For each stimulus configuration's data and cross-validation splits, we fit several candidate response models testing how well a putative neural population with monotonic or tuned responses to each stimulus feature predicts the observed responses. For monotonic response models, we convolved the sequence of presented feature amplitudes with a canonical hemodynamic response function (HRF) to give a predicted fMRI time course. We found the optimal scaling for this prediction using a general linear model (Fig. S1) to determine the baseline and β (scaling factor) for each recording site. This revealed whether responses increased or decreased with feature amplitude (the sign of β) and the response variance explained by the scaled prediction (R^2). Log(numerosity) predicts early visual EEG (22, 29, 56) and fMRI (23) response amplitudes well. To give each non-numerical feature the best possible chance of explaining these responses, we tested both logarithmic and linear scaling and chose that which gave the highest variance explained across all recording sites and participants. For each non-numerical feature, the best performing model (logarithmic or linear) was consistent with that chosen to best predict numerosity tuned responses (17, 57).

We estimated tuned response models as previously described (4, 5, 17), following a pRF modelling approach (58). For tuned response models, we only compared predictions of tuning for numerosity and aggregate Fourier power: we have previously demonstrated that numerosity tuning predicts responses better than tuning for other features (5, 17). Briefly, for

each recording site, we start with a large set of candidate neural response models describing numerosity or Fourier power tuning using logarithmic Gaussian functions (3–5) characterized by two parameters: (1) a preferred numerosity or Fourier power (mean of the Gaussian distribution) and (2) a tuning width (standard deviation of the Gaussian). Candidate neural response time courses reflected the amplitude of the Gaussian at the numerosity or Fourier power presented at each time point. We convolved these with an HRF to generate candidate fMRI time courses. We found the candidate fMRI time course best correlated with the response at each recording site, giving the neural response model parameters that best predict that recording site's response and the response variance explained by that prediction (R^2).

We restricted candidate preferred numerosities to be between 1.05–6.95, within the range of numerosities presented. Beyond the tested range, tuning function parameters cannot be determined accurately (4, 5, 58) and neural response amplitude predictions monotonically increase or decrease within the presented numerosity range, complicating comparisons to monotonic models.

Analysis of relationship between spatial and numerical responses

We also analyzed how variance explained by the best fitting response models was related to recording sites' spatial pRFs (see SI Methods). We use pRF eccentricity (distance from the visual field center, where numerosity stimuli were presented) to quantify the pRF's coverage of the stimulus area: pRF size linearly follows eccentricity (58, 59), so pRFs with shared eccentricities cover the numerosity stimulus area similarly. For each visual field map, we binned all participants' recording sites by eccentricity. We fit relationships between bin mean variance explained and eccentricity using a sigmoid cumulative Gaussian function (i.e. variance explained decreasing with eccentricity) and a quadratic function (allowing many relationships between variance explained and eccentricity). For each visual field map and response model, we chose the function that best fit the bin means.

Acknowledgments and funding sources

This work was supported by the Netherlands Organization for Scientific Research (452.17.012 to B.M.H.); and the Portuguese Foundation for Science and Technology (IF/01405/2014 to B.M.H.). We thank Serge Dumoulin for sharing data collected by his lab.

Author contributions:

Conceptualization, J.M.P. and B.M.H.; Methodology, B.M.H.; Software, J.M.P., M.A., and B.M.H.; Validation, J.M.P. and B.M.H.; Formal Analysis, J.M.P. and B.M.H.; Investigation, J.M.P. and B.M.H.; Data Curation, J.M.P., M.A., T.C.C., and B.M.H.; Writing – Original Draft, J.M.P. and B.M.H.; Writing – Review & Editing, J.M.P., M.A., T.C.C., and B.M.H.; Visualization, J.M.P. and B.M.H.; Supervision, B.M.H.; Project Administration, B.M.H.; Funding Acquisition, B.M.H.

References

1. H. M. Ditz, A. Nieder, Sensory and working memory representations of small and large numerosities in the crow endbrain. *J. Neurosci.* **36**, 12044–12052 (2016).
2. A. Nieder, D. J. Freedman, E. K. Miller, Representation of the quantity of visual items in the primate prefrontal cortex. *Science* **297**, 1708–1711 (2002).
3. A. Nieder, E. K. Miller, Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* **37**, 149–157 (2003).
4. B. M. Harvey, B. P. Klein, N. Petridou, S. O. Dumoulin, Topographic representation of numerosity in the human parietal cortex. *Science* **341**, 1123–1126 (2013).
5. B. M. Harvey, S. O. Dumoulin, A network of topographic numerosity maps in human association cortex. *Nat. Hum. Behav.* **1** (2017).
6. M. Piazza, V. Izard, P. Pinel, D. Le Bihan, S. Dehaene, Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron* **44**, 547–555 (2004).
7. E. Eger, P. Pinel, S. Dehaene, A. Kleinschmidt, Spatially invariant coding of numerical information in functionally defined subregions of human parietal cortex. *Cereb. Cortex* **25**, 1319–1329 (2015).
8. S. Cavdaroglu, A. Knops, Evidence for a Posterior Parietal Cortex Contribution to Spatial but not Temporal Numerosity Perception. *Cereb. Cortex* **29**, 2965–2977 (2019).
9. E. Castaldi, M. Piazza, S. Dehaene, A. Vignaud, E. Eger, Attentional amplification of neural codes for number independent of other quantities along the dorsal visual stream. *Elife* **8** (2019).
10. O. Tudusciuc, A. Nieder, Neuronal population coding of continuous and discrete quantity in the primate posterior parietal cortex. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14513–14518 (2007).

11. A. J. Kersey, J. F. Cantlon, Neural tuning to numerosity relates to perceptual tuning in 3– 6-year-old children. *J. Neurosci.* **37**, 512–522 (2017).

12. G. Lasne, M. Piazza, S. Dehaene, A. Kleinschmidt, E. Eger, Discriminability of numerosity-evoked fMRI activity patterns in human intra-parietal cortex reflects behavioral numerical acuity. *Cortex* **114**, 90–101 (2019).

13. F. H. Durgin, Texture density adaptation and visual number revisited. *Curr. Biol.* **18**, R855–R856 (2008).

14. S. C. Dakin, M. S. Tibber, J. A. Greenwood, F. A. A. Kingdom, M. J. Morgan, A common visual metric for approximate number and density. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 19552–19557 (2011).

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

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

17. B. M. Harvey, S. O. Dumoulin, Can responses to basic non-numerical visual features explain neural numerosity responses? *Neuroimage* **149**, 200–209 (2017).

18. B. M. Harvey, S. O. Dumoulin, Data describing the accuracy of non-numerical visual features in predicting fMRI responses to numerosity. *Data Br.* **16**, 193–205 (2018).

19. I. Stoianov, M. Zorzi, Emergence of a “visual number sense” in hierarchical generative models. *Nat. Neurosci.* **15**, 194–196 (2012).

20. S. Dehaene, J. P. Changeux, Development of elementary numerical abilities: A neuronal model. *J. Cogn. Neurosci.* **5**, 390–407 (1993).

21. T. Verguts, W. Fias, Representation of number in animals and humans: A neural model. *J. Cogn. Neurosci.* **16**, 1493–1504 (2004).

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

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

24. A. van Rinsveld, *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. U. S. A.* **117**, 5726–5732 (2020).

25. D. H. Hubel, T. N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–154 (1962).

26. B. M. Harvey, A. Fracasso, N. Petridou, S. O. Dumoulin, Topographic representations of object size and relationships with numerosity reveal generalized quantity processing in human parietal cortex. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13525–13530 (2015).

27. B. M. Harvey, S. Ferri, G. A. Orban, Comparing Parietal Quantity-Processing Mechanisms between Humans and Macaques. *Trends Cogn. Sci.* **21**, 779–793 (2017).

28. P. Viswanathan, A. Nieder, Spatial neuronal integration supports a global representation of visual numerosity in primate association cortices. *J. Cogn. Neurosci.* **32**, 1184–1197 (2020).

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

30. J. G. Daugman, Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Res.* **20**, 847–856 (1980).

31. L. He, J. Zhang, T. Zhou, L. Chen, Connectedness affects dot numerosity judgment: Implications for configural processing. *Psychon. Bull. Rev.* **16**, 509–517 (2009).

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

33. A. Ramirez-Cardenas, M. Moskaleva, A. Nieder, Neuronal representation of numerosity zero in the primate parieto-frontal number network. *Curr. Biol.* **26**, 1285–1294 (2016).

34. K. N. Kay, T. Naselaris, R. J. Prenger, J. L. Gallant, Identifying natural images from human brain activity. *Nature* **452**, 352–355 (2008).

35. J. Freeman, C. M. Ziemba, D. J. Heeger, E. P. Simoncelli, J. A. Movshon, A functional and perceptual signature of the second visual area in primates. *Nat. Neurosci.* **16**, 974–981 (2013).

36. G. Anobile, M. Turi, G. M. Cicchini, D. C. Burr, Mechanisms for perception of numerosity or texture-density are governed by crowding-like effects. *J. Vis.* **15**, 1–12 (2015).

37. M. J. Morgan, S. Raphael, M. S. Tibber, S. C. Dakin, A texture-processing model of the "visual sense of number." *Proc. R. Soc. B Biol. Sci.* **281**, 1–9 (2014).

38. J. Ross, D. Burr, Number, texture and crowding. *Trends Cogn. Sci.* **16**, 196–197 (2012).

39. J. M. Paul, R. A. Reeve, J. D. Forte, Taking a(c)count of eye movements: Multiple mechanisms underlie fixations during enumeration. *J. Vis.* **17** (2017).

40. J. M. Paul, R. A. Reeve, J. D. Forte, Enumeration strategy differences revealed by saccade-terminated eye tracking. *Cognition* **198**, 104204 (2020).

