Journal of Experimental Psychology: General

ISSN: 0096-3445

2024, Vol. 153, No. 8, 2028–2042
https://doi.org/10.1037/xge0001595

Contextual Coherence Increases Perceived Numerosity Independent of Semantic Content

Chuyan Qu[1], Michael F. Bonner[2], Nicholas K. DeWind[1], and Elizabeth M. Brannon[1]
[1] Department of Psychology, University of Pennsylvania
[2] Department of Cognitive Science, Johns Hopkins University

Number perception emerges from multiple stages of visual processing. Understanding how systematic biases in number perception occur within a hierarchy of increasingly complex feature representations helps uncover the multistage processing underlying our visual number sense. Recent work demonstrated that reducing coherence of low-level visual attributes, such as color and orientation, systematically reduces perceived number. Here, we ask when in the visual processing hierarchy coherence affects numerosity perception and specifically whether the coherence effect is exclusive to low-level visual features or instead whether it can be driven by contextual or semantic relationships. We tested adults in an ordinal numerical comparison task with contextual coherence mathematically manipulated using a statistical model of visual object co-occurrence. Across several experiments, we found that arrays with high contextual coherence were perceived as numerically larger than arrays with low contextual coherence. This contextual coherence effect was not attenuated even when we reduced objects to *texforms* (unrecognizable images that preserve midlevel visual features) or removed semantic content from the images through box scrambling and diffeomorphic warping. Together, these results suggest that visual coherence derived from natural statistics of object co-occurrence systematically alters perceived numerosity at low-level visual processing, even before later stages at which items can be explicitly categorized and identified.

Public Significance Statement

While humans possess a number sense that allows the perception of abstract number, perceived number can be systematically biased by irrelevant features. In the previously described coherence illusion, arrays of items that are similar to each other in color or line orientation, are perceived as more numerous than arrays of items that are heterogeneous. In this series of studies, we explore whether abstract features such as semantic relatedness can drive the coherence illusion. We demonstrate that the coherence effect emerges via low level visual processing stages that are independent of categorical or semantic information.

Keywords: numerosity perception, approximate number system, coherence, object, natural statistics

This article was published Online First July 11, 2024.
Timothy Vickery served as action editor.
Chuyan Qu https://orcid.org/0000-0002-1718-4884
This research was partially funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant R01 HD079106 awarded to Elizabeth M. Brannon. The authors thank Nuwar Ahmed, Echo Xu, and Ziqi Liu for help with data collection and management. The authors also thank Yi-Chia Chen and Lingqi Zhang for helpful discussion on texform generation methods. The authors have no conflicts of interest to disclose.

The ideas and data presented in this article were previously presented at the poster session at the annual meeting of the Vision Sciences Society, May 18–23, 2023, St. Petersburg, Florida. Our hypotheses, procedures, and main analyses were preregistered. Experiment 1, 2, and 3a are preregistered at https://osf.io/76shf/?view_only=830868263d4a4bbc9486eed86c033cd1. Experiment 3b is preregistered at https://osf.io/73fn9/?view_only=ba146070a19c47209d317351733a7dfa. The data and materials described in this article can be located at https://osf.io/9jbpf/?view_only=e3f24f4b475344788d6d9d0a65a3b45b.

Chuyan Qu played a lead role in conceptualization, data curation, formal analysis, investigation, methodology, project administration, supervision, validation, visualization, writing–original draft, and writing–review and editing. Michael F. Bonner played a supporting role in conceptualization, methodology, and writing–review and editing. Nicholas K. DeWind played a supporting role in conceptualization, methodology, and writing–review and editing. Elizabeth M. Brannon played a lead role in funding acquisition and resources and a supporting role in conceptualization, project administration, supervision, and writing–review and editing.

Correspondence concerning this article should be addressed to Chuyan Qu, Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104, United States. Email: qchuyan@sas.upenn.edu

Number is an abstract yet prominent property of our external environment. Humans and nonhuman species have an untrained ability to represent number without relying on symbols or verbally counting (e.g., Cantlon & Brannon, 2006; Dehaene, 1997). This evolutionarily preserved sense of number, underwritten by an approximate number system (ANS), supports essential calculation abilities, such as approximate arithmetic, and has been identified as a building block for more advanced symbolic mathematics (Barth et al., 2005, 2006, 2008; Q. Chen & Li, 2014; Halberda et al., 2008; Pica et al., 2004; Qu et al., 2021; Szkudlarek & Brannon, 2017). Despite its importance in our everyday lives, much is unknown about the underlying mechanisms by which nonverbal number representations are formed from sensory input.

Emerging evidence indicates that the formation of numerosity representations relies on multistage neuronal processing along the dorsal visual stream hierarchy from the primary visual cortex to the posterior parietal cortex (Castaldi et al., 2019; Dehaene, 1992; Paul et al., 2022; Skagenholt et al., 2018). Early research suggested that the primitive quantification system underlying our approximate number representations resides in a specialized parieto-frontal brain network (Nieder & Dehaene, 2009). For example, electrophysiological studies in monkeys identified neurons selectively tuned to numerical quantity in the intraparietal areas (Nieder et al., 2002; Nieder & Miller, 2003, 2004). Functional neuroimaging studies in humans found that the bilateral intraparietal sulcus shows selective neural response to number change (Eger et al., 2009; Piazza et al., 2004, 2006). While the intraparietal sulcus is known to be crucial for ANS representations, recent electroencephalography and functional magnetic resonance imaging (fMRI) studies demonstrate that numerosity is also encoded extremely early in the visual stream (DeWind et al., 2019; Fornaciai & Park, 2018; Park et al., 2016). For instance, using electroencephalography, Park et al. (2016) found numerosity-sensitive activity over a medial occipital site at around 75 ms, followed by a later stage around 180 ms over bilateral occipitoparietal sites when adults passively viewed arrays that varied in numerosity. Additional studies using ultra-high-field (7 T) fMRI data found numerosity-tuned responses organized into a hierarchy of topographic numerosity maps throughout human association cortices, including areas of temporal–occipital, parietal–occipital, superior parietal, and frontal cortices (Harvey & Dumoulin, 2017).

An important question that needs addressing at the computational, algorithmic, and mechanistic level is how approximate number representations emerge from different levels of processing. One promising approach that is useful for all levels of explanation is to characterize stimulus features that systematically bias our perception of number (which we term numerical illusions). By asking how such manipulations alter number perception when and where in the visual processing hierarchy these illusions emerge, we will gain a deeper understanding of approximate number representations.

Classic Visual Illusions of Numerosity

A handful of visual illusions are known to systematically influence the perception of number. For example, in the regular random illusion, arrays with regular arrangements of elements appear numerically larger than arrays with random arrangements (Ginsburg, 1976). The occupancy model offers an explanation for the regular random illusion by positing that each element occupies a circular surrounding area within a fixed radius, and the total area collectively occupied by all elements (the occupancy index) produces the correlate of the perceived numerosity (Allik & Tuulmets, 1991). When objects are randomly distributed, they are more likely to cluster together in subfields. Small spacing between two closely neighbored elements leads to overlap in their occupancy area and consequently reduces the total occupancy index, explaining why clustering reduces perceived numerosity.

Another factor that systematically biases perceived numerosity is connectedness, also known as the barbell illusion. In the connectedness illusion, perceived numerosity decreases when pairs of elements are connected by thin lines or Kanizsa-type illusory contours. The degree of underestimation is systematically influenced by the proportion of pairs that are connected (Adriano, Girelli, & Rinaldi, 2021; Adriano, Rinaldi, & Girelli, 2021; Castaldi et al., 2021; Franconeri et al., 2009; He et al., 2009; Kirjakovski & Matsumoto, 2016). The connectedness illusion provides strong evidence for the direct encoding account of numerosity perception. The connections do not change total surface area, spatial frequency, or other low-level features, yet they dramatically impact perceived number as would be expected if numerosity were directly extracted from segmented perceptual units independently from the nonnumerical visual features (Burr & Ross, 2008; Clarke & Beck, 2021; Dehaene & Changeux, 1993). Importantly, the effect of connectedness was observed 150 ms post stimulus onset in the third visual area, whereas earlier event-related potentials were modulated by the veridical number of dots regardless of the number of connections. This suggests that numerosity encoding in early visual cortex reflects raw sensory input but not the perceptual qualia of numerosity, which is influenced by connectedness at later stages (Fornaciai & Park, 2018).

The Coherence Illusion: A Novel Approach to Exploring the Contribution of Different Visual Processing Stages to the Perception of Numerosity

Perceived numerosity is also systematically altered by array coherence or inversely the entropy of an array. In the first demonstration of the coherence effect,[1] the variance in Gabor patch orientation parametrically influenced perceived numerosity with more coherent arrays perceived as numerically larger than less coherent arrays (DeWind et al., 2020). In a second set of studies, we demonstrated that the effect of coherence on perceived numerosity extends to color entropy (Qu et al., 2022). In that work, color was treated as a categorical variable, and color coherence was quantified by information entropy (Shannon, 1948). Children as young as 5 years of age showed the coherence effect, and the effect increased in strength into adulthood.

Numerosity is encoded extremely early in the visual stream and is processed throughout multiple visual stages before transmitting to higher order cortical areas (DeWind et al., 2019; Fornaciai & Park, 2018; Park et al., 2016). Given that color and orientation are coded in early visual cortex (Livingstone & Hubel, 1988; Priebe, 2016; Solomon & Lennie, 2007), one possibility is that the coherence effects driven by color or orientation arise from the interaction between low-level variance representations and numerical signals

[1] Note that the term "coherence effect" is used here to refer to array homogeneity. The inverse of coherence is variance or entropy. While variance and entropy have distinct mathematical definitions, they both generally refer to the degree of heterogeneity in arrays.

and are exclusive to low-level visual features. An alternative possibility is that coherence influences perceived numerosity at later stages and can be driven by higher level features such as category membership or midlevel visual features, such as shape information. For example, Park & Huber (2022) suggested that the color coherence effect we observed might be caused by top-down semantic influences. Yet, a third possibility is that the coherence effect can emerge at multiple levels in the visual processing hierarchy.

