Published in final edited form as:

Cognition. 2020 November ; 204: 104352. doi:10.1016/j.cognition.2020.104352.

Number sense biases children's area judgments

Rachel C. Tomlinson^{a,co}, Nicholas K. DeWind^{b,co}, Elizabeth M. Brannon^b

^[a]Department of Psychology, University of Michigan, East Hall, 530 Church Street, Ann Arbor, MI 48109.

^[b]Department of Psychology, University of Pennsylvania, 3720 Walnut St, Philadelphia, PA 19104.

Abstract

Humans are thought to use the approximate number system (ANS) to make quick approximations based on quantity even before learning to count. However, there has long been controversy regarding the salience of number versus other stimulus dimensions throughout development, including a recent proposal that number sense is derived from a sense of general magnitude. Here, we used a regression approach to disentangle numerical acuity from sensitivity to total surface area in both 5-year-old children and adults. We found that both children and adults displayed higher acuity when making numerosity judgments than total surface area judgements. Adults were largely able to ignore irrelevant stimulus features when making numerosity or total area judgements. Children were more biased by numerosity when making total area judgments than by total area when making numerosity judgments. These results provide evidence that number is more salient than total surface area even before the start of formal education and are inconsistent with the Sense of Magnitude proposal.

Keywords

approximate number system; Sense of Magnitude; number sense; numerical cognition

1. Introduction

Humans use quick approximations of number every day, from choosing the shortest line at the grocery store to estimating the number of people in a crowd. In many cases, these

Corresponding Author: Rachel C. Tomlinson, Department of Psychology, 2269 East Hall, 530 Church Street, University of Michigan, Ann Arbor, MI 48109. raclaire@umich.edu.

^{Co}Rachel C. Tomlinson and Nicholas K. DeWind contributed equally to the paper.

Credit Author Statement

Rachel C. Tomlinson: conceptualization, investigation, writing – original draft, writing – review & editing

Nicholas K. DeWind: conceptualization, methodology, formal analysis, visualization, writing – original draft, writing – review & editing

Elizabeth M. Brannon: conceptualization, resources, supervision, funding acquisition, writing – review & editing

Declarations of Interest: none.

Data linking: The complete anonymized dataset can be accessed at <https://osf.io/er9du/>.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

approximations do not involve symbolic numbers. Instead, we form an approximate sense of number that guides decision-making (Feigenson, Dehaene, & Spelke, 2004). This approximate number sense (ANS) is thought to be the internal representation of approximate number that allows us to estimate and compare quantities in the absence of language, and adult humans share the ability to approximate number with infants and non-human primates (Brannon, 2006; Brannon & Terrace, 1998; Xu & Spelke, 2000).

Numerical acuity can be defined simply as a person's ability to correctly detect numerical differences without counting. This is most easily measured by asking participants to compare the numerical value of two dot arrays and report which is larger, the ordinal comparison task. As the difference in number of dots becomes smaller and as the total number of dots becomes larger, choices become more difficult. By one theory, the ANS is thought to be an important building block as children acquire basic math skills (Dehaene, 1997; Gallistel & Gelman, 1992), and there is evidence that young children's numerical acuity is related to their symbolic math ability (for meta-analyses see Chen & Li, 2014; Fazio, Bailey, Thompson, & Siegler, 2014; Schneider et al., 2017). According to this theory, young children and even infants utilize their internal representation of approximate number as a scaffold for math skills developed later in childhood; as a result, greater numerical acuity early in life predicts easier acquisition of math upon the start of formal education (Feigenson, Libertus, & Halberda, 2013; Starr, Libertus, & Brannon, 2013). Similarly, the development of symbolic math skills is thought to in turn enhance the ANS (Suarez-Pellicioni & Booth, 2018).

However, recent research has demonstrated that non-numerical visual features, such as the size and spacing of the elements in a visual array, can affect numerical approximation and therefore estimates of numerical acuity (Cantlon, Safford, & Brannon, 2010; Gebuis, Kadosh, de Haan, & Henik, 2009; Leibovich & Henik, 2014; Soltesz, Szucs, & Szucs, 2010; Starr, DeWind, & Brannon, 2017). These findings have led to a growing number of theories proposing the existence of a more general magnitude sense rather than a unique number sense (Cantrell & Smith, 2013; Gebuis & Reynvoet, 2014; Lourenco, Ayzenberg, & Lyu, 2016; Mix, Huttenlocher, & Levine, 2002), including the recent proposal of the Sense of Magnitude Theory by Leibovich and colleagues (2017).

Although some of these theories suggest that numerosity is merely one of several magnitude perceptions, none of which has primacy, here we critically examined the implications of the Sense of Magnitude Theory, a stronger form of these general magnitude theories that holds that numerosity representations are secondary to representations of non-numerical magnitudes (Leibovich et al., 2017). We identified two tenets of the Sense of Magnitude theory that we operationalized and tested here. First is that numerical judgements are largely driven by prepotent responses to non-numerical visual magnitudes rather than a true, separate number sense. Implicit in this idea is that non-numerical representations are more closely connected to decision processes, and so these representations must be inhibited using executive control processes in order for numerical judgments to be expressed (e.g. Gilmore et al., 2013). This leads to a prediction: if non-numerical magnitude representations have greater primacy and are more difficult to suppress than numerical representations, we should

expect numerical judgments to be as influenced or more influenced by conflicting non-numerical magnitudes than non-numerical magnitudes are by numerosity.

The second tenet we operationalized is that numerical representations are derived from non-numerical visual magnitudes via statistical regularities in visual input (Leibovich et al., 2017). According to this idea, the transformation from an image of items projected onto the retina into a numerical representation in the brain is mediated by an intermediate non-numerical magnitude representation. This leads to an interesting prediction: if nonverbal numerical representations are derived from other magnitude representations, we should expect acuity for non-numerical magnitudes to be equal or greater than numerical acuity. The reason for this is because errors in the non-numerical representation would propagate to the numerical representation. We develop this logic further using a Monte Carlo simulation of internal magnitude representations.

A notorious problem for studying approximate number is that many non-numerical stimulus features are confounded with number. As the number of dots in an array changes, so do the values of many other features such as total surface area, total perimeter, and density. The approach taken up until now for assessing the influence of non-numerical visual features has been to use pairs of stimulus sets, one in which number is congruent with a non-numerical visual feature and one in which it is incongruent (Gebuis & Reynvoet, 2012; Gilmore, Cragg, Hogan, & Inglis, 2016; Nys & Content, 2012). Researchers have reasoned that if accuracy in the congruent set is higher than the incongruent set, the non-numerical feature must be affecting numerosity comparison.

