ELSEVIER

COGNITION

Clustering leads to underestimation of numerosity, but crowding is not the cause[☆]

Ramakrishna Chakravarthi[a,*], Marco Bertamini[b]

[a] *School of Psychology, University of Aberdeen, Kings College, Aberdeen AB24 3FX, UK*
[b] *Department of Psychological Sciences, University of Liverpool, Liverpool, L69 7ZA, UK*

ABSTRACT

Humans have the remarkable ability to rapidly estimate the number of objects in a visual scene without relying on counting, something referred to as the number sense. It has been well documented that the more clustered the elements are, the lower their perceived numerosity is. A recent account of this observation is the *crowding hypothesis*, which posits that the perceived underestimation is driven by visual crowding: the inability to recognise objects in clutter. Crowding can impair individuation of the elements, which would explain the underestimation. Here, we tested the crowding hypothesis by assessing numerosity estimation and crowding for the same stimulus configurations in the same participants. Experiment 1 compared the two tasks when numerosity can be considered to be estimated directly by the visual system (reference patch density = 0.12 items/deg^2), while Experiment 2 used high density stimuli (density = 0.88 items/deg^2), where numerosity may be estimated indirectly. In both cases, we found that spacing and similarity between elements affected estimation and crowding tasks in markedly different ways. These results are incompatible with a crowding account of numerosity underestimation and point to separate mechanisms for object identification and number estimation, although grouping may play a moderating role in both cases.

1. Introduction

Humans and other animals can assess, at a glance, the approximate number of objects in a scene. This ability has been attributed to the approximate number system (ANS; Dehaene, 1992; Feigenson, Dehaene, & Spelke, 2004). It has been argued that the ANS responds to numerosity per se, but it is well established that numerosity judgments can also be biased by irrelevant dimensions (Gebuis & Reynvoet, 2012; Ginsburg & Nicholls, 1988; Hurewitz, Gelman, & Schnitzer, 2006). In particular, perceived numerosity is lower for configurations with clustered elements relative to configurations with regular spacing between elements (Ginsburg, 1980, 1991; Ginsburg & Goldstein, 1987). One possibility is that this underestimation is a result of crowding, the phenomenon in which nearby objects impair the identification of a target object (Bouma, 1970; Stuart & Burian, 1962). To evaluate the role of crowding in relation to numerosity, for the first time we directly compare crowding and numerosity estimation tasks.

1.1. The approximate number sense

The human ability to judge the numerosity of large sets of elements has been called the "number sense" (Dehaene, 1992). Having access to information about approximate numerosity has clear behavioural advantages, especially given that humans can do this quickly and for large numerosities where counting the items is impossible or impractical.

Research on perceived numerosity has identified different underlying mechanisms. When the number is low (below 5) this process is precise and the mechanism, called subitising, allows tracking of individual objects (Kaufman, Lord, Reese, & Volkmann, 1949). Subitising is not a process of estimation and will not be discussed further in this paper. For larger numbers the individual relies on the ANS, which is error-prone. In this case, the ability to discriminate between two numerosities depends on the ratio between the two (Feigenson et al., 2004).

There is an ongoing debate about the characteristics of the ANS (e.g. Burr, Anobile, & Arrighi, 2018; Burr & Ross, 2008; Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011; Durgin, 2008). One recent development has been the proposal that two distinct mechanisms might be involved. First, the visual system can directly compute numerosity when the number of elements in the scene is not too high or the elements are sparse (Anobile, Cicchini, & Burr, 2013). However, larger and denser numerosities are estimated by the system using indirect methods that rely on texture properties and density estimation (Anobile et al., 2013; Dakin et al., 2011; Durgin, 1995). For example, Anobile et al. (2013) found that when density was below 0.25 dots/deg^2 sensitivity to numerosity followed Weber's law indicating reliance on a direct estimation process, but varied with the square root of density for higher densities suggesting the utilisation of density dependent mechanisms.

