Acta Psychologica 230 (2022) 103737

Contents lists available at ScienceDirect

Acta Psychologica

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

The malleable impact of non-numeric features in visual number perception

Denitza Dramkin[a,*], Cory D. Bonn[a,b], Carolyn Baer[a,c], Darko Odic[a]

[a] *Department of Psychology, University of British Columbia, Canada*
[b] *Strong Analytics, United States of America*
[c] *Department of Psychology, University of California, Berkeley, United States of America*

ARTICLE INFO

Keywords:
Number perception
Feedback
Congruency effects
Response conflicts

ABSTRACT

Non-numeric stimulus features frequently influence observers' number judgments: when judging the number of items in a display, we will often (mis)perceive the set with a larger cumulative surface area as more numerous. These "congruency effects" are often used as evidence for how vision extracts numeric information and have been invoked in arguments surrounding whether non-numeric cues (e.g., cumulative area, density, etc.) are combined for number perception. We test whether congruency effects for one such cue – cumulative area – provide evidence that it is *necessarily used* and integrated in number perception, or if its influence on number is *malleable*. In Experiment 1, we replicate and extend prior work showing that the presence of feedback eliminates congruency effects between number and cumulative area, suggesting that the role of cumulative area in number perception is malleable rather than obligatory. In Experiment 2, we test whether this malleable influence is because of *use of prior experiences* about how number naturalistically correlates with cumulative area, or the result of *response competition,* with number and cumulative area actively competing for the same behavioral decision. We preserve cumulative area as a visual cue but eliminate response competition with number by replacing one side of the dot array with its corresponding Hindu-Arabic numeral. Independent of the presence or absence of feedback, we do not observe congruency effects in Experiment 2. These experiments suggest that cumulative area is not necessarily integrated in number perception *nor* a reflection of a rational use of naturalistic correlations, but rather congruency effects between cumulative area and number emerge as a consequence of response competition. Our findings help to elucidate the mechanism through which non-numeric cues and number interact, and provide an explanation for why congruency effects are only sometimes observed across studies.

1. Introduction

With seemingly little effort, we can rapidly extract numeric information from a visual scene. A quick glance will instantly, though imprecisely, tell us whether our shopping cart likely has too many items to go to the express checkout or which line likely has the fewest people in it. Despite the apparent ease through which we can do this, how vision extracts numeric information remains unclear. One major challenge to understanding how we perceive number is that it is not based on a one-to-one correspondence between visual features and perceptual representations. Non-numeric cues, such as the area and density of objects, are often closely related to numeric information, making it difficult for number to be reliably inferred from any single perceptual feature (Anobile et al., 2016; DeWind et al., 2015; DeWind & Brannon, 2012; Gebuis, Reynvoet, 2012a, 2012b). Number perception is thus a challenge of inference: determining which features are relevant – and which can be ignored – for arriving at the correct percept of the number of objects in the external world. Yet, how vision ultimately accomplishes this has been a source of considerable debate.

Under some models, vision is thought to fundamentally rely on extracting and combining a bundle of non-numeric features (e.g., cumulative area, size, density, convex hull, luminance, etc.) in order to infer number. That is, rather than encoding number directly (e.g., through neural mechanisms that are domain-specific to number; Anobile et al., 2016; Burr & Ross, 2008; Ross & Burr, 2010), visual perception depends on some mixture of concurrent visual features to infer number indirectly, a process which may also be refined through development (Dakin et al., 2011; Gebuis et al., 2016; Gebuis & Reynvoet, 2012b, 2012c; Leibovich et al., 2017; Szucs et al., 2013). At the same time, a major question surrounding the role of non-numeric features in visual

* Corresponding author at: Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC V6T 1Z4, Canada.
E-mail address: ddramkin@psych.ubc.ca (D. Dramkin).

https://doi.org/10.1016/j.actpsy.2022.103737
Received 7 February 2022; Received in revised form 20 July 2022; Accepted 3 September 2022
Available online 9 September 2022
0001-6918/

number perception is whether theses cues used to infer number are *mandatory* and stable, or whether they are *malleable* and sensitive to practice and experience.

Here, we explore this question through testing the malleability of one specific visual feature: cumulative area. We focus on this cue over other relevant features for two key reasons. First, cumulative area has been especially well-documented as an influence on number perception via congruency effects (Aulet & Lourenco, 2021; Gebuis, Reynvoet, 2012b; Gilmore et al., 2016; Hurewitz et al., 2006). For example, when deciding which side of an array is more numerous, observers are often faster and more accurate on area "congruent" trials, where the side with the greater number of objects is also greater in cumulative area or other non-numeric cues (e.g., has the larger convex hull, higher luminance, etc.; Aulet & Lourenco, 2021; Dakin et al., 2011; Gebuis, Reynvoet, 2012b; Gebuis & van der Smagt, 2011; Hurewitz et al., 2006). On area "incongruent" trials, observers tend to choose the set that appears physically larger as having numerically more items, even if populated by fewer dots (e.g., Dakin et al., 2011; Gebuis, Reynvoet, 2012b). Findings such as these have led to arguments that cumulative area is integral to number perception and that changes in area dimensions necessarily and obligatorily influence number (Aulet & Lourenco, 2021). Consistent with this, popular neural network models deterministically code for number by weighing cumulative area with cumulative extent of edges (e.g., Stoianov & Zorzi, 2012). Second, recent theories on how numeric information might spontaneously emerge from viewing natural scenes also place cumulative area as a central feature in this learning process, as visual number *strongly* correlates with cumulative area in natural scenes (e.g., Nasr et al., 2019, Testolin, Zou, et al., 2020b). Hence, exploring whether cumulative area, itself, is a *mandatory* or *malleable* cue is important for understanding how vision ultimately arrives at number. However, our goal is not to suggest that other features do not play any role in this process. Instead, we aim to provide an experimental strategy for how one might go about teasing apart the precise role of any one visual feature on number perception.

One approach taken to disentangle the influence of mandatory versus malleable effects is through providing trial-by-trial feedback. In previous work, DeWind and Brannon (2012) found that providing trial-by-trial feedback in a speeded number discrimination task decreased observers' reliance on non-numeric cues (e.g., cumulative area), with accuracy improving on trials where the side with numerically greater dots was populated by a smaller cumulative area. In this way, feedback may have served to eliminate congruency effects through signaling to observers that the typical – albeit natural correlation – between number and cumulative area does not hold, allowing observers to readily abandon their use of this cue. At the same time, however, the authors note the lack of a control (i.e., no feedback) condition in their study, make it unclear whether feedback – itself – was used as the cue to what visual features were relevant to number, or whether practice and experience with the task was used to key in observers to what visual information was relevant (DeWind & Brannon, 2012).

In Experiment 1, we replicate and extend the work of DeWind and Brannon (2012) by directly comparing the presence vs. the absence of feedback on speeded number discrimination. If cumulative area is obligatory for number perception, then we'd expect that observers will demonstrate congruency effects regardless of whether or not they received feedback on the task. That is, we should find that observers are less accurate on Incongruent trials (i.e., where the array with numerically more dots appears visually smaller) both in the No Feedback and Feedback conditions. However, if congruency effects are reduced or eliminated with feedback, this would suggest that cumulative area is not a mandatory part of visual number perception (e.g., see Aulet & Lourenco, 2021).

But if feedback does reduce congruency effects between number and cumulative area (e.g., DeWind & Brannon, 2012), this could occur for multiple theoretical reasons. First, feedback could signal to observers that their normal *priors* about the co-occurrence of number and cumulative area are not appropriate here. For example, in typical visual experiences, cumulative area is strongly and frequently correlated with number (Aulet & Lourenco, 2021; Hurewitz et al., 2006; Nasr et al., 2019; Testolin, Dolfi, et al., 2020a; Testolin, Zou, et al., 2020b; Zorzi & Testolin, 2018). As a result, it would be rational for observers to form expectations that number and cumulative area should be correlated (i.e., the side that is more numerous also appears visually larger). However, when provided feedback that contradicts this expectation (e.g., the more numerous side is the side that actually appears smaller), this could allow observers to abandon their reliance on prior experiences and minimize the use of cumulative area as a cue to numerosity. Alternatively, feedback could serve to help reduce *response competition* between number and cumulative area. In response to feedback, observers may more readily monitor their responses, allowing them to avoid responding according to the wrong cue (i.e., cumulative area), thus minimizing the effect of cumulative area on their behavioral responses (e.g., Barth, 2008; Picon et al., 2019).

