AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

© 2026 American Psychological Association
ISSN: 0096-3445

2026, Vol. 155, No. 8, 1985–2000
https://doi.org/10.1037/xge0001973

Spontaneous Perception of Visuospatial Ensembles

Gabriel Waterhouse[1] and Sami Yousif[2]

[1] Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill
[2] Department of Psychology, Ohio State University

Look out at a forest. What do you see? You see visual masses of green; you see large numbers of leaves; you see the various sizes of the many different trees. Is one of those features more salient than the others? According to a popular view in cognitive science, number is perceptually prioritized over other visuospatial ensemble properties like density and size, such that when you look out at a forest, what you see first and foremost is the number of trees (or trunks, or leaves) rather than their size or their density, etc. The problem with these arguments is that they mostly hinge on comparisons to only one other dimension: cumulative area. Here, we ask a very simple question: Might there be some other dimension that is prioritized over number? Across five unique paradigms and 13 experiments with over 2,000 participants, we show that human adults prioritize average size over number, such that they are more likely to (a) categorize, (b) explicitly match, (c) notice changes in, and (d) identify extreme instances of that dimension. We additionally show that children behave the same way. For this reason, we argue against the universal salience of number, noting that these "competitive" approaches (i.e., in which dimensions are compared against each other) are often context-dependent.

Public Significance Statement

Humans spontaneously extract summary information from groups of objects, such as how many items are present or how large they are on average. A common argument in theories of visual and numerical cognition is that number holds a privileged status among such ensemble representations, in that it is more readily perceived than other summary statistics. Across more than a dozen experiments spanning multiple paradigms and populations, we find robust evidence against this view. Observers consistently prioritized average size over number, even when both dimensions were simultaneously available and when genuine ensemble processing was required. These findings challenge accounts in which number is automatically or obligatorily extracted from visual scenes and instead suggest that ensemble representations are flexibly weighted in a context-dependent manner. More broadly, our results call into question strong claims about the perceptual origins of numerical knowledge while providing new insight into the nature of ensemble processing.

Keywords: ensemble perception, number perception, number bias, size perception

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

We rapidly extract all sorts of information from the visual world around us. Crucially, the visual system represents not just properties of objects and scenes but also properties of collections or *ensembles*. That is: When looking out at a forest, you perceive not only the presence of trees but also the approximate number of them, their average height, the average hue of their leaves, and so on. Yet there is a long-standing idea that certain visual ensembles are more salient than others—that, among other attributes like area and density, *number* is prioritized such that it is more readily and more accurately perceived and represented (see, e.g., Cicchini et al., 2016; Cordes & Brannon, 2008; Ferrigno et al., 2017; Tomlinson et al., 2020; see Figure 1A). Its putative uniqueness has led to number being studied more extensively than other quantity dimensions, and with much greater theoretical emphasis (see, e.g., Clarke & Beck, 2022; Feigenson et al., 2004; Leibovich et al., 2017). Indeed, some have claimed that the supposed prioritization of number in perception is a

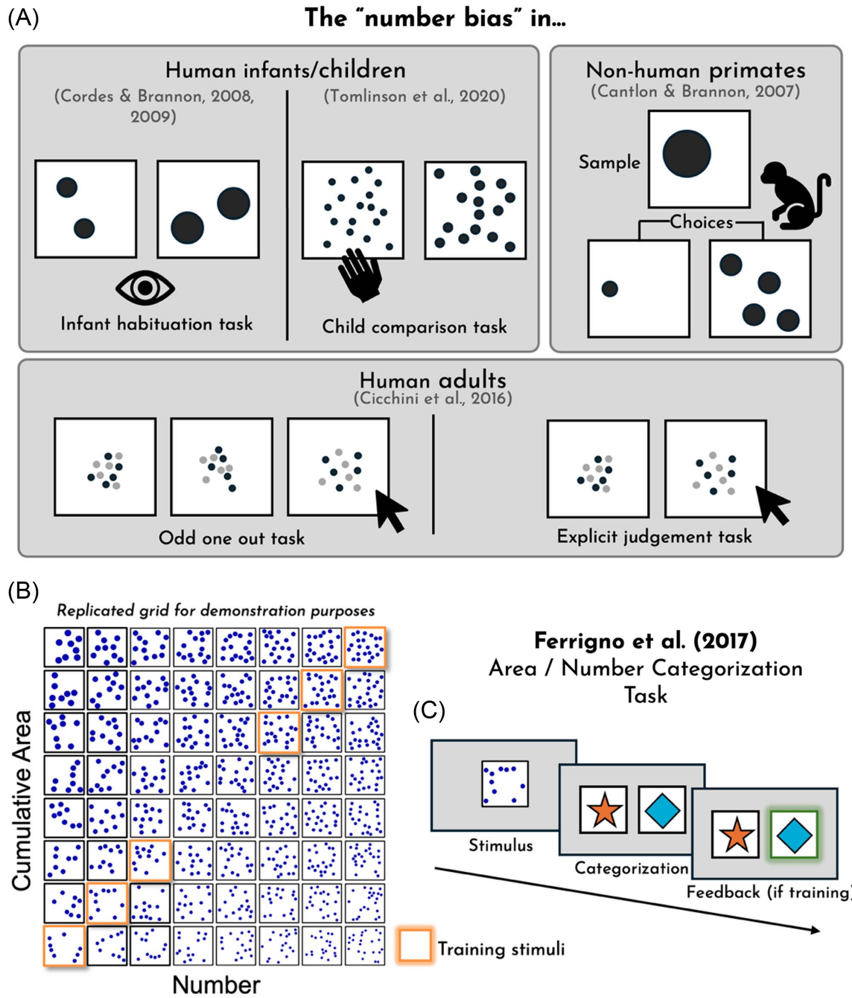
Timothy Vickery served as action editor.
Gabriel Waterhouse https://orcid.org/0009-0007-8997-3261
All stimuli used in these experiments, as well as the raw data, can be found on the Open Science Framework at https://osf.io/t7dsv/.

The authors thank the members of the Spatial Cognition Lab and the Moments Lab at Ohio State for their feedback and helpful discussion on this article and Tal Boger for helpful discussion that led to the second set of experiments. The authors also thank Stephen Ferrigno for thoughtful feedback on an earlier draft of the article and Jason Haberman for helpful comments during review.

Gabriel Waterhouse played a lead role in data curation, formal analysis, investigation, and visualization and an equal role in conceptualization, writing–original draft, and writing–review and editing. Sami Yousif played a lead role in funding acquisition, project administration, and supervision and an equal role in conceptualization, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Gabriel Waterhouse, Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, 400 Carraway Crossing, Chapel Hill, NC 27516, United States, or Sami Yousif, Department of Psychology, Ohio State University, 1835 Neil Avenue, Columbus, OH 43210, United States. Email: gwaterho@unc.edu or yousif.36@osu.edu

Figure 1
Evidence in Favor of a Number Bias



Note. (A) Evidence supporting a universal number bias comes from research on human adults, children, and nonhuman primates. (B) A replication of the stimulus space used by Ferrigno et al. (2017). Note that in the training stimuli, although number and cumulative area are both increasing, it appears as if cumulative area is only increasing *because* of an increase in number. (C) Example trial protocol for the categorization task. Participants are presented with a stimulus and then with two symbols, which they have learned to associate with either large or small quantities. During the training phase, participants receive feedback on whether they responded correctly. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

root cause of its cognitive and cultural salience, thus directly leading to the development of advanced mathematical knowledge (Ferrigno et al., 2017). Here, we question this orthodoxy: We present evidence from 13 experiments demonstrating not only that numerical information is not the most salient feature of an ensemble but also that at least one other dimension (average size) is consistently prioritized by both adults and children.

Number as a Unique Dimension

Number perception is commonly thought to arise from a universal "number sense" (Dehaene, 1997) that allows for the discrimination and estimation of sets of objects—universal in the sense that it is present early in life and throughout the animal kingdom. Current theories posit an evolutionarily ancient "approximate number system," which relates to both numerical perception and numerical cognition (see Feigenson et al., 2004; Halberda et al., 2008). This capacity for numerical representation is thought to be present in both human infants (Wynn, 1992, 1995a, 1995b; Xu & Spelke, 2000; Xu et al., 2005) and nonhuman primates (Cantlon & Brannon, 2007), suggesting that whatever system is responsible for that capacity has deep phylogenetic roots.

Though number is often studied in tandem with other visual ensemble attributes like area and density, it has received special attention relative to those dimensions. For example, number is considered to be its own domain of "core knowledge" (Feigenson

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

et al., 2004; Spelke, 2022; Spelke & Kinzler, 2007), whereas there has never been mention of core knowledge of density or size. As well, much attention has been paid to the fact that number is perceived directly (see, e.g., Burr & Ross, 2008; Durgin, 2008, but see also Negen, 2025), whereas area and density perception are essentially taken for granted. Why is that? What makes number so special, while discoveries about area and density are treated as interesting only insofar as they bear on debates about number (see, e.g., Durgin, 2008; Hurewitz et al., 2006; Mix et al., 2002)?

Perhaps number is unique because perceptual number acuity is related to mathematical ability (Halberda et al., 2008). Individual differences in number discrimination predict differences in achievement in school mathematics, suggesting a direct link between an "evolutionarily ancient, unlearned approximate number sense" and more abstract mathematical knowledge (Halberda et al., 2008, p. 665; for review, see Feigenson et al., 2013). By studying the perception of number, then, we might gain some critical insights about the "core" abilities underlying our more complex cognitive capacities. Then again, similar correlations have been observed between perceptual area acuity and mathematics achievement (Lourenco et al., 2012), yet little attention has been paid to whether there exists an ancient system of size representation that underlies uniquely human achievements in geometry.