The problem with this approach is that it does not provide an obvious quantitative way to compare the effect of a non-numerical feature on numerosity with the effect of numerosity on that non-numerical feature. These shortcomings in experimental design can be addressed by utilizing the modeling approach first described by DeWind and colleagues (2015) that disentangles the effects of numerosity on participant choice from the effects of other visual features. This approach provides independent estimates of numerical acuity and bias. The difference between acuity and bias is analogous to the difference between precision and accuracy in statistical theory. Acuity, like precision, is a measure of the random errors in judgments of numerosity. Participants with higher acuity can correctly discern the more numerous of two arrays that differ by a small numerical ratio. Bias, like statistical accuracy, is a measure of systematic deviation from reporting the correct answer caused by differences in the size or spacing of the elements in the arrays. Biased participants can be induced to consistently choose the incorrect array when the size or spacing of the elements is incongruent to number.

Here we develop a novel extension of the DeWind and colleagues (2015) model to allow an analogous analysis of the acuity and bias for the total area of the stimuli, as defined by the total number of illuminated pixels in an array. Total area acuity is a measure of the random errors in judgments of total area, and bias is a measure of the systematic deviation from reporting total area caused by differences in the number or spacing of the elements in the arrays. With this method, researchers gain the ability to measure differences in acuity and bias across multiple stimulus dimensions, and to compare them to each other.

Even given the ability to measure bias and acuity across different dimensions, there remains the issue of the prevalence and privilege of numbers within modern society. If an experiment were to reveal higher acuity and lower bias for number versus other stimulus dimensions in adults, it could be the result of enculturation through extensive formal mathematics education or the result of a more primitive privilege of numerosity. Previous work with the Mundurucú, an Amazonian indigene group whose culture includes little formal education and primarily nonverbal concepts of number, found that math education does increase the acuity of the number sense (Piazza, Pica, Izard, Spelke, & Dehaene, 2013). As mentioned previously, the ANS is considered to be an important building block as children learn symbolic mathematics (Dehaene, 1997; Gallistel & Gelman, 1992); therefore, the age at which children have yet to begin formal education, but are able to understand words describing different stimulus dimensions is of particular interest. If number sense is truly secondary to magnitude sense, this should be most evident before formal mathematics training begins.

In the present study, we use a novel adaptation of the DeWind modeling approach to test two hypotheses implied by a strong version of the Sense of Magnitude theory. First, if non-numerical magnitude representations are more prepotent than numerosity representations and require executive control to inhibit, we should expect that numerical judgments should be biased by total area cues as much as or more than total area judgments are biased by conflicting numerical cues. Second, if nonverbal numerical representations are derived from other magnitude representations, we should expect acuity for non-numerical magnitudes to be equal to or greater than numerical acuity, because errors in the representation of non-numerical magnitudes will be propagated to downstream numerosity representations. To test these predictions, we therefore asked whether five-year-old children and adults showed differences in bias or acuity when asked to make numerosity or total surface area judgments. Using the same stimulus space, we asked half of our participants to choose dot arrays based on numerosity and the other half to choose based on total surface area. Such an approach allowed us to fully separate ANS and total surface area acuity from other visual features and to quantify the extent to which the judgments of children and adults were biased by conflicting irrelevant magnitude cues.

2. Methods

2.1 Participants

A total of 32 five-year-old participants (ages 5.08–5.80 years with mean of 5.45, 17 female and 15 male) and 46 adult participants (ages 18.6–24.8 years with mean of 20.5, 18 female, 24 male and 4 not reporting) completed the experiment. One five-year-old participant was excluded due to experimenter error in the administration of the task. Adult participants and parents of child participants provided written informed consent to a protocol approved by the University of Pennsylvania Institutional Review Board (adults) or the Duke University Institutional Review Board (children) before participation. Children received small stickers and a prize for their participation, and their parents were compensated for travel. Adult participants were compensated with course credit.

2.2 Task

All child and adult participants completed a dinosaur-themed nonsymbolic comparison task. The nonsymbolic comparison task was a two-alternative-forced-choice adapted from DeWind et al. (2015) to be child-friendly (Figure 1). Participants touched a large rainbow-colored bull's-eye to initiate a trial. For each trial, two dot arrays appeared on a touchscreen monitor for 750 ms followed by the choice screen with a dinosaur cartoon character and two question-marks where the arrays had been. The dinosaur was left-right flipped on a random 50% of trials. Participants were asked to make a choice by touching one of the two question marks. Participants in the number condition were given the following instructions:

Dino wants to eat lots and lots of dots. Can you choose which side had **more** dots?
He only cares about the **number** of dots.

Participants in the total area condition were instructed as follows:

Dino wants to fill up lots of **space** in his tummy. Can you choose which dots will take up lots of **space** in his tummy? He only cares about how much the dots will fill up his tummy.

Children and adults were randomly assigned to one of the instruction conditions. Though a within-subject design would be powerful, one of our primary concerns was encouraging adherence to the assigned task instructions, and we were concerned that a within-subject design would lead to contamination of these instructions within the 5-year-old group. Children were reminded of the instructions throughout the task whereas adults were given the instructions once at the beginning of the experiment. Participants completed up to 200 trials per session; all adults completed all 200 trials, while children's performance varied (from 94–200 trials, median 180 trials).

2.3 Stimuli and difficulty titration

In both conditions the stimuli were drawn from the same stimulus space described previously (DeWind et al., 2015). Number and total area varied across 13 levels approximately evenly spaced on a log-scale covering a fourfold change from the smallest value to the largest. The numerical values were: 8, 9, 10, 11, 13, 14, 16, 18, 20, 23, 25, 29, and 32. The total area also varied across 13 levels and ranged from 2463 to 9852 pixels. In addition to number and total area, three other features also varied over a fourfold change across thirteen levels: item area, field area, and sparsity/ density. These features are defined mathematically in Table 1. The items were white on a black background. They were placed randomly (centered on a random pixel) such that they entirely fell within a circular field and did not overlap. Within an array, all items were the same size, but their size varied between arrays. Trial-by-trial feedback was provided, with a colorful dinosaur-themed screen indicating a correct answer and a blank screen indicating an incorrect answer.