To disentangle these possibilities about what may contribute to the malleability of cumulative area, we take inspiration from the work of Picon et al. (2019), who demonstrate that response competition can be eliminated by asking participants to *estimate* the number of dots rather than *discriminate* between which side of the display has 'more' dots. In typical speeded number discrimination, the labels of 'more' and 'less' can apply not only to the number of dots but also other non-numeric dimensions (e.g., the side with larger/smaller cumulative area). In cases where number and non-numeric cues do not coincide (e.g., the side with numerically more dots has a smaller surface area) this creates competition for the same response code (see also Barth, 2008; Odic et al., 2013; Pietroski et al., 2009). But, in number estimation, categorical numerical responses are not inherently possible for non-numeric dimensions. In line with this, Picon et al. (2019) showed that estimation eliminates some congruency effects, namely those between contour length and number, suggesting that at least some of these effects result from response conflicts. While their work examined several non-numeric features that are thought to be important for number perception (e.g., contour length, convex hull/density), cumulative area was not explored. This leaves open the possibility that any effects observed in Experiment 1 could be due to either a flexible prior or the result of response competition.

In Experiment 2, we preserve cumulative area as a feature for number discrimination but replace one side of the dot array with a Hindu-Arabic numeral. This eliminates any competition between non-numeric cues during the comparison of the two dot arrays, since the side represented by the numeral has no cumulative area, while still preserving the effect of congruency on the side with dots, as a high-numbered set is sometimes represented using a high cumulative area (consistent with the naturalistic prior) and sometimes using a low cumulative area (inconsistent with the naturalistic prior). Thus, if congruency effects between number and cumulative area result from response conflicts, eliminating the competition between these features should eliminate the congruency effect, independent of the presence or absence of feedback. If, instead, congruency effects arising from the influence of cumulative area remain, then we would expect the pattern of responses to be similar to Experiment 1. The latter would then suggest that cumulative area is a flexible cue to number, which stems from reliance on naturalistic priors.

2. Experiment 1

2.1. Preregistration

This experiment was fully pre-registered, including in sample size and primary analyses. The link to the pre-registration, as well as the data, stimuli, and analysis scripts can be found on the Open Science Framework (OSF; the full preregistration can be found here: https://osf.io/4fja3/?view_only=ac70869c68784a038a79105b7d4b8787;

materials can be found here: https://osf.io/jfmny/?view_only=e3c30e7858ea40649a4e7a3aa8595598).

2.2. Participants

Past research yields a wide range of effect sizes for congruency effects between number and cumulative area in number discrimination tasks with and without feedback (DeWind & Brannon, 2012; Halberda & Feigenson, 2008; Inglis & Gilmore, 2014; Tokita & Ishiguchi, 2010). Hence, we conservatively assumed that the true size of the effect was medium ($d = 0.25$). Based on an a priori power analysis, we determined that we would need 52 participants (26 per condition: No Feedback, Feedback) to reach 90 % power in order to detect a significant effect ($\alpha = 0.01$) of the interaction between feedback and area congruency on Accuracy. Any participants with a self-reported learning or developmental disorder, who failed to complete all trials (i.e., <100 %), or had chance performance (determined through a binomial test) were excluded and replaced until the target sample size was met.

Fifty-two university students participated for course credit (M = 21.4). An additional three participants were excluded and replaced: 2 experienced technical issues resulting in their inability to complete the study, and 1 did not significantly perform above chance. All experimental methods and procedures were reviewed and approved by the [EXCLUDED FOR BLINDED REVIEW] of Research Ethics.

2.3. Methods and procedures

Participants were tested in-person in a quiet room on a 22.5″ iMac running custom-made Psychtoolbox-3 scripts in MATLAB (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). They were shown 256 randomly ordered displays of blue and yellow dots, with yellow dots appearing on the left side of the screen and blue dots on the right (Fig. 1). They were instructed to judge which set was more numerous by pressing the "F" key if they thought there were more yellow dots and the "J" key if they thought there were more blue dots. Stimuli were counterbalanced such that both the yellow and blue dots were more numerous for half the total trials. Performance was measured through Accuracy, corresponding to the percentage of trials in which participants correctly identified the more numerous side of the screen.

Stimuli remained visible on the screen for a maximum of 1200 ms or until participants responded, whichever came first. To vary the degree of difficulty of the task, we manipulated the total number and ratio of yellow to blue dots across trials. The arrays varied between 10 and 50 dots. These quantities were chosen given previous work which has suggested that small quantities (e.g., those <4) are perceived by separate mechanisms compared to larger quantities (e.g., those above 5; Barth et al., 2003; Feigenson et al., 2004; Mandler & Shebo, 1982; Trick & Pylyshyn, 1994; Whalen et al., 1999; Xu, 2003; Zimmermann, 2018). For example, small numerosities are perceived exactly and show distinct psychophysical signatures including a lack of ratio dependence and being less susceptible to adaptation effects (Anobile et al., 2014; Feigenson et al., 2004; Mandler & Shebo, 1982; Zimmermann & Fink, 2016). In contrast, perception of large numerosities are frequently shown to be more imprecise and, critically, susceptible to congruency effects (e.g., Feigenson et al., 2004; Odic & Starr, 2018). At the same time, some previous work suggests that with age, perception of large quantities that are more frequently encountered (e.g., those >5 but below 25) can become increasingly precise (albeit not exact; Gebuis & Reynvoet, 2011; Izard & Dehaene, 2008; but see Cheyette & Piantadosi, 2020). To account for this, we include two "Set Sizes" for these larger numerosities.

For the "Small" set, arrays varied near-uniformly from 10 to 20 dots. For the "Large" set, arrays varied near-uniformly from 25 to 50 dots. Half of all trials consisted of Small set sizes, and the other half were Large sets. However, we made no specific a priori predictions about the potential effects of set size on performance. For Small and Large sets, we presented four identical ratios: 2.0 (e.g., 20 yellow vs. 10 blue dots; 50 yellow vs. 25 blue dots), 1.5, 1.25, and 1.07 (e.g., 16 yellow vs. 15 blue dots; 32 yellow vs. 30 blue dots). With this design, accuracy should increase as the ratio increases, so we should expect to see greatest accuracy on the highest ratio (2.0) and the worst accuracy on the lowest ratio (1.07).

To examine the impact of cumulative area on number discrimination, we varied whether cumulative area correlated or conflicted with the number of dots in the arrays. This method is consistent with several other studies (e.g., Halberda et al., 2008, 2012; Halberda & Feigenson, 2008; Libertus et al., 2011, 2012; Mazzocco et al., 2011; Odic, 2018; Odic et al., 2013).[1] For half the trials, the more numerous side of the array was also larger in cumulative area (i.e., Congruent trials), and on the other half of the trials, the more numerous array was smaller in cumulative area (i.e., Incongruent trials; Fig. 1). If the mismatch between cumulative area and number impacts discrimination, then we should expect lower Accuracy on Incongruent trials compared to Congruent trials. Alternatively, if participants ignore cumulative area as a cue to number, then we would expect no congruency effects, with Accuracy being equal on Congruent and Incongruent trials.

Recent work on cumulative area perception has suggested that observers might not perceive area veridically, but systematically underestimate it by performing an "additive" heuristic (e.g., adding the two lengths of a square rather than multiplying them; Yousif & Keil, 2020). In our stimuli, "additive cumulative area" is held constant on Incongruent trials (e.g., the side that twice as many dots is only half as large), and it covaries with side with numerically more dots on Congruent trials. Therefore, even if participants perceive area "additively", we would still expect to find a congruency effect, albeit an attenuated one.