Despite its special status, empirically isolating number is an infamously challenging task (see, e.g., Leibovich et al., 2017). In any given ensemble, numerical information is inevitably correlated with other spatial dimensions (see Leibovich & Henik, 2013). As a result, some have argued that number is only perceived indirectly via other continuous spatial magnitudes (see Gebuis & Reynvoet, 2012a, 2012b, 2014; Negen, 2025). Even if number could be mathematically isolated from other spatial dimensions (for a respectable attempt, see DeWind et al., 2015), other tricky measurement issues arise: For instance, should number be compared against *physical* area, or *perceived* area (Yousif & Keil, 2019, 2021)? Because of complications like these, it is challenging to clearly dissociate number from other spatial properties in any given visual display.[1]

Perhaps for this reason, there has been an interest in comparing the relative acuity and salience of number and other ensemble properties (see, e.g., Aulet & Lourenco, 2021; Brannon et al., 2004; Clearfield & Mix, 1999; Ferrigno et al., 2017; Hurewitz et al., 2006; Kerjean et al., 2026; Mix et al., 2002; Odic et al., 2013; Tomlinson et al., 2020; Yousif & Keil, 2020). If number cannot be isolated from all other dimensions, it can at least be measured against them. Despite nearly three decades of debate on the topic, however, there is no obvious winner: Some early work suggested that infants discriminate number using contour length (Clearfield & Mix, 1999; Mix et al., 2002), but that claim was questioned by several later studies (Brannon et al., 2004; Xu et al., 2005). Some work has shown that children perceive cumulative area with greater acuity than number (Odic et al., 2013); other work has shown that children find number to be the more salient dimension (Tomlinson et al., 2020), and yet other work has shown the latter number-salience bias is explained by a failure to account for perceived area (Aulet & Lourenco, 2021; see also Yousif et al., 2022). Despite these ambiguities, perceptual sensitivity to number is (sometimes tacitly, sometimes more explicitly) taken much more seriously than sensitivity to related visuospatial ensembles.

Is Number Truly Special?

In this ocean of mixed results, there is one study that stands out as a particularly compelling demonstration of a genuine number bias. Ferrigno et al. (2017) demonstrated that adults across multiple cultures, children, and nonhuman primates all preferentially categorized visual stimuli based on their number rather than their cumulative area. In their paradigm, participants were trained on stimuli that were orthogonal in cumulative surface area and number and learned to associate a certain symbol (either a star or a diamond) as representative of "more" or "less" (see Figure 1B–1C). After being trained, participants were shown stimuli varying in cumulative size and number. Across all populations, there was a preference to categorize based on the *number* of the displays, rather than cumulative size. Ferrigno and colleagues take this as evidence of a "universal number bias," wherein the perceived number of an ensemble is spontaneously prioritized over other spatial properties in perception. These results are compelling because (a) the study is carefully designed and the stimuli are well-controlled, (b) the task relies minimally on linguistic demands, and (c) there is a clear pattern of number bias that emerges across four distinct, diverse populations. This work represents the most persuasive evidence of a number bias to date. But is it enough to warrant the strong claims that have been made about the uniqueness and universality of number?

Notably, Ferrigno et al. (2017) compared number against only *one* other ensemble property. This choice is understandable insofar as area is the dimension most often compared against number. Yet this leaves open the possibility that there might be some other visual property that could be prioritized over number. What would be made of the fact that number is prioritized over *some* but not *all* other ensemble properties?

They conclude their article by speculating about the mechanistic role that the perceptual prioritization of number may have played in the cultural development of counting systems: "A fundamental bias to segregate numerical information from other quantitative dimensions and *use* it preferentially was likely an important catalyst for the emergence of an abstract, discrete counting system in human cultural evolution" (p. 8). Provocative as this argument may be, it only works if number is truly prioritized against all other dimensions. Any one demonstration to the contrary would undermine the theoretical weight of the conclusions. This is not to say that an innate system for number perception could not have played a role in the development of mathematical knowledge (we find this possibility quite likely; see Halberda et al., 2008), but, rather, that such a claim could not be supported by appealing to a perceptual prioritization of numerical information.

So, what dimension might be prioritized over number? Most work of this sort compares number against cumulative size. Here, we take

[1] In spite of these physical constraints, there are two established ways that number can be convincingly isolated. The first is cross-modal studies: Effects of number that extend across visual and auditory modalities, for instance, are unlikely to be explained by a confound in any one modality (e.g., Arrighi et al., 2014; Barth et al., 2005). The second is via a certain class of visual illusions wherein the *addition* of information to a display results in a clear *reduction* in perceived number (Franconeri et al., 2009; He et al., 2009; Waterhouse & Yousif, 2025). Such illusions typically escape concerns about confounds with spatial dimensions because despite adding visual information to a display (and thus increasing area and density, e.g.), perceived number is reduced. That is: Such illusions cannot be straightforwardly explained by positive congruencies with spatial dimensions.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

a different approach. Instead of comparing against *cumulative* size, we compare number against *average* size. If at first this seems like a subtle distinction, take a moment to carefully inspect the stimulus space used by Ferrigno et al. (2017; recreated in Figure 1B). Because number is pitted against cumulative area, note what happens along the main diagonal: As you move from left to right (bottom to top), both dimensions increase, but cumulative area only increases *because* number increases. The average size of the dots remains the same. Even just glancing along that diagonal, one has the clear impression that size, in some meaningful sense, is held constant. This is because in their stimulus space the relationship between the two dimensions is asymmetric: An increase in number necessarily leads to an increase in cumulative area, yet an increase in cumulative area does not necessarily lead to an increase in number. Because number and area increase at the same rate, changes in *both* dimensions may be—for whatever reason—attributed to a change in only one of those dimensions. So it may be that people notice the change in numerical information not because they perceive that information more readily, but perhaps because, as a discrete dimension, number is taken to be the underlying dimension that is altered. After all, objects tend not to spontaneously change size, but the number of items in a set changes all the time (i.e., by manual intervention). As a result, perhaps participants learn during training that number is the relevant dimension (because, in a certain light, number is the only dimension that changes). Critically, however, this would mean that if the quantity information was presented in a different way (i.e., such that the numerical information was not perfectly confounded with the area information), this bias might disappear.

Independent of the above concern, there is good reason to believe that average size is a meaningful perceptual dimension worthy of investigation unto itself. While *cumulative area* has not been studied extensively in vision science (but see Pisu et al., 2025; Yousif & Keil, 2019; Yousif et al., 2020), *average size* has been. There is consensus that people readily and accurately extract the average size of items in a display (see, e.g., Albrecht & Scholl, 2010; Allik et al., 2013; Ariely, 2001; Chong & Treisman, 2003, 2005; Raidvee et al., 2020; Sweeny et al., 2015). Further, since average size is a summary statistic over another quantity dimension (size), it can be considered a kind of quantity dimension, comparable to area or density. Even if it is not a quantity dimension, per se, it is still a meaningful variable that should, in principle, be dominated by number under a number bias account. We are not the first to compare these dimensions in this way. Recent work has shown that pigeons and humans alike find average size to be a salient quantity dimension (Diaz & Wasserman, 2026).

The focus on *average* size rather than *cumulative* size meaningfully alters the nature of the task. In our stimulus space (see Figure 2A), there is a much clearer sense in which both size and number are changing across the training items. In theory, this means that participants would then have a real choice to make between the two dimensions when they go on to categorize the off-diagonals at test. Is number just as salient in this revised stimulus space?

The Present Study

To foreshadow, we found that not only was number *not* prioritized over average size, but that the reverse was true: Average size was consistently prioritized over number—not just in a categorization task, but in five distinct paradigms. In a first experiment, we showed that altering the categorization paradigm from Ferrigno et al. (2017) to focus on number versus average size revealed a decisive average size bias. We then showed that this bias extended to a variety of other tasks. In three variations of a second experiment, we found that when shown an entire stimulus space and asked to divide it into distinct groups (thus eliminating concern about changes in number "causing" changes in area), people preferentially grouped by average size, nevertheless. In a third experiment, we showed that participants preferentially *match* items based on average size rather than number. In four variations of a fourth experiment, we showed that participants preferentially identify oddball stimuli in the same way—even when we controlled the displays in a variety of different ways. In a fifth experiment, we showed that participants were also better at detecting outlier displays based on their average size rather than their number. Finally, we show that children exhibited the same average size bias. Collectively, these results undermine strong claims made about "universal" number biases in the perception of ensembles.

Experiments 1a–1c: Categorization Task

Ferrigno et al. (2017) showed that human adults, children, and nonhuman primates preferentially categorize dot displays based on their *number* rather than their *cumulative area*, concluding that there must be a "universal" number preference. In practice, though, they showed a preference for number over only one other dimension. How robust is this preference? In Experiment 1a, using a modified version of the same paradigm (see Figure 1C), we compared number against *average size* (rather than *cumulative area*). Participants were trained on dot displays that were either "more" or "less" in average size and number relative to the stimulus space (see the training stimuli in Figure 2A and 2C). Once participants were able to accurately categorize these training displays, they were shown new displays, which placed the two dimensions in a direct contrast, to see whether they would categorize the displays based on their average size or their number (Figure 2B). In a second version of this paradigm (Experiment 1b), participants performed the same task, except with heterogeneously sized dots (see Figure 2D). We reasoned that heterogeneous dot displays would make the task harder, insofar as participants would be forced to parse the ensemble. In a third version (Experiment 1c), we replicated the findings of Experiment 1b using Gabor-like visual gratings, instead of dots, to ensure that participants were not relying on low-level visual features such as brightness (Im & Halberda, 2013; see Figure 2E).

