

Journal of Experimental Psychology: General

Visual Illusions Help Reveal the Primitives of Number Perception

Edwina Picon, Denitza Dramkin, and Darko Odic

Online First Publication, February 7, 2019. <http://dx.doi.org/10.1037/xge0000553>

CITATION

Picon, E., Dramkin, D., & Odic, D. (2019, February 7). Visual Illusions Help Reveal the Primitives of Number Perception. *Journal of Experimental Psychology: General*. Advance online publication. <http://dx.doi.org/10.1037/xge0000553>

Visual Illusions Help Reveal the Primitives of Number Perception

Edwina Picon, Denitza Dramkin, and Darko Odic
University of British Columbia

The human perceptual system is responsive to numerical information within visual and auditory scenes. For example, when shown 2 displays of dots, observers can instantly, albeit approximately, identify the set that is more numerous. Theories in perceptual and cognitive psychology have focused on 2 mechanisms for how vision accomplishes such a feat: Under the domain-specific encoding theory, number is represented as a primary visual feature of perception, much like motion or color, while under the domain-general theory, the visual system represents number indirectly, through a complex combination of features such as the size of the dots, their total cluster, and so forth. Evidence for the latter theory often comes from “congruency effects:” the finding that participants frequently select the side where the dots on the screen are denser, larger, or brighter, rather than the side that is actually more numerous. However, such effects could also stem from response conflicts between otherwise independent dimensions. Here, we test these 2 competing accounts by embedding numerical displays within visual illusions that create large conflicts between number and other non-numeric dimensions—including contour length, convex hull, and density—and contrast participants’ performance on a number discrimination task (i.e., “Which side has more dots?”) against a number estimation task (i.e., “How many dots are there?”), which should eliminate response conflicts. Across 3 experiments, we find that while contour length illusions only affect number perception in discrimination tasks, the influences of convex hull and density on number perception persist in both discrimination *and* estimation tasks, supporting a more domain-general account of number encoding.

Keywords: approximate number system, number perception, number sense, visual illusions

Supplemental materials: <http://dx.doi.org/10.1037/xge0000553.supp>

Perception gets a lot from very little. Even just a cursory glance at Figure 1 yields an instant and automatic sense of number: Without counting you can easily decide whether there are more blue dots than yellow dots. Decades of work have shown that the visual and auditory systems of human newborns and nonhuman animals are sensitive to changes in number (Izard, Sann, Spelke, & Streri, 2009; for review of the nonhuman animal literature, see Vallortigara, 2017), and that this “number sense” contributes to an assortment of other cognitive abilities, including our understanding of currency (Marques & Dehaene, 2004) and foraging behavior in nonhuman primates (Piantadosi & Cantlon, 2017). As a result, the perceptual sense of number has been of great interest to cognitive, computational, developmental, and comparative psychologists.

Our sensitivity to visual number information simultaneously showcases the efficiency and the mystery of perception: Because

number cannot be extracted from any single objective feature, such as wavelengths of light (for color), salt concentration (for taste), pressure on skin (for touch), and so forth, how do our perceptual systems encode and represent number? To date, two types of theories have been put forward to answer this question. Under the first, the *domain-specific encoding theory*, number is represented relatively early in sensory processing by dedicated and specialized neurons, thus constituting a primitive of perception, much like color, orientation, and motion. For example, in the model of Dehaene and Changeux (1993), low-level neurons instantiate a two-dimensional object map whose total activity corresponds to the number of objects in the scene: the more objects, the more populated the map, and the higher the representation of number (see also Stoianov & Zorzi, 2012). Consistent with this domain-specific encoding account, Burr and Ross (2008; Ross & Burr, 2010, 2012) have repeatedly demonstrated that we can perceptually adapt to number. In the same way that staring at a green square subsequently produces an illusory percept of a red square, staring at a display of 100 dots subsequently produces an illusory diminished sense of number compared with staring at a display of 50 dots (see Burr & Ross, 2008 for demonstration). Because adaptation effects are most often caused by the fatiguing of low-level neurons due to repeated exposure, Burr and Ross (2008; Ross & Burr, 2010, 2012) have concluded that number must therefore be a low-level feature akin to motion, color, orientation, and so forth (for other arguments in favor of number as a primary visual feature, see Anobile, Cicchini, & Burr, 2016).

Edwina Picon, Denitza Dramkin, and Darko Odic, Department of Psychology, University of British Columbia.

This research was supported by an NSERC DG provided to Darko Odic. We thank Catalina Canon for help in data collection and Justin Halberda for various conversations about the data and the article.

Correspondence concerning this article should be addressed to Darko Odic, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, British Columbia V6T 1Z4. E-mail: darko.odic@psych.ubc.ca

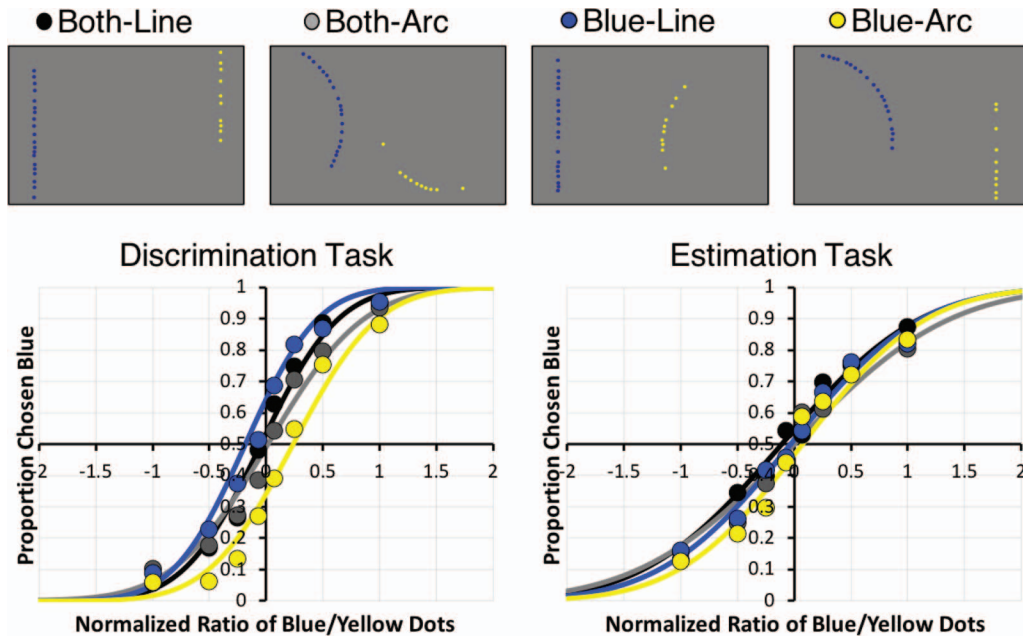


Figure 1. Experiment 1 stimuli and results. The top panel illustrates four example trials, one from each of the four conditions. The bottom panel shows the data from the Experiment 1 discrimination and estimation conditions; the lines are the best-fit psychophysical cumulative normal model. See the online article for the color version of this figure.

The alternative account—the *domain-general encoding theory*—claims that number is not a primitive feature of perception, but is instead constructed or inferred through some combination of non-numeric features, such as—in the case of vision—density, area, brightness, and so forth (e.g., Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011; Defever, Reynvoet, & Gebuis, 2013; Gebuis & Van Der Smagt, 2011; Leibovich, Katzin, Harel, & Henik, 2017; Szucs, Nobes, Devine, Gabriel, & Gebuis, 2013). Under some versions of this account, number is still represented by the visual system, albeit indirectly through a combination of several non-numeric visual features (e.g., Leibovich et al., 2017), while under other versions, number is actually not represented at all and participants' responses on number tasks are thought to be entirely based in non-numeric features (e.g., Gebuis & Van Der Smagt, 2011). These accounts rely on the findings that number is strongly correlated with many other perceptual properties: The set with more dots is also bigger, brighter, denser, and so forth, than the set with fewer dots. Hence, rather than relying on dedicated neurons, our visual sense of number may instead emerge from a learned or an innate combination of several non-numeric feature detectors (Gebuis & Reynvoet, 2012b; Leibovich et al., 2017; Szucs et al., 2013).

A central phenomenon supporting the domain-general encoding account is the wealth of “congruency effects”—instances where participants appear biased toward non-numeric features (e.g., density, area, brightness, etc.), rather than toward number itself. For example, Gebuis and colleagues (Gebuis & Reynvoet, 2012a, 2012b; Gebuis & Van Der Smagt, 2011; see also Hurewitz, Gelman, & Schnitzer, 2006) show that participants are more accurate on “congruent” trials, where the more numerous side of the dot array is also the one that visually appears “bigger” or more

spread out, compared with “incongruent” trials, where the more numerous side is the one that appears “smaller.” Dakin, Tibber, Greenwood, Kingdom, and Morgan (2011; see also Durgin, 1995) demonstrate similar biases toward non-numeric cues, finding that participants often perceive denser sets of dots as more numerous compared to displays that are sparser (for evidence supporting the domain-general account that go beyond congruency effects, see Leibovich, Al-Rubaiey Kadhim, & Ansari, in press; Leibovich-Raveh, Stein, Henik, & Salti, 2018; Salti, Katzin, Katzin, Leibovich, & Henik, 2017).

