When the weaker conquer: A contrast-dependent illusion of visual numerosity

Quan Lei
Department of Psychology, Northeastern University, Boston, MA, USA
Department of Psychology, University of Minnesota, Minneapolis, MN, USA

Adam Reeves
Department of Psychology, Northeastern University, Boston, MA, USA

Humans and many nonhuman species have the capacity to perceive the approximate numerosity of a large set of elements. In the visual domain, various factors have been found to affect perceived numerosity, including, as studied here, the effect of luminance contrast. We report a new numerosity illusion in which low-contrast elements appear to be more numerous than high-contrast ones when intermixed in the same display, an effect we have called “the weak conquer the strong.” The illusion occurred for both positive and negative contrast disks of the same sign; for example, white disks appeared less numerous than light gray disks when presented on a dark gray field (all positive contrasts). The illusion grew in size as the contrast difference between the two sets of disks became smaller but did not hold for disks with opposite signs, such as white and dark gray disks on a light gray field. The illusion was also eliminated when the disks were segregated spatially, in which case the high-contrast (white) disks and low-contrast (gray) disks appeared equally numerous. The illusion is due to the loss of some high-contrast elements; e.g., the perceived numerosity of the white disks was decreased whereas that of gray ones was preserved as shown by matching to isolated sets of white or gray disks. We speculate that failure in segregation coupled with error in contrast normalization may account for the numerosity illusion.

Introduction

Humans and other species alike are endowed with extraordinary numerical competence (Cantlon & Brannon, 2006; Dehaene, 2011; Scarf, Hayne, & Colombo, 2011), one aspect of which is the capability to approximate the numerosity of a large set of objects (Feigenson, Dehaene, & Spelke, 2004). Although controversy exists as to how large numerosities are extracted, whether directly through a dedicated mechanism (Anobile, Cicchini, & Burr, 2014; Burr, Anobile, & Arrighi, 2018; Burr & Ross, 2008; Cicchini, Anobile, & Burr, 2016; Dehaene, 2011) or indirectly by inferring from surrogate metrics, such as density and contrast energy (Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011; Durgin, 1995; Gebuis & Reynvoet, 2012), there is general consensus that numerosity is a readily perceivable dimension of a visual array despite wide variations in nonnumerical properties of individual elements. Most studies with the goal of tapping the so-called approximate number system have adopted discrimination tasks that involve relative judgment of numerosity of two patches of dots, presented either sequentially in different intervals or simultaneously at different spatial locations. Whereas numerosity discrimination is largely accurate, with its precision dictated by Weber’s law (Ross, 2003), various factors have been revealed to bias such judgment, including element size, aggregate surface, patch area or convex hull, and density (Dakin et al., 2011; Gebuis & Reynvoet, 2012; Ginsburg & Nicholls, 1988; Hurewitz, Gelman, & Schnitzer, 2006; Krueger, 1972; Miller & Baker, 1968; Sophian & Chu, 2008). Results have been inconsistent about the specific direction of the biases induced by these variables. For example, larger surface area was associated with larger numerosity in some studies (Hurewitz et al., 2006), but the opposite pattern was shown in others (Gebuis & Reynvoet, 2012). More robust biases, however, have been demonstrated to be induced by the spatial arrangement of elements, including both the spatial distribution of elements over a 2-D surface and the 3-D layout involving depth. For instance, clustering elements was found to reduce perceived numerosity (Ginsburg, 1991; Sophian & Chu, 2008) with a more clustered set appearing less
numerous, explaining such numerosity illusions as the regular–random illusion (Ginsburg, 1976; Messenger, 1903). Depth structure of a set biases numerosity so that a back surface appears to be more numerous than a front surface populated with an equal number of elements (Schütz, 2012; Tsirlin, Allison, & Wilcox, 2012), and the numerosity of elements distributed over two surfaces in depth is overestimated relative to that of elements on a single surface (Aida, Kusano, Shimono, & Tam, 2015). It is an open question whether these various observations reflect some sort of cognitive intervention at a later stage of numerosity judgment or instead speak downward to the core mechanisms of numerosity computation.

In a class of models proposing a dedicated mechanism for numerosity (e.g., Dehaene & Changeux, 1993; Trick & Pylyshyn, 1994), the segmentation of elements in a set, either from the background or from one another, constitutes an early stage during numerosity computation. To the extent that segmentation provides the prerequisite to numerosity processing, it is conceivable that factors limiting this early process would have profound impacts on numerosity perception. For example, the perceived numerosity of a small array is found to be reduced by lowering element contrast toward the threshold of visibility (Palomares & Egeth, 2010). However, the effect of suprathreshold contrast on numerosity perception is less clear. If numerosity is indeed a fundamental attribute of perception (Burr et al., 2018; Cicchini et al., 2016), to the extent it is also orthogonal to other perceptual attributes, it should show at least approximate constancy; that is, two displays of equal numbers of elements should appear equally numerous despite variations in irrelevant aspects, such as the shape, size, location, and also the contrast of the elements, just as other perceptual attributes, such as size, shape, and color, show approximate constancy. Indeed, in most cases, such displays do show approximate numerosity constancy despite the biasing factors mentioned above. However, if the computation of numerosity takes as input a raw image with a certain contrast pattern rather than a collection of segmented objects and performs some sort of contrast energy analysis as suggested in several models (Dakin et al., 2011; Morgan, Raphael, Tibber, & Dakin, 2014), then stimulus contrast might well impact numerosity representation. An analogy to this latter piece of reasoning can be found in the domain of motion perception, in which the effects of contrast on perceived motion direction and velocity are well documented (Blakemore & Snowden, 1999; Stone & Thompson, 1992; Stone, Watson, & Mulligan, 1990). Given that contrast is a fundamental attribute of almost any visual stimulus and presumably forms the input for numerosity computation, either as segmented objects shaped by contrast or as an unsegmented raw image with contrast energy, the literature contains surprisingly few mentions of any effect of contrast on numerosity perception, in particular the apprehension of large numerosities. Exceptionally, Ross and Burr (2010) found that displays of dots with lower luminances had greater perceived numerosity than displays of dots with higher luminances. However, dot luminances in the study spanned photopic, mesopic, and even scotopic levels, so the input receptor class (and, therefore, retinal response distribution) varied, making the findings difficult to interpret. Additionally, because they have varied the luminance of the dots above a dark field, the relative roles of contrast and luminance are unclear.