Method

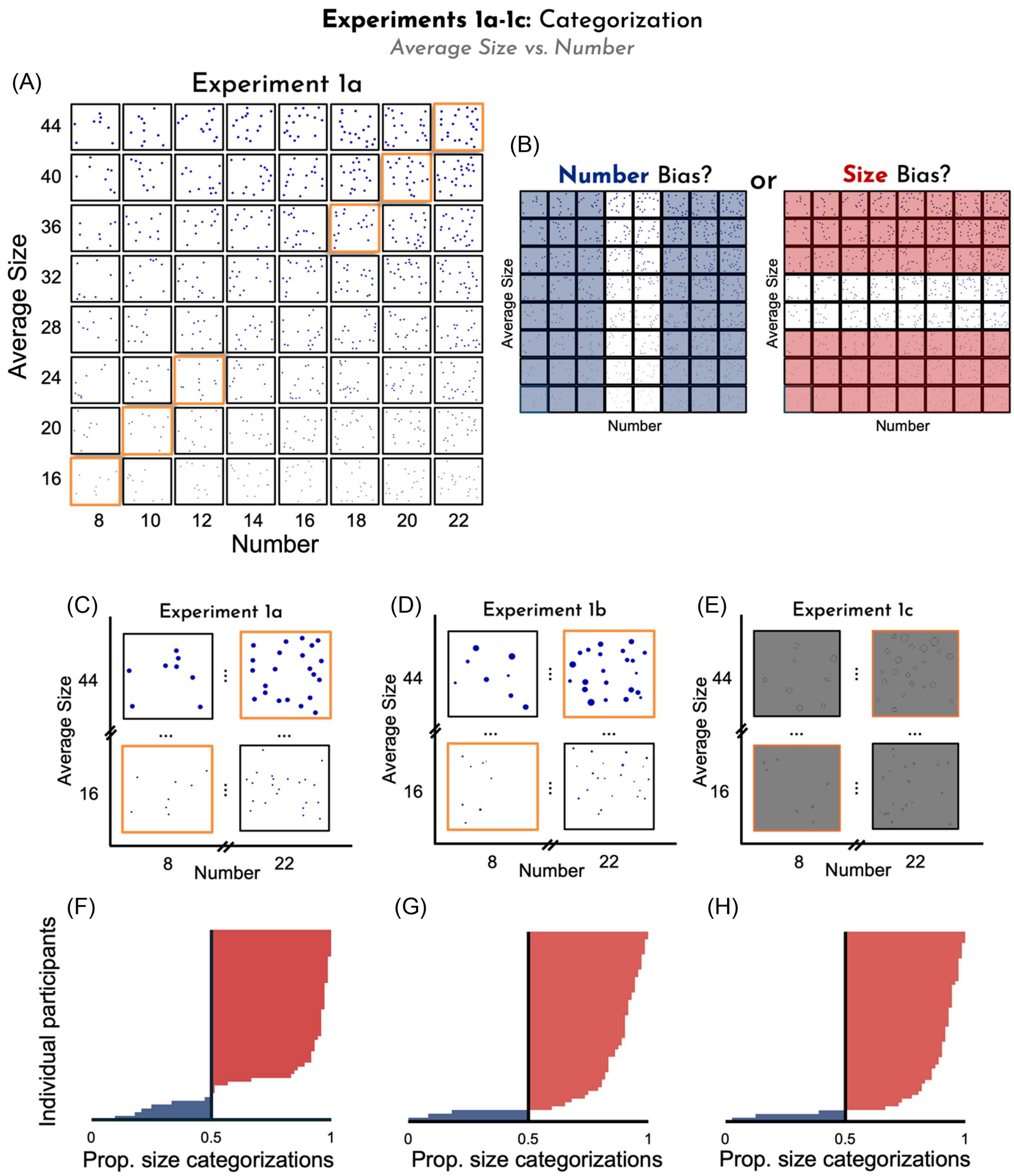
Transparency and Openness

For this experiment and all subsequent experiments, the sample sizes, primary dependent variables, and key statistical tests were chosen in advance and were preregistered. The only exception to this is Experiments 1c and 4d, which were recommended by anonymous reviewers to ensure brightness did not confound with the prioritization of particular quantity dimensions. All stimuli used in these experiments, as well as the raw data, can also be found on our Open Science Framework page (https://osf.io/t7dsv/).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 2

Design and Results of Experiments 1a–1c



Note. (A) The stimulus space for Experiment 1a pitting average size against number. Training trials are highlighted in orange. Note that unlike the stimulus space in Figure 1B, the variance in size in the training trials makes size and number appear truly orthogonalized. (B) The categorization task may result in either a number bias (wherein low number/high size stimuli are placed in the "less" quantity and high number/low size stimuli are placed in the "more" category) or a size bias (where the reverse is true). (C) A simplified version of the stimulus space used in Experiment 1a. The highlighted stimuli are representative of the "low" and "high" categories that the participants were trained on, and the off-diagonal stimuli are representative of the displays that participants were tested on. (D) A simplified version of the stimulus space used in Experiment 1b. Now, the dot displays are now heterogeneously sized. (E) A simplified version of the stimulus space used in Experiment 1c. This experiment utilized Gabor-like gratings instead of dots. (F) Results of Experiment 1a. The majority of participants tended to categorize using size information. (G) Results of Experiment 1b. Again, the majority of participants tended to categorize using size information. (H) Results of Experiment 1c. Once again, participants tended to categorize using size information. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Participants

For each version of the task, 50 participants were recruited online via Prolific, resulting in a total of 150 participants. Participants would have been excluded if they failed to complete the training phase twice in a row; however, this did not occur, and thus, no participants were excluded. This study was approved by the relevant Institutional Review Board.

Stimuli

All stimuli in Experiments 1a and 1b were displays of blue (No. 0000B1) dots, ranging in number from 8 to 22, with the average size of the dots ranging from 16 to 44 pixels in diameter.[2] In Experiment 1a, although the size of dots varied between displays, they were homogeneous within a display. In Experiment 1b, the dot sizes were randomized along a normal distribution curve surrounding the intended mean. Any sizes that were less than half or more than twice the mean were replaced. In Experiment 1c, the dots were replaced by Gabor-like visual gratings. The gratings were created using the GratingStim function in PsychoPy with a circle mask and a spatial frequency of 0.1 px. The orientation of each grating was randomly determined. An outline was also drawn around each grating to ensure boundedness. The size of the gratings was determined in an identical manner to the dots in Experiment 1b. In all versions of the task, the objects were randomly placed into a 400 px × 400 px display, with the constraint that they could not appear within five pixels of another object (from edge to edge) or within 35 pixels of the display's edge (from the center of the object). All displays were randomized such that a participant never saw the same arrangement of objects more than once throughout the experiment.

The stimuli in Experiments 1a and 1b were designed to resemble those from the categorization task performed by Ferrigno et al. (2017). As in their paradigm, training stimuli were created by matching the number and size of the dot displays. Thus, the training stimuli consisted of two "groups"—one representing the "small" quantity and one representing the "large" quantity. In the "small" quantity group, the training stimuli had 8, 9, 10, 11, or 12 objects. The average diameter of the objects (in pixels) was equivalent to twice the number of objects in the display (thus, 16, 18, 20, 22, and 24 pixels, respectively). For the "large" quantity, the training stimuli had 18, 19, 20, 21, or 22 objects, and the average diameter was once again calculated by doubling the number of objects in the display. There were four unique instances of each of the 10 number/size combinations, resulting in a total of 40 training trials. These training trials were presented in a unique random order for each participant.

The probe stimuli were effectively the "off-diagonal" of the stimulus space (see Figure 2A). "Category A" contained stimuli with low number and high size, whereas "Category B" contained stimuli with high number and low size. Category A stimuli had 8, 10, or 12 objects and average diameters of 36, 40, or 44 pixels. These features were counterbalanced with each other, resulting in nine unique Category A stimuli. Category B stimuli had 18, 20, or 22 objects and average diameters of 16, 20, or 24 pixels, resulting in nine unique Category B stimuli. There were four unique instances of each of the 18 probes, resulting in 72 probe trials. These trials were also presented in a unique random order for each participant.

Procedure

All tasks were implemented using custom JavaScript code that was accessed online. Prior to beginning the task, participants were presented with a diagram explaining how to categorize the stimuli. They were instructed to categorize the "small" quantity stimuli as a diamond and the "large" quantity stimuli as a star (see Figure 1C). The stimuli were displayed for 1s. Participants were then presented with the star and diamond on the left and right side of the screen (in a counterbalanced position). Participants were instructed to press either "q" for "left" or "p" for "right" to categorize the stimuli as either a star or a diamond. If the participant was correct, the image of their chosen shape flashed green for 500 ms; if they were incorrect, it flashed red. After the participant had completed all 40 training trials, they were allowed to continue to the probe phase if they had achieved at least 60% accuracy. If they did not, they were required to complete the training phase again.

In the probe phase, participants no longer received any feedback. Instead, after the participants' response, the selected display flashed gray for 500 ms, and the next trial began.

Results and Discussion

The results of Experiments 1a–1c can be seen in Figure 2F–2H. As is evident from the figures, participants in Experiment 1a used size information to categorize the stimuli 83% of the time, $t(49) = 9.27$, $p < .001$, $d = 1.31$. Forty-four of the 50 participants displayed a "size bias," making categorizations based on size for more than 50% of the trials (binomial test, $p < .001$). In Experiment 1b, participants used size information to categorize the stimuli 84% of the time, $t(49) = 11.27$, $p < .001$, $d = 1.59$. Forty-seven of the 50 participants displayed a "size bias" (binomial test, $p < .001$). In Experiment 1c, participants categorized on the basis of size 87% of the time, $t(49) = 13.4$, $p < .001$, $d = 1.89$. Forty-seven of the 50 participants displayed a "size bias" (binomial test, $p < .001$).