A pair of unique arrays were generated by the program for each trial. One of these was the anchor, which was always composed of 16 items and a total area of 4,926 pixels. The other array, the deviant, differed from the anchor by a specific ratio as determined by the titrated difficulty level. The starting difficulty was a 1:2 ratio. For subjects in the number comparison condition these easiest trials consisted of 8:16 or 16:32 comparisons. For

subjects in the total area condition, the easiest trials consisted of 2,463:4,926 or 4,926:9,852 pixel comparisons. There were six levels of ratio difficulty for both the numerical and total area comparison conditions. Every five trials the program would assess performance on the previous five trials. If accuracy was 3/5 or less the ratio would get one level easier, if it was 4/5 the ratio would remain the same, and if it was 5/5 the ratio would get one level more difficult. The difficulty levels were bounded by the easiest (2:1) and hardest (8:7) difficulties, and the titration algorithm would stay at these ratios even if a participant did poorly on the easiest or very well on the most difficult levels. All the other visual features of the deviant were allowed to vary randomly (within the range of the maximum and minimum values given in Table 1). Thus, the number of items was varied randomly in the total area condition, and total area was varied randomly in the number condition. This allowed the effect of all visual features (including the instructed one) on participants' choices to be modeled as described below.

2.4 Analysis – just-noticeable-difference

The just-noticeable-difference was calculated as the average (geometric mean) ratio the titration procedure reached for each participant. For participants in the number condition, this ratio is a numerical ratio and for participants in the total area condition this ratio is a total area ratio. The titration procedure took about 50 trials to home in on a difficulty ratio appropriate for each participant, and so the just-noticeable difference was calculated as the average of ratios presented on trials 51–200. For children who did not complete all 200 trials the just-noticeable-difference was calculated as the average of all trials presented after trial 50. For this and all other analyses we excluded trials with a RT longer than 5 seconds.

2.5 Analysis – generalized linear model (GLM): disentangling bias and acuity

To disentangle the effects of acuity and bias on participants' choices we fit a GLM using a probit link function and a binomial error distribution with a constant term and terms for the logarithm of the numerical ratio, size ratio, and spacing ratio (DeWind et al., 2015).

$$p_{cr} = \Phi(\beta_{Side} + \beta_{Num}\log_2(r_{Num}) + \beta_{Size}\log_2(r_{Size}) + \beta_{Spacing}\log_2(r_{Spacing})) \quad \text{Eq. 1}$$

Where p_{cr} is the proportion of trials the participant selected the stimulus on the right, Φ is the cumulative normal distribution, β_{Side} is the intercept (indicating side bias), r_{Num} , r_{Size} , and $r_{Spacing}$ are the ratios of the number, size and spacing of the right array to the number, size, and spacing of the left array, and β_{Num} , β_{Size} , and $\beta_{Spacing}$ are the regression coefficients. Size and spacing are mathematically defined features designed to serve as regressors in the model.

$$Size = Item\ area \times Total\ area \quad \text{Eq. 2}$$

$$Spacing = Field\ area \times Sparsity \quad \text{Eq. 3}$$

As a result of these definitions, the coefficients in the GLM serve as a precise indication of the acuity and bias of participants in both tasks. For example, if participants engaged in pure

numerical discrimination, the β_{Num} fit to their choice data would be significantly positive and β_{Size} and $\beta_{Spacing}$ would be zero, because the size and spacing of the stimuli had no effect on estimates. Any consistent use of size or spacing would result in non-zero coefficient fits in the model, an indication of bias, and other things being equal lower accuracy on numerical discrimination. β_{Num} serves as an index of acuity, a larger value indicates a stronger sensitivity to numerical ratio, and other things being equal a higher accuracy on numerical discrimination. In this sense we can decompose choice data into acuity and bias, two factors that combine to determine accuracy. In other words, this modeling approach is unique in its ability to present an acuity estimate that is independent from all other features within the number, size, and spacing stimulus space.

Participants engaged in pure total area discrimination would have a slightly more complex set of coefficients. They would have a significant positive and equal β_{Num} and β_{Size} but zero $\beta_{Spacing}$. Intuitively, this can be understood because increasing the size of the items and increasing their number both increase their total area, so sensitivity to both these features is necessary to discriminate total area. Mathematically it can be demonstrated that the logarithm of the total area ratio of two stimuli is equally proportional to the logarithm of their numerical ratio and their size ratio.

$$\log_2(r_{TA}) = \frac{1}{2}\log_2(r_{Size}) + \frac{1}{2}\log_2(r_{Num}) \quad \text{Eq. 4}$$

Where r_{TA} is the ratio of the total areas of the two arrays. As a result of this equality, if participants are discriminating arrays based on their total area ratio, then fitting the GLM in Eq. 1 to their choices data will result in equal positive β_{Num} and β_{Size} and zero $\beta_{Spacing}$. Similar relationships exist between item area, density, and sparsity and number, size, and spacing.

$$\log_2(r_{IA}) = \frac{1}{2}\log_2(r_{Size}) - \frac{1}{2}\log_2(r_{Num}) \quad \text{Eq. 5}$$

$$\log_2(r_{FA}) = \frac{1}{2}\log_2(r_{Spacing}) + \frac{1}{2}\log_2(r_{Num}) \quad \text{Eq. 6}$$

$$\log_2(r_{Spar}) = \frac{1}{2}\log_2(r_{Spacing}) - \frac{1}{2}\log_2(r_{Num}) \quad \text{Eq. 7}$$

r_{IA} , r_{FA} , and r_{Spar} are the item area ratio, the field area ratio, and the sparsity ratio respectively. Participants purely discriminating based on item area will have a positive β_{Size} , an equal magnitude but negative β_{Num} , and zero $\beta_{Spacing}$. Participants purely discriminating based on field area will have equal and positive β_{Num} and $\beta_{Spacing}$ and zero β_{Size} . Participants purely discriminating based on sparsity will have a positive $\beta_{Spacing}$, an equal magnitude but negative β_{Num} , and zero β_{Size} . When the beta coefficients for a participant (or mean of many participants) are plotted against each other, these relationships have easily visualized geometric properties that allow the rapid assessment of which feature participants

were discriminating, their acuity in those discriminations, and their bias away from pure discrimination (Figure 2).

