Flawed stimulus design in additive-area heuristic studies

Joonkoo Park

Department of Psychological & Brain Sciences, University of Massachusetts

Commonwealth Honors College, University of Massachusetts

* Correspondence should be addressed to:

Joonkoo Park, Ph.D.

Department of Psychological and Brain Sciences
University of Massachusetts
135 Hicks Way/Tobin Hall
Amherst, MA 01003. U. S. A.
Phone: (413) 545-0051
Email: joonkoo@umass.edu

Abstract

In a series of recently published studies purportedly on the "additive-area heuristic," Yousif & Keil (2019; 2020) argue for a systematic distortion in the perception of the cumulative area of an item array and further claim that previous findings of numerical cognition and magnitude perception in general are "at risk" (Yousif & Keil, 2021). This commentary describes serious stimulus design flaws present in all of Yousif and colleagues experiments that prevent from making such conclusions. Specifically, item arrays used in those studies demonstrate a skewed correlational structure between selected magnitude dimensions and exhibit unbalanced ranges across different magnitude dimensions of interest. Because the perception of magnitude dimensions interferes one another and because our perceptual system is sensitive to the statistical regularities of the sensory input, such a biased design makes it difficult, if not impossible, to interpret the choice behavior of an observer making magnitude judgments. By re-introducing the mathematical framework for a systematic construction of dot array stimuli (DeWind et al., 2015) and by re-analyzing the data from another recent study on area perception (Tomlinson et al., 2020), this paper explains—and introduces a MATLAB code for—an optimal method for designing and constructing dot arrays for magnitude perception studies.

Keywords

number perception; area perception; stimulus design; numerical cognition; dot arrays

1. Introduction

In a series of recently published articles, Yousif and colleagues (Yousif, Aslin, & Keil, 2020; Yousif & Keil, 2019, 2020) investigated adult participants' judgment of the cumulative area of an item array. In these studies, participants were given two item arrays and were asked to indicate which array has more cumulative area. From their results, the authors collectively make two conclusions. They first claim that the estimation of true visual area of an item array, defined by the sum of the mathematical area of all items in the array, is significantly distorted by perceived visual area which they define as the sum of the width (along horizontal axis) plus the height (along vertical axis) of all the items in that array, which they refer to as the additive-area heuristic. The authors then make a broader and stronger claim about magnitude perception in general, particularly about numerosity perception. They argue that "the ability to make area discriminations on the basis of 'additive area' … predicts number discrimination ability" (Yousif & Keil, 2020) and further claim that "the explanatory power of this heuristic … may bear on the interpretation of many seminal articles in the field of numerical cognition as well as work on area estimation, general magnitude, and various aspects of visual perception" (Yousif & Keil, 2019). Unfortunately, however, as demonstrated in the current paper, those additive-area heuristic studies contain flaws in their stimulus design that make it difficult to reach such conclusions.

2. A very brief review of the area perception literature

Area perception has been studied for more than a century, documented as early as Bolton (1898). It has long been known that area perception is significantly influenced by the shape and size of the stimulus. For instance, Anastasi (1936) reported that a diamond is relatively overestimated compared to a square of the same size. Other researchers (e.g., Teghtsoonian, 1965) found that perceived area increases at a lower rate than true area. These older studies seem to have focused on explaining these phenomena by how perception relies on multiple pieces of information to derive an estimate of area. Some suggested that area perception is based on one of the linear dimensions (Anastasi, 1936; Verge & Bogartz, 1978), while others suggested that it is based on some integration of information such as the function of perimeter (Smith, 1969). Interestingly, Anderson & Cuneo (1978) and Wolf & Algom (1987) who used Anderson & Cuneo's paradigm demonstrated that younger children (5-6 y) use the function of width and height, in other words an additive rule, while older children (10-12 y) use a multiplicative rule, although their paradigm has been criticized on statistical and methodological grounds by Bogartz (1978). Moreover, one of the more recent studies on this literature suggests a multiplicative model for area estimation and comparison, where the relative saliency of the axial dimensions is weighted accordingly in that multiplication (Krider, Raghubir, & Krishna, 2001). Overall, these earlier studies collectively demonstrate that area perception is influenced by size and shape of the item in many ways. Therefore, Yousif and colleagues' first conclusion can be considered as a reiteration of those findings. Then, the novel contribution of Yousif and colleagues' work comes from their second conclusion, as they investigated area perception using item arrays in the context of magnitude perception across other dimensions, including number.

3. Difficulties in designing item-array stimuli for magnitude perception studies

Studying magnitude perception using visual item arrays is notoriously difficult due to the intricate relationship between various dimensions including number, area, density, etc. Consider an array of identical-sized circles. One may quantify the magnitude of such an array as the number of circles, the area of each circle, the summation or cumulative area of all the circles, the perimeter of each circle, the cumulative perimeter of all the circles, the convex hull of the array defined as the area of the circumscribed region within which the circles are drawn, the density of the array defined as the convex hull divided by the number of circles, and potentially more. All of these dimensions are tightly interrelated (in fact, by their mathematical definitions), which engenders a significant challenge when studying the perception of these magnitudes. This is because when a researcher varies one dimension to examine its effect on behavior or neural measures, some of the other dimensions necessarily change accordingly, which may go unnoticed by the researcher.

As elaborated later in this paper, the sets of stimuli used in studies of additive-area heuristic (Yousif & Keil, 2019, 2020) suffer from two design problems. The first problem can be described as the skewed selection of stimulus parameters causing congruity and incongruity in unexpected dimensions. The second problem can be described as having unbalanced stimulus parameter ranges across different magnitude dimensions. These problems make it difficult, if not impossible, for one to make firm conclusions about magnitude perception across multiple dimensions, simply because the item arrays are not designed to make proper interpretations of those dimensions.

To understand the first problem, consider a number discrimination task with a binary discrimination between two dot arrays. The arrays can be constructed such that the individual dot area or the total dot area is held constant between the two arrays. In the former case, number is congruent with total dot area. In the latter case, number is incongruent with individual dot area. It has long been known that judgment on number is strongly influenced by whether and how much other dimensions are congruent or incongruent with number (e.g., see Allik & Tuulmets, 1993; Clearfield & Mix, 1999; Ginsburg & Nicholls, 1988; Miller & Baker, 1968; Xu & Spelke, 2000), making it crucial to construct stimulus sets that balance congruent and incongruent effects. As elaborated below, stimuli used in the additive-area heuristic studies show extremely skewed distributions in their stimulus parameters, eliciting strong congruity and incongruity effects in an unexpected way.

A second design choice in stimulus design construction needs to be considered. Consider again a binary-choice number discrimination task, in which the range of the number dimension is very narrow, say, between 20 and 26 dots (1.3:1 ratio). It is problematic if other salient magnitude dimensions such as cumulative area is allowed to have a much larger range (e.g., 5:1 ratio difference). This is because our sensory system is sensitive to the statistical regularities of the sensory input over time (Barlow, 2001). Thus, if one magnitude dimension changes in a much larger range than another over time, the sensitivity of the visual system is likely to become biased towards that dimension. Stimuli used in the additive-area heuristic studies also show this issue.

The current paper describes the flaws in Yousif & colleagues' stimulus design by re-introducing an innovate framework for constructing and analyzing dot array stimuli for

magnitude perception studies (DeWind, Adams, Platt, & Brannon, 2015). This framework establishes three orthogonal axes—number, size, and spacing—in which their linear combinations allow all other prominent magnitude dimensions to be explicitly defined in that stimulus parameter space. The two aforementioned problems in the additive-area heuristic studies are analyzed and visualized, and the importance of a systematic (i.e., unskewed and balanced) stimulus design is demonstrated from a re-analysis of Tomlinson, DeWind & Brannon (2020), a previous study that addressed the exact same question as in Yousif & Keil (2019; 2020).

4. Systematic construction of dot-array stimuli by DeWind et al. (2015)

DeWind and colleagues (2015) developed a seminal framework for establishing systematic relations between magnitude dimensions of a dot array. Assume an array of n number of homogenous dots with a radius of r_d randomly dispersed within an invisible circular field with a radius of r_f. Then, various other prominent magnitude dimensions can be defined mathematically by these three variables, as shown in **Table 1**. A log (base 2) transformation of these variables "linearizes" those dimensions which allows all of them to be represented as a linear combination of the three orthogonal dimensions (**Fig. 1**). In the original paper by DeWind and colleagues, number (N), size (Sz), and spacing (Sp) were chosen as the three orthogonal dimensions for ease of interpretations; however, it should be noted that analyses based on this mathematical framework works in an identical fashion for any three orthogonal dimensions chosen from the stimulus space. Importantly, size and spacing are meaningful, interpretable dimensions. Size refers to the dimension that changes both TA and IA together while holding N constant. Spacing refers to the dimension that changes both $Spar$ and FA together while holding N constant. Readers are strongly encouraged to read DeWind et al. (2015) for details.

Table 1. Mathematical relationship between various magnitude dimensions.

