Cognition 205 (2020) 104439

Contents lists available at ScienceDirect

Cognition

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

Judgments of spatial extent are fundamentally illusory: ‘Additive-area’ provides the best explanation

Sami R. Yousif*, Richard N. Aslin, Frank C. Keil

Yale University, Department of Psychology, United States of America

ARTICLE INFO

Keywords:
Additive area
Area
Perception
Density
Approximate number

ABSTRACT

How do we represent extent in our spatial world? Recent work has shown that even the simplest spatial judgments — estimates of 2D area — present challenges to our visual system. Indeed, area judgments are best accounted for by ‘additive area’ (the sum of objects' dimensions) rather than ‘true area’ (i.e., a pixel count). But is ‘additive area’ itself the right explanation — or might other models better explain the results? Here, we offer two direct and novel demonstrations that ‘additive area’ explains area judgments. First, using stimuli that are simultaneously equated for number and all other confounding dimensions, we show that area judgments are nevertheless explained by ‘additive area’. Next, we show how ‘scaling’ models of area fail to explain even basic illusions of area. By contrasting squares with diamonds (i.e., the same squares, but rotated), we show a robust tendency to perceive the diamonds as having more area — an effect that no other model of area perception would predict. These results not only confirm the fundamental role of ‘additive area’ in judgments of spatial extent, but they highlight the importance of accounting for this dimension in studies of other features (e.g., density, number) in visual perception.

1. Introduction

In some contexts we make estimates of “how many” in terms of exact or approximate number (how many pins are standing?; how many people are ahead of me?), but in other contexts we want to know “how much” in terms of area (how much of that field can I harvest before sunset?) or volume (how much of that pile of fruit can I eat?). Here we focus on area as one of the most common spatial estimation tasks we confront in our daily lives, yet one where we seem to distort our estimations in a highly predictable manner. For example, suppose you need to paint several different surfaces, and you must decide how much paint to purchase. Imagine that one surface is 20 m by 10 m that will be blue, and three surfaces are each 7 m by 3 m that will be green. To decide how much paint to buy, you could just do the math: you need $20 \times 10 = 200\ m^2$ of blue paint, and you need $(7 \times 3) \times 3 = 63\ m^2$ of green paint. Clearly, you need much more blue paint than green paint. But suppose that you let your visual system solve the same problem by estimation rather than by computation. Would this “eyeball” estimate arrive at the same answer?

Despite the ubiquity of problems like these, we know surprisingly little about how our visual system solves simple problems of area perception. When positioning furniture, when purchasing and preparing food, and when drawing diagrams or making art, we are forced to reckon with our visual system's ability to perceive space. Yet our percept of area does not reflect ‘true area’ at all (i.e., a true pixel count; Yousif & Keil, 2019; see also Carbon, 2016); instead, our percepts of area seem to reflect ‘additive area’ (the sum of the length and width of every item in the display). When observers complete area discrimination tasks for dot displays that vary in either ‘additive area’ or ‘true area’ (a true pixel count; e.g., for an array of squares, the *products* of the length and width summed over every item in the display) while controlling the other, variation in ‘additive area’ predicts essentially all of the variance in area judgments; in fact, observers were unable to discriminate displays that were equated in ‘additive area’, even when the true pixel count varied by as much as 30% (Yousif & Keil, 2019; see Fig. 1 for a graphical depiction of ‘Additive area’). The perception of area seems to be systematically biased — almost perfectly tracking variance in ‘additive area’. If you relied on your visual system alone in the paint purchasing example, you may end up purchasing equal amounts of green and blue paint — because the sums of their dimensions are equal ($20 + 10 = 30$; $[7 + 3] \times 3 = 30$), even though in reality you need more than three times as much blue paint.

That said, there are other models of area perception, most of which are roughly compatible with the ‘additive area’ view (e.g., Ekman & Junge, 1961; Nachmias, 2008, 2011; Stevens & Guirao, 1963; Teghtsoonian, 1965). One class of these models (e.g., Ekman & Junge,

* Corresponding author at: Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205, United States of America.
E-mail address: sami.yousif@yale.edu (S.R. Yousif).

https://doi.org/10.1016/j.cognition.2020.104439
Received 5 May 2020; Received in revised form 31 July 2020; Accepted 11 August 2020
Available online 16 September 2020