To test the malleability of congruency effects between cumulative area and number, participants (N = 52) were randomly assigned to either a No Feedback ($n = 26$) or Feedback ($n = 26$) condition. In the No Feedback condition, observers received no indication of their performance after each trial. In contrast, for the Feedback condition, the phrases "Correct" or "Incorrect" was visually displayed at the center of the screen following each trial. We predicted that if participants' use of cumulative area during number discrimination is malleable to the influence of feedback, then we would observe congruency effects only in the No Feedback condition. That is, if feedback eliminates congruency effects, we expected participants in the Feedback condition to perform equally well on Congruent vs. Incongruent trials, and only participants in the No Feedback condition to show worse performance on Incongruent compared to Congruent trials. Importantly, observing an impact of feedback on Incongruent trials would suggest that cumulative area is not mandatory for visual number perception, and that instead it may only be flexibly used as a cue for number (e.g., due to naturalistic priors or later decision-making processes).

[1] Manipulating cumulative area necessarily may manipulate other non-numeric features (e.g., density/spread, convex hull). In contrast to some other stimuli (e.g., Burr & Ross, 2008), here, cumulative area is not linearly related to the number of dots, so instead we calculate density as the convex hull divided by the cumulative area. For posterity, we report the correlations for both the yellow and blue dots separately. In our stimuli, convex hull and cumulative area were not significantly correlated: Yellow $r(62) = 0.03$, $p = .815$, Blue $r(62) = 0.17$, $p = .179$. Though – as would be expected based on how density is calculated here – convex hull and density are significantly related to each other, Yellow $r(62) = 0.48$, $p < .001$, Blue $r(62) = 0.40$, $p < .001$, as are cumulative area and density, Yellow $r(62) = -0.84$, $p < .001$, Blue $r(62) = -0.80$, $p < .001$. We do not make claims about whether or not participants could use other cues to numerosity for number discrimination, as even density and convex hull are not the only other non-numeric features that could possibly influence number. Though critically, our stimuli are such that cumulative area is only correlated with the number of dots on half the trials. We provide calculations for cumulative area, convex hull, and density based on pixel counts are provided for each stimulus on OSF as part of our study materials.

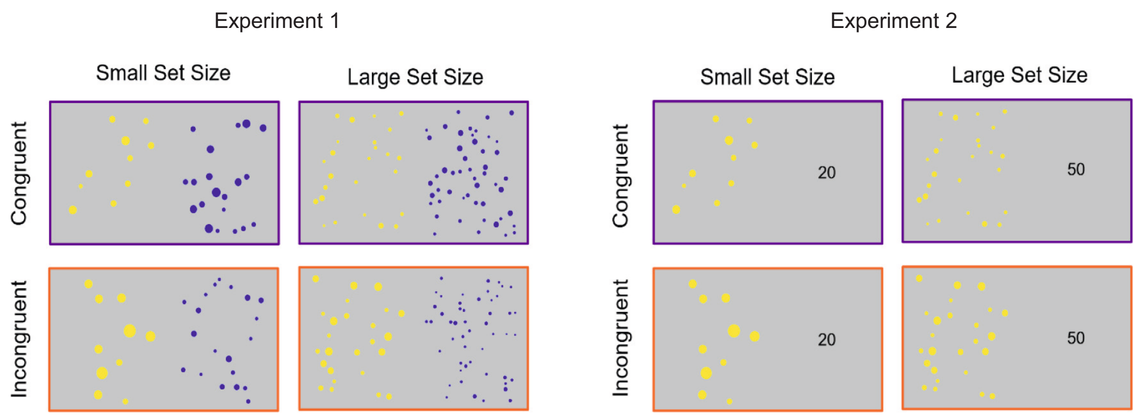

Fig. 1. Stimuli examples from Experiment 1 and 2
Note. For Experiment 1, the example shows 10 yellow vs. 20 blue dots for the Small Set Size and 25 yellow vs. 50 blue dots from the Large Set Size. In Experiment 2, the blue dots are replaced with their corresponding Arabic digit. For Congruent trials, the set with numerically more dots also had the larger cumulative area. On Incongruent trials, the set with numerically more dots had the smaller cumulative area. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.4. Results

Following our pre-registration, we analyzed our data using both traditional frequentist approaches, and a more nuanced and powerful Bayesian analysis. The results are identical under both approaches. However, we provide both analyses to give readers an alternative to frequentist statistics without a major loss of interpretability or readability.

We first examine the influence of feedback, area congruency (i.e., trial type), set size, and ratio on accuracy using classical frequentist approaches. A 2 (Condition: No Feedback, Feedback) × 2 (Area Congruency: Congruent, Incongruent) × 2 (Set Size: Small, Large) × 4 (Ratio: 2.0, 1.50, 1.25, 1.07) Greenhouse-Geisser corrected repeated measures ANOVA over Accuracy revealed a main effect of Condition, $F(1, 50) = 6.19$, $p = .016$, $\eta^2_G = 0.03$, a main effect of Area Congruency, $F(1, 50) = 11.27$, $p = .002$, $\eta^2_G = 0.02$, and a main effect of Ratio, $F(2.35, 117.34) = 1168.15$, $p < .001$, $\eta^2_G = 0.77$, but no main effect of Set Size, $F(1, 50) = 0.28$, $p = .598$, $\eta^2_G < 0.01$. There was also a significant Condition × Area Congruency interaction, $F(1, 50) = 6.52$, $p = .014$, $\eta^2_G = 0.01$, but no other significant interactions between Condition, Area Congruency, Ratio, and/or Set Size (all p's > 0.213).

To further probe these effects and interactions, we performed planned post-hoc Tukey-corrected pairwise comparisons. As can be seen in Table 1, participants in the No Feedback Condition were less accurate than those in the Feedback condition (averaging across Area Congruency, Ratios, and Set Sizes), $t(50) = -2.49$, $p = .016$. Additionally, those in the No Feedback condition were significantly worse on Incongruent trials compared to Congruent trials, $t(50) = 4.18$, $p < .001$ (Fig. 2). In contrast, participants in the Feedback condition performed equally well on the Congruent and Incongruent trials, $t(50) = 0.57$, $p = .572$. These results demonstrate that – at least for cumulative area – congruency effects are malleable and sensitive to feedback. We also observed differences in performance across ratios, with participants performing better on easier (i.e., larger) ratios in both No Feedback and Feedback conditions. However, overall, participants given feedback demonstrated greater accuracy, with accuracy, itself, being broadly consistent with prior work using similar stimuli (e.g., Odic et al., 2013; Odic, 2018; Table 1).

Table 1
Accuracy across experiments, area congruency, and conditions.

	Area congruency	No feedback *M* (SE)	Feedback *M* (SE)
Experiment 1	Congruent	83.8 (0.01)	85.1 (0.01)
	Incongruent	80.0 (0.01)	84.6 (0.01)
Experiment 2	Congruent	74.3 (0.01)	80.3 (0.01)
	Incongruent	73.8 (0.01)	79.3 (0.01)

Note. Table depicts accuracy as the mean proportion of trials correct. Standard errors are included in parentheses next to average accuracy.

2.4.1. Bayesian analyses

Our main analyses demonstrate effects of feedback and area congruency on observers' number discrimination accuracy. Accuracy measurements are themselves a combination of potential changes in response bias (e.g., preference for congruent trials) and stimulus discriminability (e.g., Weber fraction). To further decompose these effects into changes in response bias, stimulus discriminability, or both, we fit Bayesian, multilevel regression models to the raw response data using RStan (Stan Development Team, 2020). These models are straightforward extensions of multilevel (or "mixed-effects") probit regression.

A multilevel approach allows for partial pooling of information across participants to optimally estimate the effect sizes of within-subject experimental manipulations. Therefore, it is particularly appropriate for estimating group-level psychometric functions when each individual subject has a limited number of trials (on the order of hundreds) compared to more traditional psychophysics experiments, which sometimes require thousands of trials per subject to estimate.

A Bayesian approach provides a convenient form of regularization to ensure stable estimates of lapse/random guess rates, which cannot easily be obtained with commonly used packages for maximum-likelihood fitting of multilevel models. In addition, maximum-likelihood approaches to lapse-rate estimation often need to be fitted with hard constraints to avoid local minima (e.g., must be <5 % of trials); the Bayesian approach allows for softer constraints with informative priors. We evaluated experimental effects directly with parameter estimates; effects are deemed credible when 95 % of their posterior density excludes 0.

We used generalized linear models with probit transforms (i.e., normal CDF) to model the psychometric function, augmented with an additional lapse rate parameter to allow the asymptotes to be >0 and <1. We discuss the model specifications below.

