Cognition 142 (2015) 247–265

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Modeling the approximate number system to quantify the contribution of visual stimulus features

Nicholas K. DeWind [a,b,*], Geoffrey K. Adams [a,b], Michael L. Platt [a,b], Elizabeth M. Brannon [a,c]

[a] Center for Cognitive Neuroscience, Duke University, Durham, NC, USA
[b] Department of Neurobiology, Duke University, Durham, NC, USA
[c] Department of Psychology and Neuroscience, Duke University, Durham, NC, USA

ARTICLE INFO

Article history:
Received 1 April 2014
Revised 13 April 2015
Accepted 15 May 2015
Available online 6 June 2015

Keywords:
Numerical cognition
Number sense
Psychophysics
Approximate number system
Magnitude comparison

ABSTRACT

The approximate number system (ANS) subserves estimation of the number of items in a set. Typically, ANS function is assessed by requiring participants to compare the number of dots in two arrays. Accuracy is determined by the numerical ratio of the sets being compared, and each participant's Weber fraction (w) provides a quantitative index of ANS acuity. When making numerical comparisons, however, performance is also influenced by non-numerical features of the stimuli, such as the size and spacing of dots. Current models of numerosity comparison do not account for these effects and consequently lead to different estimates of w depending on the methods used to control for non-numerical features. Here we proffer a new model that teases apart the effects of ANS acuity from the effects of non-numerical stimulus features. The result is an estimate of w that is a more theoretically valid representation of numerical acuity and novel terms that denote the degree to which a participant's perception of number is affected by non-numerical features. We tested this model in a sample of 20 adults and found that, by correctly attributing errors due to non-numerical stimulus features, the w obtained was more reliable across different stimulus conditions. We found that although non-numerical features biased numerosity discriminations in all participants, number was the primary feature driving discriminations in most of them. Our findings support the idea that, while numerosity is a distinct visual quantity, the internal representation of number is tightly bound to the representation of other magnitudes. This tool for identifying the different effects of the numerical and non-numerical features of a stimulus has important implications not only for the behavioral investigation of the ANS, but also for the collection and analyses of neural data sets associated with ANS function.

1. Introduction

The approximate number system (ANS) is a nonverbal mechanism for estimating the number of items in a set that develops early in human ontogeny and is shared with a wide array of animals (Feigenson, Dehaene, & Spelke, 2004). The ANS is faster but much less accurate than verbal counting. The ANS may serve as a neural scaffold for symbolic mathematics, a proposition supported by the finding that ANS acuity (w) predicts math achievement in both children and adults (DeWind & Brannon, 2012; Gilmore, McCarthy, & Spelke, 2010; Gilmore et al., 2013; Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Halberda, Mazzocco, & Feigenson, 2008; Lyons & Beilock, 2011; Mazzocco, Feigenson, & Halberda, 2011; Piazza et al., 2010; Starr, Libertus, & Brannon, 2013; but see Holloway

* Corresponding author at: 450 Research Drive, Box 90999, Durham, NC 27708, USA.

E-mail address: nicholas.dewind@duke.edu (N.K. DeWind).

http://dx.doi.org/10.1016/j.cognition.2015.05.016
0010-0277/

& Ansari, 2009; Sasanguie, Defever, Maertens, & Reynvoet, 2013; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013) and that extensive practice on tasks that tap the ANS improves symbolic math performance (Hyde, Khanum, & Spelke, 2014; Park & Brannon, 2013).

The acuity of the ANS typically is measured by presenting arrays of dots and requiring participants to indicate which has more. When dot arrays differ in numerosity, however, other properties of the stimuli—such as dot size, dot density, and array extent—differ as well. Many prior studies have found that non-numerical visual stimulus features influence numerosity discrimination performance, thus interfering with precise estimates of ANS acuity (e.g. DeWind & Brannon, 2012; Frith & Frith, 1972; Gebuis & Gevers, 2011; Ginsburg, 1976; Sophian, 2007; Tokita & Ishiguchi, 2010).

While most researchers acknowledge that non-numerical stimulus features influence numerosity judgments, the two most commonly used models of ANS acuity do not account for these biases (Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004; Pica, Lemer, Izard, & Dehaene, 2004; Whalen, Gallistel, & Gelman, 1999). These models postulate that numerosity is represented by the ANS as a normally distributed random variable with a mean equal to the number being represented and a width proportional to ANS acuity (*w*). Errors in numerical discrimination occur when the numbers of items being compared activate overlapping internal numerosity representations. According to these models, the overlap of these representations is entirely attributable to the ratio of the numerosities being compared and the *w* term.

When the *w* parameter in these models is fit to accuracy data from a dot array comparison task, all errors in the task are implicitly assumed to be the result of imprecision of the representation of number. However, since non-numerical features also affect numerosity judgments, they sometimes cause errors (or correct responses) that cannot be attributed to numerical ratio. These responses are incorrectly attributed to imprecision or precision in the representation of number. As a result, the *w* measure derived from the current models of numerical representation conflates the acuity of the numerical representation with the biasing effects that non-numerical stimulus features have on that representation. In practice, this means *w* is influenced by idiosyncrasies in the way the experimenter has chosen to control for non-numerical stimulus features. In its most extreme form, large differences in the congruence or incongruence of non-numerical stimulus features with number result in wildly divergent estimates of *w* in the same individual, causing some to question the existence of the ANS independent of non-numerical feature cues (Szucs, Nobes, Devine, Gabriel, & Gebuis, 2013).

Here we introduce a new "stimulus space" that elucidates the dependencies and degrees of freedom inherent in dot array stimuli. Utilizing the insights provided by the stimulus space, we then propose a modification to the logarithmic model of the ANS that explicitly accounts for the effects of non-numerical stimulus features on numerosity judgments. This approach allows ANS acuity to be estimated independently of the influence of multiple non-numerical stimulus features, thus yielding a more theoretically valid estimate of *w* that is more reliable across stimulus sets. Improved estimation of *w* will help researchers elucidate the relative importance of ANS acuity on mathematical cognition, the factors that may mediate that relationship, and its developmental trajectory.

In addition to making theoretical advances in modeling *w*, our model also returns coefficients describing the influence of non-numerical stimulus features, thereby providing novel quantitative parameters useful for comparing individuals. We assessed the prevalence of non-numerical feature bias among educated adults and statistically tested the hypothesis implicit in the current models of the ANS: that *w* and numerical ratio are the only factors that determine the discriminability of dot arrays in a numerical discrimination task. The non-numerical feature coefficients also provide a straightforward and quantitative way to assess the use of "alternative strategies": that is, the reliance on non-numerical features instead of numerosity to make discriminations between stimuli. Here we are able to provide a comprehensive unbiased assessment of the role of ten non-numerical stimulus features in numerical discriminations and test the hypothesis that the ability to approximately enumerate is reducible to co-varying non-numerical cues.

Here we model choice behavior in adults based on the number, size, and spacing of dots; however, with slight modification we can use those same factors to model neural dependent variables such as neuronal firing rate, electroencephalography (EEG) scalp voltage, and blood-oxygen-level-dependent (BOLD) signal. Thus, in addition to clarifying the effect of different stimulus features on behavior, our new modeling approach can help elucidate which brain responses reflect number as opposed to other features of a stimulus.

2. Theory and calculations

We applied a novel analytical technique to model numerical discrimination performance as a function of numerosity, item size, and item spacing. Our approach relies on the insight that although arrays of dots have many different features that all co-vary, the features known to influence numerosity judgments have three degrees of freedom. Thus, numerosity discrimination performance can be modeled as a function of just three stimulus features: the number of dots in the array and two novel parameters that describe the size and spacing of the dots within the array. From the coefficients returned for these three features the influence of many other non-numerical features can then be calculated. This modeling approach allows a dissociation of ANS acuity from the biasing effects of non-numerical visual features, thus yielding a theoretically valid estimate of ANS acuity that is more reliable across different stimulus sets.

2.1. Intrinsic and extrinsic stimulus features

Our approach requires a full understanding of the relationship among numerosity, intrinsic features of the stimulus, and extrinsic features of the stimulus (Dehaene, Izard, & Piazza, 2005; Piazza et al., 2004). Intrinsic features are parameters of the individual items within an array, whereas extrinsic features are parameters of the array as a whole. When the numerosity of an array is fixed, the relationship between a given pair of intrinsic and extrinsic features is linear. For example, total surface area (an extrinsic feature) is equal to the number of items multiplied by the item surface area (an intrinsic feature). The same relationship exists between field area (the space within which the dots are drawn, sometimes referred to as the envelope or the convex hull) and sparsity (average field area per item, or the inverse of the density). For a given numerosity, increasing sparsity necessitates a linear increase in field area. Another way of describing these relations is to say that numerosity, item surface area, total surface area, sparsity, and field area are not mutually independent of each other, and describing all of them overdetermines the stimulus. A smaller subset of these features is sufficient to determine the full set of features, an idea we will return to below.

Fig. 1A and B plot "stimulus spaces" that summarize these relationships. Stimulus parameters are plotted with intrinsic features on the *x*-axis and extrinsic features on the *y*-axis. Fig. 1A shows the intrinsic and extrinsic features related to the size of the items, item surface area and total surface area, and Fig. 1B shows the intrinsic and extrinsic features related to the spacing of the items, sparsity and field area. Also apparent are what we term iso-numerosity lines (gray); all stimuli of a particular numerosity lie on a single iso-numerosity line, the slope of which is equal to the numerosity. Different points along an iso-numerosity line correspond to stimuli that differ in the intrinsic and extrinsic properties but have the same numerosity. An individual stimulus occupies a single point in both Fig. 1A and B; for example, the stimulus labeled S1 has a numerosity of 8. Location of a single stimulus in each of the two plots is constrained by numerosity; it must fall along the same iso-numerosity line in both plots. However, its location along that line is independent in each of the plots. In other words, the size of the items and spacing of the items in an array are independent of each other.

The essential confound inherent in generating pairs of numerosity stimuli can be appreciated visually in Fig. 1A and B. Two stimuli must be chosen, each from a different iso-numerosity line; however, any two such stimuli will also differ in the intrinsic variable, the extrinsic variable, or both. It is mathematically impossible for two stimuli to differ only in numerosity. For example, consider again the stimulus labeled S1 in Fig. 1A and B of numerosity 8. We may want to pair this stimulus with another of numerosity 16 in an ordinal comparison task and seek a way to control for changes in other visual features (the intrinsic and extrinsic variables). Now consider the two stimuli labeled S2′ and S2″, both of numerosity 16. S2′ has the same total surface area and field area as S1, but a different item surface area and sparsity. In contrast, S2″ has the same item surface area and sparsity as S1, but a different total surface area and field area. Other stimuli occupying other positions along the 16 iso-numerosity line