41. G. Anobile, G. M. Cicchini, D. C. Burr, Number As a Primary Perceptual Attribute: A Review. *Perception* **45**, 5–31 (2016).

42. A. Adriano, L. Girelli, L. Rinaldi, Non-symbolic numerosity encoding escapes spatial frequency equalization. *Psychol. Res.* (2021) https:/doi.org/10.1007/s00426-020-01458-2.

43. W. Zuiderbaan, B. M. Harvey, S. O. Dumoulin, Modeling center-surround

configurations in population: Receptive fields using fMRI. *J. Vis.* **12**, 1–15 (2012).

44. S. Kastner, M. A. Pinsk, Visual attention as a multilevel selection process. *Cogn. Affect. Behav. Neurosci.* **4**, 483–500 (2004).

45. M. Palomares, H. Egeth, How element visibility affects visual enumeration. *Vision Res.* **50**, 2000–2007 (2010).

46. M. A. Silver, D. Ress, D. J. Heeger, Topographic maps of visual spatial attention in human parietal cortex. *J. Neurophysiol.* **94**, 1358–1371 (2005).

47. K. Amano, B. A. Wandell, S. O. Dumoulin, Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. *J. Neurophysiol.* **102**, 2704–2718 (2009).

48. S. O. Dumoulin, *et al.*, A new anatomical landmark for reliable identification of human area V5/MT: A quantitative analysis of sulcal patterning. *Cereb. Cortex* **10**, 454–463 (2000).

49. D. J. Hagler, M. I. Sereno, Spatial maps in frontal and prefrontal cortex. *Neuroimage* **29**, 567–577 (2006).

50. M. Piazza, Neurocognitive start-up tools for symbolic number representations. *Trends Cogn Sci* **14**, 542–551 (2010).

51. A. Knops, M. Piazza, R. Sengupta, E. Eger, D. Melcher, A shared, flexible neural map architecture reflects capacity limits in both visual short-term memory and enumeration. *J. Neurosci.* **34**, 9857–9866 (2014).

52. J. R. Dyhr, C. M. Higgins, The spatial frequency tuning of optic-flow-dependent behaviors in the bumblebee Bombus impatiens. *J. Exp. Biol.* **213**, 1643–1650 (2010).

53. O. Rinner, J. M. Rick, S. C. F. Neuhauss, Contrast sensitivity, spatial and temporal tuning of the larval zebrafish optokinetic response. *Investig. Ophthalmol. Vis. Sci.* **46**, 137–142 (2005).

54. R. Over, D. Moore, Spatial acuity of the chicken. *Brain Res.* **211**, 424–426 (1981).

55. H. MaBouDi, *et al.*, Non-numerical strategies used by bees to solve numerical cognition tasks. *Proc. R. Soc. B Biol. Sci.* **288**, 20202711 (2021).

56. M. Fornaciai, J. Park, Distinct neural signatures for very small and very large numerosities. *Front. Hum. Neurosci.* **11** (2017).

57. B. M. Harvey, S. O. Dumoulin, Data describing the accuracy of non-numerical visual features in predicting fMRI responses to numerosity. *Data Br.* **16**, 193–205 (2018).

58. S. O. Dumoulin, B. A. Wandell, Population receptive field estimates in human visual cortex. *Neuroimage* **39**, 647–660 (2008).

59. B. M. Harvey, S. O. Dumoulin, The relationship between cortical magnification factor and population receptive field size in human visual cortex: Constancies in cortical architecture. *J. Neurosci.* **31**, 13604–13612 (2011).

Supplementary Information for:

Transformation from local image contrast to location-independent numerosity tuning in human extrastriate cortex

Jacob M. Paul[1*], Martijn van Ackooij[1], Tuomas C. ten Cate[1] & Ben M. Harvey[1]

[1] Experimental Psychology, Helmholtz Institute, Utrecht University, Heidelberglaan 1, Utrecht 3584 CS, Netherlands

* Corresponding author:
Jacob M. Paul
Experimental Psychology, Helmholtz Institute, Utrecht University
Heidelberglaan 1
Utrecht 3584 CS
Netherlands

SI Methods

Participants

We acquired fMRI data from eleven participants (aged 25–39 years, one female, one left-handed). All had normal or corrected-to-normal visual acuity, good mathematical abilities and were well educated. Participants were familiarized with the numerosity stimuli using tasks that required numerosity judgements before scanning.

Data from participants P1-P5 were included in a previous study (1), although we use updated preprocessing protocols here. These participants were scanned while viewing all four stimulus configurations (Fig. 1A), to test for responses to non-numerical features over a broad range of stimulus parameters. P6-P11 were only scanned while viewing the constant area and constant size configurations: previous studies show very similar responses across stimulus configurations, and non-numerical features that differ between configurations predict numerosity selective (2) and monotonic (3) responses poorly.

Visual stimuli

Following experimental protocols described elsewhere (4, 5), numerosity stimuli were presented by back-projection on a 15×9 cm screen inside the MRI bore, which participants viewed through prism glasses and a mirror attached to the head coil. Distance from the participants' eyes to the display was 41 cm, with a visible resolution of 1,024×538 pixels.

Visual stimuli were generated using MATLAB and Psychtoolbox-3 (6). Stimuli were sets of black circles (items) on a middle gray background. A diagonal cross of thin red lines intersected the center of the screen and covered the entire display throughout the experiment to aid with accurate fixation.

We used four stimulus configurations with different progressions of item size and spacing with numerosity (Fig. 1). The first three kept total item area (and display luminance), individual item size, or total item perimeter constant. These placed items randomly but

approximately homogenously within the stimulus area. The fourth ("high density") grouped the items constant total item area entirely inside a 0.375° radius circular area, randomly placed inside the stimulus area. 10% of displays showed white items instead of black, and participants responded to these displays with a button press (80-100% accurately in each run).

The numerosities one through seven were first presented in ascending order, with numerosity changing every 4200 ms (two TRs, see Fig. S1C-D inset). Within this period, a numerosity pattern was shown for 300 ms, alternating with 400 ms of gray background, repeated six times. Each numerosity pattern had items drawn in new, random positions. These short presentations prevented participants from counting. Following this, twenty items were presented in the same way for eight TRs (16.8s, 24 presentations of a pattern). These periods of twenty items served to distinguish between very small and very large tuning widths (4, 7). Then numerosities one through seven were presented as before, but in descending order, followed by another long period of twenty items. This cycle was repeated 4 times in each scanning run.

Repeated presentation of stimuli with the same numerosity likely caused some adaptation (8) or repetition suppression. To minimize effects of adaptation on estimated response models, we used a single model to summarize responses to both increasing and decreasing numerosity. This ensured each numerosity presentation was preceded by stimuli that caused both higher and lower responses, reducing dependence on preceding stimuli.

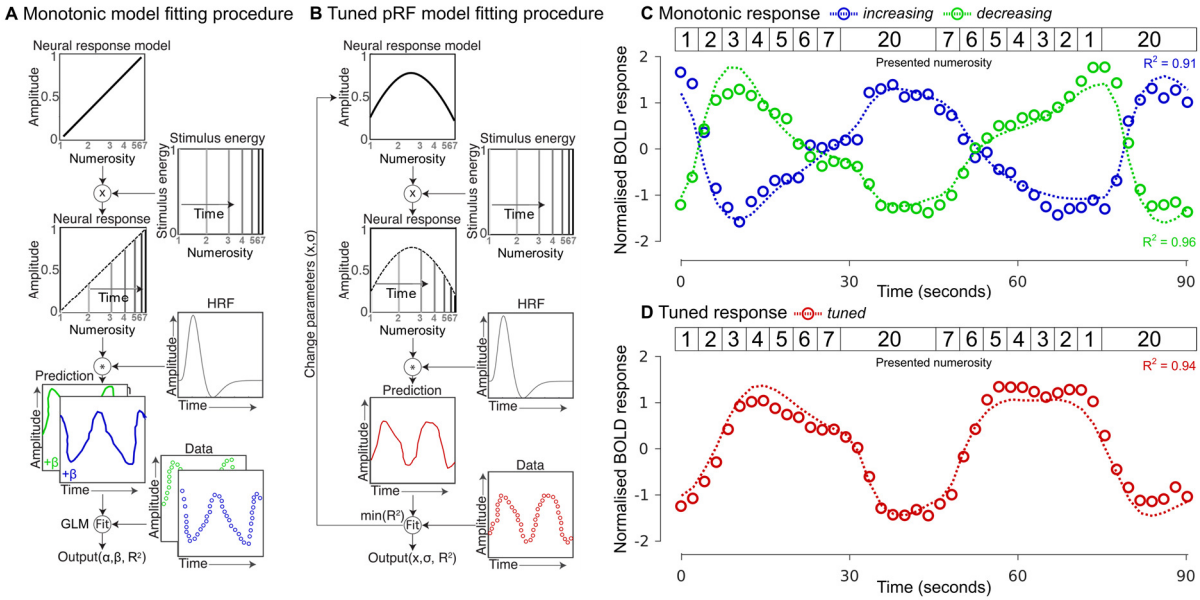

Fig. S1. Model fitting. (A, B) Monotonic model and population receptive field (pRF) fitting procedures for estimating monotonic and tuned response functions for numerosity, respectively. (C, D) Illustrative neural response time course (circles) and model prediction (dashed line) for three separate voxels best described by preferences for monotonically increasing (blue), monotonically decreasing (green) or tuned (red) responses to numerosity. Similar models were used to test predictions of response models based on non-numerical stimulus features. Numerosity varied in ascending and descending order for all stimulus configurations to counterbalance adaptation effects (see stimulus inset).