2.4.1.1. Model specifications. Responses (coded as 0 and 1) were modeled with a Bernoulli likelihood, with probability of responding 1 represented by p (see Eq. (2.1)). This probability is a function of two latent quantities modeled simultaneously: (1) an underlying probability of responding to one of the two stimuli as a function of the experimental manipulations and individual participants' biases and sensitivities, represented by a probit-transformed linear prediction $\Phi(\mu)$, and (2) the

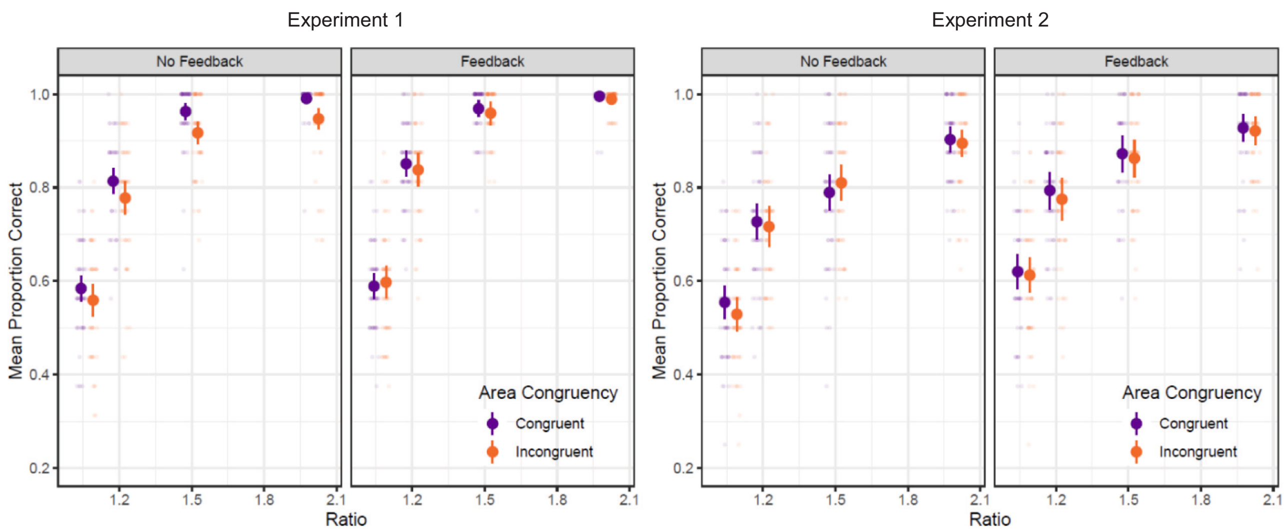

Fig. 2. Results from Experiment 1 and 2
Note. The figure depicts mean accuracy measured as the mean proportion correct (collapsed across Set Size) as a function of stimulus Ratio (x-axis: 1.07, 1.2, 1.5, 2.0; e.g., 10 yellow dots vs. 20 blue dots; Arabic digit 10 vs. 20 yellow dots), Condition (panels: No Feedback, Feedback), and Area Congruency (color). The scatter points correspond to group-level accuracy, with confidence intervals around the mean shown, with fainter points used to indicate individual participants' performance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

latent lapse rate Λ specified as in Odic (2018). This function is represented in Eq. (2.2). We discuss each of these parts of the likelihood in turn, followed by notes about specifying priors for the free parameters, including participant-specific ("random") effects.

$$response_i \sim \text{Bernoulli}(p_i) \tag{2.1}$$

$$p_i = (1 - \Lambda)(\Phi(\mu_i)) + \frac{\Lambda}{2} \tag{2.2}$$

The prediction μ is a linear function of the following coefficients (represented with the usual symbol β) and the model/design matrix (see Eq. (2.3)): (1) an intercept β_0 to account for baseline response biases, (2) a coefficient β_1 to quantify the effect of the log of the ratio of the dot arrays presented as pairs (fixed for each stimulus pair across participants), (3) a coefficient β_2 representing the effect of congruency (where incongruent trials were scored as −0.5 and congruent trials were scored as 0.5), (4) a coefficient β_3 representing the effect of feedback (no feedback was scored as −0.5, feedback was scored as 0.5), and the full set of two- and three-way interactions. Each coefficient corresponding to a within-subject manipulation was paired with a subject-level random effect, indicated with the symbol τ.

$$\begin{aligned} \mu_i = {} & \beta_0 + \tau_{0j} + \\ & (\beta_1 + \tau_{1j})(ln(\text{Ratio}_i)) + \\ & (\beta_2)(\text{Feedback}_i) + \\ & (\beta_3 + \tau_{2j})(\text{Congruency}_i) + \\ & (\beta_4)(ln(\text{Ratio}_i) \times \text{Feedback}_i) + \\ & (\beta_5 + \tau_{3j})(ln(\text{Ratio}_i) \times \text{Congruency}_i) + \\ & (\beta_6)(\text{Feedback}_i \times \text{Congruency}_i) + \\ & (\beta_7)(ln(\text{Ratio}_i) \times \text{Feedback}_i \times \text{Congruency}_i) \end{aligned} \tag{2.3}$$

Because the latent lapse rate is bounded between 0 and 1, for convenience we modeled it with a logistic regression equation, with terms corresponding to the group mean lapse rate λ and subject-specific lapse rates τ_{4j}. See Eq. (2.4).

$$\Lambda = \frac{1}{1 + exp(-(\lambda + \tau_{4j}))} \tag{2.4}$$

Since it was not possible to use fully informative priors based on past designs, for coefficients in the linear model we used weakly informative priors (see Eqs. (2.5) and (2.6)). Coefficients in the linear portion of the model were given Gaussian priors centered at 0 with a standard deviation of 2. The lapse rate λ was given a normal prior with mean below zero such that 50 % of its inverse-logit-transformed probability density would be below a rate of 0.07 and 95 % would be below a rate of 0.25.

Following Sorensen and Vasishth (2015), we specified the correlation matrix for within-subject effects by placing an LKJ prior over its Cholesky decomposition. The vector $\boldsymbol{\sigma}$ includes the estimated standard deviations of each of the participant-specific coefficients corresponding to each β. The Cholesky-decomposed matrix is given by $\boldsymbol{L}$. These are multiplied by a matrix of normally distributed random variables $\boldsymbol{z}$ to create the matrix of correlated participant-level coefficients $\boldsymbol{\tau}$.

$$\beta_k \sim \text{Normal}(0, 2) \tag{2.5}$$

$$\lambda \sim \text{Normal}(-1.5, 0.5) \tag{2.6}$$

$$\boldsymbol{\sigma} \sim \text{Cauchy}(0, 1) \tag{2.7}$$

$$\mathbf{L} \sim \text{LKJ}(2) \tag{2.8}$$

$$\mathbf{z} \sim \text{Normal}(0, 1) \tag{2.9}$$

$$\boldsymbol{\tau} = (diag(\boldsymbol{\sigma})(\mathbf{L})(\mathbf{z}))^{\mathrm{T}} \tag{2.10}$$

2.4.1.2. Bayesian results. The marginal means as a function of stimulus ratio, trial feedback, and area congruency are shown in Fig. 3. We also include a graphical table of the most important parameter values, generated using the ggdist package (Kay, 2021) in Fig. 4.

As expected, participants displayed a very high sensitivity to the (log of the) ratio between the two numbers, replicating the main effect of Ratio in the frequentist analysis. In addition, both feedback and stimulus congruency credibly moderated the effect of ratio (Ratio × Feedback and Ratio × Area Congruency interaction terms), with the slope of the psychometric function being steeper on average with feedback and when cumulative surface area was congruent with number.