Dimension	As a function of n, r_d, r_f	Relationship with N, Sz, and Sp
Individual area (IA)	πr_d^2	$\log(IA) = \frac{1}{2}\log(Sz) - \frac{1}{2}\log(N)$
Total area (TA)	$n \times \pi r_d^2$	$\log(TA) = \frac{1}{2}\log(Sz) + \frac{1}{2}\log(N)$
Field area (FA)	πr_f^2	$\log(FA) = \frac{1}{2}\log(Sp) + \frac{1}{2}\log(N)$
Sparsity (Spar)	$\pi r_f^2/n$	$\log(Spar) = \frac{1}{2}\log(Sp) - \frac{1}{2}\log(N)$
Individual perimeter (IP)	$2\pi r_d$	$\log(IP) = \log(2\sqrt{\pi}) + \frac{1}{4}\log(Sz) - \frac{1}{4}\log(N)$
Total perimeter (TP)	$n \times 2\pi r_d$	$\log(TP) = \log(2\sqrt{\pi}) + \frac{1}{4}\log(Sz) + \frac{3}{4}\log(N)$
Coverage (Cov)	$n \times r_d^2/ r_f^2$	$\log(Cov) = \frac{1}{2}\log(Sz) - \frac{1}{2}\log(Sp)$
Closeness (Close)	$\pi^2 \times r_d^2 \times r_f^2$	$\log(Close) = \frac{1}{2}\log(Sz) + \frac{1}{2}\log(Sp)$

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

One detail worthy of note is that size can be defined in different ways. It can be defined in terms of area (i.e., $\log(Sz) = \log(TA) + \log(IA)$) or in terms of perimeter (i.e., $\log(Sz) = \log(TP) + \log(IP)$), as done in Experiment 2 of Park et al. (2016). In other words, number can be orthogonalized against area or it can be orthogonalized against perimeter. When number is orthogonalized against area, TP is more correlated with number than TA. This is because one log-unit increase in N corresponds to ¾ log-unit increase in TP, while it

only corresponds to ½ log-unit increase in *TA* (see **Table 1**). Orthogonalizing number against perimeter prevents this greater confound between *N* and *TP* because in that case one log-unit increase in *N* corresponds to ½ log-unit increase in *TP*. However, this manipulation raises the range of *TA* variation to the second power compared to the range of variation in other dimensions (*TP*, *N* and *Sp*). These different ways of orthogonalizing produces slightly different statistical regularities of the dimension parameters.

Figure 1. Properties of magnitude dimensions represented in three orthogonal axes defined by log-scaled number, size, and spacing. The framework originally developed by DeWind and colleagues (2015) allows various magnitude dimensions of a dot array to be represented systematically in this 3D parameter space.

It should be noted that the field of numerical cognition has long been developing ways to construct item arrays systematically. There have been earlier efforts to better identify the relations between various magnitude dimensions (Dehaene, Izard, & Piazza, 2005; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). Besides DeWind et al. (2015), many recent studies in numerical cognition have used systematic design of item-array stimuli at least within the 2-dimensional parameter space pertinent to their research questions (Castaldi, Piazza, Dehaene, Vignaud, & Eger, 2019; Cicchini, Anobile, & Burr, 2016), and across the 3-dimensional parameter space (Park, 2018; Van Rinsveld et al., 2020).

5. Regression-based assessment of the effect of magnitude dimensions on choice behavior: re-analysis of Tomlinson et al. (2020)

A recent study by Tomlinson et al. (2020) provides a good demonstration of a systematic design of dot arrays using DeWind and colleagues' framework for investigating area perception. In that study, adult (and child) participants made number and area judgments in a between-subjects design on pairs of dot arrays. The dot array stimuli were constructed using the aforementioned mathematical framework so that the dimensional properties of the dot arrays were systematically distributed in equal ranges of *N*, *Sz*, and *Sp*. Specifically, each trial presented a reference dot array with a set of stimulus parameters (*N*=16; *TA*=4,926; *Sz*: 1.5166e+06; *Sp*: 2.4674e+08) and a deviant dot array with a set of stimulus parameters that were distributed systematically across a "cube" in the parameter space defined by the three orthogonal axes *N*, *Sz*, and *Sp* (**Fig. 2A**). This design ensured an explicit understanding of the amount of congruency and incongruency between magnitude dimensions and the range of ratios (or log differences) across them.
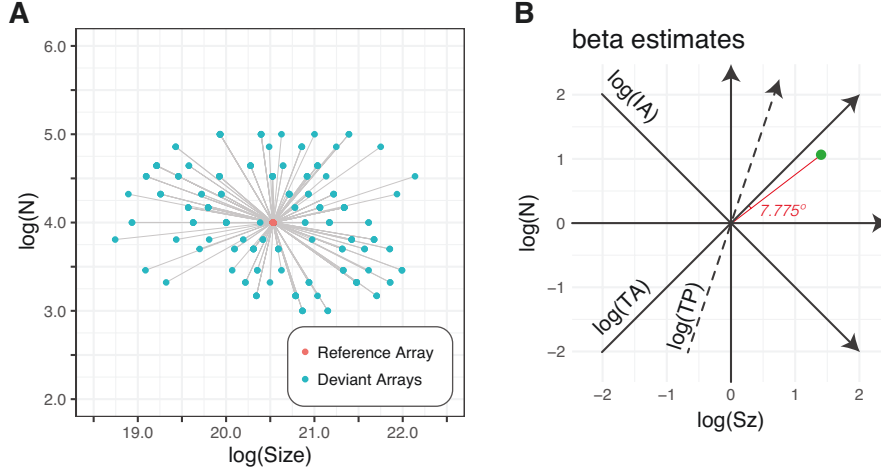
Figure 2. Stimulus design and result of Tomlinson et al. (2020). A. Distribution of stimulus parameters for dot arrays used in one adult participant's area judgment task. The red dot in the center indicates the reference dot array in that study which consisted of 16 (hence, log(N)=4) dots with the total area of 4,926 pixels. Blue dots indicate the stimulus parameters for the deviant arrays. The gray lines connecting blue dots to the red dot depicts the fact that each of the deviant arrays constructed using the unique stimulus parameter was paired up with the reference dot array in the area judgment task. The pattern of these dots and lines illustrate that deviant arrays differed from the reference array systematically across all the dimensions that can be expressed as a linear combination of log(N) and log(Sz). B. Result of the regression-based analysis of the effect of magnitude dimensions on participants' area judgment. The beta estimate vector (represented as the green dot) lied most closely to (7.775° away from) the dimension of TA.

As in the original study by DeWind et al. (2015), Tomlinson and colleagues used a regression-based approach to assess how each dimension of interest explains participants' choice behavior. In order to demonstrate this procedure here, the raw data from Tomlinson et al. (2020) were obtained from the public repository (https://osf.io/er9du/) and re-analyzed. The mean accuracy for adults was 81% in the area judgment task. Participants' responses were modeled using a generalized linear model with a probit link function, but only with the parameters for number and size:

$$p_{cr} = \boldsymbol{\Phi}(\beta_{side} + \beta_N \log_2(r_N) + \beta_{Sz} \log_2(r_{Sz})), \quad (1)$$

where p_{cr} is the proportion of trials the participants selected the stimulus on the right, $\boldsymbol{\Phi}$ is the cumulative normal distribution, β_{side} is the intercept indicating side bias, r_N and r_{Sz} are the ratios of the N and Sz of the right array to the N and Sz of the left array, and finally β_N and β_{Sz} are the regression coefficients. The spacing dimension was collapsed in order to better match the stimuli used in Yousif & Keil (2019; 2020) which do not provide spacing information. Moreover, spacing overall had little influence on both number and total area judgment tasks (Tomlinson et al., 2020).

The beta estimates for N (β_N) and Sz (β_{Sz}) produced by **Equation 1** illustrate the extent to which ratios (or log differences) in those dimensions influence choice behavior. Specifically, the degree to which the vector $\boldsymbol{\beta}=[\beta_{Sz}, \beta_N]$ is close to an axis representing a dimension (e.g., N, TA, Sz, etc.) can be interpreted as how much that dimension

determined choice behavior compared to other dimensions. For instance, $\boldsymbol{\beta}=\alpha[1,1]$ (where α is a scaling constant) would imply that participants made judgments based on total area (note $\log(TA) = \frac{1}{2} \times \log(N) + \frac{1}{2} \times \log(Sz)$) with no influence from other task-irrelevant dimensions. The result of this analysis, $\boldsymbol{\beta}=[\beta_{Sz}, \beta_N]=[1.3982, 1.0623]$, is shown as the green dot in **Figure 2B**. This vector lied most closely to the dimension of TA (7.775 deg clockwise from the TA axis), indicating that participants' choice behavior was best explained by total area.

For the completeness of the current re-analysis, the same analysis was done on the choice behavior in which a different group of participants made number judgments. This analysis resulted in $\boldsymbol{\beta}=[\beta_{Sz}, \beta_N]=[0.0851, 4.6183]$, which was only 1.0559 deg clockwise away from the N axis. Consistent with what was demonstrated in DeWind et al. (2015), these results indicate that participants' numerosity discrimination was most closely based on the number of dots in an array.