The coherence effect provides a unique opportunity to investigate the contribution of different processing levels of numerosity perception because the visual system encodes coherence for a variety of visual dimensions ranging from low-level properties to high-level properties in a hierarchical manner. Prior work demonstrates that variance of low-level visual features, such as orientation, color, and size, is directly encoded (Maule & Franklin, 2020; Norman et al., 2015; Yang et al., 2018). For example, after adapting to arrays with a specific orientation variance, individuals exhibit adaptation aftereffects whereby exposure to arrays with high variance reduces perceived variance in subsequent orientation variance judgments (Norman et al., 2015). Humans also represent variance in high-level visual features, such as facial emotion within a crowd of faces (Haberman et al., 2015). By examining the level of featural abstraction that can drive the coherence effect, we can uncover how information processed at different visual stages contributes to the formation of numerosity representations.

The Present Study

In the present study, we examined when the coherence effect arises in the visual processing hierarchy by asking whether the statistical co-occurrence of objects in visual arrays influences perceived numerosity and, if so, whether this is driven by semantic information. In everyday environments, we rarely perceive objects in isolation; instead, we perceive and interact with objects in different contexts. Some object pairs have a high probability of occurring together in natural scenes, and others do not. For example, we frequently see a bar of soap, a bathtub, and a towel together in a bathroom, while it is rare to see a bar of soap with a streetlight or a blender alongside a tree. Our perceptual experience of object co-occurrence leads to statistical expectations that may facilitate object identification and visual memory (Hollingworth, 2007; Palmer, 1975; Sadeghi et al., 2015). Prior work suggests that the statistical co-occurrence of objects derived from natural scene databases captures information about category similarities between objects (e.g., oranges and lemons are both citrus fruits that are often found together in the kitchen; Sadeghi et al., 2015). Object co-occurrence statistics also capture information about cross-category associations. For example, oranges and lemons do not share basic-level categorical similarities with kitchen appliances; nevertheless, they have higher statistical co-occurrence with kitchen items than with bedroom items because they are often observed near each other (Sadeghi et al., 2015). Objects belonging to the same taxonomic category have meaningful similarities such as taste or function but are also more likely to be similar in low- and midlevel features such as shape, size, and color (Dilkina & Lambon Ralph, 2013). In contrast, items that cut across taxonomic categories but have high statistical co-occurrence may share fewer low- and midlevel features (e.g., a toaster and a loaf of bread). Thus, statistical co-occurrence of arrays captures rich information across multiple representational levels, including low-level visual features (e.g., color), midlevel features (e.g., shape), and high-level semantic knowledge of contextual associations that goes beyond perceptual properties. Each representational level explains unique variance in the statistics of visual object context. Statistical co-occurrence is therefore well-suited for probing the representational levels at which the coherence of visual arrays affects numerosity perception along a hierarchy of increasingly complex feature representations.

In Experiment 1, we asked whether the degree of statistical co-occurrence of arrays systematically affects perceived numerosity. To mathematically manipulate the degree of statistical co-occurrence of arrays, we utilized a statistical model of visual object context ("object2vec"). In a recent neuroimaging study, Bonner and Epstein (2021) explicitly modeled real-world statistics of object context and related these statistics to object-evoked fMRI responses. Their results demonstrated that object representations in visual cortex reflect statistical regularities of object co-occurrence in natural environments, suggesting that the contextual information associated with objects is encoded in the visual system even in the absence of a surrounding scene. For the present study, we calculated the mean of all pairwise distances of objects for each set based on the embeddings from the object2vec model to quantify the *contextual coherence* of the array.

Experiment 1 revealed a strong contextual coherence effect. As is the case with orientation and color coherence, arrays with high contextual coherence were perceived as numerically larger than arrays with high variance or entropy.

In Experiment 2, we tested whether midlevel visual features were sufficient to drive the contextual coherence effect observed in Experiment 1 independent of higher level processing of object recognition. Midlevel visual features are combinations of features reflecting textural and shape information that loosely preserve local corners, junctions, and contours, while eliminating object identity (Long et al., 2016, 2018). These features are of intermediate complexity, as they are more complex than low-level features like luminance and contrast, but more primitive than high-level features, which capture semantic information like object identity. We created "texforms"—synthetic stimuli that preserve some texture and coarse form information from the original images but render the object identities unrecognizable (Deza et al., 2019; Long et al., 2018). We assessed the impact of contextual coherence on perceived numerosity for intact images and the comparable texform images. We reasoned that, if the contextual coherence effect could be driven by midlevel visual features in the absence of semantic content, then the coherence effect would be similar in the intact and texform conditions. In contrast, if the semantic features of the objects above and beyond the midlevel perceptual features contribute to the contextual coherence effect, then reducing objects to texforms should attenuate or eliminate the coherence effect. We found no attenuation in the strength of the coherence effect, definitively demonstrating that midlevel features are sufficient to drive the effect.

Finally in Experiment 3, we tested whether very low-level visual stimuli that disrupted midlevel features are sufficient to drive the contextual coherence effect. Although texform images prevent precise semantic identification, they do allow discrimination of animate versus inanimate objects and large from small objects, suggesting that midlevel visual features carry information about broad category distinctions (Long et al., 2016, 2017; Long & Konkle, 2017). Moreover, the large-scale organization of the object selective cortex along the ventral visual stream can be elicited by midlevel features without requiring explicit access to the semantic content of objects (Long et al., 2018). As such, it is possible that the basic-level category membership associated with midlevel features drove the contextual coherence effect we observed in Experiment 2. In Experiment 3, we asked whether low-level visual features isolated from both semantic knowledge and broad category membership would be sufficient to drive the contextual coherence effect. To do this, we used both a box scrambling (Experiment 3a) and a diffeomorphic scrambling algorithm (Experiment 3b) to preserve low-level visual features of the images while severely constraining or even eliminating processing of object identities and features that carry categorical information (Stojanoski & Cusack, 2014; Vogels, 1999). We found that neither type of scrambling algorithm attenuated the coherence effect, suggesting that the effect of contextual coherence was not driven by either high-level semantic information or broad-category membership but instead is driven by low-level visual features. Furthermore, we demonstrated (Experiment 3c) that the diffeomorphic scrambling algorithm we used was indeed effective at eliminating participant's ability to identify the objects.

Taken together, these findings reveal that contextual coherence, which arises from the co-occurrence statistics of vision, affects the representation of approximate numerosity at low-level visual processing, before items are explicitly categorized and identified. In more general terms, our results provide a blueprint for parsing how visual illusions can emerge within the hierarchy of increasingly complex feature representations.

Transparency and Openness

In each experiment, we report how we determined the sample size, data exclusions, all manipulations, and measures. Hypotheses, procedures, and main analyses for all three experiments were preregistered at https://osf.io/76shf/?view_only=830868263d4a4bbc9486eed86c033cd1 (Experiments 1, 2, and 3a) and https://osf.io/73fn9/?view_only=ba146070a19c47209d317351733a7dfa (Experiment 3b). All data and materials are available in the Open Science Framework repository at https://osf.io/9jbpf/?view_only=e3f24f4b475344788d6d9d0a65a3b45b.

Experiment 1

Experiment 1 was designed to examine whether and how contextual coherence of arrays of objects affects perceived numerosity. We mathematically manipulated contextual coherence based on the pairwise co-occurrence statistics of objects in natural scenes using a statistical model of visual object context ("object2vec"; Bonner & Epstein, 2021). We predicted that arrays with high contextual coherence would be perceived as more numerous than arrays with low contextual coherence.

Method

Participant

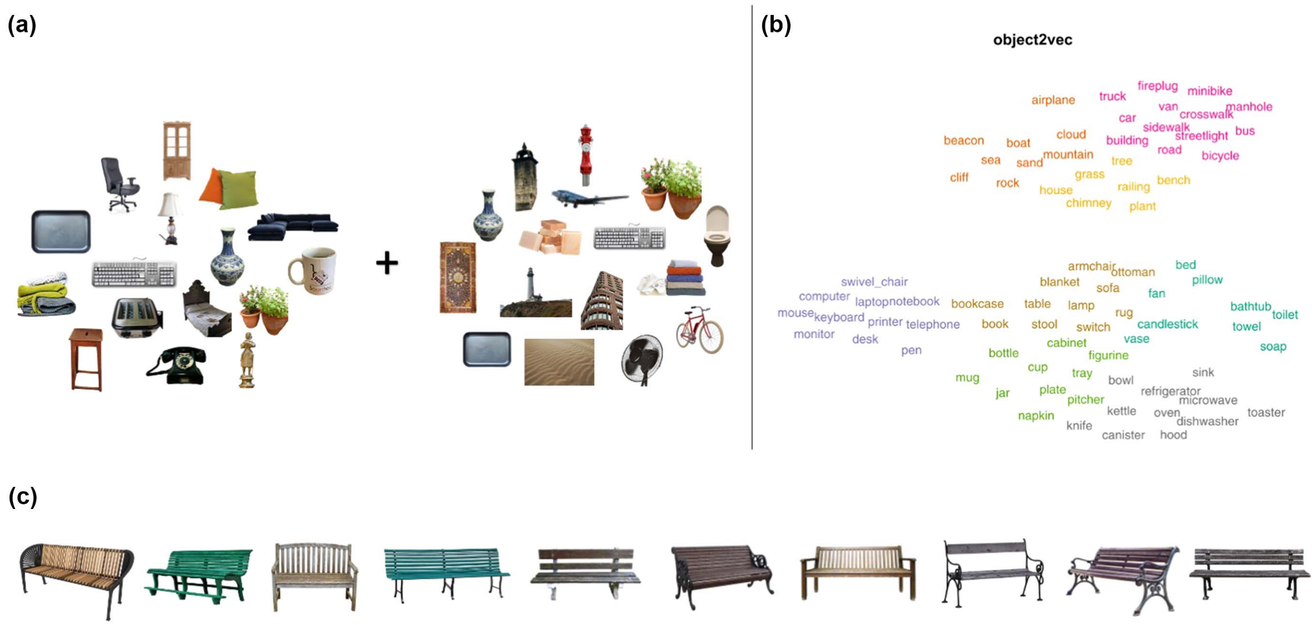
For determining the sample size, we set a Type I error at .05 and an expected Cohen's d at 0.845 based on a pilot study conducted in our laboratory. These values led us to a minimum target sample size of 21 for reaching 95% statistical power. To accommodate the potential for technical problems or inattentive online participants, a sample of 27 adults was recruited online through Prolific. Participants were compensated $10.47/hr. All participants reported normal or corrected-to-normal visual acuity and color vision. Following our preregistered exclusion criteria, we excluded four participants who did not complete all experimental blocks and one participant whose mean task accuracy was lower than 3 *SD*s from the mean of the sample, leaving a final sample size of 22 (M_{age} = 24.33 years, SD = 3.77 years, 16 female participants, six male participants). Participants provided informed consent to a protocol approved by the institutional review board at the University of Pennsylvania before starting the experiment.