In a separate scanning session, visual field mapping was used to delineate visual field maps and determine the position selectivity of our recording sites, following protocols described previously (7, 9). Briefly, a bar filled with a moving checkerboard pattern stepped across a 6.35° (radius) circle in the display center in eight (cardinal and diagonal) directions. Participants fixated the same central fixation cross, pressing a button when this changed color to ensure fixation and attention.

fMRI acquisition and data pre-processing

We acquired MRI data on a 7T Philips Achieva scanner, following protocols described fully in our previous studies (1, 5, 10). Briefly, we acquired T1-weighted anatomical scans, automatically segmented these with Freesurfer, then manually edited labels to minimize segmentation errors using ITK-SNAP. This provided a highly accurate cortical surface model at the grey-white matter border to characterize cortical organization. We acquired T2*-

weighted functional images using a 32-channel head coil at a resolution of 1.77×1.77×1.75 mm, with 41 interleaved slices of 128×128 voxels. The resulting field of view was 227×227×72 mm. TR was 2100 ms, TE was 25 ms, and flip angle was 70 degrees. We used a single shot gradient echo sequence with SENSE acceleration factor 3.0 and anterior-posterior encoding. This covered most of the brain but omitted anterior frontal and temporal lobes, where 7T fMRI has low response amplitudes and large spatial distortions. Each functional run acquired 182 images (382.2s) of which the first six (12.6s) were discarded to ensure a steady signal state. In each scan session, we acquired six to eight repeated runs in one stimulus configuration, plus a top-up scan with the opposite phase-encoding direction to correct for image distortion in the gradient encoding direction (11), and a T1-weighted anatomical image with the same resolution, position and orientation as the functional data. Different stimulus configurations were tested in different sessions.

Co-registration of functional data to the high-resolution anatomical space was performed using AFNI (afni.nimh.nih.gov; (12)), which differs from our previous studies. A single transformation matrix was constructed, incorporating all the steps from the raw data to the cortical surface model to reduce the number of interpolation steps to one. No other spatial or temporal smoothing procedures were applied. A T1 image with the same resolution, position and orientation as the functional data was first used to determine the transformation to a higher resolution (1 mm isotropic) whole-brain T1 image (3dUnifize, 3dAllineate). For the fMRI data, we first applied motion correction to two series of images that were acquired using opposing phase-encoding directions (3dvolreg). Subsequently, we determined the distortion transformation between the average images of these two series (3dQwarp). We then determined the transformation in brain position between and within functional scans (3dNwarpApply). Then we determined the transformation that co-registers this functional data to the T1 acquired in the same space (3dvolreg). We applied the product of all these transformations to every

functional volume to transform our functional data to the whole-brain T1 anatomy. We repeated this for each fMRI session to transform all their data to the same anatomical space.

We then imported these data into mrVista (github.com/vistalab/vistasoft). We averaged visual field mapping scan runs. We averaged scans in the same numerosity stimulus configuration together, and also averaged data from odd and even runs from the same configuration for cross-validation. Within these averaged runs, we averaged repeated stimulus cycles.

Analyses of visual features co-varying with numerosity

We also examined whether responses to other low-level non-numerical features that co-varied with numerosity could better explain our data, using methods analogous to those described fully in (2). For monotonic models, we tested predictions of responses monotonically increasing or decreasing linearly or logarithmically (whichever fit best) with: [1] numerosity; individual item [2] area and [3] perimeter; total item [4] area (13) and [5] perimeter (14); [6] area and [7] perimeter of the convex hull (15); density of [8] luminance, [9] edges, and [10] number within the convex hull; root mean square contrast [11] across the display and [12] in the convex hull; and [13] aggregate Fourier power. We also tested a model proposed to capture numerosity perception using the responses of a high spatial frequency Laplacian of Gaussian center-surround filter (16). We used a filter with a 2-pixel standard deviation here, but other filter sizes made very similar predictions (Fig. S12). We convolved this filter with each image and summed the response over all image locations. We also tested the ratio of responses of high and low spatial frequency filters (the response ratio model (16)). We used filters with standard deviations of 1 and 34 pixels here as the prediction of this filter pair was most strongly correlated with numerosity in our displays. We also tested spatial pRF models following the visual positions of [14] edges and [15] luminance, summing these across all displays of the same numerosity.

We used the following MATLAB code to determine the aggregate Fourier power in the first harmonic, from images placed in the image buffer using Psychtoolbox-3 (6):

```
ItemColor = 0; %Intensity values in the first color channel
BackgroundColor = 128;
GetImageSize = 768; %Image must be square, must be an even number of pixels

%Get image from image buffer
NumImage = Screen('GetImage', display.windowPtr, [display.numpixels(1)/2-
GetImageSize/2 display.numpixels(2)/2-GetImageSize/2
display.numpixels(1)/2+GetImageSize/2
display.numpixels(2)/2+GetImageSize/2]);

%Normalise image so background is zero and item is one
NumImage = (BackgroundColor-single(NumImage(:,:,1)))./(BackgroundColor-
ItemColor);

%Take 2D Fourier transform, shift zero frequency component to centre
FourierImage = abs(single(fftshift(fft2(NumImage, GetImageSize,
GetImageSize))));

%Sum over all orientations for each frequency
[x,y] = meshgrid(-GetImageSize/2:GetImageSize/2-1, -GetImageSize/2:
GetImageSize/2-1);
[~,Radius] = cart2pol(x,y);
Radius = round(Radius);
FreqBins = zeros(1,(GetImageSize/2));
for RadCounter = 1:(GetImageSize/2)
      FreqBins(RadCounter) = sum(FourierImage(Radius==RadCounter));
end

%Find the limit of the first harmonic. For robustness, use the 1st and 2nd
derivative, and use whichever returns the lowest frequency.
[maxVal, maxPos] = max(FreqBins);
deriv1 = [diff(FreqBins)];
deriv1(FreqBins>(maxVal/4)) = 0;
deriv1(1:maxPos) = 0;
[~, whichMin1] = max(deriv1);

deriv2 = [0 diff(diff(FreqBins))];
deriv2(FreqBins>(maxVal/4)) = 0;
deriv2(1:maxPos) = 0;
[~, whichMin2] = max(deriv2);
whichMin = min([whichMin1 whichMin2]);

%Sum to this limit to get the aggregate Fourier power in the first harmonic
AggregatePower = sum(FreqBins(1:whichMin));

%Compensate for the image resolution
AggregatePower = AggregatePower/(GetImageSize^2*sqrt(2));

%To convert this to a numerosity estimate if preferred. This is somewhat
approximate and descriptive of the observed relationship.
NumerosityEstimate = AggregatePower^2.2;
```

We also determined how well the relationship between aggregate Fourier power and numerosity generalizes across a much greater range of item sizes, spacings and shapes that were not included in our fMRI experiments. We tested single circles from 1-240 pixels diameter placed at the image center. We then tested images of 7 circles, each of diameter 16 pixels, spaced randomly but evenly within a 50 to 528 pixel diameter circular area. The minimum center-to-center distance increased with the group area. We then tested images of 16-pixel diameter circles from numerosity 1-175 within a 700-pixel diameter circular area, placed at least 24 pixels apart (edge-to-edge).

We then tested regular convex polygons from 3 corners (equilateral triangle) to 10 corners (regular decagon), plus a circle (i.e. infinite corners), with all corners 30 pixels (radius) from the image center. We also tested regular stars with 3 to 10 points (at 30 pixels radius), interleaved with concave corners at 10 pixels radius. Spatial frequency differs with orientation in these shapes, so we determined the frequency of the first harmonic separately at each orientation and summed the PSD within the resulting area.

We finally tested pairs of dots with and without connecting elements to determine whether the effects of these connecting elements on Fourier power predict their effects on perception. All dots we 30 pixels in diameter. All bars were 4 pixels wide and the same length as the (edge-to-edge) distance between dots. Illusory contour inducers were 4 pixels wide, the same color as the background, and extending 10 pixels into the dot. To separate effects of connected configurations from changes to image properties, we included conditions where the bar or illusory contour inducer was split in the middle and rotated around the dot center. As well as analyzing the aggregate Fourier power of these images, we evaluated the predictions of the response ratio model for numerosity perception which has previously been shown to predict some perceptual effects of connectedness. Here we used 2-pixel and 23-pixel standard deviations for the high and low frequency filters respectively as previously described (16).

To model predictions of the mean fMRI response to aggregate Fourier power (or any other non-numerical feature) we took the mean value over all displays of the same numerosity and stimulus configuration. We calculated the proportion of variance in the responses to all stimulus configurations (Area, Size, Perimeter, Density) that these models explained, rather than simply averaging the variance explained values from each stimulus configuration. This ensured that configurations with larger fMRI signal variance were weighted more heavily in the average across stimulus configurations. Arbitrary differences in fMRI response amplitude between scan sessions (17) prevented us from quantifying differences in neural response amplitude between stimulus configurations.

We compared these model fits under cross-validation between odd and even scanning runs per condition: we fit the model's free parameters on one half of the data and evaluated the resulting model's fit on the complementary half. This primarily compensates for the free parameters in the visual position selective pRF models and the tuned models: the monotonic models have no free parameters. Using cross-validated model fits allows an unbiased comparison between the monotonic model fits, visual position selective pRF model fits, and tuned model fits. We refit the scaling factor to evaluate the prediction, as scaling factors change between scans and sessions, but did not allow the scaling factor of monotonic models to change sign.

We then compared the cross-validated variance explained in visual field maps V1-V3 between the monotonic numerosity model and alternative monotonic and visual position selective pRF models. We first excluded from this comparison any recording site where no model explained more than 20% of the response variance. We then took the mean cross-validated fit from each visual field map, in each hemisphere, and from both cross-validation halves. We compared the resulting averaged model fits using Wilcoxon signed-rank tests between each pair of candidate response models. We used the same procedure to compared the

cross-validated variance explained in the numerosity maps between the numerosity-tuned model and aggregate Fourier power tuned model, as we have previously demonstrated that the numerosity-tuned model fits better than models tuned to other non-numerical visual features.