A credible Ratio × Feedback × Area Congruency interaction revealed the effect of most interest: the presence of feedback eliminated the Ratio × Area Congruency interaction (the effect of area congruency on participants' sensitivity to ratio). Ratio × Area Congruency interaction was

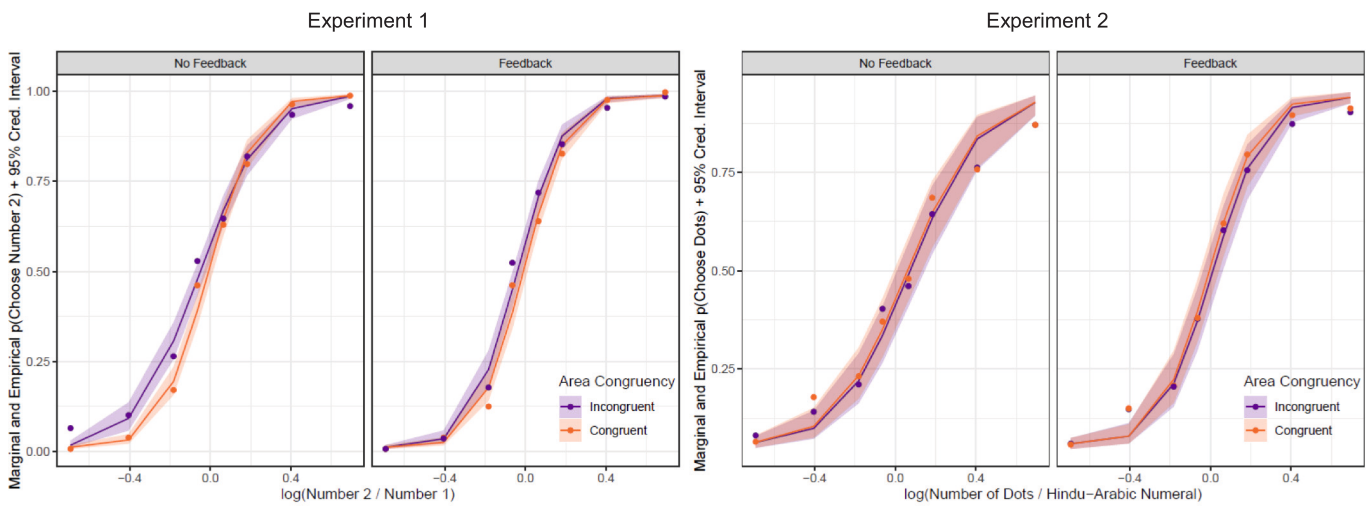

Fig. 3. Selection across experiments.
Note. The figure depicts the marginal means (lines), 95 % credible intervals (ribbons), and empirical proportions (scatterplot points) as a function of stimulus Ratio (x-axis), Condition (panels), and Area Congruency (color). For Experiment 1, dot arrays assigned numbers 1 and 2 were fixed across participants; 'congruent' means that the dot array with larger cumulative area was number 2, while 'incongruent' means that the dot array with the larger cumulative surface area was number 1. For Experiment 2, one side of the dot array was replaced with a Hindu-Arabic numeral, and was fixed across participants; 'congruent' means that the dot array was more numerous and consisted of a larger cumulative area; 'incongruent' means that the dot array was less numerous but consisted of a visually large cumulative area (or vice versa).

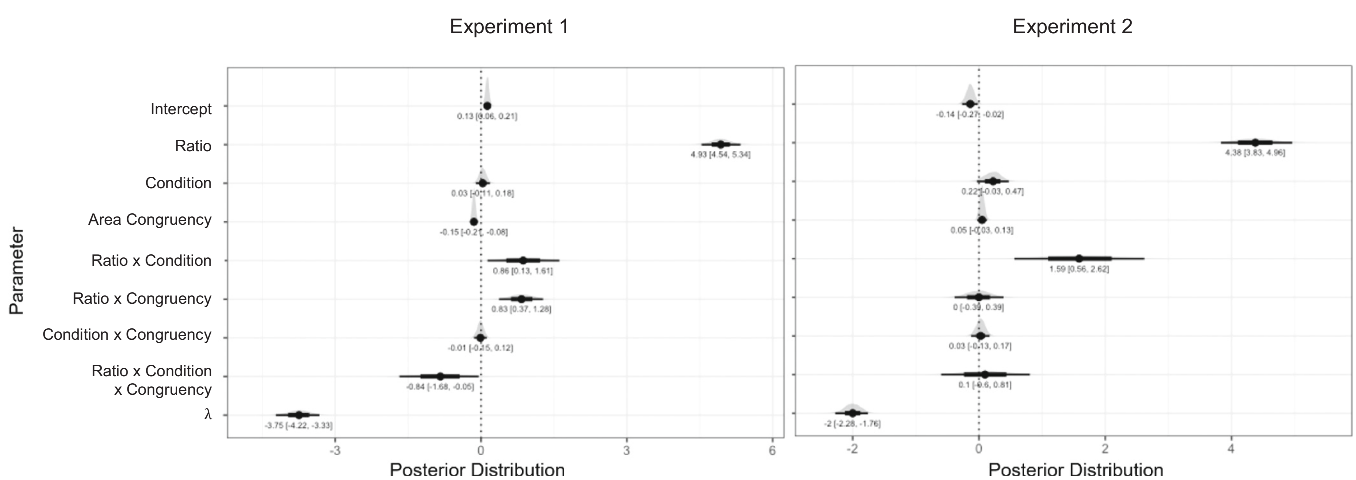

Fig. 4. Parameters and corresponding posterior distributions plotted with density estimates
Note. Parameters are show on the x-axis, with the posterior distribution for these parameters shown on the y-axis. Below each posterior, its corresponding mean and 95 % credible interval are printed.

not credibly different from 0 in the presence of feedback (simple slope mean = 0.174; 95 % credible interval = [−0.491, 0.851]). It was only credibly different from 0 without feedback (simple slope mean = 1.22; 95 % credible interval = [0.639, 1.77]).

2.5. Discussion

In Experiment 1, participants showed strong congruency effects in the absence of feedback: when number and cumulative area conflicted (e.g., the side with the larger cumulative area was populated by numerically fewer dots), participants were less accurate. However, this effect was eliminated when participants were given feedback. This suggests that congruency effects between cumulative area and number are malleable to the influence of feedback and that number may not strictly depend on cumulative area as a non-numeric visual feature. In other words, these findings challenge the notion that cumulative area is obligatory for number perception. We return to and expand upon this point in the General Discussion. Furthermore, while these findings are consistent with those reported by DeWind and Brannon (2012), our data clearly point to the role of feedback rather than mere training/practice for reducing congruency effects through the inclusion of a no feedback condition.

But if the influence of cumulative area on number is malleable, where does it exert its influence? For example, we find evidence that cumulative area (as well as dot size) influences numeric judgments. In particular, the Bayesian analyses highlight a bias observers had for overestimating the smaller dots and underestimating the larger dots, which was absent on congruent trials. However, what role feedback played in how participants used cumulative area is unclear. One possibility is that feedback served to eliminate biases in incongruent trials because it allowed participants to update their priors about the way in which visual features are correlated with the displays. Given that cumulative area is frequently correlated with number in natural scenes (e.g., Nasr et al., 2019; Testolin, Zou, et al., 2020b), rational observers may normally rely on these naturalistic correlations between number and cumulative area, except when they have clear reasons to believe that the stimuli they are viewing are not consistent with the distributions of features they learned through prolonged experience (Gebuis et al., 2016; Gebuis, Reynvoet, 2012b; Leibovich et al., 2017). In this case, receiving feedback that their use of cumulative area is resulting in low accuracy may have allowed participants to update their expectations. An alternative is that feedback eliminated *response competition* between cumulative area and number for the same behavioral response (e.g., 'more'/ 'less'; Barth, 2008; Picon et al., 2019). However, much like how the classic Stroop effect (Stroop, 1935) can be eliminated when participants can point to the color rather than reading it (e.g., Durgin, 2000), feedback could have served as a way to reduce response competition on incongruent trials, allowing participants to avoid biases towards cumulative area.

There is also a way in which the results of Experiment 1 could be seen as consistent with a mandatory integration view of cumulative area in number perception (e.g., Aulet & Lourenco, 2021). Perhaps participants who were given feedback recognized that on some trials their (mis) perception of the dots misaligned with the feedback and chose to willingly invert their responses on those trials, despite still using cumulative area to perceive number. By analogy, consider how participants might eventually give correct answers for visual illusions without actually changing their perception of them.