Materials and Design

This experiment was programmed and implemented via an online PsychoPy routine (Peirce, 2007). On each trial, two arrays of objects were presented simultaneously for 750 ms. Participants were instructed to judge which array contained the greater number of objects. Each stimulus array was composed of eight to 16 objects (see Figure 1a for an example display). Numerical values were chosen to be approximately evenly spaced on a base-2 log scale while rounding to whole numbers (8, 10, 11, 13, and 16). The numerical ratios between the left-side arrays and the right-side arrays were: 1.4:1, 1.2:1, 1:1, 1:1.2, and 1:1.4. Each participant completed 600 trials in total, with an approximately equal number of trials at each ratio level over two blocks.

Stimuli were taken from a published stimulus set that contains 81 different object categories (e.g., airplane, bottle, and bed), with 10 unique images per category, resulting in a total of 810 experimental images (Bonner & Epstein, 2021). We first computed the pairwise contextual distance of object categories in natural scenes. To accomplish this, we used a statistical model of visual object context ("object2vec") developed by Bonner and Epstein (2021). This model was trained on the image annotations from ADE20K data set to characterize the co-occurrence statistics of object categories in natural images and thus yielded an embedding (eight dimensions) that represents statistical information about the natural contexts for each object category (see Figure 1b for illustration). For example, blenders have a high probability of co-occurring with toasters in the images that were fed into this algorithm, whereas airplanes and toasters had a low probability of co-occurring. An advantage of the object2vec model, for our purposes that differentiates it from the language-based word2vec model, is that the object2vec model characterizes not only categorical similarities between objects within a category but also similarities based on visual contextual information (e.g., between fruit and kitchen items; Sadeghi et al., 2015). We created a contextual distance matrix by calculating pairwise comparisons of the representational vectors of the 81 object categories. To obtain arrays of objects that varied in contextual coherence, we first generated 100,000 random sets of objects from the stimulus set at each numerosity. Then we calculated the mean of all pairwise contextual distances for each set based on the contextual distance matrix of the object categories. A histogram was then plotted for the mean contextual distances of arrays at each numerical value. We chose high contextual coherence from the bottom 20% of the distribution of mean contextual distances, whereas to obtain arrays with low contextual coherence, we selected arrays from the

Figure 1
Stimulus in Experiment 1



Note. (a) Example of two sample arrays in Experiment 1. Both arrays contain 16 unique objects. The array on the left side has greater contextual coherence than the right-side array. (b) A two-dimensional visualization of the object2vec representations for 81 object categories adapted from "Object Representations in the Human Brain Reflect the Co-Occurrence Statistics of Vision and Language," by M. F. Bonner and R. A. Epstein, *Nature Communications*, *12*(1), p. 3 (https://doi.org/10.1038/s41467-021-24368-2). CC BY 4.0. The object2vec representations characterize the co-occurrence statistics of objects in real-world scenes. The colors reflect object clusters that tend to co-occur within scene contexts (e.g., indoor vs. outdoor, urban vs. natural, kitchen vs. office). Note that words were not used in this research, this image is purely for illustration of the object2vec model concept. (c) The stimulus set we used contains 81 different categories, with 10 unique tokens per category. This panel shows 10 exemplars from a sample category "bench." All stimuli in this figure are presented for illustrative purposes, utilizing images for which we possess the appropriate publishing license. See the online article for the color version of this figure.

top 20% of that distribution. There were four different trial types: (a) left side had high and right side had low contextual coherence, (b) right side had high and left side had low contextual coherence, (c) both sides had low contextual coherence, (d) both sides had high contextual coherence. Over 600 trials, each participant viewed 1,200 unique stimulus arrays, approximately 300 arrays at each trial type.

Procedure

Participants completed a consent and demographic form before the start of the experiment. Each trial began with a gray fixation cross presented on a white screen for 500 ms. Then two arrays of objects appeared simultaneously on the white screen for 750 ms followed by the choice screen where participants were instructed to press the "f" or "j" key to indicate which side had more objects. Response time was unlimited, but participants were instructed to respond as fast as possible when they saw the choice screen. The experiment lasted around 20–30 min.

Data Analysis—Generalized Linear Model

To quantify the effects of numerical ratio and contextual coherence on participants' numerosity judgments, we fit a generalized linear model (GLM) using a probit link function and a binominal error distribution with a constant term and regressor for logarithm of the numerical ratio and the difference in contextual coherence between the two arrays. The GLM was derived from previous research on numerosity perception (DeWind et al., 2015; Qu et al., 2022; Tomlinson et al., 2020). The use of the GLM allows for the decomposition of the trial-level choice data into acuity and bias, two factors that combine to determine accuracy.

$$p(\text{choose right}) = \Phi(\beta_{\text{side}} + \beta_{\text{num}}\log_2(r_{\text{num}}) + \beta_{\text{coh}}\text{CoherenceDiff}). \tag{1}$$

In the formula, p(*choose right*) is the proportion of trials that participants chose the right side, Φ is the cumulative normal distribution, and β_{side} is the y-intercept indicating the subject's side bias for choosing one side over the other regardless of the number of objects or contextual coherence. r_{num} is the ratio of the number of the right array to the number of the left array. *CoherenceDiff* is the regressor indicating difference in contextual coherence between the two arrays, which has three possible values (-1, 0, and 1) coded by the following rules: 1 if the right side has greater contextual coherence, -1 if the left side has greater contextual coherence, or 0 if two sides are of equal contextual coherence. Accordingly, the coefficients fit to the numerical ratio and contextual coherence regressors (β_{num} and β_{coh}) for each participant serve as a precise indication of their numerosity discrimination acuity and the effect of contextual coherence on their numerosity judgments. A positive

β_{num} indicates that participants mostly choose the side with greater number of items; therefore, a large value of β_{num} would be associated with sharper acuity of the ANS. If participants engaged in pure numerical discrimination regardless of the contextual coherence, the β_{coh} fit to the CoherenceDiff regressor would be zero. Conversely, any consistent effect of contextual coherence on numerosity perception would lead to a nonzero β_{coh}, where a positive value of β_{coh} indicates that the side with higher contextual coherence is perceived to be more numerous.

Results and Discussion

All 22 participants performed the task with above chance accuracy[2] (accuracy $M = 88.5\%$, $SD = 0.04$, 95% CI [.87, .90], $t_{21} = 46.1$, $p < .001$; one sample t test). We fit the GLM (Equation 1) to each participant's binary response data. We then ran a one-sample t test to determine whether the coefficients' fit to the GLM regressors were significantly different from zero across participants. There was no side bias for the left or right side (β_{side} $M = 0.04$, $SE = 0.06$, 95% CI [−.08, .16], $t_{21} = 0.76$, $p = .46$). We found a significant effect of numerical ratio on participants' binary responses (β_{side} $M = 3.63$, $SE = 0.15$, 95% CI [3.33, 3.94], $t_{21} = 24.84$, $p < .001$). We also found that the mean regression coefficient of the contextual coherence β_{coh} was significantly positive, indicating that participants perceived the more contextually coherent side as more numerous (β_{coh} $M = 0.08$, $SE = 0.02$, 95% CI [0.04, 0.18], $t_{21} = 4.57$, $p < .001$). As the histogram and density plot shows (Figure 2), 18 of 22 (81.8%) participants showed a positive β_{coh} coefficient.

To further determine whether the effect of contextual coherence was modulated by numerical ratio, we next ran a binomial generalized linear mixed-effect model (GLMM) using a logit link function to predict participants' item-level choices. The numerical ratio, the difference in contextual coherence, and the interaction between these two factors were entered into the GLMM as fixed effects while participant was treated as a random effect. We found a significant fixed effect of the numerical ratio ($\beta_{num} = 6.10$, $SE = 0.10$, $Z = 57.48$, $p < .001$, 95% CI [5.89, 6.31]) and the contextual coherence difference ($\beta_{coh} = 0.12$, $SE = 0.03$, $Z = 3.82$, $p < .001$, 95% CI [0.06, 0.19]) and no significant interaction between these two factors ($\beta_{Num \times Coh} = 0.05$, $SE = 0.15$, $Z = 0.34$, $p = .73$). The lack of an interaction indicates that the numerical ratio did not modulate the strength of the contextual coherence effect. To quantify the effect of one unit increase in right-to-left difference in contextual coherence on the likelihood of choosing the right side, we calculated the exponentiated model coefficient of the contextual coherence difference. Note that there were three levels of right-to-left difference in contextual coherence (−1, 0 and 1). We found that one unit increase in the right-to-left difference in contextual coherence led to a 13.0% increase in the odds of choosing the right side.

Experiment 2

The results of Experiment 1 clearly demonstrate that arrays with high contextual coherence were perceived as numerically larger than arrays with low contextual coherence. An intriguing possibility is that the effect of contextual coherence was driven by high-level semantic information such that the conceptual knowledge of object occurrence gained from long-term visual experience biased rapid numerosity perception. However, given that in natural environments objects that tend to co-occur within the same context generally have a higher probability of sharing similar perceptual features such as shapes, sizes, or texture, a more parsimonious possibility is that the contextual coherence effect observed in Experiment 1 was driven by perceptual factors that operate independently from semantic content. Thus, our next experiment was designed to ask whether images that preserve midlevel visual features of intermediate complexity could drive the contextual coherence effect in the absence of explicit object recognition. Midlevel visual features are combinations of features reflecting texture and coarse shape of objects, while eliminating object identity (Long et al., 2016, 2018).

To accomplish this, we used a texture synthesis algorithm to create a class of stimuli known as "texforms," which preserve midlevel features including coarse curvature and some texture information from the original images while eliminating object recognizability (Deza et al., 2019; Long et al., 2018). We also created a comparison group of grayscale intact objects controlling for low-level image properties including contrast and luminance. If the midlevel visual features preserved in texforms are sufficient to drive the contextual coherence illusion independent of semantic content, we would expect to see a similar effect of contextual coherence in both intact and texform conditions.

Method

Experiment 2 used a within-subject design with two conditions: texform-object condition versus intact-object condition. The order of the two conditions were counterbalanced across participants.