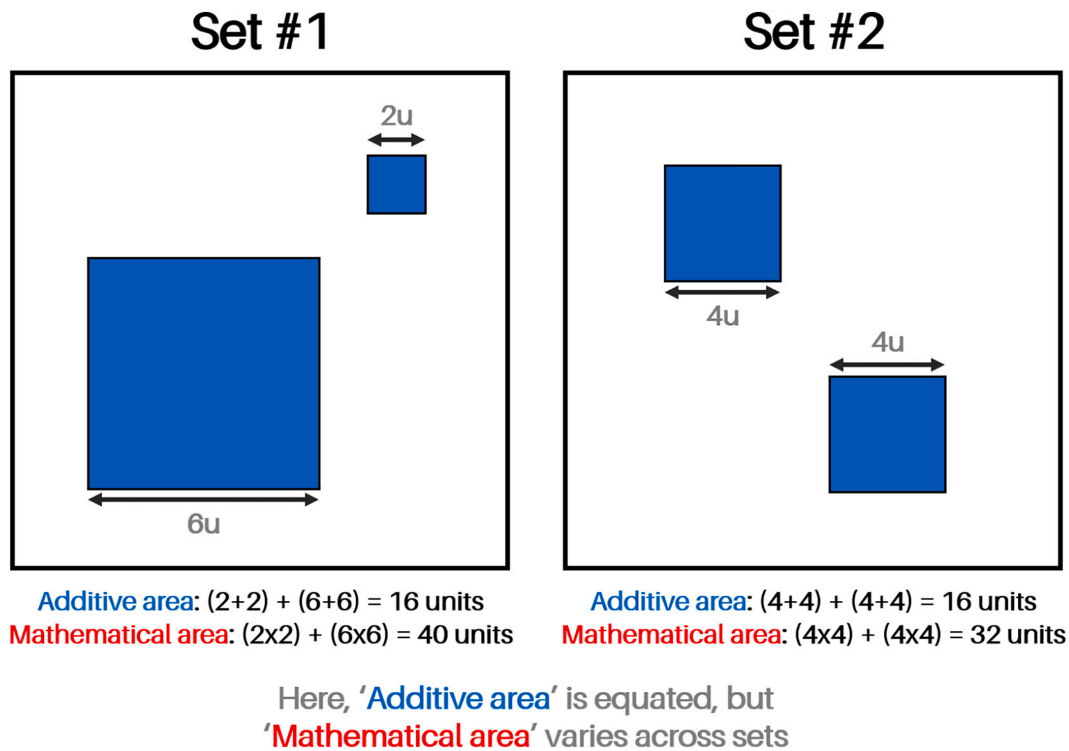

Fig. 1. A visual explanation of the relationship between 'Additive area' and 'Mathematical area'. Note that 'Mathematical area' refers to the true, objective area value; yet prior work demonstrates that 'Additive area' better captures subjective impressions of area. In this figure, we demonstrate how 'Additive Area' and 'Mathematical area' can be teased apart by varying the sizes of objects within a display.

1961; Stevens & Guirao, 1963; Teghtsoonian, 1965) which we will refer to generally as 'scaling models' holds that perceived area is equal to true area raised to the power of some number less than 1. For example, Stevens and Guirao (1963) claimed that area is scaled with an exponent of 0.7 (but there is disagreement about the exact value of this exponent; see Ekman & Junge, 1961; Teghtsoonian, 1965). In some respects, scaling models are like the 'additive area' model. Both views agree, for example, that area perception is not veridical — and that area perception will be distorted most for large shapes. Furthermore, for simple shapes (e.g., circles and squares), 'additive area' predicts the same behavior as a scaling model with an exponent of .5 (We note however at least study that has investigated area perception for irregular shapes; see Odic et al., 2013. In this work, area discriminations are shown to be ratio-dependent based on the true, mathematical area of the shapes. It is unclear, however, to what extent 'additive area' and 'mathematical area' are confounded in these cases. It is possible that 'additive area' does not apply to irregular shapes, or it is possible that the two are highly correlated in this work. In the present work, we focus only on regular shapes — circles and squares — because these are by far the most representative in the literature on quantity estimation.)

1.1. Why we must understand area perception

One reason to study area estimation is because of its relation to quantity perception more broadly. For example, much work has investigated the perception of numerosity (e.g., Anobile et al., 2016; DeWind et al., 2015; Halberda, Mazzocco, & Feigenson, 2008; Lourenco et al., 2012) and many confounded spatial dimensions such as area (see Odic et al., 2013), perimeter (sometimes also referred to as 'contour length'; Clearfield & Mix, 1999), density (the degree of compactness of objects within a space; see Durgin, 1995, 2008; Dakin, Tibber, Greenwood, & Morgan, 2011; see also Anobile et al., 2014), and convex hull (the spatial envelope of the objects within a space; Clayton et al., 2015). To understand how these cues collectively contribute to quantity estimation (see also see Gebuis & Reynvoet, 2011; Leibovich et al., 2017), we need to understand how we perceive each in isolation (see also Barth, 2008; Lourenco et al., 2012; Odic et al., 2013). Suppose, for example, you want to study the relative contribution of area and numerosity to quantity judgments. You might create stimuli that vary in number but are equated in area and vice versa — but what does it mean to equate area?

Virtually all existing work on quantity estimation rests on an unproven assumption: that *perceived* area is equal to *true* area (i.e., the actual number of pixels on the screen). If area perception is illusory (as recent work suggests; Yousif & Keil, 2019), then our understanding of the relation between area perception and quantity estimation may be confused. In fact, it has been shown that accounting for perceived area (i.e., 'additive-area') substantively changes conclusions one would draw about area, number, and quantity estimation more broadly (Yousif et al., 2019; Yousif & Keil, 2020).

Another reason we must study area perception is because illusions of area may speak to a fundamental constraint of our visual system. Indeed, one explanation for illusions like these is that we struggle to perceptually integrate multiple spatial dimensions (e.g., Carbon, 2016). Perhaps this view explains classic illusions of volume (in which a tall glass appears more voluminous than a shorter glass; e.g. Frayman & Dawson, 1981; Raghubir & Krishna, 1999) and possibly even some illusions of numerosity perception (e.g., DeWind et al., 2020). Because the way in which we perceive area has such widespread implications, it is crucial to understand the spatial features that explain judgments of area. The answer to this question speaks not only to how we spontaneously perceive quantity in our spatial world, but also directly informs ongoing debates about whether (or the extent to which) number plays a privileged role in human cognition (e.g., Leibovich et al., 2017). So: is 'additive-area' the right explanation of area perception?

1.2. Current study

The original study on AA addressed several alternative explanations that may have explained area judgments but did not address all of them directly. Therefore, it is possible that some feature other than 'additive-area' explains area judgments (or that some other model better captures the observed data). For example, in the original work, number was manipulated in an indirect manner, and number was not manipulated at all in several of the experiments (Yousif & Keil, 2019). Density and convex hull were not addressed, even though each of these dimensions has been discussed as playing a crucial role in the perception of number. To find out whether 'additive area' truly explains area judgments, here we first ask whether 'additive area' still explains variance in area judgments even when simultaneously accounting for number, density, and convex hull. Observers viewed displays that vary in either 'additive area' (henceforth, AA) or 'mathematical area' (i.e., true area; henceforth, MA) while the other dimension was equated. Across these displays, number was always set at a fixed quantity of 6 in Experiment 1a and 10 in Experiment 1b. By designing the stimuli in this way, density and convex hull are intrinsically equated on average (by virtue of number and area being simultaneously equated). We chose these dimensions because they are some of the most important in the study of number perception, and because other dimensions (e.g., perimeter) were already addressed in prior work (e.g., Yousif & Keil, 2019).