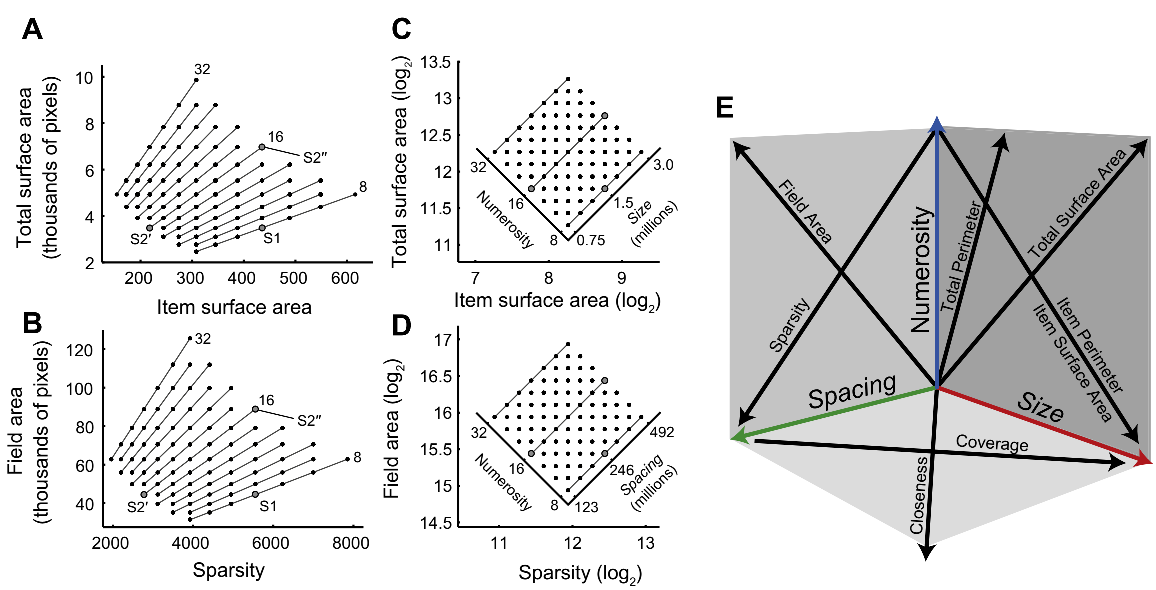

Fig. 1. Features thought to influence numerical estimation can be represented as different axes in a three dimensional stimulus space. (A and B) Stimuli (black dots) used in this experiment plotted as item surface area by total surface area (A) and sparsity by field area (B). Stimuli of the same numerosity fall along iso-numerosity lines (gray). The slope of these lines is equal to the numerosity of the stimuli that fall along them. Example stimuli referred to in the text are labeled and indicated by gray dots with black outlines. (C and D) The same plots as in (A) and (B) but with x and y axes log scaled. Changes in numerosity occur along a linear axis, and two orthogonal dimensions, *Size* and *Spacing* are apparent. The alternative axis dimensions within the plot area provide an equally descriptive quantitative account of stimulus features as the external axes do. (E) The log of numerosity, *Size*, and *Spacing* plotted as cardinal axes of a 3D stimulus space. Log of non-numerical stimulus features are also plotted as arrows to indicate the direction in the space in which they increase. Any dot array stimulus can be uniquely defined with respect to numerosity, *Size*, *Spacing*, item surface area, total surface area, item perimeter, total perimeter, field area, sparsity (and density), coverage, and apparent closeness on the basis of its location in this space.

would differ in both intrinsic and extrinsic stimulus features. All stimuli that differ in numerosity from S1 must also differ in either item surface area or total surface area and must also differ in either sparsity or field area.

2.2. Logarithmic scaling and deriving orthogonal regressors

Fig. 1A and B are redrawn in Fig. 1C and D, but with the intrinsic and extrinsic axes log scaled. Log scaling the axis affords two critical advantages: it makes the iso-numerosity lines parallel (for clarity, only the iso-numerosity lines for 8, 16, and 32 are shown), and it makes the distance between stimulus points in the space proportional to the ratios of their various features (numerical and non-numerical). As a result, changes in numerosity are represented as movement along a linear numerosity dimension, and changes in the extrinsic or intrinsic features that do not result in a change in numerosity are represented as movement along an orthogonal linear axis. We represent these linear orthogonal stimulus dimensions as alternative axes in Fig. 1C and D to emphasize that they represent a quantitative way to specify the location of a stimulus in "stimulus space", which contains all the same information as specifying the values of the intrinsic and extrinsic features. Furthermore, these three alternative axes represent a minimally sufficient set of features for describing the numerosity, as well as both sets of intrinsic and extrinsic features of a stimulus. For the rest of this manuscript we will refer to these linear orthogonal dimensions as *Size* for the intrinsic and extrinsic variables item surface area and total surface area and *Spacing* for sparsity and field area. We use capitalization and italics to make it clear that we are referring to a rigorously defined mathematical construct, but we also wish to emphasize the close relationship of these terms to the everyday concepts of size and spacing. Intuitively, changes in *Size* are equivalent to changing the size of a fixed number of items with fixed distances between their centers, and changes in *Spacing* are equivalent to changing the distances between a fixed number of items of fixed size.

We algebraically define *Size* and *Spacing* by examining the relationship between intrinsic and extrinsic stimulus parameters and finding the dimension that is orthogonal to numerosity.

$$\log_2(n) = \log_2\left(\frac{TSA}{ISA}\right) = \log_2\left(\frac{FA}{Spar}\right) \tag{1}$$

where n is the number of items, TSA is the total surface area, ISA is the item surface area, FA is the field area, and $Spar$ is the sparsity. The dimensions orthogonal to log number are

$$\log_2(Size) = \log_2(TSA) + \log_2(ISA) \tag{2}$$

$$\log_2(Spacing) = \log_2(FA) + \log_2(Spar) \tag{3}$$

Size and *Spacing* capture the aspects of dot size and spacing that are independent of numerosity. These definitions support the basic logic of the regression model that we will formulate in Section 3.6. The effect of the numerical ratio on choice behavior can be assessed, as in previous models, by fitting the w term. We can expand that model, however, by adding terms that will quantify the effect of the *Size* ratio and *Spacing* ratio. By including terms that capture the perceptual effects of item size and spacing that are independent of the perceptual effect of numerosity itself, we lay the basis for independently assessing their contributions to numerosity discrimination performance.

2.3. Non-numerical stimulus features can be reduced to linear combinations of numerosity, Size, and Spacing in a logarithmic stimulus space

Fig. 1E represents numerosity, *Size*, and *Spacing* as cardinal axes in a three-dimensional stimulus space with log scaled axes. We can imagine generating this space by taking the two two-dimensional spaces in Fig. 1C and D and intersecting them along the numerosity axis at right angles to each other (in the third dimension). Any given array of dots is described by a single position within this three-dimensional space. The three dimensions (log of numerosity, log of *Size*, and log of *Spacing*) are independent of each other, and these three variables fully determine the extrinsic and intrinsic parameters discussed above (as illustrated in Fig. 1C and D). Importantly, several other stimulus features are also fully specified by numerosity, *Size*, and *Spacing*. For example, the item perimeter and total perimeter are determined by the numerosity and *Size* parameters. Coverage and apparent closeness are two features that depend on *Size* and *Spacing* and are not related to the numerosity of the stimulus. Coverage, sometimes referred to as density (e.g. Gebuis & Reynvoet, 2011), is the total surface area per field area. Apparent closeness is the overall scaling of the stimulus, and increasing it is equivalent to zooming in on a stimulus such that it subtends a larger visual angle without changing its relative proportions. Appendix A contains the equations that relate each of these features to numerosity, *Size*, and *Spacing*. These equations demonstrate that our stimulus space is very descriptive. With just three values it specifies a stimulus's numerosity, item surface area, total surface area, sparsity (and density), field area, item perimeter, total perimeter, coverage, apparent closeness, *Size*, and *Spacing*. This descriptiveness is important for the modeling approach that we describe below, because it provides the basis by which our model can infer the effect of any stimulus feature on discrimination performance while only containing terms for numerosity, *Size*, and *Spacing*.

It is worth noting that log scaling is not merely a mathematical trick; it has important behavioral, neurobiological and theoretical bases as well. Behavioral "same-different" judgments in humans and monkeys are best fit by assuming a log compressed mental number line rather than a linear one with scalar variability (Merten & Nieder, 2009). Furthermore, response functions of single neurons tuned to individual numerosities found in prefrontal cortex in monkeys are also logarithmically compressed (Nieder & Miller, 2003). Theoretically, the Weber–Fechner law states that the discriminability of two stimuli is linearly related to their ratio, equivalent to their distance on the logarithmic scale. For example, according to Weber–Fechner, a stimulus of numerosity 8 and one of numerosity 16 are equally discriminable as a stimulus of 16 and one of 32, because both pairs have a 1:2 ratio. In logarithmic stimulus space the distances along the numerosity dimension between 8 and 16 and between 16 and 32 are equal. Thus, we can use the difference in log numerosity as a regressor in a generalized linear model of numerical discrimination. Indeed, this is the approach of the logarithmic model developed by Piazza et al. (2004).

Critically, on a logarithmic scale, the equations relating the numerosity, *Size*, and *Spacing* to the other non-numerical features are all linear equations (Appendix A). Geometrically, this means that the dimensions along which different non-numerical features increase are straight lines in the three-dimensional stimulus space illustrated in Fig. 1E. Furthermore, the distance along any of these dimensions that separates a pair of stimuli is proportional to the ratio difference of that feature. Thus, one of the benefits of the new stimulus space introduced here is that just as a 1:2 ratio of numerosity corresponds to a fixed distance along the numerosity dimension, here a fixed ratio of any feature corresponds to a fixed distance along its own dimension. For example, the distance between two stimuli along the total surface area dimension, one of which is comprised of 2000 pixels and the other of 4000 pixels, will be the same as the distance between a stimulus comprised of 4000 pixels and one of 8000 pixels, since both these stimulus pairs differ by a 1:2 total surface area ratio. Therefore, we can extend the logic of the Piazza et al. (2004) model to non-numerical stimulus features. Instead of assuming that only the log of the numerical ratio affects numerical judgments, we can determine which stimulus feature ratios are affecting judgments, and we can do so in a manner that does not favor any particular feature.

It may be tempting to simply include regressors for the log ratio of all of the non-numerical features in a generalized linear model and have them compete with numerosity to explain the variance in behavioral discrimination performance. Although any two stimulus features are only partially collinear, some combinations of two features are fully collinear with a third, making such a model overspecified. Instead, we can take advantage of the linear equations that relate the log ratios of all the other features to the log ratios of numerosity, *Size*, and *Spacing* (Appendix A). These linear relationships mean that we can use the log of the S1 to S2 numerosity ratio, the log of the S1 to S2 *Size* ratio, and the log of the S1 to S2 *Spacing* ratio as regressors in a generalized linear model of numerosity discrimination. From the coefficients returned we can then infer the effect of a ratio difference of any feature on numerosity discrimination performance.

In short, previous models of numerical comparison predict that accuracy in a numerical ordinal comparison task is a function of the numerical ratio. The model, which we introduce in the methods section below, allows instead that accuracy is a function of the numerical ratio, the *Size* ratio, and the *Spacing* ratio. The effects of *Size* and *Spacing* ratio on accuracy would be of little interest by themselves, since they are merely novel mathematical constructs. However, by virtue of the relationship between *Size* and *Spacing* and the other non-numerical features, estimating the effect of *Size* and *Spacing* on accuracy is mathematically equivalent to estimating the effects of all the non-numerical stimulus features on accuracy.

To evaluate our new modeling approach, we tested 20 adult participants using a standard non-symbolic numerical ordinal comparison task. Participants were instructed to choose the array that contained more dots. Our stimuli were generated such that numerical ratio, *Size* ratio, and *Spacing* ratio were varied independently across stimulus pairs. We developed a generalized linear model that allowed us to fit choice curves that modeled each participant's sensitivity to each of those ratios. We hypothesized that numerosity ratio would be the main determinant of choices given the task instructions, but that *Size* and *Spacing* ratio would have some influence.

3. Methods and materials

3.1. Participants

Participants were 20 adults (mean 22.9 years, range 19.6–26.8 years) recruited from the Duke University community. Eleven of the 20 participants were female. All participants gave written informed consent in accordance with a Duke IRB approved protocol.

3.2. Design

Five participants completed ten sessions within 11 days and performed a maximum of three sessions in one day. Another 15 participants completed a single session in one day. Each session lasted about 1 h and consisted of 750 trials broken into three blocks of 250 trials each. Participants were required to take a five-minute break between blocks. Participants were compensated 10 USD for each session.

3.3. Task

Participants were seated in front of a computer and instructed to indicate the side of the screen that contained the greater number of dots using the arrow keys on a standard keyboard. Instructions were given verbally at the beginning of the session and in written format on the computer screen at the beginning of each block. At the beginning of each trial, a readiness cue was presented in the center of the screen (500 ms) followed by two arrays of white dots on a black background presented simultaneously to the right and left of the readiness cue (eccentricity $\sim$8.5 degrees) for 250 ms. A response prompt was then presented, and responses were followed by a 2 second inter-trial interval.

Participants were given eight easy practice trials (1:4 numerical ratio) at the beginning of each block. Practice trials were identical to the experimental trials except that they had a longer readiness cue time (1 s), longer stimulus display time (1 s) and a longer inter-trial interval (4 s). In the rare event that a participant responded incorrectly on any practice trial, the script terminated with a prompt to see the experimenter. The experimenter then repeated the instructions and the block restarted.