It should be emphasized that the psychometric modeling can be done using any two dimensions without affecting the conclusion. For example, choice behavior can be modeled using IA and TA as in:

$$p_{cr} = \boldsymbol{\Phi}(\beta_{side} + \beta_N \log_2(r_{IA}) + \beta_{Sz} \log_2(r_{TA})). \quad (2)$$

Analyzing Tomlinson and colleagues' adult data using this model yields $\boldsymbol{\beta}=[\beta_{TA}, \beta_{IA}]=[2.7981, 0.3369]$. This vector is 7.792 deg clockwise from the TA axis. This is exactly how much the vector $\boldsymbol{\beta}=[\beta_{Sz}, \beta_N]$ was rotated away from the TA axis within rounding error (see above).

Of importance is that β coefficients should be interpreted collectively. Hence, for a straightforward interpretation of the results and to avoid multicollinearity, it is recommended to use orthogonal dimensions in the psychometric model. For instance, the same data analyzed above yields the estimate of $\boldsymbol{\beta}=[\beta_{TP}, \beta_N]=[5.5961, -3.1349]$ when TP and N are used, and yields the estimate of $\boldsymbol{\beta}=[\beta_{TA}, \beta_N]=[2.7981, -0.3367]$ when TA and N are used. Note how β_N in these two different models are different from each other, which are also different from $\boldsymbol{\beta}=[\beta_{Sz}, \beta_N]=[1.3982, 1.0623]$. But, again, these β coefficients cannot be interpreted individually.

In fact, $\boldsymbol{\beta}=[\beta_{TP}, \beta_N]=[5.5961, -3.1349]$, $\boldsymbol{\beta}=[\beta_{TA}, \beta_N]=[2.7981, -0.3367]$, and $\boldsymbol{\beta}=[\beta_{Sz}, \beta_N]=[1.3982, 1.0623]$ are all essentially the same results after some vector algebra. First, for an easy comparison, let us normalize $[\beta_{Sz}, \beta_N]$ which yields $[0.7962, 0.6050]$. Then, take the first result, $[\beta_{TP}, \beta_N]=[5.5961, -3.1349]$. Given that the dimension TP is defined by one log-unit increase in Sz and three log-unit increase in N (see **Table 1**), $\beta_{TP} = 5.5961$ refers to a vector defined by that magnitude along the $[Sz, N] = [1, 3]$ axis. Separately, $\beta_N = -3.1349$ refers to a vector defined by that magnitude along the $[Sz, N] = [0, 1]$ axis. Thus, some algebra yields the normalized vector $[\beta_{TP}, \beta_N]=[\beta_{Sz}, \beta_N]=[0.6313, 0.7755]$. The same can be done with the second result, which yields the normalized vector $[\beta_{TA}, \beta_N]= [\beta_{Sz}, \beta_N]=[0.7695, 0.6386]$. Thus, the results from using $[\beta_{TP}, \beta_N]$ and $[\beta_{TA}, \beta_N]$ can be transformed to the results using the orthogonal dimensions $[\beta_{Sz}, \beta_N]$. Some numerical differences between the three results exists, however, likely caused by the spurious effects of multicollinearity when two correlated dimensions are used as regressors, which is another important reason for using orthogonal dimensions in the psychometric model.

6. Problems with the stimulus design in Yousif & Keil (2019; 2020)

6.1. Skewed sampling of stimulus parameters

Now that an example of a systematic stimulus design is introduced and explained in the previous section, this section provides an elaborate description of the stimulus design by Yousif & Keil (2019; 2020) using the mathematical framework developed by DeWind and colleagues (2015). The stimulus design from Yousif & Keil (2019) was obtained from the public repository (https://osf.io/dc5t8/). The file contained stimulus information about number, additive area, and true area for pairs of dot arrays used in area judgment tasks. In most cases (as long as the average aspect ratio between the width and height of each item is one), additive area referred to the summation of the diameters across all the dots in an array (or width in case of squares). That said, this dimension is equivalent to total perimeter (TP) in the aforementioned framework (differing only in scale). True area referred to the summation of the area across all the dots in an array, which is equivalent to total area (TA) in the framework. The terms used by Yousif & Keil, namely true area and additive area, are now referred to as $TrueA$ and $AddA$, in order to avoid confusion with the terms used in DeWind and colleagues' framework. Again, $TrueA$ is equivalent to TA and $AddA$ is equivalent to TP. No information about spacing was available, so that third orthogonal dimension was excluded from further analysis, as in the re-analyses of Tomlinson et al. (2020) in the previous section.
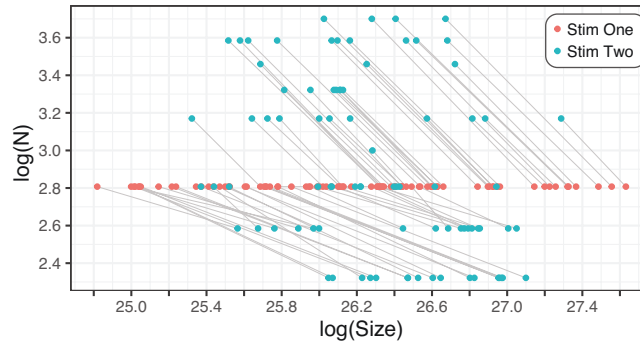

Figure 3. Distribution of stimulus parameters for dot arrays used in Experiment 1 & 2 of Yousif & Keil (2019). The red dots locate the stimulus parameters of the first array of each pair, and the blue dots locate the stimulus parameters of the second array of each pair. The grey line connects the two arrays within each pair. As can be seen, there was a systematic anti-correlation between N and Sz in that stimulus design. This kind of skewed sampling of stimulus parameters may yield biased results.

Stimulus properties between all the pairs of dot arrays used in Exp. 1 & 2 of Yousif & Keil (2019) were appropriately transformed to be represented in $\log(N)$ and $\log(Sz)$. Then, the stimulus properties of those pairs were visualized in a 2-dimensional space in **Figure 3**. As shown, all the pairs differed in their stimulus parameters almost exclusively along the IA axis, although with some variations in the angle. This is a prototypical example of a skewed selection of stimulus parameters, that emerges when experimenters attempt to manipulate just two dimensions ($TrueA$ and $AddA$ in this case). Specifically, this stimulus design produces a strong incongruity between N and Sz.

For a different perspective of this stimulus design, the first array of each pair was compared with the second array of each pair in *N*, *TrueA*, and *AddA*. Specifically, the difference between log-transformed *N*, *TrueA*, and *AddA* were taken between arrays within each pair, in order to quantify how the two arrays differed in those dimensions. The distribution of this difference across all stimulus pairs as well as cross-correlations between the differences across the three dimensions are visualized in **Figure 4**. Although the mean of log difference was comparable across the dimensions of N (-0.115), TrueA (-0.106), and AddA (-0.105), the distribution of log difference showed a greater variance along the *N* dimension than along the *TrueA* and *AddA* dimensions (**Fig. 4A**). Moreover, there was a strong negative correlation between how much the two arrays in each pair differed in *TrueA* and *N*, and a strong positive correlation between how much the two arrays in each pair differed in *AddA* and *N* (**Fig. 4B**). This means that the stimuli were selectively created to impose a very strong incongruency between *TrueA* and *N* and a very strong congruency between *AddA* and *N*. If dot arrays were constructed systematically across the mathematically defined parameter space, *N* should be positively correlated with *TA* (i.e., $\log(TA) = \frac{1}{2} \times \log(N) + \frac{1}{2} \times \log(Sz)$) and slightly more positively correlated with *TP* (i.e., $\log(TP) = \log(2 \times \mathrm{sqrt}(\pi)) + \frac{3}{4} \times \log(N) + \frac{1}{4} \times \log(Sz)$) (see **Figure S1** in Supplementary Materials for the same scatterplot matrix acquired from the stimuli set used in Tomlinson et al. 2020). In contrast, *N* was highly negatively correlated with *TrueA* (equivalent of *TP*). In sum, while the authors' intention may have been to equate the ratios between *TrueA* and *AddA* on average, this design artificially yielded a strong incongruity between *TrueA* and *N*. It should be noted that all of their experiments in Yousif & Keil (2019; 2020) showed an extreme skewness in their stimulus parameters. See Supplementary Materials (**Fig. S2-S11**) for the analyses of the stimulus design in the rest of the experiments.
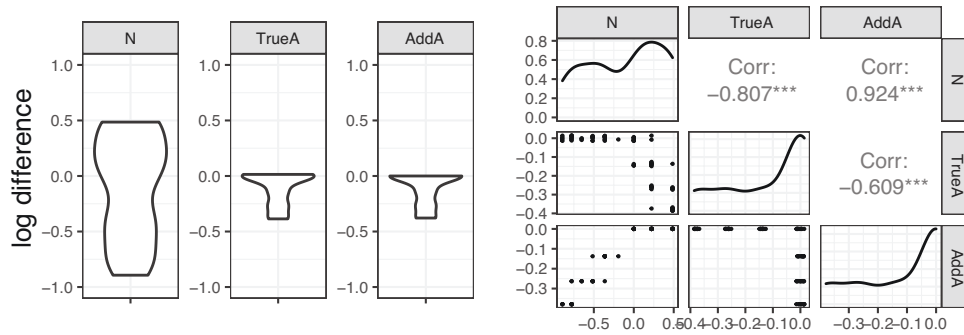

Figure 4. Stimulus design from Yousif & Keil (2019, Exp. 1 & 2). A. Distribution of log differences (or ratios) within pairs of arrays in number (*N*), true area (*TrueA*), and additive area (*AddA*). B. Scatterplot matrix illustrating the correlational structure of those log differences across the three dimensions.