One possibility is that AA does not explain area judgments — and that the previous results were due to a confound with some unaccounted for dimension (like density or convex hull). In other words, it is possible that the original study documenting AA (Yousif & Keil, 2019) failed to properly account for all possible dimensions that could explain area judgments. If true, we should expect that observers will be able to discriminate displays that vary in MA but not in AA (once these other dimensions are accounted for). On the other hand, if AA *does* explain area judgments, then we should expect that observers will be able to discriminate displays that vary in AA but not in MA (mirroring the original AA results, but after accounting for these new dimensions). Of course, it is also possible that the truth lies somewhere in between: that observers will successfully discriminate stimuli that vary in both AA and MA, or that they will fail to discriminate stimuli that vary in both AA and MA. Even if these results are mixed (i.e., observers discriminate using both AA and MA, or neither), there would still be cause for concern: this would nevertheless mean that area judgments are fundamentally illusory.

Furthermore, the original work did not address alternative models of area perception, most notably classic 'scaling' models (e.g., Ekman & Junge, 1961; Stevens & Guirao, 1963; Teghtsoonian, 1965). Here, we test one of the scaling models' most essential predictions: that objects of equal area should be perceived as having equal area. To do so, we compare squares vs. diamonds. According to scaling models, a square and an equivalent square rotated 45 degrees (i.e., a symmetrical diamond) should be perceived as equal. But the 'additive area' perspective may predict something more interesting: that diamonds are perceived as having *more* area than equivalent squares (if the horizontal and vertical axes of shapes are prioritized; as in Li, Peterson, & Freeman, 2003; Yousif, Chen, & Scholl, 2020). In a second experiment, we test this hypothesis directly.

2. Experiment 1a: equating number (6 items)

Previous studies on AA used displays that varied in numerosity (Yousif & Keil, 2019). Although that work controlled numerosity indirectly, it failed to test for differences in AA vs. MA when numerosity was held constant. Here, we tested area perception in the same way except that all stimuli had a fixed numerosity of six (see Fig. 2A and B). Does AA still best explain area discriminations?

2.1. Method

This experiment mirrored the design of previous studies on 'additive area' (Yousif & Keil, 2019). This experiment was pre-registered, and raw data are posted on our OSF page.

2.1.1. Participants

100 observers were recruited via Amazon's Mechanical Turk, though 2 observers were excluded because they did not complete the task. All observers consented prior to participation, and these studies were approved by the IRB at Yale University.

2.1.2. Materials

All of the stimuli were generated via custom software written in Python with the PsychoPy libraries (Peirce et al., 2019). The aim was to create pairs of stimuli that varied in either AA or MA while the other was equated. Virtually all the details mirror those of the original design (see Yousif & Keil, 2019). For each stimulus pair, we randomly generated an initial set of discs and then pseudo-randomly generated a second set of objects based on a given AA ratio. The dots ranged in from 20 pixels to 120 pixels in diameter (though the exact size depends on the participants display). Unlike the original experiments, these displays *always* had exactly 6 items. Stimulus pairs were generated randomly until a pair met both the AA criterion *and* the MA criterion, at which point that pair would be rendered another time and saved. The second stimulus (i.e., the one that was pseudo-randomly generated to match the first) always had more area (whether AA or MA) than the initial stimulus. While density and convex hull were not explicitly constrained, they were intrinsically equated for the following reasons. Insofar as both area and number were equated, density must also be equated. (We note that our own work raises questions about how density ought to be equated in the first place; if perceived area does not track MA, then what about perceived density? For our purposes, we ignore this complication, accepting that either controlling MA or AA while controlling number must have accounted for density.) Similarly, the average convex hull did not vary across the two stimulus types (i.e., those controlled for AA vs. MA; and mathematically this must be the case insofar as number, density, and area are all equated).

Of note, there are a limited number of ways to de-confound AA and MA. When equating numerosity across displays, the only way to tease these two dimensions apart is to manipulate the variance in size across the items. For example, imagine some simple squares. Consider a 2×2 square and a 6×6 square. The combined MA of these shapes would be 40 ($2 \times 2 + 6 \times 6$). The combined AA of these shapes would be 16 ($2 + 2 + 6 + 6$). Now consider you have two 4×4 squares. The combined MA of these shapes would be 32 ($4 \times 4 + 4 \times 4$) but the combined AA would *also* be 16 ($4 + 4 + 4 + 4$). In this example, two displays with equal AA vary in MA. Using the same principle, it is also possible to have displays that vary in AA but are equated in MA. In other words, dissociating AA and MA in this experiment necessarily involves a difference in the variance of object sizes. Note however that this is not true when numerosity is allowed to vary across stimuli (see Yousif & Keil, 2019). For more information about how AA, MA, and number covaried, see the "Stimulus Details" files on the OSF page. The images depicted in Fig. 2A and B are representative, as they were actual images used in this experiment.

While AA was controlled, MA could vary in either a 1.10 or 1.15 ratio (and vice versa for AA while MA was controlled). As there are mathematic constraints on how much AA and MA can differ, these ratios were selected to maximize the differences between them. Because of the pseudo-random nature of stimulus creation and the mathematical constraints involved in creating such stimuli, MA was never perfectly matched with the stated ratio; it could vary ± 1%. That is, if the MA ratio for a given trial was 1.10, then we allowed the difference in MA to fluctuate between 1.09 and 1.11.

