Cognition 239 (2023) 105565

Contents lists available at ScienceDirect

Cognition

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

Fundamental units of numerosity estimation

Ramakrishna Chakravarthi [a,*], Andy Nordqvist [a], Marlene Poncet [b], Nika Adamian [a]

[a] School of Psychology, University of Aberdeen, Aberdeen, United Kingdom
[b] School of Psychology & Neuroscience, University of St Andrews, St Andrews, United Kingdom

ARTICLE INFO

Keywords:
Numerosity estimation
Gestalt grouping
Approximate number system
Numerical cognition
Visual perception

ABSTRACT

Humans can approximately enumerate a large number of objects at a single glance. While several mechanisms have been proposed to account for this ability, the fundamental units over which they operate remain unclear. Previous studies have argued that estimation mechanisms act only on topologically distinct units or on units formed by spatial grouping cues such as proximity and connectivity, but not on units grouped by similarity. Over four experiments, we tested this claim by systematically assessing and demonstrating that similarity grouping leads to underestimation, just as spatial grouping does. Ungrouped objects with the same low-level properties as grouped objects did not cause underestimation. Further, the underestimation caused by spatial and similarity grouping was additive, suggesting that these grouping processes operate independently. These findings argue against the proposal that estimation mechanisms operate solely on topological units. Instead, we conclude that estimation processes act on representations constructed after Gestalt grouping principles, whether similarity based or spatial, have organised incoming visual input.

1. Introduction

Humans and members of several other species can estimate the number of objects at a single glance (Dehaene, 1992; Nieder, 2005). This estimation ability is intimately tied to numerical cognition in humans and is considered foundational for the acquisition of mathematical competence, particularly among children (Anobile et al., 2019; Starr, Libertus, & Brannon, 2013). Yet, there is an ongoing debate about how estimation is implemented in humans. Some have attributed it to an innate 'number sense' that directly apprehends numerosity (Anobile, Cicchini, & Burr, 2013; Burr & Ross, 2008; Dehaene, 1992). This number sense has been proposed to operate independently of other factors that often co-vary with numerosity, such as occupied area or density (Burr & Ross, 2008; Cicchini, Anobile, & Burr, 2016; DeWind, Park, Woldorff, & Brannon, 2019). On the other hand, it has been proposed that there is no innate number sense per se. Some have argued that the visual system first evaluates continuous properties of the visual input such as density or spatial frequency content and then computes numerosity on the basis of such low-level visual quantities (Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011; Paul, van Ackooij, ten Cate, & Harvey, 2022). Others have posited that numerical and non-numerical magnitude information (such as size, cumulative surface area, or density) are conjointly represented. Numerosity perception would be the late-stage read-out of this conjoint representation (Gebuis & Reynvoet, 2012a, 2012b; Leibovich, Katzin, Harel, & Henik, 2016), that might be executed by selective attention (Aulet & Lourenco, 2021; Lourenco & Aulet, 2022).

Irrespective of the precise mechanisms underlying estimation, it nevertheless remains unclear what the *inputs* to the estimation mechanisms are. Specifically, it is unknown whether these mechanisms operate over individual objects or segmented collections of features. Computational models of numerosity perception often simply register individual objects (Allik & Raidvee, 2021; Cheyette & Piantadosi, 2020; Im, Zhong, & Halberda, 2016) and add constraints to account for the variability in human enumeration performance. For example, one model argues that each object is represented with some probability depending on factors such as proximity to the nearest neighbour (Allik & Raidvee, 2021). Another model posits that groups or clusters of objects are the units on which estimation mechanisms operate (Im et al., 2016). These constraints indicate that the fundamental units of estimation depend on specific inter-object relationships. Indeed, connecting objects with lines leads to underestimation (Anobile, Cicchini, Pomè, & Burr, 2017; Franconeri, Bemis, & Alvarez, 2009; He, Zhou, Zhou, He, & Chen, 2015; Yu, Xiao, Bemis, & Franconeri, 2019). Similarly, other ways of spatially

* Corresponding author.
E-mail addresses: rama@abdn.ac.uk (R. Chakravarthi), a.nordqvist.22@abdn.ac.uk (A. Nordqvist), mfp7@st-andrews.ac.uk (M. Poncet), nika.adamian@abdn.ac.uk (N. Adamian).

https://doi.org/10.1016/j.cognition.2023.105565
Received 23 October 2022; Received in revised form 22 May 2023; Accepted 13 July 2023
Available online 22 July 2023

grouping objects, such as enclosing them within a common region or spatially clustering them also lead to underestimation (He et al., 2015; Yu et al., 2019). On the other hand, it was found that grouping by non-spatial cues such as by colour or shape similarity did not appear to lead to underestimation (He et al., 2015; Yu et al., 2019). Based on such findings, He et al. (2015) proposed that distinct topological units form the inputs to estimation mechanisms. They argued that topologically connected objects are represented as single objects and hence underestimated. Others have asserted that estimation mechanisms operate over units generated by spatial grouping (Im et al., 2016; Yu et al., 2019). The latter does not invoke topological principles but is not incompatible with it. According to this idea, objects bound to each other by Gestalt principles like connectivity, closure, and common region are represented as single units, which leads to underestimation. Other Gestalt principles that lead to appearance-based grouping (e.g., by colour similarity or common fate) might lead to perceptually grouped units, but these do not form the inputs to numerosity estimation mechanisms. In such a case, the estimation mechanisms would instead operate on individual objects.

However, there is considerable evidence that grouping by similarity supports segmentation and guides attention (e.g. Treisman, 1982). Grouping by similarity is known to be centrally involved in figure-ground processing, contour integration, border assignment, and in the formation of object representations (see Wagemans et al., 2012 for an extensive review). Such segmentation could potentially affect numerosity processing. Importantly, similarity in appearance has been shown to modulate estimation processes. For example, similarly oriented Gabors were overestimated relative to randomly oriented Gabors (DeWind, Bonner, & Brannon, 2020). Relatedly, there is tentative evidence that grouping by similarity causes underestimation. When objects with (two) distinct colours or motion directions were intermixed, their overall numerosity was underestimated but not when the objects were uniform (Poom, Lindskog, Winman, & van den Berg, 2019). However, this appeared to be the case only for some numerosities (20) but not for others (16). Surprisingly, spatial clustering in the same paradigm led to overestimation, contrary to what has been established by numerous studies (Allik & Tuulmets, 1991; Bertamini, Zito, Scott-Samuel, & Hulleman, 2016; Yu et al., 2019). These peculiar results could be due to the unusual stimulus (limited lifetime displays, which gave the stimulus a 'twinkling' appearance) and protocol (unlimited viewing with an adjustment task) adopted in this study. More importantly, their manipulations of colour, size and motion directions did not lead to perceptual segregation of objects into distinct groups, as the objects with different properties were intermixed. Hence, while indicative, the results cannot be taken to suggest a general influence of appearance similarity on estimation processes. A different finding that provides more compelling evidence of the effect of similarity grouping on segmentation is that humans can simultaneously select up to three subsets of objects by their colour and subsequently enumerate any one of them (Halberda, Sires, & Feigenson, 2006). This result suggests that objects that appear similar can be selected as a unit, and multiple (up to three) such units can be represented at any given time.

The above considerations would lead one to expect that units formed by similarity grouping can also potentially be inputs to estimation mechanisms. It is therefore surprising that previous examinations of this question did not find an effect of similarity grouping on estimation. This lack of evidence might indicate that estimation is genuinely a spatial process, or it might be driven by confounding factors such as mismatched strengths between different grouping cues. That is, the strength of similarity-based grouping might have been inadvertently weaker than that of spatial grouping, perhaps because of using colours and shapes that were too similar to each other to allow adequate segmentation.

The proposals that topological units (He et al., 2015) or spatially grouped clusters (Im et al., 2016) are the fundamental units for numerosity estimation mechanisms *require* that non-topological or non-spatial properties, such as appearance, should not affect estimation. Hence, it is critical to rigorously test if similarity grouping can also lead to underestimation before these proposals are considered viable. In four experiments, we systematically examined the effect of a range of similarity-based and spatial grouping cues on estimation to determine the units of estimation. If similarity-based grouping also causes underestimation, it would support the contention that all forms of Gestalt grouping create bound units that are then processed by numerosity estimation mechanisms.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Twenty-five participants (20 females, 5 males; 3 left-handed) aged 18–60 years (mean = 29; standard deviation = 14.6) took part in this experiment. All participants had self-reported normal colour vision and normal or corrected-to-normal visual acuity. Participants provided written informed consent and were reimbursed with £15 for their time. The study was approved by the Psychology Ethics Committee at the University of Aberdeen and was conducted in accordance with the Declaration of Helsinki.

2.1.2. Materials and stimuli

The stimuli were generated and presented using MATLAB with Psychtoolbox extensions (Kleiner et al., 2007; Pelli, 1997). We created and e-mailed MATLAB executable files to the participants to conduct the experiment remotely on their personal computer (laptop or desktop computer running the Windows operating system). Participants were provided detailed instructions about how to run the experiment by themselves. Given the inevitable variability of the setup across participants, stimulus sizes were programmed and expressed as fractions of the screen width.

The stimuli consisted of patches of non-overlapping, coloured circles presented on a mid-grey background (RGB = 128, 128, 128). There were two kinds of patches: the *reference* patch and the *test* patch. The reference patch, in which grouping cues were manipulated, consisted of a fixed number (24) of circles. The test patch consisted of black circles (RGB = 0, 0, 0), whose numerosity was controlled by a 1-up, 1-down staircase on a trial-by-trial basis. The numerosity of the test patch was restricted to be between 6 and 49. Each circle had a diameter of 2% of the width of the screen (w_s). They were presented within a 7 × 7 grid centred on the screen centre (Fig. 1). Each cell in the grid had a length and breadth of 4% of w_s (which would therefore be the shortest distance between two circles). The locations of the circles in this grid were chosen randomly on each trial. Each of the circles were jittered by ±0.5% of w_s (corresponding to 0 to 25% of the cell length within the grid) in the vertical and horizontal directions to prevent regular arrangement of the circles. A fixation cross, comprising of two lines of length 0.5% of w_s, was presented at the centre of the screen.

Each participant was tested on nine grouping conditions (Fig. 2). In two spatial-grouping conditions, either two pairs of circles (connect-4) or six pairs of circles (connect-12) out of 24 were connected to each other. In the connect-4 condition, one of the 24 circles in the reference patch was chosen randomly and a connecting line between it and its nearest neighbour was drawn. Then one of the remaining 22 circles was chosen randomly and connected to its nearest neighbour as long as it was not one of the previously connected circles. The connecting lines were black and 10 pixels in width, with the caveat that computers that could not draw lines of this thickness would use the highest possible thickness. In the connect-12 condition, the set of 24 circles was divided into six non-overlapping sectors using k-means clustering. Within each sector, a circle was randomly chosen and connected to its nearest neighbour within the same sector.