Visual field mapping analysis

Visual field mapping data were analyzed following a standard pRF analysis (7, 9). We identified visual field map borders based on reversals in the cortical progression of the polar angle of recording sites' visual field position preferences, drawing these by hand on the cortical surface of an inflated rendering of each participant's brain (Fig. 2A and Fig. S3-S4). As well as the early visual field maps (V1, V2, V3), we identified higher visual field maps in the lateral/temporal occipital (LO1-LO2, TO1-TO2), parietal association (V3A/B, IPS0-IPS5) and frontal (sPCS1-sPCS2, iPCS) cortices with reference to landmarks identified in previous studies (18–21).

We binned recording sites in each visual field map by pRF eccentricity, at 0.2˚ intervals from 0˚ to 5.5˚. We took the mean and standard error of the variance explained by each response model among the recording sites within each bin in plotted this against bin eccentricity. We fit relationships between bin mean variance explained in each numerosity response model and eccentricity using cumulative Gaussian sigmoid functions with 4 parameters: the point of inflection (cumulative Gaussian mean), slope (standard deviation), maximum asymptote (variance explained at fixation) and minimum asymptote (variance explained at 5.5˚ eccentricity). We also fit quadratic curves to these, with three parameters: intercept, slope and quadratic term. This includes the possibility of a constant relationship, linear relationship, and more complex relationships that allow for an increase in variance explained in the middle of the eccentricity range, which was often observed. For both fit functions, we bootstrapped these fits using 1000 samples drawn from the unbinned variance explained data across all participants. We took the median of these bootstrapped fit parameters as the best fitting curve.

We determined 95% confidence intervals from the 2.5 and 97.5 percentile from the distribution of bootstrapped fits at each eccentricity. For each visual field map and response model, we chose the function that gave the best correlation between the fit and the bin means.

SI Results

Defining visual field maps and comparing monotonic vs. tuned numerosity model fit

Eccentricity preferences of each recording site identified from the visual field mapping task for all participants are shown in Fig. S2. Likewise, polar angle preferences of each recording site also identified from the visual field mapping task for all participants are shown in Fig. S3. Reversals in polar angle preferences were used to define visual field map borders. Whole-brain comparisons of cross-validated model fits (R^2) between monotonically increasing, monotonically decreasing, and tuned responses to numerosity for all participants are shown in Fig. S4.

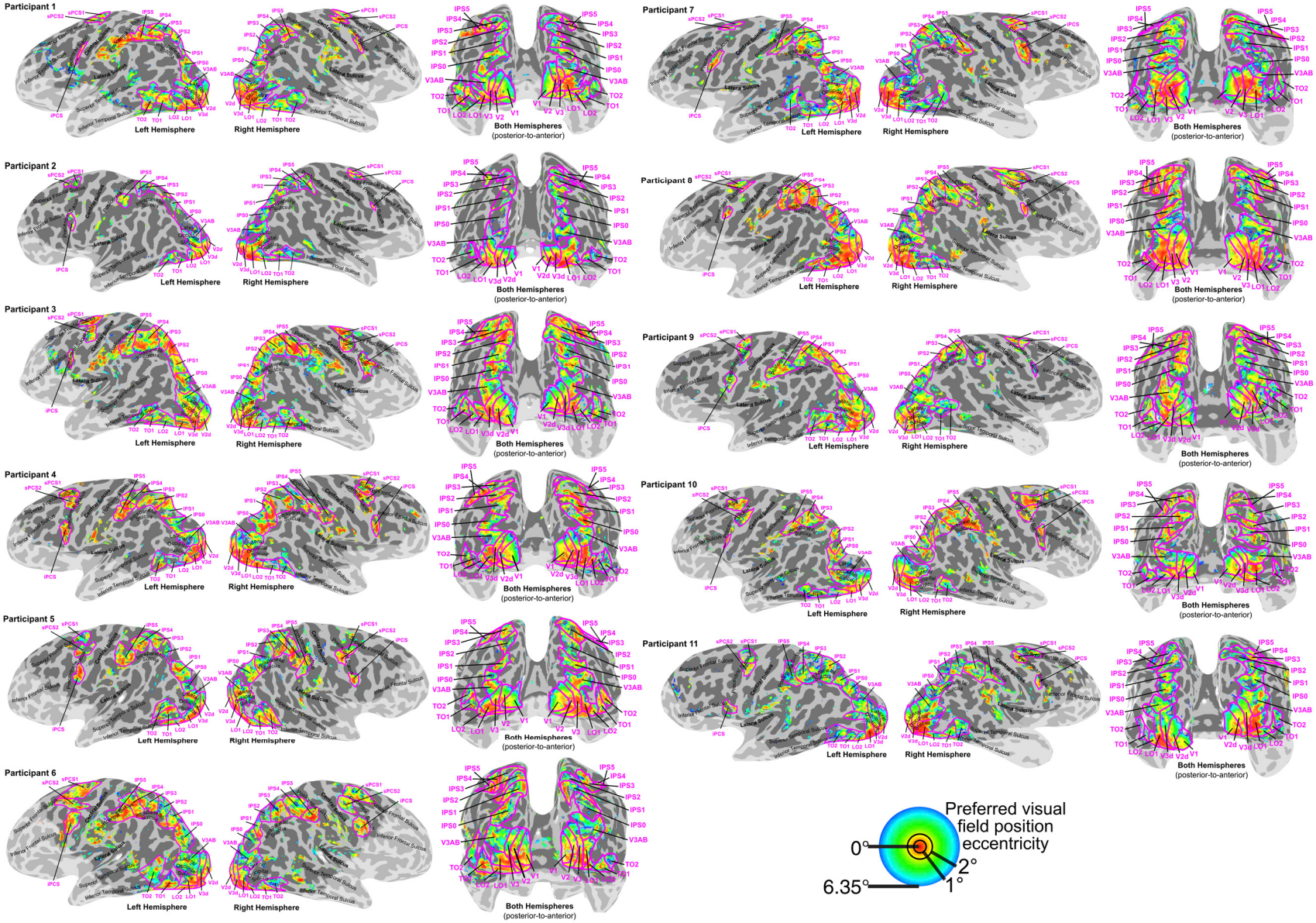
Participant 1
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 2
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 3
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 4
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 5
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 6
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 7
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 8
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 9
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 10
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 11
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Preferred visual
field position
eccentricity
0°
2°
1°
6.35°

Fig. S2. Preferred visual field position eccentricities determined from the visual field mapping task. All colored sites have greater than 10% model variance explained. Magenta lines show visual field map borders and magenta labels highlight visual field map names. The light shaded region is outside the fMRI recording volume.

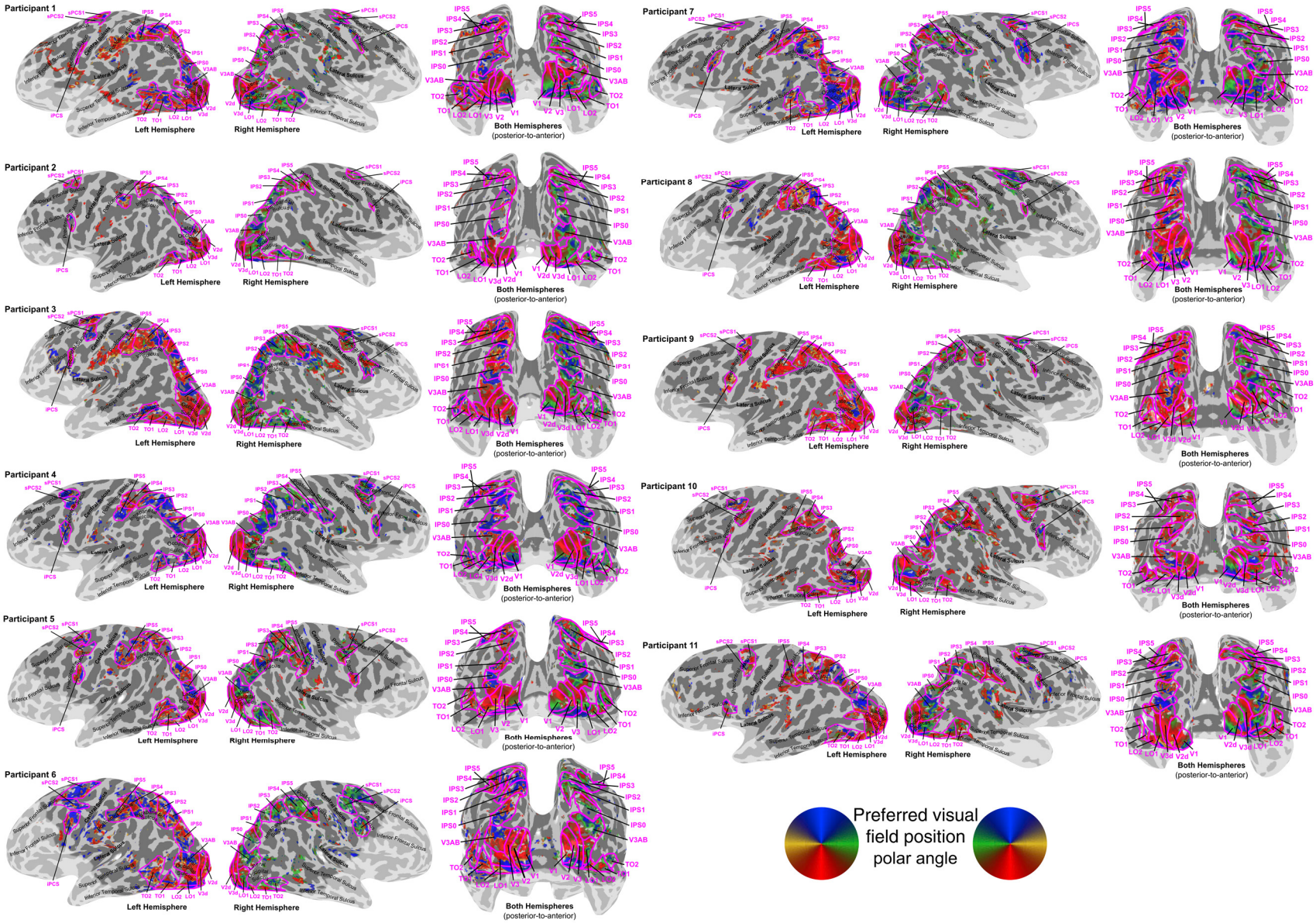
Participant 1
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 2
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 3
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 4
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 5
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 6
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 7
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 8
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 9
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 10
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 11
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Preferred visual
field position
polar angle

Fig. S3. Preferred visual field position polar angles determined from the visual field mapping task. All colored sites have greater than 10% model variance explained. Magenta lines show visual field map borders and magenta labels highlight visual field map names. The light shaded region is outside the fMRI recording volume.