These results indicate not only a lack of a "number bias" but also a clear preference for categorizing based on the average size of the ensemble. Experiment 1b provides strong evidence that participants interpret the ensemble in categorizing the stimuli (insofar as no one dot is diagnostic of the average size of the entire set), and Experiment 1c ensures that a bias away from numerical information was not the result of low-level confounds, such as brightness.

Why do we observe such a dramatic difference from Ferrigno et al. (2017)? One possibility is that there simply is not a "universal" number bias. After all, such a claim would require clear evidence that number is prioritized over *all* other plausible dimensions—not just cumulative area. Another possibility is that one of the comparisons made is somehow "unfair." Perhaps one could argue that comparing number against average size is not a natural comparison. However, we think there are at least two reasons why this should be considered a valid contrast class: First, it is taken for granted in vision science that average size is automatically represented as part of visual ensemble representations (see Albrecht & Scholl, 2010; Allik et al., 2013; Ariely, 2001; Chong & Treisman, 2003, 2005; Raidvee et al., 2020;

[2] We manipulated object diameter because prior work has shown that perceived size is better predicted by diameter rather than true area (see Albrecht & Scholl, 2010; Allik et al., 2013; Ariely, 2001; Chong & Treisman, 2003, 2005; Raidvee et al., 2020; Yousif & Keil, 2019, 2021).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Sweeny et al., 2015). Indeed, more attention has been paid to ensemble representations of average size than cumulative area (at least within that literature). Second, we see at least one reason for thinking that cumulative area might be a *worse* contrast class. As noted above, in Ferrigno et al.'s (2017) specific stimulus space, area and number are asymmetric. The stimulus spaces used here do not have this problem. The sets increase in both number and average size along the key diagonal, such that a participant would perceive both changes independently (and subsequently choose between them). We do not mean to argue that average size is a strictly "better" comparison than cumulative area. To us, these data simply represent one obvious case where number information does not seem to be "universally" prioritized.

Experiments 2a–2c: Grouping Task

The categorization tasks in Experiment 1 are relatively demanding, requiring the maintenance of two separate categories in memory throughout the experiment. Furthermore, presenting the stimuli one-at-a-time only exaggerates any asymmetries between the tested dimensions (see our point above about changes in number "causing" changes in area). To sidestep these challenges, we presented participants with a similar categorization task, but instead of displaying each stimulus individually, we showed the participant the entire stimulus space at once and asked them to divide it into two distinct categories (see Figure 3A–3C). Thus, participants were able to interpret the stimulus space holistically and divide it based on any one of several possible ensemble dimensions (e.g., number, cumulative area, average size).

In the first version of the task (Experiment 2a, Figure 3A), we created a stimulus space based on that of Ferrigno et al. (2017) wherein number is contrasted with cumulative area. In the second version (Experiment 2b, Figure 3B), we use the same stimulus space as in Experiment 1a, wherein number is contrasted with average size. In the third version (Experiment 2c, Figure 3C), we use a stimulus space that contrasts number and average size, but with heterogeneously sized dots within each individual stimulus. Notably, the latter version eliminates the possibility that average size is not being perceived as an ensemble property and thus is not fit to be compared to number or cumulative area. A consistent bias toward a certain visual dimension across all three versions of this task would strongly imply that the said dimension was perceptually prioritized by participants.

Method

This experiment was identical to the previous experiment except as noted.

Participants

For each version of the task, 500 new participants were recruited online via Prolific, resulting in a total of 1,500 participants.

Stimuli

Each stimulus space was composed of an 8×8 grid of dot displays (see Figure 3A–3C). In all versions of the task, the number of dots in each display ranged along the *x*-axis from 8 to 22 (in steps of 2). In the area version (Experiment 2a), the proportion of the display covered by the dots ranged from ∼7% to ∼19% along the *y*-axis (this was designed to capture the same relative ratio as is being manipulated by number). In the average size versions (Experiments 2b and 2c), the size of the dots ranged from 16 to 44 px in diameter (in steps of 4) along the *y*-axis. To create the heterogeneous dot displays in Experiment 2c, the dot sizes were randomized along a normal distribution curve surrounding the intended mean. Any sizes that were less than half or more than twice the mean were replaced. There was one image for each combination of number and area/size; however, in all versions, the diagonals of the stimulus space were removed, thus resulting in 48 dot stimuli per display (this was done to allow participants to divide the stimulus space along the oblique axes).

Procedure

Participants were first shown the stimulus space without any dividing line. They were then prompted to use the arrow keys to cycle through four different ways of grouping the stimuli (horizontally or vertically through the center of the stimulus space, or diagonally along either of the oblique axes). All groupings involved separating the space into two sets with an equal number of items in each. The participant was instructed to select the line orientation that best divided the stimulus space into two distinct groups. In all three versions, the participant viewed an example display (using a different stimulus space divided by color) before performing a single trial in which they decided how to group the stimuli.

Baseline Experiment

As a sanity check, and in response to feedback from an anonymous reviewer, we ran a separate baseline experiment to see if people had a default bias to divide the space along a certain axis. A total of 200 participants completed a version of the task in which they saw an empty grid and were asked to divide it in the most natural way. Approximately 40% of people chose to divide the space vertically (i.e., into left and right halves), and approximately 20% of people chose each of the other three options. As is evident from the data below, it is thus highly unlikely that a default bias would explain the results we observed.

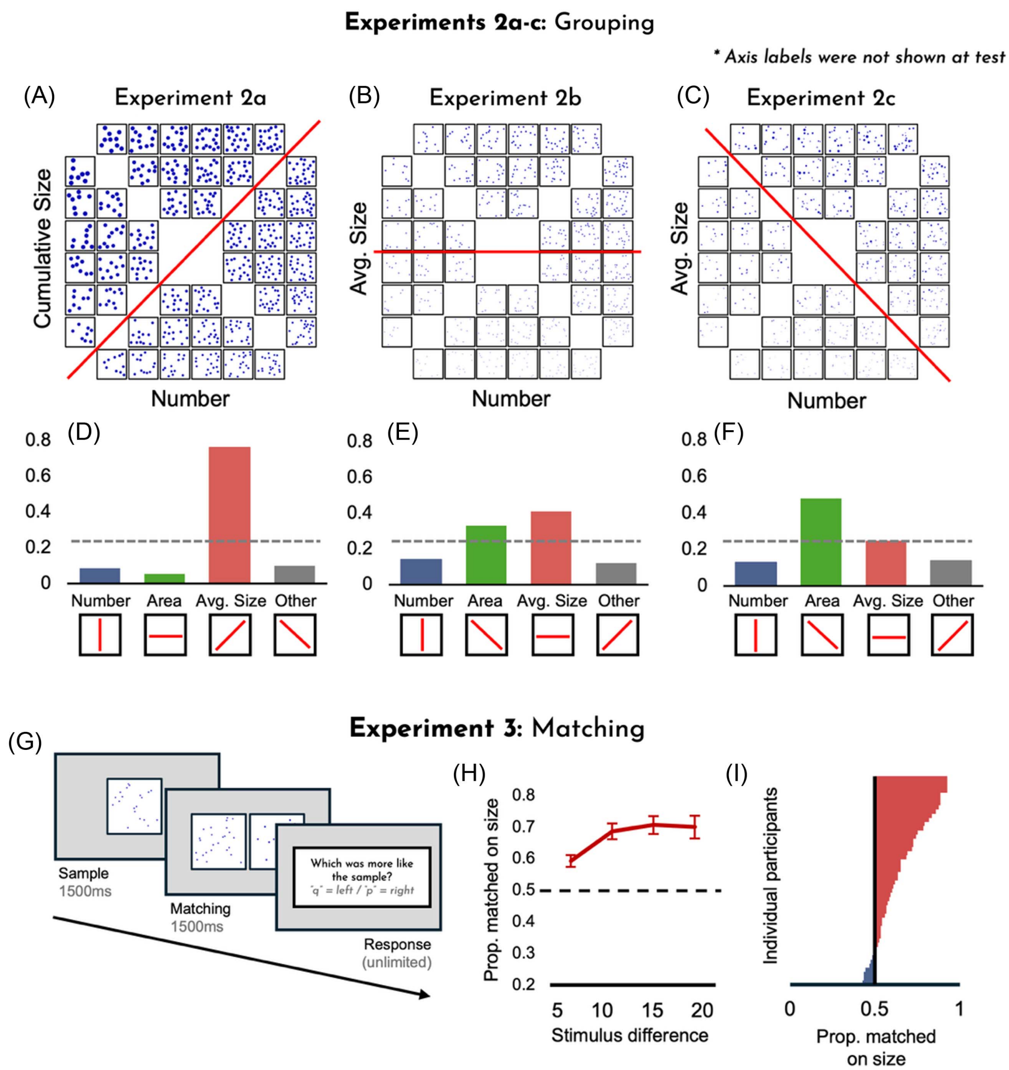
Results and Discussion

The results of Experiments 2a–2c can be seen in Figure 3D–3F. We calculated statistical significance based on 10,000 random simulations of 500 participants each. In Experiment 2a, participants selected the main diagonal (i.e., dividing the space by average size) 76% of the time (well above chance; permutation test, $p < .001$). Participants selected the vertical division (i.e., dividing the space by number) only 9% of the time (*less* than would be expected by random chance; permutation test, $p < .001$).

The results of Experiment 2b were similar, with participants grouping by average size 41% of the time (above chance; permutation test, $p < .001$) and grouping by number only 14% of the time (below chance; permutation test, $p < .001$). Notably, participants also grouped by cumulative area 33% of the time (above chance, permutation test, $p < .001$).