While congruency effects have been one of the main sources of evidence for the domain-general encoding account, these effects are unfortunately insufficient for settling the number encoding debate, as congruency effects may either be evidence of shared encoding *or* the byproduct of competition for the same behavioral response (Hurewitz et al., 2006; Odic & Starr, 2018; Van Opstal & Verguts, 2013). In other words, participants may do worse on incongruent trials because—in a typical speeded number discrimination task—*independent* non-numeric features, such as density or area, are actively competing for the opposite response from the response for number. For example, if there are more dots on the left, but a salient density signal is detected on the right, the two responses will conflict and need to be resolved (much like in the classic Stroop task, where two independent dimensions—reading and color perception—compete for the same response), leading to a reduction in accuracy and an increase in response time. As a result, congruency effects may not actually demonstrate a reliance on non-numeric cues during number perception, as has been previously argued, but may instead point to active competition between otherwise independent dimensions that are difficult to inhibit (Hurewitz et al., 2006). In other words, congruency effects

are consistent with and predicted by *both* the domain-general (i.e., the encoding of number through non-numeric features) and the domain-specific (i.e., multiple dimensions compete for the same response) accounts of number encoding.

In order to understand the contributions of non-numeric features to number perception, we have to disentangle the contributions these features make to encoding as compared to response competition. Some recent work has attempted to do this through novel tasks or mathematical models, but the findings remain inconclusive. For example, Cantlon, Safford, and Brannon (2010) and Ferrigno, Jara-Ettinger, Piantadosi, and Cantlon (2017) trained children and monkeys to categorize stimuli that vary in number and size and found that both children and monkeys spontaneously chose to focus on and categorize by number over other dimensions. At the same time, however, EEG signatures during passive viewing of dot stimuli, in which number is confounded with other variables, show that attention is clearly drawn toward non-numeric dimensions (Gebuis & Reynvoet, 2012a), opening the possibility that spontaneous categorization begins through non-numeric dimensions that are then combined into a numerical representation.

Another approach to controlling response conflicts is to try to mathematically model them. DeWind, Adams, Platt, and Brannon (2015), for example, developed a linear regression model where a participant's number discrimination response—a decision of which side of the screen has more dots—is a combination of their use of numeric, size, and spacing information, with each dimension having its own independent weighted value. The authors found that adult participants mostly responded on the basis of number and not size or spread (see also Starr, DeWind, & Brannon, 2017). However, models such as these inherently carry many assumptions about how participants are completing the task (e.g., that the contributions of each non-numeric dimension are independent of the rest), many of which have not been empirically validated.

To alleviate some of these challenges, we take a complimentary, but simpler and model-free, approach to measuring the contributions of non-numeric features to number encoding versus response conflicts. Taking inspiration from research showing that the Stroop effect is entirely eliminated when reading no longer competes for the same response as color identification (e.g., when participants click on a color patch to select the font color, rather than verbally report it; Durgin, 2000), we aimed to examine whether non-numeric features continue to affect number perception when these dimensions no longer compete for the same response as number itself. To eliminate the possibility of response conflicts, we relied on a *number estimation* task, in which participants are asked to estimate the total number of dots (e.g., “15” blue dots), rather than choose the side that is more numerous, as with standard *number discrimination* tasks. As scaled numerical responses are impossible responses for non-numeric dimensions (i.e., “15” density or “eight” area is not a meaningful response), estimation tasks should not allow for response competition between number and various other non-numeric dimensions. On the one hand, if congruency effects are eliminated in number estimation tasks, we would have good evidence that the congruency effects observed in standard number discrimination tasks are actually driven by a Stroop-like response conflict. If, on the other hand, congruency effects persist despite the estimation prompt, we would have evidence that non-numeric dimensions are actively recruited during number perception itself, independent of any response conflicts.

A second challenge in the literature on the role of non-numeric features in number perception is understanding precisely *which* non-numeric features may contribute to number encoding. In the majority of existing studies, researchers often try to manipulate one dimension (e.g., density) without affecting the others (e.g., the size of the dots). However, this is often impossible, as manipulations of one dimension inevitably lead to changes in another (see DeWind et al., 2015). To more precisely target particular non-numeric features, we embedded our stimuli into well-known strong visual-illusions: the plug-hat illusion, which selectively targets the perception of contour length (Simanek, 1996), and the Ebbinghaus illusion, which selectively affects the perceived convex hull of objects. In the plug-hat illusion, participants judge a circular contour/arc to be significantly shorter in length than a straight line that is identically long; in the Ebbinghaus illusion, participants judge an object surrounded by smaller circles to be significantly larger than the identical object surrounded by larger circles. By relying on these visual illusions, we not only magnify potential congruency effects (which tend to be small, thus minimizing the chance of false negatives), but also maintain the objective differences in our stimuli, while allowing the participant's subjective perception of non-numeric dimensions to be affected (Im & Chong, 2009). Therefore, we can more precisely target particular non-numeric features—that is, those affected by the visual illusions—without compromising the remaining objective non-numeric features held within the display.

In three experiments, we compare performance on number discrimination versus number estimation tasks by embedding number displays within visual illusions while manipulating two specific non-numeric features: contour length and convex hull/density. We selected these two dimensions both because they have previously been shown to produce congruency effects in number discrimination tasks (Clearfield & Mix, 1999, 2001; Dakin et al., 2011; Durgin, 1995), and because they are thought to tap into distinct, nonoverlapping low-level neural mechanisms, with contour length relating to neurons encoding orientation (Li, 1998) and convex hull/density to neurons encoding low-spatial frequency (Dakin et al., 2011). In our experiments, participants saw dot displays where the side with *more* dots *also* had an (illusory) longer contour length or an (illusory) larger convex hull (i.e., the congruent trials), and dot displays where the side with *fewer* dots had the longer (illusory) contour length or larger (illusory) convex hull (i.e., the incongruent trials). Consistent with previous work, we expected participants to do better on congruent trials in the discrimination tasks, because binary left/right responses create Stroop-like response conflicts between number, contour length, and convex hull/density. The key test, therefore, was whether these same biases for the congruent trials would persist in the estimation task when binary responses are eliminated.

Experiment 1: Contour Length and the Plug-Hat Illusion

Method

Participants. Based on pilot testing, we determined that the approximate sample size we would need to achieve 80% power for the key interaction between the two incongruent conditions (blue line vs. blue arc) was 22 participants per condition. Once we met

that goal, we allowed any remaining participants who had already signed up to participate in the study to complete it for university course credit, giving us a sample of 90 participants. As described in detail below, a third of the participants completed the discrimination task, a third completed the precue estimation task, and the final third completed the postcue estimation. Two subjects in the precue and two subjects in the postcue estimation task were excluded for having made random guesses on every trial. This left us with a final sample of 86 subjects ($n = 28$ in the discrimination, $n = 30$ in the precue estimation, and $n = 28$ in the postcue estimation condition). All experimental methods and procedures for all reported experiments were reviewed and approved by the University of British Columbia Office of Research Ethics (#H14-01968).

Procedure. Participants were individually tested in a quiet room on a 22.5" iMac running custom-made Psychtoolbox-3 scripts (Brainard, 1997). These scripts are freely available online at <https://osf.io/npj6g/>. Participants completed one of two tasks (i.e., the discrimination task or the estimation task), over identical sets of stimuli.

In the discrimination task, participants were shown displays of blue and yellow dots for 500 ms (too quick to allow for counting, Cordes, Gallistel, Gelman, & Latham, 2007; Figure 1), with blue dots appearing on the left side of the screen and yellow dots on the right. They were instructed to focus only on number and were not informed about the plug-hat illusion prior to testing. Participants completed a total of 512 trials, taking around 10–12 min to complete the task. Across each trial, they had to determine whether “there are more blue or yellow dots,” indicated through pressing the F key on the keyboard if they believed there to be “more blue” dots or the J key for “more yellow” dots. To vary the difficulty of the task, we altered the number and ratio of the blue to yellow dots, with the total number of dots varying near-uniformly from 10–20 and the ratio of blue to yellow dots being either 2.0 (i.e., 20 blue vs. 10 yellow dots), 1.50, 1.25, 1.07, 0.93, 0.80, 0.66, or 0.50 (i.e., 10 blue vs. 20 yellow dots). Based on this design, as the ratio increases, participants should be more and more likely to select the blue set as the more numerous one (see also Feigenson, Dehaene, & Spelke, 2004; Odic, 2017; Odic & Starr, 2018).

Critically, to vary the influence of the contour length on number perception, the blue and yellow dots were presented along a line that was either straight or curved/arc, consistent with the plug-hat illusion (see Figure 1). This led to four conditions: both-line (both the blue and yellow dots were arranged along a line), both-arc (both the blue and yellow dots were arranged along an arc), blue-line (the blue dots were arranged on a line, and yellow dots along an arc), and blue-arc (the blue dots were arranged along an arc, and the yellow dots along a line). The blue-line and blue-arc conditions effectively act as our Incongruent conditions, predicting opposite responses: If participants attend to the contour length they should be biased to respond “blue” in the blue-line condition and “yellow” in the blue-arc condition, independent of ratio. Pilot testing ($N = 12$) in which we continuously varied the length of the straight line versus the arc and asked participants to select the longer segment revealed that observers on average perceived the arc as shorter than the straight line by 15.1% ($SE = 0.03$; range 2.9%–37%), consistent with the plug-hat illusion. In order to prevent any possible perception of the dots as actually creating a complete curve, the spacing between the dots was also

randomized and adjusted (wiggled). Because the dots were randomly spaced on the line versus arc, some of the trials would appear to have one of the sets physically longer than the other; however, as this was equally likely to occur for the line versus arc stimuli, we did not analyze these trials separately. Consistent with prior work (e.g., Dakin et al., 2011), the primary dependent variable in this task was the proportion of trials in which participants selected blue as the more numerous set. Though we report average accuracy, we do not use it as the primary dependent measure of interest, as accuracy cancels out any biases toward the blue or yellow sets in the two critical conditions (e.g., if a participant in the blue-line condition chooses “blue” 75% of the time and chooses “yellow” in the blue-arc condition 75% of the time they would show an equivalent accuracy between the two conditions of around 62.5%, even though their bias is in opposite directions).