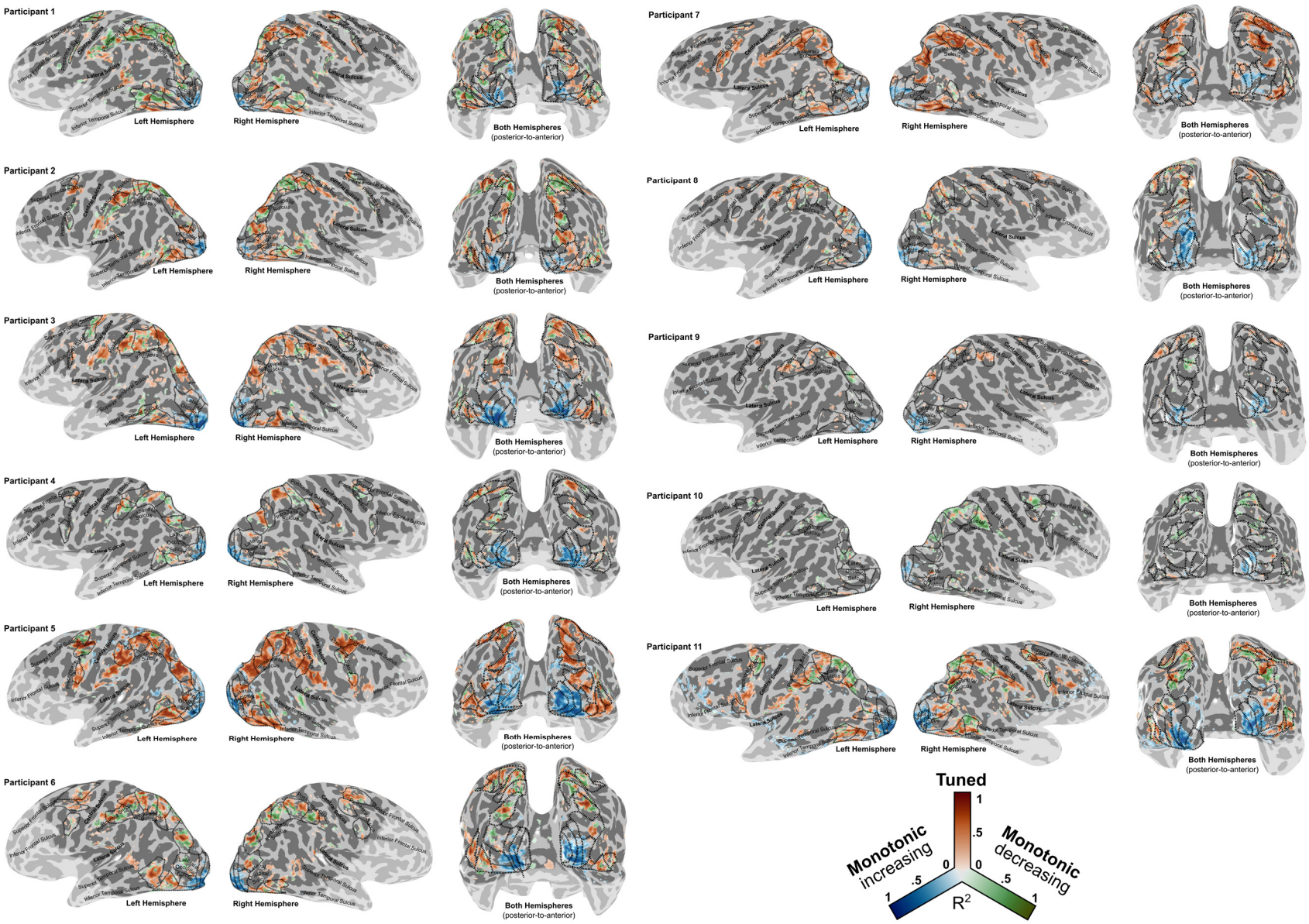
Participant 1
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 2
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 3
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 4
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 5
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 6
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 7
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 8
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 9
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 10
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Participant 11
Left Hemisphere
Right Hemisphere
Both Hemispheres
(posterior-to-anterior)
Tuned
1
.5
0
Monotonic
increasing
Monotonic
decreasing
R2

Fig. S4. Distribution of best-fitting numerosity response model across the cortical surface for all subjects: winning response model is monotonically increasing (blue), monotonically decreasing (green), or tuned (red). Color intensity shows variance explained by winning response model. All colored sites have greater than 20% model variance explained. Data are cross-validated model fits from odd and even scanning runs. Dashed black lines show visual field map borders. The light shaded region is outside the fMRI recording volume.

Monotonically increasing responses in early visual cortex are location-dependent

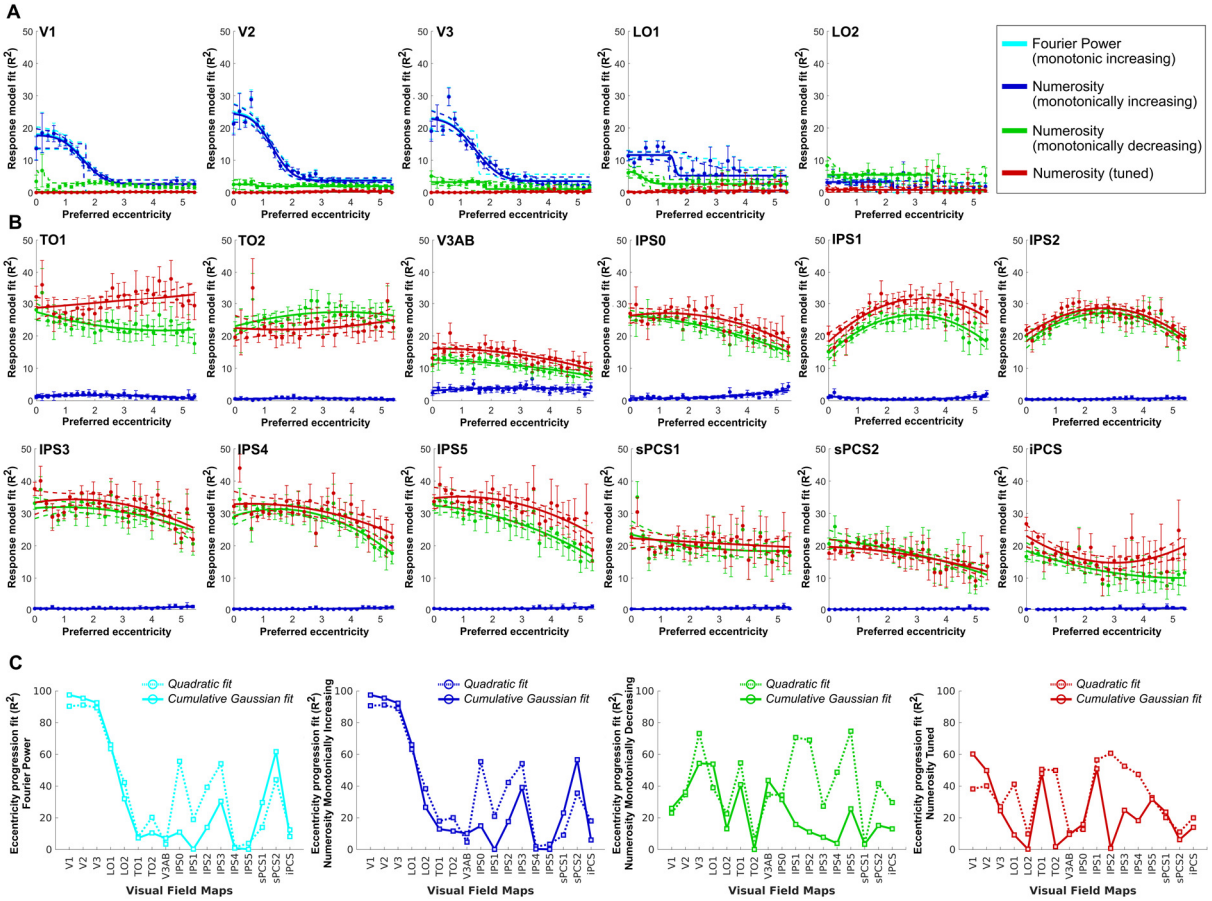

Fig. S5. Progression of eccentricity preference plotted against different numerosity and Fourier power model fits across visual field maps. Data are cross-validated model fits from odd and even scanning runs. Filled circles show mean variance explained per eccentricity bin, error bars show the standard error of the mean. Solid lines show best fit to changes with eccentricity, dashed lines are bootstrap 95% confidence intervals determined by bootstrapping. (A) Model fit variance explained as a function of preferred visual field position eccentricity for visual field maps in early visual cortex (V1-V3) and lateral-occipital (LO1-LO2) visual field maps for participants P1-P5 with Fourier power model data. (B) Visual field maps in temporal-occipital (TO1-TO2), parietal association (V3AB, IPS0-IPS5) and frontal (sPCS1-sPCS2, iPCS) areas for all participants. (C) Progression of monotonically increasing Fourier power and monotonically increasing numerosity response model fits with eccentricity was best captured by a cumulative Gaussian sigmoid function in V1–V3 and LO1 visual field maps. For all other maps and models, the eccentricity progression was better fit by a quadratic function. Solid and dashed lines show the goodness of fit of the cumulative Gaussian and quadratic functions to the eccentricity progression, respectively.