3.4. Stimuli

We constructed a stimulus set that divided two octaves of numerosity, *Size*, and *Spacing* into 13 levels, approximately evenly spaced on a logarithmic scale. The range of stimulus parameters is shown in Fig. 1A–D, with all of the stimuli plotted in the stimulus space described in Section 2.2. For 7 of the 13 numerosities, stimuli were generated at 7 different *Sizes* and 7 different *Spacings* yielding a total of $7 \times 7 \times 7 = 343$ stimuli. For the other 6 numerosities, stimuli were generated at 6 *Sizes* and 6 *Spacings* for a total of $6 \times 6 \times 6 = 216$ stimuli. Thus there were 559 unique stimulus parameter combinations. On each trial the experimental program randomly picked one of 4 different numerical ratios (closest whole numbers to $1{:}2^{1/6}$, $1{:}2^{1/3}$, $1{:}2^{1/2}$, or 1:2 ratios), one of 13 *Size* ratios (all possible pairings), and one of 13 *Spacing* ratios (all possible pairings).

In order to spread stimuli evenly along a logarithmic scale, the values were rounded to the nearest whole number. For example, 32 is 2^5 and 8 is 2^3. However, we wanted 11 more powers of two spaced evenly between 2^3 and 2^5, such as $2^{4.5}$. $2^{4.5}$ is approximately 22.627 which we rounded to 23. Similarly, dot diameters and field diameters were rounded to the nearest whole pixel so they could be drawn properly on a monitor.

After defining the number, *Size*, and *Spacing* of a stimulus, the algorithm created an instantiation of that stimulus. First the field area and item surface area were calculated (see Appendix A for the relations between number *Size* and *Spacing* and the other visual magnitudes). Dots of the appropriate size were drawn at random locations within a circular field of the appropriate area. The only constraint on placement was that all dots were separated by at least one pixel and that all the dots were completely within the circle defining the stimulus field. It is worth noting that the circular field was not necessarily the smallest circle that could encompass all the dots in the array, although across multiple stimuli the field area and the smallest encompassing circle area were closely correlated.

3.5. Modeling choice behavior with existing models

We compared our model to the two standard models for estimating numerical acuity (w). The first model, termed here the "logarithmic model," assumes numerosities are represented as normally distributed random variables on a log compressed mental number line with means equal to the logarithm of the number represented and a fixed standard deviation (Piazza et al., 2004, 2010). That model was used to fit data in which participants compared a deviant value to a fixed standard value (either 16 or 32). The probability of choosing "larger" for the deviant stimulus was the proportion of the numerosity distribution lying on the greater side of the standard. The probability of this occurring at different numerosities is a cumulative normal distribution with a standard deviation that is equal to the standard deviation of the representation of numerosity, w.

In contrast participants in our task were asked to pick the larger of two numerosities that both varied from trial to trial with no fixed reference value. To accommodate this change in paradigm, we modified the model used by Piazza et al. (2010) to include two numerosity distributions on a log-compressed number line, each with equal variance w. According to this version of the logarithmic model, the probability of choosing a stimulus was the proportion of its numerosity distribution lying on the greater side of the other stimulus distribution (not a fixed referent). The probability of this happening at different log right to left numerosity ratios is a cumulative normal distribution with standard deviation of w multiplied by root two.

$$p(ChooseRight) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\log_2(r_{num})}{\sqrt{2}(\sqrt{2}w)}\right)\right] \quad (4)$$

where r_{num} is the ratio of the right side to the left side stimulus and *erf* is the error function.

The second model, termed here the "linear model," assumes that number is represented on a mental number line that is linearly spaced but has variance that scales linearly with magnitude (Halberda et al., 2008; Pica et al., 2004). In this model w is the scalar that relates the numerosity to the standard deviation.

$$p(ChooseRight) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{(r_{num} - 1)}{\sqrt{2}w\sqrt{r_{num}^2 + 1}}\right)\right] \quad (5)$$

It should be noted that w refers to different mathematical constructs in the logarithmic and linear models, making direct comparisons meaningless. Indeed, the same accuracy data fit by these two models produces different numerical values for w.

3.6. A novel model of numerosity discrimination that accounts for the effect of non-numerical features

We compared the two models above with the model we developed that was designed to accommodate the empirical fact that the size and spacing of dots within an array affect subjective numerosity. We fit a generalized linear model to choice data with regressors for the log of the ratio of numerosity, *Size*, and *Spacing* of the stimulus appearing on the right and the stimulus appearing on the left. The model formatted as a function of a linear expression:

$$p(ChooseRight) = (1-\gamma)\left(\frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\beta_{side} + \beta_{num}\log_2(r_{num}) + \beta_{Size}\log_2(r_{Size}) + \beta_{Spacing}\log_2(r_{Spacing})}{\sqrt{2}}\right)\right] - \frac{1}{2}\right) + \frac{1}{2} \quad (6)$$

This equation looks rather different, but it can be thought of as simply an elaboration of Eq. (4). This can be better appreciated if we rearrange it:

$$p(ChooseRight) = (1-\gamma)\left(\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{\log_2(r_{num}) - \left(\frac{-\beta_{side}-\beta_{Size}\log_2(r_{Size})-\beta_{Spacing}\log_2(r_{Spacing})}{\beta_{num}}\right)}{\sqrt{2}\frac{1}{\beta_{num}}}\right)\right]-\frac{1}{2}\right)+\frac{1}{2} \quad (7)$$

The $\log_2(r_{num})$ term is equivalent between Eqs. (4) and (7). The standard deviation of Eq. (4) was w multiplied by root two, whereas here the standard deviation is the reciprocal of β_{num}:

$$\sigma = \frac{1}{\beta_{num}} \quad (8)$$

Therefore we can compute w from the new model parameters according to:

$$w = \frac{\sigma}{\sqrt{2}} = \frac{1}{\sqrt{2}\beta_{num}} \quad (9)$$

This makes sense: the w term depends only on β_{num}, the term capturing participants' sensitivity to number. There are also meaningful differences between the models. In Eq. (4) $\log_2(r_{num})$ alone determines the x-axis position along a single choice curve where the indifference point is at 0 (a 1:1 ratio). The greater the numerical ratio of right to left, the greater the probability of choosing right. In Eq. (7) this is still true, but there is now a large term subtracted from the x-axis position. This value determines the indifference point of the choice curve, which can now vary according to several new terms. In the context of a cumulative normal choice curve, the indifference point is the mean and is given by:

$$\mu = \frac{-\beta_{side} - \beta_{Size}\log_2(r_{Size}) - \beta_{Spacing}\log_2(r_{Spacing})}{\beta_{num}} \quad (10)$$

These new terms include the log ratios of *Size* and *Spacing* as well as all the β terms. β_{side} is an offset term that accounts for any side bias a participant might have. β_{Size} and $\beta_{Spacing}$ modulate the degree to which the *Size* and *Spacing* ratios affect the indifference point, and β_{num} scales the effect of all factors such that the greater the numerical acuity the smaller the effect of everything else.

The other new term in the model is the γ. Because our task was fast paced to allow many trials to be collected within a reasonable amount of time, we assumed that participants occasionally looked away from the screen or were momentarily distracted and failed to process the stimuli. In this case participants' choices would be random and not related to any stimulus characteristics. To accommodate this we included γ, a guessing term (Halberda & Feigenson, 2008; Pica et al., 2004). This term allows choice curves to asymptote below 100% and above 0%, since the more a participant guesses, the more the entire choice curve is compressed toward 50%. In the extreme example of a participant who responded randomly, the γ term would be 1, the proportion of rightward responses would be 50%, and no other term in the model would matter.

It is worth noting that if γ and all the β terms besides β_{num} are zero, the new model completely reduces to the logarithmic model in Eq. (4). However, if the β terms for *Size* and *Spacing* are non-zero, the indifference point will not be at a 1:1 numerical ratio. In other words, the participant can be biased, choosing one of two numerically equal arrays more than 50% of the time if that array has, for example, more spaced out dots.

The logarithmic model of choices in Eq. (4) is based on a particular hypothesis regarding the underlying internal representation of numerosity (Piazza et al., 2004). According to this hypothesis an approximate numerosity is represented as a normally distributed random variable. The distribution is centered on the actual value it is representing, but it is imprecise and probabilistic. The standard deviation of the numerosity random variable is the term w. When two numerosities are compared, as in a task like the one used here, the overlap in the two distributions causes confusability. Thus, the distance between the two numbers on the logarithmic mental number line (equivalent to the ratio) and the w term determine the confusability of two numerosities and therefore the error rate.

The changes proposed to get from Eq. (4) to Eq. (7) correspond to an equivalent change in the hypothetical underlying mental representation of numerosity. Instead of the numerosity normal random variable being centered on the actual number being represented, we propose that, in people who are biased by non-numerical features, the mean can vary depending on the *Size* and *Spacing* of the stimulus. Thus, the size and spacing of the items in a

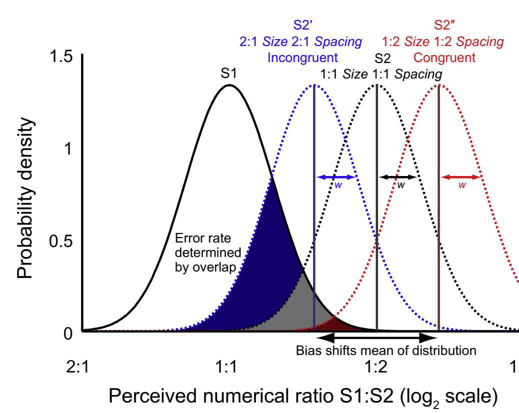

Fig. 2. The internal representation of numerosity in a hypothetical participant according to the new model. This hypothetical participant has the following coefficients: $w = 0.3$ ($\beta_{num} \approx 2.36$), $\beta_{Size} = 0.5$, $\beta_{Spacing} = 0.5$, $\beta_{side} = 0$. Stimulus S1 and S2 are represented internally as normally distributed random variables on a logarithmically compressed mental number line. The standard deviation of these representations is fixed and equal to w. When *Size* and *Spacing* are equal in S1 and S2, the model is equivalent to the logarithmic model (Eq. (4); Piazza et al., 2010). However, when *Size* and *Spacing* are incongruent to number as in S2′ the perceived numerical ratio decreases and confusability increases (as represented by the overlap of the distributions). When *Size* and *Spacing* are congruent (S2″) perceived numerical ratio increases and confusability decreases. The shift of the mean of the S2 distribution is given by Eq. (10).

stimulus array can be thought of as increasing or decreasing the perceived numerosity depending on whether a particular participant has a positive or negative β_{Size} and $\beta_{Spacing}$.

Fig. 2 helps to elucidate the effects of *Size* and *Spacing* on the overlap of the internal representation of two numerosities, S1 and S2, in a hypothetical participant. If S1 and S2 differ by a numerical ratio of 1:2, but do not differ in *Size* and *Spacing*, then the model reduces to the logarithmic model (Piazza et al., 2004, 2010) expressed in Eq. (4). The overlap will depend only on w as illustrated by the S2 distribution in black in Fig. 2. If the *Size* and *Spacing* ratios of S1 to S2 are both 2:1, however, the participant's bias causes the mean of the numerosity representation of S2′ to shift to the left as given by Eq. (10) (blue distribution). As a result, the overlap between S1 and S2′ increases and accuracy decreases, just as is actually observed in experiments in which non-numerical features are incongruent with numerosity. Conversely if the *Size* and *Spacing* ratios are 1:2, as in congruent trials, the distributions grow farther apart and accuracy improves (distribution S2″ in red). These changes in accuracy occur despite no change in numerical ratio and no change in w, and therefore cannot be modeled using previous approaches. Only a framework that takes non-numerical stimulus features into account can model these effects on error rate.

We have provided code in the supplementary materials that computes the *Size* and *Spacing* parameters and will fit the model in Eq. (7) to behavioral data sets.