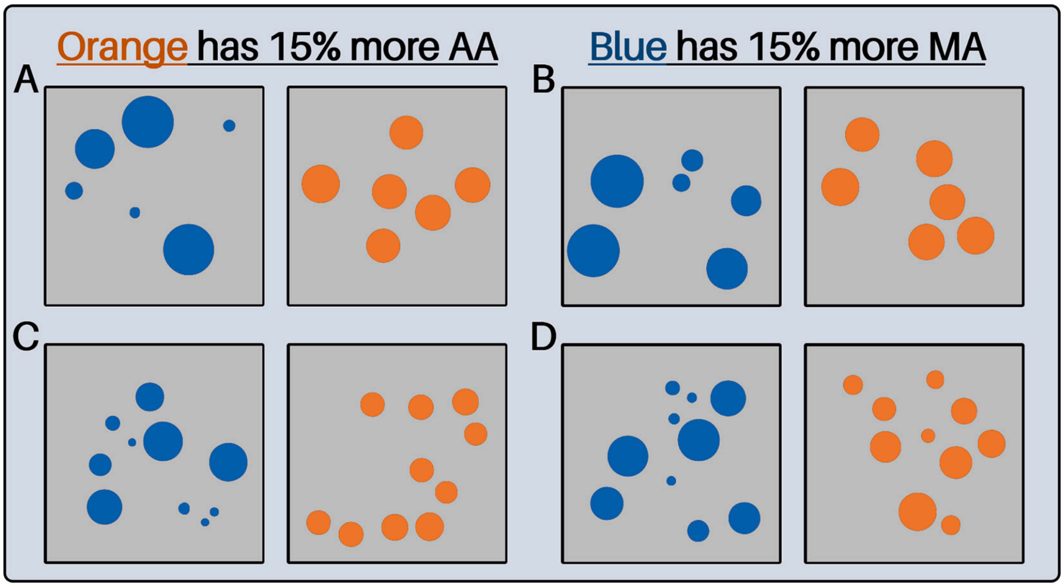

Fig. 2. Example displays from Experiment 1a (A and B) and Experiment 1b (C and D). Panels A and C depict trials in which true area is equated across the two displays. However, additive area is 15% greater in the stimuli on the right. Panels B and D depict trials in which additive area is equated across the two displays. Here, mathematical area is 15% greater in the stimuli on the left. The stimuli appear here exactly as they would have to observers in the task. Additive area in each case is equal to the sum of the objects' height and width. For circles, additive area for each shape is equal to twice the diameter.

2.1.3. Procedure

The task itself was administered online via Amazon Mechanical Turk, using custom software. On each trial, observers saw two spatially separated displays consisting of sets of blue- or orange-colored dots, presented side-by-side in the center of the screen, with 50 pixels of space in between (blue always appeared on the left, orange on the right). Each display was 400 pixels by 400 pixels. The side that contained the set with more area (either AA or MA) was counterbalanced such that half the time the left side had more cumulative area (i.e., summed over the 6 dots) and half the time the right side had more cumulative area. Observers were instructed to press ‘q’ if the image on the left had more cumulative area, and ‘p’ if the image on the right had more cumulative area. Observers were told the following: “Your task is simply to indicate which set of circles has **more cumulative area**. In other words: if you printed the images out on a sheet of paper, which would require more total ink?” The stimuli appeared for only 700 ms but there was no time limit on responses. Between each trial, there was a 1000 ms ITI. Observers completed 96 trials, 24 of each of 4 trial types (MA varying in a 1.10 or 1.15 ratio while AA was held constant; AA varying in a 1.10, 1.15 ratio while MA was held constant). All trials were presented in a unique random order for each participant. Observers completed two representative practice trials with feedback before beginning the actual task. Because over half of the trials had no objectively correct answer (because MA did not vary), we measured accuracy as a *propensity to choose ‘more’* – whether that be more AA or more MA.

2.2. Results

The results are shown in Fig. 3A and B. (Per our pre-registered exclusion criteria, trials with RTs greater than 10 s were excluded from analyses.) Observers were indeed more accurate in making discriminations on the basis of AA rather than MA. A repeated-measures ANOVA conducted on accuracy with two factors (condition: AA vs. MA; ratio: 1.10 and 1.15) revealed main effects of both condition ($F[1,97] = 4.55$; $p = .04$) and ratio ($F[1,97] = 21.98$; $p < .001$), as well an interaction between the two ($F[1,97] = 4.20$; $p = .04$). Post-hoc tests revealed that overall performance was above chance in the AA condition for both ratios (1.15: $t[97] = 6.16$, $p < .001$; $d = 0.62$; 1.10: $t[97] = 4.21$, $p < .001$; $d = 0.43$). However, observers were unable to make discriminations on the basis of MA alone, even for the larger ratio (1.15: $t[97] = 1.97$, $p = .05$; $d = 0.20$; 1.10: $t[97] = 1.26$, $p = .21$; $d = 0.13$). A separate ANOVA conducted on response times revealed a small but significant (44 ms) advantage for the AA trials ($F[1,97] = 4.82$; $p = .03$), but no effect ratio ($F[1,97] = 0.79$; $p = .38$) and no interaction ($F[1,97] = 0.98$; $p = .33$).

2.3. Discussion

These results validate and extend previous work showing that area judgments are best explained by variation in ‘additive area’. Here, we observed a robust accuracy and response time advantage for trials that varied in AA as opposed to MA — even when we directly controlled number (and consequently equated density and convex hull). And AA was not just a better predictor than MA: observers were unable to discriminate displays that differed in MA, even at the highest ratios tested. This suggests that AA is not merely one dimension correlated with area judgments, but that it is instead the singular dimension that seems to capture performance on these tasks. This pattern is noteworthy for several reasons. *First*, these results lay to rest any questions about the relation and/or interaction between AA and number, as objective numerosity is set at a fixed quantity (6) across all 96 stimuli. (We note however that *subjective* numerosity may not be equated across displays; for more on the relation between perceived area and perceived numerosity, see Yousif & Keil, 2020.) *Second*, these results speak to additional dimensions that have been discussed in the approximate number literature (see, e.g., Leibovich et al., 2017) but were not addressed in the original study (Yousif & Keil, 2019). *Third*, these results more generally suggest that the AA effect is robust across a great deal of variation in stimulus design (as the parameters here differed slightly from those used in the original study). *Fourth*, these results suggest that the effect of AA is robust even at smaller ratios; observers achieve 67% accuracy for ratio differences as small as 15%. This is important in practice because number, MA, and AA are only mathematically dissociable to about this extent — yet clearly this amount of variability is consequential. *Fifth*, the lack of an effect of MA suggests that equating this dimension — by far the most common practice in hundreds of approximate number studies — is insufficient to account for the percept