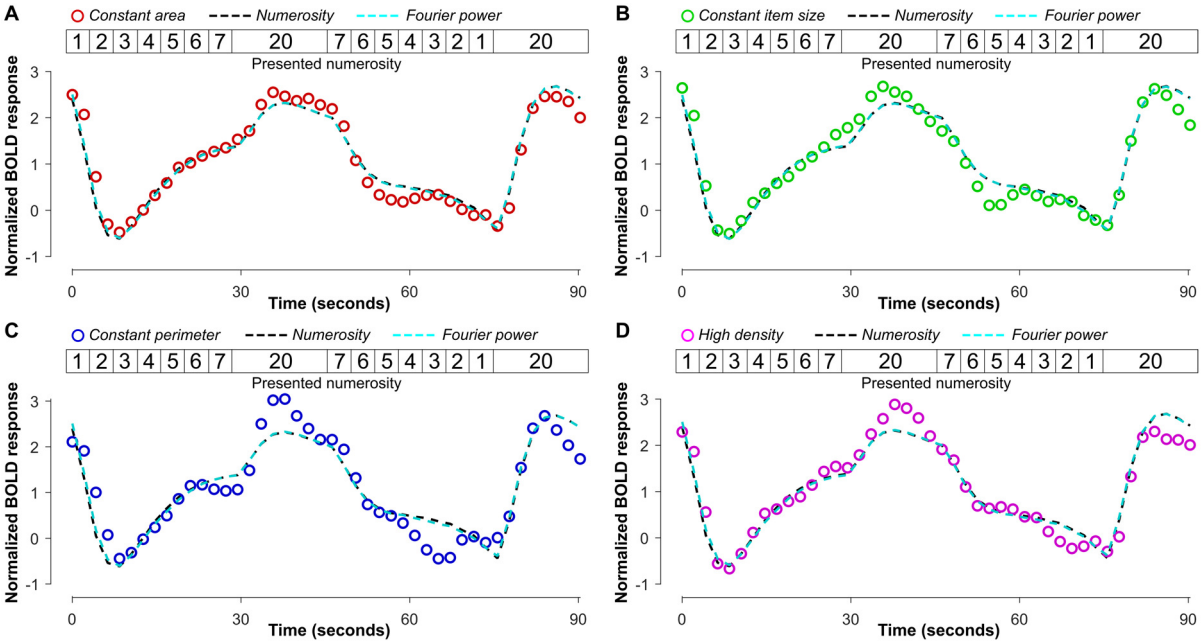

Fig. S6. Monotonic brain responses across the seven numerosities (and baseline numerosity 20) for each configuration together with numerosity and aggregate Fourier power model predictions for the same conditions. (A) Constant area (B) Constant item size (C) Constant perimeter (D) High density.

Fourier power generalization following parametric changes to stimulus features

Aggregate Fourier power as a function of parametric changes to item numerosity (Fig. S7), item size (Fig. S8), item spacing (Fig. S9), and item shape (polygons and stars, Fig. S10).

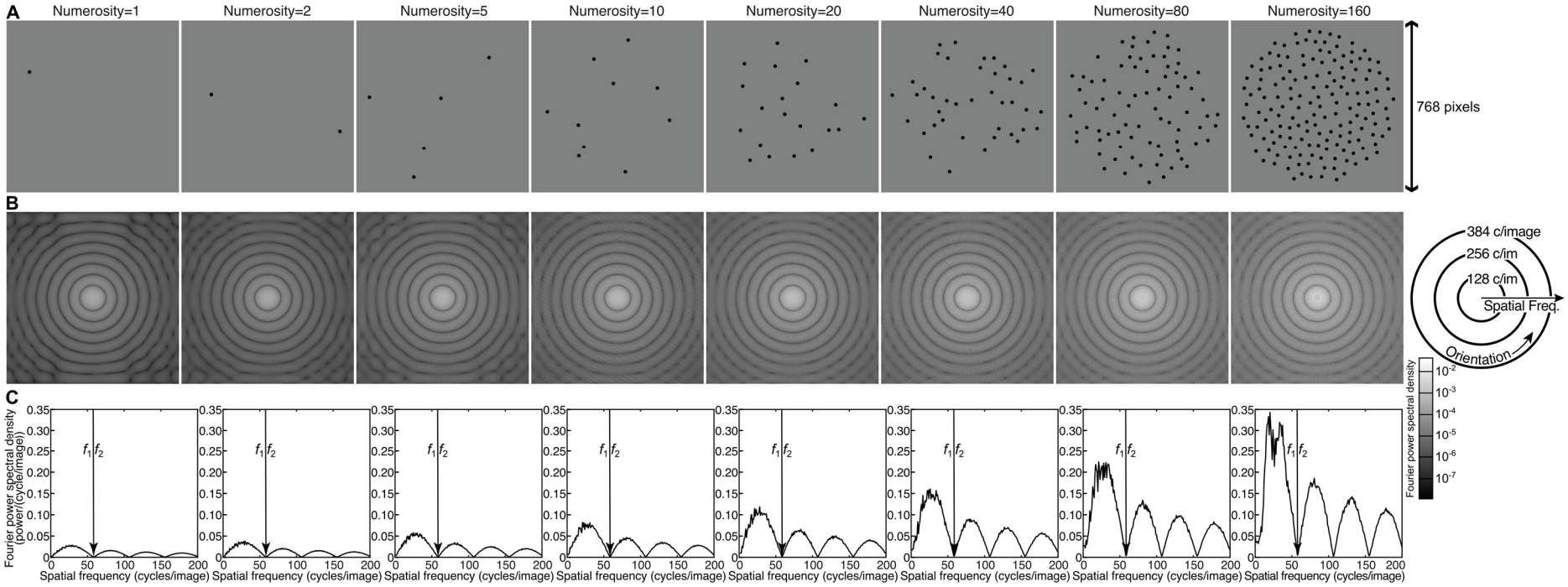

Fig. S7. Fourier power spectra and display numerosity. (A) Randomly generated images with a parametric increase in numerosity from 1 to 160 items, using the same iterative procedure used to draw the stimulus configurations shown in our experiment (constant item size, non-overlapping, sampled from a central circular area). (B) Fourier transforms of each corresponding stimulus image shown in panel A. (C) Fourier power spectral density each spatial frequency, collapsed over orientation, of the corresponding transformed images shown in B. The limit of the first harmonic (f_1) is used to determine the aggregate Fourier power of each image.

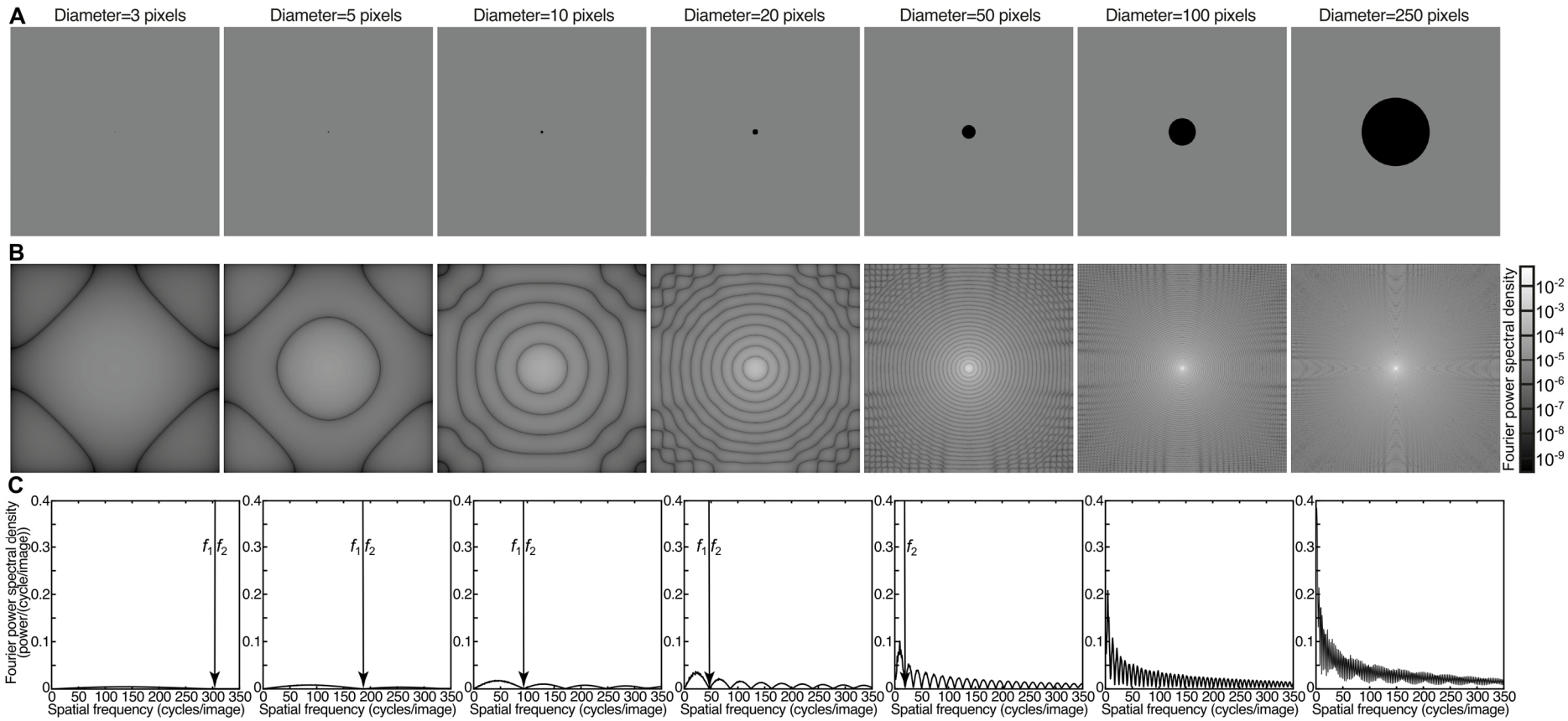

Fig. S8. Fourier power spectra and item size. (A) Illustrative images of a parametric increase in item diameter for a single circle placed at the image center ranging from 3 to 250 pixels. (B) Fourier transforms of each corresponding stimulus image shown in panel A. (C) Fourier power spectral density each spatial frequency, collapsed over orientation, of the corresponding transformed images shown in B. The limit of the first harmonic (f_1) is used to determine the aggregate Fourier power of each image.