3.7. The discrimination vector, discrimination dimension, and testing for non-numerical alternative strategies

The three value vector defined by β_{num}, β_{Size}, and $\beta_{Spacing}$ reflects the degree to which the distance between two stimuli along the three cardinal dimensions in Fig. 1E affect the probability of choosing a particular stimulus as the more numerous one. We will refer to this vector as the participant's discrimination vector and the dimension it defines in Fig. 1E stimulus space as the discrimination dimension. Pairs of stimuli that differ along the discrimination dimension are most easily discriminated, and participants are indifferent between pairs of stimuli that differ along the dimensions orthogonal to the discrimination dimension. If a participant has no bias, then her discrimination dimension will be identical to the numerosity dimension, and the magnitude of the discrimination vector will be identical to β_{num}. However, if β_{Size} or $\beta_{Spacing}$ is not zero then the discrimination dimension will differ from the numerosity dimension. A participant who has a significant β_{num} ($p < 0.01$) and no significant effect of β_{Size} or $\beta_{Spacing}$ ($p > 0.1$) can be considered to be making unbiased numerosity judgments.

Participants who fail to meet the criteria for unbiased numerical discrimination may be primarily relying on numerosity but have a non-numerical bias; alternatively, they may be responding primarily on the basis of one of many possible non-numerical stimulus features. Geometrically, this is equivalent to asking which of the named dimensions in the Supplementary Animated Figure is closest to the discrimination vector. To test this statistically we projected the discrimination vector onto the numerosity dimension (equal to β_{num}) and onto each of the other dimensions. Participants whose numerosity vector projection was significantly greater than all other vector projections ($p < 0.05$) were considered to be primarily relying on numerosity but biased by a non-numerical feature. Those whose numerosity vector projection was significantly smaller than another vector projection ($p < 0.05$) were considered to be primarily relying on a non-numerical strategy. Any participant whose numerosity vector projection was not significantly different from another vector projection was categorized as having an indeterminate response strategy.

4. Results

We fit the accuracy data of individual participants performing an ordinal approximate number discrimination task with choice curves with terms for side (left or right), guessing rate, numerical ratio, *Size* ratio and *Spacing* ratio. The *Size* and *Spacing* variables are defined mathematically in Section 2.2. Intuitively, *Size* can be thought of as the aspect of the stimulus that changes with the size of a fixed number of items at fixed locations, and *Spacing* can be thought of as the parameter that changes when a fixed number of items of fixed size are spread out over a greater or lesser area of space.

4.1. Model fits account for performance variations due to non-numerical stimulus features

Fig. 3A shows the model fit for the five participants who were tested with 7500 trials. As the numerical ratio of items in the right array to items in the left array increased, participants became more likely to choose the right stimulus, as instructed. The effect of numerical ratio on the probability of choosing "right" was well fit by the model across trials (black data points and fit lines in Fig. 3A). In order to examine the effects of *Size* and *Spacing* and to evaluate how well the model accounted for these effects, we examined the subset of trials in which the non-numerical features differed dramatically. The red markers and green markers in Fig. 3A reflect trials with large *Size* and *Spacing* ratios respectively (greater than an 8:3 or less than a 3:8 ratio). Critically, the model was only fit once for each participant to his or her full dataset; the red and green lines represent the predictions of the model for these subsets of trials. As can be visualized in the offset of the red and green lines from the black lines, all of these participants were influenced by *Size*, *Spacing*, or both. These red and green lines represent an explanation of variance in numerosity judgments that cannot be accounted for with either the logarithmic or linear models of the ANS used in previous studies.

On any given trial, the influence of *Size* or *Spacing* may help or hinder performance. If a participant has a significant positive effect of *Size* ratio or *Spacing* ratio, as do most of the participants in our sample, then larger and more spaced-out dots are perceived as more numerous (a notable exception is participant 1 who has a negative *Size*

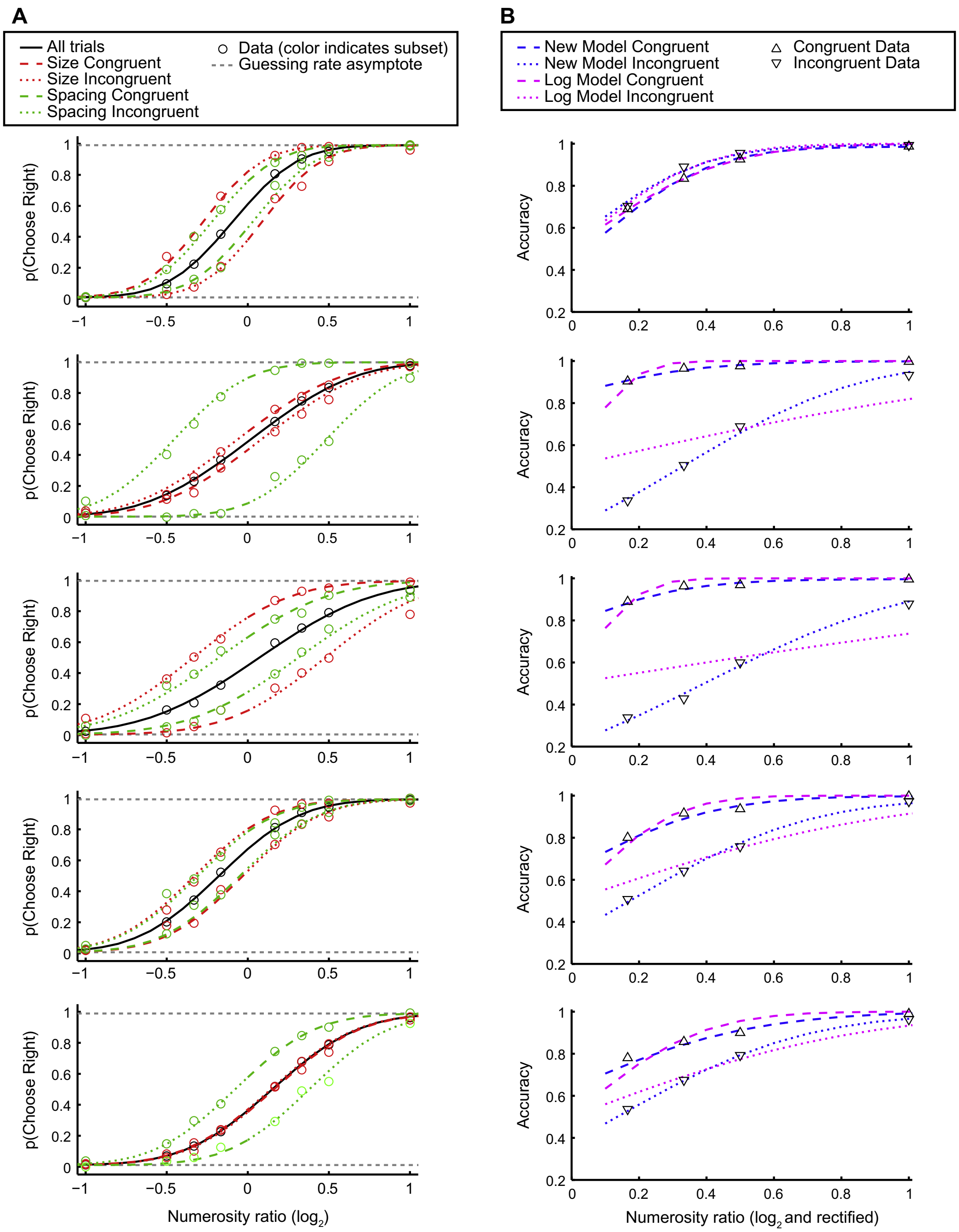

Fig. 3. Modeling *Size* and *Spacing* explains congruity effects. Each row of plots is a single participant's data. (A) Data (open circles) and model fit (black, red, and green lines). Dashed gray lines indicate model asymptote due to guessing rate (γ). The probability of choosing the stimulus array presented on the right is modeled as a function of the log of the left array to right array ratio of numerosity, *Size*, and *Spacing*. Black indicates the average of all data and the corresponding model fit. Red shows data and model fit for the third of trials with the greatest absolute ratio in *Size* and green shows the same for *Spacing*. Dashed lines indicate *Size* or *Spacing* was congruent with number, and dotted lines indicate incongruent. All model predictions (lines) are derived from the parameters fit once to the entire participant's dataset. (B) Data from the same participants plotted as accuracy. Upward pointing triangles indicate average data from all trials on which both *Size* and *Spacing* were congruent with numerosity, and downward pointing triangles indicate data from incongruent trials. Dashed lines are models fit only to congruent data points, and dotted lines are fit only to incongruent data points. Blue lines were fit using the new model presented in this paper, and magenta lines using the standard logarithmic model based on Piazza et al. (2010, 2004); linear model fit lines were excluded because they overlapped almost perfectly with the predictions of the log model and were difficult to visually distinguish.

coefficient). As a result, when numerosity is congruent with *Size* or *Spacing*, a participant will be more likely to correctly identify the stimulus with the larger numerosity, as illustrated by the dashed red and green model fit lines in Fig. 3A. In contrast, when *Size* or *Spacing* is incongruent, performance decreases as shown by the dotted red and green lines. In these trials, the larger and more spaced out dots make the less numerous stimulus appear more numerous and thus reduce accuracy. When the numerical ratio is difficult and the changes in *Size* and *Spacing* are large and incongruent to numerosity, participants can be induced to consistently incorrectly choose the less numerous stimulus at a rate greater than chance.

4.2. Modeling the effect of Size and Spacing improves w reliability across stimulus sets

One clear inadequacy of the previous models of numerical discrimination is that they are incapable of modeling below chance performance. In particular, when stimulus features are very incongruent to numerosity and the numerical ratios are difficult, the current models of numerosity discrimination will sometimes fail to converge, essentially estimating an absurdly large or infinite *w* (Szucs, Nobes, et al., 2013). We fit our model as well as the standard logarithmic (Piazza et al., 2010) and linear models (Pica et al., 2004) separately to only the congruent trials and only the incongruent trials to see if accounting for non-numerical feature bias helped reduce the variability in *w* estimates obtained from different stimulus sets. We considered a congruent trial as one in which the array containing more dots also had a larger *Size* and larger *Spacing* and incongruent as one in which the array containing more dots had the smaller *Size* and *Spacing*. As shown in Fig. 3B, our model provides much better fits to the data than the standard logarithmic model (the fit of the linear model was not plotted because it overlapped so closely with the log model that it was difficult to distinguish). The fit is especially better on the difficult incongruent trials, on which some participants performed consistently below chance on the difficult numerical ratios.

Furthermore, the inter-method reliability of *w* was higher and *w* was more similar to the *w* obtained from fitting the full stimulus set for our model compared to the two other models. Fig. 4 summarizes these results. When

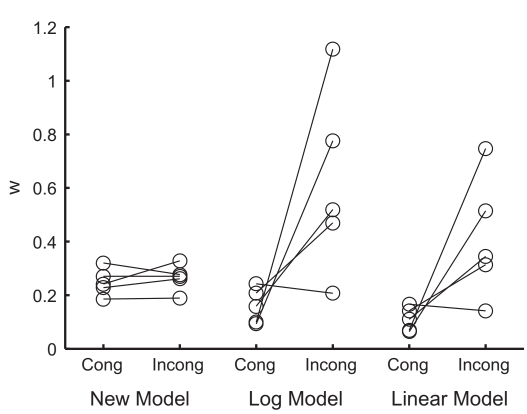

Fig. 4. Modeling the effects of non-numerical features increases the inter-method reliability of *w* over changes in stimulus set. The *w* coefficients calculated for the five participants who completed 7500 trials by fitting the model presented in this paper, the logarithmic model (Piazza et al., 2004, 2010), and the linear model (Pica et al., 2004). Models were fit separately to just the data from congruent trials and incongruent trials.

using the logarithmic and linear models, *w* obtained from incongruent trials tended to be much larger than for congruent trials; however, *w* obtained from our model was quite similar across stimulus sets, with no discernable increasing or decreasing trend across the five participants. Thus, our model is capable of explaining large differences in performance originating from non-numerical features. Furthermore, it properly attributes these performance effects to non-numerical model parameters with no systematic impact on the estimation of numerical acuity.