In one set of similarity grouping manipulations, we changed the colour of four, eight or twelve circles in the reference patch (always

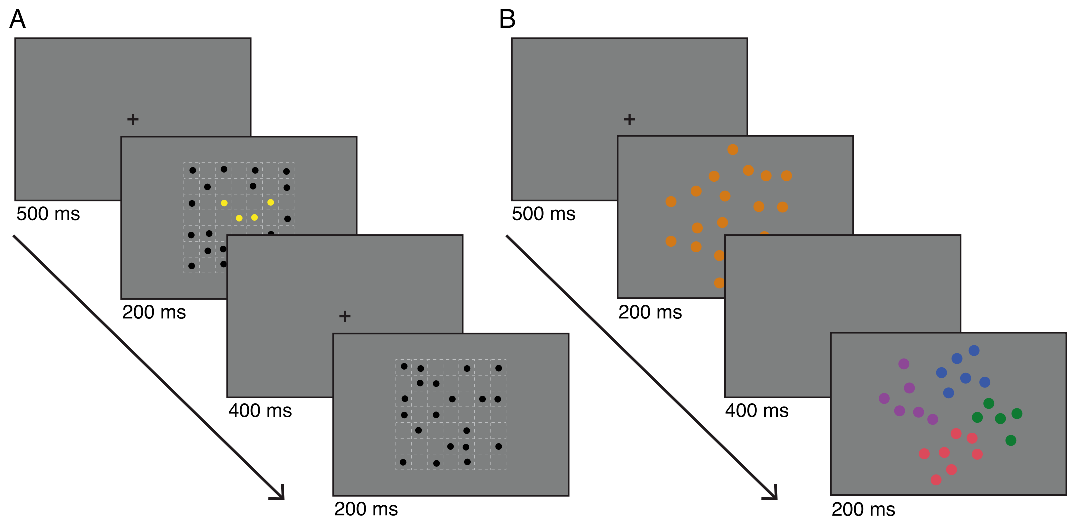

Fig. 1. Trial protocol in Experiments 1 through 4. A. Protocol in Experiments 1 and 2. B. Protocol in Experiments 3 and 4. In all experiments, after a fixation period of 500 ms, a patch of circles was presented for 200 ms followed by a gap of 400 ms (during this period, the fixation cross was present in Experiments 1 and 2, but not in 3 and 4). Subsequently a second patch was presented for 200 ms. Participants then had unlimited time to report which patch had more circles. The light grey grid shown in A is for illustration purposes and was not visible to the participants. The location of the circles was jittered (not shown) to avoid perceiving them as aligned.

composed of 24 circles) to a highly salient yellow colour (RGB: 200, 220, 50). In each of these three conditions (yellow-4, yellow-8, yellow-12), a circle was first randomly chosen. It and its n-1 nearest neighbours were coloured yellow.

In the other set of similarity grouping manipulations, we segregated subsets of objects using an increasing number of colours. In this manipulation, the 24 circles in the reference patch were segmented and coloured with either one, two, four or six distinct colours (1-colour, 2-colours, 4-colours, 6-colours conditions). That is, the appropriate number of colours were chosen randomly (sampled without replacement) on each trial from a set of six colours that were visually distinct from each other (yellow: 200, 220, 50; blue: 80, 180, 230; pink: 230, 120, 130; green: 50, 230, 150; purple: 150, 50, 220; light violet: 180, 200,250). Except in the case where a single colour was given to all circles, the circles in the reference patch were divided into the relevant number of sectors (2, 4 or 6) by k-means clustering. All circles within each sector were assigned one colour from the chosen subset.

The process of creating the reference patches was the same across all conditions. The grouping manipulations were imposed on the reference patch after creating the layouts. Hence, many of the low-level properties, such as density, convex-hull, and aggregate area, should be approximately the same across all conditions; similarly, any differences along these dimensions between reference and test patches would be comparable across the grouping manipulations. The outcomes in these conditions should therefore reflect differences in grouping among them.

2.1.3. Procedure

Participants' numerosity perception was measured using a two-interval forced-choice task (Fig. 1A). Each trial started with a 500 ms fixation period. Then, the reference and test patches were presented in succession. Each display was presented for 200 ms with a gap of 400 ms between the offset of the first and the onset of the second stimulus (Inter Stimulus Interval). The order of presentation of the two patches was randomised on each trial. Participants were asked to report the patch that appeared more numerous through keypresses. There was no time restriction for their response; no feedback was provided.

Participants began with 20 practice trials where the reference patch consisted of 24 white circles. Each of the nine condition was subsequently tested in a separate block. Their order was randomised across participants. Two one-up, one-down staircases were run for each condition, one where the test patch started with a higher numerosity (randomly chosen between 26 and 34) and one that started with a lower numerosity (randomly chosen between 14 and 22) than the reference patch. On each trial, the staircase that controlled the test patch numerosity was randomly chosen. Each staircase terminated after 20 consecutive reversals. Participants were encouraged to take a short break every forty trials and between each of the nine blocks.

2.1.4. Data analysis

Data (along with scripts for stimulus presentation and data analysis) from all experiments are available at OSF: https://osf.io/wmxqn/. For each participant and condition, we pooled single-trial responses from the two staircases and fitted a logit psychometric function to all responses. The point of subjective equality (PSE) extracted from the fit indicated the number of circles needed in the test patch to appear numerically equivalent to that in the reference patch. We then used the bootstrap approach to assess goodness-of-fit (Wichmann & Hill, 2001) and removed PSE estimates where bootstrapped standard deviation exceeded 2. This resulted in the removal of two thresholds (0.9% of data) in Experiment 1, fourteen thresholds (6.6% of data) in Experiment 2, and four thresholds (5.9% of data) in Experiment 3. The exclusion criterion for Experiment 4 was slightly different (please see section 5.2.2).

The remaining PSEs were entered into a linear mixed effects model implemented using *lme4* and *lmerTest* packages in R (Bates, Mächler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2017) with grouping condition as a fixed factor and participant as a random factor (see Supplementary Statistics for details and output). Significance testing was carried out using Satterthwaite approximation for the degrees of freedom (Satterthwaite, 1946). Planned comparisons were performed among conditions within a given kind of grouping manipulation (e.g., number of colours) and are reported in Supplementary Statistics. We conducted additional (traditional) data analyses similar to approaches used in previous studies (e.g., Yu et al., 2019), which yielded the same results (see Supplementary Data Analysis).

Following Pomè, Anobile, Cicchini, Scabia, and Burr (2019) we computed the Just Noticeable Difference (JND) using a similar procedure as for the PSE. We first determined the numerosity at which the test patch appears more numerous than the reference patch 75% of the time. We then calculated JND as the difference between this numerosity and the PSE (which indexes the numerosity at which the test patch appears more numerous 50% of the time). JND is a measure of precision (slope of the psychometric curve). It indicates if and how various forms of grouping affect the sensitivity of the numerosity mechanisms.

All plots in the results sections were created using the *gramm* package for MATLAB (Morel, 2018).

2.2. Results

We compared PSEs (FDR corrected for multiple comparisons; Benjamini & Hochberg, 1995) for each of the nine conditions against 24 to determine if numerosity was either over- or under-estimated (Fig. 2; Table 1). Among the two spatial-grouping conditions where subsets of circles were connected, estimation did not differ from 24 when two pairs of circles were connected (connect-4), but numerosity was under-estimated when more pairs of circles (connect-12) were connected. The lack of underestimation in the connect-4 conditions is in contrast to previous studies (Franconeri et al., 2009; He et al., 2015), but might be explained by the connection being between very closely spaced circles. These might already have been perceived as grouped units because of proximity cues and hence the effect of connectivity might have been weaker in this condition. Nevertheless, we found the classic effect of connectivity on estimation when more pairs of circles were connected.

In the first similarity manipulation, we changed the colours of 4, 8 or 12 circles out of 24 from black to yellow. No underestimation was observed in any of these conditions, which replicates and extends He et al. (2015)'s finding that grouping by changing the appearance of some objects to a different colour does not lead to underestimation.

In the second similarity manipulation, we assigned 1, 2, 4 or 6 colours to clusters of circles within the reference patch. We observed underestimation in the 2, 4, and 6-colour conditions indicating that, unlike previous reports, similarity-based grouping does lead to underestimation. Interestingly, underestimation was strongest when four colours were assigned to the circles compared to when one or two colours were assigned (one vs four: $t_{(72)} = 4.18$, $p = .0005$, $d = 1.18$; two vs four: $t_{(72)} = 3.05$, $p = .01$, $d = 0.86$). On the other hand, underestimation was slightly weaker with six colours but not significantly different than with four colours ($t_{(72)} = -1.55$, $p = .17$, $d = -0.44$).

To test whether the underestimations caused by spatial and similarity grouping in our setup were similar, we compared the maximum underestimation caused by similarity grouping (4-colours condition) with that caused by spatial grouping connectivity (connect-12 condition); we found that similarity grouping caused greater underestimation ($t_{(190)} = 2.195$, $p = .029$, $d = 0.62$). This shows that similarity grouping can be at least as strong as, if not more than, spatial grouping in inducing underestimation, in contrast to previous findings (He et al., 2015; Yu et al., 2019).

The JNDs (a measure of precision based on the slope of the psychometric curve) were comparable across almost all tested conditions (Supplementary Data Analysis). That is, despite the effect of grouping on the accuracy of estimation, the sensitivity of the numerosity mechanisms was not affected by grouping. The only exception was that the JND for the 4-connect condition was higher than that for the 12-connect condition ($t_{(22.1)} = 2.215$, $p = .0374$), indicating that increased connectivity increases the sensitivity of the numerosity estimation mechanism. Previous studies (Adriano, Girelli, & Rinaldi, 2022; He, Zhang, Zhou, & Chen, 2009) have not observed a similar change in sensitivity as a result of real or illusory connectedness. This finding has to be systematically tested further.

2.3. Discussion

The main finding that similarity grouping leads to underestimation contradicts the topological hypothesis (He et al., 2015), which predicts that a non-topological change should not affect perceived numerosity. However, it clearly does, thus ruling out the possibility that numerosity mechanisms act *only* on topological units. The results instead suggest that numerosity mechanisms are sensitive to a range of grouping principles, including both spatial- and similarity-based ones.