In Experiment 2c, participants once again grouped by number at a below-average rate (13%, permutation test, $p < .001$). However, in

Figure 3
Design and Results of Experiments 2a–2c and 3



Note. (A) Stimulus space for Experiment 2a. Participants were prompted to choose an orientation to divide the space into two distinct groups. The red line indicates the orientation most frequently chosen by participants (this is true of Panels B and C as well). (B) Stimulus space for Experiment 2b. (C) Stimulus space for Experiment 2c. (D) Results of Experiment 2a. Participants tended to divide the stimulus space along the main diagonal, thus grouping by average size. (E) Results of Experiment 2b. Participants tended to group by average size. (F) Results of Experiment 2c. Participants tended to group by average size or cumulative area. (G) Trial protocol for Experiment 3. Participants were shown a sample image and then two matching images (one equal in number but differing in size and one equal in size and differing in number). (H) Results of Experiment 3. Error bars represent the standard error of the mean. The preference for matching based on size information was present for all variances in the stimuli. Here, the *x*-axis indicates the difference in number/size from the sample to the matching displays (e.g., a sample of 20 dots/20 px and matching displays of 30 dots/20 px and 20 dots/30 px would be a trial with difference 10). (I) Results of Experiment 3. The majority of participants tended to select the stimulus that matched in size. See the online article for the color version of this figure.

contrast to the prior experiments, participants tended to group most frequently by cumulative area, selecting the off-diagonal 48% of the time (above chance; permutation test, $p < .001$). Participants grouped by average size 25% of the time (permutation test, $p = .44$); however, the rate at which participants grouped by average size was still significantly higher than that rate at which they grouped by number (permutation test, $p < .001$).

These results provide further evidence that people do not categorize ensembles based on their numerical information. As is evident in Figure 3A–3C, the stimulus spaces used were highly variable, and yet, in all of them, participants consistently failed to group by number. Even using a recreated version of the exact stimulus space from Ferrigno et al.'s (2017) paradigm (Experiment 2a), we find that participants categorize by average size almost

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

exclusively. Even if we set aside the salience of average size, note that the next-most-selected dimension overall was not number, but *cumulative area*. In other words, there seem to be at least some cases in which people spontaneously attend to cumulative area rather than number. Of course, this raises a question about why we would observe seemingly opposite biases in this grouping task as opposed to the original categorization task used by Ferrigno et al. We think this discrepancy validates our concern that changes in area are effectively *caused* by changes in number in the original paradigm. Here, when we remove this impression (i.e., by showing the entire stimulus space at once), we no longer observe a number bias.

Experiment 3: Matching Task

Seemingly, when given direct instructions to categorize stimuli, people tend to make their decisions based on the *average size* of the ensembles. But will this "average size bias" persist in a task with a less explicit reliance on the visuospatial properties of the ensembles? One could imagine, for example, that in previous experiments participants were able to rely on some *nonperceptual bias* toward one dimension over the other (as we argue is the case in Ferrigno et al., 2017). That is, because the tasks were unspeeded, participants may well have made a conscious judgment about which dimension to prioritize, regardless of any perceptual proclivity. Here, therefore, we wanted to test the relative salience of number and average size in an experiment that more explicitly relied on perceptual processing. Participants were presented with a sample image and then shown two similar displays—one which matched the average size of the dots but differed in number and one which matched in number but differed in the average size. With less time to think, participants' responses would be more indicative of perceptual processing.

Method

This experiment was identical to the previous experiment except as noted.

Participants

Fifty new participants were recruited online via Prolific.

Stimuli

All stimuli in this experiment were displays of blue (No. 0000B1) dots, ranging in number from 20 to 40 dots and in diameter from 20 to 40 pixels, each in steps of five.

A "sample" display was made using each pair of equivalent number and size (e.g., 30 dots/30 px diameter, 40 dots/40 px diameter), resulting in five types of displays. For the to-be-matched displays, the deviation of number and size from the sample was equivalent (i.e., if the sample was 30 dots/30 px diameter, a possible combination would be a display with 30 dots/20 px diameter and a display with 20 dots/30 px diameter). Importantly, in each matching stimulus, only *one* of the variables differed from the sample. Each possible deviation from the sample within the range was tested (i.e., a sample value of 40 was compared against test values of 35, 30, 25, and 20), resulting in four trials per sample type. The side with the size-unique versus number-unique stimuli was counterbalanced such that they appeared equally on the left and right throughout the experiment. Each of these unique combinations was repeated three times, resulting in a total of 120 trials.

Procedure

The sample stimulus flashed on the screen for 1,500 ms. Immediately after, the two to-be-matched stimuli were presented side by side for 1,500 ms (see Figure 3G). Participants were asked to indicate which of the displays was more "similar" to the sample display by pressing either "q" for "left" or "p" for "right." Participants did not receive any feedback, since there was technically no correct answer, and thus were simply instructed to make their choice solely on which display they believed to be more similar in whatever way they saw fit. Before completing the task, participants performed four representative practice trials with identical instructions, the data from which were not recorded.

Results and Discussion

The results from Experiment 3 can be seen in Figure 3H–3I. Overall, participants chose the stimulus that matched in size (but differed in number) 65% of the time, $t(49) = 7.10$, $p < .001$, $d = 1.00$. Forty-two of the 50 participants chose the size-similar stimuli more than 50% of the time (binomial test, $p < .001$, see Figure 3B). Thus, these results indicate that when comparing dot displays, people are more likely to judge similarity based on the average size of the dots, rather than their number. This experiment provides converging evidence with the previous experiments that, if anything, people have a default "average size bias."

Experiments 4a–4d: Odd-One-Out Tasks

The first three experiments show that people preferentially categorize and match stimuli based on their average size rather than their number. To provide further converging evidence of this pattern, here, we used another distinct paradigm: an "odd-one-out" task. In this task, participants were shown three displays quickly in succession and asked to determine which they believe was the "odd-one-out." In other words, this task is effectively the opposite of the previous task (in which participants were asked to identify which of the two stimuli in a set of three were most like each other). We reasoned that this task would provide critical converging evidence that the size bias observed in the prior experiments was not the result of any sort of response bias or arbitrary task strategy.

Then, in a second version of this task (Experiment 4b), we addressed a possible concern about the structure of the previous experiments. Instead of comparing number to average size by matching the pixels of diameter, we measured average size by the true area of the dots. In practice, this meant that the *relative* size differences were smaller than before (ensuring that the bias we observed is not simply because the average size differences were more perceptible).

Next, in a third version of this task (Experiment 4c), we varied the sizes of the dots within each display, effectively making average size less salient. We reasoned that if we continued to observe an "average size bias," we could be reasonably confident that the bias is both genuine and robust.

Finally, in Experiment 4d, we replicated Experiment 4c using Gabor-like visual gratings (as in Experiment 1c) to ensure that the bias toward average size was not reliant on low-level visual features like brightness (Im & Halberda, 2013).

Method

This experiment is identical to Experiment 3 except as noted below. For each version of the task, 50 new participants were recruited online via Prolific, resulting in a total of 200 participants.

For Experiment 4a, we used the exact same stimuli as we had in Experiment 3. Instead of selecting which one of two dot displays was more “similar” to a sample, participants were prompted to indicate which of the displays was the “odd-one-out.” All three stimuli in a set (the sample and its two corresponding matching stimuli) were presented sequentially for 1s each (see Figure 4A). Afterward, participants indicated which one was the odd-one-out by pressing either 1, 2, or 3 on their keyboard, corresponding to the first, second, and third stimulus that they saw. It was once again emphasized to participants that there was no correct answer, but that they should make their judgment solely on which display they believed to be different from the rest (and thus, there was no feedback given after each response).

For Experiment 4b, the exact same procedure was used as Experiment 4a, but the stimulus space was modified so that average size was matched to number based on true physical area, rather than diameter. The number of dots still ranged from 20 to 40, but the physical area of the dots in each display ranged from 500 to 1,000 pixels in increments of 125. For the purposes of this experiment, “equivalent stimuli” were those in which the mathematical area was equivalent to 25 times the number of dots (e.g., 20 dots/500 px area). In practice, this meant that average dot diameter ranged from ∼25 to ∼36 px (compared to the previous range of 20–40 px).

For Experiment 4c, the exact same procedure was used as Experiment 4a, but the dots were heterogeneously sized. The number and average size of the displays still ranged from 20 to 40, but within a single display, the average size of the dots varied. The dot sizes were randomized along a normal distribution curve surrounding the intended mean. Any sizes that were less than half or more than twice the mean were replaced.

Experiment 4d was identical to Experiment 4c, except with Gabor-like visual gratings instead of dots. The gratings were created using the GratingStim function in PsychoPy with a circle mask and a spatial frequency of 0.1 px. The orientation of each grating was randomly determined. An outline was also drawn around each grating to ensure boundedness.

In all four experiments, participants completed four representative practice trials with identical instructions, the data from which were not recorded.

Results and Discussion

The results of Experiments 4a–4d can be seen in Figure 4B. In Experiment 4a, participants selected the size-unique stimulus 66% of the time, the number-unique stimulus 16% of the time, and the sample stimulus 18% of the time. The size-unique stimuli were selected more frequently than the number-unique stimuli, $t(49) = 11.69$, $p < .001$, $d = 1.65$. Thus, these results show a clear bias toward using size information to determine an “odd” stimulus.

In Experiment 4b, we found a similar pattern (see Figure 4B): Participants selected the size-unique stimulus 46% of the time, the number-unique stimulus 30% of the time, and the sample stimulus 24% of the time. The size-unique stimuli were selected more frequently than the number-unique stimuli, $t(49) = 3.68$, $p < .001$, $d = .52$. Thus, even when the differences in area were much more subtle, participants still seem to identify stimuli based on average size rather than number. However, it is notable that the strength of this effect is weaker than Experiment 4a, indicating (perhaps unsurprisingly) that perceptual discriminability also affects perceptual salience. Thus, this result validates the importance of accounting for perceptual discriminability as much as possible when comparing quantity dimensions.