In many known cases numerosity judgments are biased by irrelevant

Fig. 1. In the random-regularity illusion the regular spacing of the elements (left) leads to a larger perceived numerosity compared to a random configuration (right). These stimuli are redrawn based on the example in Ginsburg (1980).

dimensions such as regularity, object size, and area of the overall configuration (Ginsburg & Nicholls, 1988; Hurewitz et al., 2006; Tibber, Greenwood, & Dakin, 2012). An ideal number system should be able to abstract from all these dimensions by means of a process of normalisation. The fact that these biases exist has led to discussions regarding the nature of ANS (Gebuis, Cohen Kadosh, & Gevers, 2016; Harvey & Dumoulin, 2017). For example, some argue that the interference from such irrelevant dimensions arise at later stages (Harvey & Dumoulin, 2017), whereas others suggest that irrelevant dimensions are processed and lead to interference at early stages (Balas, 2016; Gebuis et al., 2016).

One interesting finding, which has been observed right from the earliest work on perceived numerosity, is that regularly spaced configurations appear more numerous than irregular and clustered configurations. Ginsburg called this the random-regularity numerosity illusion (Fig. 1; Ginsburg, 1980, 1991; Ginsburg & Goldstein, 1987). Even when the area occupied by the elements is the same, the perceived number of elements decreases with decreasing inter-object spacing (Bertamini, Zito, Scott-Samuel, & Hulleman, 2016; Valsecchi, Toscani, & Gegenfurtner, 2013). That is, clustering leads to underestimation. The random-regularity numerosity illusion led to the development of the *occupancy* model. In this model, each element has a region of influence, and these regions can overlap. The total region under the influence of the elements is used to compute numerosity (Allïk & Tuulmets, 1991; Vos, Van Oeffelen, Tibosch, & Allik, 1988). When elements are close to each other, the regions overlap and hence their contribution is reduced. This leads to underestimation.

A final interesting aspect is that, even when distance is kept constant (and therefore occupancy does not change), simple manipulations of distant elements, such as linking them by lines, leads to underestimation (Franconeri, Bemis, & Alvarez, 2009; He, Zhou, Zhou, He, & Chen, 2015; Yu, Xiao, Bemis, & Franconeri, 2019). This and similar findings led to the proposal that grouping between clustered elements is the cause for underestimation. Indeed, there is evidence supporting the idea that various forms of grouping between elements can lead to underestimation (Anobile, Cicchini, Pomè, & Burr, 2017; Franconeri et al., 2009; He et al., 2015).

1.2. Crowding and numerosity

Recently, it has been posited that the link between clustering and underestimation can be explained by visual crowding (Valsecchi et al.,

2013), at least at high element densities (Anobile et al., 2017). Visual crowding is the phenomenon where the identification of a stimulus is affected by nearby flankers (Bouma, 1970). Crowding is sensitive to the spacing between stimuli - the closer the stimuli are, the stronger the crowding between them. Further, crowding scales with eccentricity; that is, for a given inter-stimulus spacing, crowding is stronger farther in the periphery. Crowding has been argued to work through texture formation of closely dispersed elements (Balas, Nakano, & Rosenholtz, 2009; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Hence, crowding is thought to prevent individuation of elements (Intriligator & Cavanagh, 2001), leading to underestimation. Valsecchi et al. (2013) found that perceived numerosity decreases with a decrease in inter-element distance and with an increase in eccentricity. This has been taken to support the *crowding hypothesis*. Further, the dependency of estimation on eccentricity is modulated by centre-to-centre inter-stimulus distance and is independent of size (Anobile, Turi, Cicchini, & Burr, 2015), which accords well with the crowding hypothesis, since crowding is also independent of size (Pelli, Palomares, & Majaj, 2004; Tripathy & Cavanagh, 2002), but is sensitive to centre-to-to centre spacing (compared to edge-to-edge spacing; Pelli et al., 2004; Rosen, Chakravarthi, & Pelli, 2014). Additionally, under crowded conditions, observers often fail to report the presence of a subset of target elements (Sayim & Wagemans, 2017), lending support to the idea that crowding can lead to underestimation.