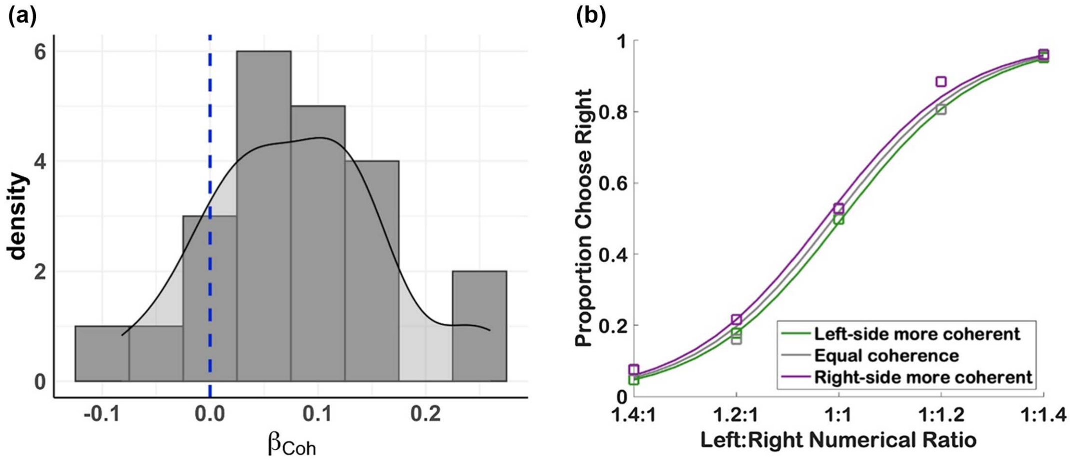
We processed the 810 images used in Experiment 1 using the following procedure adapted from Y.-C. Chen et al. (2022). First, all images were resized to 512 pixel × 512 pixel. Next, we converted all images to grayscale using Rec.601 Luma coding formula and then equalized all images across luminance and contrast using the spectrum, histogram, and intensity normalization (SHINE) toolbox (Willenbockel et al., 2010). Images were then placed in the center of a gray background of 640 pixel × 640 pixel. These processed 810 images were then used to generate 1,200 arrays of intact objects that varied in contextual coherence following the same procedure as in Experiment 1.

To create the corresponding texform images, we used an accelerated texform generation pipeline (Deza et al., 2019) modified from a previous texform model using a texture synthesis algorithm (Long et al., 2018). Specifically, this procedure was done by placing the processed intact images in a simulated visual periphery as input entering into a metamer model (Freeman & Simoncelli, 2011). Then a metamer image was synthesized by iteratively coercing random noises to match the texture statistics of the input image for every overlapping receptive field in the periphery in addition to approximately matching the structure given a low-pass residual of the input image. To produce a final texform image, we ran this procedure for 50 iterations using a variant of gradient descent. Accordingly, we obtained 810 texform images that corresponded to

[2] Accuracy was calculated as the proportion of correctness across trials excluding those with a 1:1 numerical ratio, where there were no correct answers. These trials were nevertheless included for regression modeling analyses because they were highly informative of the effect of the contextual coherence on numerosity judgment.

Figure 2

Arrays With High Contextual Coherence Were Perceived as Numerically Larger



Note. (a) The histogram and density plot indicate the distribution of the regression coefficient β_{coh} across participants. (b) The proportion of choosing the right side as a function of the left-to-right numerical ratio. The psychometric curves denote smooth fit of the generalized linear model in Equation 1 to the pooled data. Green line indicates that the left side had higher contextual coherence than the right side. Purple line indicates that the right side had higher contextual coherence than the left side. Grey line indicates equal contextual coherence of both sides. The squares represent the mean proportion of choosing the right side at each combination of numerical ratio levels and right-to-left difference in contextual coherence across participants. See the online article for the color version of this figure.

the intact objects and generated a stimulus set of texform-object arrays following the same procedure detailed for Experiment 1. Other methods, procedure, and data analyses were identical to Experiment 1 (Figure 3).

Participant

Sixty-three adults (M_{age} = 24.37 years, SD = 1.23 years, 16 female participants, 46 male participants, one nonbinary participant) were recruited via Prolific participated in Experiment 2. Following the preregistered exclusion criteria, we excluded an additional 19 participants either because they did not complete all blocks or because their mean task accuracy was lower than 3 *SD*s from the mean of the sample. All participants reported normal or corrected-to-normal visual acuity and color vision, provided informed consent, and were compensated for their time.

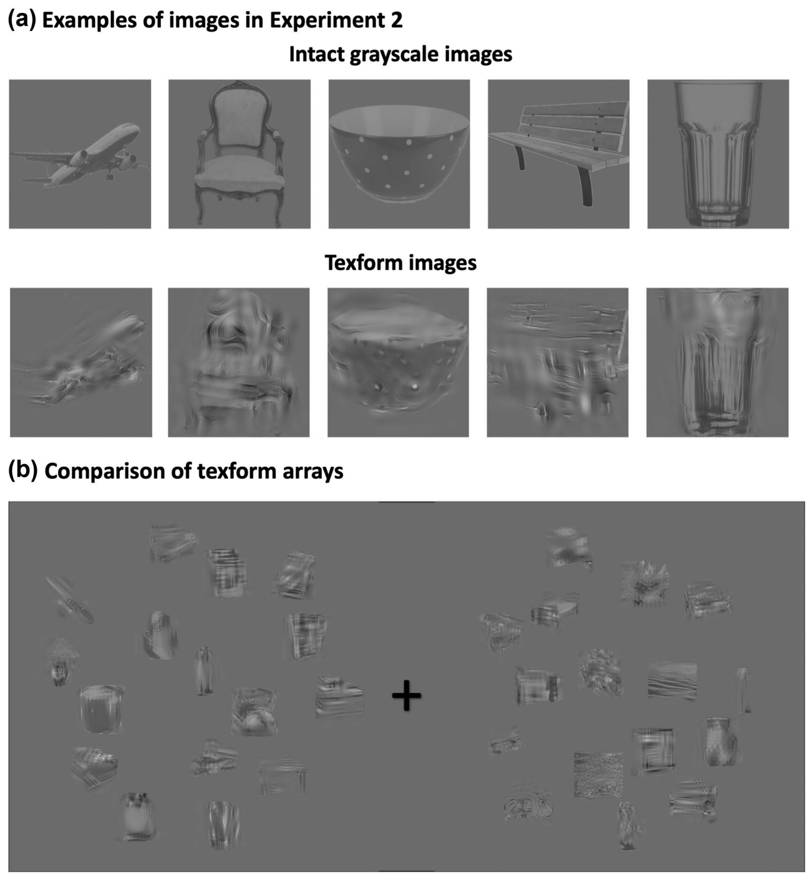
Results and Discussion

All participants performed the numerosity comparison task with above chance accuracy in both the intact-object condition (accuracy M = 87.80%, SD = 0.05, 95% CI [.87, .89], $t_{62} = 58.47$, $p < .001$; one sample t test) and the texform-object condition (accuracy M = 88.47%, SD = 0.05, 95% CI [.87, .90], $t_{62} = 65.26$, $p < .001$; one sample t test). We quantified the contextual coherence effect on numerosity judgments by modeling each participant's binary choice data using Equation 1. As predicted, we observed a significant effect numerical ratio on participants' numerosity judgment in both conditions (intact-object condition: β_{num} M = 3.56, SE = 0.10, 95% CI [3.35, 3.76], $t_{62} = 34.57$, $p < .001$; texform condition: β_{num} M = 3.70, SE = 0.11, 95% CI [3.49, 3.91], $t_{62} = 34.51$, $p < .001$). Forty-six out of 63 participants (73.02%) exhibited a positive β_{coh}. The mean β_{coh} was significantly positive in the intact-object condition (β_{coh} M = 0.04, SE = 0.01, 95% CI [0.02, 0.06], $t_{62} = 3.77$, $p < .001$), indicating that object arrays with high contextual coherence were perceived as more numerous even when color and contrast were eliminated as cues. In the texform condition, 40 of 63 participants (63.49%) exhibited a positive β_{coh}. We also found a significantly positive β_{coh} across participants (β_{coh} M = 0.03, SE = 0.01, 95% CI [0.01, 0.05], $t_{62} = 2.67$, $p < .01$), indicating that midlevel visual features preserved in texforms were sufficient to drive the contextual coherence effect even when the high-level processing of object identities was disrupted.

A binominal GLMM with participant treated as a random effect confirmed the above results. The numerical ratio and difference in contextual coherence were both significant fixed effects on predicting the item-level responses of choosing the right side (intact-object condition: β_{num} = 6.20, SE = 0.19, Z = 31.79, p < .001, 95% CI [5.81, 6.58]; β_{coh} = 0.07, SE = 0.02, Z = 3.45, p < .001, 95% CI [0.03, 0.10]; texform condition: β_{num} = 6.13, SE = 0.06, Z = 97.49, p < .001, 95% CI [6.01, 6.26]; β_{coh} = 0.04, SE = 0.02, Z = 2.35, p = .019, 95% CI [0.007, 0.08]). One unit increase in the right-to-left difference in contextual coherence was associated with a 6.8% increase in the odds of choosing the right side in the intact condition and a 4.5% increase in the odds of choosing the right side in the texform condition.

Although the coherence effect was observed in the texform condition suggesting that midlevel features are sufficient to drive the coherence effect, we next tested whether the magnitude of the coherence effect was attenuated by reducing images to texforms. If semantic information is processed in the intact condition, it might afford an additional basis for the coherence effect above and beyond midlevel features. To test this, we compared the magnitude of the contextual coherence effect quantified by the GLM coefficient

Figure 3
Stimulus in Experiment 2



Note. (a) Examples of Grayscale Intact Images and Texform Images in Experiment 2 and (b) Example Display of Trial-Level Comparison Between Two Texform Arrays.

β_{coh} between the intact and the texform conditions across participants. As shown in Figure 4, the pairwise *t* test showed no significant difference on the contextual coherence effect between these two conditions, intact-object condition: $M = 0.042$, $SD = 0.09$; texform condition $M = 0.030$, $SD = 0.09$; $t(62) = 0.72$, $p = .475$, Cohen's $d = 0.09$. To ensure that a nonsignificant result in conventional significance testing was not simply due to data insensitivity, we next conducted Bayesian factor analyses to determine whether nonsignificant results support a theoretically meaningful null hypothesis that no difference exists between condition (Dienes, 2014; Rouder et al., 2009). Bayes factor (BF) analyses provided moderate evidence in favor of the null hypothesis ($BF_{10} = 0.177$), indicating a similar strength of the contextual coherence illusion in both conditions. Thus, the semantic content in the intact condition did not increase the magnitude of the coherence effect.