4.3. Numerosity is the best explanation of performance but bias is universal

Fig. 5 summarizes the effect of numerical ratio, *Size* ratio, and *Spacing* ratio on choice behavior for all of our 20 participants. Beta estimates are plotted in pairs as $\beta_{num} \times \beta_{Size}$, $\beta_{num} \times \beta_{Spacing}$, and $\beta_{Size} \times \beta_{Spacing}$ in Fig. 5A–C respectively (see the Supplementary Animated Figure for a fully three dimensional plot). The coefficients are plotted as standard error ellipses to denote the confidence of the estimate. The small ellipses represent more precise beta estimates derived from the five participants who performed 7500 trials (some errors are so small they may appear as points), and larger ellipses reflect the 15 participants who performed 750 trials each. These three beta estimates comprise the discrimination vector of each participant. The direction of the discrimination vector (the discrimination dimension) represents what stimulus features a participant is utilizing to make her choices, and the magnitude of the discrimination vector represents the participant's acuity in discriminating that feature (see Section 3.7 for further explanation).

The hypothesis that numerosity is the sole factor driving behavior is equivalent to the hypothesis that the beta for numerosity is significantly different from zero, and *Size* and *Spacing* betas are not different from zero. Although the choices of all of the participants in our sample were significantly influenced by numerosity ($p \ll 0.001$), none of them met our criterion for "pure" numerosity discriminators. In other words, we could not rule out the possibility that *Size* and *Spacing* might also be influencing behavior ($p < 0.1$ for all participants), and so we categorized them as "biased". Thus, we could rule out what we consider to be the implicit hypothesis of the two dominant models of numerosity discrimination: that the ratio of the numerosities and ANS acuity (*w*) are the only factors affecting numerosity discrimination performance.

The beta space represented in Fig. 5A–C, Supplementary Fig. 1, and the Supplementary Animated Figure is analogous to the stimulus space in Fig. 1E. In particular, the log ratios of the non-numerical stimulus features can be expressed as linear combinations of the log ratios of numerosity, *Size*, and *Spacing*. Thus, the hypothesis that a particular stimulus feature contributes to choice behavior is equivalent to the hypothesis that a particular linear combination of log numerosity, *Size*, and *Spacing* shape choice behavior. These linear combinations are represented as "feature" dimensions in Fig. 5A–C, Supplementary Fig. 1,

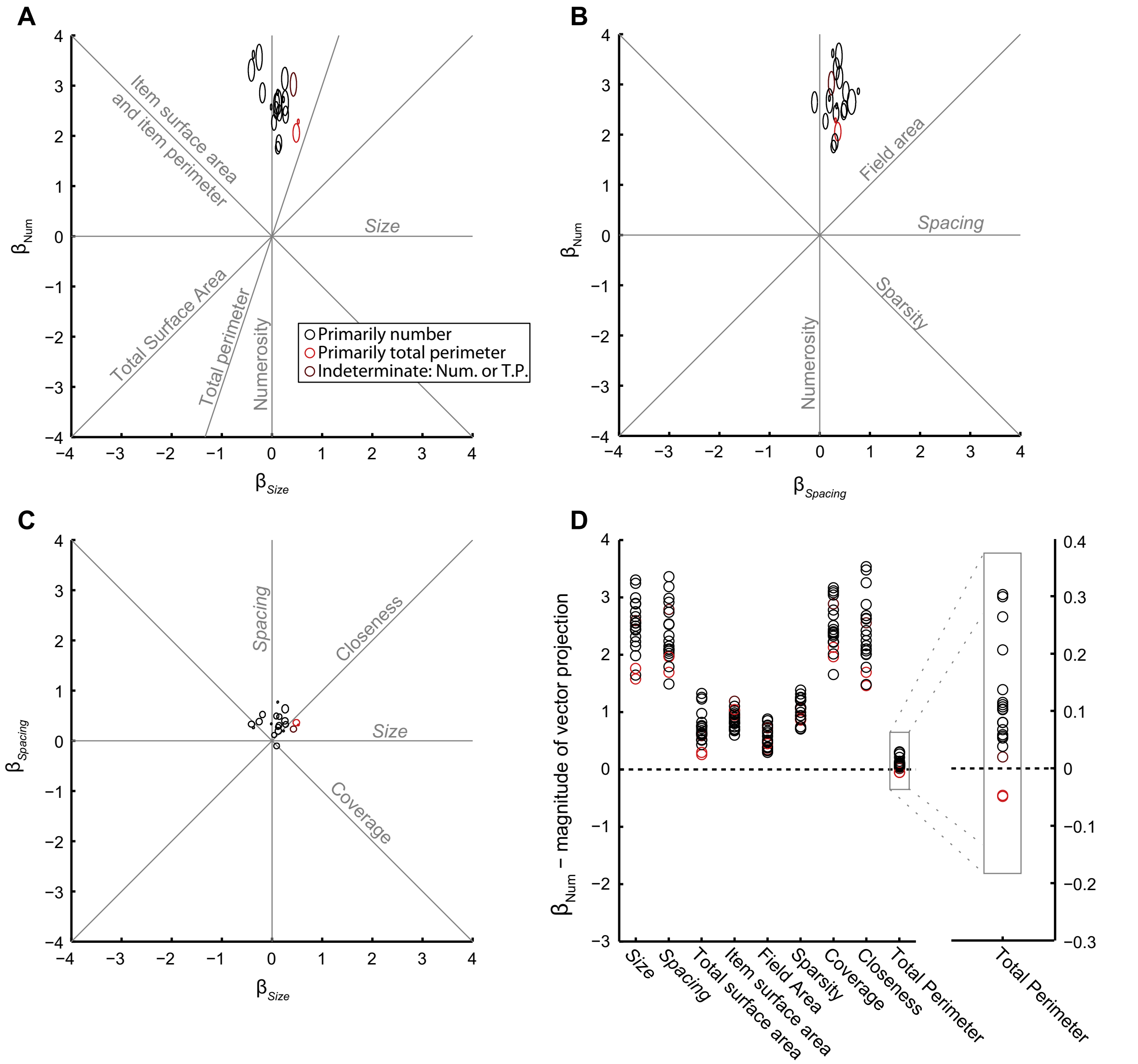

Fig. 5. No single feature fully explains choice behavior in any participant, but numerosity is the closest approximation for 18 out of 20 participants. (A–C) Estimates of effect sizes of the log of the ratio of numerosity, *Size*, and *Spacing* plotted in pairs for all 20 participants represented as standard error ellipses. Stimulus feature dimensions are shown as gray lines. The proximity of participants' coefficients (discrimination vector) to the gray dimension lines indicates how well that feature describes their choices. Black ellipses indicate the 17 participants whose choices were significantly better explained by numerical ratio than by any other stimulus feature, red indicates the two significantly better explained by total perimeter, and dark red the one participant who may have been choosing either on the basis of number or total perimeter (the trend is toward number). (D) The magnitude of the projection of the discrimination vector onto each non-numerical dimension subtracted from the magnitude of the projection onto the numerosity dimension. Positive values indicate that number was a better explanation of choices than the feature it is being contrasted against. Negative values indicate that the non-numerical feature was a better explanation of choices. Colors indicate the same participant groups as above. Axis on the right side is the same data from the contrast between number and total perimeter expanded for easier inspection. Fig. 5 is redrawn in Supplementary Fig. 1 showing more detail for the five participants who completed 7500 trials. The same data is recapitulated again for all participants in the Supplementary Animated Figure showing the three dimensional relationships between the dimensions and participant data.

and the Supplementary Animated Figure and labeled with the stimulus feature to which they correspond.

This produces a simple graphical representation of the features driving individual participants' choice behavior. For example, if we had found a participant who made choices based only on the number of dots while ignoring the dots' size and spacing, the discrimination vector error ellipse for that participant would lie on the numerosity feature dimension in both Fig. 5A and B and at the origin in Fig. 5C. Alternatively, if a participant always and only relied on total surface area to discriminate stimuli, her beta parameters for numerosity and *Size* would be significantly positive and equal, but the parameter for *Spacing* would be near zero. The exact numerical values of the numerosity and *Size* beta parameters would indicate the participant's acuity in discriminating total surface area. As a result such a participant's discrimination vector would fall along the "total surface area" dimension in Fig. 5A. The slope of each of these feature dimension lines in Fig. 5 is determined by the linear equations in Appendix A. Geometrically, asking which stimulus feature is determining behavior is equivalent to asking which feature dimension a participant's discrimination vector lies on. If the discrimination vector is significantly offset from all features lines, we may ask which feature *best* explains behavior. This is the equivalent of asking to which feature dimension the discrimination vector is closest. The three dimensional nature of the coefficient space can be better appreciated in the Supplementary Animated Figure.

To determine if a participant's discrimination vector was significantly closer to a non-numerical feature dimension or to the numerosity dimension, we projected the discrimination vectors onto each feature dimension. We then tested whether β_{num} was significantly larger than all the other vector projections using linear contrast hypothesis testing. If a particular vector projection was significantly larger than β_{num} we could conclude that the participant was more influenced by that parameter than by numerosity. Fig. 5D shows the difference between β_{num} and the magnitude of the discrimination vector projected onto each of the other stimulus features. Two out of twenty participants were significantly better described as basing choices on total perimeter rather than numerosity ($p = 0.009$ and $p \ll 0.001$ for linear contrast). One additional participant could not be categorized and was either a numerosity or a total perimeter discriminator ($p = 0.31$). All 17 other participants were significantly better described as discriminating numerosity than as discriminating any other stimulus feature ($p < 0.05$).

4.4. Advantages of new model hold with fewer trials

We ran our participants on many trials to ensure that we were able to precisely quantify their response strategies and bias terms. However, most studies of numerical cognition run fewer trials. If we hope for broader adoption of our modeling approach then it would be useful to know if the main advantages outlined here apply when fewer trials are used. We reran our analyses on reduced numbers of trials for all participants. Fig. 6A shows the proportion of the 20 participants that met the criteria outlined above for the effect of non-numerical features. We tested for bias: that some factor besides numerosity affects choices, and for strategy: that numerosity or some other factor is the statistically significant best determinate of choices. Although the ability to detect bias and strategy decreases with fewer trials, both measures were effective to as few as 250 trials, which our participants completed in about fifteen minutes. It is worth noting that a study designed with larger *Size* and *Spacing* ratios would be more sensitive to bias using even fewer trials.

Fig. 6B recapitulates the finding in Fig. 4 for all participants and using fewer trials. Inter-method reliability of w is higher using our new model than for either the logarithmic or linear models, both of which overestimate w when non-numerical stimulus features are incongruent with number. On average this is true regardless of the number of trials, but the standard error of the ratio increases with fewer trials.

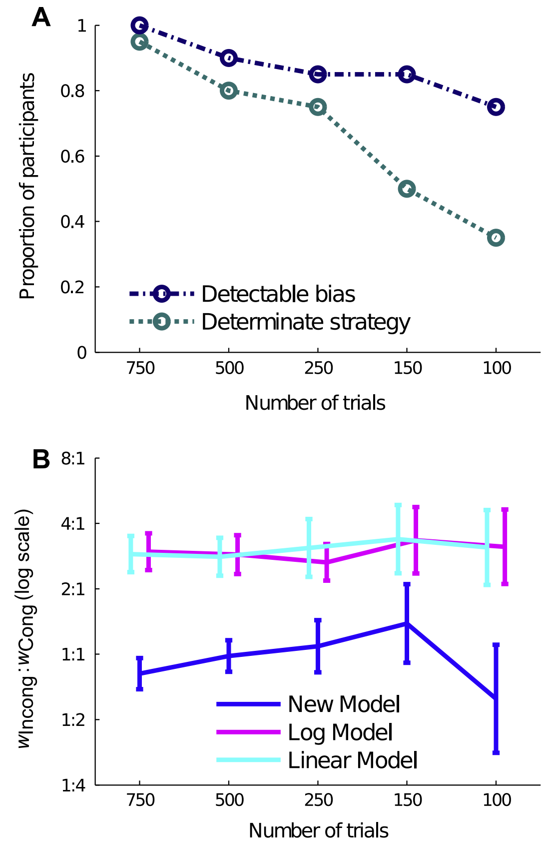

Fig. 6. The advantages of the new model are evident even with fewer trials. (A) The proportion of participants in whom we could detect bias is plotted against number of trials (navy). The proportion of participants for whom one of the ten features explains the significantly greatest portion of the variance (determinate strategy; teal) is also plotted against the number of trials used in the analysis. (B) Extending the analysis in Fig. 4 to all participants, w was calculated separately for congruent and incongruent trials (w_{Cong} and w_{Incong} respectively). The mean of the ratio of w_{Incong} to w_{Cong} is plotted on a log scale for the three models. Error bars denote standard errors of the means ($n = 20$ participants).