In the two estimation tasks, participants saw identical displays as in the discrimination task, but were asked to estimate the number of dots: either “how many blue dots” or “how many yellow dots.” The stimuli once again appeared for 500 ms (too quick to allow for counting), then disappeared, allowing participants to respond by typing in any number using the top row of keys on the keyboard (participants could erase their response and retype it if they made a mistake before moving on to the next trial). In the precue version, participants were told which color to estimate before seeing the stimuli, as is typical in number estimation tasks (e.g., Cordes, Gelman, Gallistel, & Whalen, 2001; Odic, Im, Eisinger, Ly, & Halberda, in press). However, to make sure that participants attended to both sides of the screen rather than just focusing only on the target set, we also ran a postcue version of the task, in which participants were cued to which color they needed to estimate *after* the stimuli had already disappeared. Each trial was repeated twice in a randomized order—once for estimating yellow dots and once for estimating blue dots—so that participants would inevitably estimate both sides of the array. Participants completed 512 trials, which included 256 unique trials repeated twice, and took 20–25 min to complete this condition. As in the discrimination task, participants were not informed about the plug-hat illusion and were instructed to only focus on number. Additionally, they were not informed that any trials would repeat or what the total range of the number of dots would be. As described in detail below, the dependent variables in these two estimation tasks were the estimation error rate (i.e., the difference between the true number and the participant’s response, divided by the true number), and the proportion of trials on which participants estimated the blue set to be more numerous, allowing us to make direct comparisons in across tasks.

Results

Number discrimination results. Participants showed excellent accuracy in this task, averaging 78.3% ($SE = 0.97$). A 4 (Condition: Both-Line, Both-Arc, Blue-Line, Blue-Arc) $\times$ 8 (Ratio: 2.0, 1.50, 1.25, 1.07, 0.93, 0.80, 0.66, 0.50) Gaisser-Greenhouse corrected repeated-measures ANOVA over the proportion of trials on which participants selected the blue side as more numerous revealed significant main effects of condition, $F(3, 81) = 59.18$; $p < .001$; $\eta_p^2 = .69$; ratio, $F(7, 189) = 448.27$; $p < .001$; $\eta_p^2 = .94$; and a Condition $\times$ Ratio interaction, $F(21, 567) = 6.39$; $p < .001$; $\eta_p^2 = .19$. As shown in Figure 1 and Table 1, these

effects were carried by a clear bias that participants had for indicating that the blue side was more numerous in the blue-line condition ($M = 56.7\%$; $SE = 1.64\%$) compared with the blue-arc condition ($M = 38.9\%$; $SE = 1.16\%$). In other words, we observed a clear congruency effect of contour length on number discrimination.

Number estimation results. To test whether the discrimination congruency effect stemmed from response conflicts or from shared encoding procedures, we next turn to analyzing the estimation task data. If participants also show a bias toward estimating the dots along the line as more numerous, we should find that participants *overestimate* the number of blue dots (relative to yellow) in the Blue-Line condition, and *underestimate* the number of blue dots (relative to yellow) in the Blue-Arc condition. To determine the degree of over/underestimation, we calculated—for each condition and each set of dots—the average error rate (i.e., the difference between the participant's estimate and the true number of dots, divided by the true number of dots). Error rates are the most common measure of over/underestimation in the literature (e.g., Crollen, Castronovo, & Seron, 2011; Libertus, Odic, Feigenson, & Halberda, in press), with numbers greater than 0 indicating overestimation, numbers below 0 indicating underestimation, and numbers around 0 indicating excellent performance on matching the estimate to the true number of dots (i.e., no over/under estimation).

As shown in Figure 2, we found that participants' error rates for blue versus yellow dots were equivalent across the four conditions in both versions of the estimation task: A 2 (Version: Precue, Postcue) $\times$ 4 (Condition: Both-Line, Both-Arc, Blue-Line, Blue-Arc) $\times$ 2 (Dot Set: Blue, Yellow) Gaisser-Greenhouse corrected mixed-measures ANOVA over error rates showed no significant main effects of condition, $F(3, 168) = 1.89$; $p = .13$; $\eta_p^2 = .03$; dot set, $F(1, 56) < 1$; $p = .34$; $\eta_p^2 = .02$; and most importantly, no significant Condition $\times$ Dot Set interaction, $F(3, 168) < 1$; $p = .56$; $\eta_p^2 = .01$. To further quantify the magnitude of this null effect, we also computed the Bayes Factor (BF) for the critical Condition $\times$ Dot Set interaction, finding strong evidence in favor of the null model (BF = 68.51). In addition, we found no main effect of version, $F(1, 56) < 1$; $p = .41$; $\eta_p^2 = .01$, nor a Version $\times$ Condition interaction, $F(1, 56) < 1$; $p = .41$; $\eta_p^2 = .01$, nor a three-way Version $\times$ Condition $\times$ Dot Set interaction, $F(3, 168) = 1.44$; $p = .24$; $\eta_p^2 = .02$. Thus, participants appeared to successfully ignore the contour length of the dots in the two versions of the estimation tasks, producing no congruency effects.

This null result cannot be explained by participants randomly guessing on the number of blue versus yellow dots: We find that the average slope of the estimates plotted against the objective number of dots is significantly higher than zero in both the precue ($M = .66$, $SE = .09$; $t(29) = 7.25$; $p < .001$) and the postcue versions ($M = .73$, $SE = .16$; $t(27) = 4.78$; $p < .001$), suggesting typical overall estimation performance (Izard & Dehaene, 2008; Krueger, 1972; Sullivan & Barner, 2014).

Across-task comparison results. To more directly compare the results of the two tasks, we converted the estimation data into an isomorphic format to the discrimination task. Because each trial was repeated twice for the estimation task, participants made two estimates: one for the number of blue dots and another for the number of yellow dots. By pairing identical trials, we examined whether each participant's estimate was higher for the blue dots or

for the yellow dots (as in the discrimination task). In cases where the participant's estimate was higher for the blue dots, we converted their response into a single, discrimination-like response of "more blue," while in cases where their estimate was higher for the yellow dots, we converted their response into a "more yellow" discrimination-like response. For the few cases (less than 0.5% of the trials), where the estimates were the same for the number of blue and yellow dots, we eliminated those trials from further analyses. By doing this conversion, we could reanalyze the estimation task data as if those participants were completing the discrimination task, allowing us to make direct comparisons across the two tasks.¹ If the two tasks produce an equal amount of bias given the contour length illusion, we should expect to find no interaction between task and condition.

To maximize sample size, we collapsed the pre- and postcue versions into one global estimation task. As can be seen in Figure 1 and Table 1, a 2 (Task: Discrimination, Estimation) $\times$ 4 (Condition: Both-Line, Both-Arc, Blue-Line, Blue-Arc) $\times$ 8 (Ratio) Gaisser-Greenhouse corrected mixed-measures ANOVA over the proportion of trials on which blue was selected as more numerous revealed no main effect of task, $F(1, 83) = 2.34$; $p = .13$; $\eta_p^2 = .03$, but a significant main effect of ratio, $F(7, 581) = 571.2$; $p < .001$; $\eta_p^2 = .87$; a significant main effect of condition, $F(3, 249) = 42.18$; $p < .001$; $\eta_p^2 = .34$; and a Task $\times$ Condition interaction, $F(3, 249) = 19.16$; $p < .001$; $\eta_p^2 = .19$.² This suggests that while participants showed a strong congruency effect in the discrimination task, they failed to show such an effect in the estimation task. Together, this set of tests is the strongest evidence that, while a congruency effect of contour length over number was clearly present in the discrimination task, it was eliminated in the estimation task, independent of testing participants in the pre- or the postcue version.

Discussion

Experiment 1 highlights a clear contrast between the influence of non-numeric features on number discrimination versus number estimation: Embedding blue and yellow dots within the plug-hat illusion created a very strong bias toward contour length when discriminating number, but this same bias was eradicated when response conflicts were alleviated through the estimation task (i.e., through having participants report the number of dots they saw). In other words, participants did not *perceive* the dots arranged along

¹ One key prediction of this method is that it should yield a significant effect of ratio: Because high ratios are, by definition, trials on which the blue and yellow sets are most divergent in numbers, we should expect participants to most often clearly delineate them in their estimates. For example, when shown 10 blue versus 20 yellow dots, we should expect that participants will almost always estimate a higher number for the yellow compared with the blue set, and therefore that trials with these ratios should consistently be converted to a "less blue" response (and vice-versa for trials on which there are 20 blue vs. 10 yellow dots).

² This effect held even when we analyzed the Precue and the Postcue versions of the Estimation task separately: A 2 (Task: Discrimination, Precue Estimation) $\times$ 4 (Condition) $\times$ 8 (Ratio) Gaisser-Greenhouse corrected mixed-measures ANOVA also showed a significant Task $\times$ Condition interaction, $F(3, 168) = 19.26$; $p < .001$; $\eta_p^2 = .26$, as did a parallel 2 (Task: Discrimination, Postcue Estimation) $\times$ 4 (Condition) $\times$ 8 (Ratio) Gaisser-Greenhouse corrected mixed-measures ANOVA, $F(3, 159) = 12.48$; $p < .001$; $\eta_p^2 = .19$.