Therefore, a stronger test of whether cumulative area is an obligatory cue for number perception would be a demonstration that, even without feedback, participants can spontaneously (i.e., flexibly) drop their usage of cumulative area information, resulting in no congruency effects. In Experiment 2, we probe this by replacing one side of the dot array shown in Experiment 1, with its corresponding Hindu-Arabic numeral (Fig. 1). This allows us to preserve cumulative area as a visual cue to the numerosity of the dot array (on one side of the screen), but it eliminates the response competition with cumulative area (on the other side), since Hindu-Arabic numerals do not take up physical space in the way that dots do. As a result, if congruency effects between cumulative area and number result from response competition, then eliminating the conflict should result in no difference in accuracy between participants on the No Feedback and Feedback conditions across trial types. That is, we should observe no feedback by congruency interaction.

However, if we still observe congruency effects even after eliminating the competition between cumulative area and number, then this would provide evidence that is consistent with cumulative area being a flexible cue to number perception, but with congruence effects emerging due to a rational use of priors. For the side of the array that is not replaced by a Hindu-Arabic numeral, we would expect that the cumulative area of the dots will continue to influence number discrimination. For example, if presented with a display of 20 dots, observers should (mis)perceive it as being >20 on incongruent trials where the cumulative area is low (Fig. 1). If this is the case, then cumulative area should still bias participants' judgments, resulting in a Condition (No Feedback, Feedback) by Trial Type (Congruent, Incongruent) interaction in Experiment 2, as cumulative area would still be rationally recruited for number, independent of any response conflicts.

3. Experiment 2

The methods, stimuli, and corresponding analyses were identical to Experiment 1, with the only change being that one side of the dot array was replaced by its corresponding Arabic digit (Fig. 1). The link to the data and materials can be found on the Open Science Framework (OSF; https://osf.io/jfmny/?view_only=e3c30e7858ea40649a4e7a3aa8595598).

3.1. Participants

None of the participants who completed Experiment 1 were in Experiment 2. Consistent with the power analysis and pre-registration from Experiment 1, 52 university students (26 per condition: No Feedback, Feedback) completed a speeded number discrimination task for course credit ($M = 20.6$, 35 female).[2] We used the same exclusion and replacement criteria as in Experiment 1. Two participants were excluded and replaced due to a technical issue that resulted in their inability to complete all trials.

Participants were shown 256 randomly ordered displays with the dot array on one side of the screen and an Arabic digit on the other half. For half of the participants ($n = 26$) the Arabic digit appeared on the left side of the screen (corresponding to the number of yellow dots from Experiment 1), with the blue dots on the right. For the other half of participants ($n = 26$), the Arabic digit appeared on the right side of the screen (representing the corresponding to number of blue dots from Experiment 1), with yellow dots on the left. Participants were asked to judge which side of the screen was more numerous by pressing the "F" key for the left side (digit/yellow dots) and the "J" key for the right side (digit/ blue dots). The trials were counterbalanced such that the left side of the screen was the correct answer for half the trials.

The stimuli remained visible for 1200 ms or until participants responded, whichever came first. The Set Sizes were identical to Experiment 2. The Small set size varied near-uniformly from 10 to 20 dots/numerals; the Large set varied near-uniformly from 25 to 50 dots/ numerals. Half of all trials were Small set arrays, and the other half were Large set, and we made no prior predictions about the potential impact of mere set size on performance. We again presented Set Sizes across four identical Ratios: 2.0 (e.g., 20 yellow or blue dots vs. digit 10; 50 yellow or blue dots vs. digit 25), 1.5, 1.25, and 1.07 (e.g., 16 yellow or blue dots vs. digit 15; 32 yellow or blue dots vs. digit 30). As with Experiment 1, we expected that as the ratio increased, Accuracy (i.e., proportion of trials in which participants correctly selected the more numerous side) should increase. Half of the participants ($n = 26$) were also given feedback on their performance (Feedback condition: the phrase "Correct" or "Incorrect" was visually shown on the screen following each trial), and the other half ($n = 26$) were not (No Feedback condition).

3.2. Results

As with Experiment 1, the primary analysis consisted of a 2 (Condition: No Feedback, Feedback) × 2 (Area Congruency: Congruent, Incongruent) × 2 (Set Size: Small, Large) × 4 (Ratio: 2.0, 1.50, 1.25, 1.07) Greenhouse-Geisser corrected repeated measures ANOVA over Accuracy (i.e., proportion of total trials in which participants correctly selected the more numerous side of the display), allowing us to examine how feedback, congruency, set size, and ratio impacts performance. Consistent with Experiment 1, we find a main effect of Condition, $F(1, 50) = 10.63$, $p = .002$, $\eta^2_G = 0.06$, a main effect of Ratio, $F(2.73, 136.68) = 343.82$, $p < .001$, $\eta^2_G = 0.53$, and no main effect of Set Size, $F(1, 50) = 1.55$, $p = .218$ $\eta^2_G < 0.01$. We also find a Ratio × Set Size interaction, $F(2.68, 134.01) = 3.55$, $p = .020$, $\eta^2_G = 0.01$. However, in an important contrast to Experiment 1, we observe no main effect of Area Congruency, $F(1, 50) = 1.12$, $p = .295$, $\eta^2_G < 0.01$, nor a Condition × Area Congruency interaction, $F(1, 50) = 0.10$, $p = .756$, $\eta^2_G < 0.01$, and no other significant interactions between Condition, Area Congruency, Ratio, and/or Set Size (all p's > 0.100).

Feedback provided an *overall* boost in accuracy: participants were significantly more accurate in the Feedback condition compared to No Feedback, $t(50) = 3.26$, $p = .002$ (Fig. 2). We suspect that this effect occurred because participants normally under-estimate the number of dots on the screen, even independent of area (e.g., Izard & Dehaene, 2008), and that feedback effectively "calibrated" their responses (e.g.,

[2] One participant did not report their date of birth, so average age is calculated based on information from 51 participants.

DeWind & Brannon, 2012). Critically, when averaged across Ratio and Set Size, post-hoc Tukey-corrected pairwise comparisons also show no significant differences in how participants performed across Congruent and Incongruent trials within each condition (Table 1). Performance on Congruent vs. Incongruent trials was not significantly different within neither the Feedback, $t(50) = 0.95$, $p = .346$, nor No Feedback conditions, $t(50) = 0.54$, $p = .593$. Thus, unlike Experiment 1, we find no congruency effects in the No Feedback condition when one side the array is replaced by an Arabic digit. In other words, without a comparable cumulative area cue on both sides of the display, participants did not over- or under-estimate the number of dots independent of their cumulative area, suggesting that they did not use cumulative area even in the No Feedback condition.

3.2.1. *Bayesian analyses*

The same approach from Experiment 1 was adopted here. The marginal means as a function of stimulus ratio, condition (i.e., feedback, no feedback), and area congruency for Experiment 2 are shown in Fig. 3 and a graphical table of the most important parameter values are indicated in Fig. 4.

In contrast to Experiment 1, we find that there is no three-way interaction between Ratio, Condition, and Area Congruency ($M = 0.25$, 95 % credible interval = [−0.46, 0.97]) and no Condition × Area Congruency interaction ($M = 0.024$, 95 % credible interval = [−0.12, 0.20]), therefore we do not report the simple slopes (Fig. 4). However, we do find that like with Experiment 1, participants displayed a very high sensitivity to the Ratio between the number of dots and the Hindu-Arabic numerals. Yet only Condition credibly moderated the effect of Ratio, with the slope of the psychometric function being steeper on average with feedback ($M = 1.4$, 95 % credible interval = [0.29, 2.53]; Fig. 4). This indicated that participants were more sensitive to the numerical ratio with feedback.

4. General discussion

Several lines of work have suggested that visual number perception relies on a combination of non-numeric features, including cumulative area. But whether these cues can be used flexibly during number perception or are obligatory is unclear. Here, we test whether observers necessarily and *obligatorily* rely on cumulative area for inferring number (i.e., Experiment 1), or whether use of cumulative area in number perception simply reflects either a *naturalistic prior* about the typical correlation between cumulative area and number, or whether cumulative area act as a *response competition* with number (i.e., Experiment 2). We report two key findings.