In Experiment 4c, we once again found a similar pattern (see Figure 4B): Participants selected the size-unique stimulus 56% of the time, the number-unique stimulus 19% of the time, and the sample stimulus 25% of the time. The size-unique stimuli were selected more frequently than the number-unique stimuli, $t(49) = 11.95$, $p < .001$, $d = 1.69$. The results of this experiment are particularly revealing because in this task participants had to calculate the average size of the ensemble, rather than being able to base their judgment off the size of a single dot. Therefore, the convergence of these results with the prior odd-one-out tasks indicates that the differences in selection of size-unique and number-unique stimuli did not result from the homogeneity of the displays.

In Experiment 4d, we replicated the pattern found in Experiment 4c: Participants selected the size-unique stimulus 56% of the time, the number-unique stimulus 20% of the time, and the sample stimulus 24% of the time. The size-unique stimuli were selected more frequently than the number-unique stimuli, $t(49) = 13.11$, $p < .001$, $d = 1.85$. Thus, it is clear that the bias observed here is not due to confounds with low-level visual properties like brightness.

Additionally, the lack of difference between the sample group and number-unique group across all versions of the task could indicate that not only is size being prioritized, but that number is effectively irrelevant in participants' decisions. Since the sample and number-unique groups were selected relatively equally, this would indicate that participants treated the stimuli as similar, despite their varying number. Overall, these experiments provide further converging evidence of not only the absence of a “number bias” but also a possible “average size bias.”

Experiment 5: Min/Max

The previous experiments have shown that when given a choice, participants tend to preferentially categorize displays based on average size rather than number. Here, we wanted to assess this putative size bias using an entirely different approach. Inspired by recent work (Odic et al., 2024), we wondered whether people would also preferentially extract the minimum and maximum sizes in a set. Participants were simultaneously shown five displays of dots and asked to determine which one had either the minimum or the maximum average size or number.

Method

This experiment was identical to the previous experiment except as follows.

Participants

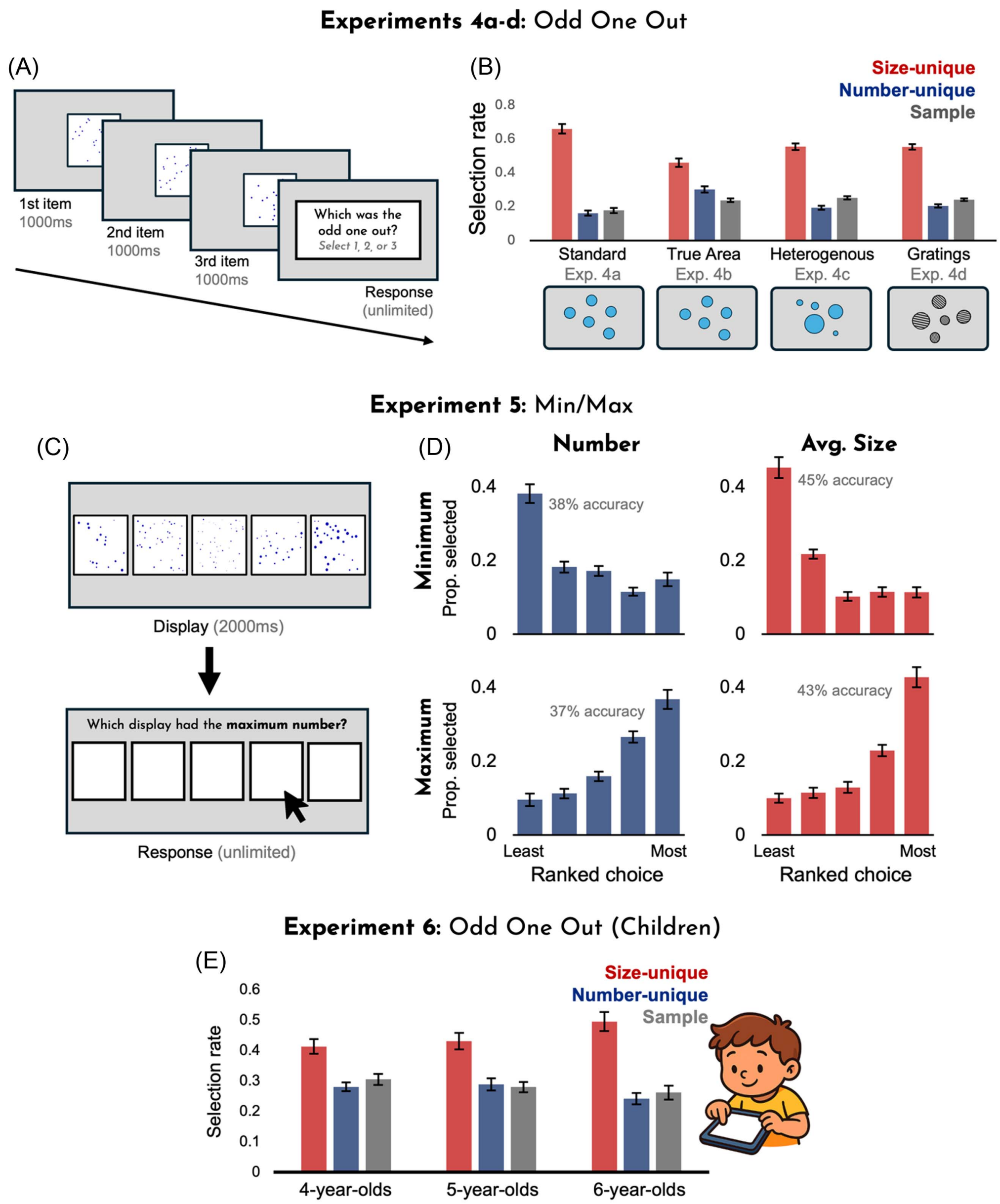
Fifty new participants were tested online via Prolific.

Stimuli

All stimuli in this experiment were displays of blue (No. 0000B1) dots placed randomly on a 400 px × 400 px display with no overlap.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 4
Design and Results of Experiments 4a–4d, 5, and 6



Note. Error bars represent the standard error of the mean. (A) Trial protocol for Experiments 4a–4d. Participants were shown three displays sequentially and asked to indicate which they believed to be the odd-one-out. (B) Results of Experiments 4a–4d. Participants tended to choose the size-unique stimuli as the odd-one-out in all four experiments, indicating that size information was more salient than number. (C) Design of Experiment 5. Participants were shown five dot displays simultaneously. Then, the images were removed, and they were asked to indicate the image that contained either the minimum or the maximum average size or number. (D) Results of Experiment 5. Participants were more accurate when making judgments on average size than number. (E) Results of Experiment 6. Four-, 5-, and 6-year-old children were more likely to select the size-unique stimuli than the number-unique stimuli. See the online article for the color version of this figure.

The size of the dots within a display was heterogeneous. All displays were randomized such that a participant never saw the same arrangement of dots more than once throughout the experiment.

Each trial was composed of five stimuli displayed side by side (see Figure 4C). Within a trial, the average number of dots across the five displays was either 20, 30, or 40, and the average diameter of the dots was either 20 px, 30 px, or 40 px. The five displays in each trial had either 60%, 80%, 100%, 120%, or 140% of the average number/size. The position of each of the five unique items in the sequence was randomized. The average number across all five displays

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

(20, 30, or 40 dots) and the average size across all five displays (20 px, 30 px, or 40 px) were counterbalanced, resulting in nine trial types. Additionally, whether the participant was judging based on size or number and whether the participant was judging based on the minimum or maximum were counterbalanced. Finally, each of the conditions was repeated twice, resulting in a total of 72 trials.

Within each display, the dot sizes were randomized following a normal distribution around the intended mean. Any sizes that were less than half or more than twice the mean were replaced.

Procedure

On each trial, the five displays flashed on the screen in a line for 2s (see Figure 4C). Participants were then asked to indicate which display contained either the minimum average size, maximum average size, minimum number, or maximum number (18 trials for each condition). Participants responded by pressing a number 1–5, indicating the position of the image they were selecting (1 being the leftmost and 5 being the rightmost). Participants did not know what dimension they were to judge until after the stimuli had disappeared. Participants completed four representative practice trials, the data from which were not recorded. The specific condition (e.g., minimum average size, maximum number) for the practice trials was randomized.

Results and Discussion

The results of Experiment 5 can be seen in Figure 4D. When determining the displays containing the minimum and maximum size, participants selected the correct display 44% (45% minimum, 43% maximum) of the time. When determining the displays containing the minimum and maximum number, participants selected the correct display 37% (38% minimum, 37% maximum) of the time. Participants were slightly more accurate when judging based on average size than number, $t(49) = 2.07$, $p = .044$, $d = .29$. Additionally, a ranked score was calculated by assigning a value to the position of the chosen stimulus, where 1 represented the stimulus with the least size/number and 5 represented the stimulus with the greatest size/number (e.g., a score of 2 would indicate that the participant selected the second least stimulus on average). The average ranked score for size judgments was 2.2 for the minimum and 3.8 for the maximum. For number judgments, the average ranked score was 2.5 for the minimum and 3.7 for the maximum. The differences between the ranked scores between number and size judgments were significant for the minimums, $t(49) = 2.20$, $p = .03$, $d = .31$, but not for the maximums, $t(49) = .65$, $p < .52$, $d = .09$. These results suggest that participants are slightly better at identifying extreme sizes rather than extreme numbers. However, the magnitude of these effects is substantially weaker than the previous experiments (despite equivalent discriminability as Experiment 4a), perhaps suggesting that these results should be interpreted with some caution. After all, this task was significantly more complicated (and more challenging) than any of the others employed here.