6.2. Unbalanced stimulus parameter ranges

Besides the problem of skewed sampling of stimulus parameters, Yousif & Keil's (2019; 2020) stimulus design suffers from another problem, namely having unbalanced stimulus parameter ranges across different magnitude dimensions. This issue is already shown in their stimulus design of Yousif & Keil (2019) (see **Fig. 4A**), but it becomes exacerbated

in Yousif & Keil (2020). In that study, participants were asked to make judgments on area and number, as the authors were interested in evaluating how area affects number judgment and vice versa. The design of dot arrays used in those experiments revealed that the distribution of stimulus parameters across trials was vastly different across dimensions. **Figure 5A** illustrates how much N and Sz varied across all the dot arrays used in Exp 2 of Yousif & Keil (2020) (obtained from https://osf.io/kx3sv/). While the range of N across all the arrays was only about 0.5 in log scale (i.e., about 1.4:1 ratio), the range of Sz was close to 2.5 in log scale (i.e., 5.6:1 ratio). Such a difference between the two dimensions was also apparent when the log difference between two arrays within each pair was quantified. Even though the mean of the log-difference measures was comparable between the two dimensions (0.111 for N and 0.110 for Sz), the distribution of this log difference was fairly narrow (with a range below 0.4) in N, while the distribution of this log difference was vastly wide (with a range about 2.5) in Sz. This means that across all the trials, participants saw dot arrays with much larger differences in Sz than in N, and that they saw pairs that differed much more substantially in Sz than in N. Remember that $\log(TA) = \log(N) + \log(Sz)$, indicating that large ranges of those measures in Sz results in large ranges in TA. The same pattern of unbalanced ranges was observed in all of Yousif & Keil's experiments (see **Fig. 4** and **Fig. S7-S11** in Supplementary Materials). In contrast to Yousif & Keil's design, distributions of stimulus parameters in Tomlinson et al. (2020) were more balanced and comparable (**Fig. 5B**).
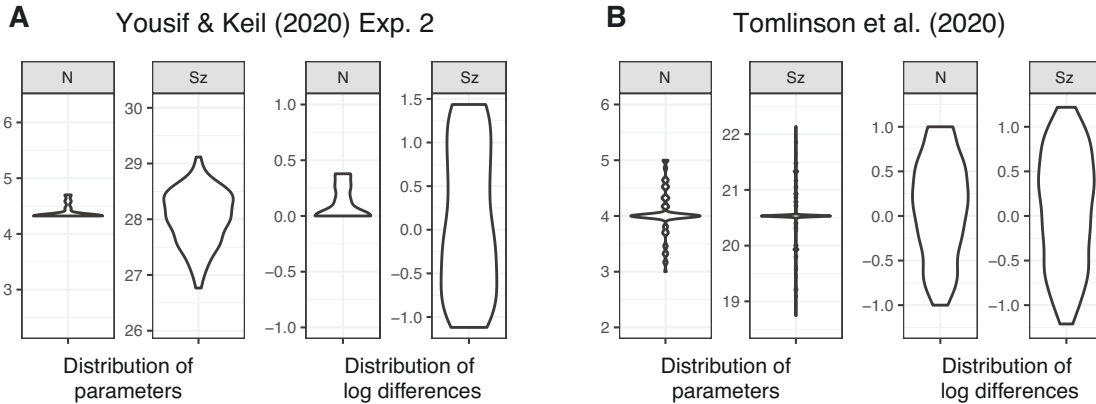

Figure 5. Distribution of parameter values across all the stimuli and distribution of log differences (or ratios) across all the pairs of arrays used in the two experiments (Yousif & Keil, 2020; Tomlinson et al., 2020). Yousif & Keil's arrays varied much widely along the Sz dimension compared to the N dimension, while Tomlinson and colleagues' dot arrays varied in a comparable range between the two dimensions. The same was true with log difference measures between the two arrays within each pair.

6.3. Interpretation of Yousif and Keil's (2019; 2020) data considering their design problems

The skewed and unbalanced sampling of stimulus parameters prevents one from making an unbiased interpretation about magnitude perception. Consider Exp. 1 of Yousif & Keil (2019) where N is highly negative correlated with $TrueA$ but highly positively correlated with $AddA$. This means that when participants view two dot arrays where one (say Array

Left) of them is greater in *TrueA* than the other (say Array Right), Array Left is smaller in *N*. Simultaneously, the range of differences in *N* is so much larger than the range of differences in *TrueA* making *N* a more salient dimension that could possibly influence choice behavior. Then, participants are likely to make substantially more incorrect decisions than when such a strong incongruity did not exist. In contrast, when Array Left is greater than Array Right in *AddA*, Array Left is also greater in *N*. In such a case if participants' magnitude perception again is biased by the saliency of number, their perception would be boosted by the especially high correlation between *AddA* and *N* and would likely to be correct more so than when such a strong congruity did not exist.

Figure 6A explains this hypothetical but plausible scenario in graphics. The strongly skewed distribution of stimulus parameters can be illustrated as the cyan cloud in the dimensional space defined by *N* and *Sz*. Let us assume that an observer makes an unbiased judgment of area (i.e., based on *TrueA*) as shown in adult participants of Tomlinson et al. (2020). Then, that observer's choice behavior would be modeled as beta estimates that would lie on the axis of *TA* (point a in **Fig. 6A**). Now, a strong negative correlation between *TA* and *N* presumably influences the observer's choice behavior, in that he or she perceives more numerous arrays to contain more area. This is a reasonable presumption because number is the most salient visual feature processed in the earliest level of cortical processing (Fornaciai, Brannon, Woldorff, & Park, 2017; Park, DeWind, Woldorff, & Brannon, 2016) and because log difference in *N* is about three times as much of that in *TA* in Yousif & Keil (2019). This perceptual interference, then, would make the choice behavior rely less on *Sz* but more on *N*, pulling the beta estimates accordingly (point b in **Fig. 6A**). Simultaneously, the direction to which choice behavior is influenced is the opposite, because in reality, more numerous arrays always had smaller *TA*. Therefore, the beta estimate for *N* would decrease, lying closest to the *TP* (*AddA*) axis (point c in **Fig. 6A**).

Yousif & Keil (2019) performed a linear regression to explain participants' choice behavior with *AddA*, *TrueA*, and *N* as regressors. However, there are two important caveats to consider in their regression analysis. First, as shown in **Figure 4**, some of those dimensions are extremely highly correlated (up to r = 0.924) with each other. Such a severe multicollinearity makes it difficult to properly interpret the coefficients. Second, Yousif & Keil (2019) interpreted beta coefficients individually; however, as explained in Section 3, those beta coefficients must be interpreted in the context of each other. Their analysis in Exp. 1 showed that *AddA*, but not *TrueA* or *N*, predicted observer responses. However, this pattern can be fully explained by the skewed and unbalanced sampling of stimulus parameters as described above and illustrated in **Figure 6A**, considering that a virtually identical experiment but with a systematic stimulus design revealed an entirely different pattern of results (Tomlinson et al., 2020).

Note that under a strong negative correlation between *N* and another dimension, such as *TrueA* or *AddA*, it is even possible for the beta estimate for *N* to be negative in an area judgment experiment. This is because a larger value in *TrueA* or *AddA* is always associated with a smaller value in *N* so that participants need to choose the array with fewer items in order to choose the array with greater area. This pattern was indeed evident in Yousif & Keil's studies. In Exp. 4 of Yousif & Keil (2019) (see **Fig. S5**), their regression analysis showed a negative effect of *N* specific to the trials in which *TrueA*

varied along *N*. In Exp. 2b of Yousif & Keil (2020) (see **Fig. S9**), their regression analysis showed that participants were significantly below chance when selecting between arrays equal in *AddA* but differing in *N*, indicating that they perceived less numerous arrays to contain more area. Thus, these negative beta coefficients for *N*, as well as all the beta coefficients from Yousif & Keil's (2019; 2020) report, should be interpreted with caution considering the skewed distribution of stimulus parameters.