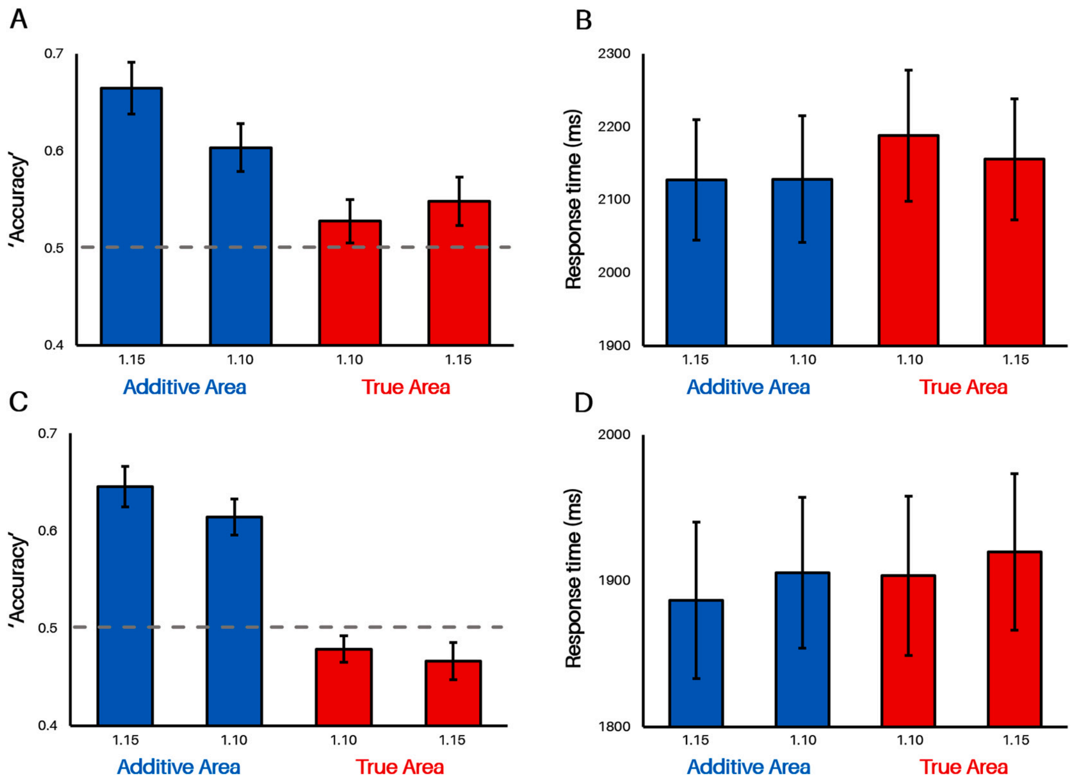

Fig. 3. Results from Experiment 1a (A and B) and Experiment 1b (C and D). Panels A and C depict the proportion of trials for which observers select the option with 'more' – whether that was more true area or more additive area — for each of the four additive area/true area ratios tested. The dashed lined represents at-chance performance. Panels B and D depict response times for each of the seven ratios tested. In all graphs, the x-axis represents the ratio. While additive area varied, true area remained constant. While true area varied, additive area remained constant. Thus, blue correspond to additive area trials, red bars correspond to true area trials. Error bars represent ± 1 SE. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of spatial extent gathered in area judgments. *Finally*, these results raise the prospect that perceived area (here operationalized as AA) is a confound in prior studies on approximate number; indeed, this confound may actually explain some well-known results (as demonstrated in Yousif & Keil, 2020).

3. Experiment 1b: equating number (10 items)

Experiment 1a used stimuli with a relatively small number of dots (6 per display). This number of items is smaller than many studies on number perception, and is close to subitizing range (in which their numerosity would be automatically rather than approximately perceived; e.g., Kaufman et al., 1949). Here, we directly replicated the findings of Experiment 1a except with slightly more numerous displays (i.e., 10 dots per display as opposed to 6).

3.1. Method

This experiment was identical to Experiment 1a except as otherwise noted. 100 new observers participated (1 excluded for failing to complete the task). To accommodate the increased number of dots, the minimum and maximum dot sizes were decreased; as a result, the dots ranged from 15 to 95 pixels in diameter.

3.2. Results

The results are shown in Fig. 3C and D. Observers were indeed more accurate in making discriminations on the basis of AA rather than MA. A repeated-measures ANOVA conducted on accuracy with two factors (condition: AA vs. MA; ratio: 1.10 and 1.15) revealed a main effect of condition ($F[1,98] = 27.63$; $p < .001$) but not ratio ($F[1,98] = 0.84$; $p = .36$), as well an interaction between the two ($F[1,98] = 4.09$; $p = .046$). Post-hoc tests revealed that overall performance was above chance in the AA condition for both ratios (1.15: $t[98] = 7.00$, $p < .001$; $d = 0.70$; 1.10: $t[98] = 6.20$, $p < .001$; $d = 0.62$). However, observers were unable to make discriminations on the basis of MA alone, even for the larger ratio (1.15: $t[98] = 1.79$, $p = .08$; $d = 0.18$; 1.10: $t[97] = 1.59$, $p = .12$; $d = 0.16$). A separate ANOVA conducted on response times revealed no significant differences (condition: $F[1,98] = 1.94$; $p = .17$; ratio: $F[1,98] = 0.01$; $p = .91$; interaction: $F[1,98] = 2.59$; $p = .11$).

3.3. Discussion

This experiment replicates the findings of Experiment 1a, demonstrating once again the AA and not MA best captures area judgments. Simultaneously, the increased number of dots in each stimulus ensures that the present effects cannot be explained by the stimuli's proximity to subitizing range.