An unbiased participant only chooses stimuli based on the instructed dimension. Bias indicates a participant's reliance on a non-instructed feature to make responses. For example, if a participant in the number condition is more likely to respond to a stimulus that has a greater total area or vice-versa, that participant would be biased. Acuity refers to a participant's precision in making their selection. Acuity indexes the degree to which a participant can differentiate between arrays at close ratios after taking bias into account. We took advantage of the geometrical relationships indicated in Figure 2 to quantify bias and acuity according to the following equations. Bias was calculated as the angular deviation of the coefficient vector from the numerosity line in the β_{Num} by β_{Size} coefficient space if the participant was in the number condition, or from the total area coefficient line if the participant was in the total area condition. By convention bias in the number condition towards total area (arrays with larger total area perceived as more numerous) were given positive bias angles and bias away from total area (arrays with larger total area perceived as less numerous) were given negative bias angles. Similarly, participants in the total area condition who perceived more numerous arrays as having more total area were given positive bias angles and those who perceived more numerous arrays as having smaller total area were given negative bias angles.

$$Bias_{Num} = \tan^{-1}\left(\frac{\beta_{Size}}{\beta_{Num}}\right) \quad \text{Eq. 8}$$

$$Bias_{Total\ area} = -\tan^{-1}\left(\frac{\beta_{Size}}{\beta_{Num}}\right) - 45^\circ \quad \text{Eq. 9}$$

$$Acuity = \beta_{Num} + \beta_{Size} + \beta_{Spacing} \quad \text{Eq. 10}$$

2.6 Analysis – Monte Carlo simulations: numerosity derived from total area representations

We ran Monte Carlo simulations to more clearly demonstrate how a Sense of Magnitude theory of numerosity perception would affect the derived behavioral metric of acuity. In these simulations we used the real stimuli from the experiment and modeled the formation of internal representations of numerosity and total area from the visual input. We then took this simulated behavioral output and fit the behavioral regression model described above.

We explored a cognitive model in which numerosity is derived from non-numerical visual magnitudes via learned associations. In particular, we explored what it would mean for acuity measurements if numerosity representations were derived from total area and item area representations. This is theoretically possible because total area divided by average item area is equal to numerosity. However, if numerosity were derived in this way noise in the

total area representation would propagate to the numerosity representation. We suspected that this would be apparent in the acuity measure.

Our cognitive model was based on the logarithmic model of numerosity representation developed by Piazza and colleagues (2004). According to this model numerosity is represented internally as a normally distributed random variable on a logarithmically compressed mental number line. All numerosities are represented by random variables with the same standard deviation, w . This theory is notable for predicting ratio dependent performance empirically observed in numerosity discrimination tasks, otherwise known as Weber's law. The only further development of the cognitive model we made here, was to assume that total area and item area are similarly represented by normally distributed random variables on a logarithmically compressed total area and item area mental ordinate, respectively. This is consistent with the finding that these magnitudes also follow Weber's law of ratio dependent performance in discrimination tasks (Matthews, 1969; Ono, 1967).

For each trial we transformed the true total area of each array into a simulated noisy internal representation of total area by taking the logarithm and adding a random number drawn from a normal distribution centered on zero with a standard deviation equal to $w_{total\ area}$. We repeated the experiment with a few different values for $w_{total\ area}$: 0.05, 0.1, and 0.3. We then calculated an internal representation of item area using the same method with possible values of $w_{item\ area}$ equal to 0, 0.05, 0.1, and 0.3. In the case of item area, we included the very optimistic scenario that item area was represented with perfect precision ($w_{item\ area} = 0$). Internal numerosity representations were generated by subtracting the item area representation from the total area representation. Subtraction was used instead of division, because both the total area and the item area representations had already been log-transformed. Again optimistically, we assumed that the subtraction operation added no noise to the numerosity representation. Finally, we simulated the decision-making phase by simply assuming that participants choose the array with the larger internal numerosity representation with no further influence from total area or item area. After repeating this procedure for each trial, we fit the behavioral regression model and calculated acuity using Eq. 10.

3. Results

Both five-year-old children and adults performed better than would be expected by chance in both the number instruction condition (children: 79.9%; adults: 85.2%) and the total area instruction condition (children: 77.8%; adults: 81.1%). Both five-year-old children and adults were titrated to a stable difficulty level within the first 50 trials.

3.1 Just-noticeable-difference

The difficulty was titrated to maintain accuracy around 80%, and so the average ratio (numeric or total area ratio depending on condition) that each participant reached was more informative of performance than was accuracy. Here we term this ratio the just-noticeable-difference, and because it is a ratio, it is directly comparable between the number and total area tasks. The just-noticeable difference was better for number discrimination than total area discrimination (main effect of task $F(1,73) = 44$, $p \ll 0.001$; Figure 3), and adults could

discriminate more difficult ratios than children (main effect of age group $F(1,73) = 147$, $p \ll 0.001$; Figure 3). Follow-up pair-wise comparisons of the four groups (t-tests) demonstrated that adults performed better than children within each task condition and that participants discriminating number performed better than participants discriminating total area within each age group (all p 's < 0.001 ; Figure 3). While it is clear from these analyses that both adults and children are better at discriminating arrays based on numerosity, these analyses do not differentiate between bias and acuity. In other words, it is unclear from the just-noticeable-difference whether participants could differentiate closer ratios of numerosity because they had a more precise representation of numerosity than total area (acuity) or because they were less distracted by non-numerosity visual features in the numerosity condition than they were distracted by non-total-area visual features in the total area condition (bias).

3.2 Generalized linear model (GLM): disentangling bias and acuity

To address our hypotheses and differentiate the role of bias from acuity we adapted and advanced the modeling approach developed previously and described above (DeWind et al., 2015). In short, this regression model measures the relative influence of number of elements, the size of the elements, and the spacing of the elements on participant choice. We fit the model to each participant's choice data, and plotted the coefficients returned against each other to create a "coefficient space" (Figure 4).

Figure 4 allows for a qualitative assessment of participants' performance by examining the position of the mean coefficient vectors. Adults in the number condition had the highest acuity and lowest bias, indicated by the relatively large distance of the coefficient vector from the origin and the proximity to the "number" feature line in both plots. Adults in the total area condition had similarly low bias, indicated by the proximity of the coefficient vector to the "total area" feature line in Figure 4A. However, adult total area acuity was lower, indicated by the smaller distance to the origin compared to the adult number participants. Overall, children had lower acuity and greater bias than adults. However, children in the number condition were less biased than children in the total area condition, as indicated by the relative proximity of the child number coefficient vector to the number feature line compared to the child total area coefficient vector to the total area feature line.