In the present study, a numerosity illusion was reported that clearly showed an effect of suprathreshold luminance contrast on numerosity perception under quite different conditions from those of Ross and Burr (2010). We used photopic stimuli and gray backgrounds so that cone contrast, not just luminance, can be determined. Critically, instead of presenting segregated dot arrays, we have intermixed elements of different luminances (contrasts) within a single patch. The illusion that we first report is that when equal numbers of white disks and light gray disks are intermixed on a dark gray background, there appear to be more light gray disks than white ones (Figure 1A). In several experiments, attempts were made to quantify and characterize the illusion by examining its preconditions and probing the nature of the interaction that leads to the illusion. In Experiments 1A and 1B, disks all of the same contrast polarity, either positive or negative, were used as stimuli in a numerosity discrimination task to test the generality of the contrast-based illusion. In Experiments 2A and 2B, the contrast polarities of the two sets of disks were manipulated to be either the same or opposite to each other, and the difference in contrast magnitude between the two sets was varied to investigate how the illusion depends on polarity congruence and relative contrast. In Experiment 3, as opposed to intermixing them, the two disk sets were spatially segregated to examine if a contrast difference alone is sufficient to induce the illusion. In Experiments 4A and 4B, by employing a subset and a superset discrimination task, respectively, we set out to determine the genesis of the illusion. In the subset task, we asked how the perceived numerosity of each disk set was affected by intermixing with the other set compared to when the set was presented alone, and in the superset task, we asked how the perceived total numerosity of the intermixed patch would compare to a homogenous counterpart. The results from these experiments served as evidence to support or reject several potential explanations of the illusion, such as misclassification and occlusion. To the best of our knowledge, the numerosity illusion reported...
in this study is a novel one that cannot be satisfactorily explained by any existing theories of numerosity perception in their current forms (as explained in the General discussion), thus providing new opportunities for revisiting the computations underlying our “number sense.”

**General methods**

**Subjects**

Observers in all experiments were students enrolled in an introductory psychology course and participated for course credit. They all had normal or corrected-to-normal vision and were naïve as to the purpose of the experiments. Written consent was obtained from the subjects before the experiments. The study was approved by the institutional review board at Northeastern University.

**Stimuli**

The stimuli in all experiments consisted of disks (0.5° in diameter) distributed over an imaginary square patch measuring 10° each side (Figure 1A). The patch was divided into a 12 by 12 grid of cells. On each trial, a subset of the 144 cells was randomly selected to place the disks. To increase randomness, the actual location of each disk was jittered both horizontally and vertically as much as 0.2° about the center of its cell. In Experiments 1 and 2, two sets of disks with different contrasts were intermixed in a single patch; the patch was presented at the center of the screen and subjects discriminated the numerosities of the two sets in that patch. In Experiments 3 and 4, two disk patches were used, one presented to the left and the other to the right of the screen center, the inner edge of each patch being 1° and the outer edge 11° from the screen center; subjects discriminated numerosities across the two patches. The stimuli were generated using Matlab in conjunction with Psychtoolbox (Brainard, 1997) and presented on a Dell LCD monitor viewed at 60 cm. The luminances of the disks and field were varied across experiments to vary contrasts as noted for each experiment. Because the disks were small relative to the field, physical contrasts are defined in Weberian terms as $\delta L/L$, where $L$ is the field luminance and $\delta L$ is the luminance difference between the disks and the field.

**Procedures**

Subjects completed a numerosity discrimination task and were asked to judge the relative numerosity of two sets of disks, either within one single patch or across two patches. A constant stimulus method was employed. In each experiment, unless otherwise noted, one set of disks with a fixed contrast was designated as the
standard set and the other set with a different fixed contrast designated as the comparison. The standard set, used as reference, had a standard (constant) numerosity of 50. The numerosity of the comparison set was varied between 30 and 70 in steps of five, resulting in nine different levels. The test numerosity on each trial was randomly selected from these nine levels; each numerosity was judged against the standard numerosity and was repeated 10 times for each subject and experimental condition. A different spatial pattern was generated for each repetition of the same physical numerosity. Each trial began with a fixation cross presented at the center of the screen for 1 s, followed by the disk display for 1.5 s. The disk display was then replaced by a blank screen, upon which event subjects had to press one key to report perceiving more disks in one set (e.g., white disks) than in the other set (e.g., gray disks) and pressing another key for the reverse. The next trial started 3 s after subjects had responded. Eye movements were not tracked, and subjects were allowed to freely scan the disk display.

Data analysis

For data analysis, the percentage of the comparison set being selected as more numerous than the standard set was plotted against the physical numerosity of the comparison set and the data fit with a Weibull function separately for each experimental condition and each subject (Figure 1B). Fitting was implemented using the Psignifit toolbox, based on Bayesian methods (Schütz, Harmeling, Macke, & Wichmann, 2016). In this study, the point of subjective equality (PSE), which was derived as the physical numerosity of the comparison set that corresponds to a 50% probability on the Weibull curve, was used as the dependent variable in all experiments. A PSE less than 50 indicates an overestimation of the comparison set relative to the standard set, and a PSE greater than 50 indicates the opposite. In the example given in Figure 1B, 45.7 gray disks were matched to 50 white ones as if the gray disks were perceived as somewhat more numerous than the white ones and, therefore, had to be reduced in number to match them. We had no theoretical expectations concerning the slope of the Weibull curves; inspection revealed that the curves were generally steep. Therefore, for the purpose of this study, only PSEs were reported, but we refer to the slopes when they are relevant. For statistical analysis, a one-sample t test was conducted to test the significance of a difference between the mean PSE and the standard numerosity at the alpha level of 0.05; repeated-measure ANOVAs and other analyses were conducted as noted.