Table 1
The Percentage of Time That Blue Was Selected as the More Numerous Set, Across the Three Experiments and Two Conditions

	Condition	Discrimination task	Estimation task (Pre/Postcue)
Experiment 1 (Contour length)	Both-Line	52.5 (1.3)	53.6 (1.04) (53.9/53.2)
	Both-Arc	49.1 (1.1)	50.2 (.90) (50.4/49.9)
	Blue-Line	56.7 (1.6)	51.3 (1.0) (51.7/50.9)
	Blue-Arc	38.9 (1.2)	48.7 (.99) (49.0/48.2)
Experiment 2 (Convex hull)	Both-Small	48.4 (2.0)	44.6 (1.2)
	Both-Large	49.9 (1.1)	44.7 (1.4)
	Blue-Small	67.5 (2.5)	61.4 (1.7)
	Blue-Large	33.3 (2.0)	29.6 (1.6)
Experiment 3 (Density/convex hull)	Both-Expanded	46.1 (1.2)	45.7 (1.3)
	Both-Compressed	45.4 (1.0)	49.9 (1.1)
	Blue-Expanded	56.3 (2.3)	58.4 (1.8)
	Blue-Compressed	37.5 (2.3)	36.9 (1.9)

Note. If there is no bias, we would expect that blue should be selected 50% of the time; values above 50% indicate a bias towards selecting blue as more numerous, while values below 50% indicate a bias towards selecting yellow as more numerous.

a line as being more numerous in the estimation task, but they did select them as such in the discrimination task. This powerfully demonstrates that congruency effects can stem from response conflicts generated in typical discrimination tasks, and that by eliminating these conflicts, the congruency effect is eliminated as well. Furthermore, it suggests that contour length is *not* used during number encoding (and, as we discuss in the online supplementary materials, a control experiment also demonstrates that these results cannot be accounted for by more general task differences between estimation and discrimination). This lack of bias in estimation was present both when participants were told which set of dots to estimate *before* seeing the stimuli (i.e., as in classic estimation experiments) and in a version when they were told which set to estimate *after* seeing the stimuli (which assured that they attended to both colors equally).

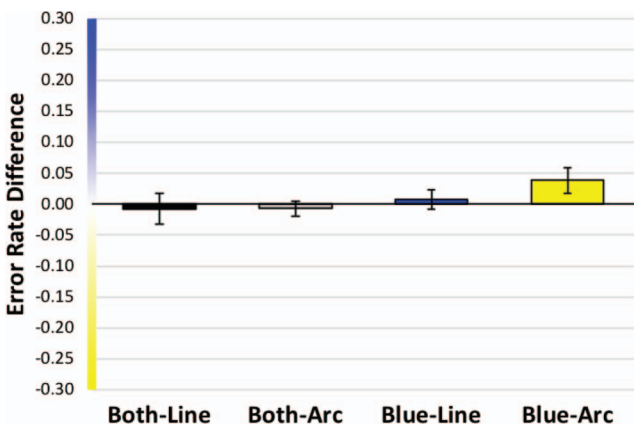


Figure 2. Error rate differences for the blue versus the yellow set for the Experiment 1 estimation condition. If participants are showing a positive error rate difference, they are estimating the blue set higher than the yellow set; if participants are showing a negative error rate difference, they are estimating the yellow set higher than the blue set. Bars indicate standard error. See the online article for the color version of this figure.

In Experiment 2, we use this same logic to test a different non-numeric feature that is often thought to contribute to perception's sensitivity to numerical information: convex hull/density.

Experiment 2: Convex Hull and the Ebbinghaus Illusion

Method

Participants. Thirty-six participants completed Experiment 2 for university course credit. Because we reached the end of the semester at the time of testing, we did not test the target 22 participants per condition, but believed that our sample size ($n = 36$) was large enough to conclude the study and proceeded with the data analysis. Half of the participants completed the discrimination task ($n = 18$), while the other half completed the postcue estimation task ($n = 18$; we did not run participants in the precue version of the estimation task, as we found no differences between the two conditions in Experiment 1). None of the participants in Experiment 2 had previously completed Experiment 1.

Procedure. The procedures in Experiment 2 are identical to those in Experiment 1, with the only change being the stimuli. Instead of varying contour length, we manipulated the convex hull through the Ebbinghaus illusion, in which an object's perceived size (in our case, the size of the set of dots) is affected by surrounding "context circles." The total number of dots in each set varied from 15–34, and the ratios were 1.07, 1.25, 1.50, and 2.0 (as in Experiment 1). For this experiment, however, the blue and yellow dots were arranged within a circle with a diameter of 300 pixels (see Figure 3), with each side surrounded by either five large circles (each with a diameter of 80 pixels) or eight small circles (each with a diameter of 30 pixels). This resulted in four conditions: both-small (both the blue and yellow dots were surrounded by small context circles), both-large (both the blue and yellow dots were surrounded by large context circles), blue-small (the blue dots were surrounded by small context circles, and the yellow dots by large context circles), and blue-large (the blue dots were surrounded by large context circles, and yellow dots by small context

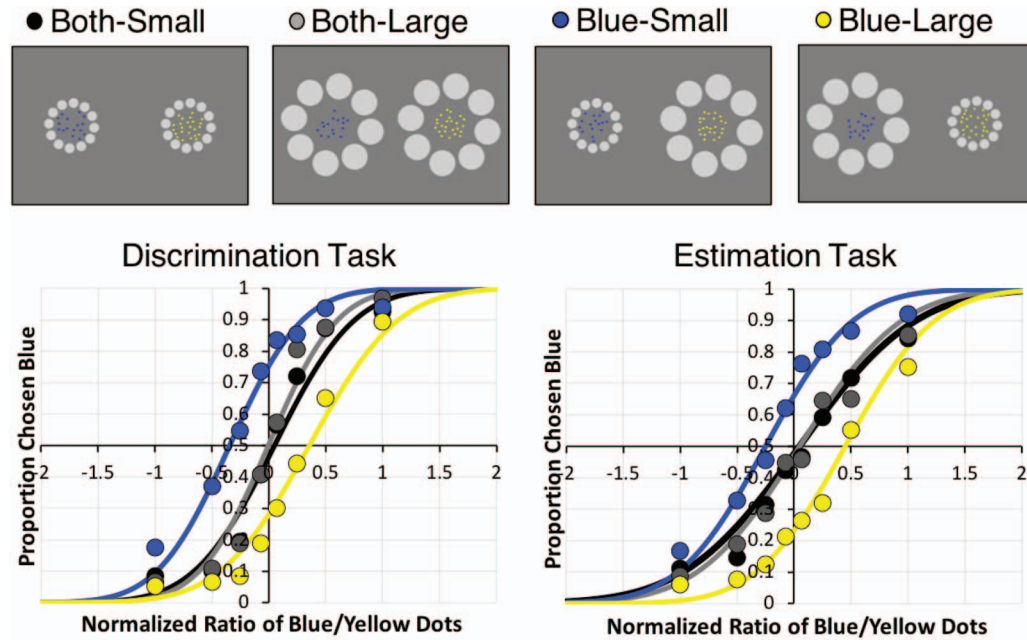


Figure 3. Experiment 2 stimuli and results. The top panel illustrates four example trials, one from each of the four conditions. The bottom panel shows the data from the Experiment 2 discrimination and estimation conditions; the lines are the best-fit psychophysical cumulative normal model. See the online article for the color version of this figure.

circles). Given that the Ebbinghaus illusion results in the perception of the object within smaller context circles appearing as further extended, we predicted that if participants used convex hull cues in their decision, they would be biased toward the side with small context circles, perceiving it as a more numerous, independent of number.

Results

Number discrimination results. Participants showed excellent accuracy in this task, averaging 76.3% ($SE = 0.40$). A 4 (Condition: Both-Small, Both-Large, Blue-Small, Blue-Large) $\times$ 8 (Ratio: 2.0, 1.50, 1.25, 1.07, 0.93, 0.80, 0.66, 0.50) Gaiser-Greenhouse corrected repeated-measures ANOVA over the proportion of trials on which blue was selected as more numerous revealed a significant main effect of condition, $F(3, 90) = 16.94$; $p < .001$; $\eta_p^2 = .36$; a significant main effect of ratio, $F(7, 210) = 380.71$; $p < .001$; $\eta_p^2 = .93$; and a significant Condition $\times$ Ratio interaction, $F(21, 630) = 7.29$; $p < .001$; $\eta_p^2 = .20$. These effects were carried by a clear bias participants had for selecting the blue side as more numerous in the blue-small condition ($M = 67.5\%$; $SE = 2.5\%$), and the yellow side as more numerous in the blue-large condition ($M = 33.3\%$; $SE = 2.0\%$), suggesting that our manipulation of convex hull had a large impact on performance (see Table 1 and Figure 3). Taken together, these results show a clear congruency effect based on the subjectively perceived convex hull.