To quantify these observations and statistically test them we calculated bias according Eq. 8 and Eq. 9 and acuity according to Eq. 10 for each participant (Figure 5). For both children and adults, bias and acuity differed depending on task. Adults in the number condition demonstrated higher acuity than adults in the total area condition ($t(44) = 2.2$, $p = 0.03$). Adults showed no detectable bias from total area in the numerical condition ($t(25) = 1.0$, $p = 0.32$), and only a marginally significant bias in the total area condition such that smaller numerosities were judged as having a larger total area ($t(19) = -2.0$, $p = 0.06$). Like adults, children in the number condition demonstrated higher acuity than children in the total area condition ($t(29) = 2.3$, $p = 0.03$). In contrast to adults, five-year-old children showed significant bias for both numerosity and total area judgments. Children in the total area condition demonstrated substantial bias such that more numerous arrays appeared to have greater total area ($t(14) = 8.5$, $p < 0.001$), while children in the number condition

demonstrated a somewhat smaller bias such that arrays with greater total area appeared more numerous ($t(15) = 5.0, p < 0.001$). Children's bias in the total area task was significantly greater than children's bias in the number task ($t(29) = 4.0, p < 0.001$).

In summary, children were more biased by number than by total area, while adults showed little bias overall. Both children and adults showed higher acuity for number than for total area.

3.3 Monte Carlo simulations: numerosity derived from total area representations

We ran Monte Carlo simulations to confirm our reasoning that if numerosity representations are derived from total area representations, this would result in acuity in the total area task being equal to or greater than acuity in the numerosity task. Figure 6A shows the cognitive model represented by the simulation. We ran 12 simulations examining a range of possible $w_{total\ area}$ and $w_{item\ area}$, the terms that denote the noisiness of the total area and item area representations. In all 12 simulations the total area acuity was greater than or equal to the numerosity acuity (Figure 6B), the opposite of the pattern observed in the real data.

3.4 Post-hoc analysis – Congruency conditions

To ensure that differences in congruency conditions were not driving the difference between the number and total area tasks, we conducted a post-hoc check of the number of congruent and incongruent trials for each task. We defined a “congruent” trial as one in which the correct choice given a participant's task would have also been the correct choice in the other condition (e.g. in the number task, the array with greater numerosity also had larger total surface area). We defined an “incongruent” trial as the inverse (e.g. in the number task, the array with greater numerosity had smaller total surface area). We defined “neutral” trials as trials in which the dimension of non-interest for the trial was the same for both arrays (e.g. in the number condition, the array with greater numerosity had the same total surface area as the other array presented). The proportion of “congruent”, “incongruent”, and “neutral” trials did not differ by condition. In the number condition, number was “congruent” with area 67.5% of the time, “incongruent” 26.0% of the time, and neutral (area same in both arrays) 6.6% of the time. The area condition was very similar: area was “congruent” with number 67.7% of the time, “incongruent” 24.8% of the time, and neutral 7.5% of the time.

4. Discussion

The goal of our study was to compare sensitivity to the numerosity versus the total area of an array and examine how this relative sensitivity may or may not change over development. We sought to test a strong form of the Sense of Magnitude Theory, and more generally the proposal that numerosity is a derived visual feature calculated from the integration of area and other non-numerical visual features (Leibovich et al., 2017). Specifically, we tested two predictions derived from the Sense of Magnitude Theory: first, that numerical judgments should be biased by area as much as or more than the reverse, and second, that acuity for total area should be equal or greater than numerical acuity. We reasoned that such evidence may be more apparent before the start of formal mathematics education, in which number is emphasized more than area. To test these hypotheses, we asked five-year-old children and

adults to make numerosity or total surface area judgments given similar stimuli drawn from the same counter-balanced stimulus space. We used a novel modeling approach to disentangle acuity along these two dimensions, number and total surface area, from bias to all other continuous stimulus dimensions within the size, spacing, and number stimulus space. In contrast with the hypotheses implied by a strong interpretation of the Sense of Magnitude proposal, we found that bias, which was primarily observed in children, was to number, not to other continuous stimulus dimensions, and that numerical acuity was higher than total area acuity for both children and adults.

First, we tested the assumption that numerical judgments should be biased by total area as much as or more than the reverse. One major tenet of the Sense of Magnitude Theory is that inhibitory control is necessary to suppress prepotent responses to continuous magnitudes for numerical representations to gain behavioral expression (Leibovich et al., 2017). Indeed, individual differences in inhibitory control may relate to individual bias estimates, as inhibition is known to be important for developmental dyscalculia (Szucs, Devine, Soltesz, Nobes, & Gabriel, 2013) and to relate to performance on incongruent trials (Gilmore et al., 2013). However, if numerical magnitude representations are derivative of and secondary to non-numerical magnitude representations, we would predict that this bias effect would be symmetric at a minimum. In other words, we would expect that numerical judgments would be biased by total area at least as much as total area judgements are biased by number. Given the developmental trajectory of inhibitory control (Davidson, Amso, Anderson, & Diamond, 2006), this prediction would be even stronger in children, in which the prepotent, or innate, response should emerge as bias. However, we found that adults were largely able to ignore irrelevant stimulus features, demonstrating a highly differentiated sense of number and of area. Children were more susceptible to bias, in line with previous work examining the influence of bias on numerical discriminations throughout development (Starr et al., 2017). However, in contrast to what might be expected under a strong interpretation of the Sense of Magnitude theory, children in the number condition were only mildly biased by area, while children in the area condition were strongly biased by number. This finding fits with previous work showing that young children struggle to judge surface area when it is incongruent with number (Rousselle & Noel, 2008). Therefore, while we do see mutual bias among stimulus features in children, this novel adaptation of the DeWind (2015) modeling approach allows us to determine that total area judgements are more biased by number than vice versa. Thus number, rather than total surface area, may be the prepotent response that children struggle to inhibit.

Second, we tested the assumption that that acuity for total area should be equal or greater than numerical acuity. We reasoned that if numerosity representations were derived from continuous magnitude representations, then numerosity representations would also inherit the noisiness of those dimensions. To confirm this reasoning, we developed a novel Monte Carlo simulation. This simulation had simple assumptions: total area and item area are both represented as normally distributed random variables on a log-compressed mental ordinate, and numerosity representations are generated by subtracting the item area representation from the total area representation. Even though we modeled this subtraction operation as completely noiseless, the numerosity representations were, nevertheless, noisier than total area and item area representations. This is because when adding or subtracting random

variables, their variances always sum. After simulating trial level representations of total area and numerosity, we were able to model the decision process as a simple comparison of the internal representations. We then fit the behavioral regression model to the simulated choice data and calculated acuity for both the number task and the total area task. As expected, we found that the acuity on the numerosity task was always lower than or equal to the acuity on the total area task for the simulated data. However, this was the opposite of the pattern observed in both the adult and child empirical data sets we collected. Instead, we found that both adults and children demonstrated higher numerical acuity than total area acuity. This provides strong evidence against the model posited in Figure 6A, that numerosity representations are derived from total area representations.