First, we replicate and extend work demonstrating that congruency effects – at least between cumulative area and number – can be eliminated in number discrimination tasks with feedback (DeWind & Brannon, 2012). In Experiment 1, we show that participants were more accurate on trials in which number coincided with cumulative area (e.g., when the side with the larger cumulative area also had numerically more dots; Congruent trials) compared to trials in which number and cumulative area were in conflict (e.g., the side with a larger cumulative area contained numerically fewer dots; Incongruent Trials). Nevertheless, these effects were eliminated with trial-by-trial feedback about accuracy, resulting in no difference in accuracy on Congruent and Incongruent trials in the Feedback condition. This suggests that cumulative area is *malleable* in visual number perception, as participants' use of cumulative area as a cue to number could be flexibly adjusted when participants realized (via feedback) that their use of cumulative area is negatively impacting their performance (i.e., reducing accuracy). This finding is a significant challenge for computational or theoretical models that use evidence of congruency effects to place cumulative area as part of the bundle of cues early vision uses that represent number (Aulet & Lourenco, 2021; Dakin et al., 2011; Gebuis et al., 2016; Gebuis, Reynvoet, 2012b; Gebuis & Reynvoet, 2012c; Leibovich et al., 2017; Szucs et al., 2013).

Second, we found that when *response competition* between cumulative area and number for the relative-magnitude responses were eliminated, so were congruency effects. In Experiment 2, we replaced one side of the dot array with its corresponding Hindu-Arabic numeral, allowing us to preserve cumulative area as a visual cue to number on that side but eliminate the possibility of a conflicting cumulative area cue on the other. If congruency effects between cumulative area and number could arise from observers' priors expectations about how number and non-numeric visual features are naturally correlated (Gebuis & Reynvoet, 2012c; Leibovich et al., 2017), then eliminating response conflicts should have had no impact on the congruency effect in the no-feedback condition (e.g., the trials on which 20 dots were drawn with large dots should have been perceived differently than trials on which 20 dots were drawn with smaller ones). Yet, independent of the presence or absence of feedback, we found no congruency effects in Experiment 2, suggesting that these congruency effects likely stem from higher-level response competition between number and cumulative area. Hence, despite the strength of cumulative area as a cue to numerosity, its influence on number perception appears malleable, either through the presence of feedback or eliminating the response competition for the same decision. These results provide several novel insights for disentangling the contributions of non-numeric features on visual number perception.

Given findings that number discrimination is often influenced by non-numeric features (e.g., cumulative area, density, contour, convex hull, etc.), this has been used to argue that number must be inferred indirectly through other visual features (Dakin et al., 2011; Gebuis, Reynvoet, 2012a; Leibovich et al., 2017). That is, "number" is entirely or at least partially derived through combining some mixture of concurrent visual features rather than mechanisms that are domain-specific to number, itself (see Leibovich et al., 2017). While our findings do not eliminate this possibility entirely, these results do challenge the conclusions that can be drawn from evidence based on congruency effects alone.

First, the fact that participants can ignore strong cumulative area cues in the presence of feedback (Experiment 1) suggests that cumulative area is likely not a *necessary* component of the integrated features used to encode number, given that its influence could be readily ignored. Second, the finding that participants can spontaneously ignore cumulative area even without feedback (Experiment 2) – i.e., when participants have no reason to abandon it – puts a constraint even on models of number perception that suggest that observers can flexibility decide *which* features to use depending on local stimuli properties. Our stimuli had rich cumulative area information, and yet, even in the absence of feedback in Experiment 2, this information was not used by participants to arrive at their number estimates.[3]

An important direction for future work is to similarly examine the contributions of other non-numeric features and their role on visual number perception. Our focus here was on cumulative area, a cue that has repeatedly been shown to influence number perception, and is the strongest candidate for a non-numeric feature that strongly predicts number in natural scenes (Aulet & Lourenco, 2021; Hurewitz et al., 2006; Nasr et al., 2019; Testolin, Dolfi, et al., 2020a; Zorzi & Testolin, 2018). In line with this, we do find the size of the effect of cumulative area on number perception is consistent with prior work (e.g., DeWind & Brannon, 2012; Halberda & Feigenson, 2008; Inglis & Gilmore, 2014; Tokita & Ishiguchi, 2010). But, of course, other cues are also relevant for number, and it is not clear to what extent other cues can be used flexibly

[3] If congruency effects between number and cumulative area are the product of response conflicts, why are they sometimes observed in passive viewing tasks (Gebuis & Reynvoet, 2012c; Hurewitz et al., 2006; Tokita & Ishiguchi, 2010)? Consistent with other accounts of response conflicts, we believe that the most parsimonious explanation is that cumulative area and number compete for attention as well (Aulet & Lourenco, 2021; Yousif & Keil, 2020).

or are obligatory. For example, despite eliminating response conflicts between number and density and convex hull, Picon et al. (2019) still observed robust congruency effects during number estimation. Hence, although cumulative area may be a malleable cue, other non-numeric dimensions (e.g., density, convex hull) may indeed be necessary components of early number perception (e.g., Gebuis, Reynvoet, 2012a). At the same time, an alternative is that several non-numeric features simply share the same low-level encoding cues with number (Anobile et al., 2014; Dakin et al., 2011). For instance, density and number might both be computed through low-spatial frequency detectors (Dakin et al., 2011; Paul et al., 2022). If number also relies on these low-level features, any changes in low-spatial frequency will necessarily impact number perception (see Paul et al., 2022). In this way, density may influence number only to that extent it shares the same encoding features. Disentangling which features – and where – they exert their influence on number perception represents an important challenge for future work.

Furthermore, it is important to consider the extent to which certain non-numeric cues are correlated with each other and whether the influence of certain cues can be examined independently of others. Several lines of work have sought to demonstrate perceptual sensitivity specifically to number through controlling for non-numeric features (e.g., cumulative area, density, convex hull, contour, etc.; Clayton et al., 2015; DeWind et al., 2015; Inglis & Gilmore, 2014). Though, factors like average dot size and density are often correlated with cumulative area (Gebuis, Reynvoet, 2012b), including in our stimuli (see Footnote 1). Emerging work also suggests that attempts to control the objective presence of some of these features (e.g., via pixel count) may not always control for subjective perception (Clayton et al., 2015; Yousif & Keil, 2019, 2020). For example, in the case of area, rather than computing the true area of an array, Yousif and Keil (2019, 2020) suggest that participants rely on "additive" area (i.e., the sum, rather than the product, of a shape's dimensions). Accordingly, they find that perceived (additive) area is much more likely to impact number discrimination (Yousif & Keil, 2020). It is worth emphasizing that the construction of our stimuli leads to Incongruent trials being *equal* in additive area, still predicting a congruency effect if additive area is used throughout our tasks. Furthermore, participants' performance at levels well above chance on these trials shows that number perception is stable even when additive area is a reliable cue (i.e., Congruent trials) or when it is controlled for (i.e., Incongruent trials). Adapting models like that of DeWind et al. (2015) for stimuli generation – where different stimulus properties can be assessed and controlled over hundreds of trials – may provide a fruitful direction for future work that seeks to disentangle the individual contributions of different non-numeric features and test the malleability of other non-numeric features on number perception.

Ultimately, while our data suggest that cumulative area may be flexibly used during number perception, this does not eliminate the possibility that other cues are obligatory. Manipulating the presence vs. absence of feedback in number discrimination tasks like those in Experiment 1 and 2 may provide a useful methodological tool for testing which – and how – non-numeric features influence number perception, as knowing which dimensions represent response conflicts with number, and how to eliminate them (e.g., via providing feedback), is important for our abilities to evaluate visual acuity, specifically for number.

Our results also provide a possible explanation for why congruency effects between cumulative area and number are not always observed even across similar tasks and stimuli. Experiment 1 highlights that trial-by-trial feedback alleviates the competition between number and non-numeric features, independent of any improvements with task performance over time (DeWind & Brannon, 2012). Hence, contrary to proposals that the conflicting findings regarding congruency effects result from a failure to control for certain non-numeric cues (e.g., convex hull when controlling for cumulative area; Clayton et al., 2015), we find that the influence of at least some non-numeric features on number can be readily abandoned when the competition with number is eliminated via feedback or changes in task parameters (Picon et al., 2019). This can explain the differences between studies that fail to observe congruency effects for cumulative area in number discrimination tasks from those that do: those that find no congruency effects provided feedback (e.g., DeWind & Brannon, 2012; Halberda & Feigenson, 2008; Odic et al., 2016; Wang et al., 2016), while those with congruency effects did not (e.g., Aulet & Lourenco, 2021; Inglis & Gilmore, 2014; Tokita & Ishiguchi, 2010).