Number estimation results. In contrast to Experiment 1, however, we also found evidence for congruency effects in the estimation task: A 4 (Condition: Both-Small, Both-Large, Blue-Small, Blue-Large) $\times$ 2 (Dot Set: Blue, Yellow) Gaiser-

Greenhouse corrected repeated-measures ANOVA over error rates showed a significant main effect of condition, $F(3, 51) = 26.52$; $p < .001$; $\eta_p^2 = .61$; no significant main effect of dot set, $F(1, 17) = 27.34$; $p < .001$; $\eta_p^2 = .62$; and, most importantly, a significant Condition $\times$ Dot Set interaction, $F(3, 51) = 42.86$; $p < .001$; $\eta_p^2 = .72$. As shown in Figure 4, participants were much more likely to show positive error rates (i.e., overestimation) for the blue dots in the blue-small condition, and more likely to show negative error rates (i.e., underestimation) for the blue dots in the

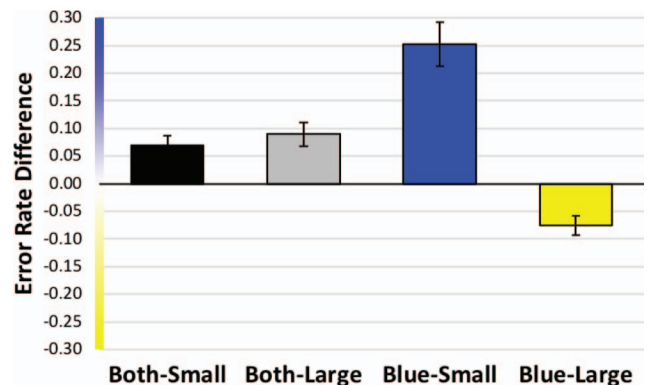


Figure 4. Error rate differences for the blue versus the yellow set for the Experiment 2 estimation condition. If participants are showing a positive error rate difference, they are estimating the blue set higher than the yellow set; if participants are showing a negative error rate difference, they are estimating the yellow set higher than the blue set. Bars indicate standard error. See the online article for the color version of this figure.

blue-large condition. These result cannot be accounted for by random guessing, as the estimation slopes were significantly higher than 0 ($M = 0.72$; $SE = 0.08$; $t(17) = 8.56$; $p < .001$).

Across-task comparison results. As with Experiment 1, in order to directly compare the results of the two tasks, we converted the estimation data into an isomorphic format to the discrimination task. A 2 (Task: Discrimination, Estimation) $\times$ 4 (Condition: Both-Small, Both-Large, Blue-Small, Blue-Large) $\times$ 8 (Ratio) Mixed-Measures ANOVA with the proportion of trials on which blue was selected as more numerous showed no main effect of condition, $F(3, 102) = 1.86$; $p = .14$; $\eta_p^2 = .05$; but a main effect of ratio, $F(7, 238) = 320.7$; $p < .001$; $\eta_p^2 = .90$; task, $F(1, 34) = 4.51$; $p < .05$; $\eta_p^2 = .12$; and a Task $\times$ Condition interaction, $F(3, 102) = 66.06$; $p < .001$; $\eta_p^2 = .66$. As can be seen in Figure 4, this interaction was carried by an even *larger effect* in the estimation task compared with the discrimination task, even in the blue-small and blue-large conditions.

Across-experiment comparison. Finally, to further understand the differences between the two experiments, we performed an across-experiment analysis, using the combined estimation task (i.e., the combined pre- and postcue versions) from Experiment 1; the results reported below remain identical if we only use the postcue version. A 2 (Experiment: 1, 2) $\times$ 2 (Task: Discrimination, Estimation) $\times$ 4 (Condition) mixed-measures ANOVA with the proportion of trials on which blue was selected as more numerous revealed a main effect of experiment, $F(1, 117) = 30.68$; $p < .001$; $\eta_p^2 = .20$, carried by a greater overall bias for yellow in Experiment 2, no main effect of task, $F(1, 117) < 1$; $p = .38$; $\eta_p^2 = .01$; a main effect of condition, $F(3, 351) = 20.32$; $p < .001$; $\eta_p^2 = .15$; and, most importantly, a three-way Condition $\times$ Experiment $\times$ Task interaction, $F(3, 351) = 101.78$; $p < .001$; $\eta_p^2 = .47$, demonstrating that Experiment 2 carried a congruency effect in both estimation and discrimination while Experiment 1 did not. This suggests that both tasks showed a significant bias effect, but that the estimation condition carried an even higher overall degree of bias compared with the discrimination condition in the blue-small and blue-large conditions. Hence, unlike in Experiment 1, whereby congruency effects in contour length over number were eliminated during the estimation task, in Experiment 2 we find evidence for congruency effects in convex hull over number persisting across tasks.

Discussion

In Experiment 2, we used the Ebbinghaus illusion to test whether congruency effects involving convex hull stem from response biases. We once again found large congruency effects in the discrimination task: Participants selected the side with the larger perceived convex hull (that is, smaller context circles) as the “more numerous” side. However, in contrast to Experiment 1, we found that this congruency effect persisted in the Estimation task: Participants significantly *overestimated* the number of dots presented in a larger convex hull and *underestimated* the number of dots presented in the smaller convex hull. Beyond suggesting that congruency effects in convex hull are not a mere byproduct of Stroop-like response conflicts, this positive result also showcases that the estimation task *can* produce biases when appropriate non-numeric features are present. In other words, the Ebbinghaus

illusion produced a difference in how number was *perceived*, not only in how responses were selected.

In Experiment 3, we control for two alternative explanations to these findings. First, in Experiment 2, the condition with the larger convex hull also had more small context circles (see Figure 3). Hence, one explanation for our finding is that participants may have been distracted by the context circles and included them in their estimate of blue/yellow dots, therefore inflating their small context circle estimates. To account for this possibility, Experiment 3 removes the Ebbinghaus illusion and instead physically manipulates the convex hull/density by having the dots placed within physically larger or smaller areas (see Figure 5). Second, recent work by Anobile, Cicchini, and Burr (2014) has shown that convex hull and density may only play a role in number encoding when the dot stimuli are highly crowded, such as in situations where more than 35 dots are on the screen. To test whether convex hull/density affects number perception across a large range of numbers, Experiment 3 has participants perform the discrimination and estimation tasks both with stimuli with fewer than 35 dots (e.g., 10 vs. five dots for ratio 2.0), and with stimuli made up of well over 35 dots (e.g., 80 vs. 40 dots for ratio 2.0).

Experiment 3: Density and Dot Number

Method

Participants. Sixty-two participants completed Experiment 3, with none having participated in either Experiment 1 or 2. We once again had a target of 22 participants per condition but allowed remaining participants who had already signed-up, prior to us reaching this goal, to complete the study for university course credit. This left us with a final sample of 31 per condition: Half the participants ($n = 31$) completed the discrimination task, and half completed the postcue estimation task ($n = 31$).

Procedure. The procedures in Experiment 3 are consistent with Experiments 1 and 2, with the only change being the stimuli. Rather than relying on a visual illusion that changed the subjective experience of convex hull and density, we created stimuli in which the dots were objectively presented within a larger convex hull (i.e., higher density) or within a smaller convex hull (i.e., lower density). The large convex hull stimuli were drawn within an invisible circle measuring 300 pixels in radius, while the small convex hull stimuli were drawn within an invisible circle measuring 170 pixels in radius. This generated four conditions: both-expanded (i.e., both blue and yellow dot arrays having low density), both-compressed (i.e., both blue and yellow dots arrays having high density), blue-expanded (i.e., blue dots having low density, and yellow dots having a high density), and blue-compressed (i.e., blue dots having a high density, and yellow dots having a low density; see Figure 6). We expected that participants would be biased toward the set with the higher density if our manipulation had an effect on performance.

To also test whether the effect of density is independent across various numbers of dots (i.e., low vs. high), we varied the total number of dots in each set to either be low on both sides (i.e., less than 35 cumulative dots) or high on both sides (i.e., between 85 and 150 cumulative dots). As in Experiments 1 and 2, we generated four ratios, each of which was made up of four different number combinations ranging from low to high: 2.0 (low: 10:5 or

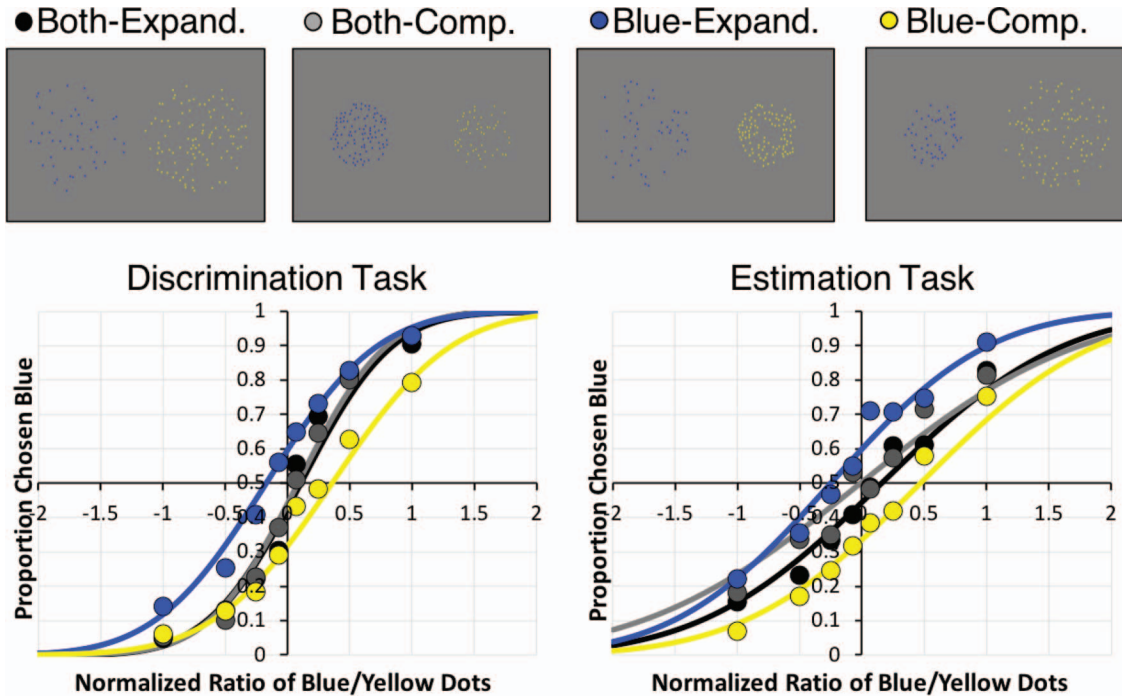


Figure 5. Experiment 3 stimuli and results. The top panel illustrates four example trials, one from each of the four conditions. The bottom panel shows the data from the Experiment 3 discrimination and estimation conditions; the lines are the best-fit psychophysical cumulative normal model. See the online article for the color version of this figure.