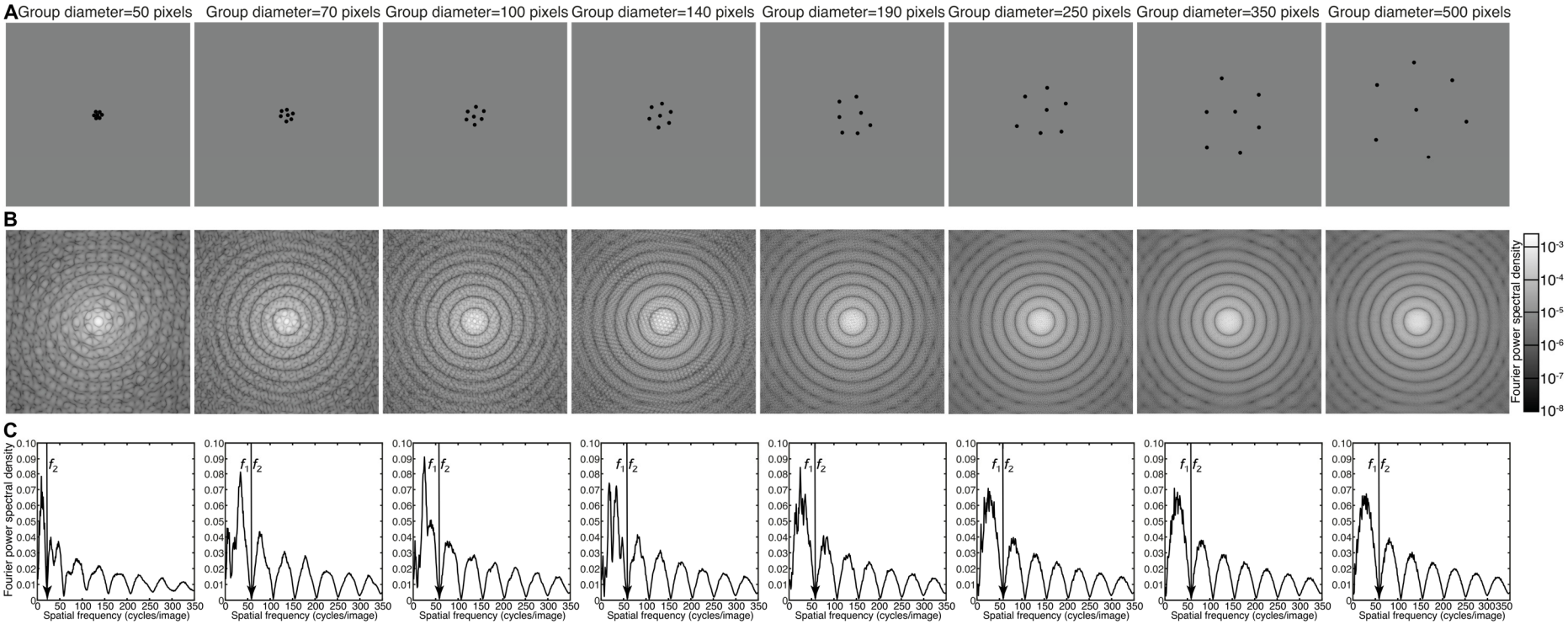

Fig. S9. Fourier power spectra and item spacing. (A) Illustrative images of a parametric increase in item spacing for 7 circles, each of 16 pixels diameter, placed at the image center ranging from with changes in the diameter of the stimulus area from 50 to 500 pixels. (B) Fourier transforms of each corresponding stimulus image shown in panel A. (C) Fourier power spectral density each spatial frequency, collapsed over orientation, of the corresponding transformed images shown in B. The limit of the first harmonic (f_1) is used to determine the aggregate Fourier power of each image.

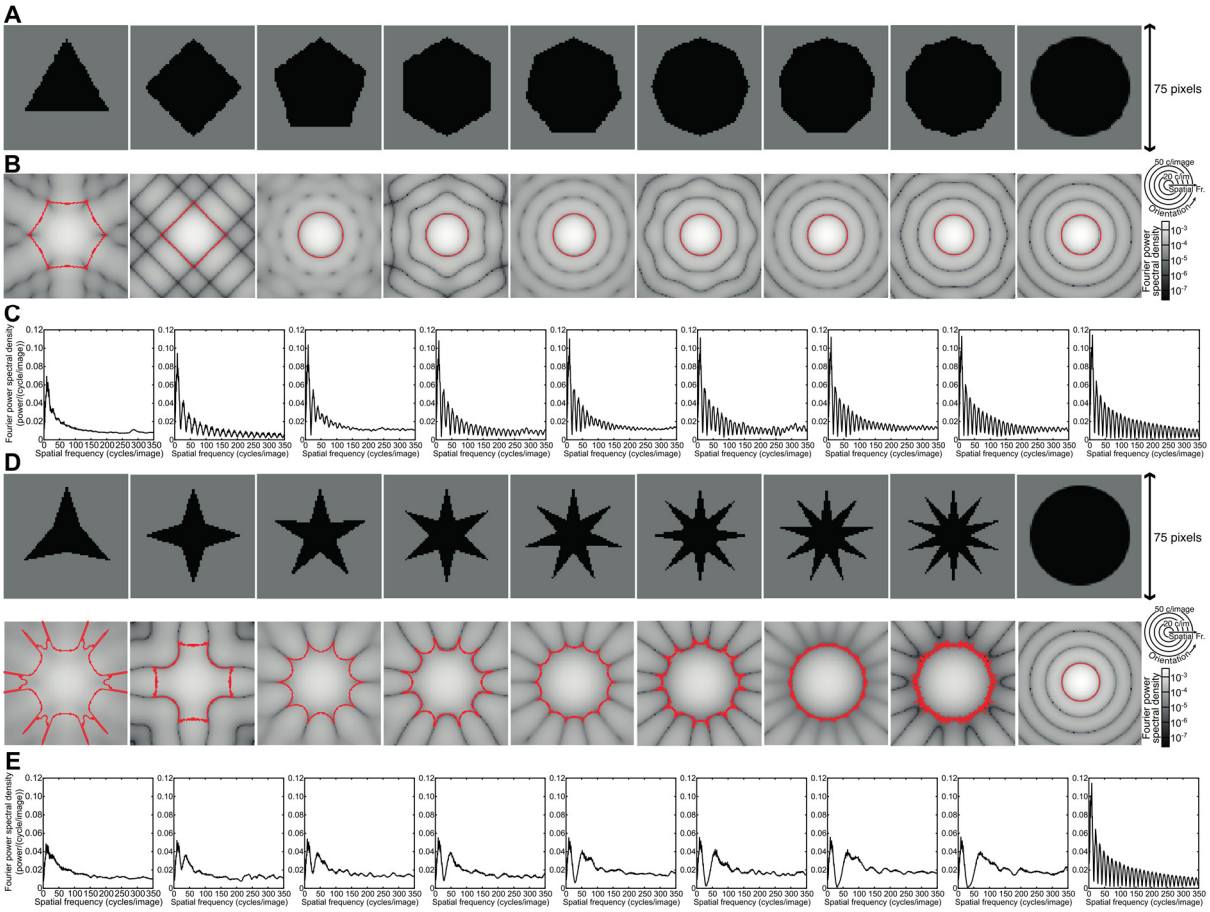

Fig. S10. Fourier power spectra and item shape. (A) Regular polygons with a parametric increase in the number of corners from 3 to 10, as well as a circle (infinite corners). (B) Fourier transforms of each corresponding stimulus image shown in panel A, above. The red line shows the limit of the first harmonic at each orientation, and we determined the aggregate Fourier power within that limit. (C) Fourier power spectral density each spatial frequency, collapsed over orientation, of the corresponding transformed images shown in B. The limit of the first harmonic differs by orientation, so is not at a single spatial frequency. (D) Stars with a parametric increase in the number of points from 3 to 10, as well as a circle (infinite points). (E) Fourier transforms of each corresponding stimulus image shown in panel C, above. The red line shows the limit of the first harmonic at each orientation, and we determined the aggregate Fourier power within that limit. (F) Fourier power spectral density each spatial frequency, collapsed over orientation, of the corresponding transformed images shown in E. The limit of the first harmonic differs by orientation, so is not at a single spatial frequency.