It is interesting to note that the two similarity manipulations appear to give somewhat contradicting results. Although the yellow-12 and the 2-colours conditions are ostensibly the same, underestimation was observed only in the latter. This might be taken to argue against the grouping explanation of underestimation. However, there are two major differences between the two sets of manipulations that might account for this discrepancy. First, in the 'yellow' conditions, the circles were black and yellow, whereas in the 2-colour condition, the two colours were randomly chosen on each trial. That is, participants could use a strategy to discount potential grouping processes in the former but not in the latter condition (Adam & Serences, 2021; Geng, Won, & Carlisle, 2019). Second, in the 'yellow' conditions, the circles that were coloured yellow were randomly chosen and hence were likely to have been embedded among (or encircled by) black circles (as illustrated in Fig. 2), whereas in the 2-colour condition, two non-overlapping sectors were identified and

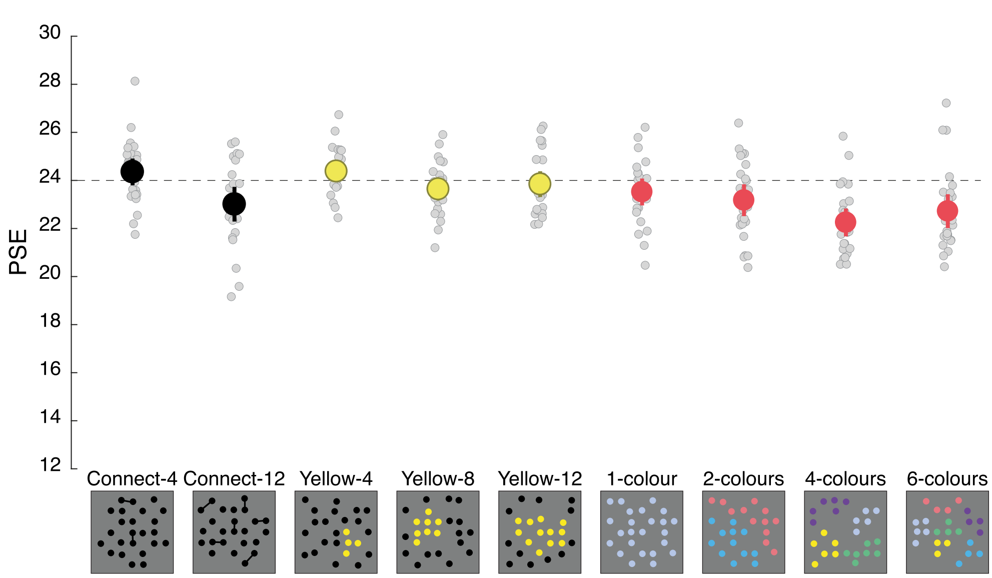

Fig. 2. Points of subjective equality (PSEs) in Experiment 1. Example stimuli, one for each grouping condition, are shown along the x-axis (note, however, that the locations of circles within each frame were jittered in the actual experiment). Black circles in the plot represent spatial grouping conditions where either 4 or 12 circles were connected. Yellow circles represent similarity grouping conditions where 4, 8 or 12 circles were coloured yellow. Red circles represent the colour grouping conditions, where the 24 circles were perceptually clustered with 1, 2, 4, or 6 colours. In this and all subsequent figures, the large circles represent the mean PSE across participants for each of the tested conditions. Error bars represent 95% confidence intervals. Each small, grey dot represents an individual participant's PSE for that condition. The dashed horizontal line represents the numerosity of the reference patch. PSE below this line indicate underestimation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Results from Experiment 1. The number of circles needed in the test patch to appear equally numerous as the 24 circles in the reference patch (point of subjective equality: PSE) for each of the nine grouping manipulations are presented, along with their corresponding confidence intervals (CI). Also presented are the results of statistical comparisons against 24: the t-statistic, the corresponding *p*-values with and without corrections for multiple comparisons based on the false discovery rate (FDR) approach, and the estimated effect size.

Condition	PSE	95%CI	t-statistic	p-value	FDR corrected p-value	Estimated standardized effect size (*d*)
Connect-4	24.4	23.8–24.9	1.25	.21	.23	0.25
Connect-12	23.0	22.5–23.6	−3.49	.0006	.002	−0.7
Yellow-4	24.4	23.8–24.9	1.34	.18	.23	0.27
Yellow-8	23.6	23.0–24.1	−1.41	.16	.23	−0.29
Yellow-12	23.8	23.3–24.4	−0.59	.55	.55	−0.12
1-colour	23.5	23.0–24.0	−1.69	.09	.16	−0.34
2-colours	23.2	22.6–23.7	−2.90	.004	.009	−0.58
4-colours	22.3	21.7–22.8	−6.16	<.0001	<.0001	−1.23
6-colours	22.7	22.2–23.3	−4.51	<.0001	<.0001	−0.9

assigned different colours. It is possible that underestimation occurs only when grouped segments are segregated and not embedded within other segments. Indeed, when the yellow circles are embedded among black circles, the display might be akin to the solitaire illusion (Bertamini, Guest, Contemori, & Zito, 2023; Frith & Frith, 1972). In one example of this illusion (Agrillo, Parrish, & Beran, 2016), small sets of black objects surround a set of white objects. This makes it appear as if the white objects are embedded within the peripheral black clusters. In such a stimulus, the peripheral black objects are underestimated, whereas the central white objects are overestimated. This is not just relative to each other, but relative to their actual numerosities. These misestimations are roughly equal. Under these circumstances, if participants are asked to estimate the *total* number of objects, as we did in this experiment, the estimation might be veridical, even though each subset is incorrectly enumerated. Because of this solitaire illusion, circles would not be underestimated in the 'yellow' conditions but they would in the 2-colour condition. Similarly, it can be said that the embedded (yellow) objects are seen as the foreground and the surrounding (black) objects are seen as the background, whereas in the 2-colour condition, both segments would be the foreground. Although speculative, this difference might potentially underlie the difference in underestimation between the two types of similarity-based grouping manipulations.

3. Experiment 2

3.1. Introduction

The second experiment was designed to further test and extend the finding that similarity grouping leads to underestimation. First, we sought to determine if the findings generalise to similarity grouping cues other than colour. Hence, we tested if grouping by shape would also lead to underestimation. Second, we examined the cause for the discrepancy between our findings and previous results. Specifically, assigning six colours among 24 circles led to clear underestimation in our study but not in the study by Yu et al. (2019). This discrepancy might be due to Yu and colleagues using smaller circles in their displays (in addition to using less distinct colours, which we corrected for in our study). Colours appear less saturated when object size is small (Gordon & Abramov, 1977), which could have led to weaker grouping. We tested this possibility by assessing underestimation with circles of different sizes. Third, we pitted a particularly strong form of spatial grouping, closure, known to produce substantial underestimation (Yu et al., 2019), against similarity grouping to assess the comparability between the two forms of grouping. Given previous reports that the two forms of grouping act independently (Kubovy & van den Berg, 2008; Wagemans et al., 2012), we also assessed if their effects were additive.

3.2. Methods

3.2.1. Participants

Twenty-one participants (15 females, 6 males; 2 left-handed) between the ages of 19 and 61 years (mean = 28.2, standard deviation = 13.9) took part in this experiment. Twelve of these participants had also taken part in Experiment 1. All participants had self-reported normal colour vision and normal or corrected-to-normal visual acuity. Participants provided written informed consent and were reimbursed with £15 for their time. The study was approved by the Psychology Ethics Committee at the University of Aberdeen and was conducted in accordance with the Declaration of Helsinki.

3.2.2. Materials, stimuli, and procedure

The materials and procedure were the same as in Experiment 1. Participants were tested in ten conditions (Fig. 3). The test patch was always composed of black circles whose numerosity varied. Grouping cues were manipulated in the reference patch. In the baseline condition, the reference patch consisted of 24 black circles (identical to those in the test patch but with different locations). In one set of manipulations, we assigned four colours to clusters of circles, just as in the 4-colours condition of Experiment 1 and manipulated their size: the circles had a diameter of 0.8%, 1.4% or 2% of w_s (size-small, size-medium, and size-large conditions, respectively). The large size used here was the same as the size of the circles in the test patch and in Experiment 1. Note that 'size' is relative here, since each participant's screen size and viewing distance would have been different. But the diameter of the circles in the size-large condition was 2.5 times that of the circles in the size-small condition; there would have been a corresponding and substantial 6.25 times difference in area. Hence, any effect of size should be visible across these levels, even if a given 'size' was not constant across participants.

In a second set of conditions, we grouped objects by shape similarity rather than colour. Two, four or six shapes were assigned to the objects in the same manner as colours were assigned in Experiment 1 (2-shape, 4- shape and 6-shape conditions). Shapes were selected randomly on each trial from a set of seven character symbols: +, >, &, Y, §, ø, and #. We chose these symbols because they were visually distinct from each other. The characters were drawn in Bold Courier font (an equal width font) at a font size equivalent to 3% of w_s.

In the final set of three conditions, 12 circles out of 24 were connected. In the closure-condition, four triplets of circles were connected into non-overlapping triangles. This was done as follows. The 24 circles were divided into four clusters using k-means clustering. One circle near the centre of each cluster was selected. Then, its two nearest neighbours were selected, and lines were drawn between these three circles to form a closed triangle, with one constraint. The angle between any two sides of the triangle was restricted to be between 40 and 120 degrees to prevent creating collinear or near-collinear connections. In the remaining two conditions, a similarity cue was added to the closure cue.

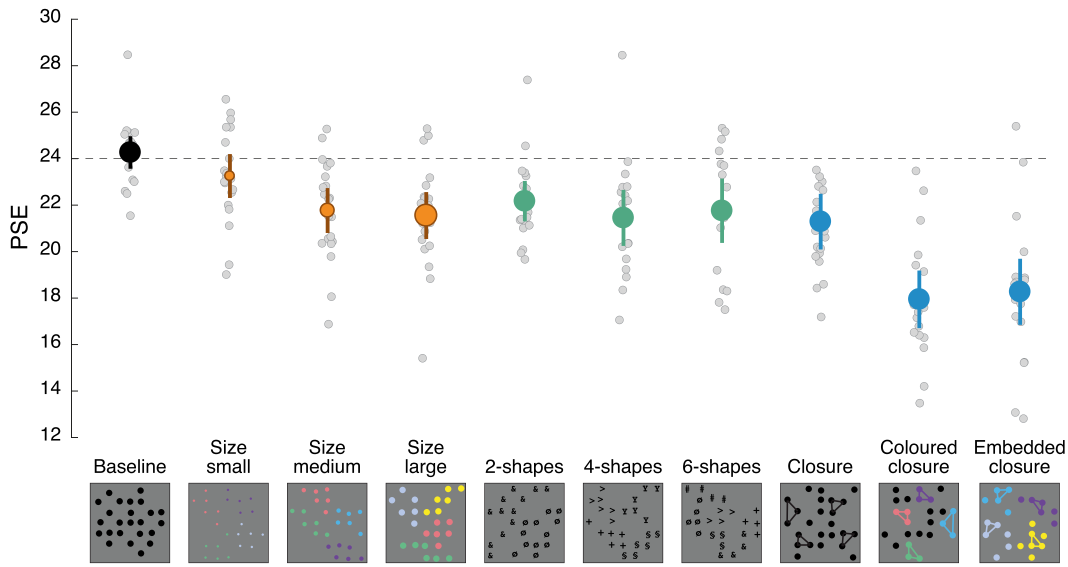

Fig. 3. Results from Experiment 2. This experiment tested nine grouping conditions in addition to a baseline condition (black circle) without any grouping manipulation. Orange circles represent conditions where the size of the circles was manipulated. The size manipulation is illustrated by the size of the marker of the mean PSE (small, medium, large). Green circles depict conditions where shape similarity was manipulated by using 2, 4, or 6 symbols as objects. Blue circles depict conditions where closure was manipulated along with colour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In the coloured-closure condition, each of the closed triangles (the circles and lines connecting them) was assigned a colour chosen randomly and without replacement from the set of six colours used in Experiment 1. The remaining 12 circles were black. In the embedded-closure condition, these remaining circles were of the same colour as the triangles they surrounded. Circles within the clusters identified by k-means clustering were given the same colour. That is, closure was embedded within groups of similar coloured objects.