Experiment 1A: Positive contrast

As a first attempt to quantify the illusion, the discrimination task described in the General methods section was employed to measure how many light gray disks (the comparison set) would match 50 white disks (the standard set) in perceived numerosity when the two sets were intermixed on a dark gray background, both sets having positive contrast. The PSE was predicted to be somewhere below 50 to reflect the informal observation that gray disks appeared to be more numerous than white ones when there were equal numbers of them. However, there is one caveat. Because the numerosity of white disks, which served as the reference, was fixed at 50 throughout, numerosity adaptation to a constant value might reduce the perceived numerosity of white disks and, thus, potentially account for the result. This seemed unlikely, however, even though Burr and Ross (2008) demonstrated that numerosity can adapt, because most naïve observers appreciate the illusion on first presentation and because adaptation to the mean numerosity rather than to particular numerosity contrasts would affect gray and white disks equally. However, to assess whether numerosity adaptation might yet influence the illusion, in separate blocks, 50 gray disks served as the fixed reference while the white disks served as the comparison. If adaptation to a constant numerosity selectively reduces the perceived numerosity of the standard, then less than 50 white disks should now be required to match the standard 50 gray disks. If, on the other hand, the illusion is independent of adaptation and whichever set serves as the standard, we should expect the same pattern of perceiving more gray disks than white ones in both reference conditions; moreover, the absolute size of the illusion should correlate between the two conditions.

Methods

Eight subjects participated in the experiment. Disks of two different contrasts were intermixed on a dark gray background with a luminance of 18.2 cd/m². The luminance of the low-contrast (gray) and the high-contrast (white) disks was 26.1 cd/m² and 75.1 cd/m², respectively. The Weber contrasts were (26.1 – 18.2)/18.2 = 0.43 for the gray disks and (75.1 – 18.2)/18.2 = 3.13 for the white disks, both well above threshold. A single patch of intermixed disks was used for numerosity judgments.

The procedure as described in the General methods was followed except that in separate blocks either white disks or gray disks served as the standard (reference) set with a fixed numerosity of 50. There were 90 trials for each reference type, which were divided evenly into two
blocks with a short break in between. Half of the subjects completed the white reference blocks first and then the gray reference blocks, and the other half received the opposite order. For both reference types, the subjects performed the same forced-choice numerosity discrimination task, comparing the numerosity of white versus gray disks, so the manipulation of reference set was transparent to them. Before the experiment, subjects completed 10 trials of practice. The experiment lasted about 30 min.

Results and discussion

For both reference types, average PSEs across subjects were significantly different from the standard numerosity of 50 (Figure 2A), t(7) = 3.47, p = 0.010, and t(7) = 4.50, p = 0.003, for white and gray reference, respectively. On average, 45 gray disks matched 50 white disks in perceived numerosity in the white reference condition, and 59 white disks were required to match the 50 gray disks in the gray reference condition. The illusion held independent of which disk set served as the reference: in both cases, the numerosity of gray disks was overestimated relative to white ones. Therefore, adaptation to the standard numerosity cannot explain the main result although there was an effect of reference condition on the overall size of the illusion (10% for white reference vs. 15% for gray reference), t(7) = 3.06, p = 0.018. Inspecting individual subjects (Figure 2B), all but one had PSEs less than 50 in the white reference condition and PSEs greater than 50 in the gray reference condition, thus showing the same illusion (perceiving more gray disks than white ones) in both conditions. Subjects who experienced a strong (or weak) numerosity illusion in one condition also did so in the other condition, with a significant correlation of PSEs: r(6) = −0.85, p = 0.007, so the illusion was stable within individuals. Because adaptation to constant (as opposed to varied) numerosity was minor if any and because the PSEs were highly correlated between reference conditions, in subsequent experiments we only used one set of disks (usually white) as the reference.

Experiment 1B: Negative contrast

In Experiment 1A, both disk sets had positive contrast. Although the term “contrast” was used, the illusion can equally be interpreted as an effect of luminance because on a dark field those disks with higher contrast also had higher luminance and vice versa. So far, the effect we observed is qualitatively similar to that reported by Ross and Burr (2010) in that dimmer stimuli were reportedly more numerous than brighter ones. To disentangle the effect of contrast from that of luminance, in this experiment disks with negative contrast were used as stimuli. If the illusion measured in Experiment 1A is a luminance effect, black disks should appear more numerous than dark gray disks when equal numbers of them are intermixed on a light gray field, but if the illusion is instead a contrast effect, the opposite should occur.

Methods

Nine subjects participated in this experiment. As illustrated in Figure 3A, the disk display consisted of
black disks (2.0 cd/m²) and dark gray disks (8.2 cd/m²) intermixed on a light gray field (35.4 cd/m²). Black disks served as the reference and had a standard numerosity of 50, and the numerosity of gray disks varied on each trial between one of nine levels. The procedure was otherwise identical to that used in Experiment 1A. Subjects were instructed to press the left arrow key for perceiving more gray disks than black ones and press the right arrow key for the opposite. The experiment lasted about 15 min.

Results and discussion

All but one subject had PSEs below 50 (Figure 3B), so that the numerosity of gray disks were overestimated relative to black ones, indicating an illusion similar to that observed with positive contrast in which the disks with lower contrast (but now higher luminance) appeared to be more numerous. The average PSE of all subjects indicated that 44 gray disks matched 50 black disks in perceived numerosity, \( t(8) = 3.70, p = 0.006 \), amounting to an illusion of 12%, a magnitude quite similar to that measured in Experiment 1A, namely 10%. Experiments 1A and 1B together indicate that the numerosity illusion depends on stimulus contrast, not on luminance.

Experiment 2A: Same versus opposite contrast polarity

In Experiments 1A and 1B, the disks in a display all had the same sign of contrast, either positive or negative. It was unclear whether the numerosity illusion would hold for two sets of disks that differ not only in contrast magnitude, but also in polarity. Because rectification is a canonical operation in visual processing (Sperling, 1989), one might suggest that the relative magnitude of contrast between the two sets of disks is the key factor, and their respective polarities should not matter. On the other hand, polarity might have an effect if the illusion relies on an interaction between the two sets of items within the same (on or off) input channel before rectification takes place. In this experiment, two sets of disks with either the same or opposite contrast polarities were intermixed in a display to investigate the possible role of polarity congruency in the numerosity illusion. We also manipulated the contrast magnitude of the comparison disk set to examine the generality of the illusion over contrast variations.