Monotonically decreasing responses outside early visual cortex

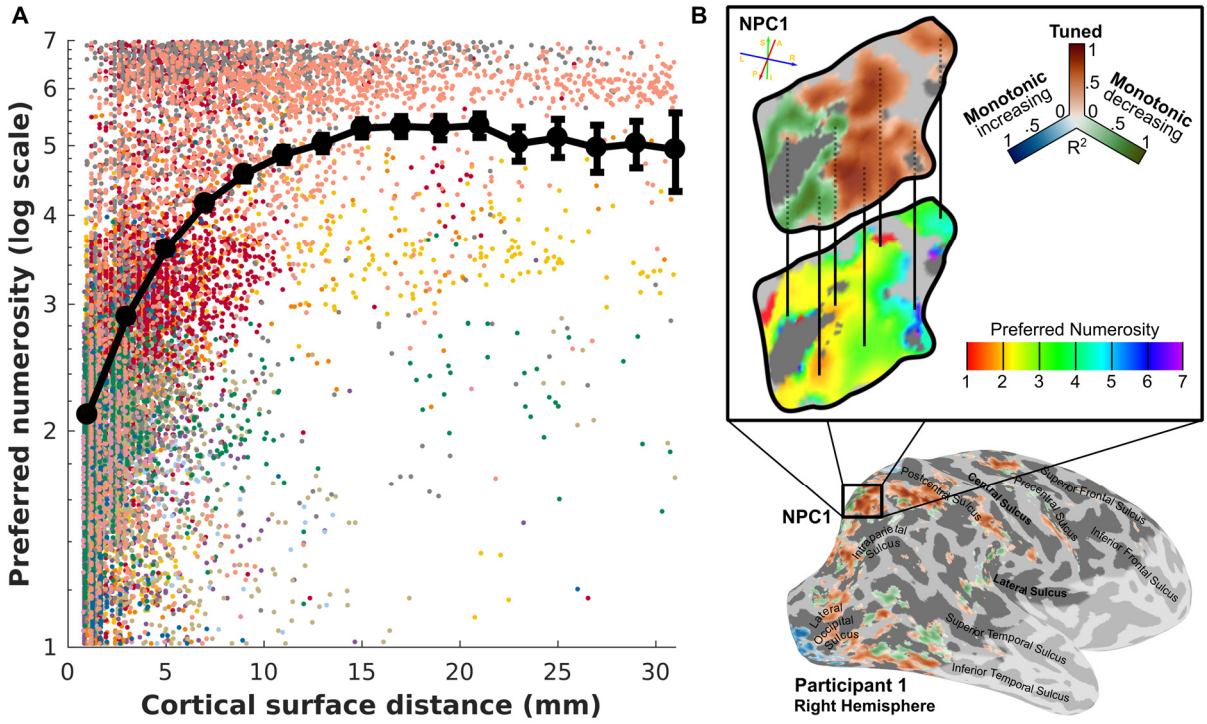

Fig. S11. (A) Preferred numerosity of the nearest recording site to a monotonically decreasing response as a function of distance along the cortical surface (recording sites are organized into 2mm bins; black circles show group mean ± 2 × standard error). Individual points represent recording sites color-coded by participant. (B) Illustrative example of relationship between tuned numerosity preferences and monotonically decreasing responses across cortical distance. Selected region-of-interest (inset) is one of the previously identified topographic maps of tuned numerosity preferences in the right hemisphere of Participant 1; NPC1 (1). Vertical black lines highlight the correspondence between lower tuned numerosity preferences occurring closer to recording sites identified as better fit as monotonically decreasing, and vice versa.

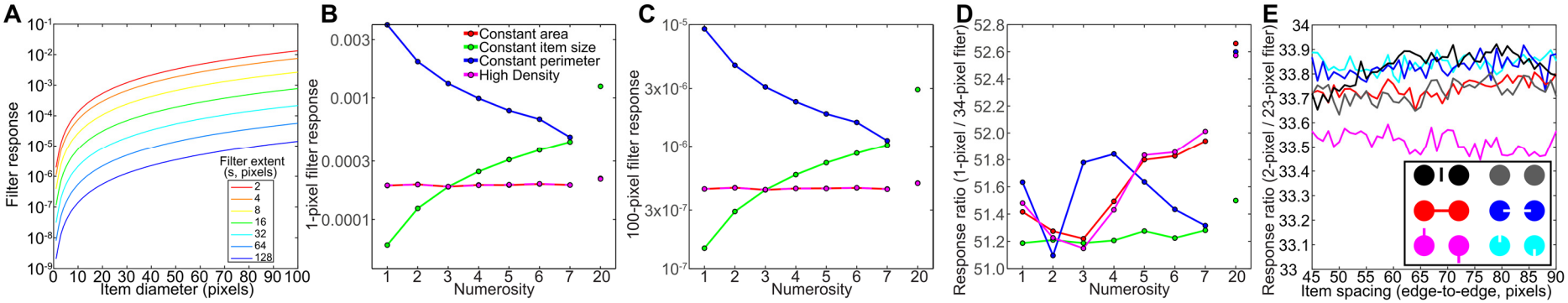

Fig. S12. (A) Responses of Dakin and colleagues' high spatial frequency filter response metric to a single circle of different sizes, for a large range of filter extents. Regardless of filter extent, the filters responses increases monotonically with circle size, so no filter of this type can give a response proportional to numerosity if item size changes with numerosity. (B & C) Responses of Dakin and colleagues' high spatial frequency filter response metric to different numerosities in each of our stimulus configurations. This is proportional to numerosity when item size is fixed for all numerosities (as Dakin and colleagues show), but this relationship does not generalize to other stimulus configurations. This is because this metric does not follow numerosity regardless of item size (A). The same result is found whether the filter extent is 1 pixel, 100 pixels, or any value in between. Note the red and magenta points and lines lie on top of each other. (D) Responses of Dakin and colleagues' response ratio metric, which they propose subjects actually perceive, for the pair of filter extents giving the highest correlation to numerosity. This metric does not closely follow numerosity, although subjects accurately and spontaneously perceive numerosity in these stimuli with minimal effects of object size and spacing. (E) Effects of connections by a bar and an illusory contour on the response ratio metric, together with control dot pairs with the same change to the dot but no connection and no reduction in perceived numerosity. Compared to unconnected dots and bars, bars touching the dots decrease the response ratio metric (regardless of connectedness) while illusory contour inducers do not affect this metric. Compared to dots only, a connected pair does not affect the response ratio and illusory contour inducers increase the response ratio. These differences are not consistent with perception of the connectedness illusion.

SI References

1. B. M. Harvey, S. O. Dumoulin, A network of topographic numerosity maps in human association cortex. *Nat. Hum. Behav.* **1** (2017).
2. B. M. Harvey, S. O. Dumoulin, Can responses to basic non-numerical visual features explain neural numerosity responses? *Neuroimage* **149**, 200–209 (2017).
3. J. Park, N. K. Dewind, M. G. Woldorff, E. M. Brannon, Rapid and Direct Encoding of Numerosity in the Visual Stream. *Cereb. Cortex* **26**, 748–763 (2016).
4. B. M. Harvey, B. P. Klein, N. Petridou, S. O. Dumoulin, Topographic representation of numerosity in the human parietal cortex. *Science* **341**, 1123–1126 (2013).
5. B. M. Harvey, A. Fracasso, N. Petridou, S. O. Dumoulin, Topographic representations of object size and relationships with numerosity reveal generalized quantity processing in human parietal cortex. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13525–13530 (2015).
6. M. Kleiner, *et al.*, What's new in Psychtoolbox-3? *Perception* **36**, 1–16 (2007).
7. S. O. Dumoulin, B. A. Wandell, Population receptive field estimates in human visual cortex. *Neuroimage* **39**, 647–660 (2008).
8. D. Burr, J. Ross, A Visual Sense of Number. *Curr. Biol.* **18**, 425–428 (2008).
9. B. M. Harvey, S. O. Dumoulin, The relationship between cortical magnification factor and population receptive field size in human visual cortex: Constancies in cortical architecture. *J. Neurosci.* **31**, 13604–13612 (2011).
10. B. M. Harvey, S. O. Dumoulin, A. Fracasso, J. M. Paul, A Network of Topographic Maps in Human Association Cortex Hierarchically Transforms Visual Timing-Selective Responses. *Curr. Biol.* **30**, 1424-1434.e6 (2020).
11. J. L. R. Andersson, S. Skare, J. Ashburner, How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *Neuroimage* **20**, 870–888 (2003).

12. R. W. Cox, AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).

13. L. Feigenson, S. Carey, E. Spelke, Infants' discrimination of number vs. continuous extent. *Cogn. Psychol.* **44**, 33–66 (2002).

14. M. W. Clearfield, K. S. Mix, Amount Versus Number: Infants' Use of Area and Contour Length to Discriminate Small Sets. *J. Cogn. Dev.* **2**, 243–260 (2001).

15. T. Gebuis, B. Reynvoet, The role of visual information in numerosity estimation. *PLoS One* **7** (2012).

16. S. C. Dakin, M. S. Tibber, J. A. Greenwood, F. A. A. Kingdom, M. J. Morgan, A common visual metric for approximate number and density. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 19552–19557 (2011).

17. D. J. McGonigle, *et al.*, Variability in fMRI: An examination of intersession differences. *Neuroimage* **11**, 708–734 (2000).

18. K. Amano, B. A. Wandell, S. O. Dumoulin, Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. *J. Neurophysiol.* **102**, 2704–2718 (2009).

19. W. E. Mackey, J. Winawer, C. E. Curtis, Visual field map clusters in human frontoparietal cortex. *Elife* **6** (2017).

20. M. A. Silver, D. Ress, D. J. Heeger, Neural correlates of sustained spatial attention in human early visual cortex. *J. Neurophysiol.* **97**, 229–237 (2007).

21. B. A. Wandell, S. O. Dumoulin, A. A. Brewer, Visual field maps in human cortex. *Neuron* **56**, 366–383 (2007).