Our finding that numerosity acuity was greater than total area acuity in adults and children adds to previous findings that infants can discriminate changes in non-numerical magnitude dimensions, such as total surface area, but with less sensitivity than they can discriminate changes in number (Brannon, Lutz, & Cordes, 2006; Cordes & Brannon, 2009; Lourenco & Longo, 2010; Starr & Brannon, 2015) and that children appear to spontaneously attend to number rather than total surface area (Cantlon et al., 2010). Indeed, we see that even within the same counter-balanced stimulus set, and even when equating the ratio difference of number and area, number sticks out to both children and adults as a more salient feature than surface area.

Taken together, our results do not support the idea that number sense is secondary to magnitude sense. Instead, they provide evidence that number may be privileged even before the start of formal education. While our methodology differs substantially from previous approaches to the number sense debate, our findings are not without precedent. For example, our finding that bias and acuity differ substantially between number and surface area conditions is in line with recent work demonstrating differing developmental trajectories for performance on number versus other magnitude tasks (Odic, 2018), as well as evidence that number and surface area within a dot array are encoded differently (Odic & Halberda, 2015). Within the adult literature, our results are in line with previous evidence of mutual interference between numerosity and continuous magnitudes (Hurewitz, Gelman, & Schnitzer, 2006), as well as recent work demonstrating that numerosity signal can be differentially influenced by area in a context-dependent manner (Fornaciai, Farrell, & Park, 2019). Our findings also reflect those of Cicchini and colleagues (2016) in which subjects were found to be much more sensitive to numerosity than density or field area at all but the highest densities (Cicchini et al., 2016). Though the investigators did not probe sensitivity to total surface area, their findings are similar to ours in that adults based discriminations on numerosity even when asked to discriminate density, and that field area discriminations were also influenced by numerosity.

In contrast, our findings are not consistent with what would be expected given a strong interpretation of the Sense of Magnitude Theory put forth by Leibovich and colleagues (2017), which holds that the representation of continuous magnitudes is more prepotent and that number representation is derived from magnitude representations. Given this theory, we might expect similar acuity and mutual interference for number and surface area, especially for younger children. We might also expect preferential acuity for number and bias to

number to increase with age. Instead, we find that both five-year-olds and adults are better at discriminating number than total surface area. Furthermore, in five-year-olds number appears to be more prepotent than total area, as number has a greater influence on area discrimination than vice-versa. This inverse relationship between math exposure and bias would indicate that enculturation to number is not the primary driver of our results. Rather than supporting the Sense of Magnitude theory, these findings suggest that number is privileged, and that this privilege is not merely the result of extensive formal math education.

Future work regarding the privilege of number versus non-numerical visual features should explore the relative salience of magnitude dimensions other than total surface area. We chose total surface area in part because judgment along this dimension is more intuitive than for other visual features; however, additional evidence involving other features would lend support to our conclusions regarding number as privileged. In particular, a follow-up study investigating participants' ability to discriminate convex hull might be of interest, as recent findings have indicated that this dimension affects numerosity discrimination (Katzin, Salti, & Henik, 2019; Salti, Katzin, Katzin, Leibovich, & Henik, 2017). However, we note that our results do not indicate bias to convex hull in any age group or condition. This is best visualized in Figure 4b, where the Field Area line can be considered a close approximation of convex hull. There is no clear bias to Field Area, and no condition is making judgments based on that dimension alone, which would be indicated by a group data point falling along that line. Additional work could also explore the ability of even younger children to discriminate number and non-numerical visual features, as most 5-year-old children in the United States have already been exposed to numbers for several years. However, creative adjustments would be necessary to make the protocol more accessible and ensure successful task completion. Similarly, our findings may be affected by our decision to include trial-by-trial feedback. Though participants generally reached peak performance quickly and did not consistently improve from that level, it is possible that there was some learning throughout the task.

As with any experiment investigating non-symbolic numerosity discrimination, idiosyncrasies in the stimulus set may have influenced our results. Of particular interest is how the degree of congruence or incongruence between number and total area may influence the degree to which participants are influenced by the non-instructed visual feature. Previous work has found that different stimulus sets result in different degrees of convex-hull bias in a numerosity discrimination task, depending on whether the stimulus set includes incongruent trials designed to penalize reliance on convex-hull (DeWind & Brannon, 2016). Whether total area discrimination is similarly affected by the degree of congruence or incongruence with numerosity is an important area for future research.

5. Conclusions

While much more data is needed to fully understand the role non-numerical variables play in the development of number discrimination, our findings suggest that number representations are unlikely to be secondary to other magnitude representations. While our findings are not inconsistent with the idea that representation of number is related to representation of other magnitudes, they are inconsistent with the implications of the Sense of Magnitude Theory

that number is less prepotent than and derived from other magnitude representations. Instead, we find that number is more salient than total surface area even before the start of formal education.

Acknowledgements

This research was supported by the National Institutes of Health via RO1HD059108 awarded to EMB. RCT was supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1256260. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation or the National Institutes of Health.