Figure 6. Schematics of how skewed and unbalanced stimulus design may explain Yousif & Keil's (2019; 2020) results. See Section 6.3 for how these designs may influence choice behavior and hypothetical beta estimates. A. Hypothetical illustration of the case of area judgment. In many of Yousif & Keil's stimulus design, there was a strong positive correlation between *N* and *TP* and a strong negative correlation between *N* and *TA*, and this pattern can be graphically illustrated as the illustrated cyan cloud. Take the right-most (*p*) and left-most (*q*) points of the cyan cloud. The projection of points *p* and *q* onto the dimensional axes gives clues about the correlational structure between dimensions. For example, a dot array with its stimulus parameters represented by point *p* is greater in *TA* than a dot array with its stimulus parameters represented by point *q* (indicated by the direction of the purple arrow). The former array, however, is smaller in *N* than the latter array (indicated by the direction of the red arrow). B. Hypothetical illustration of the case of number judgment. In some other designs, *N* was negatively correlated with both *TA* and *TP*, which as a corollary created a much more unbalanced distribution of stimulus parameters. As illustrated in both cases, such a skewed and unbalanced design can explain idiosyncratic choice behavior patterns in area and number judgments even if an observer is assumed to have an unbiased representation of area and number. As a comparison, a systematically constructed set of dot arrays would be represented by a circular or square cloud (see Fig. 2A).

Another design problem explained in Section 4.2 was unbalanced parameter ranges across dimensions. This pattern usually happens together with skewed distribution of stimulus parameters as described in Section 4.1. Unbalanced parameter ranges can be problematic because our perceptual system is sensitive to statistical regularities of the sensory input (Barlow, 2001; Kohn, 2007; Wark, Lundstrom, & Fairhall, 2007). There is, in fact, empirical evidence for it in the domain of numerosity perception. When participants are passively viewing dot arrays systematically designed across equal ranges

of *N*, *Sz*, and *Sp*, their early visual cortical activity shows almost no modulations by *Sz* or *Sp* (Park et al., 2016), but when they view dot arrays that exclusively change in *Sz* (with values of *N* and *Sp* held constant) or *Sp* (with values of *N* and *Sp* held constant), their visual cortical activity becomes modulated by those exclusive changes in one dimension (Park, Godbole, Woldorff, & Brannon, in press). Hence, our perceptual system becomes more sensitive to dimensions that change relatively larger in scale over time.

Consider Exp. 2 of Yousif & Keil (2020), where they designed the stimuli to "directly [pit] *AddA* and *N* against each other" in number judgment and area judgment tasks. However, their stimulus design unfortunately became extremely skewed and unbalanced, which is schematically represented in **Figure 6B** (for real values, see **Fig. S8 and S9**). Having this skewed and unbalanced stimulus design in mind, consider the case of number judgment. It is well known that, when dot arrays are systematically designed, participants are most sensitive to changes in *N* than changes in *Sz* or *Sp* with a slight bias towards the positive *Sz* direction (e.g., DeWind et al., 2015; Tomlinson et al., 2020). Hence, if an observer's choice behavior for number judgment is modeled, the beta estimates would lie close to the *N* axis but slightly toward the positive *Sz* direction (point a in **Fig. 6B**). In Yousif & Keil's (2020) design, however, the range of log difference in *N* was extremely small compared to the range in *Sz*. This extreme unbalance in stimulus parameters, thus, is likely to make the choice behavior rely less on *N* but more on *Sz* and even towards the negative direction along the *TA* axis, pulling the beta estimates accordingly (point b in **Fig. 6B**). This effect will make the beta estimates to be closer to the *TP* axis. This pattern was, in fact, what was reported in Exp. 2a: "Observers judged images containing more discs as more numerous … However, [they] also judged images with greater perceived area (but equal in number) to be more numerous" (Yousif & Keil, 2020). To reiterate, however, those beta estimates measuring the effect of magnitude dimensions on choice behavior should not be interpreted individually.

In sum, all of Yousif & Keil's results can be explained by the skewed and unbalanced sampling of their parameters. It should be emphasized that previous studies with systematic stimulus design all point to a completely different conclusion (e.g., DeWind et al., 2015; Cicchini et al., 2016; Tomlinson et al., 2020). Therefore, the results from Yousif & Keil (2019; 2020) including their regression coefficients should be interpreted with caution. Moreover, their sampling of the stimulus parameters differs sometimes minimally and other times substantially across studies (compare all the figures in Supplementary Materials). Such a variability of stimulus structure makes it even more difficult, if not impossible, to make a coherent interpretation about the obtained results.

7. Issues in the reanalysis of Tomlinson et al. (2020) by Aulet & Lourenco (2020)

As introduced earlier, Tomlinson et al. (2020) addressed the exact same research question as Yousif & Keil (2019; 2020) in both children and adults with a set of systematically constructed dot arrays and reached a completely different conclusion. Specifically, Tomlinson and colleagues found that the acuity for making a number judgment was much higher than the acuity for making a total area (*TA*) judgment both in children and adults. In addition, adult participants' choice behavior model indicated that the beta coefficients lied more or less on the *N* and *TA* axes for number and area judgment tasks, respectively,

indicating that their judgments were fairly unbiased. In contrast, children showed a mutual bias between *N* and *TA* in both tasks.

Interestingly, Aulet and Lourenco (2020) reanalyzed Tomlinson and colleagues' data specifically questioning the role of *AddA* in choice behavior and reached slightly different conclusions. For instance, they used a similar choice behavior model (**Eq. 1**) including *N* and *AddA* as regressors to conclude that "when using a measure of perceived area, rather than mathematical area, in the analyses, acuity was comparable for number and area in both children and adults" (p. 3, Aulet & Lourenco, 2020). However, some of these interpretations are premature. This is because Aulet & Lourenco (2021) analyzed the data using correlated regressors but did not account for those correlations in their interpretation of the *β* coefficients. For instance, they treated individual beta coefficients directly as a measure of acuity or bias, when in fact all the beta coefficients need to be interpreted together within the parameter space. See Sections 3 and 4 for detailed explanations about the importance of using orthogonal dimensions in the psychometric model.

Despite issues with their analytic method, one of Aulet & Lourenco's (2021) conclusions about the effect of *AddA* on children's choice behavior does hold. Remember that *TP* is on the same scale as *AddA* defined by Yousif & Keil (2019). However, Tomlinson et al. (2020) unfortunately did not discuss *TP* as a central dimension of interest in their paper. One most reasonable approach to assess how *AddA* influences one's area (or number) judgment in Tomlinson and colleagues' study would have been to model the data using two orthogonal dimensions (e.g., *N* and *Sz*) and evaluate the degree to which the beta estimate vector deviates away from the *TP* axis. When this exact analysis was done here with Tomlinson and colleagues' children data on area judgment, the estimate vector $\boldsymbol{\beta}=[\beta_{Sz}, \beta_N]=[0.4015, 1.2442]$ was 17.89 deg away clockwise from the *N* axis, and very close to the *TP* axis (0.5497 deg away from it), suggesting that children's area judgment was best explained by *TP* or *AddA*.

For completeness, the same analysis was done with children's number judgment data from Tomlinson et al. (2020), which resulted in $\boldsymbol{\beta}=[\beta_{Sz}, \beta_N]=[0.3237, 1.9456]$. This vector lied in between the *N* and *TP* axes, 9.448 deg away from the *N* axis and 8.990 deg away from the *TP* axis, suggesting some bias from *TP* or *AddA* for children's number judgment. However, another study incorporating the same analysis in a different sample of children demonstrated that children's number judgment, best explained by *N,* had little bias from other dimensions (Starr, DeWind, & Brannon, 2017). Thus, future research is needed to better determine the role of perimeter or additive area in children's perception of magnitude.

8. Discussion of Yousif, Aslin, and Keil (2020)

While Yousif & Keil's (2019; 2020) manipulation of stimulus parameters involved changes in number, Yousif, Aslin, & Keil (2020) constructed item arrays that were constant in number. In that study, participants made area judgments on two item arrays that equated for *N* but differed in *AddA* and *TrueA*. In Exp. 1, this manipulation, that is de-confounding *AddA* and *TrueA*, was achieved by creating differences in the variability in area across items within an array. In Exp. 2, this manipulation was achieved by

presenting squares in one array and diamonds containing the same area in the other array. As number was kept constant between the arrays to be compared, there was no correlational structure between N and $TrueA$ or between N and $AddA$ to be considered; therefore, those experiments have little relevance to number perception. Hence, stimulus design in Yousif, Aslin, & Keil (2020) is not directly subject to the main criticisms raised in this paper. However, it is worth discussing that study in the broader context of magnitude perception.