5. Discussion

Here we put forward an elaboration of the logarithmic model of approximate enumeration comparisons (Piazza et al., 2004) to account for the effect of non-numerical features associated with the size and spacing of items in a visual array. At its heart the logarithmic model is a generalized linear model of error rate using the log of the numerical ratio as its sole regressor. Our key innovation is that we added regressors for the size and spacing of the items, so that variance in choice probability can be attributed to non-numerical features of the stimulus. The main complication arises from the multitude of visual features related to size (item surface area, total surface area, item perimeter and total perimeter), spacing (field area and sparsity/density), or both (coverage and apparent closeness) that could influence discrimination. To disentangle the roles of these various features we designed the size and spacing regressors such that they were linearly related to all features. As a result, discriminations based on a particular visual feature, such as total perimeter, result in a unique ratio of number, *Size*, and *Spacing* regression coefficients returned by the model (these ratios are geometrically equivalent to the gray feature dimension lines in Fig. 5 and the Supplementary Animated Figure). By testing which of these unique ratios of coefficients is most similar to a particular set of coefficients (equivalent to which feature dimension line a participant's performance is closest to), we can determine whether a participant is discriminating based on number or some other visual feature of the stimuli. By examining how far the coefficients are from the unique coefficient combination associated with number, we can quantify the extent of non-numerical bias.

This model and the stimulus space on which it is based represent an advance over previous approaches in four important ways. First, the stimulus space itself identifies, for the first time, the three degrees of freedom available to nonverbal numerical cognition researchers in designing stimuli, and elucidates the tradeoffs and partial collinearities inherent to arrays of dots. Second, the model reconciles the concept of numerosity as an internal random variable on a log scale mental number line with the fact that non-numerical features also affect discrimination performance by allowing the mean of the numerosity random variable to shift with changes in the size and spacing of the dots within the stimulus arrays. Third, by correctly attributing correct and incorrect responses caused by congruence or incongruence of non-numerical features to the *Size* and *Spacing* parameters, the model yields a *w* that is a more valid estimate of numerical acuity and is more reliable over different stimulus sets. Fourth, our model provides a quantitative assessment of the role of non-numerical stimulus features in numerosity judgments. Rather than attempting to control for non-numerical features by equating different dimensions in different trial subsets, our approach is to intentionally vary non-numerical features and to model their effects on performance.

Our model applied to the data set presented here demonstrates that all participants were influenced, to some extent, by non-numerical features while attempting to perform a numerical discrimination. Although the effect of non-numerical features on numerical estimation and comparison has been well documented in the literature, our data show that these effects are nearly universal even among educated adults. Our data further show that for at least 17 out of 20 subjects, number, out of the comprehensive list of ten stimulus features tested, best explained behavior. We consider this strong evidence that numerosity does exist as an internal magnitude. Number cannot be explained away as "merely" the derived effect of other features even if other features inarguably affect numerical perception.

5.1. Stimulus space and modeling

The two dominant models of numerosity discrimination do not adequately account for the effects of non-numerical stimulus dimensions on accuracy. Both the logarithmic (Piazza et al., 2004, 2010) and linear (Halberda et al., 2008; Pica et al., 2004) models posit that numerosity is internally represented as a distribution or random variable along a mental number line, with a mean equal to the number represented. The width of the distribution may be fixed (log model) or may vary with the magnitude being represented (linear model). In either case *w* is proposed to be a measure of the fuzziness of the internal representation of number intrinsic to the individual. Critically, both models posit that performance in a numerical ordering task is determined only by *w* and the numerical ratio. Empirically, however, many groups have demonstrated that non-numerical stimulus features do indeed affect performance in numerical ordering tasks (e.g. DeWind & Brannon, 2012; Frith & Frith, 1972; Gebuis & Gevers, 2011; Ginsburg, 1976; Sophian, 2007; Tokita & Ishiguchi, 2010).

Here we extend the standard logarithmic model of numerosity perception and discrimination to include terms that capture the effects of the size and spacing of the dots in the stimulus arrays. In our new revised logarithmic model, numerosity is represented as a random variable on a log compressed mental number line. However, the size and spacing of the dots in the stimulus array can cause the mean of this distribution to be shifted to a position greater or less than the actual number of items in the stimulus. As a result, the overlap of two numerosity distributions, and therefore the predicted error rate, may be larger or smaller depending not only on numerical ratio but also on whether non-numerical features are congruent or incongruent with numerosity (see Fig. 2 for a hypothetical example). By accommodating the effects of non-numerical features, our model is able to capture variance in numerical discrimination behavior that went unaccounted for in previous models.

Extending the logarithmic model of numerosity, however, is not as straightforward as simply adding regressors for each non-numerical feature that might influence numerical perception. Such a model would be overdetermined due to the partial collinearity of these features. It was therefore essential to identify the mutually independent regressors that fully describe the stimulus features that could affect performance on a numerosity discrimination task.

Relying heavily on the framework of intrinsic and extrinsic features pioneered by Dehaene et al. (2005), we developed a novel stimulus space. For the first time we provide a comprehensive description of dot array stimuli that encompasses the critical features affecting numerical discrimination. This space has three dimensions that describe the number, size, and spacing of the dots in an array, a formulation that is complete but not redundant. This space provides ANS researchers with a powerful new tool for understanding the tradeoffs and collinearities inherent to dot array stimuli and a basis for quantifying the effects of ten different stimulus features on numerical discrimination.

5.2. A more valid and reliable *w*

In previous studies fitting the logarithmic or linear models, estimates of *w* have failed to account for the effect of non-numerical features in a systematic and quantifiable way. As a result, *w* estimates derived from these models are an amalgamation of the effects of number and non-numerical features. Our model estimates *w* independently of the effects of *Size* and *Spacing*, and in this sense it is a more valid measure of numerical acuity itself.

The practical corollary of a more valid measure of numerical acuity is an increase in "inter-method" or "alternate-form" reliability. This type of reliability refers to the tendency of different tests to generate the same result. In this case the different tests of numerical acuity are different stimulus sets that vary non-numerical features in different ways. To assess inter-method reliability we compared the *w* estimates for the new revised logarithmic model to the two standard models for congruent and incongruent trials separately. For both standard models, *w* estimates were much higher in the incongruent compared to the congruent condition. In contrast, our new revised logarithmic model returned similar estimates for the two trial types and therefore showed greater inter-method reliability. The discrepancy in *w* estimates for incongruent and congruent trials observed under the standard models has been observed previously and has been interpreted to mean there is no stable internal representation of numerical magnitude; it has even been offered as evidence against the existence of the ANS (Szucs, Nobes, et al., 2013). Our model demonstrates that *w* is in fact stable over these stimulus conditions and that the instability observed in previous studies was due to the conflation of acuity and bias.

There are several potential benefits of a more reliable, valid, and cross-paradigm comparable measure of numerical acuity. Recently, there has been interest in the predictive power of numerical acuity on mathematical achievement (DeWind & Brannon, 2012; Gilmore et al., 2010, 2013; Halberda et al., 2008, 2012; Lyons & Beilock, 2011; Mazzocco et al., 2011; Park & Brannon, 2013; Piazza et al., 2010; Starr et al., 2013). These correlations, however, are relatively weak and only predict a small amount of variance in mathematical performance. Some researchers have also suggested that non-symbolic numerical abilities are part of a larger suite of visual-perceptual abilities that predict mathematics performance (Tibber et al., 2013). Others have argued that ANS acuity provides unique variance to predicting mathematical performance and that other similar perceptual tasks do not (Agrillo, Piffer, & Adriano, 2013), or that both ANS acuity and other perceptual tasks provide unique variance (Lourenco, Bonny, Fernandez, & Rao, 2012). Parsing out non-numerical bias from numerical acuity may improve these correlations by reducing the effect of bias on *w* estimates. Alternatively, bias itself might be a mediating factor. Participants who cannot clearly differentiate numerosity from other magnitudes may have impaired performance on other perceptual tasks or with symbolic mathematics itself. For example, some have suggested that the "stroop like" aspect of numerosity discriminations with strong non-numerical feature incongruity reveals difficulties inhibiting prepotent responses (Fuhs & McNeil, 2013; Szucs, Devine, Soltesz, Nobes, & Gabriel, 2013). They suggest that previous correlations between *w* and math achievement may be mediated by failure to inhibit responses to other stimulus features. Isolating bias from numerical acuity will allow these hypotheses to be tested more directly.

5.3. Stimulus control

The goal of stimulus control in numerosity experiments has been to ensure that numerosity is driving choice behavior. In the literature there are two standard ways of accomplishing stimulus control in dot array comparison tasks. Both of these methods have drawbacks, and neither gives an objective measure of non-numerical feature bias. Our modeling approach, in contrast, provides a clear quantitative measure of both numerical acuity and bias and can detect alternative response strategies that are based primarily on non-numerical features of the stimulus.

The most common method for controlling non-numerical features is to divide trials into sets that each control for a different non-numerical stimulus feature, an approach adopted by many research groups (e.g. Ansari & Dhital, 2006; Halberda et al., 2008; Izard, Sann, Spelke, & Streri, 2009; Libertus, Woldorff, & Brannon, 2007; Piazza et al., 2010; Santens, Roggeman, Fias, & Verguts, 2010). For example, if total surface area were fixed in one set of trials, numerosity and item surface area would vary together. In another set of trials, item surface area would be fixed and total surface area would vary with numerosity. If a participant were relying on one of these features as a proxy for numerosity, then choice behavior would be at chance on the subset of trials on which that feature is fixed. This basic logic certainly works for ruling out total reliance on a particular feature; however, the analysis is underpowered since it relies on a subset of trials. Furthermore, for practical reasons, most studies do not control all possible parameters in different trial subsets, a problem which is particularly salient given our finding that total perimeter, a rarely controlled parameter, is subserving a non-numerical strategy in some people.

Another common approach to stimulus control is to have subsets of trials in which a particular non-numerical stimulus feature is varied in a manner either congruent with or incongruent with numerosity (e.g. Cantlon & Brannon, 2005; DeWind & Brannon, 2012; Hurewitz, Gelman, & Schnitzer, 2006; Rousselle & Noël, 2008; Szucs, Nobes, et al., 2013). Ruling out a non-numerical strategy is particularly problematic using this paradigm. It depends on observing above chance performance in incongruent trials, but as can be seen in Fig. 3B, participants with any bias at all can be induced to consistently choose the incorrect stimulus when numerical ratios are very difficult and the ratio of incongruent non-numerical features is very large. Thus, the test for non-numerical strategies is too sensitive and can interpret small effects of non-numerical features as total reliance on them (as in Szucs, Nobes, et al., 2013). The sensitivity of the test depends on paradigm idiosyncrasies such as the difficulty of the numerical ratios and the degree of variation in non-numerical features. Furthermore, like the first method mentioned above, a design that attempts to control for all non-numerical stimulus features using a congruent-incongruent paradigm would require a multitude of conditions (e.g., perimeter congruent, surface area incongruent, etc.). A recent paper claimed to have resolved the problem of stimulus control for dot arrays using this congruent-incongruent approach (Gebuis & Reynvoet, 2011). While their approach succeeds in reducing the overall correlation between number and non-numerical features, it still suffers from the same intrinsic problem outlined above.