References

- Brannon EM (2006). The representation of numerical magnitude. *Curr Opin Neurobiol*, 16(2), 222–229. doi:10.1016/j.conb.2006.03.002 [PubMed: 16546373]
- Brannon EM, Lutz D, & Cordes S (2006). The development of area discrimination and its implications for number representation in infancy. *Dev Sci*, 9(6), F59–64. doi:10.1111/j.1467-7687.2006.00530.x [PubMed: 17059447]
- Brannon EM, & Terrace HS (1998). Ordering of the Numerosities 1 to 9 by Monkeys. *Science*, 282(5389), 746–749. doi:10.1126/science.282.5389.746 [PubMed: 9784133]
- Cantlon JF, Safford KE, & Brannon EM (2010). Spontaneous analog number representations in 3-year-old children. *Developmental Science*, 13(2), 289–297. doi:10.1111/j.1467-7687.2009.00887.x [PubMed: 20136925]
- Cantrell L, & Smith LB (2013). Open questions and a proposal: A critical review of the evidence on infant numerical abilities. *Cognition*, 128(3), 331–352. doi:10.1016/j.cognition.2013.04.008 [PubMed: 23748213]
- Chen Q, & Li J (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, 148, 163–172. doi:10.1016/j.actpsy.2014.01.016 [PubMed: 24583622]
- Cicchini GM, Anobile G, & Burr DC (2016). Spontaneous perception of numerosity in humans. *Nat Commun*, 7, 12536. doi:10.1038/ncomms12536 [PubMed: 27555562]
- Cordes S, & Brannon EM (2009). The relative salience of discrete and continuous quantity in young infants. *Dev Sci*, 12(3), 453–463. doi:10.1111/j.1467-7687.2008.00781.x [PubMed: 19371370]
- Davidson MC, Amso D, Anderson LC, & Diamond A (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, 44(11), 2037–2078. doi:10.1016/j.neuropsychologia.2006.02.006 [PubMed: 16580701]
- Dehaene S (1997). *The Number Sense: How the Mind Creates Mathematics*. New York: Oxford University Press.
- DeWind NK, Adams GK, Platt ML, & Brannon EM (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition*, 142, 247–265. doi:10.1016/j.cognition.2015.05.016 [PubMed: 26056747]
- DeWind NK, & Brannon EM (2016). Significant Inter-Test Reliability across Approximate Number System Assessments. *Front Psychol*, 7, 310. doi:10.3389/fpsyg.2016.00310 [PubMed: 27014126]
- Fazio LK, Bailey DH, Thompson CA, & Siegler RS (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, 123, 53–72. doi:10.1016/j.jecp.2014.01.013 [PubMed: 24699178]
- Feigenson L, Dehaene S, & Spelke E (2004). Core systems of number. *Trends Cogn Sci*, 8(7), 307–314. doi:10.1016/j.tics.2004.05.002 [PubMed: 15242690]
- Feigenson L, Libertus ME, & Halberda J (2013). Links Between the Intuitive Sense of Number and Formal Mathematics Ability. *Child Development Perspectives*, 7(2), 74–79. doi:10.1111/cdep.12019 [PubMed: 24443651]

- Fornaciai M, Farrell A, & Park J (2019). Looking for more food or more people? Task context influences basic numerosity perception. *Cortex*, 114, 67–75. doi:10.1016/j.cortex.2018.05.021 [PubMed: 30032863]
- Gallistel CR, & Gelman R (1992). Preverbal and verbal counting and computation. *Cognition*, 44(1–2), 43–74. doi:10.1016/0010-0277(92)90050-r [PubMed: 1511586]
- Gebuis T, Kadosh RC, de Haan E, & Henik A (2009). Automatic quantity processing in 5-year olds and adults. *Cognitive Processing*, 10(2), 133–142. doi:10.1007/s10339-008-0219-x [PubMed: 18607652]
- Gebuis T, & Reynvoet B (2012). The Role of Visual Information in Numerosity Estimation. *Plos One*, 7(5). doi:10.1371/journal.pone.0037426
- Gebuis T, & Reynvoet B (2014). The Neural Mechanism Underlying Ordinal Numerosity Processing. *Journal of Cognitive Neuroscience*, 26(5), 1013–1020. doi:10.1162/jocn_a_00541 [PubMed: 24345168]
- Gilmore C, Attridge N, Clayton S, Cragg L, Johnson S, Marlow N, ... Inglis M (2013). Individual Differences in Inhibitory Control, Not Non-Verbal Number Acuity, Correlate with Mathematics Achievement. *Plos One*, 8(6). doi:10.1371/journal.pone.0067374
- Gilmore C, Cragg L, Hogan G, & Inglis M (2016). Congruency effects in dot comparison tasks: convex hull is more important than dot area. *Journal of Cognitive Psychology*, 28(8), 923–931. doi:10.1080/20445911.2016.1221828 [PubMed: 28163886]
- Hurewitz F, Gelman R, & Schnitzer B (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences of the United States of America*, 103(51), 19599–19604. doi:10.1073/pnas.0609485103 [PubMed: 17159143]
- Katzin N, Salti M, & Henik A (2019). Holistic processing of numerical arrays. *J Exp Psychol Learn Mem Cogn*, 45(6), 1014–1022. doi:10.1037/xlm0000640 [PubMed: 30070564]
- Leibovich T, & Henik A (2014). Comparing performance in discrete and continuous comparison tasks. *Q J Exp Psychol (Hove)*, 67(5), 899–917. doi:10.1080/17470218.2013.837940 [PubMed: 24070376]
- Leibovich T, Katzin N, Harel M, & Henik A (2017). From “sense of number” to “sense of magnitude”: The role of continuous magnitudes in numerical cognition. *Behav Brain Sci*, 40, e164. doi:10.1017/S0140525X16000960 [PubMed: 27530053]
- Lourenco SF, Ayzenberg V, & Lyu J (2016). A general magnitude system in human adults: Evidence from a subliminal priming paradigm. *Cortex*, 81, 93–103. doi:10.1016/j.cortex.2016.04.013 [PubMed: 27179917]
- Lourenco SF, & Longo MR (2010). General Magnitude Representation in Human Infants. *Psychological Science*, 21(6), 873–881. doi:10.1177/0956797610370158 [PubMed: 20431048]
- Matthews ML (1969). Visual size difference discrimination: Effect of disc size and retinal locus. *Perception & Psychophysics*, 6(3), 160–162. doi:10.3758/bf03210084
- Mix KS, Huttenlocher J, & Levine SC (2002). Multiple cues for quantification in infancy: Is number one of them? *Psychological Bulletin*, 128(2), 278–294. doi:10.1037/0033-2909.128.2.278 [PubMed: 11931520]
- Nys J, & Content A (2012). Judgement of discrete and continuous quantity in adults: Number counts! *Quarterly Journal of Experimental Psychology*, 65(4), 675–690. doi:10.1080/17470218.2011.619661
- Odic D (2018). Children’s intuitive sense of number develops independently of their perception of area, density, length, and time. *Dev Sci*, 21(2). doi:10.1111/desc.12533
- Odic D, & Halberda J (2015). Eye movements reveal distinct encoding patterns for number and cumulative surface area in random dot arrays. *Journal of Vision*, 15(15). doi:10.1167/15.15.5
- Ono H (1967). Difference threshold for stimulus length under simultaneous and nonsimultaneous viewing conditions. *Perception & Psychophysics*, 2(5), 201–207. doi:10.3758/bf03213050
- 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. doi:10.1016/j.neuron.2004.10.014 [PubMed: 15504333]