In Exp. 1 of Yousif, Aslin, & Keil (2020), the way that $TrueA$ and $AddA$ was de-confounded creates another confound between homogeneity of the array and whether or not $AddA$ or $TrueA$ is greater. When $TrueA$ between the two arrays is equated, $AddA$ is larger for the more homogeneous array. When $AddA$ is equated, $TrueA$ is smaller for the more homogeneous array. Therefore, if participants are, for any reason, biased to respond to the more homogeneous array of the two given, then the results in Yousif, Aslin, & Keil (2020) could be explained by that effect. This scenario is hypothetical but plausible given that the ensemble representation literature demonstrates that "judgments of mean set size become less accurate when set size [in other words, number] increases and the heterogeneity of the item sizes increases" (Marchant, Simons, & de Fockert, 2013). This previous work may also explain why participants in Yousif, Aslin, & Keil (2020) showed even below chance accuracy for true area trials when the set size increased from 6 (Exp. 1a) to 10 (Exp. 1b). An interesting future direction is to explicitly test for an effect of homogeneity of items on area perception, which could open a new research avenue that connects magnitude perception and ensemble representation (Ariely, 2001; Chong & Treisman, 2005).

As reviewed in Section 2, the demonstration that diamonds are perceived to be larger than squares in Exp. 2 of Yousif, Aslin, & Keil (2020) is not novel. In fact, how area perception is influenced by shape (including individually presented star, diamond, rhombus, triangle, square, circle, and other polygons) has been studied since nearly a century ago (e.g., Anastasi, 1936; Fisher & Foster, 1968; Krider, Raghubir, & Krishna, 2001; Warren & Pinneau, 1955). A novel contribution of Exp. 2, then, is the use of an *array* of items of different shapes. However, such a shape bias is difficult to be directly contextualized in the number perception literature because shape has been a constant factor in most previous number perception studies.

9. Conclusions

This paper raises issues with recent "additive area heuristic" studies published by Yousif & Keil (2019; 2020) and demonstrates two flaws in their stimulus design. The selection of stimulus parameters was extremely skewed towards making congruity and incongruity structures between some dimensions, and the stimulus parameters across different dimensions were extremely unbalanced. Such a design is usually a consequence of the intention to control for certain dimensions without knowing how such manipulations impact the other dimensions. From the standpoint that the authors aimed to interpret their results and made strong claims about general magnitude perception across multiple dimensions, such a design is flawed because, as elaborated in Section 6, it does not allow a proper interpretation of how various magnitude dimensions together influence

participants' choice behavior. Moreover, considering the markedly different conclusions reached by Yousif & Keil (2019; 2020) and by Tomlinson et al. (2020) as well as other previous studies that used systematically constructed dot-array stimuli (e.g., DeWind et al., 2015; Cicchini et al., 2016; Starr, DeWind, & Brannon, 2017), conclusions reached by Yousif & Keil (2019; 2020; 2021) are likely invalid.

Nevertheless, while Yousif & Keil's (2019; 2020) stimulus design is problematic, the jury may still be out on how cumulative area of an item array is perceived. For singly presented items, older literature has shown mixed results. Some studies demonstrated that perceived size of a 2-dimensional image increases more slowly than the physical size even for adult observers (e.g., Stevens & Guirao, 1963; Teghtsoonian, 1965), but it was not always the case (e.g., Ekman & Junge, 1961). Also, studies suggested that only young, but not older, children use the "width + height" rule for estimating the area (Anderson & Cuneo, 1978; Wolf & Algom, 1987). Collectively, however, previous work suggests that the perception of area of a *singly* presented item is biased to some extent (see Krider, Raghubir, & Krishna, 2001 for a concise review). Thus, Tomlinson and colleagues' (2020) finding that adults' estimation of cumulative area of an *array* of items is relatively unbiased could rather be interpreted as a somewhat surprising and interesting finding. For instance, if adults rely on perimeter or the additive rule when estimating the area of a singly presented item, but they rely on the mathematical area when estimating the area of an array of items, that would be an interesting pattern of results worthy of further investigations.

Arrays of items have been used to study perception of magnitude in various dimensions for more than half a century, with number as a main enterprise. However, systematic modeling of those magnitude dimensions was scant until recently. DeWind and colleagues' (2015) framework provides one elegant approach to gain an explicit understanding of the stimulus structure and visualize the stimulus parameters and analyze the data accordingly. Researchers who study magnitude perception using item arrays are encouraged to consider DeWind and colleagues' framework for stimulus construction or at least for the verification of stimulus dimensional properties. A MATLAB code published in a public repository (https://osf.io/s7xer/) may be used for an easy implementation of this framework.

References

Allik, J., & Tuulmets, T. (1993). Perceived numerosity of spatiotemporal events. *Perception & Psychophysics*, *53*(4), 450–459. https://doi.org/10.3758/BF03206789

Anastasi, A. (1936). The Estimation of Area. *The Journal of General Psychology*, *14*(1), 201–225. https://doi.org/10.1080/00221309.1936.9713146

Anderson, N. H., & Cuneo, D. O. (1978). The height?+?width rule in children's judgments of quantity. *Journal of Experimental Psychology: General*, *107*(4), 335–378. https://doi.org/10.1037/0096-3445.107.4.335

Ariely, D. (2001). Seeing Sets: Representation by Statistical Properties. *Psychological Science*, *12*(2), 157–162. https://doi.org/10.1111/1467-9280.00327

Barlow, H. (2001). The exploitation of regularities in the environment by the brain. *Behavioral and Brain Sciences*, *24*(4), 602–607. https://doi.org/10.1017/S0140525X01000024

Bogartz, R. S. (1978). Comment on Anderson and Cuneo's "The height?+?width rule in children's judgments of quantity." *Journal of Experimental Psychology: General*, *107*(4), 379–387. https://doi.org/10.1037/0096-3445.107.4.379

Bolton, F. E. (1898). A Contribution to the Study of Illusions, with Special Reference to (a) The Effect of Size upon Estimations of Weight, (b) The Effect of Contour upon Estimations of Area. *The American Journal of Psychology*, *9*(2), 167. https://doi.org/10.2307/1411756

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

Chong, S. C., & Treisman, A. (2005). Statistical processing: computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900. https://doi.org/10.1016/j.visres.2004.10.004

Cicchini, G. M., Anobile, G., & Burr, D. C. (2016). Spontaneous perception of numerosity in humans. *Nature Communications*, *7*(1), 12536. https://doi.org/10.1038/ncomms12536

Clearfield, M. W., & Mix, K. S. (1999). Number Versus Contour Length in Infants' Discrimination of Small Visual Sets. *Psychological Science*, *10*(5), 408–411. https://doi.org/10.1111/1467-9280.00177

Dehaene, S., Izard, V., & Piazza, M. (2005). Control over non-numerical parameters in numerosity experiments. Retrieved from http://www.unicog.org/docs/DocumentationDotsGeneration.doc

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition*, *142*, 247–265. https://doi.org/10.1016/j.cognition.2015.05.016

Ekman, G., & Junge, K. (1961). Psychophysical relations in visual perception of length, area and volume. *Scandinavian Journal of Psychology*, *2*(1), 1–10.

https://doi.org/10.1111/j.1467-9450.1961.tb01215.x

Fisher, G. H., & Foster, J. J. (1968). Apparent Sizes of Different Shapes and the Facility with which they can be identified. *Nature*, *219*(5154), 653–654. https://doi.org/10.1038/219653c0

Fornaciai, M., Brannon, E. M., Woldorff, M. G., & Park, J. (2017). Numerosity processing in early visual cortex. *NeuroImage*, *157*, 429–438. https://doi.org/10.1016/j.neuroimage.2017.05.069

Ginsburg, N., & Nicholls, A. (1988). Perceived Numerosity as a Function of Item Size. *Perceptual and Motor Skills*, *67*(2), 656–658. https://doi.org/10.2466/pms.1988.67.2.656

Kohn, A. (2007). Visual Adaptation: Physiology, Mechanisms, and Functional Benefits. *Journal of Neurophysiology*, *97*(5), 3155–3164. https://doi.org/10.1152/jn.00086.2007

Krider, R. E., Raghubir, P., & Krishna, A. (2001). Pizzas: π or Square? Psychophysical Biases in Area Comparisons. *Marketing Science*, *20*(4), 405–425. https://doi.org/10.1287/mksc.20.4.405.9756

Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica*, *142*(2), 245–250. https://doi.org/10.1016/j.actpsy.2012.11.002

Miller, A. L., & Baker, R. A. (1968). The Effects of Shape, Size, Heterogeneity, and Instructional Set on the Judgment of Visual Number. *The American Journal of Psychology*, *81*(1), 83. https://doi.org/10.2307/1420810

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

Park, J., DeWind, N. K., Woldorff, M. G., & Brannon, E. M. (2016). Rapid and Direct Encoding of Numerosity in the Visual Stream. *Cerebral Cortex (New York, N.Y. : 1991)*, *26*(2), 748–763. https://doi.org/10.1093/cercor/bhv017

Park, J., Godbole, S., Woldorff, M. G., & Brannon, E. M. (2021). Context-dependent modulation of early visual cortical responses to numerical and non-numerical magnitudes. *Journal of Cognitive Neuroscience*.