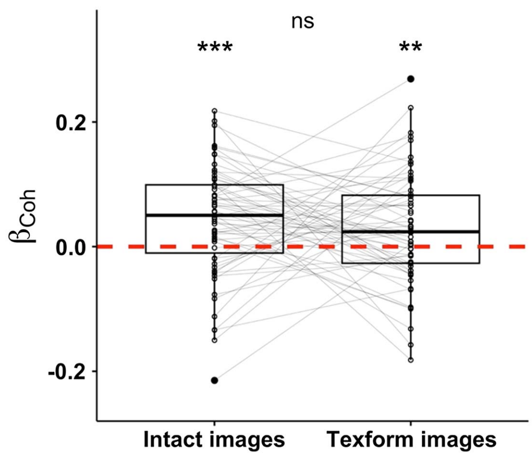
Experiments 3a, 3b, and 3c

In Experiment 2, the magnitude of the coherence effect elicited by texform images that preserved midlevel features but largely disrupted semantic processing was similar to the coherence effect elicited by intact images. This finding suggests that the contextual coherence effect is not driven by semantic processing and instead is driven by lower level features shared by objects that co-occur in real-world scenes. However, given that midlevel perceptual features are sufficient to differentiate broad classes of objects, such as animacy and object size (Konkle & Oliva, 2012; Long et al., 2016, 2018), another possibility is that the basic-level categorical information that is carried by the midlevel features could drive the contextual coherence effect. We thus designed Experiments 3a and 3b to assess if the contextual coherence effect can be observed with merely low-level visual features, even when course semantic information is disrupted.

In Experiments 3a and 3b, we tested whether scrambling the object images, rendering them unrecognizable, eliminates or attenuates the coherence effect compared to the intact-object condition. Both scrambling methods preserve basic low-level visual features of images while removing semantic content and features that carry any type of categorical information. We reasoned that, if the contextual coherence effect was driven solely by low-level visual features, the coherence effect should be similar in the scrambled and intact conditions. Experiments 3a and 3b were otherwise identical, except for the scrambling methods we used to disrupt the semantic information. In response to a reviewer critique, we ran Experiment 3c to verify that the diffeomorphic scrambling images were not recognizable.

Figure 4

The Effect of Contextual Coherence (Quantified by the Regression Coefficient β_{coh}) for the Intact and Texform Conditions



Note. In each box, the central band is the median, the edges of the box are the first and third quartiles, and the whiskers indicate ±1.5 × interquartile range. Asterisk indicates a significant effect of contextual coherence for each condition. *ns* indicates no significant difference between these two conditions. Dots and lines represent participants. See the online article for the color version of this figure.

Method

Participants

Fifty-one adults (M_{age} = 24.26 years, SD = 3.58 years, 19 female participants, 32 male participants, zero nonbinary participants) participated in Experiment 3a, and 62 adults (M_{age} = 23.17 years, SD = 3.61 years, 26 female participants, 36 male participants, zero nonbinary participants) participated in Experiment 3b. An additional 41 adults (M_{age} = 24.16 years, SD = 2.41 years, 17 female participants, 20 male participants, four nonbinary participants) participated in Experiment 3c. All participants were recruited via Prolific and tested online. All participants reported normal or corrected-to-normal visual acuity and color vision and provided informed consent before starting the experiment. An additional 15 and 20 participants were excluded from Experiments 3a and 3b, respectively, either because they did not complete all blocks or because their mean accuracy was lower than 3 *SD*s from the mean of the sample. All experiments were approved by the institutional review board in a large research university.

Materials and Design

Experiments 3a and 3b. Experiments 3a and 3b both used a within-subject design with two conditions: scrambled-object condition versus intact-object condition. The intact-object condition in each experiment was a direct replication of Experiment 1. For both experiments, the order of the two conditions were counter-balanced across participants.

In Experiment 3a, we used a box scrambling algorithm to alter the 810 object images used in Experiments 1 (Vogels, 1999). The scrambling function divided each of the original images into a 5 × 5 matrix of squares and randomly repositioned each square to a new location within the boundary of the image (see Figure 5). In Experiment 3b, we used a more rigorous scrambling method by applying a diffeomorphic transformation to the images used in Experiment 1 (Stojanoski & Cusack, 2014). The diffeomorphic transformation algorithm was designed to match the neural activity at the early stages of visual processing between the scrambled images and intact images while rendering objects unrecognizable. This procedure was done by applying a 2D-flow field to the images over 20 iterations with a maximum distortion of 30. It effectively distorted images without altering any major features of the nonbackground pixels as well as the topography of the image.

Experiment 3c. In Experiment 3c, we tested the recognizability of the images used in Experiments 3b to assess whether diffeomorphic transformation effectively eliminated semantic content of the images. We used a between-subject design with two conditions: a diffeomorphic scrambled condition (n = 22, M_{age} = 24.72 years, SD = 2.82 years, seven female participants, 11 male participants, four nonbinary participants) versus an intact-object condition (n = 19, M_{age} = 23.63 years, SD = 1.86 years, 10 female participants, nine male participants, zero nonbinary).

Stimuli were taken from images used in Experiment 3b, which contained 81 different categories. We randomly selected one stimulus from each of the 81 categories for each of the two conditions. We further randomly selected one object from each of the stimulus arrays and highlighted that object with a red lasso to indicate which object the participants should attend. Thus, the stimulus set for this experiment included 81 unique images for the scrambled condition and 81 unique images for the intact condition.

In the instruction phase, participants were told that they would see intact or scrambled versions of object arrays, and their task was, to the best of their ability, to identify the targeted object in the image—"limited to a word or two." We gave some example answers (e.g., "airplane," "toaster," "soap") to cover a wide range of possible categories. Each trial started with a 200-ms fixation screen. Then an array of objects with a target object highlighted with a red lasso was presented for 2,000 ms. Participants were instructed to report whether they thought they could identify the target object by pressing "f" for yes or "j" for no. After participants made the binary choice as to whether they could identify the object, a prompt was presented "Please label the object with 1 or 2 words. If you did not recognize the object please make your best guess. Type your answer here."

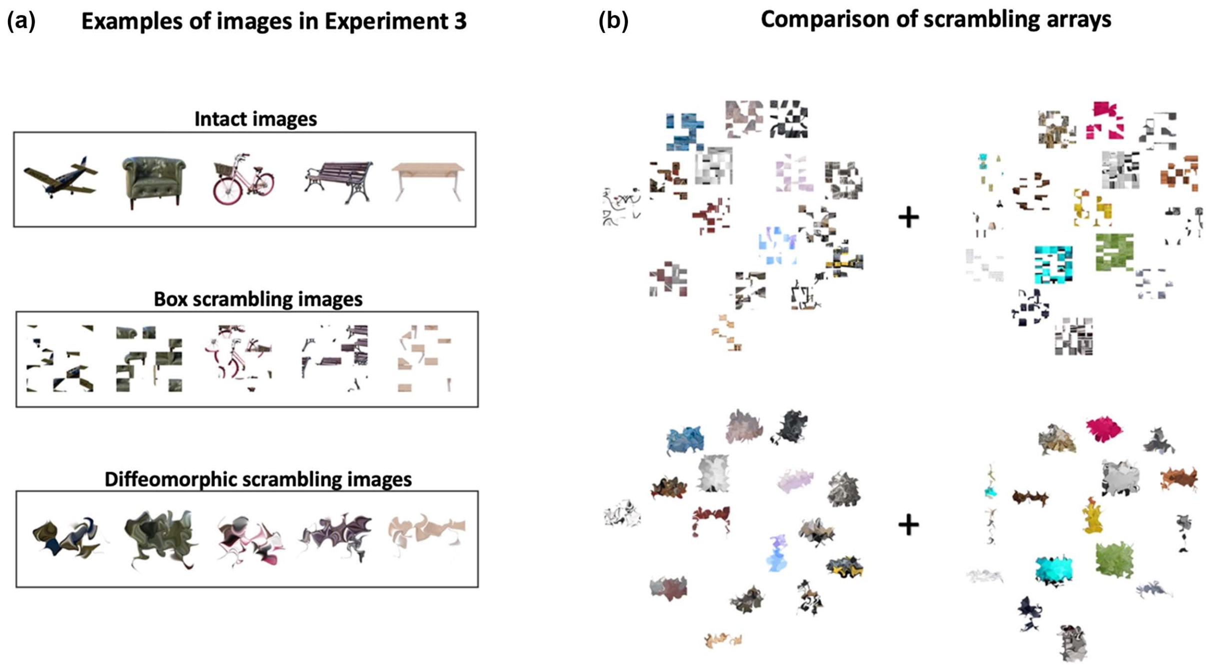
Two independent raters coded the written responses as correct or incorrect. The coders were shown the original intact object image for both the scrambled and the intact stimuli, the ground truth labels from the Bonner and Epstein (2021) data set, and the responses that participants provided. Raters indicated whether the responses were adequate descriptors of the original target object. For example, if the object was a bicycle, correct answers might include bicycle, bike, 10-speed, mountain bike. We reexamined any trials for which the two raters did not agree and reported reliability.

Results and Discussion

Experiments 3a and 3b

Participants performed the numerosity comparison task with above chance accuracy in both the intact-object condition (Experiment 3a: accuracy M = 89.17%, SD = 0.04, 95% CI [.88, .90], t_{50} = 65.06,

Figure 5
Stimulus in Experiments 3a and 3b



Note. (a) Examples of intact and scrambling images from Experiments 3a and 3b. (b) Example displays of trial-level comparison of scrambling arrays. The top panel represents the box scrambling condition in Experiment 3a. The bottom panel represents the diffeomorphic scrambling condition in Experiment 3b. See the online article for the color version of this figure.

$p < .001$; Experiment 3b: accuracy $M = 89.52\%$, $SD = 0.03$, 95% CI [.89, .90], $t_{61} = 98.67$, $p < .001$; one sample t test) and the scrambled-object condition (Experiment 3a: accuracy $M = 88.49\%$, $SD = 0.05$, 95% CI [.87, .90], $t_{50} = 57.91$, $p < .001$; Experiment 3b: accuracy $M = 90.42\%$, $SD = 0.03$, 95% CI [.90, .91], $t_{61} = 104.5$, $p < .001$; one sample t test). To quantify the effect of contextual coherence, each participants' binary choice data in the two conditions was modeled separately using Equation 1. We then ran a one-sample t test to determine whether the coefficients of the GLM regressors were significantly different from zero across participants.