Understanding what congruency effects signal about the relationship between number and non-numeric visual features has been an important subject of debate. Our findings challenge the use of congruency effects to support claims about the role of non-numeric cues on number perception. While congruency effects between cumulative area and number have been found, our data suggests that these findings do not necessitate cumulative area being mandatory for number, but rather these congruency effects could stem from response conflicts. Work that has relied on congruency effects for arguments about number being fundamentally reliant on specific non-numeric features will need to consider whether congruency effects with other non-numeric cues may similarly emerge because of competition for the same decision-making components as number, rather than being obligatory cues for number perception (e.g., Leibovich et al., 2017). The methodology here provides one potential way to disentangle these possibilities.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgments

This work is supported by a Natural Sciences and Engineering Discovery Grant awarded to D.O., the Faculty of Graduate Studies Doctoral Fellowship awarded to D.D., a Social Sciences and Humanities Research Council of Canada Postdoctoral Fellowship awarded to C.B., and a University of British Columbia, Data Science Institute grant awarded to C.D.B.

References

Anobile, G., Cicchini, G. M., & Burr, D. C. (2014). Separate mechanisms for perception of numerosity and density. *Psychological Science, 25*(1), 265–270.

Anobile, G., Cicchini, G. M., & Burr, D. C. (2016). Number as a primary perceptual attribute: A review. *Perception, 45*(1–2), 5–31.

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

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition, 86*(3), 201–221.

Barth, H. C. (2008). Judgments of discrete and continuous quantity: An illusory stroop effect. *Cognition, 109*(2), 251–266.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology, 18*(6), 425–428.

Cheyette, S. J., & Piantadosi, S. T. (2020). A unified account of numerosity perception. *Nature Human Behaviour, 4*(12), 1265–1272. https://doi.org/10.1038/s41562-020-00946-0

Clayton, S., Gilmore, C., & Inglis, M. (2015). Dot comparison stimuli are not all alike: The effect of different visual controls on ANS measurement. *Acta Psychologica, 161*, 177–184. https://doi.org/10.1016/j.actpsy.2015.09.007

Dakin, S. C., Tibber, M. S., Greenwood, J. A., & Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences, 108*(49), 19552–19557.

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition, 142*, 247–265.

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: Effects of feedback and training. *Frontiers in Human Neuroscience, 6*, 68.

Durgin, F. H. (2000). The reverse stroop effect. *Psychonomic Bulletin & Review, 7*(1), 121–125.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.

Gebuis, T., Kadosh, R. C., & Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: A critical review. *Acta Psychologica, 171*, 17–35.

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods, 43*(4), 981–986. https://doi.org/10.3758/s13428-011-0097-5

Gebuis, T., & Reynvoet, B. (2012a). Continuous visual properties explain neural responses to nonsymbolic number. *Psychophysiology, 49*(11), 1649–1659.
Gebuis, T., & Reynvoet, B. (2012b). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General, 141*(4), 642.
Gebuis, T., & Reynvoet, B. (2012c). The role of visual information in numerosity estimation. *PloS One, 7*(5), Article e37426.
Gebuis, T., & van der Smagt, M. J. (2011). Incongruence in number–luminance congruency effects. *Attention, Perception, & Psychophysics, 73*(1), 259–265.
Gilmore, C., Cragg, L., Hogan, G., & Inglis, M. (2016). Congruency effects in dot comparison tasks: Convex hull is more important than dot area. *Journal of Cognitive Psychology, 28*(8), 923–931.
Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the" number Sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457.
Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proceedings of the National Academy of Sciences, 109*(28), 11116–11120.
Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*(7213), 665–668.
Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences, 103*(51), 19599–19604.
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*(3), 1221–1247.
Kay, M. (2021). *ggdist: Visualizations of distributions and uncertainty, R package version 2.2.0*.
Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception, 36*(14), 1–16.
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*.
Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*(6), 1292–1300.
Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica, 141*(3), 373–379.
Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111*(1), 1.
Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development, 82*(4), 1224–1237.
Nasr, K., Viswanathan, P., & Nieder, A. (2019). Number detectors spontaneously emerge in a deep neural network designed for visual object recognition. *Science Advances, 5*(5), Article eaav7903.
Odic, D. (2018). Children's intuitive sense of number develops independently of their perception of area, density, length, and time. *Developmental Science, 21*(2), Article e12533.
Odic, D., Lisboa, J. V., Eisinger, R., Olivera, M. G., Maiche, A., & Halberda, J. (2016). Approximate number and approximate time discrimination each correlate with school math abilities in young children. *Acta Psychologica, 163*, 17–26.
Odic, D., Pietroski, P., Hunter, T., Lidz, J., & Halberda, J. (2013). Young Children's understanding of "More" and discrimination of number and surface area. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 39*(2), 451–461. https://doi.org/10.1037/a0028874
Odic, D., & Starr, A. (2018). An introduction to the approximate number system. *Child Development Perspectives, 12*(4), 223–229.
Paul, J. M., van Ackooij, M., Ten Cate, T. C., & Harvey, B. M. (2022). Numerosity tuning in human association cortices and local image contrast representations in early visual cortex. *Nature Communications, 13*(1), 1–15.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.
Picon, E., Dramkin, D., & Odic, D. (2019). Visual illusions help reveal the primitives of number perception. *Journal of Experimental Psychology: General, 148*(10), 1675.
Pietroski, P., Lidz, J., Hunter, T., & Halberda, J. (2009). The meaning of 'Most': Semantics, numerosity and psychology. *Mind & Language, 24*(5), 554–585. https://doi.org/10.1111/j.1468-0017.2009.01374.x
Ross, J., & Burr, D. C. (2010). Vision senses number directly. *Journal of Vision, 10*(2), 10-10.
Sorensen, T., & Vasishth, S. (2015). *Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists*. ArXiv Preprint ArXiv:1506.06201.
Stan Development Team. (2020). *RStan: The R interface to Stan. R package version 2.17. 3*.
Stoianov, I., & Zorzi, M. (2012). Emergence of a'visual number sense'in hierarchical generative models. *Nature Neuroscience, 15*(2), 194–196.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643.
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.
Testolin, A., Dolfi, S., Rochus, M., & Zorzi, M. (2020). Visual sense of number vs. sense of magnitude in humans and machines. *Scientific Reports, 10*(1), 10045. https://doi.org/10.1038/s41598-020-66838-5
Testolin, A., Zou, W. Y., & McClelland, J. L. (2020). Numerosity discrimination in deep neural networks: Initial competence, developmental refinement and experience statistics. *Developmental Science, 23*(5), Article e12940.
Tokita, M., & Ishiguchi, A. (2010). How might the discrepancy in the effects of perceptual variables on numerosity judgment be reconciled? *Attention, Perception, & Psychophysics, 72*(7), 1839–1853.
Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review, 101*(1), 80.
Wang, J. J., Odic, D., Halberda, J., & Feigenson, L. (2016). Changing the precision of preschoolers' approximate number system representations changes their symbolic math performance. *Journal of Experimental Child Psychology, 147*, 82–99.
Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*(2), 130–137.
Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition, 89*(1), B15–B25.
Yousif, S. R., & Keil, F. C. (2019). The additive-area heuristic: An efficient but illusory means of visual area approximation. *Psychological Science, 30*(4), 495–503.
Yousif, S. R., & Keil, F. C. (2020). Area, not number, dominates estimates of visual quantities. *Scientific Reports, 10*(1), 1–13.
Zimmermann, E. (2018). Small numbers are sensed directly, high numbers constructed from size and density. *Cognition, 173*, 1–7.
Zimmermann, E., & Fink, G. R. (2016). Numerosity perception after size adaptation. *Scientific Reports, 6*(1), 1–7.
Zorzi, M., & Testolin, A. (2018). An emergentist perspective on the origin of number sense. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1740), Article 20170043.