3.3. Results

PSEs and the corresponding confidence intervals and statistics for each condition are presented in Fig. 3 and Table 2. There was, reassuringly, no underestimation in the baseline condition. Similarly, the JNDs did not differ across the tested conditions (Supplementary data analysis), indicating that the sensitivity of the estimation mechanisms was not affected by these grouping manipulations.

We replicated the strong underestimation induced by grouping by colour similarity found in Experiment 1. Interestingly, numerosity estimation was modulated by the size of the circles. There was no underestimation when the size of the circles was the smallest, but substantial underestimation was observed when the size of the circles was medium or large (medium vs small: $t_{(39.5)} = -2.8$, $p = .02$, $d = -0.88$; large vs small: $t_{(39.5)} = -3.2$, $p = .008$, $d = -1$). This effect of size might explain the lack of underestimation due to colour clustering in previous studies (Yu et al., 2019).

Importantly, we found that underestimation was also observed when similarity between objects was driven by shapes instead of colour. As observed in Experiment 1 with the number of colour clusters, underestimation was strongest when objects were clustered by four shapes compared to two or six shapes, but differences between conditions did not reach statistical significance (four vs two: $t_{(32.7)} = 1.47$, $p = .38$, $d = 0.48$; four vs six: $t_{(33.3)} = -1.15$, $p = .39$, $d = -0.39$).

Finally, closure led to underestimation in all three manipulations that we tested. Crucially, adding similarity cues to the closure cue substantially increased underestimation. This was the case for both manipulations: whether the elements of the closed triangles were similar to each other (coloured-closure vs closure: $t_{(35.9)} = -5.32$, $p < .0001$, $d = -1.76$) or whether closure was embedded within clusters of similar objects (embedded-closure vs closure: $t_{(35.7)} = -5.62$, $p < .0001$. $d = -1.7$). Clusters of four colours led to an underestimation of two to three objects. Closure by itself led to an underestimation of about two to three objects. Combining closure and colour similarity led to an underestimation of about 5–6 objects, which is the sum of the effects of individual cues. Supporting this contention is the finding that there was a strong correlation between the sum of underestimation observed in the closure and size-large conditions and the underestimation in the embedded-closure condition ($r = 0.63$, $p < .004$). In fact, the slope of this correlation was 1.05, indicating an almost perfect match between the two quantities (Fig. 4). That is, for any given participant, the

Table 2
PSEs for conditions tested in Experiment 2. Also reported are the corresponding confidence intervals (CI), results of *t*-tests conducted against 24 with and without corrections for multiple comparisons using the false discovery rate (FDR) approach and the estimated effect size.

Condition	PSE	95% CI	t-statistic	p-value	FDR corrected p-value	Estimated standardized effect size (*d*)
Baseline	24.3	23.3–25.4	0.59	.55	.55	0.13
Size-small	23.2	22.2–24-2	−1.51	.13	.15	−0.33
Size-medium	21.7	20.8–22.8	−4.35	< .0001	< .0001	−0.95
Size-large	21.6	20.6–22.5	−4.78	< .0001	< .0001	−1.04
2-shapes	22.1	21.1–23.2	−3.47	.0007	.0008	−0.79
4-shapes	21.3	20.3–22.4	−4.99	< .0001	< .0001	−1.14
6-shapes	21.7	20.6–22.4	−4.08	< .0001	.0001	−0.98
Closure	21.2	20.3–22.3	−5.29	< .0001	< .0001	−1.15
Coloured-Closure	17.9	16.9–19.0	−11.24	< .0001	< .0001	−2.57
Embedded-Closure	18.2	17.3–19.3	−10.95	< .0001	< .0001	−2.44

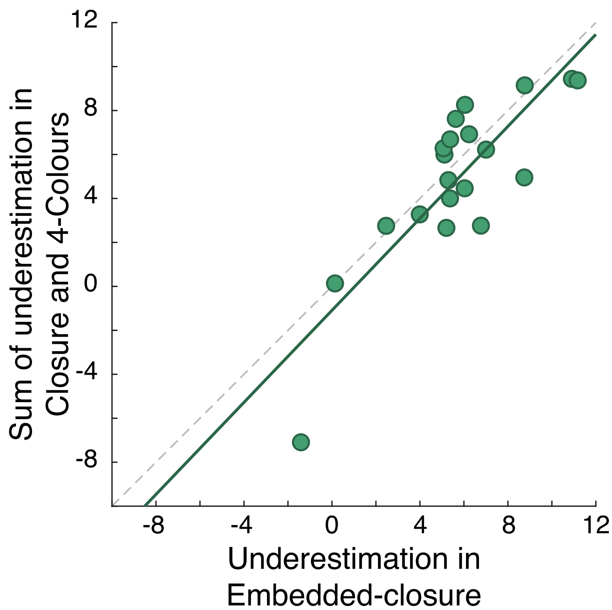

Fig. 4. Additivity of spatial and similarity-based grouping. Correlation between underestimation observed in the embedded-colour condition and the sum of underestimation observed in the closure and size-large conditions. The values are the difference between 24 and the estimated PSE. Positive numbers indicate underestimation and negative numbers indicate overestimation. Each green circle represents one participant's data. The solid green line is the best fitting straight line through these data points. The dashed grey line represents 'equality' where both quantities are the same. Participants' data are distributed closely along this equality line, with a small offset. The slope is nearly 1 indicating that the effect of combined cues is almost fully predictable from the sum of the effects of individual cues. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

underestimation due to similarity and closure cues present conjointly (embedded-closure condition) was the sum of the underestimation due to each cue taken separately. Thus, the underestimation due to similarity grouping and closure is additive, indicating that they contribute to estimation through distinct processes.

3.4. Discussion

The results of this experiment show that similarity grouping, whether by colour or shape, leads to underestimation. Importantly, underestimation due to colour and shape grouping seems to be comparable to that caused by closure, one of the most powerful spatial grouping principle (Kovács & Julesz, 1993). Further, the effects of spatial and similarity grouping are additive, indicating that their mechanisms are likely to be independent.

Our results that similarity grouping leads to underestimation differ from previous findings of veridical estimation in the presence of both colour (He et al., 2015; Yu et al., 2019) and shape similarity. The reason for this difference can probably be attributed to (inadvertent) weak similarity grouping in the previous studies. For example, the size of objects used by Yu et al. (2019) to test the effect of colour similarity might have been too small to be grouped strongly. Our results show that indeed, reducing the size of the circles in the reference patch relative to the test patch objects, led to a loss of (or a reduction in) underestimation for the smallest size objects.

Previous studies have reported that the perceived numerosity of objects decreases with increasing object size (Ginsburg & Nicholls, 1988; Tokita & Ishiguchi, 2010). In our experiment, the 'large' circles in the reference patch were the same size as the circles in the test patch. The underestimation we observed in that condition was thus not due to size, but due to grouping by colour similarity. We attribute the lack of underestimation for the smallest objects to weak grouping between them. Alternatively, it could be due to a combination of similarity grouping and the difference in size between reference and test patch objects. Similarity grouping would lead to underestimation, which would be cancelled by the overestimation due to a difference in size between reference and test objects; the former were substantially smaller than the latter and would thus have been overestimated. Our study cannot tease apart these two explanations. Both cases, however, support the idea that similarity grouping leads to underestimation.

4. Experiment 3

4.1. Introduction

The results of Experiments 1 and 2 indicate that similarity grouping leads to underestimation just as spatial grouping does and that these effects are additive. However, there are potential confounds in the stimuli that could explain these results. First, the coloured stimuli have a different luminance (and contrast) than the comparison black circles in the test patch. One might argue that the average luminance of circles in a given trial when two colours are used might be different than when four colours are used. We do not think that this is the case because of the sampling process we used to create the stimuli. In our method, each colour is likely to be picked an equal number of times within a condition. Hence, the average luminance should be the same across conditions. However, the variability of mean luminance across conditions might be different, but it is unclear how this variability cue can modulate estimation. We also think that this explanation is unlikely for two further reasons: a) the yellow-4, 8 and 12 conditions had changing mean luminance, but this was not sufficient to induce any underestimation; b) we found underestimation even with shape similarity, where luminance and contrast did not vary across shapes. Moreover, previous studies using spatial-grouping cues have shown that underestimation is robust to stimulus variations (Adriano, Girelli, & Rinaldi, 2022; Adriano, Rinaldi, & Girelli, 2022a, 2022b). Nevertheless, it is important to empirically test the role of low-level stimulus properties. We did so by testing participants with a set of equiluminant colours, which preclude luminance changes across trials and conditions. We also introduced a condition where differently coloured circles were intermixed rather than segregated. In this condition the low-level properties would remain constant while only grouping is manipulated.

A second potential confound argues that since black circles were always the comparison stimulus, which participants had to estimate on each trial, they might have been attentionally prioritised by the visual system. Hence, the other, non-black, objects would have received less attentional resources and hence been underestimated (Cheyette & Piantadosi, 2019). This *attentional prioritisation* hypothesis can explain the underestimation observed in both the colour similarity and shape similarity conditions. In the latter, to the extent that some of the shapes were dissimilar to the black circles, they would not be attentionally prioritised and would have been underestimated. We do not think that this alternative explanation captures our findings since, as noted above, there was no underestimation in the 'yellow' conditions, where there were several non-black circles in each condition. Further, this theory predicts that there should have been more underestimation in the Embedded-Closure condition than in the Coloured-Closure condition, since the latter had several black circles, while the former had none. However, there was no difference in the amount of underestimation between the two, arguing against the attentional prioritisation hypothesis.