Intact-Object Condition. We replicated the findings of Experiment 1 in the intact-object condition in both Experiments 3a and 3b. There was a significant effect of numerical ratio on participants' numerosity judgments in both experiments (Experiment 3a: β_{num} $M = 3.75$, $SE = 0.11$, 95% CI [3.52, 3.98], $t_{50} = 32.92$, $p < .001$; Experiment 3b: β_{num} $M = 3.81$, $SE = 0.09$, 95% CI [3.64, 3.98], $t_{61} = 44.48$, $p < .001$). Again replicating Experiment 1, the mean β_{coh} for both experiments was significantly positive, indicating that the more contextually coherent side was perceived as more numerous (Experiment 3a: β_{coh} $M = 0.05$, $SE = 0.01$, 95% CI [0.02, 0.07], $t_{50} = 3.97$, $p < .001$; Experiment 3b: β_{coh} $M = 0.06$, $SE = 0.01$, 95% CI [0.04, 0.08], $t_{61} = 5.68$, $p < .001$). Thirty-five of 51 participants (68.62%) in Experiment 3a and 50 of 62 (80.64%) participants in Experiment 3b showed a positive β_{coh}. This finding was confirmed using a binomial GLMM to control for random effect of participants. Again, we observed a significant fixed effect of numerical ratio (Experiment 3a: $\beta_{num} = 6.57$, $SE = 0.21$, $Z = 31.25$, $p < .001$, 95% CI [6.16, 6.99]; Experiment 3b: $\beta_{num} = 6.65$, $SE = 0.16$, $Z = 42.72$, $p < .001$, 95% CI [6.34, 6.95]) and contextual coherence (Experiment 3a: $\beta_{coh} = 0.08$, $SE = 0.02$, $Z = 3.42$, $p < .001$, 95% CI [0.03, 0.12]; Experiment 3b: $\beta_{coh} = 0.10$, $SE = 0.02$, $Z = 5.10$, $p < .001$, 95% CI [0.06, 0.14]) and no interaction between these two factors (Experiment 3a: $\beta_{Num\times Coh} = 0.16$, $SE = 0.10$, $Z = 1.63$, $p = .10$; Experiment 3b: $\beta_{Num\times Coh} = -0.10$, $SE = 0.09$, $Z = -1.10$, $p = .28$). One unit increase in the right-to-left difference in contextual coherence was associated with an 8.2% increase in the odds of choosing the right side in Experiment 3a and an 10.7% increase in the odds of choosing the right side in Experiment 3b. Taken together, the results of the intact-object conditions replicated all the essential patterns observed in Experiment 1.

Scrambled-Object Condition. One-sample t tests showed that the coefficients of numerical ratio were significantly above zero in the scrambled-object conditions of both experiments (Experiment 3a—box scrambling: β_{num} $M = 3.65$, $SE = 0.12$, 95% CI [3.40, 3.90], $t_{50} = 29.38$, $p < .001$; Experiment 3b—diffeomorphic scrambling: β_{num} $M = 4.00$, $SE = 0.09$, 95% CI [3.83, 4.18], $t_{61} = 45.47$, $p < .001$). The coefficients for contextual coherence difference were also significantly positive in both experiments (Experiment 3a—box scrambling: β_{coh} $M = 0.05$, $SE = 0.01$, 95% CI [0.03, 0.07], $t_{50} = 4.69$, $p < .001$; Experiment 3b—diffeomorphic scrambling: β_{coh} $M = 0.05$, $SE = 0.01$, 95% CI [0.03, 0.08], $t_{61} = 4.70$, $p < .001$). Thirty-nine of 51 participants (76.47%) in Experiment 3a and 42 of 62 participants (67.74%) in Experiment 3b showed a positive β_{coh}. A binominal GLMM with participant treated as a random effect revealed similar results. The numerical ratio and difference in contextual coherence were both significant predictors of the item-level responses of choosing the right side (Experiment 3a—box scrambling: $\beta_{num} = 6.36$, $SE = 0.22$, $Z = 28.45$, $p < .001$, 95% CI [5.92, 6.80]; $\beta_{coh} = 0.07$, $SE = 0.02$, $Z = 3.14$, $p < .01$, 95% CI [0.03, 0.11]; Experiment 3b—diffeomorphic scrambling: $\beta_{num} = 7.02$, $SE = 0.16$, $Z = 42.85$, $p < .001$, 95% CI [6.70, 7.34]; $\beta_{coh} = 0.09$, $SE = 0.02$, $Z = 4.35$, $p < .001$, 95% CI [0.05, 0.13]). One unit increase in the right-to-left difference in contextual coherence was associated with a 7.3% increase in the odds of choosing the right side in Experiment 3a and a 9.2% increase in the odds of choosing the right side in Experiment 3b.

Comparison Between Intact and Scrambled Conditions. To investigate whether the low-level visual processing could fully account for the contextual coherence effect observed in the intact-object condition, we then compared the magnitude of the contextual coherence effect quantified by β_{coh} between the intact-object condition and the scrambled condition. In Experiment 3a, pairwise *t* test revealed no significant difference in β_{coh} between the box scrambling condition ($M = 0.048$, $SD = 0.09$) and the intact condition ($M = 0.046$, $SD = 0.07$), $t(50) = 0.13$, $p = .90$, Cohen's $d = 0.02$, as can be seen in Figure 6a. Bayesian analyses showed moderate evidence in favor of the null hypothesis ($BF_{10} = 0.15$), suggesting that the contextual coherence effect (β_{coh}) of the intact condition was not different from that of the box scrambling condition.

In Experiment 3b, again we found no significant difference in β_{coh} between the diffeomorphic scrambling condition ($M = 0.054$, $SD = 0.09$) and the intact condition ($M = 0.062$, $SD = 0.09$), $t(61) = 0.516$, $p = .608$, Cohen's $d = 0.066$, pairwise *t* test (see Figure 6b). Bayesian analyses again revealed moderate evidence for the null hypothesis over the alternative hypothesis ($BF_{10} = 0.16$), indicating that the magnitude of the contextual coherence effect observed in the diffeomorphic scrambling condition was not reduced compared with the intact-object condition.

Experiment 3c: Image Identification Results

The interrater reliability was calculated using Cohen's κ. The calculated κ coefficient was 0.91, indicating that the two raters were quite consistent. To determine the trial-level labeling accuracy, we adopted a lenient criterion from previous research: if a response received a correct rating from either of the two raters, it was considered as correct (Lin et al., 2021). This ensured that the labeling scores showed the upper estimate of people's ability to recognize the images.

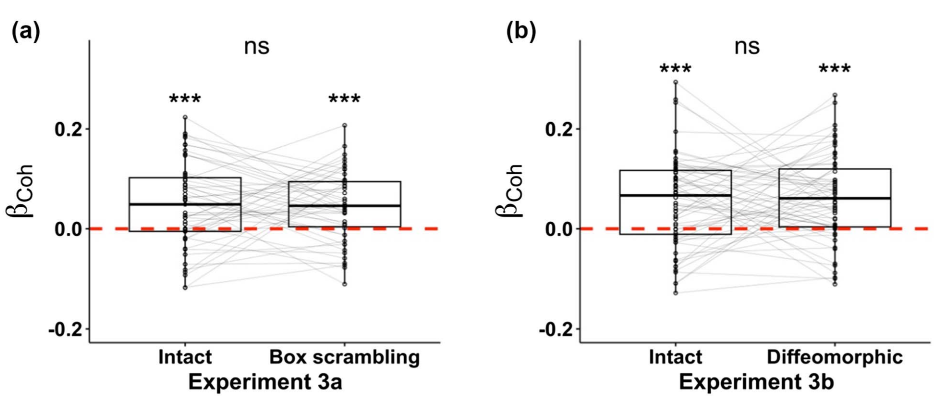
We first calculated a correct labeling score for each image as the percentage of responses rated as correct. Figure 7a shows the relative frequency distribution of the correct labeling scores for intact images and diffeomorphic images. People correctly identified most images in the intact condition ($M = 91.9\%$, $SD = 0.13$), whereas people failed to identify most images in the diffeomorphic scrambled condition ($M = 7.8\%$, $SD = 0.13$). Independent samples *t* test revealed that mean labeling accuracy for diffeomorphic images was significantly lower than for the intact images, $t(39) = 43.71$, $p < .0001$, Cohen's $d = 13.9$, see Figure 7b. We next conducted a Pearson correlation to examine the relationship between labeling accuracy and participants' self-reported recognition ability (yes/no) in the diffeomorphic condition, after removing three outliers with *z* scores beyond 2 *SD*s. There was no correlation between participants' labeling accuracy in the diffeomorphic condition and their self-reported ability to recognize the images ($r = 0.18$, $p = .11$). This implies that, even for the small number of correct responses to scrambled images, participants did not express greater confidence than when their responses were incorrect. Together, these results suggest that the diffeomorphic scrambling transformation effectively eliminated the ability to identify or semantically interpret the images.

General Discussion

Numerosity is an abstract feature of a set and number perception emerges from multiple stages of visual processing (Harvey & Dumoulin, 2017; Paul et al., 2022). By examining when in the visual processing hierarchy systematic biases in numerosity perception emerge, we can learn more about the different stages of the visual processing hierarchy in shaping numerosity representations. Our prior work demonstrated that reducing coherence of low-level visual

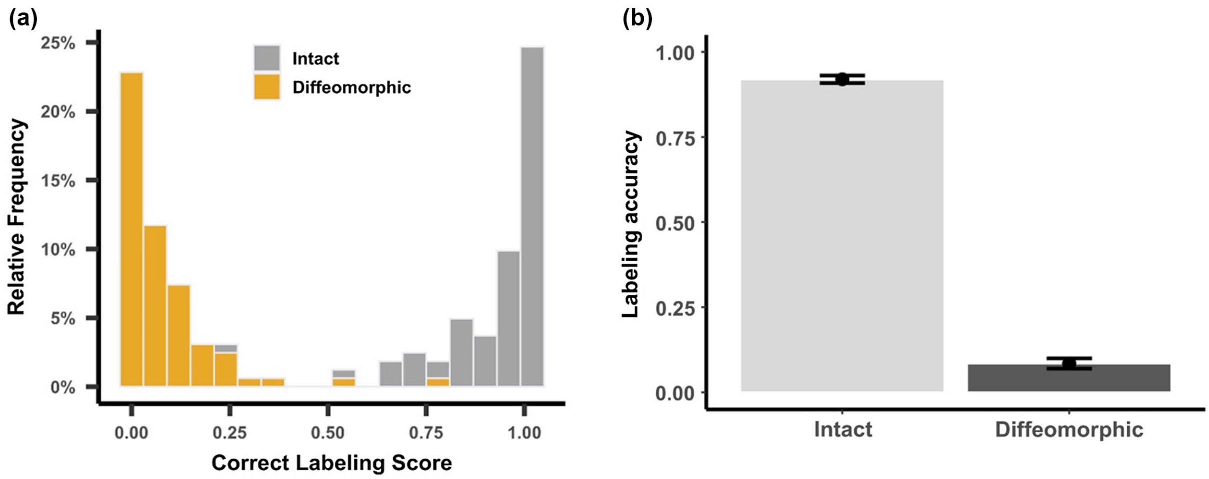
Figure 6

The Effect of Contextual Coherence (Quantified by the Regression Coefficient β_{coh}) for (a) Intact and Box Scrambling Conditions and (b) Intact and Diffeomorphic Scrambling Conditions



Note. In each box, the central band is the median, the edges of the box are the first and third quartiles, and the whiskers indicate ±1.5 × interquartile range. Asterisk indicates a significant effect of contextual coherence for each condition. *ns* indicates no significant difference between intact and each of the two scrambling conditions. Dots and lines represent participants. See the online article for the color version of this figure.