Experiment 6: Odd-One-Out—Children

Part of the elegance of the paradigm posed by Ferrigno and colleagues, and a reason we were initially hesitant to make a claim against the universal "number bias," is that they demonstrated a bias that was robust across adults of various cultures, children, and nonhuman primates. The converging evidence across these groups is hard to deny. Indeed, many arguments for the universality of a number bias rely on its presence in early development (see, e.g., Tomlinson et al., 2020; Cordes & Brannon, 2008, 2009). Thus, if our results are to have any bearing on those findings, we must ensure that they generalize beyond at least this one group. Given that children have been shown to attend to ensemble average size in a similar capacity to adults (Sweeny et al., 2015), we might also expect children to display an "average size bias." Thus, in this experiment, we performed the odd-one-out-task (Experiment 4a) on a group of 4-, 5-, and 6-year-old children to see whether the size bias we observed previously is also robust across populations.

Method

This experiment was identical to Experiment 4a except as noted.

Participants

Instead of adult participants, the subject pool was twenty-five 4-, 5-, and 6-year-old children (75 total) tested in person at a local museum.

Stimuli

The task was modified to be more child-friendly by displaying all three stimuli simultaneously and keeping them on the screen until a response was made. Additionally, the stimulus space was reduced to only include numbers/sizes of 25, 30, and 35, thus resulting in 36 trials.

Procedure

The task was shown to the child on an iPad. On each trial, the child was shown the three images side by side and asked to point to the display that they thought "didn't belong," at which point the experimenter would click on the display. Before beginning the task, the child completed four representative practice trials with identical instructions, the data from which were not recorded.

Results and Discussion

The results of Experiment 6 can be seen in Figure 4E. Overall, children selected the size-unique stimulus 45% of the time, the number-unique stimulus 27% of the time, and the sample stimulus 29% of the time. The size-unique stimuli were selected more frequently than the number-unique stimuli, $t(24) = 7.08$, $p < .001$, $d = .82$. The 6-year-olds selected the size-unique stimulus the most (50%), and the 4-year-olds selected it the least (41%), $t(48) = 2.09$, $p = .04$, $d = .59$. These results indicate that not only do children *not* display a bias toward number, but that they instead exhibit the same "size bias" as adults.

General Discussion

Across 13 experiments, we document not only a lack of a number bias but also a possible "size bias" in people's ability to categorize, match, and rapidly identify displays based on their ensemble properties. First, in an experiment designed to mimic the categorization

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

task used by Ferrigno et al. (2017), we showed that most participants predominantly use size information to categorize dot displays (Experiment 1a) even when the sizes of the individual dots varied considerably (Experiment 1b), and that this effect is robust even when accounting for low-level visual features like brightness (Experiment 1c). Similarly, we found that when asked to divide a stimulus space contrasting both number/area (Experiment 2a) and number/size (Experiments 2b and 2c), participants tended to group by average size rather than number. We then found converging results in a matching task, showing that participants make similarity judgments using size information rather than numerical information (Experiment 3). In four separate odd-one-out tasks, we showed that participants are more sensitive to changes in size than number (Experiment 4a), even when the veridical area of the displays is accounted for (Experiment 4b), when the displays are heterogeneous in size (Experiment 4c), and when brightness is controlled (Experiment 4d). Then, we extended these findings to an ensemble recognition task, in which we found that participants were slightly more accurate at identifying the minimum and maximum size compared to the minimum and maximum number (Experiment 5). Finally, we extended the results of the odd-one-out task to 4-, 5-, and 6-year-old children, showing that this preference for average size may be, as Ferrigno and colleagues claimed for number, a "universal" feature of human perception (Experiment 6). Collectively, these results undermine the existence of a universal "number bias," providing evidence for a size bias instead.

It Is Not a Competition

To claim that number is uniquely perceptually salient, it is not sufficient to show that number is prioritized over *some* other dimension; there is a need to show that number is prioritized over *all* related ensemble properties. The present data indicate that this is not the case. Instead, it seems as if the salience of different ensembles may vary depending on context. If anything, though, the available evidence seems to be most consistent with a general average size bias.

One unique aspect of average size is that the ease in which it can be extracted from an ensemble can vary, unlike number or cumulative area. Notably, the average size of a homogeneously sized ensemble can be ascertained with greater accuracy and less effort than that of a heterogeneously sized ensemble since the information can be obtained from a single dot. Thus, some may argue that comparing number to average size is unfair in that respect. However, the fact that the "size bias" persists (and is equally robust) in both heterogeneously and homogeneously sized displays (see Experiments 1b–1c, 2c, 4c–4d, and 5) seems to suggest that the preference for average size is not merely born of convenience.

That being said, our point is not really about the specialness of average size. In fact, we think that all arguments about the specialness of one feature over another should be tempered, because such claims are almost always context-dependent (cf. Aulet & Lourenco, 2023). This is not to say that there is nothing "special" about number. The simple fact that any species is capable of spontaneously perceiving numerical information is intriguing on its own—regardless of how they do so and what other dimensions they perceive at the same time. Even so, the sweeping claims often made about the specialness of number have often led to theoretical confusions (see, e.g., Yousif et al., 2024, 2025; Yousif & Clarke, 2026). The "size bias" we document here does not undermine views about the spontaneity of number perception. We still view the study of number as necessary for revealing the foundations of abstract thought, whether in the form of mathematical knowledge or otherwise. Likewise, though, we think the perception of size may reveal just as much about higher order cognitive achievements (see, e.g., Lourenco et al., 2012).

Additionally, we are not meaning to argue that number has *not* played a meaningful role in cultural evolution, as Ferrigno et al. (2017) suggested. It surely has. Our point is merely that the supposed "universality" of spontaneous number perception (a) cannot be easily gleaned from tasks like these, especially if it is compared against only a single dimension, and (b) is not dependent on number being prioritized over other dimensions. Perhaps number and size are *both* salient, "universal" dimensions that contributed meaningfully to the cultural evolution of quantity concepts.

Ensemble Coding

Here, we have primarily focused on a debate within the quantity perception literature concerning the relative prioritization of different visual ensemble dimensions. However, the present findings may also have broader implications for ensemble processing more generally.

First, these results strongly reinforce the now well-established idea that average size is a readily and robustly extracted ensemble property (see, e.g., Albrecht & Scholl, 2010; Allik et al., 2013; Ariely, 2001; Chong & Treisman, 2003, 2005; Raidvee et al., 2020; Sweeny et al., 2015). Perhaps more so than in much prior work, average size was put to a particularly stringent test here: It was directly compared against multiple other ensemble dimensions across a wide range of paradigms. Despite this, average size was reliably prioritized across tasks, stimulus formats, and populations.

Second, these results suggest that ensemble processing may not constitute a single, uniform stage in the visual processing stream. Instead, different ensemble properties may be extracted at different stages or with different efficiencies, leading some dimensions to dominate behavior under time pressure or task uncertainty. On this view, a complete account of ensemble coding would require not only a general theory of ensemble representations but also a dimension-specific understanding of how different summary statistics are computed. Average size representations, for example, may differ systematically from representations of average orientation, numerosity, or other ensemble properties.

Finally, these findings suggest that ensemble coding is context-sensitive, at least to some extent. It remains unclear, for instance, why a robust average size bias emerges in stimulus spaces that previously elicited a number bias in other paradigms. Subtle differences in task structure, stimulus presentation, or response demands may bias observers toward prioritizing one ensemble property over another. For this reason, we urge caution in interpreting results from paradigms of this kind: What appears to be a robust perceptual bias in one context may not generalize beyond it.

Perceptual or Conceptual?

Regardless of whether there is a "number bias" or a "size bias" or neither, we think it may be worth unpacking what causes people to prioritize one dimension over another in a categorization task (like those used by Ferrigno et al., 2017, and throughout the experiments here). One possibility is that people automatically *perceive* number

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

with greater salience, leading them to notice changes in that dimension, latch on to them, and categorize stimuli accordingly. A related but distinct possibility is that people perceive both dimensions as equally salient, but one with greater precision than another (see Odic et al., 2013). These possibilities are distinct but difficult to disentangle.

Another possibility is that people *conceive* of number as special in some way—perhaps because number, unlike size, is a discrete dimension. For all we know, people tend to prioritize discrete dimensions over continuous ones (the present results notwithstanding). Or maybe people do not prioritize number at all, but they find themselves aware of a heavy-handed experimental manipulation. Forced to choose one dimension over the other, they make the conscious inference that, for whatever reason, the experimenters are looking for them to categorize based on number rather than area (perhaps because number is a discrete dimension).

The bottom line here is that categorization tasks like these can, at best, reveal that a dimension is prioritized at *some* level, but they can never reveal exactly how. In this way, we think the min/max task (Experiment 5) adds something that the approach by Ferrigno and colleagues lacks: a measure of actual perceptual ability (as in some prior work; see Aulet & Lourenco, 2021; Odic et al., 2013). But this approach also raises other questions: If size is perceived with greater acuity, is it prioritized only for that reason? The only way to address this question is by creating stimuli that are tailor-made to account for each individual participant's perceptual abilities (see, e.g., Aulet & Lourenco, 2023). Given the robustness of the biases we have observed here, we would be surprised if they could be explained entirely by differences in perceptual precision. But, for now, this remains an open question.