- Piazza M, Pica P, Izard V, Spelke ES, & Dehaene S (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychol Sci*, 24(6), 1037–1043. doi:10.1177/0956797612464057 [PubMed: 23625879]
- Rousselle L, & Noel M-P (2008). The development of automatic numerosity processing in preschoolers: Evidence for numerosity-perceptual interference. *Developmental Psychology*, 44(2), 544–560. doi:10.1037/0012-1649.44.2.544 [PubMed: 18331143]
- Salti M, Katzin N, Katzin D, Leibovich T, & Henik A (2017). One tamed at a time: A new approach for controlling continuous magnitudes in numerical comparison tasks. *Behav Res Methods*, 49(3), 1120–1127. doi:10.3758/s13428-016-0772-7 [PubMed: 27439534]
- Schneider M, Beeres K, Coban L, Merz S, Schmidt SS, Stricker J, & De Smedt B (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science*, 20(3). doi:10.1111/desc.12372
- Soltész F, Szucs D, & Szucs L (2010). Relationships between magnitude representation, counting and memory in 4-to 7-year-old children: A developmental study. *Behavioral and Brain Functions*, 6. doi:10.1186/1744-9081-6-13
- Starr A, & Brannon EM (2015). Evidence against continuous variables driving numerical discrimination in infancy. *Front Psychol*, 6, 923. doi:10.3389/fpsyg.2015.00923 [PubMed: 26191028]
- Starr A, DeWind NK, & Brannon EM (2017). The contributions of numerical acuity and non-numerical stimulus features to the development of the number sense and symbolic math achievement. *Cognition*, 168, 222–233. doi:10.1016/j.cognition.2017.07.004 [PubMed: 28715669]
- Starr A, Libertus ME, & Brannon EM (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences of the United States of America*, 110(45), 18116–18120. doi:10.1073/pnas.1302751110 [PubMed: 24145427]
- Suarez-Pellicioni M, & Booth JR (2018). Fluency in symbolic arithmetic refines the approximate number system in parietal cortex. *Hum Brain Mapp*, 39(10), 3956–3971. doi:10.1002/hbm.24223 [PubMed: 30024084]
- Szucs D, Devine A, Soltész F, Nobes A, & Gabriel F (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex*, 49(10), 2674–2688. doi:10.1016/j.cortex.2013.06.007 [PubMed: 23890692]
- Xu F, & Spelke ES (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74(1), B1–B11. doi:10.1016/s0010-0277(99)00066-9 [PubMed: 10594312]

Figure 1. Nonsymbolic comparison task.

The nonsymbolic comparison task was adapted from DeWind et al. (2015). For each trial, two dot arrays appeared on a touchscreen monitor. Half the participants were instructed to choose the dot array with the greater number of dots, and half were instructed to choose the dot array with the greater total surface area. In the figure above, the dot array on the left has greater total surface area, while the dot array on the right has a greater number of dots. Feedback was provided after each response.

Figure 2. Hypothetical examples illustrate coefficient vectors.

A participant who discriminated arrays based on their number, total area, or item area would fall along the lines respectively labeled in (a), and a participant who discriminated arrays based on number, field area, or sparsity would fall along the lines respectively labeled in (b). Participants choosing the array with the greater quantity of that feature would fall on the arrow-head side of the origin, and those choosing the lesser quantity would fall on the arrow-tail side of the origin. A participant with a coefficient vector at the origin would be uninfluenced by the number, size, or spacing of the stimuli (e.g., random responses or always responding on the same side). The city-block distance from the origin is the acuity and the angular deviation from the instructed feature lines is the bias. Two hypothetical examples are shown, a participant instructed to discriminate number (blue) and a participant instructed to discriminate total area (red). According to Eq. 8–10 their acuities are 2.0 and 1.5 respectively, and their biases are 7° and -14° respectively. Neither hypothetical subject was influenced by element spacing.

Figure 3. Adults perform better than children and performance is better on number comparison than on total area comparison.

Adults reached a higher mean difficulty (ratio) than children, and both adults and children reached higher mean difficulty in the number than in the total area task. *** $p < 0.001$. Error bars indicate SEMs across subjects.

Figure 4. GLM coefficient estimates demonstrate more acute number comparison than total area comparison in adults and children and strong bias towards number in the total area comparison in children.

β_{Num} plotted against β_{Size} (a) and β_{Num} plotted against $\beta_{Spacing}$ (b). Filled markers are group means and open markers are individual subject data. Triangles are adult data and circles are child data. Blue markers are participants performing the number discrimination task and red markers are participants performing the total area discrimination task. Labeled black lines indicate where coefficients would fall if a participant was discriminating the labeled visual feature without bias.

Figure 5. Across all participants, acuity was higher for number comparison than for total area comparison. Children comparing total area were more biased than children comparing number.

The left panel shows acuity in each group calculated according to Eq. 10. The right panel shows bias for each group. Bias in the number task was calculated according to Eq. 8 and bias in the total area task was calculated according to Eq. 9. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Error bars indicate standard errors across subjects.

Figure 6. Numerical acuity is greater than total area acuity, inconsistent with a cognitive model in which numerosity representations are derived from total area representations.

(a) The cognitive model underlying the Monte Carlo simulation. For each trial the array images are transformed into simulated internal representations of total area and item area by adding noise. The numerosity representation is calculated from these two non-numerical representations resulting in additive noise in the numerosity representation. (b) Numerical acuity and total area acuity (calculated by Eq. 10) exhibited by different Monte Carlo simulations derived from the cognitive model. Multiple simulations were run assuming different levels of item area noise ($w_{item\ area}$ of 0, 0.2, 0.3, and 0.4 indicated by markers ranging from purple to pink respectively) and different levels of total area noise ($w_{total\ area}$ of 0.2, 0.3, and 0.4 indicated by the squares, 5-point stars and 6-point stars respectively). The black line indicates equality of numerical and total area acuity. Unlike the empirical data (black triangle and black circle), the simulation assumption that numerosity is derived from total area and item area means total area acuity is always greater than or equal to numerical acuity.

Table 1.

Definition of counterbalanced visua features.

Feature	Formula	Minimum value	Maximum value
Number	n	8	32
Total area	$n\pi r^2$	2,463 pixels	9,852 pixels
Item area	πr^2	156 pixels	616 pixels
Field area	πr_f^2	31,416 pixels	125,664 pixels
Sparsity	$\frac{\pi r_f^2}{n}$	1,963 pixels	7,854 pixels

n refers to the number of items in the stimulus, r refers to the radius on one item, and r_f refers to the radius of the field within which the items are drawn. Density can be defined as the reciprocal of sparsity. Number and total area were the dimensions of interest for the present tasks.