Methods

Eight subjects participated in the experiment. The stimuli and procedure were identical to that of Experiment 1A except that the comparison set now had one of four luminances: 2.0, 11.8, 26.1, and 45.8 cd/m², chosen randomly on each trial. On a field of 18.2 cd/m², the corresponding Weber contrasts of the disks were \(-0.89, -0.35, 0.43, \) and \(1.52\), denoted as the high negative, low negative, low positive, and high positive conditions, respectively. The Weber contrast of the standard set, i.e., the white disks, remained 3.13, higher in magnitude than any of the comparison sets. Subjects were instructed to press one key for perceiving more white disks than gray (or black) disks and another key for the opposite. The experiment lasted about 1 hr.
Results and discussion

One subject was excluded from data analysis due to poor fitting of the data for all conditions. The average PSEs of the remaining subjects are shown in Figure 4. On average, 49.5 comparison disks in the high negative condition and 51.5 disks in the low negative condition matched 50 white disks in perceived numerosity; neither of the differences was significant, \( t(6) = 0.30, p = 0.776 \) and \( t(6) = 0.86, p = 0.423 \), respectively. For the low positive condition, 43.2 comparison disks matched 50 standard disks, the difference being significant, \( t(6) = 2.48, p = 0.048 \), which was a successful replication of the white reference condition in Experiment 1A. The PSE was 39.6 for the high positive condition, which was, however, not significant, \( t(6) = 2.38, p = 0.055 \). One subject was identified as an outlier for this particular condition, with a PSE of 62.4. This same subject also complained about having difficulty completing the task in the high positive condition, in which the contrasts of the two sets of disks were rather similar. If this subject was excluded, the average PSE of the remaining six subjects was 35.8 for the high positive condition, significantly different from 50, \( t(5) = 5.60, p = 0.003 \). A paired \( t \) test was then conducted to compare the size of the illusion between the low and high positive conditions; despite the seemingly larger effect in the high condition, the difference was not significant, \( t(5) = 2.16, p = 0.084 \). Furthermore, to confirm the effect of polarity congruency, a \( 2 \times 2 \) repeated-measure ANOVA was conducted, with polarity (positive vs. negative) and magnitude (low vs. high) as the within-subject factors. The main effect of polarity was significant, \( F(1, 6) = 12.07, p = 0.013 \); neither the main effect of magnitude nor the interaction between polarity and magnitude was significant.

These results indicate that the numerosity illusion only occurs for two sets of disks with the same sign of contrast and not for disks with opposite contrast polarities. Apparently, interaction among the elements within the same input channel contributes to the illusion. As is discussed later, by engaging different input channels, contrast polarity may have played a role in the segregation of the two sets of disks in the incongruent condition, thereby eliminating the illusion. It is worth noting that the slopes of the psychometric functions were steep and comparable between the two negative conditions and the low positive condition, so the different results cannot be attributed to a difference in task difficulty and associated change in precision of numerosity discrimination. However, we did observe a decrease in slope as the condition changed from low positive to high positive (mean slopes of 4.3% vs. 2.6% per disk), suggesting that the task became more difficult as the two sets of disks became closer in contrast. Nevertheless, the overall bias to perceive low-contrast elements as more numerous than high-contrast ones seems to be robust for two sets with the same polarity.

Experiment 2B: Effect of relative contrast

Experiment 2A clearly showed that the numerosity illusion is restricted to two sets of elements with the same contrast polarity. Although we also manipulated the contrast magnitude of the comparison set in a same-polarity display and found no evidence of a difference in illusion size between different contrast levels, the experiment was inconclusive regarding the effect of relative contrast between the comparison and the standard sets. As we noted, the slope of psychometric curves had dropped substantially, and subjects reported difficulty in completing the task as the contrast difference between the two sets became smaller. This is not too surprising given the nature of the task. Because contrast is the only feature subjects could rely on to distinguish the two sets, numerosity discrimination is inevitably limited by one’s sensitivity to tell the two contrasts apart, which decreases with diminishing contrast difference. It is possible that the high positive condition in Experiment 2A had approached the limit of a reliable segregation of the two disk sets based on contrast, which could explain the larger than normal individual differences among subjects in that condition. Despite this complication, there was indeed a trend toward a larger illusion as the
contrasts of the two sets became more similar. In Experiment 2B, we, therefore, manipulated the contrast magnitude of the comparison set over an expanded range to investigate how the illusion would change with the luminance contrast between the comparison and the standard sets while keeping in mind the issue of reliability associated with possible changes in the slope of psychometric curves.

**Methods**

Six subjects participated in this experiment. The stimuli and procedure were identical to that of Experiment 2A except that the field now had a luminance of 8.2 cd/m², and the comparison set had one of four luminances: 11.8, 21.6, 35.4, and 52.4 cd/m². The corresponding Weber contrasts of the comparisons were 0.44, 1.63, 3.32, and 5.39, all of positive sign. With a luminance of 75.1 cd/m², the standard set of white disks had a contrast of 8.16. The experiment lasted about 1 hr.

**Results and discussion**

PSEs were outside the stimulus range in three out of four conditions for one subject, indicating unreliable estimation, so this subject was excluded from further analysis. The average PSEs of the remaining subjects for each of the four contrast conditions are shown in Figure 5A. Apparently, as the contrast of the comparison set increased toward that of the standard set, there was an initial decrease and then a surge in PSE, indicating an initial increase in the magnitude of the illusion followed by a decrease or even a reversal. The mean PSEs were significantly different from the standard of 50 for the three lower contrasts, which were 43.3, $t(4) = 2.85, p = 0.046$; 39.4, $t(4) = 7.03, p = 0.002$; and 35.4, $t(4) = 28.65, p < 0.001$, respectively, from low to high contrast. For the highest contrast, the mean PSE, being 57.5, was not significant, $t(4) = 2.04, p = 0.111$. Inspection of individual psychometric functions reveals that fitting is generally poor for this condition as indicated by very shallow slopes (mean slope: 1.7%, compared to means of 4.9%, 4.0%, and 3.3%, respectively, for the three lower contrasts as shown in Figure 5B) and large deviance scores. A repeated-measure ANOVA was conducted, confirming the change in slope across contrast levels, $F(1.042, 4.169) = 10.03, p = 0.032$, the Greenhouse–Geisser correction being applied to the degrees of freedom. Pairwise comparisons with Bonferroni correction indicated that the slope for the highest contrast was significantly smaller than those for the lower contrasts whereas there were no significant differences in slope among the three lower contrasts. Therefore, subjects might have difficulty completing the task when the contrast difference between the two sets was too small and the PSE estimates are most likely unreliable for the highest contrast condition. This suggests that the numerosity illusion is limited by the discriminability of the constituent sets of elements.