However, we feel that this possibility needs to be tested explicitly. We therefore tested two different comparison conditions against our grouped stimuli – a) a constant equiluminant grey, or b) an equiluminant colour that changed on each trial. The attentional prioritisation hypothesis would predict that the grey circles, being constant across trials, should be prioritised and the coloured stimuli should not be. Thus, the grouped stimuli should be underestimated. Whereas when the comparison colour changes on each trial, no attentional prioritisation would be possible and hence no underestimation is expected. We included a further control condition where we asked participants to compare the numerosity of single-coloured circles with grey circles. The colour of the coloured circles changed on each trial. Here, the attentional prioritisation hypothesis would predict underestimation of the coloured circles (relative to the grey circles), since the grey will be prioritised and estimated accurately, whereas the changing colours will not be. On the other hand, the grouping hypothesis would predict no underestimation. This condition, incidentally, also serves to examine the validity of our staircasing procedure and the comparability between grey and single-colour test patches.

Finally, our first two experiments were conducted on personal devices, where we did not have control over any of the stimulus parameters (colours, luminance, displays, size, etc.) as well as motivational and attentional factors (distraction, etc.). Hence, we also wanted to examine if our results replicate in a tightly controlled environment. We conducted a laboratory experiment using well controlled stimuli and environment.

4.2. Methods

4.2.1. Participants

Seventeen participants (9 females, 5 males, 3 other) between the ages of 21 and 47 years (mean = 27.9, standard deviation = 6.7) took part in this experiment. All participants had self-reported normal colour vision and normal or corrected-to-normal visual acuity. Participants provided written informed consent. The study was approved by the Psychology Ethics Committee at the University of Aberdeen and was conducted in accordance with the Declaration of Helsinki.

4.2.2. Materials and stimuli

The experiment was conducted on a Dell computer, using PsychToolbox extensions (Kleiner et al., 2007; Pelli, 1997) for MATLAB (Mathworks, Natick, MA). A Cambridge Research Systems 32″ Display++ LCD monitor set to 1920 × 1080 pixels resolution and a 120 Hz refresh rate was used for stimulus presentation and viewed at 57 cm. Participants' head position was secured with a chin rest. This display is hardware linearised and calibrated with high precision 10-bit colour depth. The experiment was conducted in a darkened room, where the only source of light was the monitor.

The stimuli consisted of patches of non-overlapping, coloured circles presented on a mid-grey background. We selected one grey and five equiluminant colours (see Table 3). Note, however, that these were not perceptually equiluminant, as determined by a procedure like heterochromatic flicker photometry, but were nominally equiluminant as measured by a Spectrometer (SpectroCal Mark II Spectroradiometer, CRS, The United Kingdom).

As in previous experiments, there were two kinds of patches: the reference patch and the test patch. The reference patch, in which grouping cues were manipulated, consisted of a fixed number (24) of circles. The numerosity of the test patch was controlled by a 1-up, 1-down staircase on a trial-by-trial basis and was restricted to be between 8 and 40. The circles in each patch had a diameter of 1 deg. These circles were presented within an imaginary circle of diameter 14 deg. First, six locations, separated by at least 20 angular degrees (1/18th of the circle circumference), were randomly chosen on the circumference of this imaginary circle. We then iteratively introduced one location at a time within the imaginary circle, with the constraint that no new location was within 1.25 deg. of its nearest neighbour. This procedure ensured that the convex hull was approximately circular on each trial and that there were no overlapping circles.

We tested four conditions, whose order was randomised across participants (Fig. 5). In the first condition, 4-colours-colourTest, we once again tested grouping with four colours. On each trial four colours were chosen randomly from the set of five equiluminant colours. We applied k-means clustering to the reference patch and determined four segregated regions. If any subcluster had less than five circles, the entire set of locations was dropped, and a new set was generated. The process was repeated until this criterion was met. The circles within each region were assigned one of the four colours. The test patch circles were assigned the remaining equiluminant colour that was not present in the reference patch. In the second condition, 4-colours-greyTest, the reference patch was generated as in the 4-colours-colourTest condition, but the test patch circles were assigned an equiluminant grey. In the 4-colours-ungrouped-colourTest condition, the process of generating the reference patch was the same as above, but after generating the four segregated areas, the colour assignments were randomised across the circles, creating a display where circles of different colours were intermixed. Thus, everything was kept constant in the generation of the stimulus with the exception of colour assignment at the end. Finally, we included a control condition, 1-colour-greyTest, where we painted all circles in the reference patch with a single colour randomly chosen on each trial from among the five possible equiluminant colours. Estimation of these circles was compared against equiluminant grey circles.

4.2.3. Procedure

We used the same procedure as in Experiment 1 (Fig. 1B) with two small differences, noted below. Each participant completed four blocks, one for each grouping condition, the order of which was randomised across participants. The first change introduced in this experiment was that there was a set of practice trials before each block. In the first block, there were 30 practice trials, whereas there were 18 practice trials in the subsequent 3 blocks. Among the 30 trials of the first practice set, feedback was provided on the first 10 trials. The fixation turned green if the response was accurate and red if it was inaccurate. Among these trials,

Table 3
Colours of the circles used in Experiments 3 and 4. Chromatic coordinates (1931 CIE space) along with the luminance of the colours used in Experiments 3 (left set) and 4 (right set). In Experiment 4, we used three sets of three colours each, which were equiluminant within each set. The background always had a higher luminance than the circles.

Experiment 3				Experiment 4			
Colour	x	y	Y(cd/m^2)	Colour	x	y	Y(cd/m^2)
Background (grey)	0.2985	0.3101	58.5	Background (grey)	0.2975	0.3087	91.7
Pink	0.3715	0.302	43.8	Red	0.4808	0.3172	28.3
Green	0.2957	0.456	43.6	Green	0.3189	0.3975	28.2
Blue	0.2233	0.198	43.9	Blue	0.1998	0.1468	28.2
Orange	0.4607	0.4177	43.7	Brown	0.3882	0.377	43.2
Purple	0.2949	0.2207	43.9	Purple	0.2814	0.2427	43.6
Grey	0.2976	0.3088	43.9	Grey	0.2976	0.3088	43.9
				Cyan	0.2366	0.2949	67.4
				Tangerine	0.4224	0.4301	67.1
				Lime	0.3496	0.467	67.2

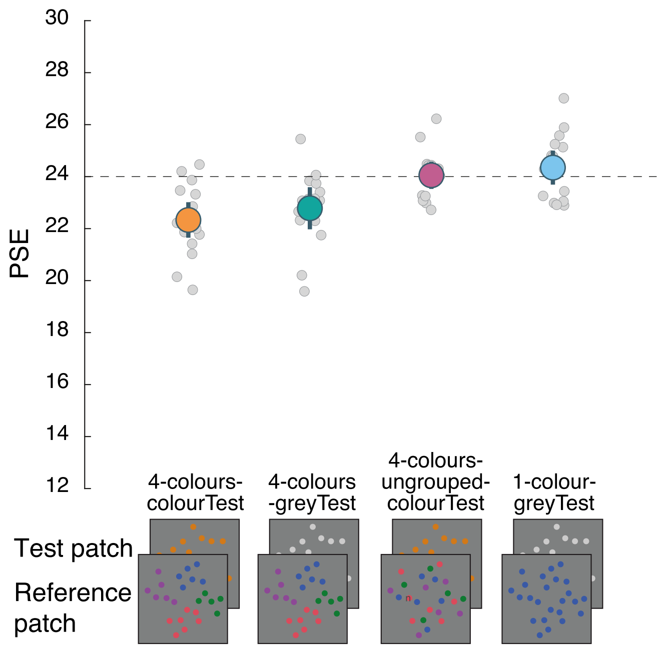

Fig. 5. Results from Experiment 3. This experiment tested four grouping conditions while controlling for potential low-level stimulus properties. The orange circle represents the condition where the objects were grouped with four colours in the reference patch while the test patch consisted of objects with a colour that was not presented in the reference patch. The green circle represents the condition where the reference patch had the same grouping manipulation as above, but the test patch contained grey objects. The purple circle represents the condition where the objects in the reference patch were not grouped by colour. The light blue circle represents the condition where the reference patch contained objects of a single colour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the numerosity in the test patch was deliberately chosen to be far from that of the reference patch (test patch: 8–12, 36–42; reference patch: 24). In the remaining practice trials, no feedback was provided (test patch numerosities: 12–36). The second small change was that a fixation cross (0.6 deg) was presented at the centre of the screen only before the first stimulus was presented and after the second was presented and not in between these events.

4.3. Results

PSEs and the corresponding confidence intervals and statistics for each condition are presented in Fig. 5 and Table 4. JNDs did not differ across the four grouping conditions (Supplementary data analysis).

We replicated the underestimation induced by similarity grouping found in Experiments 1 and 2. This was the case irrespective of whether the test patch consisted of grey circles or circles whose colour changed on each trial. The magnitude of underestimation was comparable to that found in Experiments 1 and 2, with uncontrolled stimulus, screen, and environmental parameters, indicating that the underestimation was driven by our grouping manipulation and not any of the multiple factors that might change in such uncontrolled experiments. Interestingly, estimation was veridical when the circles were assigned the same four colours but were intermixed such that no grouping could occur between them. That is, the absence of similarity grouping prevented underestimation. Finally, when all circles in the reference patch were of the same single colour, which changed on each trial, there was no underestimation. These findings argue against underestimation caused by luminance or contrast differences or the attentional prioritisation of the most frequent objects.

There was no difference in underestimation when the test patches were grey or coloured (4-colours-colourTest vs 4-colours-greyTest conditions; $t_{(44.6)} = 0.65$, $p = .17$, $d = -0.5$). But estimation was lower in both of these conditions relative to when the circles were ungrouped (4-colours-colourTest vs 4-colours-ungrouped-colourTest: $t_{(43.9)} = -1.73$, $p = .0004$, $d = -1.5$; 4-colours-greyTest vs 4-colours-ungrouped-colourTest: $t_{(43)} = -1.08$, $p = .02$, $d = -0.94$). Incidentally, estimation was comparable in the 1-colour-greyTest and 4-colours-ungrouped-colourTest conditions ($t_{(43)} = 0.32$, $p = .42$, $d = 0.28$), indicating that estimation in the ungrouped condition was the same as when a single colour was presented.