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. https://doi.org/10.1016/j.neuron.2004.10.014

Smith, J. P. (1969). The effects of figurai shape on the perception of area. *Perception & Psychophysics*, *5*(1), 49–52.

Starr, A., DeWind, N. K., & Brannon, E. M. (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. https://doi.org/10.1016/j.cognition.2017.07.004

Stevens, S. S., & Guirao, M. (1963). Subjective scaling of length and area and the matching of length to loudness and brightness. *Journal of Experimental Psychology*, *66*(2), 177–186. https://doi.org/10.1037/h0044984

Teghtsoonian, M. (1965). The Judgment of Size. *The American Journal of Psychology*, *78*(3), 392. https://doi.org/10.2307/1420573

Tomlinson, R. C., DeWind, N. K., & Brannon, E. M. (2020). Number sense biases children's area judgments. *Cognition*, *204*, 104352. https://doi.org/10.1016/j.cognition.2020.104352

Van Rinsveld, A., Guillaume, M., Kohler, P. J., Schiltz, C., Gevers, W., & Content, A. (2020). The neural signature of numerosity by separating numerical and continuous magnitude extraction in visual cortex with frequency-tagged EEG. *Proceedings of the National Academy of Sciences*, *117*(11), 5726–5732. https://doi.org/10.1073/pnas.1917849117

Verge, C. G., & Bogartz, R. S. (1978). A functional measurement analysis of the development of dimensional coordination in children. *Journal of Experimental Child Psychology*, *25*(2), 337–353. https://doi.org/10.1016/0022-0965(78)90087-5

Wark, B., Lundstrom, B. N., & Fairhall, A. (2007). Sensory adaptation. *Current Opinion in Neurobiology*, *17*(4), 423–429. https://doi.org/10.1016/j.conb.2007.07.001

Warren, J. M., & Pinneau, S. R. (1955). Influence of Form on Judgment of Apparent Area. *Perceptual and Motor Skills*, *5*(1), 7–10. https://doi.org/10.2466/pms.1955.5.g.7

Wolf, Y., & Algom, D. (1987). Perceptual and memorial constructs in children's judgments of quantity: A law of across-representation invariance. *Journal of Experimental Psychology: General*, *116*(4), 381–397. https://doi.org/10.1037/0096-3445.116.4.381

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11. https://doi.org/10.1016/s0010-0277(99)00066-9

Yousif, S. R., Aslin, R. N., & Keil, F. C. (2020). Judgments of spatial extent are fundamentally illusory: 'Additive-area' provides the best explanation. *Cognition*, *205*, 104439. https://doi.org/10.1016/j.cognition.2020.104439

Yousif, S. R., & Keil, F. C. (2019). The Additive-Area Heuristic: An Efficient but Illusory Means of Visual Area Approximation. *Psychological Science*, *30*(4), 495–503. https://doi.org/10.1177/0956797619831617

Yousif, S. R., & Keil, F. C. (2020). Area, not number, dominates estimates of visual quantities. *Scientific Reports*, *10*(1), 13407. https://doi.org/10.1038/s41598-020-68593-z

Yousif, S. R., & Keil, F. C. (2021). How We See Area and Why It Matters. *Trends in Cognitive Sciences*, *25*(7), 554–557. https://doi.org/10.1016/j.tics.2021.03.017
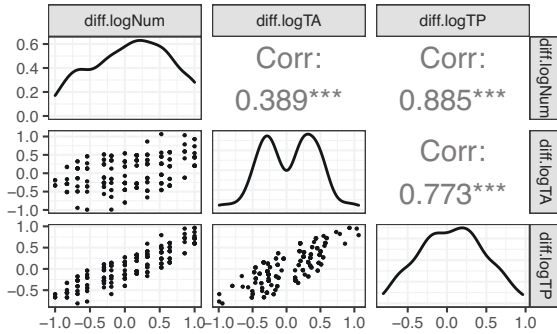
Supplementary Materials

Figure S1. Scatterplot matrix illustrating the correlational structure of log difference measures in number (diff.logNum), total area (diff.logTA), and total perimeter (diff.logTP) between the first and second array within each pair of arrays in an exemplary stimulus set from Tomlinson et al. (2020).
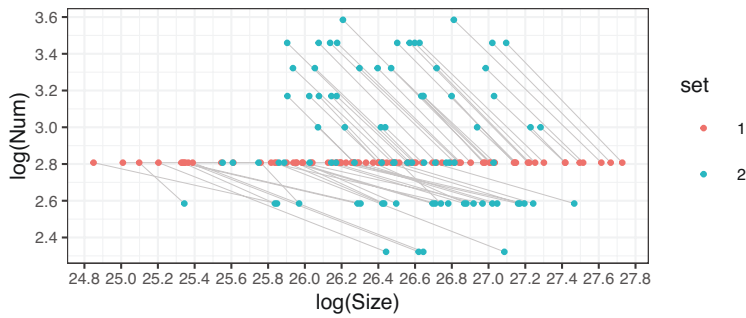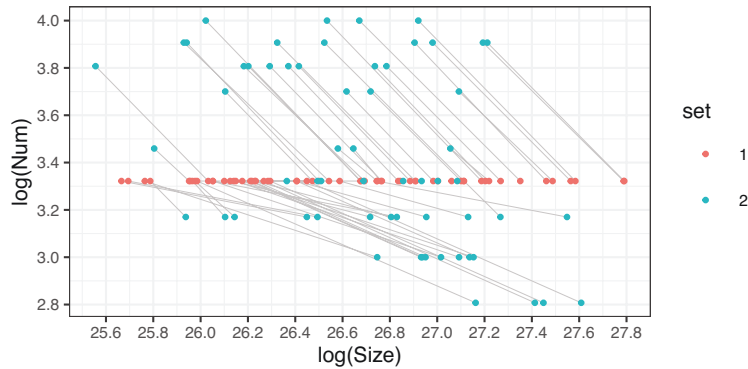

Figure S2. Distribution of stimulus parameters for dot arrays used in Experiment 3 of Yousif & Keil (2019). The red dots locate the stimulus parameters of the first array of each pair, and the blue dots locate the stimulus parameters of the second array of each pair. The grey line connects the two arrays within each pair.
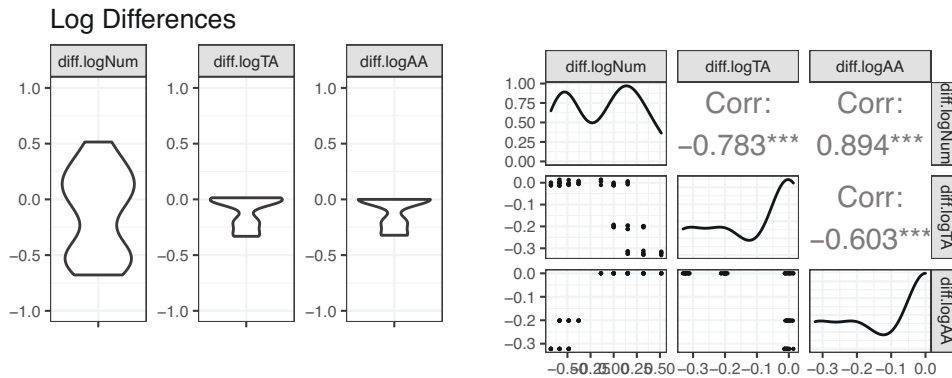

Figure S3. Distribution of log differences (or ratios) within pairs of arrays in number (*N*), true area (*TrueA*), and additive area (*AddA*), and the scatterplot matrix illustrating the

correlational structure of those log differences across the three dimensions, from Experiment 3 of Yousif & Keil (2019).
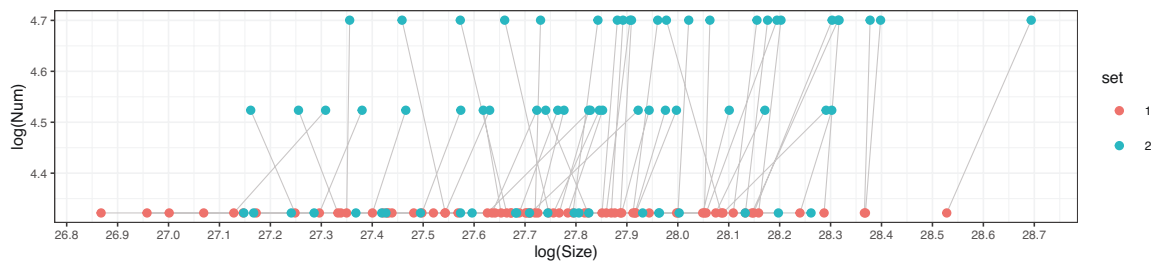