4. Experiment 2: squares vs. diamonds

In the previous experiments, we compared AA against 'true', or 'mathematical' area. However, this may be an unfair comparison. We have long known that area perception is not veridical. Originally proposed over fifty years ago, 'scaling' models of area perception suggest that perceived area is equal to 'true' area scaled with an exponent of ~0.7 (e.g., Ekman & Junge, 1961; Stevens & Guirao, 1963; Teghtsoonian, 1965). In some ways, AA and scaling models make similar predictions. For example, both models predict that one shape with twice as much area as another will be *perceived* as having less than that amount. However, these models differ in how they try to explain this non-veridical effect. We propose that this distortion occurs because the visual system is unable to integrate the horizontal and vertical dimensions; in contrast, scaling models offer no mechanistic account of area perception. So, here, we test what should be one straightforward prediction of scaling models, regardless of the precise value of the exponent: that two shapes of equal area are perceived as being equal. To test this hypothesis, we compare squares vs. symmetrical diamonds (i.e., rotated squares). However, mirroring the design of the previous study, we still use sets of shapes rather than individual shapes. Although we do not hold numerosity constant across displays, numerosity is held constant across conditions. If scaling models are correct, then we should not expect that the perceived area of squares and diamonds differs. Therefore, if we *do* observe a difference between diamonds and squares, we have reason to doubt such models — and more reason to embrace models such as AA that can explain such illusions.

4.1. Method

This experiment was identical to Experiment 1 except as otherwise noted. 100 new observers participated. 1 observer was excluded for failing to complete the task. Unlike the previous experiment, we used squares instead of circles (see Fig. 4). This is solely because squares can be rotated, creating objects (i.e., diamonds) with a different vertical and horizontal extent that are otherwise equal in MA. Unlike the previous experiment, AA and MA differed in two slightly larger ratios (1.15 and 1.25 vs. 1.10 and 1.15); there were also trials that varied in neither AA nor MA. These trials in which neither MA nor AA varied were included as a clean test case for whether diamonds are perceived as having more area than squares. Note that we calculated AA *as if* the shapes were squares. In other words, all AA calculations assume that the shapes were in the same canonical orientation; we rotated the squares only after these variables were calculated for each display.

Trials were divided into two distinct types: ones in which the squares had 'more' (whether more AA or more MA) and ones in which the diamonds had 'more' (whether more AA or more MA). Every comparison was between a stimulus array with all squares and a stimulus array with all diamonds. There were 40 unique trials of each trial type (5 different ratios × 8 instances of each ratio), resulting in a total of 80 trials. Note that the parameters of the diamonds-more trials and the squares-more trials were also equated; i.e., number and other

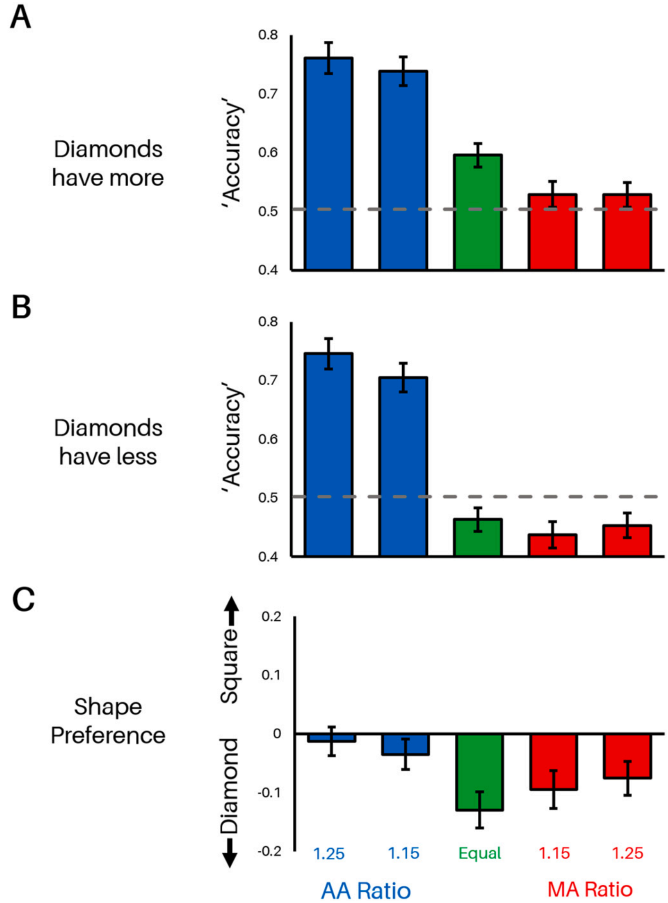

Fig. 5. Results from Experiment 2. The proportion of trials for which observers select the option with 'more' – whether that was more true area or more additive area — for each of the five additive area/true area ratios tested. The data here are broken down by trials in which diamonds had more (A), squares had more (B), and the difference scores between these two trial types (C). In (C), note that values below the x-axis correspond to a diamond preference. The dashed lined represents at-chance performance.

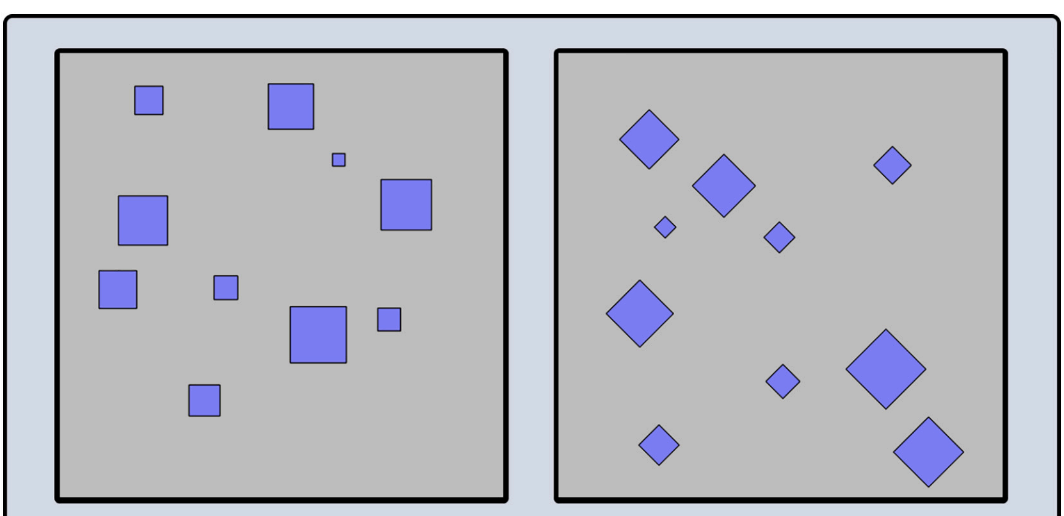

Fig. 4. An example display from Experiment 2. Here, both displays have equal additive area and equal mathematical area (if we calculate additive area as if all objects were squares).

stimulus dimensions should be equated across these trial types, and the only difference across the two trial types is whether the squares were rotated.