To confirm the trend of an increase in illusion magnitude across the three lower contrasts, a repeated-measure ANOVA was conducted. Indeed, there was a significant difference in PSE among the three conditions, $F(1.164, 4.657) = 14.08, p = 0.014$, Greenhouse–Geisser correction applied. As mentioned above, there was a statistically nonsignificant decrease in slope of the psychometric functions across these three contrast levels as the contrast difference between the comparison and the standard became smaller. However, the fitted curves were still sufficiently steep to allow reliable
estimates of the PSEs so as not to obscure the major effect. Our results suggest that the numerosity illusion is parametrically modulated by the contrast difference between the two sets of elements in an intermixed display: the illusion increases in strength as the contrast difference decreases but to the extent that contrast-based segregation between the two sets is still possible.

**Experiment 3: Spatial segregation**

The experiments above have clearly showed that numerosity judgment is affected by element contrast. Low-contrast (gray) elements appeared to be more numerous than high-contrast (black or white) ones. In this experiment, two disk sets like those used in Experiment 1A were segregated into two patches occupying separate spatial regions to investigate whether a mere contrast difference (without intermixing) is sufficient to induce the illusion. We expect from numerosity constancy (see Introduction) that contrast per se would have no effect, just as long as all disks were distinctly visible.

**Methods**

Six subjects participated in this experiment. On each trial, two uniform patches of disks were presented, one gray and the other white, with one patch located to the left and the other to the right of the screen center (Figure 6A), the side of presentation being balanced across trials. Each disk patch was created in a similar fashion to those used in Experiment 1A. The comparison set of gray disks had one of two luminances: 26.1 and 45.8 cd/m², chosen randomly on each trial, and denoted, respectively, as having low or high contrast. The luminance of the standard set, the white disks, remained at 75.1 cd/m². All disks had positive contrasts on the dark gray (18.2 cd/m²) background. In the numerosity discrimination task, subjects were asked to pick the side with more disks by pressing the corresponding left or right arrow key. All other aspects of the experiment were identical to that of Experiment 1A. The experiment lasted about 30 min.

**Results and discussion**

PSEs averaged across subjects are shown as the segregated conditions in Figure 6B. Because the same contrast levels were also tested for intermixed patches in Experiment 2A, the results for the two contrast levels from that experiment are replotted here for comparison and are denoted as the intermixed conditions. Whereas a robust illusion was observed for intermixed patches, the illusion was eliminated altogether when the two disk sets were segregated into two separate patches. The PSEs for both contrast levels (51.1 for low and 51.4 for high) were not significantly different from the standard numerosity of 50, $t(5) = 2.01, p = 0.100$ and $t(5) = 2.06, p = 0.095$, for low and high, respectively. Therefore, subjects showed veridical comparative judgment of numerosity in the segregated conditions. To further confirm the effect of spatial segregation, a mixed-design ANOVA was conducted, with segregation condition (intermixed vs. segregated) as the between-subject and contrast level (low vs. high) as the within-subject factor. There was indeed a significant main effect of segregation, $F(1, 11) = 20.33$, $p = 0.001$. Neither the main effect of contrast level nor the interaction between contrast level and segregation was significant. These results suggest that the mere presence of a contrast difference between two sets is insufficient to induce the illusion and that intermixing may be necessary for it to occur.
Experiment 4A: Subset matching

In above experiments, an illusion was revealed in which low-contrast (gray) elements have higher apparent numerosity than high-contrast (white or black) elements if they all had the same contrast polarity. It was further shown that the illusion occurs only when the two sets of elements are intermixed, but not when they are spatially segregated. The question remains: What causes the numerosity illusion when the two sets of disks are not segregated? Here we highlight two possibilities: misclassification and occlusion. According to the misclassification hypothesis, to perform numerosity judgment with our stimulus configuration, subjects would have to classify each disk as belonging to the gray or the white set. When disks that only differ in contrast are intermixed, it is possible that some disks are misclassified; some gray disks are misclassified as white ones if at all and some white disks misclassified as gray with the latter occurring more frequently—hence, the illusion of more gray disks than white ones.

According to the occlusion hypothesis, contrast may serve as a depth cue in our display so that white disks tend to be seen in front and the gray disks at back, forming two surfaces separated in depth (e.g., Dresp-Langley & Reeves, 2014; O’Shea, Blackburn, & Ono, 1994). Because objects in the real world can occlude each other, the visual system may infer that some gray disks are occluded by white ones, and these inferred gray disks may be included with the visible ones for numerosity estimation. Therefore, both hypotheses may potentially explain the illusion.

This experiment was aimed at uncovering the interaction between gray and white elements while they are in an intermixture, the condition that gives rise to the numerosity illusion. Specifically, we are interested in knowing how the presence of one element set affects the perceived numerosity of the other set. Although both the misclassification and occlusion hypotheses predict the change in relative numerosity between the two sets, the former implies that the perceived numerosity of white disks would decrease while that of gray disks would increase when the two sets are intermixed whereas the latter implies that the perceived numerosity of gray disks would increase while that of white ones would remain unchanged. To explicitly test these predictions, the perceived numerosities of both the white and the gray disk sets in the intermixed patch were measured by selectively comparing each set to a set of disks presented in isolation in a second patch. Comparison with isolated elements is justified experimentally because Experiment 3 showed that segregated disks do not suffer from the numerosity illusion.