20:10 dots; high: 80:40 or 100:50 dots), 1.5 (low: 12:8 or 21:14; high: 45:30 or 75:50 dots), 1.25 (low: 10:8 or 15:12; high: 50:40 or 60:48), and 1.07 (low: 15:14; high: 30:28 or 45:42). For half of the trials, the blue dots were more numerous, and for the other half the yellow dots were more numerous. Therefore, each of the four main conditions could also be crossed with either a low-number or high-number condition.

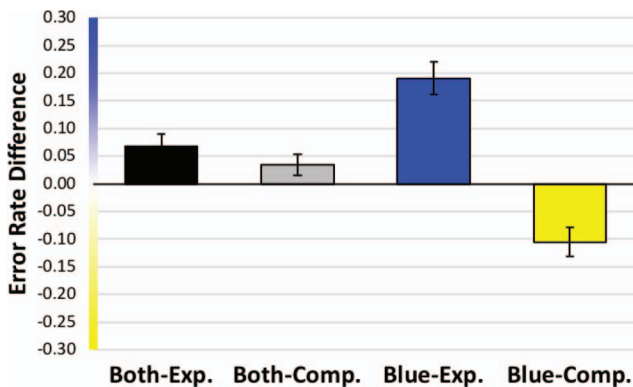


Figure 6. Error rate differences for the blue versus the yellow set for the Experiment 3 estimation condition. If participants are showing a positive error rate difference, they are estimating the blue set higher than the yellow set; if participants are showing a negative error rate difference, they are estimating the yellow set higher than the blue set. Bars indicate standard error. See the online article for the color version of this figure.

Results

Number discrimination results. Participants showed excellent accuracy in this task, averaging 78.3% ($SE = 1.43$). A 4 (Condition: Both-Compressed, Both-Expanded, Blue-Expanded, Blue-Compressed) $\times$ 8 (Ratio: 2.0, 1.50, 1.25, 1.07, 0.93, 0.80, 0.66, 0.50) Gaisser-Greenhouse corrected repeated-measures ANOVA over the proportion of trials on which blue was selected as more numerous revealed a significant main effect of condition, $F(3, 90) = 16.93$; $p < .001$; $\eta_p^2 = .36$; a significant main effect of ratio, $F(7, 210) = 380.71$; $p < .001$; $\eta_p^2 = .93$; and a significant Condition $\times$ Ratio interaction, $F(21, 630) = 7.29$; $p < .001$; $\eta_p^2 = .20$. These effects appear to be carried by a clear bias in participants for selecting the set of blue dots as more numerous in the blue-expanded condition ($M = 56.3\%$; $SE = 2.3\%$), and the yellow dots as more numerous in the blue-compressed condition ($M = 37.5\%$; $SE = 2.3\%$; see Figure 5). This suggests that our manipulation of objective convex hull and density had a large impact on performance.

As noted earlier, recent work has suggested that large sets of dots—around 35 or more—may be encoded through density, while small sets of dots may be encoded through number-specific features (Anobile et al., 2014). To test this, we split our data into trials into low-number trials (i.e., displays with fewer than or exactly 35 dots on each side) and high-number trials (i.e., displays with more than 35 dots on each side). Contrary to previous work, we found congruency effects for both the large and small sets of dots: a 4 (Condition: Both-Compressed, Both-Expanded, Blue-Expanded, Blue-Compressed) $\times$ 2 (Number: Low-, High-Number) Gaisser-

Greenhouse corrected repeated-measures ANOVA over the proportion of trials participants judged blue as more numerous revealed a main effect of condition, $F(3, 90) = 16.54$; $p < .001$; $\eta_p^2 = .36$, but no main effect of number, $F(1, 30) = 1.75$; $p = .20$; $\eta_p^2 = .06$; and no Condition $\times$ Number interaction, $F(3, 90) = 1.91$; $p = .13$; $\eta_p^2 = .06$. Hence, we find that the congruency effect for density occurs across the entire tested range for the discrimination task.

Number estimation results. As in Experiment 2, we also found evidence for a significant bias in the Estimation task: a 4 (Condition: Both-Compressed, Both-Expanded, Blue-Expanded, Blue-Compressed) $\times$ 2 (Dot Set: Blue, Yellow) Gaisser-Greenhouse corrected repeated-measures ANOVA over error rates showed a significant main effect of condition, $F(3, 90) = 10.32$; $p < .001$; $\eta_p^2 = .26$; a significant main effect of dot set, $F(1, 30) = 6.91$; $p = .013$; $\eta_p^2 = .19$; and, most importantly, a significant Condition $\times$ Dot Set interaction, $F(3, 90) = 42.13$; $p < .001$; $\eta_p^2 = .58$. Participants were much more likely to show positive error rates (i.e., overestimation) for the blue dots in the blue-expanded condition, and more likely to show negative error rates (i.e., underestimation) for the blue dots in the blue-compressed condition (see Figure 5). We again find that these results persist when examining estimation slopes and average guesses for each set, and cannot be accounted for by random guessing as evidenced in participants' estimation slopes being significantly greater than zero ($M = 0.45$; $SE = 0.04$; $t(30) = 11.00$; $p < .001$).

We also examined whether participants show distinct biases when the total number of dots is higher than 35 in estimation. Accordingly, we split the estimation data into high-number trials (i.e., more than 35 dots on each side) and low-number trials (i.e., fewer than or exactly 35 dots on each side). A 2 (Condition: Blue-Expanded, Blue-Compressed) $\times$ 2 (Number: High/Low) $\times$ 2 (Dot Set: Blue, Yellow) repeated-measures ANOVA over average error rates for the blue versus yellow set revealed a main effect of number, $F(1, 30) = 291.32$; $p < .001$; $\eta_p^2 = .91$, as participants tended to overestimate the low sets of numbers and underestimate the high sets of numbers. We found no main effect of condition, $F(1, 30) < 1$; $p = .63$; $\eta_p^2 = .01$; or dot set, $F(1, 30) = 3.51$; $p = .07$; $\eta_p^2 = .11$; though, we did find a marginally significant Condition $\times$ Number $\times$ Dot Set interaction, $F(1, 30) = 4.28$; $p = .047$; $\eta_p^2 = .13$, that was driven by a larger difference in error rates between blue and yellow dots in the blue-expanded low-number condition, compared with a larger difference between the blue and yellow dots in the blue-compressed high-number condition. In both conditions, however, the estimates suggest a strong degree of blue dot overestimation in the blue-expanded condition and blue dot underestimation in the blue-compressed condition, consistent with the discrimination data.

Across-task comparison results. As with Experiment 1 and 2, we again converted the estimation data into an isomorphic format to the discrimination task, allowing for more direct comparisons across tasks. A 2 (Task: Discrimination, Estimation) $\times$ 4 (Condition: Both-Compressed, Both-Expanded, Blue-Expanded, Blue-Compressed) $\times$ 8 (Ratio) mixed-measures ANOVA over the proportion of trials blue was selected as more numerous showed a main effect of condition, $F(3, 180) = 49.76$; $p < .001$; $\eta_p^2 = .45$; and ratio, $F(7, 329) = 439.8$; $p < .001$; $\eta_p^2 = .90$; but no main effect of task, $F(1, 60) = 1.17$; $p = .28$; $\eta_p^2 = .020$; nor a Task $\times$ Condition interaction, $F(3, 180) = 1.06$; $p = .37$; $\eta_p^2 = .02$. These

results suggest that the estimation and discrimination tasks both showed an equivalent amount of bias in the blue-expanded and blue-compressed conditions. Therefore, in replication of Experiment 2, we find strong evidence for a congruency effect in both the discrimination and estimation tasks for convex hull/density across both small and large dot number sets

Across-experiment comparison results. Finally, a 2 (Experiment: 1, 3) $\times$ 2 (Task: Discrimination, Estimation) $\times$ 4 (Condition) Mixed-Measures ANOVA with the proportion of trials on which blue was selected as more numerous revealed a main effect of experiment, $F(1, 143) = 12.98$; $p < .001$; $\eta_p^2 = .08$; no main effect of task, $F(1, 143) = 3.09$; $p = .08$; $\eta_p^2 = .02$; a main effect of condition, $F(3, 429) = 10.46$; $p < .001$; $\eta_p^2 = .07$; and, most importantly, a three-way Condition $\times$ Experiment $\times$ Task interaction, $F(3, 429) = 4.67$; $p < .01$; $\eta_p^2 = .03$, demonstrating that unlike Experiment 1, Experiment 3 carried a congruency effect in both Estimation and Discrimination.