A critical insight derived from our stimulus space and modeling approach is that non-numerical bias, which we define as the marginal effect of a non-numerical feature on choices, and a non-numerical strategy, which we define as the primary reliance on non-numerical features, exist on a continuum. By varying numerical and non-numerical stimulus features and modeling their effect on choices, our paradigm provides a quantitative measure of non-numerical feature bias. If these terms are sufficiently large, then choices will be better described by the ratio of a non-numerical feature than by the ratio of numerosity itself, and we consider such a participant to be utilizing a non-numerical strategy. Furthermore, these analyses are made based on the entire dataset, rather than subsets of trials, and therefore have more statistical power.

This study represents the most comprehensive effort of which we are aware to simultaneously quantify the effect of as many non-numerical features as possible on the internal representation of number. Thus, although bias was universal among our participants, it is worth noting that 17 out of 20 participants used numerosity more than any other feature to make their discriminations. We take this as evidence that numerosity is not reducible to "merely" the effects of other stimulus features as suggested by some (Gebuis & Reynvoet, 2012b,c; Szucs, Nobes, et al., 2013), but is itself an important determinant of behavior. Two of the twenty participants, however, had such large *Size* bias that it was more parsimonious to describe them as discriminating total perimeter than as discriminating numerosity. This finding makes the importance of controlling total perimeter in ANS studies apparent.

We also found a large and relatively consistent effect of *Spacing* on numerosity judgments. Nineteen out of twenty participants viewed arrays with more spaced out dots as more numerous. This effect has been noted before (Allïk & Tuulmets, 1991; Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011; Kramer, Di Bono, & Zorzi, 2011) and may provide some insight into the processes by which numerosity is extracted from the visual scene.

5.4. Approximate number system or an approximate magnitude system?

We use the term ANS throughout this paper; however, it is important to emphasize that our model and data set are not designed to test for the existence or lack of an independent representational system for number. A recent study purported to provide evidence against the existence of an ANS based on low within subject reliability between stimulus sets for which non-numerical variables were congruent or incongruent with number (Szucs, Nobes, et al. 2013). However, as we explained in Sections 5.2 and 5.3, the low reliability they obtained can be attributed to failing to model non-numerical features, not the instability of the numerical representation. Instead, we found that number was a significant determinate of choices, and non-numerical features had a ubiquitous but secondary role.

One of the advantages of our stimulus space and model is that it illustrates the close relationship between number and other features of the stimulus. In Fig. 5A–C especially, it is clear that a small effect of *Size* or *Spacing* can be considered a marginal biasing effect on numerical discrimination, but a sufficiently large effect is better described as an alternative response strategy (albeit likely an unconscious one). This continuum of effects and the fact that a minority do indeed rely on perimeter in making numerical judgments may lend support to the idea that, rather than an ANS, there is a more general approximate magnitude system that allows approximate enumeration, but also subserves approximation of other continuous properties of a stimulus. From this perspective, our findings can be seen as supporting Walsh's theory of magnitude (ATOM), which suggests that all magnitudes share a common currency, or at least overlapping representation in the brain (Bueti & Walsh, 2009; Cantlon, Platt, & Brannon, 2009; Walsh, 2003).

A related question is how the representation of number and continuous variables emerges over human development. One possibility is that numerosity is conflated with other magnitudes early in development, but that over development, numerosity becomes more differentiated (Lourenco & Longo, 2010; Walsh, 2003). Within the context of our model, confusion of different stimulus dimensions would manifest itself in the magnitude of the *Size* and *Spacing* coefficients. The classic Piagetian view is that early in development children attend to size and volume and only later come to appreciate number as an abstract variable. Seemingly consistent with this view, a handful of studies found that perimeter or area is more readily encoded by infants than number (Clearfield & Mix, 1999, 2001; Feigenson, Carey, & Spelke, 2002). However, other data is inconsistent with this view and suggests that infants spontaneously encode both kinds of information. For example, Libertus, Starr, and Brannon (2014) used a visual change detection paradigm and found that when infants were shown two streams of visual images where one stream alternated numerically and the other alternated in total surface area, infants preferred to look at the numerically changing stream (see also Cordes & Brannon, 2009). We hope that our model can be used to assess the relative influence of number and continuous variables in young children's decision-making and track changes in numerical sensitivity and bias over development.

Relatedly, comparative studies of other species have examined relative use of number and other features. Monkeys and many other animal species can be trained to attend to number and largely ignore other visual features (e.g. Brannon & Terrace, 1998; Cantlon & Brannon, 2006). It has been suggested, however, that this ability is not part of animals natural behavioral repertoire and only results from extensive training (e.g. Seron & Pesenti, 2001). Cantlon and Brannon (2007) offered evidence against this view. They trained rhesus monkeys to match stimuli based on numerosity and a redundant non-numerical variable such as color, shape, or surface area. Once monkeys reliably matched these redundant cue stimuli, they were given a choice between one stimulus that matched the sample numerically and another stimulus that matched based on the previously redundant variable (e.g., color, shape, or surface area). The monkeys' decisions were strongly influenced by the numerical distance between the sample and incorrect numerical match (Cantlon & Brannon, 2007). Furthermore, this was true even for one monkey who had no prior numerical training. These data suggest that number is a salient discrimination cue even when monkeys can rely on other visual features. In contrast, research with some other species such as mosquito fish suggests that their quantitative judgments may be more influenced by continuous variables (Agrillo, Piffer, & Bisazza, 2011).

Thus, the preponderance of recent developmental and comparative evidence suggests that number is more than a "last resort" strategy for disambiguating stimuli. We hope that our model can be used to quantify the role of number versus non-numerical variables on behavioral decisions, as well as to study species differences and changes as a function of development and experience.

5.5. Future directions

An important future direction is to see how well non-numerical bias can be estimated in previously collected published datasets on numerical cognition and to see how these bias estimates, as well as the new estimates of numerical acuity that account for bias related errors, change or clarify previous hypotheses. Our modeling approach does not depend on a particular esoteric arrangement of stimulus parameters. We orthogonalized number, *Size*, and *Spacing* ratios to increase power; however, as long as these features are not perfectly collinear, our modeling approach can be applied to data sets acquired using diverse stimulus control paradigms. To facilitate the adoption of our model, we have included computer code in a supplement to this research article.

An important advantage inherent to our model of choice behavior is that it easily accommodates more regressors to model other important aspects of choice behavior. We argue that the advance made in this paper is the observation that the continuous parameters thought to affect numerical discriminations can be reduced to three regressors that can be varied independently. A fourth regressor was added to account for side bias. More regressors could be added for other variables that can be varied independently from numerosity, *Size*, and *Spacing*: for example, brightness, contrast ratio, or item shape.

The standard way of measuring w is to present pairs of dot arrays and require participants to make an ordinal judgment. The number of parameters that can freely vary between research groups without being expressly modeled is shrinking. Here we modeled the effects of *Size* and *Spacing* for the first time and, by extension, all of the derived features in Appendix A. Our model, however, cannot explain some other features known to affect the perception of numerosity. First, our model does not account for the effect of stimulus exposure time. Inglis and Gilmore (2013) demonstrated that stimulus exposure time is a critical variable that must be accounted for when estimating w, and they provide a model for doing so. Second, although our model contains a term for item spacing, it cannot account for the effect of items "clumping" within the array. The solitaire illusion (Frith & Frith, 1972) and the regular-random illusion (Ginsburg, 1976) demonstrate that clumping does affect numerosity estimates. The occupancy model (Allïk & Tuulmets, 1991) provides a modeling framework for explaining these effects. Other effects that our model does not address are the effect of the absolute magnitude of the values being compared separate from ratio (Prather, 2014) and hysteresis, whereby the difficulty and perceptual qualities of the previous trial affect current discrimination (Cicchini, Anobile, & Burr, 2014; Odic, Hock, & Halberda, 2014). Integrating the effects of these visual features into our model is beyond the scope of this paper. However, future work should explore the interactions between exposure time, clumping, absolute magnitude, hysteresis, and numerical acuity and bias to further reconcile different paradigms, aid in comparisons across paradigms and research groups, and deepen our understanding of the mechanisms of approximate enumeration.

Although most investigations into the ANS use static arrays of dots, similar stimulus control problems exist for aural or visual numerical stimuli presented sequentially. The extrinsic variables analogous to total surface area and field area would be total event duration and total stimulus duration respectively. The intrinsic variables analogous to item surface area and sparsity would be individual event duration and mean event period (equivalently total stimulus duration per event or the reciprocal of frequency). Numerosity-independent variables analogous to *Size* and *Spacing* could be generated by the same equations, and a regression model closely analogous to the one presented here could be adopted.

We focused on choice behavior in this study; however, the stimulus space and model could be used on any dependent variable that might vary with *Size*, *Spacing* and numerosity. For example, various studies have looked at the effects of dot array numerosity on BOLD signal (Cantlon, Brannon, Carter, & Pelphrey, 2006; Jacob & Nieder, 2009; Piazza, Pinel, Le Bihan, & Dehaene, 2007; Piazza et al., 2004), EEG (Gebuis & Reynvoet, 2012a), and the firing rates of individual neurons (Nieder & Miller, 2004; Roitman, Brannon, & Platt, 2007). Currently, non-numerical features are treated as nuisance variables that must be controlled. Our approach of quantifying non-numerical features allows the stimulus space dimensions affecting neurological dependent variables to be teased apart. We hope that applying similar modeling approaches to the one used here will lead to a better understanding of how low level visual features processed early in the cortical visual stream are transformed into the numerosity signals seen in the intraparietal sulcus and prefrontal cortex.

Finally, we anticipate that this model will be useful for looking at changes in the salience of non-numerical features over development and individual differences in the influence of non-numerical variables on numerical discrimination at a given age.

6. Conclusions

We extended the logarithmic model of numerical acuity to dissociate the biasing effects of *Size* and *Spacing* from *w*. Instead of merely controlling for non-numerical stimulus features, the model allows a quantification of the effect of non-numerical stimulus features on choices. The model applied to our data set demonstrates that non-numerical features widely affect numerical discriminations in adults, but that for most individuals these effects are relatively small.

Acknowledgements

We would like to thank Vanessa Bermudez for assistance in collecting the data, and Dr. Joonkoo Park, Dr. John Pearson, Monica Carlson, Ariel Starr, Caroline Drucker, and Rosa Li for helpful discussions. We would also like to acknowledge our funding sources: Broad Research Award for Graduate Students to N.K.D., McDonnell Scholar award to E.M.B., and NEI 1R01EY019303 to M.L.P.

Appendix A

Stimulus features in terms of the three cardinal features, numerosity (n), *Size* (Sz), and *Spacing* (Sp), or the ratios of those features ($r_{feature}$).