Methods

Six subjects participated in this experiment. The stimuli consisted of two patches of disks (Figure 7A). One patch, i.e., the intermixed patch, contained both gray and white disks created in the same fashion as the stimuli used in Experiment 1A. The numerosities of both gray disks and white disks were fixed at 50. The other patch, i.e., the unmixed patch, consisted of either gray or white disks that varied in numerosity. Subjects discriminated the numerosities of disks with the same color between the two patches. For this example, subjects were instructed to choose the patch with more white disks. (B) Mean PSEs ($n = 6$) are plotted for matching white and matching gray disks, showing how many disks in the unmixed patch matched 50 white or gray disks in the intermixed patch.
Results and discussion

The results are shown in Figure 7B. On average, 42 isolated white disks matched 50 intermixed white disks, the difference being significant, \( t(5) = 5.96, p = 0.002 \). In contrast, 51 isolated gray disks matched 50 intermixed gray disks with no significant difference, \( t(5) = 1.40, p = 0.221 \). A paired \( t \) test indicated that the PSE differed significantly between the two matching conditions, \( t(5) = 12.88, p < 0.001 \). Therefore, white disks decreased in perceived numerosity when intermixed with gray disks whereas the perceived numerosity of gray disks was unaffected by intermixing with white disks; an apparent loss of white disks explains the entire illusion in an intermixed display.

This pattern of results is not consistent with either the misclassification or the occlusion hypothesis. Although misclassification does predict a loss in white disks, this loss should have led to a gain in gray disks. For different reasons, occlusion also predicts an increase in the numerosity of gray disks. Surprisingly, our data showed no change for gray disks. Indeed, the fact that white disks are missing to a large extent is counterintuitive in the sense that white disks are readily visible in both patches so that a veridical match should be expected. Moreover, the white disks are more salient than the gray ones, so if anything, the gray disks might have been the ones to be suppressed and fall “unseen.” Indeed, previous studies suggest that more salient items tend to appear more numerous than less salient ones, perhaps due to the higher attentional priority assigned to salient items (Tokita & Ishiguchi, 2010). Our finding is at odds with this effect of salience.

Most previous studies, while employing a numerosity discrimination task, have generally used two patches of elements either temporally or spatially separated, thus providing no opportunities to probe the type of interaction between two intermixed sets such as those in our stimulus configuration. One study, however, employing a magnitude estimation task, has investigated how spatially intermixed sets are enumerated, using a stimulus configuration similar to ours (Cordes, Goldstein, & Heller, 2014). It was found that, when two sets of differently colored dots were intermixed, the estimated numerosity of one set of the dots was decreased in the presence of the other set compared to when it was presented alone. The authors described this as a “subset” effect in that the numerosity of one subset of dots in an array was depressed by the other subset. A presumed symmetrical effect of this sort could not explain the asymmetry in our results—that only the intermixed white disks, not the gray ones, decreased in numerosity, suggesting different interactions between the intermixed sets of elements in their study and in ours.

Experiment 4B: Superset matching

Experiment 4A revealed an asymmetrical interaction between the gray and the white disk sets in an intermixed patch with a pattern that is consistent with the illusion we introduced in Experiment 1A but is not consistent with the predictions of the misclassification or occlusion hypotheses or the subset effect reported by Cordes et al. (2014). However, there is indeed one possibility that all these effects work together to account for the results. Misclassification or occlusion may still be the cause of a change in relative numerosities of the two disk sets, leading to the illusion of more gray disks than white ones in an intermixture. A subset effect could then operate to reduce the perceived numerosities of both sets, pushing them down to the values we measured in Experiment 4A. We set out to test this possibility in the current experiment.

Instead of comparing each subset from the intermixed patch with an unmixed set, subjects were asked to regard all the elements in the intermixture as a unitary superset and compare its numerosity to the numerosity of a homogenous set in a second patch. Whereas classifying the disks into different subsets is not explicitly necessary for this task, to the extent that misclassification still occurs in an automatic manner, it would only have an impact on the relative numerosity of the gray versus white disks but not the total numerosity of the superset as all the disks are well above the visibility threshold and only subject to within-set conversion. On the other hand, if occlusion is still assumed, the total numerosity of the superset should increase because some hypothetically occluded gray disks are counted in. Therefore, a superset matching task shows the potential to further tell apart the misclassification and the occlusion hypotheses. Moreover, the subset effect described by Cordes et al. (2014) is no longer relevant for a superset task, thus allowing a more clear-cut test of the two hypotheses.

Methods

Eight subjects participated in this experiment. The stimuli and procedure were similar to that in Experiment 4A. The intermixed patch consisted of 50 white disks and 50 gray disks throughout. The other patch consisted of disks of a homogenous contrast and varied in numerosity in the range of 80–120 by steps of five. The luminance of disks in this homogenous patch was 45.8 cd/m², approximately halfway between the luminances of the gray and white disks in the intermixed patch. Subjects were asked to select the patch with more disks regardless of color (gray or white) by
pressing the corresponding left or right arrow key. The experiment lasted about 20 min.

Results and discussion

The PSE in this experiment indicates how many disks in the homogenous patch would match 100 disks in the intermixed patch in perceived numerosity. As seen in Figure 8, all subjects had PSEs below 100, indicating that the numerosity of the intermixed patch was underestimated relative to the homogenous patch. The average PSE across subjects was 91, which was significantly different from 100, $t(7) = 3.56$, $p = 0.009$. This result was clearly at odds with the prediction of either the misclassification or the occlusion hypothesis: misclassification predicts no change whereas occlusion predicts an increase in the total numerosity of the intermixed patch. Apparently, some disks were missing from the intermixed patch, and this is consistent with the results of Experiment 4A, in which a few white disks were missing, but the gray disks were intact.

It should be noted that, in that same study, Cordes et al. (2014) also reported a superset effect, showing that subjects overestimated the total numerosity of an intermixed patch containing two subsets of colored disks relative to that of a homogenous patch. Clearly, such an effect was absent in our study. It is even more unlikely that this superset effect combined with either misclassification or occlusion could account for our results as the joint effect would only result in an overestimation rather than an underestimation of the superset. Taking all these results together, both the misclassification and the occlusion hypotheses should be rejected as the explanations of our illusion.