Discussion

In Experiment 3, we objectively manipulated the convex hull and density of the arrays of dots, making sure that both the context circles used in the Ebbinghaus illusion and the use of visual illusions more generally did not contribute to us finding congruency effects in estimation in Experiment 2. We found that participants showed a strong congruency effect in both discrimination and estimation tasks, consistent with the use of these cues during number encoding. Furthermore, we found that these effects were identical for smaller (i.e., less than 35 dots) versus larger (i.e., more than 35 dots) displays, suggesting that convex hull/density contributes to number encoding even when crowding is controlled for (Anobile et al., 2014).

General Discussion

What information does our perceptual system use to distinguish number in visual displays? To resolve issues surrounding the competing domain-specific versus domain-general encoding theories, we compared participants' performance over identical sets of stimuli on a number discrimination task (i.e., "Which side has more dots?") against their performance on a number estimation task (i.e., "How many blue/yellow dots are there?") for contour length and convex hull/density, two dimensions that have both previously been argued to play a key role in number perception. Importantly, the number estimation task allowed us to eliminate the possibility that congruency effects could be stemming from Stroop-like response conflicts (e.g., see Smets, Sasanguie, Szűcs, & Reynvoet, 2015). By embedding visual illusions (e.g., contour length and convex hull) in our stimuli, we were able to magnify the influence of non-numeric dimensions on number and precisely target the subjective perception of the two non-numeric features, leading to three key findings: (1) while embedding dots into a contour length illusion produced a large congruency effect in the discrimination task, no such effect was observed in the estimation task, suggesting that contour length creates a Stroop-like response conflict with number rather than contributing directly to its encoding; (2) embedding dots into a display that varied in convex hull and/or density, produced a large congruency effect in *both* discrimination and estimation tasks, suggesting that response biases

are not the sole explanation for why participants use convex hull/density during number perception; and (c) the convex hull/density results are observed for small sets of dots (i.e., less than 35) as well as for large sets of dots, suggesting that these findings are not simply due to the effects of perceived crowding at higher number ranges (Anobile et al., 2014). Taken together, these results provide several novel insights for disentangling the contributions of non-numeric features to number perception while controlling for response conflicts.

First, our results provide evidence for the domain-general encoding account: The visual features used to encode convex hull and/or density appear to be used to encode number, contrary to the domain-specific encoding theory (Anobile et al., 2016; Burr & Ross, 2008). Additionally, the influence of density appears to hold even for sets of dots well below 35. Given that density has been shown to be computed through low-spatial frequency detectors in early vision (Dakin et al., 2011), we hypothesize that convex hull and density must share these low-level features with number, such that a change in density carries a change in the low-spatial frequency signal that also impacts our perceived sense of number. Our results are therefore broadly consistent with previous work that has also shown that non-numeric dimensions contribute to number encoding (e.g., Gebuis & Reynvoet, 2012b; Inglis & Gilmore, 2014; Leibovich et al., in press; Leibovich-Raveh et al., 2018; Salti et al., 2017).

Furthermore, the asymmetry between the contributions of contour length and convex hull/density also provide constraints for any future models of number encoding: Because convex hull and density have been shown to be strongly associated with low-spatial frequency detectors (Dakin et al., 2011), any model of number encoding will necessarily have to account for the contribution of these detectors to number perception. On the other hand, given that contour length taps into low-level detectors for orientation (Li, 1998), any model of number encoding should function without the use of this featural information.

Second, our results show a novel method for disentangling the influence of non-numeric dimensions for number encoding versus response conflicts. We believe this contrast is useful not only for further understanding which visual features contribute to our encoding of number, but also for being able to test how participants' ability to deal with response conflicts changes and affects number perception across development. Therefore, given that we know that some illusions (e.g., the plug-hat illusion) generate a difference in response conflicts, while others (e.g., Ebbinghaus illusion) generate differences in encoding processes, we can distinguish the influences of each when studying number perception across a variety of disciplines and platforms—developmental work, neurocognitive experiments, and so forth.

Future work can easily extend this method to other non-numeric dimensions, allowing us to disentangle their contributions during encoding versus response selection. For example, to dissociate the influence of size on number perception, researchers can embed number stimuli within the classic Ponzo illusion (i.e., whereby two identically sized dots are superimposed over a picture of a receding hallway, with the dot that is “further away” subsequently being perceived as much larger). In this way, identically sized arrays of dots could be superimposed to either appear further (and therefore subjectively larger) or closer. Similarly, superimposing gray dots on a gradient of white to black would create an apparent contrast

illusion, allowing researchers to decipher the contributions of contrast to number encoding versus response conflicts. Given that the early visual properties of these illusions are well studied, understanding how they affect number perception should allow researchers to further understand which early visual features are used to represent number as well.

Thus far, we have discussed congruency effects as stemming from either response competition or from the active use of specific low-level features for number encoding. However, previous work has shown that number perception can also be manipulated by contexts that are independent of congruency effects. For example, emotionally arousing stimuli, such as displays of happy or angry faces, have been shown to affect number perception (Baker, Rodzon, & Jordan, 2013; Lewis, Zax, & Cordes, 2017; Young & Cordes, 2013). We believe that such effects are largely consistent with our argument here: given that emotionally arousing stimuli are thought to modulate attention and/or arousal, which in turn may either affect encoding procedures (e.g., speeding up an internal number accumulator) or number representations themselves (Hamamouche, Niemi, & Cordes, 2017; Lewis et al., 2017; Young & Cordes, 2013), and that they create effects in both discrimination and estimation tasks (Hamamouche et al., 2017), it is clear that emotional stimuli have an effect over number *perception*, not decision-making components. This pattern of results raises the possibility that the convex hull/density effects observed here are also somehow modulating the level of representations rather than encoding. Although, in the absence of a model that shows how these low-level features could modulate representations in a similar vein to attention and/or arousal, we believe that the most parsimonious explanation of our data is that these features contribute to encoding itself (after all, several feasible models of number encoding that invoke low-spatial frequency have already been proposed, e.g., Dakin et al., 2011).

One concern for the interpretation of our results is that the estimation and discrimination tasks may vary in many more ways than just response selection. For example, number estimation has previously been shown not to be a direct read-out of the number sense signal, but instead requires categorization of a continuous signal to a discrete, symbolic one (Odic, Le Corre, & Halberda, 2015). Number estimation has also been shown to be susceptible to feedback and expectation effects (Izard & Dehaene, 2008; Sullivan & Barner, 2014), and, compared with discrimination, very slow to develop (Le Corre & Carey, 2007). However, we have two reasons to suspect that these general task differences cannot account for our results. First, as discussed in the online supplementary materials, we find that a reintroduction of the binary “blue”/“yellow” responses (i.e., a response set that allows for non-numeric dimensions to compete for the same response as number) reinstates the congruency effects: When asked to identify which set has a particular number of dots (e.g., “15”), participants show congruency effects with contour length stimuli despite having to map the continuous perceptual number signal to discrete number words. Second, given that our three experiments varied only in the non-numeric dimensions being shown, it is clear that estimation does not necessarily eliminate or induce congruency effects by itself, but that the patterns observed here are primarily driven by the saliency of the non-numeric dimensions shown in particular illusions. In other words, whatever additional processes are required for participants to estimate, they do not automatically lead to the

elimination of congruency effects—as shown by Experiments 2 and 3—nor to creating them—as shown by Experiment 1.

A more significant limitation of our current work, however, is that it is entirely focused on adult observers. Under some accounts, number encoding may begin as a domain-general collection of features that slowly integrates and becomes more domain-specific with age and experience (e.g., Leibovich et al., 2017). Given that children do not robustly learn to estimate number displays until at least age 5, our method is also only appropriate for children from preschool onward.

In conclusion, by examining the effects of various non-numeric dimensions while controlling for the influence of response biases, we find support for a more domain-general account of number encoding. Our results suggest that number is most likely not a primary visual feature of perception, but rather is derived from a set of other features that are broadly shared with convex hull and density encoding. In that sense, numerical perception appears to be made up of the same building blocks as other midlevel representations (e.g., as with face or object perception; Fiorentini, Maffei, & Sandini, 1983; Vuilleumier, Armony, Driver, & Dolan, 2003), thereby placing number in a broader category of visual representations.

This work is part of a broader theory explored by the authors on how our mind represents quantity in general and number in particular. Previous work in our lab has shown that number representations show distinct developmental and individual differences compared with other dimensions, such as area or length (e.g., Odic, 2017). However, while this work focused exclusively on the level of number representations, a major question was left unexplored regarding how early perceptual processes actually encode these representations. Here, we attempt to disentangle the contributions of specific perceptual features on number encoding versus response competition by approaching this challenge in a novel way. We emphasize the important insights that can be gained by examining the variety of operations that observers can perform over their number representations, rather than studying number representations in isolation. Future work will seek to extend these findings across broader contexts (e.g., across development, other non-numeric features, etc.).