Stimulus feature	Feature in terms of three cardinal features	Log of feature in terms of log of three cardinal features	Log of feature ratio in terms of log of three cardinal ratios
Total surface area (TSA)	$TSA = \sqrt{Sz \cdot n}$	$\log(TSA) = \frac{1}{2}\log(Sz) + \frac{1}{2}\log(n)$	$\log(r_{TSA}) = \frac{1}{2}\log(r_{Sz}) + \frac{1}{2}\log(r_n)$
Item surface area (ISA)	$ISA = \sqrt{\frac{Sz}{n}}$	$\log(ISA) = \frac{1}{2}\log(Sz) - \frac{1}{2}\log(n)$	$\log(r_{ISA}) = \frac{1}{2}\log(r_{Sz}) - \frac{1}{2}\log(r_n)$
Field area (FA)	$FA = \sqrt{Sp \cdot n}$	$\log(FA) = \frac{1}{2}\log(Sp) + \frac{1}{2}\log(n)$	$\log(r_{FA}) = \frac{1}{2}\log(r_{Sp}) + \frac{1}{2}\log(r_n)$
Sparsity (Spar)	$Spar = \sqrt{\frac{Sp}{n}}$	$\log(Spar) = \frac{1}{2}\log(Sp) - \frac{1}{2}\log(n)$	$\log(r_{Spar}) = \frac{1}{2}\log(r_{Sp}) - \frac{1}{2}\log(r_n)$
Total perimeter (TP)	$TP = 2\sqrt{\pi} \cdot Sz^{\frac{1}{4}} \cdot n^{\frac{3}{4}}$	$\log(\mathrm{TP}) = \log(2\sqrt{\pi}) + \frac{1}{4}\log(Sz) + \frac{3}{4}\log(n)$	$\log(r_{TP}) = \frac{1}{4}\log(r_{Sz}) + \frac{3}{4}\log(r_n)$
Item perimeter (IP)	$IP = 2\sqrt{\pi} \cdot Sz^{\frac{1}{4}} \cdot n^{-\frac{1}{4}}$	$\log(IP) = (\log(2\sqrt{\pi}) + \frac{1}{4}\log(Sz) - \frac{1}{4}\log(n)$	$\log(r_{IP}) = \frac{1}{4}\log(r_{Sz}) - \frac{1}{4}\log(r_n)$
Coverage (Cov)	$Cov = \sqrt{\frac{Sz}{Sp}}$	$\log(Cov) = \frac{1}{2}\log(Sz) - \frac{1}{2}\log(Sp)$	$\log(r_{Cov}) = \frac{1}{2}\log(r_{Sz}) - \frac{1}{2}\log(r_{Sp})$
Apparent closeness (AC)	$AC = \sqrt{Sz \cdot Sp}$	$\log(AC) = \frac{1}{2}\log(Sz) + \frac{1}{2}\log(Sp)$	$\log(r_{AP}) = \frac{1}{2}\log(r_{Sz}) + \frac{1}{2}\log(r_{Sp})$

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2015.05.016.

References

Agrillo, C., Piffer, L., & Adriano, A. (2013). Individual differences in non-symbolic numerical abilities predict mathematical achievements but contradict ATOM. *Behavioral and Brain Functions, 9*(1), 26.

Agrillo, C., Piffer, L., & Bisazza, A. (2011). Number versus continuous quantity in numerosity judgments by fish. *Cognition, 119*(2), 281–287. http://dx.doi.org/10.1016/j.cognition.2010.10.022.

Allïk, J., & Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Perception & Psychophysics, 49*(4), 303–314. http://dx.doi.org/10.3758/BF03205986.

Ansari, D., & Dhital, B. (2006). Age-related changes in the activation of the intraparietal sulcus during nonsymbolic magnitude processing: An event-related functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience, 18*(11), 1820–1828. http://dx.doi.org/10.1162/jocn.2006.18.11.1820.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by Monkeys. *Science, 282*(5389), 746–749. http://dx.doi.org/10.1126/science.282.5389.746.

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1525), 1831–1840. http://dx.doi.org/10.1098/rstb.2009.0028.

Cantlon, J. F., & Brannon, E. M. (2005). Semantic congruity affects numerical judgments similarly in monkeys and humans. *Proceedings of the National Academy of Sciences of the United States of America, 102*(45), 16507–16511.

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science, 17*(5), 401–406.

Cantlon, J. F., & Brannon, E. M. (2007). How much does number matter to a monkey (*Macaca mulatta*)? *Journal of Experimental Psychology: Animal Behavior Processes, 33*(1), 32–41. http://dx.doi.org/10.1037/0097-7403.33.1.32.

Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biology, 4*(5), e125.

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences, 13*(2), 83–91. http://dx.doi.org/10.1016/j.tics.2008.11.007.

Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences, 111*(21), 7867–7872. http://dx.doi.org/10.1073/pnas.1402785111.

Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science, 10*(5), 408–411. http://dx.doi.org/10.1111/1467-9280.00177.

Clearfield, M. W., & Mix, K. S. (2001). Amount versus number: Infants' use of area and contour length to discriminate small sets. *Journal of Cognition and Development, 2*(3), 243–260. http://dx.doi.org/10.1207/S15327647JCD0203_1.

Cordes, S., & Brannon, E. M. (2009). The relative salience of discrete and continuous quantity in young infants. *Developmental Science, 12*(3), 453–463. http://dx.doi.org/10.1111/j.1467-7687.2008.00781.x.

Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences, 108*(49), 19552–19557. http://dx.doi.org/10.1073/pnas.1113195108.

Dehaene, S., Izard, V., & Piazza, M. (2005). Control over non-numerical parameters in numerosity experiments. *Unpublished Manuscript*, <www. Unicog. Org>.

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: Effects of feedback and training. *Frontiers in Human Neuroscience, 6*, 68. http://dx.doi.org/10.3389/fnhum.2012.00068.

Feigenson, L., Carey, S., & Spelke, E. (2002). Infants' discrimination of number vs. continuous extent. *Cognitive Psychology, 44*(1), 33–66. http://dx.doi.org/10.1006/cogp.2001.0760.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314. http://dx.doi.org/10.1016/j.tics.2004.05.002.

Frith, C. D., & Frith, U. (1972). The solitaire illusion: An illusion of numerosity. *Perception & Psychophysics, 11*(6), 409–410.

Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: Contributions of inhibitory control. *Developmental Science, 16*(1), 136–148. http://dx.doi.org/10.1111/desc.12013.

Gebuis, T., & Gevers, W. (2011). Numerosities and space; indeed a cognitive illusion! A reply to de Hevia and Spelke (2009). *Cognition, 121*(2), 248–252.

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods, 43*(4), 981–986. http://dx.doi.org/10.3758/s13428-011-0097-5.

Gebuis, T., & Reynvoet, B. (2012a). Continuous visual properties explain neural responses to nonsymbolic number. *Psychophysiology, 49*(11), 1649–1659. http://dx.doi.org/10.1111/j.1469-8986.2012.01461.x.

Gebuis, T., & Reynvoet, B. (2012b). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General, 141*(4), 642–648. http://dx.doi.org/10.1037/a0026218.

Gebuis, T., & Reynvoet, B. (2012c). The role of visual information in numerosity estimation. *PLoS ONE, 7*(5), e37426. http://dx.doi.org/10.1371/journal.pone.0037426.

Gilmore, C. K., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., et al. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE, 8*(6), e67374. http://dx.doi.org/10.1371/journal.pone.0067374.

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition, 115*(3), 394–406. http://dx.doi.org/10.1016/j.cognition.2010.02.002.

Ginsburg, N. (1976). Effect of item arrangement on perceived numerosity: Randomness vs regularity. *Perceptual and Motor Skills, 42*(43), 663–668.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457–1465. http://dx.doi.org/10.1037/a0012682.

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences of the United States of America, 109*(28), 11116–11120. http://dx.doi.org/10.1073/pnas.1200196109.

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*(7213), 665–668. http://dx.doi.org/10.1038/nature07246.

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology, 103*(1), 17–29. http://dx.doi.org/10.1016/j.jecp.2008.04.001.

Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences, 103*(51), 19599–19604. http://dx.doi.org/10.1073/pnas.0609485103.

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition, 131*(1), 92–107. http://dx.doi.org/10.1016/j.cognition.2013.12.007.

Inglis, M., & Gilmore, C. K. (2013). Sampling from the mental number line: How are approximate number system representations formed? *Cognition, 129*(1), 63–69.

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences, 106*(25), 10382–10385. http://dx.doi.org/10.1073/pnas.0812142106.

Jacob, S. N., & Nieder, A. (2009). Tuning to non-symbolic proportions in the human frontoparietal cortex. *European Journal of Neuroscience, 30*(7), 1432–1442. http://dx.doi.org/10.1111/j.1460-9568.2009.06932.x.

Kramer, P., Di Bono, M. G., & Zorzi, M. (2011). Numerosity estimation in visual stimuli in the absence of luminance-based cues. *PLoS ONE, 6*(2), e17378. http://dx.doi.org/10.1371/journal.pone.0017378.

Libertus, M. E., Starr, A., & Brannon, E. M. (2014). Number trumps area for 7-month-old infants. *Developmental Psychology, 50*(1), 108–112. http://dx.doi.org/10.1037/a0032986.

Libertus, M. E., Woldorff, M. G., & Brannon, E. M. (2007). Electrophysiological evidence for notation independence in numerical processing. *Behavioral and Brain Functions, 3*(1), 1–15.

Lourenco, S. F., Bonny, J. W., Fernandez, E. P., & Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proceedings of the National Academy of Sciences, 109*(46), 18737–18742. http://dx.doi.org/10.1073/pnas.1207212109.

Lourenco, S. F., & Longo, M. R. (2010). General magnitude representation in human infants. *Psychological Science, 21*(6), 873–881. http://dx.doi.org/10.1177/0956797610370158.

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition, 121*(2), 256–261. http://dx.doi.org/10.1016/j.cognition.2011.07.009.

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development, 82*(4), 1224–1237. http://dx.doi.org/10.1111/j.1467-8624.2011.01608.x.

Merten, K., & Nieder, A. (2009). Compressed scaling of abstract numerosity representations in adult humans and monkeys. *Journal of Cognitive Neuroscience, 21*(2), 333–346. http://dx.doi.org/10.1162/jocn.2008.21032.

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron, 37*(1), 149–157. http://dx.doi.org/10.1016/S0896-6273(02)01144-3.

Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences of the United States of America, 101*(19), 7457–7462.

Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology: General, 143*(1), 255–265. http://dx.doi.org/10.1037/a0030825.

Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science, 24*(10), 2013–2019. http://dx.doi.org/10.1177/0956797613482944.

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116*(1), 33–41. http://dx.doi.org/10.1016/j.cognition.2010.03.012.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron, 44*(3), 547–555.

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron, 53*(2), 293–305.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science, 306*(5695), 499–503.

Prather, R. W. (2014). Numerical discrimination is mediated by neural coding variation. *Cognition, 133*(3), 601–610. http://dx.doi.org/10.1016/j.cognition.2014.08.003.

Roitman, J. D., Brannon, E. M., & Platt, M. L. (2007). Monotonic coding of numerosity in macaque lateral intraparietal area. *PLoS Biology, 5*(8), e208. http://dx.doi.org/10.1371/journal.pbio.0050208.

Rousselle, L., & Noël, M.-P. (2008). The development of automatic numerosity processing in preschoolers: Evidence for numerosity-perceptual interference. *Developmental Psychology, 44*(2), 544–560. http://dx.doi.org/10.1037/0012-1649.44.2.544.

Santens, S., Roggeman, C., Fias, W., & Verguts, T. (2010). Number processing pathways in human parietal cortex. *Cerebral Cortex, 20*(1), 77–88.

Sasanguie, D., Defever, E., Maertens, B., & Reynvoet, B. (2013). The approximate number system is not predictive for symbolic number processing in kindergarteners. *Quarterly Journal of Experimental Psychology (2006)*. http://dx.doi.org/10.1080/17470218.2013.803581.

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology, 114*(3), 418–431. http://dx.doi.org/10.1016/j.jecp.2012.10.012.

Seron, X., & Pesenti, M. (2001). The number sense theory needs more empirical evidence. *Mind & Language, 16*(1), 76–88.

Sophian, C. (2007). Measuring spatial factors in comparative judgments about large numerosities. In *Foundations of augmented cognition* (pp. 157–165). Springer. Retrieved from <http://link.springer.com/chapter/10.1007/978-3-540-73216-7_18>.

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences*, 201302751. http://dx.doi.org/10.1073/pnas.1302751110.

Szucs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex, 49*(10), 2674–2688. http://dx.doi.org/10.1016/j.cortex.2013.06.007.

Szucs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Frontiers in Psychology, 4*. http://dx.doi.org/10.3389/fpsyg.2013.00444.

Tibber, M. S., Manasseh, G. S. L., Clarke, R. C., Gagin, G., Swanbeck, S. N., Butterworth, B., et al. (2013). Sensitivity to numerosity is not a unique visuospatial psychophysical predictor of mathematical ability. *Vision Research, 89*, 1–9. http://dx.doi.org/10.1016/j.visres.2013.06.006.

Tokita, M., & Ishiguchi, A. (2010). How might the discrepancy in the effects of perceptual variables on numerosity judgment be reconciled? *Attention, Perception & Psychophysics, 72*(7), 1839–1853. http://dx.doi.org/10.3758/APP.72.7.1839.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences, 7*(11), 483–488.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*(2), 130–137.