Figure 7
Results of Image Identification



Note. (a) Frequency distribution of correct labeling scores for diffeomorphic scrambled images and intact images. (b) Bar plot of mean labeling accuracy for diffeomorphic and intact conditions. All error bars are standard error of the mean. See the online article for the color version of this figure.

features such as color and orientation systematically reduces perceived numerosity (DeWind et al., 2020; Qu et al., 2022). Here, we asked whether the coherence effect is exclusive to low-level visual features or instead whether it can be driven by contextual or semantic relationships. To this end, we tested adults in an ordinal numerical comparison task and mathematically manipulated contextual coherence using a statistical model of visual object co-occurrence ("object2vec"; Bonner & Epstein, 2021).

In an initial experiment, we found that arrays composed of items with high statistical co-occurrence were indeed perceived as more numerous than arrays with low statistical co-occurrence. This contextual coherence effect could have been driven by top-down semantic information. However, in two additional studies, we found evidence against this interpretation. In Experiment 2, we found that texform images that preserved midlevel visual features were sufficient to elicit the contextual coherence effect, and there was in fact no difference in the strength of the contextual coherence effect between intact and texform conditions.

Given that prior studies have shown that visual features that were preserved in unrecognizable texforms contain reliable cues to animacy and real-world object size (Long et al., 2016, 2017), we considered the possibility that low-level visual features actually cue higher level properties. For example, categories such as faces and bikes are relatively invulnerable to scrambling; thus, it is likely that visual cues retained in scrambling conditions might result in residual identification (Stojanoski & Cusack, 2014). This raises the possibility that the contextual coherence effect obtained in Experiment 2 was driven by a combination of low-level visual features and higher level semantic properties cued by these visual features. In Experiment 3, we used box scrambling and diffeomorphic scrambling to eliminate such categorical information. A control study (Experiment 3c) confirmed that the diffeomorphic scrambling was effective in eliminating the participant's capacity to label the objects. Importantly, not only did the arrays composed of scrambled object images still elicit the coherence effect but there was no attenuation in the strength of the effect compared to the intact images. Thus, low-level visual features appear to fully account for the contextual coherence effect observed throughout our three studies.

The most parsimonious explanation of our results is that the coherence effect is driven by low-level perceptual features. Additional research is necessary to pinpoint precisely which visual features drive the contextual coherence effect we obtained. For example, although objects that statistically co-occur together generally have higher color similarity than objects that do not co-occur, we found that contextual coherence still significantly affected perceived numerosity even when all images were converted to grayscale and equalized across luminance and contrast (Experiment 2). Thus, although the prior work shows that color is sufficient to elicit the coherence effect, the present work demonstrates that color is not necessary to obtain the effect (Qu et al., 2022). Second, although objects from similar taxonomic categories tend to have similar shapes, the contextual coherence effect cannot be solely attributed to the perceived curvature or the global forms of objects given that scrambling removed such curvature and the contextual coherence effect remained (Stojanoski & Cusack, 2014; Vogels, 1999).

Our results suggest that low-level visual features (e.g., texture, spatial frequency) might be shared by objects that co-occur together in nature. For example, objects that tend to co-occur in natural scenes like flowers, trees, and shrubs may have fewer straight edges and lower diversity in hue compared to objects frequently seen in urban scenes. This is supported by studies that find that our subjective perception of the naturalness of visual context can be predicted by low-level visual features such as the density of contrast changes in the scene, the density of straight lines in the scene, the average color saturation in the scene, and the average hue diversity in the scene (Berman et al., 2014). Indeed, although high-level visual areas in the ventral temporal cortex are thought to be tightly linked to categorical or semantic content of images (Clarke & Tyler, 2014; Connolly et al., 2012; Kriegeskorte et al., 2008), recent neuroimaging data has shown that similarities in low-level image properties like spatial frequency and orientation could predict similarities in category-selective patterns of responses within the ventral stream, suggesting that the patterns of neural responses for object categories in high-level visual areas might be related to lower level image features rather than abstract semantic categories (Andrews et al., 2015). Because we observed coherence effects for texforms and scrambled images, which do not contain higher level category information, our findings suggest that co-occurring object categories may tend to have shared lower level features. Future work could explore the degree to which lower level image features and higher level semantic categories explain unique variance in the effects of visual object context.

Why does entropy reduce perceived numerosity or, conversely, why does coherence increase perceived numerosity? We are currently testing the hypothesis that the coherence effect is a result of attention being systematically impacted by entropy in low-level visual features. Under a limited capacity cognitive system, elements within a visual scene compete for visual attention (Desimone & Duncan, 1995; Kastner & Ungerleider, 2001). Computational models of visual attention imply that the visual system computes the differences between spatially neighboring elements via a pre-attentive, massively parallel stage that extracts early visual features (e.g., color, orientation; Itti & Koch, 2001; Wolfe, 1994; Wolfe & Horowitz, 2004). Within the visual field, regions with high differences in early visual features between neighboring items are associated with high local contrast, which leads to greater neural activation at early visual processing and thus guides focal attention (Itti & Koch, 2001; Wolfe, 1994). As such, arrays with high entropy or variance in low-level features may have more high-contrast regions, which consequently garner more focal attention toward individual items that comprise the arrays (Utochkin & Yurevich, 2016). Conversely, arrays with high visual coherence may elicit more evenly distributed attention to the individual segmented elements that comprise an array.

Array heterogeneity decreases precision in representing summary statistics of a set (Marchant et al., 2013). Relatedly, in a dual task situation when the secondary task requires focal or distributed attention, this differentially impacts precision in statistical encoding. Participants are more precise at encoding summary statistics when their secondary task requires distributed attention than when the secondary task requires more focal attention (Chong & Treisman, 2005). Following those findings, we hypothesize that a potential explanation for our results may be that array heterogeneity engages more focal attentional processes. Future manipulations of attentional processes and perhaps eye tracking could be helpful in testing this hypothesis.

Using computational modeling of 7 T fMRI data, Paul et al. (2022) showed that the monotonic responses of V1 populations follow aggregate Fourier power in the spatial frequency domain more closely than numerosity, while truly numerosity-tuned responses emerge in the lateral occipital cortex. As originally suggested by the Dehaene and Changeux (1993) model, contrast normalization may transform early visual Fourier power responses into numerosity-tuned responses (Dehaene & Changeux, 1993; Paul et al., 2022). One possibility that deserves further research is that high entropy in local image contrast within a display impairs global normalization processes which in turn reduces performance in numerosity discrimination and leads to underestimation (Morgan et al., 2014; Paul et al., 2022).

In summary, our findings demonstrate that visual coherence derived from natural statistics of object co-occurrence systematically alters perceived numerosity. This contextual coherence effect was not attenuated even in texforms and scrambled images that have been largely deprived of their semantic content, suggesting that this effect emerged at low-level visual processing stages, even before downstream processing steps at which items can be explicitly recognized and categorized. Future research is needed to identify the specific low-level features that give rise to the contextual coherence effect. A fine-grained characterization of how systematic biases in perception occurs within a hierarchy of increasingly complex feature representations will help to uncover the multistage computational processing underlying our approximate number sense and more broadly, the formation of high-level abstract visual features.

Constraints on Generality

For all experiments, participants were recruited online through Prolific. Consequently, the characteristics of our participants—specifically, individuals in early to midadulthood in the United States with internet access and a computing device—constrain the generality of our findings. For example, it remains an open question whether the effect of contextual coherence on perceived numerosity extends outside this age range or to individuals from other countries, or socioeconomic backgrounds. Given the universality of our visual number sense and the lack of semantic processing requirement for the coherence effect, we speculate that our results have potential for broader applicability. However, firm conclusions on these matters require empirical evidence.

References

Adriano, A., Girelli, L., & Rinaldi, L. (2021). Non-Symbolic numerosity encoding escapes spatial frequency equalization. *Psychological Research*, *85*(8), 3061–3074. https://doi.org/10.1007/s00426-020-01458-2

Adriano, A., Rinaldi, L., & Girelli, L. (2021). Visual illusions as a tool to hijack numerical perception: Disentangling nonsymbolic number from its continuous visual properties. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(3), 423. https://doi.org/10.1037/xhp0000844

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

Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *Journal of Vision*, *15*(7), Article 3. https://doi.org/10.1167/15.7.3

Barth, H., Beckmann, L., & Spelke, E. S. (2008). Nonsymbolic, approximate arithmetic in children: Abstract addition prior to instruction. *Developmental Psychology*, *44*(5), 1466–1477. https://doi.org/10.1037/a0013046

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, *98*(3), 199–222. https://doi.org/10.1016/j.cognition.2004.09.011

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

Berman, M. G., Hout, M. C., Kardan, O., Hunter, M. R., Yourganov, G., Henderson, J. M., Hanayik, T., Karimi, H., & Jonides, J. (2014). The perception of naturalness correlates with low-level visual features of environmental scenes. *PLOS ONE*, *9*(12), Article e114572. https://doi.org/10.1371/journal.pone.0114572

Bonner, M. F., & Epstein, R. A. (2021). Object representations in the human brain reflect the co-occurrence statistics of vision and language. *Nature*

Communications, *12*(1), Article 4081. https://doi.org/10.1038/s41467-021-24368-2

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. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science*, *17*(5), 401–406. https://doi.org/10.1111/j.1467-9280.2006.01719.x

Castaldi, E., Piazza, M., Dehaene, S., Vignaud, A., & Eger, E. (2019). Attentional amplification of neural codes for number independent of other quantities along the dorsal visual stream. *eLife*, *8*, Article e45160. https://doi.org/10.7554/eLife.45160