Figure S4. Distribution of stimulus parameters for dot arrays used in Experiment 4 of Yousif & Keil (2019). The red dots locate the stimulus parameters of the first array of each pair, and the blue dots locate the stimulus parameters of the second array of each pair. The grey line connects the two arrays within each pair.
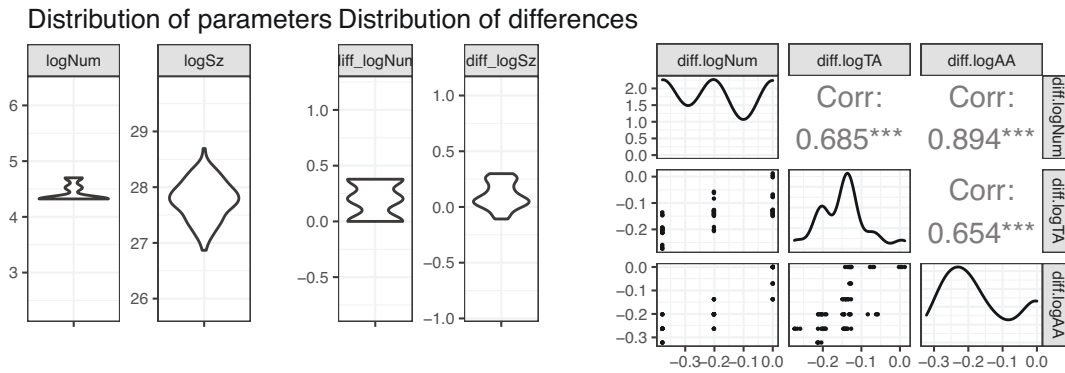

Figure S5. Distribution of log differences (or ratios) within pairs of arrays in number (*N*), true area (*TrueA*), and additive area (*AddA*), and the scatterplot matrix illustrating the correlational structure of those log differences across the three dimensions, from Experiment 4 of Yousif & Keil (2019).
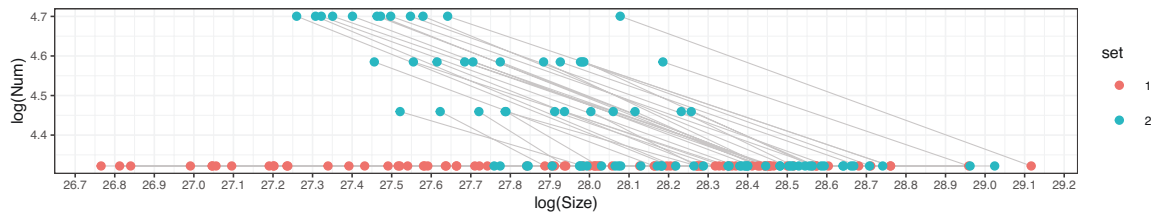

Figure S6. Distribution of stimulus parameters for dot arrays used in Experiment 1 of Yousif & Keil (2020). The red dots locate the stimulus parameters of the first array of

each pair, and the blue dots locate the stimulus parameters of the second array of each pair. The grey line connects the two arrays within each pair.

Figure S7. Distribution of log differences (or ratios) within pairs of arrays in number (*N*) and size (*Sz*), and the scatterplot matrix illustrating the correlational structure of log differences across number (*N*), total area (*TA*), and additive area (*AA*), from Experiment 1 of Yousif & Keil (2020).

Figure S8. Distribution of stimulus parameters for dot arrays used in Experiment 2 of Yousif & Keil (2020). The red dots locate the stimulus parameters of the first array of each pair, and the blue dots locate the stimulus parameters of the second array of each pair. The grey line connects the two arrays within each pair.
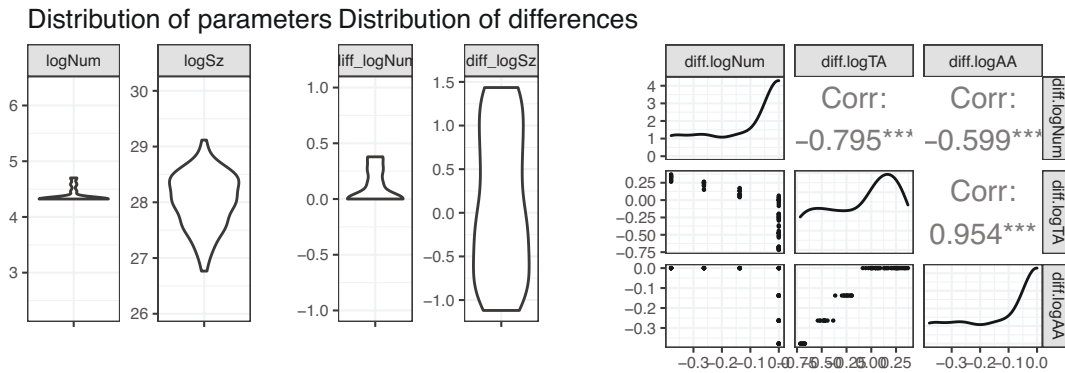

Figure S9. Distribution of log differences (or ratios) within pairs of arrays in number (*N*) and size (*Sz*), and the scatterplot matrix illustrating the correlational structure of log

differences across number (*N*), total area (*TA*), and additive area (*AA*), from Experiment 2 of Yousif & Keil (2020).
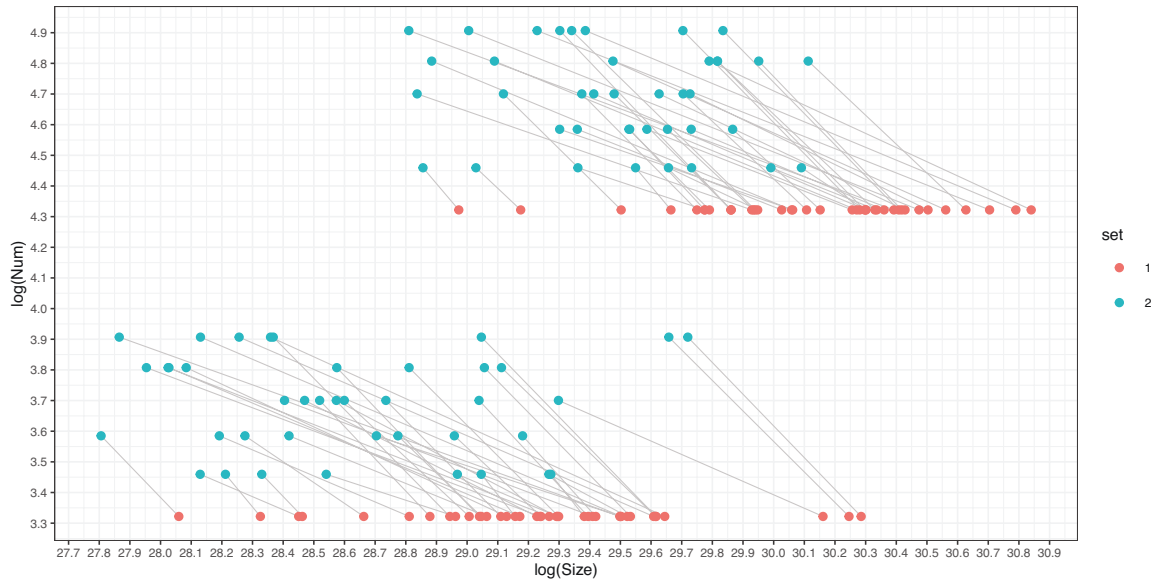

Figure S10. Distribution of stimulus parameters for dot arrays used in Experiment 3 & 4 of Yousif & Keil (2020). The red dots locate the stimulus parameters of the first array of each pair, and the blue dots locate the stimulus parameters of the second array of each pair. The grey line connects the two arrays within each pair.
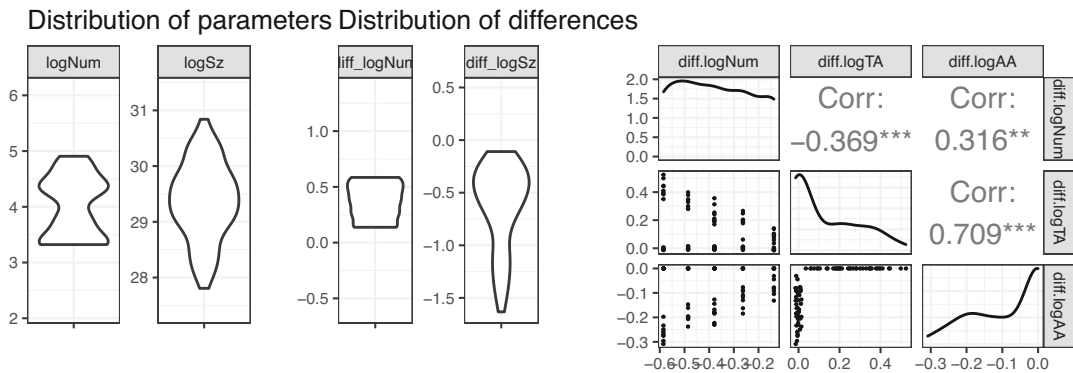

Figure S11. Distribution of log differences (or ratios) within pairs of arrays in number (*N*) and size (*Sz*), and the scatterplot matrix illustrating the correlational structure of log differences across number (*N*), total area (*TA*), and additive area (*AA*), from Experiment 3 & 4 of Yousif & Keil (2020).