5. Experiment 4

5.1. Introduction

When we varied the number of perceptual groups in Experiments 1 and 2, we observed that underestimation was greater with four and six groups than with two groups. There was also a tendency towards higher underestimation with four groups compared to six groups, although this was not statistically supported. This was the case for grouping by both colour and shape similarity. This pattern of results might lead one to consider a potential link to subitizing, the ability to rapidly, accurately, and confidently enumerate a small number (3–4) of objects in a scene (Kaufman, Lord, Reese, & Volkmann, 1949). We might expect that underestimation increases with the number of perceptual groups until the subitizing limit is reached and then either plateaus or reduces. The subitizing limit has been attributed to the limited ability of attention to individuate a small number of objects (Chakravarthi & Herbert, 2019; Mazza, Pagano, & Caramazza, 2013; Olivers & Watson, 2008; Xu & Chun, 2009). If attention plays a role in estimation (Cheyette & Piantadosi, 2019; Pomè et al., 2019), it could be that a small number of groups are segregated at a glance and these sub-groups drive the underestimation. Indeed, there is evidence that three to four groups of similar objects can be subitized (Watson, Maylor, & Bruce, 2005). Thus, underestimation would increase up to the subitizing limit (roughly four groups). The groups beyond the subitizing limit would not be segregated and act as if they were ungrouped. When the total number of objects is constant, as in Experiments 1 and 2, this proposal argues that as the number of colours increases beyond the subitizing range, an increasing number of objects will be unsegregated and not be part of any perceptual groups. Thus, the amount of underestimation would reduce beyond 3–4 groups of objects (or at best plateau). A related argument was made by Yu et al. (2019), who contended that attention was needed for grouping by similarity. They posited that only a single set of similar objects could be segmented at a glance. Therefore, they argued that since multiple such groups could not be segmented, grouping by similarity would not

Table 4
PSEs conditions tested in Experiment 3. Also reported are the corresponding confidence intervals (CI), results of *t*-tests conducted against 24 with and without corrections for multiple comparisons using the false discovery rate (FDR) approach and estimated effect size.

Condition	PSE	95% CI	t-statistic	p-value	FDR corrected p-value	Estimated standardized effect size (*d*)
4-colours-colourTest	22.3	21.7–22.9	−5.3	<.0001	<.0001	−1.32
4-colours-greyTest	22.9	22.3–23.6	−2.9	.004	.008	−0.82
4-colours-ungrouped-colourTest	24.0	23.4–24.7	0.2	.87	.87	0.04
1-colour-greyTest	24.4	23.8–24.9	1.2	.23	.31	0.3

lead to underestimation, unlike spatial grouping, which does not require attention. Our results so far contradict this prediction. Here, we extend a test of this prediction by examining if underestimation is observed with a larger range of perceptual groups.

On the other hand, increasing the number of connected objects increases underestimation (Franconeri et al., 2009; He et al., 2009). There does not seem to be a perceptual limit to the effect of spatial connectivity. It could be that similarity grouping also acts similarly and its effect would increase with the number of colours used to segregate the objects. However, there is one difference in the way grouping by connectivity and by colour are implemented. Increasing the number of connections increases the number of objects that participate in grouping, whereas increasing the number of colours used to group objects does not (at least if grouping is implemented as here and in previous studies). All objects participate in grouping by similarity, only the number of groups changes. If the number of objects participating in similarity grouping is the relevant factor, then the effect of changing the number of perceptual groups would not affect the amount of underestimation.

5.2. Methods

5.2.1. Participants

The same 17 participants who took part in Experiment 3 completed Experiment 4 in the second half of the same session. One participant did not complete two blocks due to a technical issue.

5.2.2. Materials, stimuli, and procedure

The material, stimuli and procedure were the same as in Experiment 3, except for the following differences. The reference number of circles was increased from 24 to 48 to allow us to test a large range (2–8) of perceptual groups, with at least 4 circles per group. These circles were presented within an imaginary circle of diameter 16 deg. Initially, ten locations, separated by at least 15 angular degrees, were randomly chosen on the circumference of this imaginary circle. We then iteratively introduced one location at a time within the imaginary circle, with the constraint that no new location was within 1.125 deg. of its nearest neighbour.

We used nine easily distinguishable colours for painting the circles (see Table 3). Making all colours equiluminant would make some of them hard to distinguish, which would impair perceptual segregation. Hence, we created three sets of three equiluminant colours. This approach would mitigate, to some extent, any effects of luminance differences.

In seven conditions, we tested underestimation in 2 to 8 perceptual groups, in separate blocks. The order of blocks was randomised. We applied k-means clustering to the reference patch to specify the segregated regions with a minimum of four circles in each region. The circles within each region were assigned one of the chosen colours. The test patch circles were assigned one of the colours that was not present in the reference patch.

The data analysis was the same as in previous experiments, except that we removed PSE estimates where bootstrapped standard deviation exceeded 8, instead of 2 as in previous experiments, because of the higher reference numerosity (48) used here. We excluded one participant, since we could obtain only two reliable thresholds from their data (out of 7 conditions). Among the remaining participants, we excluded 15 thresholds (13.6% of data), which is a relatively larger number of thresholds. However, the more traditional statistics (see Supplementary Data Analysis), where we did not exclude any data, also produced similar outcomes.

5.3. Results

PSEs and the corresponding confidence intervals and statistics for each condition are presented in Fig. 6 and Table 5. JNDs did not differ across the grouping conditions (Supplementary data analysis).

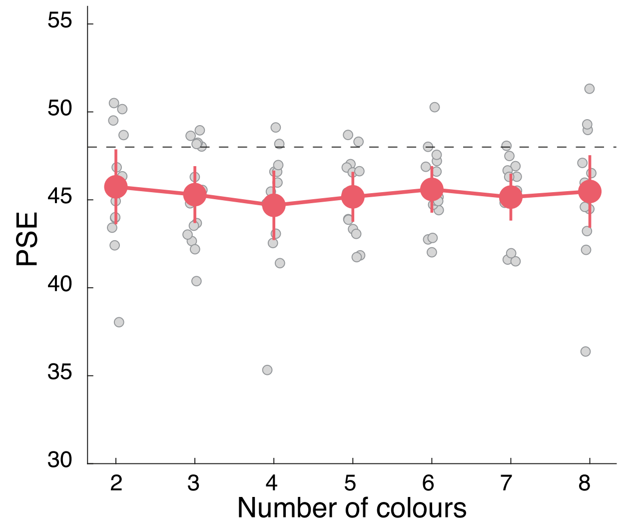

Fig. 6. Results from Experiment 4. PSEs are plotted as a function of the number of colours used to group objects in the reference patch.

We once again found that grouping by similarity led to underestimation, this time with a different reference numerosity (48). Thus, underestimation in the previous experiments could not have been due to the choice of numerosity. More importantly, underestimation was comparable across all tested numbers (from 2 to 8) of colours used to group the objects ($F(1,78.8) = 0.07$, $p = .79$, ${}_{p}\eta^2 = 0.008$). This finding suggests that segregating the objects into subitizable number of clusters does not modulate underestimation, thus limiting the role of attention in similarity grouping related underestimation.

6. General discussion

This study sought to determine the fundamental units feeding into the numerosity estimation mechanisms. We found that grouping by similarity in colour or shape as well as by spatial cues led to underestimation, indicating that estimation mechanisms are sensitive to all the Gestalt principles of grouping tested here. These effects were solely due to similarity grouping as ungrouped stimuli did not lead to underestimation. We also found that the two kinds of grouping cues (spatial and appearance-based) can independently and additively contribute to underestimation. These results contrast with the predictions of a purely topological account of numerosity processing. Instead, the findings argue that Gestalt-principle driven grouping leads to the formation of bound units that are processed by estimation mechanisms, and hence perceived as being less numerous than they are.

6.1. The relationship between spatial and similarity grouping

Our results suggest that the effects of spatial grouping (e.g., closure) and colour similarity on estimation are additive. This is in line with previous findings where, over a large range of stimulus manipulations, the effects of proximity (a spatial grouping principle) and similarity grouping were additive (Kubovy & van den Berg, 2008; Wagemans et al., 2012). There seem to be further differences in how the two forms of grouping affect estimation. For example, it has been reported that increasing the number of objects that are grouped, say by connections between them, leads to increased underestimation (Franconeri et al., 2009; He et al., 2009). However, this is not the case with similarity grouping ('yellow' conditions in Experiment 1 in our study; He et al., 2015). On a similar note, increasing the number of perceptual groups created by similarity grouping did not lead to a change in the amount of underestimation (Experiment 4). It is currently unknown, however,

Table 5
PSEs for conditions tested in Experiment 4. Also reported are the corresponding confidence intervals (CI), results of *t*-tests conducted against 48 with and without corrections for multiple comparisons using the false discovery rate (FDR) approach and the effect size.

Number of colours	PSE	95% CI	t-statistic	p-value	FDR corrected p-value	Estimated standardized effect size (*d*)
2	45.9	44.4–47.5	−2.5	.015	.015	−0.67
3	45.1	43.6–46.6	−3.6	.0006	.001	−0.95
4	44.5	43.0–46.0	−4.4	<.0001	.0002	−1.16
5	45.0	43.5–46.5	−3.7	.0004	.001	−1
6	45.7	44.1–47.2	−2.9	.004	.006	−0.77
7	45.3	43.8–46.9	−3.3	.001	.002	−0.9
8	45.8	44.2–47.3	−2.9	.005	.007	−0.75

whether changing the number of perceptual groups through spatial grouping modulates underestimation (manipulations of spatial grouping always changed both the number of perceptual groups as well as the number of objects participating in grouping). Further, distant objects linked spatially (e.g., through a line) are underestimated, whereas similar objects need to be spatial adjacent to induce underestimation (Experiment 3), indicating that similarity grouping has spatial limitations. These considerations suggest that the two grouping mechanisms operate independently or at least differently. The output after both mechanisms have organised incoming visual input are then fed into the estimation mechanisms, leading to the observed substantial underestimation in the presence of multiple grouping cues.

6.2. Alternative explanations

The first two experiments were conducted with uncontrolled stimuli, displays, and environments. Nevertheless, we found consistent underestimation in the presence of grouping between objects. One could argue that some of these results could be driven by low-level stimulus differences across grouping conditions. Although there are reasons to believe that this might not be the case (see section 4.1), in Experiment 3, we assessed the role of these potential confounds by testing participants in a controlled setup and used equiluminant colours to group objects. We found underestimation with this setup, and interestingly, of the same magnitude as in the other two experiments. This indicates that possible low-level confounds could not explain our results. We also introduced a condition where the coloured circles were intermixed in a way to prevent grouping. This condition would share the low-level properties with the grouped objects, except for the grouping variable. There was no underestimation in this condition, pointing to grouping by similarity as the sole explanation of underestimation.

The robustness of the underestimation effect despite differences in low-level properties has also been shown for spatial grouping. For example, Adriano and colleagues (Adriano, Girelli, & Rinaldi, 2022; Adriano, Rinaldi, & Girelli, 2022a,b) used a Kanizsa-like illusion to connect pairs of objects and found that underestimation increased with the number of such connections. Importantly, objects were underestimated both when low-level properties were fully controlled for and when they introduced variations in low-level properties such as contrast polarity and convex hull. These findings suggest that grouping is a fundamental step before numerosity processing and low-level confounds are unlikely to explain the effects of grouping.