References

- Anobile, G., Cicchini, G. M., & Burr, D. C. (2014). Separate mechanisms for perception of numerosity and density. *Psychological Science*, 25, 265–270. <http://dx.doi.org/10.1177/0956797613501520>
- Anobile, G., Cicchini, G. M., & Burr, D. C. (2016). Number as a primary perceptual attribute: A review. *Perception*, 45, 5–31. <http://dx.doi.org/10.1177/0301006615602599>
- Baker, J. M. P. D., Rodzon, K. M. A., & Jordan, K. P. D. (2013). The impact of emotion on numerosity estimation. *Frontiers in Psychology*, 4. <http://dx.doi.org/10.3389/fpsyg.2013.00521>
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436. <http://dx.doi.org/10.1163/156856897X00357>
- Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology*, 18, 425–428. <http://dx.doi.org/10.1016/j.cub.2008.02.052>
- Cantlon, J. F., Safford, K. E., & Brannon, E. M. (2010). Spontaneous analog number representations in 3-year-old children. *Developmental Science*, 13, 289–297. <http://dx.doi.org/10.1111/j.1467-7687.2009.00887.x>
- Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science*, 10, 408–411.
- Clearfield, M. W., & Mix, K. S. (2001). Amount versus number: Infants' use of area and contour length to discriminate small sets. *Journal of Cognition and Development*, 2, 243–260.
- Cordes, S., Gallistel, C. R., Gelman, R., & Latham, P. (2007). Nonverbal arithmetic in humans: Light from noise. *Attention, Perception, & Psychophysics*, 69, 1185–1203. <http://dx.doi.org/10.3758/BF03193955>
- Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, 8, 698–707.
- Crollen, V., Castronovo, J., & Seron, X. (2011). Under- and over-estimation: A bi-directional mapping process between symbolic and non-symbolic representations of number? *Experimental Psychology*, 58, 39–49. <http://dx.doi.org/10.1027/1618-3169/a000064>
- Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 19552–19557. <http://dx.doi.org/10.1073/pnas.1113195108>
- Defever, E., Reynvoet, B., & Gebuis, T. (2013). Task- and age-dependent effects of visual stimulus properties on children's explicit numerosity judgments. *Journal of Experimental Child Psychology*, 116, 216–233. <http://dx.doi.org/10.1016/j.jecp.2013.04.006>
- Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, 5, 390–407. <http://dx.doi.org/10.1162/jocn.1993.5.4.390>
- 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. <http://dx.doi.org/10.1016/j.cognition.2015.05.016>
- Durkin, F. H. (1995). Texture density adaptation and the perceived numerosity and distribution of texture. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 149–169.
- Durkin, F. H. (2000). The reverse Stroop effect. *Psychonomic Bulletin & Review*, 7, 121–125. <http://dx.doi.org/10.3758/BF03210730>
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8, 307–314. <http://dx.doi.org/10.1016/j.tics.2004.05.002>
- Ferrigno, S., Jara-Ettinger, J., Piantadosi, S. T., & Cantlon, J. F. (2017). Universal and uniquely human factors in spontaneous number perception. *Nature Communications*, 8, 13968. <http://dx.doi.org/10.1038/ncomms13968>
- Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception*, 12, 195–201. <http://dx.doi.org/10.1068/p120195>
- Gebuis, T., & Reynvoet, B. (2012a). Continuous visual properties explain neural responses to nonsymbolic number. *Psychophysiology*, 49, 1481–1491. <http://dx.doi.org/10.1111/j.1469-8986.2012.01461.x>
- Gebuis, T., & Reynvoet, B. (2012b). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General*, 141, 642–648. <http://dx.doi.org/10.1037/a0026218>
- Gebuis, T., & van der Smagt, M. J. (2011). False approximations of the approximate number system? *PLoS ONE*, 6, e25405. <http://dx.doi.org/10.1371/journal.pone.0025405>
- Hamamouche, K. A., Niemi, L., & Cordes, S. (2017). Quantifying a threat: Evidence of a numeric processing bias. *Acta Psychologica*, 177, 1–9. <http://dx.doi.org/10.1016/j.actpsy.2017.04.001>
- Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 19599–19604. <http://dx.doi.org/10.1073/pnas.0609485103>

- Im, H. Y., & Chong, S. C. (2009). Computation of mean size is based on perceived size. *Attention, Perception & Psychophysics*, 71, 375–384. <http://dx.doi.org/10.3758/APP.71.2.375>
- Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, 145, 147–155.
- Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, 106, 1221–1247. <http://dx.doi.org/10.1016/j.cognition.2007.06.004>
- Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 10382–10385. <http://dx.doi.org/10.1073/pnas.0812142106>
- Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, 105, 395–438.
- Leibovich, T., Al-Rubaiey Kadhim, S., & Ansari, D. (in press). Beyond comparison: physical size affects non-symbolic but not symbolic number line estimation. *Acta Psychologica*.
- Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From “sense of number” to “sense of magnitude:” The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences*, 40, 1–62. <http://dx.doi.org/10.1017/S0140525X16000960>
- Leibovich-Raveh, T., Stein, I., Henik, A., & Salti, M. (2018). Number and continuous magnitude processing depends on task goals and numerosity ratio. *Journal of Cognition*, 1, 19.
- Lewis, E. A., Zax, A., & Cordes, S. (2017). The impact of emotion on numerical estimation: A developmental perspective. *The Quarterly Journal of Experimental Psychology*, 71, 1300–1311. <http://dx.doi.org/10.1080/17470218.2017.1318154>
- Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation*, 10, 903–940. <http://dx.doi.org/10.1162/089976698300017557>
- Libertus, M. E., Odic, D., Feigenson, L., & Halberda, J. (in press). Number estimation abilities uniquely correlate with formal math. *Journal of Experimental Child Psychology*.
- Krueger, L. E. (1972). Perceived numerosity. *Attention, Perception, & Psychophysics*, 11, 5–9.
- Marques, J. F., & Dehaene, S. (2004). Developing intuition for prices in euros: Rescaling or relearning prices? *Journal of Experimental Psychology: Applied*, 10, 148–155. <http://dx.doi.org/10.1037/1076-898X.10.3.148>
- Odic, D. (2017). Children’s intuitive sense of number develops independently of their perception of area, density, length, and time. *Developmental Science*, 21, e12533. <http://dx.doi.org/10.1111/desc.12533>
- Odic, D., Im, H. Y., Eisinger, R., Ly, R., & Halberda, J. (in press). PsiMLE: A maximum-likelihood approach to estimating psychophysical scaling and variability more reliably, efficiently, and flexibly. *Behavior Research and Methods*.
- Odic, D., Le Corre, M., & Halberda, J. (2015). Children’s mappings between number words and the approximate number system. *Cognition*, 138, 102–121.
- Odic, D., & Starr, A. (2018). An introduction to the approximate number system. *Child Development Perspectives*, 12, 223–229. <http://dx.doi.org/10.1111/cdep.12288>
- Piantadosi, S. T., & Cantlon, J. F. (2017). True numerical cognition in the wild. *Psychological Science*, 28, 462–469. <http://dx.doi.org/10.1177/0956797616686862>
- Ross, J., & Burr, D. C. (2010). Vision senses number directly. *Journal of Vision*, 10, 1–8. <http://dx.doi.org/10.1167/10.2.10>
- Ross, J., & Burr, D. (2012). Number, texture and crowding. *Trends in Cognitive Sciences*, 16, 196–197. <http://dx.doi.org/10.1016/j.tics.2012.01.010>
- Salti, M., Katzin, N., Katzin, D., Leibovich, T., & Henik, A. (2017). One tamed at a time: A new approach for controlling continuous magnitudes in numerical comparison tasks. *Behavior Research Methods*, 49, 1120–1127.
- Simanek, D. E. (1996). *The principles of artistic illusions—Adding depth to illusions*. Retrieved from <https://www.lockhaven.edu/~dsimanek/3d/illus1.htm>
- Smets, K., Sasanguie, D., Szűcs, D., & Reynvoet, B. (2015). The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison. *Journal of Cognitive Psychology*, 27, 310–325. <http://dx.doi.org/10.1080/20445911.2014.996568>
- Starr, A., DeWind, N. K., & Brannon, E. M. (2017). The contributions of numerical acuity and non-numerical stimulus features to the development of the number sense and symbolic math achievement. *Cognition*, 168, 222–233. <http://dx.doi.org/10.1016/j.cognition.2017.07.004>
- Stoianov, I., & Zorzi, M. (2012). Emergence of a ‘visual number sense’ in hierarchical generative models. *Nature Neuroscience*, 15, 194–196. <http://dx.doi.org/10.1038/nn.2996>
- Sullivan, J., & Barner, D. (2014). Inference and association in children’s early numerical estimation. *Child Development*, 85, 1740–1755.
- Szucs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Frontiers in Psychology*, 4, 444. <http://dx.doi.org/10.3389/fpsyg.2013.00444>
- Vallortigara, G. (2017). An animal’s sense of number. In J. Adams, P. Barmby, & A. Mesoudi (Eds.), *The Nature and Development of Mathematics: Cross Disciplinary Perspectives on Cognition, Learning and Culture* (pp. 43–60). Florence, KY: Routledge.
- Van Opstal, F., & Verguts, T. (2013). Is there a generalized magnitude system in the brain? Behavioral, neuroimaging, and computational evidence. *Frontiers in Psychology*, 4, 435. <http://dx.doi.org/10.3389/fpsyg.2013.00435>
- Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience*, 6, 624–631. <http://dx.doi.org/10.1038/nn1057>
- Young, L. N., & Cordes, S. (2013). Fewer Things, Lasting Longer. *Psychological Science*, 24, 1057–1059. <http://dx.doi.org/10.1177/0956797612465294>

Received April 24, 2018

Revision received November 16, 2018

Accepted November 27, 2018 ■