4.2. Results

The results are shown in Fig. 5. First, we separately analyze the trials in which diamonds had more area and trials in which squares had more area. The basic AA effect replicates for both trial types. In both cases, observers were more likely to choose displays which had more AA than MA (collapsing across ratio; diamonds-more: $t(98) = 7.64$, $p < .001$, $d = 0.77$; squares-more: $t(98) = 8.21$, $p < .001$, $d = 0.83$). For the trials in which diamonds had more area, observers were not above chance selecting the display with true area ($t(98) = 1.57$, $p = .12$, $d = 0.16$) even at the higher ratio (1.25; $t(98) = 1.51$, $p = .13$, $d = 0.15$). For the trials in which squares had more area, observers were actually below chance selecting the display with true area (collapsed across ratios; $t(98) = 2.41$, $p = .018$, $d = 0.24$). The reason for this below-chance performance is that observers had a tendency to select diamonds for all the area ratios we tested; collapsed across ratios, there was a significant tendency to choose the stimulus with diamonds instead of squares ($t(98) = 3.31$, $p = .001$, $d = 0.33$). Notably, this difference was most pronounced for trials in which both AA and MA were equal ($t(98) = 4.29$, $p < .001$, $d = 0.43$); these trials therefore offer a clear test case of this diamond preference.

4.3. Discussion

This experiment put one of the critical predictions of 'scaling' models of area perception to the test: that shapes of equal area should be perceived as equal. Surprisingly, even a simple manipulation like rotating a square 45 degrees is sufficient to induce a relatively large illusion of area; observers were more than 10% more likely to indicate that the diamonds had more area than the squares when the two were equal. This pattern of results makes sense in light of the AA model. The critical insight of this perspective is that the perceptual system is independently perceiving (and summing) the spatial dimensions of an object. But how does the visual system decide which dimensions should be added together? The purpose of this paper is not to definitively answer this question. However, we expected that the vertical and horizontal axes may be prioritized, as in other known illusions of space (oblique effect; Li et al., 2003; Yousif, Chen, & Scholl, 2020). If true, we may expect that diamonds would be perceived as larger than equivalent squares (given that the diagonal is longer than the side length) — and that is exactly what we find.

However, the purpose of this experiment was not to test a positive prediction of the AA model, but to test a prediction that scaling models of area perception must make (i.e., that squares and diamonds should be perceived as equal in area). At the least, these results demonstrate that the visual system's percept of area is easily fooled, even by exceedingly simple manipulations. Once again, these findings are consistent with AA but inconsistent with any other current model of area perception (to our knowledge), including scaling models.

5. General discussion

In two experiments, 'additive area' was uniquely capable of explaining perceived area. In Experiment 1, we showed that AA explains area judgments even when we carefully control other stimulus dimensions (like number, density, and convex hull). These findings are especially important for understanding how area perception relates to number perception (a relationship that, in our view, has been misunderstood by failing to account for perceived area). In Experiment 2, we tackled the issue of area perception more directly by testing one of the critical predictions of 'scaling' models of area. We showed that merely rotating squares 45 degrees dramatically increases their perceived area. While these findings do not prove that AA is the only possible model, they cast doubt on the very premise of scaling models. In other words, we think these results collectively demonstrate that AA is the only current model capable of explaining the observed behavior. Given this — and the sheer magnitude of the illusion, in this work and elsewhere — we believe AA merits further inquiry (see also Yousif et al., under review; Yousif & Keil, 2020).

5.1. Other models of area perception

In our first experiment, we compared the AA model against the most obvious alternative a priori: that area perception is veridical. The data here and in other work suggest rather conclusively that AA captures area judgments *better* than at least this one alternative model (i.e., true, mathematical area). In principle, though, there could be an infinite number of possible models of area perception; how do we know that this one is the right one? One of the key aspects of our design is that we not only compare AA vs. MA but that we also control AA directly. In other words, we have trials that vary along other dimensions but are held constant for AA. Our view makes a strong, specific prediction: so long as AA does not vary, area discrimination performance should not vary — and that is exactly what we find. In this way, we have reason to believe that AA is not merely a model that performs well, but that it very closely approximates our true percepts.

Nevertheless, it is useful to compare the AA model against other models that have been proposed. In our second experiment, we specifically addressed 'scaling' models of area perception (e.g., Ekman & Junge, 1961; Stevens & Guirao, 1963; Teghtsoonian, 1965). In other words, if the true area of a stimulus is X, the *perceived* area of a stimulus is X raised to the power of 0.7. This view is generally consistent with ours: we *also* propose that perceived area scales non-linearly with true area. Therefore, this becomes a question of which model is correct at the margin (as they will often predict similar behavior). While we have already addressed these models empirically, let us also approach them in principle. One issue with scaling models is that they are based on area judgments of single objects; it remains unclear how these models should be applied to sets of objects. For example, scaling models fail to explain exactly how scaling occurs (i.e., over what *units* scaling operates; items vs. sets). If we assume that area is scaled for *each* item, then scaling models are largely indistinguishable from the AA model, which is essentially equivalent to a scaling model with an exponent of 0.5 (although this is not true for all shapes; these two models could be dissociated for rectangles, for example — or by using rotated shapes as we do in Experiment 2). If we instead assume that scaling happens over the entire set, then we can look to the present data for answers. If perceived area was equal to X raised to the power of 0.7, for example, then we should nevertheless predict above-chance performance on trials when AA is controlled and MA varies. We should also predict at-chance performance when AA varies and MA is controlled. Neither of these things proved to be true.