Although various factors have been implicated in previous studies to impact numerosity judgment (Aida et al., 2015; Dakin et al., 2011; Gebuis & Reynvoet, 2012; Ginsburg, 1991; Hurewitz et al., 2006; Schütz, 2012; Sophian & Chu, 2008), to the best of our knowledge, this is the first report of a robust effect of suprathreshold contrast. We showed that when two sets of stimuli with different contrasts were intermixed, the one with low contrast (gray) appears to be more numerous than the one with high contrast (white or black)—hence, a phenomenon of “the weak conquer the strong.” The illusion seems due entirely to the missing of high-contrast (e.g., white) disks as it was found that the perceived numerosity of white disks in an intermixed patch decreased while that of gray ones remained unaffected when comparing each set to an isolated set.\footnote{This claim was further corroborated by the finding that the perceived total numerosity of the intermixed patch, when considered as a superset, was reduced compared to a homogenous set, suggesting that indeed some disks are missing from the intermixture. These results point to an asymmetrical interaction between the two sets of elements within an intermixed patch. However, the mere presence of a contrast difference is not sufficient to induce the illusion. We showed that the illusion is limited to two sets with the same contrast polarity and does not occur for two sets with opposite polarities. Moreover, the illusion depends critically on intermixing; once the two sets of elements were spatially segregated, the illusion disappeared. Last but not least, the magnitude of the illusion increases as the contrast difference between the two sets decreases.}

As explained before, the illusion we reported here is clearly an effect of contrast and is different from the luminance effect reported by Ross and Burr (2010) in important ways although, in both cases, it is the weaker stimuli that appear more numerous. Because our illusion is characterized by an asymmetrical interaction between two sets of intermixed elements, it cannot be explained by the subset effect reported by Cordes et al. (2014) as no such asymmetry is implicated by their findings. Two plausible explanations of the illusion, misclassification and occlusion, can also be rejected as explained above. It should be noted that several studies have explicitly manipulated the depth structure of a dot array to evaluate its effect on numerosity perception, all showing that the numerosity of elements lying on a back surface within a 3-D configuration was overestimated relative to one display with no depth layering and, thus, pointing to a potential role of occlusion (Aida et al., 2015; Schütz, 2012) although occlusion does not explain the additional observation that there was an underestimation of elements on a front surface (Schütz, 2012). If depth layering exists between the
white and gray disks in our stimuli and plays a major role in our illusion, we should expect an overestimation of the numerosity of gray disks as was found for back-surface elements in previous studies. Our study nevertheless showed that the perception of numerosity remained largely unaffected for the gray disks in the intermixture, therefore rejecting any explanations that resort solely to contrast-induced depth separation and, in particular, to occlusion.

As contrast is a fundamental property of almost any visual input, the discovery of a contrast effect on perceived numerosity calls for a revisit of the mechanisms of numerosity perception. Controversy exists in literature as to whether numerosity can be directly perceived (Burr et al., 2018; Burr & Ross, 2008; Cicchini et al., 2016; Dehaene, 2011) or can only be inferred from surrogate measures, such as density (Dakin et al., 2011; Durgin, 1995; Gebuis & Reynvoet, 2012). On the other hand, there have also been discussions regarding whether the numerosity mechanism operates on segmented objects (Franconeri, Bemis, & Alvarez, 2009; He, Zhang, Zhou, & Chen, 2009; Kirjakovski & Matsumoto, 2016) or on the raw visual image with its spatial frequency contents as the basis for numerosity computation (Dakin et al., 2011; Morgan et al., 2014). For example, it has been demonstrated that, by connecting pairs of items in a set by either physical (Franconeri et al., 2009; He et al., 2009) or illusory connectors (Kirjakovski & Matsumoto, 2016) so that they were not segmented, the perceived numerosity of that set was reduced relative to a set in which all items were segmented from one another. This finding suggests that the segmentation of individual items in a set is a necessary early stage of numerosity processing, consistent with models with a dedicated mechanism for numerosity (e.g., Dehaene & Changeux, 1993). On the other hand, some studies disputed the idea of a dedicated number sense and showed that a mechanism exploiting the output of general-purpose neural filters is capable of implementing numerosity judgments, producing biases and performance characteristics very similar to that of human observers (Dakin et al., 2011; Morgan et al., 2014). Specifically, they advocate that the spatial frequency structure of a raw image provides the basis for numerosity computation, and the segmentation of individual objects is not explicitly necessary; to them, numerosity processing is essentially contrast energy analysis. Whereas the present study was not intended to settle all these disputes, our findings contribute important insights into the issue. Our stimuli consist of disks with contrast well above threshold, and they are clearly segmented from one another. In an attempt to explain our illusion, two hypotheses—misclassification and occlusion—were proposed, both of which implicitly assume that segmented objects underlie numerosity computation; that is, misclassification or occlusion occurs between individual objects. However, both hypotheses were falsified by our data. We are not suggesting that any theory of numerosity perception based on segmented objects is wrong but that there is no obvious way this class of theories can explain our findings.

We instead turn to consider the plausibility of mechanisms that take a raw image as input and perform contrast energy analysis as a means to gauge numerosity. It is worth noting that most existing models on numerosity are not designed to handle the kind of stimulus and task such as ours in which human observers have to discriminate numerosities of two sets of elements embedded within one another. Whereas lab studies on numerosity judgment have generally employed isolated sets of elements as the stimuli, in real-world situations this is rarely the case; it is more the rule than the exception that different sets for enumeration are intermixed with one another. Imagine that one wants to estimate the number of bright ripe plums on a plum tree; however, these do not naturally cluster together but instead are interspersed among duller ones. There is indeed one model that has proposed a mechanism to implement numerosity discrimination between two sets of elements within an intermixed patch (Morgan et al., 2014). According to the model, for two sets of elements differing in features such as orientation and size, outputs from different channels can be utilized to calculate contrast energy as the basis for numerosity discrimination. For two sets of elements (Gabor patches) differing only in contrast, however, the differently thresholded outputs from a single channel are utilized. The model, without explicitly specifying a mechanism for contrast normalization, makes no direct prediction regarding any potential effect of contrast on numerosity judgment; therefore, in its current form, it fails to explain our illusion.