Constraints on Generality

There are a few factors that may constrain the generality of these results. First, insofar as adult participants were taken from an online convenience sample, there is some uncertainty in how these results might generalize across all populations. Second, when attempting to measure the relative contributions of various quantity dimensions, there are invariably issues with controlling every possible visual confound at once (Leibovich et al., 2017). Thus, although we have been faithful to prior work in our experimental design, we fully accept that there may be other "hidden" dimensions at play. Beyond the high-level perceptual features that plague work in quantity perception (e.g., density, perimeter, convex hull), there are many low-level visual features, such as brightness or spatial frequency, that may interfere with judgments of number or average size. Although we attempted to account for some of these properties (e.g., brightness; see Experiments 1c and 4d), invariably, some of these features may be almost hopelessly intertwined (e.g., spatial frequency and average size). That said, even if it were the case that some low-level visual feature was driving the observed phenomenology, this would not change our primary conclusion, for, in that case, we would nevertheless have found that number is less perceptually salient than at least one other visual "summary statistic." Third, this work only explores the relative visual salience of number and average size in human adults and children with explicit number knowledge. Previous work on the salience of ensemble dimensions has shown (weaker, but robust) number biases in both nonhuman primates (Cantlon & Brannon, 2007; Ferrigno et al., 2017) and nonnumerate indigene individuals (Ferrigno et al., 2017), implying that the prioritization of numerical information occurs regardless of mathematical knowledge. Since we did not test these populations, it remains unclear how exactly nonnumerate populations would perform in these paradigms. However, we note that recent work has documented a size bias much like the one documented here, and in even more distant phylogenetic relatives—pigeons (Diaz & Wasserman, 2026).

Conclusion

Despite everything that we have said here, it appears that number is special in at least one way: It has captured the imagination of thousands of scholars for several decades in a way that other dimensions have not. There has been a theoretical intrigue with number that is unlikely to dissipate anytime soon. In every other way, though, it seems to us that the evidence of a number bias is lacking: Evidence accumulated over the past several decades is mixed, at best— and here, we have shown a particularly decisive case of a bias *against* number. Fortunately for number fans, our main point is not that number should be removed from its pedestal—but that nothing is gained from placing it there. The deep, timeless questions of numerical cognition will endure all the same.

References

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science*, *21*(4), 560–567. https://doi.org/10.1177/0956797610363543

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research*, *83*, 25–39. https://doi.org/10.1016/j.visres.2013.02.018

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

Arrighi, R., Togoli, I., & Burr, D. C. (2014). A generalized sense of number. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1797), Article 20141791. https://doi.org/10.1098/rspb.2014.1791

Aulet, L. S., & Lourenco, S. F. (2021). Numerosity and cumulative surface area are perceived holistically as integral dimensions. *Journal of Experimental Psychology: General*, *150*(1), 145–156. https://doi.org/10.1037/xge0000874

Aulet, L. S., & Lourenco, S. F. (2023). No intrinsic number bias: Evaluating the role of perceptual discriminability in magnitude categorization. *Developmental Science*, *26*(2), Article e13305. https://doi.org/10.1111/desc.13305

Barth, H., La Mont, K., Lipton, J., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(39), 14116–14121. https://doi.org/10.1073/pnas.0505512102

Brannon, E. M., Abbott, S., & Lutz, D. J. (2004). Number bias for the discrimination of large visual sets in infancy. *Cognition*, *93*(2), B59–B68. https://doi.org/10.1016/j.cognition.2004.01.004

Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology*, *18*(6), 425–428. https://doi.org/10.1016/j.cub.2008.02.052

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

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404. https://doi.org/10.1016/s0042-6989(02)00596-5

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*, Article 12536. https://doi.org/10.1038/ncomms12536

Clarke, S., & Beck, J. (2022). The number sense represents (rational) numbers. *Behavioral and Brain Sciences*, *44*, Article e178. https://doi.org/10.1017/s0140525x21000571

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

Cordes, S., & Brannon, E. M. (2008). The difficulties of representing continuous extent in infancy: Using number is just easier. *Child Development*, *79*(2), 476–489. https://doi.org/10.1111/j.1467-8624.2007.01137.x

Cordes, S., & Brannon, E. M. (2009). Crossing the divide: Infants discriminate small from large numerosities. *Developmental Psychology*, *45*(6), 1583–1594. https://doi.org/10.1037/a0015666

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. Oxford University Press.

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

Diaz, F., & Wasserman, E. A. (2026). Nonnumerical stimuli exert surprisingly strong behavioral control in an unconstrained numerical discrimination learning task. *Cognition*, *274*, Article 106552. https://doi.org/10.1016/j.cognition.2026.106552

Durgin, F. H. (2008). Texture density adaptation and visual number revisited. *Current Biology*, *18*(18), R855–R856. https://doi.org/10.1016/j.cub.2008.07.053

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

Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Development Perspectives*, *7*(2), 74–79. https://doi.org/10.1111/cdep.12019

Ferrigno, S., Jara-Ettinger, J., Piantadosi, S. T., & Cantlon, J. F. (2017). Universal and uniquely human factors in spontaneous number perception. *Nature Communications*, *8*, Article 13968. https://doi.org/10.1038/ncomms13968

Franconeri, S. L., Bemis, D. K., & Alvarez, G. A. (2009). Number estimation relies on a set of segmented objects. *Cognition*, *113*(1), 1–13. https://doi.org/10.1016/j.cognition.2009.07.002

Gebuis, T., & Reynvoet, B. (2012a). Continuous visual properties explain neural responses to nonsymbolic number. *Psychophysiology*, *49*(11), 1481–1491. https://doi.org/10.1111/j.1469-8986.2012.01461.x

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

Gebuis, T., & Reynvoet, B. (2014). The neural mechanism underlying ordinal numerosity processing. *Journal of Cognitive Neuroscience*, *26*(5), 1013–1020. https://doi.org/10.1162/jocn_a_00541

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

He, L., Zhang, J., Zhou, T., & Chen, L. (2009). Connectedness affects dot numerosity judgment: Implications for configural processing. *Psychonomic Bulletin & Review*, *16*(3), 509–517. https://doi.org/10.3758/pbr.16.3.509

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

Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception & Psychophysics*, *75*(2), 278–286. https://doi.org/10.3758/s13414-012-0399-4

Kerjean, E., Howard, S. R., & Avarguès-Weber, A. (2026). *A numerical bias in honeybees: Numerousness is more salient than space and size non-numerical cues during quantity discrimination* [Preprint]. bioRxiv. https://doi.org/10.64898/2026.03.25.714149

Leibovich, T., & Henik, A. (2013). Magnitude processing in non-symbolic stimuli. *Frontiers in Psychology*, *4*, Article 375. https://doi.org/10.3389/fpsyg.2013.00375

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*, Article e164. https://doi.org/10.1017/s0140525x16000960

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 of the United States of America*, *109*(46), 18737–18742. https://doi.org/10.1073/pnas.1207212109

Mix, K. S., Huttenlocher, J., & Levine, S. (2002). Multiple cues for quantification in infancy. Is number one of them? *Psychological Bulletin*, *128*(2), 278–294. https://doi.org/10.1037/0033-2909.128.2.278

Negen, J. (2025). Number blindness in human vision. *Attention, Perception, & Psychophysics*, *87*(6), 1939–1947. https://doi.org/10.3758/s13414-025-03113-7

Odic, D., Knowlton, T., Wellwood, A., Pietroski, P., Lidz, J., & Halberda, J. (2024). Observers efficiently extract the minimal and maximal element in perceptual magnitude sets: Evidence for a bipartite format. *Psychological Science*, *35*(2), 162–174. https://doi.org/10.1177/09567976231223130

Odic, D., Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Developmental changes in the acuity of number and area representations. *Developmental Psychology*, *49*(6), 1103–1112. https://doi.org/10.1037/a0029472

Pisu, V., Mehraeen, S., Graf, E. W., Ernst, M. O., & Adams, W. J. (2025). Biases in the perceived area of different shapes: A comprehensive account and model. *Journal of Experimental Psychology: Human Perception and Performance*, *51*(9), 1167–1177. https://doi.org/10.1037/xhp0001322

Raidvee, A., Toom, M., Averin, K., & Allik, J. (2020). Perception of means, sums, and areas. *Attention, Perception & Psychophysics*, *82*(2), 865–876. https://doi.org/10.3758/s13414-019-01938-7

Spelke, E. S. (2022). *What babies know: Core knowledge and composition* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780190618247.001.0001

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Sweeny, T. D., Wurnitsch, N., Gopnik, A., & Whitney, D. (2015). Ensemble perception of size in 4-5-year-old children. *Developmental Science*, *18*(4), 556–568. https://doi.org/10.1111/desc.12239

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

Waterhouse, G. C. L., & Yousif, S. R. (2025). The "crowd size illusion" and the relativity of number perception. *Cognitive Science*, *49*(12), Article e70148. https://doi.org/10.1111/cogs.70148

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*(6389), 749–750. https://doi.org/10.1038/358749a0

Wynn, K. (1995b). Origins of numerical knowledge. *Mathematical Cognition*, *1*, 35–60.

Wynn, K. (1995a). Infants possess a system of numerical knowledge. *Current Directions in Psychological Science*, *4*(6), 172–177. https://doi.org/10.1111/1467-8721.ep10772615

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

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

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science*, *8*(1), 88–101. https://doi.org/10.1111/j.1467-7687.2005.00395.x

Yousif, S. R., Alexandrov, E., Bennette, E., Aslin, R. N., & Keil, F. C. (2022). Do children estimate area using an "additive-area heuristic"? *Developmental Science*, *25*(4), Article e13235. https://doi.org/10.1111/desc.13235

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*, Article 104439. https://doi.org/10.1016/j.cognition.2020.104439

Yousif, S. R., & Clarke, S. (2026). *Number, adaptation, and perception* [Preprint]. PhilArchive. https://philpapers.org/rec/YOUNAA-3

Yousif, S. R., Clarke, S., & Brannon, E. M. (2024). Number adaptation: A critical look. *Cognition*, *249*, Article 105813. https://doi.org/10.1016/j.cognition.2024.105813

Yousif, S. R., Clarke, S., & Brannon, E. M. (2025). Seven reasons to (still) doubt the existence of number adaptation: A rebuttal to Burr et al. and Durgin. *Cognition*, *254*, Article 105939. https://doi.org/10.1016/j.cognition.2024.105939

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), Article 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

Received October 9, 2025
Revision received May 10, 2026
Accepted May 21, 2026 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.