Experiment 3 also rules out a potential 'attentional prioritisation' explanation of our results. It could be that since the most common colour in Experiments 1 and 2 was black, participants learnt to prioritise this colour, which would have led to underestimation of the less common colours. Note that, if this were the case, we should have observed underestimation in the 'yellow' conditions of Experiment 1, as well as a difference in the underestimation between the two conditions where we added spatial and similarity grouping (embedded-closure versus coloured-closure). In Experiment 3, we compared, for the same reference patch, a condition in which the test objects changed colour on a trial-by-trial basis such that no colour could be prioritised (4-colours-colourTest) with a condition where the test objects were always grey (4-colour-greyTest). We found the same underestimation in these conditions, indicating that prioritisation of a specific colour could not have explained the results of Experiments 1 and 2.

The underestimation due to similarity-based grouping cues observed in our study is in contrast with findings of earlier studies (He et al., 2015; Yu et al., 2019). We believe that this is because, in these earlier studies, the similarity cues were weaker, particularly in comparison to connectivity-based cues. For example, Yu et al. (2019) used six colours to group 24 circles and examined if numerosity was underestimated. Grouping by colour similarity could have been weak here because the objects were small and far from each other or because some of the colours were similar to each other making it difficult to segment groups. Similarly, changing the colour of only a small subset of objects, as in the study by He et al. (2015), would not have been enough to observe an effect of appearance-based grouping on estimation, as our results show (the absence of underestimation in the 'yellow' conditions).

6.3. Attention in similarity grouping induced underestimation

Experiment 4 examined whether underestimation was affected by the number of perceptual groups in the display. It is possible for example that attentional segregation processes such as subitizing play a role in grouping-based underestimation. If this were the case, underestimation should peak when the number of perceptual groups were at the subitizing limit (around four). Instead, we found substantial and a constant amount of underestimation irrespective of the number of colours used to group objects. This result argues against a role of attention in similarity-grouping induced estimation. Interestingly, Yu et al. (2019) argued, on the basis of not detecting any underestimation when objects were grouped by similarity, that attention can group objects by only one similarity cue at a time. That is, similarity-based grouping requires attention and occurs sequentially across cues (e.g., one colour at a time). In contrast, spatial grouping occurs automatically, does not require attention, and occurs simultaneously at all locations and participating objects. Our results contradict this conclusion by showing that similarity-based grouping also leads to underestimation and importantly that increasing the number of perceptual groups do not reduce underestimation. These results imply that similarity-based grouping, just like spatial grouping, does not require attention, can occur simultaneously across locations, and is likely automatic.

6.4. Mechanisms of underestimation due to similarity grouping

Our view implicitly argues for rapid, automatic, perhaps feedforward, *base* grouping of elements detected by the visual system without the need to allocate attention (Roelfsema, 2006; Roelfsema & Houtkamp, 2011), before estimation occurs. There are other situations where grouping requires effort and attention, and it is likely that such *incremental* grouping (Roelfsema & Houtkamp, 2011) does not lead to underestimation. We believe that estimation mechanisms receive their input after an initial segmentation process directed by base Gestalt principles. This is supported by the finding that the underestimation induced by (similarity) grouping does not seem to diminish with increasing number of segmented regions (up to 8 colours; Experiment 4).

Note that this argument does not *require* that Gestalt grouping is feed-forward. However, recent evidence supports this conceptualisation of the processing pipeline. Fornaciai and Park (2018, 2021) tested participants with displays containing the same number of objects that were either connected (and therefore perceived as less numerous) or unconnected to each other. They observed that neural responses to these two displays diverged relatively early in the processing stage (150 ms in EEG; V3 in fMRI), although not earlier. They further found that the segmentation of the objects by connectedness was immune to interruptions of feedback signals. These results were taken to argue that the stimuli are first registered in the early visual cortex, followed by a feedforward segmentation of the objects by spatial grouping. Only then are the spatially segmented objects estimated. Our findings are consistent with this proposal and extend the argument to similarity-based grouping.

What might explain the clear underestimation induced by similarity grouping that we observed in this study? The processes underlying underestimation have been studied primarily by manipulating inter-object proximity, which is well-known to cause underestimation (Allik & Tuulmets, 1991). Several mechanisms have been proposed to account for this finding, all of which can be extended to the effect of similarity grouping on estimation. It was posited that proximity-induced underestimation might be caused by mutual interference between nearby objects, an effect called visual crowding (Valsecchi, Toscani, & Gegenfurtner, 2013). However, recently, doubt has been cast on this hypothesis by a study that found that crowding and estimation displayed dissociable characteristics when stimulus properties were varied indicating that the underlying mechanisms are likely to be different (Chakravarthi & Bertamini, 2020). The underestimation observed in the periphery and due to proximity was instead attributed to spatial grouping. An explanation along these lines is that proximity leads to fewer perceived clusters in the display, which causes the visual system to underestimate the total number of objects (Im et al., 2016). Accordingly, similarity grouping cues can lead to fewer perceived clusters leading to underestimation. A different argument contends that estimation depends on the total area under the influence of individual objects. If so, closely spaced objects have overlapping regions of influence, which leads to a reduced overall area of influence across all objects, and hence underestimation (Allik & Tuulmets, 1991). An updated version of this ‘occupancy model’ posits that the probability of registering any individual object is binomially distributed and that this probability is modulated by proximity (Allik & Raidvee, 2021). Similarly, the probability of registering a given object might be modulated by Gestalt principles other than proximity. In all cases, grouping plays a central role in shaping the input to estimation mechanisms.

6.4.1. Neural mechanisms

Recent neuroimaging and computational work suggest an intriguing possibility about how similarity grouping can lead to underestimation. It has been shown that, surprisingly, the early visual cortex tracks numerosity irrespective of size and the spacing between objects (DeWind et al., 2019; Park, DeWind, Woldorff, & Brannon, 2016; Paul et al., 2022). Such numerosity tracking ‘neurons’ have been observed to spontaneously develop in convolutional neural networks, whether trained for object recognition or untrained (Kim, Jang, Baek, Song, & Paik, 2021; Nasr, Viswanathan, & Nieder, 2019). It has been argued that neurons in the early visual cortex can track numerosity because they represent the contrast energy of the stimulus (Park & Huber, 2022; Paul et al., 2022). This energy computation could potentially be implemented through centre-surround contrast (Gabor) filters summed across spatial scales or the aggregate Fourier spectral power (sum of power of orientation filters at multiple spatial scales). These two versions are variants of the same process. Early visual neurons in animals and machines are well suited to perform these computations. However, this spectral analysis is not sufficient to account for perceived numerosity, for example the underestimation induced by connecting objects (Fornaciai & Park, 2018). A further step of divisive (Park & Huber, 2022) or contrast (Paul et al., 2022) normalisation with an additional nonlinear step is necessary to account for human behaviour. If this is the right description of the neural representation of numerosity, a similar step of normalisation might be needed to account for our results. The precise mechanisms of how this is implemented remains to be investigated. However, this idea should be considered with caution for a few reasons. First, White, Rolfs, and Carrasco (2015) argued that divisive normalisation occurs after both spatial and feature based attention have increased responses independently (that is, both sets of factors compete for resources). Hence, the effects of these two grouping factors might not be additive, unlike what we found. Second, underestimation was observed for grouped but not ungrouped sets of objects. This suggests that feature-based normalisation should be spatially bound, unlike feature-based attention (Martinez-Trujillo & Treue, 2004; Maunsell, 2015). Thus, feature based normalisation must act on top of spatial normalisation (or be a subset of it), not independent of it. Third, as acknowledged by Park and Huber (2022), the amount of underestimation (due to, say, connectedness) explained by divisive normalisation is smaller than that observed in human observers. They argued that more complex filters might be needed to account for the full effect. Hence, at best, this model needs further development and testing to be persuasive.

6.5. Relation to other phenomena

We note that our results are not related to a seemingly similar phenomenon known as groupitizing (Anobile, Castaldi, Moscoso, Burr, & Arrighi, 2020; Ciccione & Dehaene, 2020; Maldonado Moscoso, Castaldi, Burr, Arrighi, & Anobile, 2020; Starkey & McCandliss, 2014; Wege, Trezise, & Inglis, 2022). Groupitizing refers to improved performance in enumerating objects that are grouped into several chunks, when the number of chunks is lower than the subitizing limit (about 4) and the number of objects within each chunk is less than the subitizing limit. Both similarity and spatial grouping lead to groupitizing. However, groupitizing is a temporally extended process, where participants are asked to report a specific numerosity. Importantly, it relies on subitizing and mathematical operations (e.g., adding the subitized number of objects from each chunk or using multiplication appropriately). In contrast, estimation is visual, extremely fast, does not require specifying the exact numerosity, and does not rely on mathematical manipulations. Interestingly, neither the number of objects in a cluster nor the number of clusters need to be within the subitizing range for grouping to cause underestimation, as was demonstrated by our study (Experiment 4). Groupitizing, on the other hand, is only possible for a small number of clusters with a small number of objects within each cluster. Finally, and importantly, grouping leads to *underestimation* (i.e., inaccurate reporting), as observed in this study and elsewhere, whereas the power of groupitizing lies in *improving* performance by leveraging subitizing and basic arithmetic. Thus, even though the stimulus setup in both groupitizing and grouping-induced underestimation might seem related, we consider them to be distinct processes operating under different circumstances.

7. Conclusion

In summary, we found that both spatial and similarity grouping lead to underestimation. In addition, we showed that their effects are additive implying independent mechanisms. The output after both spatial and similarity grouping have organised incoming visual input is then fed into the estimation mechanisms, leading to underestimation. In conclusion, the units over which numerosity estimation mechanisms operate are the outputs of Gestalt grouping processes.

CRediT authorship contribution statement

Ramakrishna Chakravarthi: Conceptualization, Methodology, Data curation, Project administration, Resources, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Andy Nordqvist:** Methodology, Investigation, Writing – review & editing. **Marlene Poncet:** Methodology, Conceptualization, Writing – review & editing. **Nika Adamian:** Methodology, Investigation, Formal analysis, Visualization, Writing – review & editing.

Data availability

The data are available at OSF : https://osf.io/wmxqn/

Acknowledgements

We are grateful to two anonymous reviewers and Michele Fornaciai for suggesting alternative explanations (low-level features, attentional prioritisation effect) and ideas about mechanisms (divisive normalisation).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105565.