Furthermore, from a purely computational perspective, AA is a more efficient method than a scaling model with an exponent of 0.5. Whereas the scaling model would involve at least three distinct steps of computation (multiplication, scaling, addition), the AA model involves only one (addition). In this way, the AA model is also consistent with some basic perceptual illusions (see Experiment 2 here; see also Carbon, 2016). We believe that the starting point for understanding area perception should be to consider the visual system's inability to integrate multiple dimensions — and to consider the most basic mathematical operations that can be performed over those dimensions. Because the AA model is better equipped to explain perceptual illusions like the 'folded paper size illusion' (an illusion we encourage readers to experience for themselves!), we believe the AA model is more viable than both 'true area' models and 'scaling models'.

Finally, suppose that perceived area is equal to true area scaled to some exponent and we make no predictions about what that exponent should be. Virtually any data could be 'explained' by such models. The challenge, of course, is that such models fail to account for basic illusions of area/volume perception (as Experiment 2 demonstrates rather plainly). But the debate between these models is not so much about the answer as it is about *how we arrive there*. In other words, we care about the mechanism by which we perceive area, regardless of how we can best mathematically capture behavior. Our work emphasizes the perceptual system's inability to properly integrate multiple dimensions (as opposed to merely 'counting pixels' or 'scaling' the visual input; see also Carbon, 2016). In this way, we think the distinction between these two models speaks to a fundamental constraint on our visual system.

In short, 'additive area' uniquely accounts for impressions of visual area — outperforming other models of area perception, even when we increasingly constrain the stimulus spaces to minimize the influence of other dimensions. These findings further emphasize the importance of AA when studying both area and number perception. More generally, the results here speak to a fundamental question: how is it that we (mis) perceive our spatial world? These results strongly suggest that 'additive area' may be a definitive — and valuable — model of area perception.

Data

The data for this project are available here: https://osf.io/njmkg/.

Author note

For helpful conversation and/or comments on previous drafts, we thank Justin Halberda, Frank Durgin, Brynn Sherman, and the members of the Yale Cognition and Development Laboratory. This project was funded by an NSF Graduate Research Fellowship awarded to SRY.

Open practices statement

This experiment was pre-registered. All materials and raw data are available on our OSF page: https://osf.io/njmkg/.

CRediT authorship contribution statement

Sami R. Yousif: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **Richard N. Aslin**: Conceptualization, Writing - review & editing, Supervision. **Frank C. Keil**: Conceptualization, Writing - review & editing, Supervision.

References

Anobile, G., Cicchini, G. M., & Burr, D. C. (2014). Separate mechanisms for perception of numerosity and density. *Psychological Science, 25*, 265–270.
Anobile, G., Cicchini, G. M., & Burr, D. C. (2016). Number as a primary perceptual attribute: A review. *Perception, 45*, 5–31.
Barth, H. C. (2008). Judgments of discrete and continuous quantity: An illusory Stroop effect. *Cognition, 109*, 251–266.
Carbon, C.-C. (2016). The folded paper size illusion: Evidence of inability to perceptually integrate more than one geometrical dimension. *i-Perception, 7*(4), https://doi.org/10.1177/2041669516658048.
Clayton, S., Gilmore, C., & Inglis, M. (2015). Dot comparison stimuli are not all alike: The effect of different visual controls on ANS measurement. *Acta Psychologica, 161*, 177–184.
Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science, 10*, 408–411.
Dakin, S. C., Tibber, M. S., Greenwood, J. A., & Morgan, M. J (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences, 108*, 19552–19557.
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.
DeWind, N. K., Bonner, M. F., & Brannon, E. M. (2020). Similarly oriented objects appear more numerous. *Journal of Vision, 20*, 1–11.
Durgin, F. H. (1995). Texture density adaptation and the perceived numerosity and distribution of texture. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 149.
Durgin, F. H. (2008). Texture density adaptation and visual number revisited. *Current Biology, 18*, R855–R856.
Ekman, G., & Junge, K. (1961). Psychophysical relations in visual perception of length, area and volume. *Scandinavian Journal of Psychology, 2*, 1–10.
Frayman, B. J., & Dawson, W. E. (1981). The effect of object shape and mode of presentation on judgments of apparent volume. *Perception & Psychophysics, 29*, 56–62.
Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods, 43*, 981–986.
Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in nonverbal number acuity correlate with maths achievement. *Nature, 455*, 665–668.
Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology, 62*, 498–525.
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. *Behavioral and Brain Sciences, 40*, 1–62. https://doi.org/10.1017/S0140525X16000960.
Li, B., Peterson, M. R., & Freeman, R. D (2003). Oblique effect: a neural basis in the visual cortex. *Journal of Neurophysiology, 90*, 204–217.
Lourenco, S. F., Bonny, J. W., Fernandez, E. P., & Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proceedings of the National Academy of Sciences, 109*, 18737–18742.
Nachmias, J. (2008). Judging spatial properties of simple figures. *Vision Research, 48*, 1290–1296.
Nachmias, J. (2011). Shape and size discrimination compared. *Vision Research, 51*, 400–407.
Odic, D., Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Developmental change in the acuity of approximate number and area representations. *Developmental Psychology, 49*, 1103–1112.
Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203.
Raghubir, P., & Krishna, A. (1999). Vital dimensions in volume perception: Can the eye fool the stomach? *Journal of Marketing Research, 36*, 313–326.
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*, 177.
Teghtsoonian, M. (1965). The judgment of size. *The American Journal of Psychology, 78*, 392–402.
Yousif, S. R., Alexandrov, E., Bennette, E., Aslin, R., & Keil, F. C. (2020a). *Children estimate area using an "Additive-Area Heuristic".* (under review).
Yousif, S. R., Alexandrov, E., Bennette, E., & Keil, F. C. (2019). Perceived area plays a dominant role in visual quantity estimation. *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 1241–1246). Montreal, QB: Cognitive Science Society.
Yousif, S. R., Chen, Y. C., & Scholl, B. J. (2020b). Systematic angular biases in the representation of visual space. *Attention, Perception & Psychophysics, 82*, 3124–3143.
Yousif, S. R., & Keil, F. C. (2019). The additive area heuristic: An efficient but illusory means of visual area approximation. *Psychological Science, 30*, 495–503.
Yousif, S. R., & Keil, F. C (2020). Area, not number, dominates estimates of visual quantities. *Scientific Reports, 10*, 1–13.