Here we provide a tentative explanation for the illusion. Like others (Dakin et al., 2011; Morgan et al., 2014), we assume that an input image is sampled by a mosaic of localized Gaussian-like filters, and numerosity is signaled by the pooled response of filters tuned to relatively high spatial frequency, ideally matching the size of individual items. With our stimulus configuration, however, the numerosity system has to somehow segregate the two sets of disks and perform selective pooling of filter responses in order to discriminate their numerosities. This can be achieved by a thresholding mechanism, similar in spirit but different in implementation to that proposed by Morgan et al. (2014). On the other hand, to gauge numerosity irrespective of nonnumerical properties, contrast normalization is needed to account for the overall larger responses induced by high-contrast elements than by low-contrast ones, just as normali-
numerosity or some surrogate measure for each set, gray disks (Figure 9). The goal is to derive the perceived set at the level of but simply pools the filter responses based on their

\[ N \]

Cl

observed in this study. Contrast-based thresholding and normalization is applied to other domains of visual processing (Carandini & Heeger, 2012). We suggest that it is this combination of contrast-based thresholding and normalization that leads to the numerosity illusion we observed in this study.

Suppose the case of \( N \) white disks intermixed with \( N \) gray disks (Figure 9). The goal is to derive the perceived numerosity or some surrogate measure for each set, denoted as \( N_h' \) and \( N_l' \), respectively, for white and gray disks. For simplicity, we treat the contrast of gray disks and the corresponding filter response as the same thing because the two should scale. We denote it as \( C_h \) and similarly for white disks, this is denoted as \( C_l \) with \( C_h > C_l \). The system does not select the disks one by one, but simply pools the filter responses based on their amplitudes. To pool for the gray disks, a threshold was set at the level of \( C_h \), and any filters with response greater than the threshold are suppressed, thereby excluding those filters responding to white disks from being pooled. The pooled response for the gray disks, denoted as \( E_h \), is, therefore, proportional to the product of \( N \) and \( C_h \), i.e., \( E_h \propto N \times C_h \). To pool for the white disks, however, the visual system adopts a subtraction operation, such that \( C_l \) is subtracted from the responses of all filters, thereby effectively silencing the filters responding to gray disks while reducing the filter response to white disks to \( (C_h - C_l) \). Essentially, the gray disks are treated as background noise, the elimination of which by subtraction reduces the effective contrast of the white disks. The pooled response for the white disks, denoted as \( E_l \), is, therefore, proportional to the product of \( N \) and the adjusted contrast, i.e., \( E_l \propto N \times (C_h - C_l) \). We note that, because the illusion applies to both positive and negative contrast displays, as shown in Experiments 1A and 1B, the terms \( E_h \) and \( E_l \) should be taken as depending on the square root of contrast energy rather than on signed contrast. Following the selective pooling of filter responses, contrast normalization is performed to derive numerosity or its surrogate measure. We suggest that the normalization factor is proportional to \( C_h \) for the former and to \( C_l \) for the latter. Therefore, contrast normalization is done almost correctly for the gray disks, such that the perceived numerosity of this set is an increasing function of physical numerosity alone, i.e., \( N_l' = f(N) \). On the other hand, the pooled response for the white disk set is thresholded, thus reduced, but is normalized by the unthresholded local contrast so that \( N_h' = f(N \times (C_h - C_l)/C_h) \). As a result, the perceived numerosity of white disks is reduced relative to that of gray disks—hence, the illusion we reported in this study. We acknowledge that our proposal is ad hoc, inevitably including some hypothesized “black box” operation, and is also simplistic in nature with many components yet to be developed more quantitatively (Reeves & Lei, 2016). Nevertheless, such a mechanism can explain the observation that only the high-contrast elements but not the low-contrast ones were missing from an intermixed patch and, more importantly, the fact that the illusion increases in magnitude as the contrast of the low-contrast elements increases toward that of the high-contrast ones.

Put in a broader context, many visual processes, including orientation and spatial frequency selectivity, and, at the higher level, face and object recognition, depend on contrast near threshold but are mostly independent of contrast above threshold (e.g., Avidan et al., 2002; Skottun, Bradley, Sclar, Ohzawa, & Freeman, 1987). This independence is typically explained by contrast normalization, which, when misapplied, accounts for various unexpected contrast effects, such as found in motion perception (Blakemore & Snowden, 1999; Stone & Thompson, 1992; Stone et al., 1990). It is not too surprising that numerosity perception, if involving some sort of contrast energy analysis, could also suffer from a misapplication of normalization, such as demonstrated by the numerosity illusion we discovered. The illusion is present when two sets of elements with clearly above-threshold contrasts are intermixed but is absent when they are spatially segregated, which implies a role for segregation in numerosity perception. Our observers in this study could generally segregate the disks on the basis of contrast even though such segregation can be incomplete (Beck, Graham, & Sutter, 1991). In our model, segregation by contrast-based thresholding, coupled with the misapplication of contrast normalization, explains the numerosity illusion. If segregation is achieved by utilizing cues other than contrast and, hence, does not rely on thresholding, energy pooling for...
each set can then proceed independently of one another, therefore precluding erroneous normalization and eliminating the illusion. Spatial segregation is just one example; the effect of another cue—that of contrast polarity—was also reported in Experiment 2A of the current study. Indeed, the illusion can be abolished by many other types of cues to segregation, such as motion and stereoscopic depth (Lei, 2015). As stimulus contrast is a fundamental factor in visual processing, the contrast-dependent numerosity illusion reported in this study provides an important test case for models of numerosity perception and introduces new constraints on numerosity theories.

Keywords: numerosity perception, contrast, segregation, illusion

Acknowledgments

Commercial relationships: none.
Corresponding author: Quan Lei (雷泉).
Email: qlei@umn.edu; LeiQ2009@gmail.com.
Address: Department of Psychology, University of Minnesota, Minneapolis, MN, USA.

Footnote

1 Although the data was not reported here, the same pattern was also found with negative contrast stimuli, such that in an intermixed patch of black disks and dark gray disks on a light gray field, the perceived numerosity of black disks decreased while that of the gray disks remained unchanged as compared to an isolated counterpart.

References


Franconeri, S. L., Bemis, D. K., & Alvarez, G. A.