Castaldi, E., Pomè, A., Cicchini, G. M., Burr, D., & Binda, P. (2021). The pupil responds spontaneously to perceived numerosity. *Nature Communications*, *12*(1), Article 5944. https://doi.org/10.1038/s41467-021-26261-4

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, *148*, 163–172. https://doi.org/10.1016/j.actpsy.2014.01.016

Chen, Y.-C., Deza, A., & Konkle, T. (2022). How big should this object be? Perceptual influences on viewing-size preferences. *Cognition*, *225*, Article 105114. https://doi.org/10.1016/j.cognition.2022.105114

Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*(1), 1–13. https://doi.org/10.3758/BF03195009

Clarke, A., & Tyler, L. K. (2014). Object-specific semantic coding in human perirhinal cortex. *The Journal of Neuroscience*, *34*(14), 4766–4775. https://doi.org/10.1523/JNEUROSCI.2828-13.2014

Clarke, S., & Beck, J. (2021). The number sense represents (rational) numbers. *Behavioral and Brain Sciences*, *44*, e178. https://doi.org/10.1017/S0140525X21000571

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., Abdi, H., & Haxby, J. V. (2012). The representation of biological classes in the human brain. *The Journal of Neuroscience*, *32*(8), 2608–2618. https://doi.org/10.1523/JNEUROSCI.5547-11.2012

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1–2), 1–42. https://doi.org/10.1016/0010-0277(92)90049-N

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

Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, *5*(4), 390–407. https://doi.org/10.1162/jocn.1993.5.4.390

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

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

DeWind, N. K., Bonner, M. F., & Brannon, E. M. (2020). Similarly oriented objects appear more numerous. *Journal of Vision*, *20*(4), Article 4. https://doi.org/10.1167/jov.20.4.4

DeWind, N. K., Park, J., Woldorff, M. G., & Brannon, E. M. (2019). Numerical encoding in early visual cortex. *Cortex*, *114*, 76–89. https://doi.org/10.1016/j.cortex.2018.03.027

Deza, A., Chen, Y.-C., Long, B., & Konkle, T. (2019, September 15). *Accelerated texforms: Alternative methods for generating unrecognizable object images with preserved mid-level features* [Paper presentation]. Cognitive Computational Neuroscience Conference, Berlin, Germany. https://ccneuro.org/2019/proceedings/0000879.pdf

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, Article 781. https://doi.org/10.3389/fpsyg.2014.00781

Dilkina, K., & Lambon Ralph, M. A. (2013). Conceptual structure within and between modalities. *Frontiers in Human Neuroscience*, *6*, Article 333. https://doi.org/10.3389/fnhum.2012.00333

Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., & Kleinschmidt, A. (2009). Deciphering cortical number coding from human brain activity patterns. *Current Biology*, *19*(19), 1608–1615. https://doi.org/10.1016/j.cub.2009.08.047

Fornaciai, M., & Park, J. (2018). Early numerosity encoding in visual cortex is not sufficient for the representation of numerical magnitude. *Journal of Cognitive Neuroscience*, *30*(12), 1788–1802. https://doi.org/10.1162/jocn_a_01320

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

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201. https://doi.org/10.1038/nn.2889

Ginsburg, N. (1976). Effect of item arrangement on perceived numerosity: Randomness vs regularity. *Perceptual and Motor Skills*, *43*(2), 663–668. https://doi.org/10.2466/pms.1976.43.2.663

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision*, *15*(4), Article 16. https://doi.org/10.1167/15.4.16

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

Harvey, B. M., & Dumoulin, S. O. (2017). A network of topographic numerosity maps in human association cortex. *Nature Human Behaviour*, *1*(2), Article 0036. https://doi.org/10.1038/s41562-016-0036

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

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 31–47. https://doi.org/10.1037/0096-1523.33.1.31

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203. https://doi.org/10.1038/35058500

Kastner, S., & Ungerleider, L. G. (2001). The neural basis of biased competition in human visual cortex. *Neuropsychologia*, *39*(12), 1263–1276. https://doi.org/10.1016/S0028-3932(01)00116-6

Kirjakovski, A., & Matsumoto, E. (2016). Numerosity underestimation in sets with illusory contours. *Vision Research*, *122*, 34–42. https://doi.org/10.1016/j.visres.2016.03.005

Konkle, T., & Oliva, A. (2012). A familiar-size Stroop effect: Real-world size is an automatic property of object representation. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 561–569. https://doi.org/10.1037/a0028294

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, Article 4. https://doi.org/10.3389/neuro.06.004.2008

Lin, Q., Yousif, S. R., Chun, M. M., & Scholl, B. J. (2021). Visual memorability in the absence of semantic content. *Cognition*, *212*, Article 104714. https://doi.org/10.1016/j.cognition.2021.104714

Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, *240*(4853), 740–749. https://doi.org/10.1126/science.3283936

Long, B., & Konkle, T. (2017). A familiar-size Stroop effect in the absence of basic-level recognition. *Cognition*, *168*, 234–242. https://doi.org/10.1016/j.cognition.2017.06.025

Long, B., Konkle, T., Cohen, M. A., & Alvarez, G. A. (2016). Mid-level perceptual features distinguish objects of different real-world sizes. *Journal of Experimental Psychology: General*, *145*(1), 95–109. https://doi.org/10.1037/xge0000130

Long, B., Störmer, V. S., & Alvarez, G. A. (2017). Mid-level perceptual features contain early cues to animacy. *Journal of Vision*, *17*(6), Article 20. https://doi.org/10.1167/17.6.20

Long, B., Yu, C. P., & Konkle, T. (2018). Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(38), E9015–E9024. https://doi.org/10.1073/pnas.1719616115

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

Maule, J., & Franklin, A. (2020). Adaptation to variance generalizes across visual domains. *Journal of Experimental Psychology: General*, *149*(4), 662–675. https://doi.org/10.1037/xge0000678

Morgan, M. J., Raphael, S., Tibber, M. S., & Dakin, S. C. (2014). A texture-processing model of the 'visual sense of number'. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1790), Article 20141137. https://doi.org/10.1098/rspb.2014.1137

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, *32*(1), 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550

Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science*, *297*(5587), 1708–1711. https://doi.org/10.1126/science.1072493

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

Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(19), 7457–7462. https://doi.org/10.1073/pnas.0402239101

Norman, L. J., Heywood, C. A., & Kentridge, R. W. (2015). Direct encoding of orientation variance in the visual system. *Journal of Vision*, *15*(4), Article 3. https://doi.org/10.1167/15.4.3

Palmer, t. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*(5), 519–526. https://doi.org/10.3758/BF03197524

Park, J., DeWind, N. K., Woldorff, M. G., & Brannon, E. M. (2016). Rapid and direct encoding of numerosity in the visual stream. *Cerebral Cortex*, *26*(2), 748–763. https://doi.org/10.1093/cercor/bhv017

Park, J., & Huber, D. E. (2022). A visual sense of number emerges from divisive normalization in a simple center-surround convolutional network. *Elife*, *11*, e80990. https://doi.org/10.7554/eLife.80990

Paul, J. M., van Ackooij, M., ten Cate, T. C., & Harvey, B. M. (2022). Numerosity tuning in human association cortices and local image contrast representations in early visual cortex. *Nature Communications*, *13*(1), Article 1340. https://doi.org/10.1038/s41467-022-29030-z

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555. https://doi.org/10.1016/j.neuron.2004.10.014

Piazza, M., Mechelli, A., Price, C. J., & Butterworth, B. (2006). Exact and approximate judgements of visual and auditory numerosity: An fMRI study. *Brain Research*, *1106*(1), 177–188. https://doi.org/10.1016/j.brainres.2006.05.104

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499–503. https://doi.org/10.1126/science.1102085

Priebe, N. J. (2016). Mechanisms of orientation selectivity in the primary visual cortex. *Annual Review of Vision Science*, *2*(1), 85–107. https://doi.org/10.1146/annurev-vision-111815-114456

Qu, C., DeWind, N. K., & Brannon, E. M. (2022). Increasing entropy reduces perceived numerosity throughout the lifespan. *Cognition*, *225*, Article 105096. https://doi.org/10.1016/j.cognition.2022.105096

Qu, C., Szkudlarek, E., & Brannon, E. M. (2021). Approximate multiplication in young children prior to multiplication instruction. *Journal of Experimental Child Psychology*, *207*, Article 105116. https://doi.org/10.1016/j.jecp.2021.105116

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Sadeghi, Z., McClelland, J. L., & Hoffman, P. (2015). You shall know an object by the company it keeps: An investigation of semantic representations derived from object co-occurrence in visual scenes. *Neuropsychologia*, *76*, 52–61. https://doi.org/10.1016/j.neuropsychologia.2014.08.031

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Skagenholt, M., Träff, U., Västfjäll, D., & Skagerlund, K. (2018). Examining the triple code model in numerical cognition: An fMRI study. *PLOS ONE*, *13*(6), Article e0199247. https://doi.org/10.1371/journal.pone.0199247

Solomon, S. G., & Lennie, P. (2007). The machinery of colour vision. *Nature Reviews Neuroscience*, *8*(4), 276–286. https://doi.org/10.1038/nrn2094

Stojanoski, B., & Cusack, R. (2014). Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *Journal of Vision*, *14*(12), Article 6. https://doi.org/10.1167/14.12.6

Szkudlarek, E., & Brannon, E. M. (2017). Does the approximate number system serve as a foundation for symbolic mathematics? *Language Learning and Development*, *13*(2), 171–190. https://doi.org/10.1080/15475441.2016.1263573

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

Utochkin, I. S., & Yurevich, M. A. (2016). Similarity and heterogeneity effects in visual search are mediated by "segmentability". *Journal of Experimental Psychology: Human Perception and Performance*, *42*(7), 995–1007. https://doi.org/10.1037/xhp0000203

Vogels, R. (1999). Effect of image scrambling on inferior temporal cortical responses. *Neuroreport*, *10*(9), 1811–1816. https://doi.org/10.1097/00001756-199906230-00002

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. https://doi.org/10.3758/BF03200774

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495–501. https://doi.org/10.1038/nrn1411

Yang, Y., Tokita, M., & Ishiguchi, A. (2018). Is there a common summary statistical process for representing the mean and variance? A study using illustrations of familiar items. *I-Perception*, *9*(1). https://doi.org/10.1177/2041669517747297

Received June 27, 2023
Revision received February 14, 2024
Accepted March 12, 2024 ■