References

Adam, K. C. S., & Serences, J. T. (2021). History modulates early sensory processing of salient distractors. *The Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.3099-20.2021

Adriano, A., Girelli, L., & Rinaldi, L. (2022). Number is not just an illusion: Discrete numerosity is encoded independently from perceived size. *Psychonomic Bulletin & Review, 29*(1), 123–133. https://doi.org/10.3758/s13423-021-01979-w

Adriano, A., Rinaldi, L., & Girelli, L. (2022a). Nonsymbolic numerosity in sets with illusory-contours exploits a context-sensitive, but contrast-insensitive, visual boundary formation process. *Attention, Perception, & Psychophysics, 84*(1), 205–220. https://doi.org/10.3758/s13414-021-02378-y

Adriano, A., Rinaldi, L., & Girelli, L. (2022b). Spatial frequency equalization does not prevent spatial–numerical associations. *Psychonomic Bulletin & Review, 29*(4), 1492–1502. https://doi.org/10.3758/s13423-022-02060-w

Agrillo, C., Parrish, A. E., & Beran, M. J. (2016). How illusory is the solitaire illusion? Assessing the degree of misperception of Numerosity in adult humans. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01663

Allik, J., & Raidvee, A. (2021). Proximity model of perceived numerosity. *Attention, Perception, & Psychophysics, 83*(5), 2061–2070. https://doi.org/10.3758/s13414-021-02252-x

Allik, J., & Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Perception & Psychophysics, 49*(4), 303–314. https://doi.org/10.3758/BF03205986

Anobile, G., Castaldi, E., Moscoso, P. A. M., Burr, D. C., & Arrighi, R. (2020). "Groupitizing": A strategy for numerosity estimation. *Scientific Reports, 10*(1), Article 1. https://doi.org/10.1038/s41598-020-68111-1

Anobile, G., Cicchini, G. M., & Burr, D. (2013). Separate mechanisms for perception of numerosity and density. *Psychological Science, 25*(1), 265–270. https://doi.org/10.1177/0956797613501520

Anobile, G., Cicchini, G. M., Pomè, A., & Burr, D. (2017). Connecting visual objects reduces perceived numerosity and density for sparse but not dense patterns. *Journal of Numerical Cognition, 3*(2), 133–146. https://doi.org/10.5964/jnc.v3i2.38

Anobile, G., Guerrini, G., Burr, D. C., Monti, M., Del Lucchese, B., & Cicchini, G. M. (2019). Spontaneous perception of numerosity in pre-school children. *Proceedings of the Royal Society B: Biological Sciences, 286*(1906), 20191245. https://doi.org/10.1098/rspb.2019.1245

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). *Lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2014*.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B: Methodological, 57*(1), 289–300.

Bertamini, M., Guest, M., Contemori, G., & Zito, M. (2023). What the solitaire illusion tells us about perception of numerosity. *British Journal of Psychology, 114*(2), 393–414. https://doi.org/10.1111/bjop.12627

Bertamini, M., Zito, M., Scott-Samuel, N. E., & Hulleman, J. (2016). Spatial clustering and its effect on perceived clustering, numerosity, and dispersion. *Attention, Perception, & Psychophysics, 78*(5), 1460–1471. https://doi.org/10.3758/s13414-016-1100-0

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

Chakravarthi, R., & Bertamini, M. (2020). Clustering leads to underestimation of numerosity, but crowding is not the cause. *Cognition, 198*, Article 104195. https://doi.org/10.1016/j.cognition.2020.104195

Chakravarthi, R., & Herbert, A. (2019). Two's company, three'sa crowd: Individuation is necessary for object recognition. *Cognition, 184*, 69–82.

Cheyette, S. J., & Piantadosi, S. T. (2019). A primarily serial, foveal accumulator underlies approximate numerical estimation. *Proceedings of the National Academy of Sciences, 116*(36), 17729–17734. https://doi.org/10.1073/pnas.1819956116

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

Cicchini, G. M., Anobile, G., & Burr, D. C. (2016). Spontaneous perception of numerosity in humans. *Nature Communications, 7*(1). https://doi.org/10.1038/ncomms12536. Article 1.

Ciccione, L., & Dehaene, S. (2020). Grouping mechanisms in numerosity perception. *Open Mind, 4*, 102–118.

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, 108*(49), 19552–19557. https://doi.org/10.1073/pnas.1113195108

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

DeWind, N. K., Bonner, M. F., & Brannon, E. M. (2020). Similarly oriented objects appear more numerous. *Journal of Vision, 20*(4), 4.

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

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

Fornaciai, M., & Park, J. (2021). Disentangling feedforward versus feedback processing in numerosity representation. *Cortex, 135*, 255–267. https://doi.org/10.1016/j.cortex.2020.11.013

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

Frith, C. D., & Frith, U. (1972). The solitaire illusion: An illusion of numerosity. *Perception & Psychophysics, 11*(6), 409–410. https://doi.org/10.3758/BF03206279

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

Gebuis, T., & Reynvoet, B. (2012b). The role of visual information in Numerosity estimation. *PLoS One, 7*(5), Article e37426. https://doi.org/10.1371/journal.pone.0037426

Geng, J. J., Won, B.-Y., & Carlisle, N. B. (2019). Distractor ignoring: Strategies, learning, and passive filtering. *Current Directions in Psychological Science, 28*(6), 600–606. https://doi.org/10.1177/0963721419867099

Ginsburg, N., & Nicholls, A. (1988). Perceived Numerosity as a function of item size. *Perceptual and Motor Skills, 67*(2), 656–658. https://doi.org/10.2466/pms.1988.67.2.656

Gordon, J., & Abramov, I. (1977). Color vision in the peripheral retina. II. *Hue and saturation**. *JOSA, 67*(2), 202–207. https://doi.org/10.1364/JOSA.67.000202

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science, 17*(7), 572–576.

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

He, L., Zhou, K., Zhou, T., He, S., & Chen, L. (2015). Topology-defined units in numerosity perception. *Proceedings of the National Academy of Sciences, 112*(41), E5647–E5655. https://doi.org/10.1073/pnas.1512408112

Im, H. Y., Zhong, S., & Halberda, J. (2016). Grouping by proximity and the visual impression of approximate number in random dot arrays. *Vision Research, 126*, 291–307. https://doi.org/10.1016/j.visres.2015.08.013

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology, 62*(4), 498. https://doi.org/10.2307/1418556

Kim, G., Jang, J., Baek, S., Song, M., & Paik, S.-B. (2021). Visual number sense in untrained deep neural networks. *Science Advances, 7*(1). https://doi.org/10.1126/sciadv.abd6127. eabd6127.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception, 36*(14), 1–16.

Kovács, I., & Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Sciences, 90*(16), 7495–7497.

Kubovy, M., & van den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review, 115*(1), 131–154. https://doi.org/10.1037/0033-295X.115.1.131

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). *lmerTest Package: Tests in Linear Mixed Effects Models (lmer objects of lme4 package)*.

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2016). From 'sense of number' to 'sense of magnitude'–the role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences*, 1–62.

Lourenco, S. F., & Aulet, L. S. (2022). A theory of perceptual number encoding. *Psychological Review*. https://doi.org/10.1037/rev0000380. No Pagination Specified-No Pagination Specified.

Maldonado Moscoso, P. A., Castaldi, E., Burr, D. C., Arrighi, R., & Anobile, G. (2020). Grouping strategies in number estimation extend the subitizing range. *Scientific Reports, 10*(1), Article 1. https://doi.org/10.1038/s41598-020-71871-5

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology, 14*(9), 744–751.

Maunsell, J. H. R. (2015). Neuronal mechanisms of visual attention. *Annual Review of Vision Science, 1*(1), 373–391. https://doi.org/10.1146/annurev-vision-082114-035431

Mazza, V., Pagano, S., & Caramazza, A. (2013). Multiple object individuation and exact enumeration. *Journal of Cognitive Neuroscience, 25*(5), 697–705. https://doi.org/10.1162/jocn_a_00349

Morel, P. (2018). Gramm: Grammar of graphics plotting in Matlab. *Journal of Open Source Software, 3*(23), 568. https://doi.org/10.21105/joss.00568

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). https://doi.org/10.1126/sciadv.aav7903. eaav7903.

Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence. *Nature Reviews Neuroscience, 6*(3), 177–190. https://doi.org/10.1038/nrn1626

Olivers, C. N., & Watson, D. G. (2008). Subitizing requires attention. *Visual Cognition, 16* (4), 439–462.

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

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

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

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Pomè, A., Anobile, G., Cicchini, G. M., Scabia, A., & Burr, D. C. (2019). Higher attentional costs for numerosity estimation at high densities. *Attention, Perception, & Psychophysics, 81*(8), 2604–2611. https://doi.org/10.3758/s13414-019-01831-3

Poom, L., Lindskog, M., Winman, A., & van den Berg, R. (2019). Grouping effects in numerosity perception under prolonged viewing conditions. *PLoS One, 14*(2), Article e0207502. https://doi.org/10.1371/journal.pone.0207502

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience, 29*, 203–227.

Roelfsema, P. R., & Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Attention, Perception, & Psychophysics, 73*(8), 2542–2572. https://doi.org/10.3758/s13414-011-0200-0

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*(6), 110–114. https://doi.org/10.2307/3002019

Starkey, G. S., & McCandliss, B. D. (2014). The emergence of "groupitizing" in children's numerical cognition. *Journal of Experimental Child Psychology, 126*, 120–137. https://doi.org/10.1016/j.jecp.2014.03.006

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences, 110*(45), 18116–18120. https://doi.org/10.1073/pnas.1302751110

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. https://doi.org/10.3758/APP.72.7.1839

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance, 8* (2), 194–214. https://doi.org/10.1037/0096-1523.8.2.194

Valsecchi, M., Toscani, M., & Gegenfurtner, K. R. (2013). Perceived numerosity is reduced in peripheral vision. *Journal of Vision, 13*(13), 7. https://doi.org/10.1167/13.13.7

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin, 138*(6), 1172.

Watson, D. G., Maylor, E. A., & Bruce, L. A. (2005). The efficiency of feature-based subitization and counting. *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1449.

Wege, T. E., Trezise, K., & Inglis, M. (2022). Finding the subitizing in groupitizing: Evidence for parallel subitizing of dots and groups in grouped arrays. *Psychonomic Bulletin & Review, 29*(2), 476–484.

White, A. L., Rolfs, M., & Carrasco, M. (2015). Stimulus competition mediates the joint effects of spatial and feature-based attention. *Journal of Vision, 15*(14), 7. https://doi.org/10.1167/15.14.7

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics, 63*(8), 1314–1329.

Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences, 13*(4), 167–174. https://doi.org/10.1016/j.tics.2009.01.008

Yu, D., Xiao, X., Bemis, D. K., & Franconeri, S. L. (2019). Similarity grouping as feature-based selection. *Psychological Science, 30*(3), 376–385. https://doi.org/10.1177/0956797618822798