On the other hand, results of some studies speak against the *crowding hypothesis*. There is evidence that contrast polarity (or colour) of the elements does not affect estimation (e.g., Dakin et al., 2011; Tibber et al., 2012). Items of the same colour (say all black) are perceived to have about the same numerosity as those with mixed colour (say half white and half black). However, this has not been convincingly demonstrated. Some of the results that supported this conclusion were based on methods not intended to directly compare estimation of mixed versus uniform elements. For example, Tibber et al. (2012) did not fix the polarity of the *reference* patch in a 2IFC task while modulating the polarity condition (uniform or mixed) of the *test* patch, but randomly assigned polarity conditions to either the test or the reference patch on each trial. This makes it difficult to determine if estimation is worse in the uniform polarity condition or in the mixed polarity condition.

Crowding, instead, is substantially affected by contrast polarity (Chakravarthi & Cavanagh, 2007; Kooi, Toet, Tripathy, & Levi, 1994). In general, targets that are dissimilar to the flankers in one or more

feature dimensions (colour, depth, polarity, motion, spatial frequency, complexity, etc.) are less susceptible to crowding. For example, red targets are crowded more by red flankers than by green flankers (Kennedy & Whitaker, 2010). According to the crowding hypothesis stimulus characteristics that reduce crowding should also reduce underestimation. When this is not observed, this discrepancy casts doubt on the crowding hypothesis.

Nevertheless, it is possible that dissimilarity among elements does not reduce crowding when the number of elements is large, as is typical in estimation studies. For example, it has been shown that when elements with two different colours alternate with each other, forming a pattern, crowding is not reduced (Manassi, Sayim, & Herzog, 2012; Rosen & Pelli, 2015). Similarly, when large numbers of elements are present, they might form a pattern or texture, irrespective of the similarity between adjacent elements, leading to strong crowding. However, a texturisation model, which explains both crowding and peripheral numerosity judgements, was not able to explain the underestimation due to clustering (Balas, 2016), suggesting that texture-formation and crowding cannot explain the observed underestimation.

Another fundamental problem is that crowding does not impair detection of elements or objects, but substantially impairs their identification (Pelli et al., 2004; Strasburger, 2005; Strasburger, Harvey, & Rentschler, 1991). Both objectively (detection tasks) and anecdotally, participants are quite clear that they can see a target, but they are unable to say what it is. The target appears to be a high-contrast jumble (Pelli et al., 2004; Tyler & Likova, 2007). Further, although Sayim and Wagemans (2017) documented that observers fail to report the presence of target elements in some trials, observers also report the presence of non-existent elements in others. Interestingly, the observers never seem to report the absence of entire objects, which goes against what the crowding hypothesis would predict. Additionally, Greenwood, Bex, and Dakin (2010) reported that flankers can induce an observer to perceive an object where there is none. They found that when four oriented Gabors surrounded a blank space, participants experienced an orientation after-effect at the location of the target, and failed at a change-blindness task where the display alternated with an actual oriented target presented in that location. Hence, it appears that crowding does not impair detection, but might, sometimes, lead to the perception of extra elements. Hence it is plausible that crowding does not lead to underestimation of clustered elements

In this study, for the first time, we directly tested the crowding hypothesis by examining numerosity underestimation and crowding for the same configurations of elements.

1.3. The rationale of our study

To study the role of crowding in numerosity we designed stimuli in such a way that both numerosity and crowding tasks could be performed. Typically, stimuli in estimation tasks are dots presented within a region. However, such dots do not carry any identity information and hence cannot be tested for evidence of crowding. Therefore, we used oriented T's and thetas (Experiment 1) or Verniers and lines (Experiment 2) as elements. We asked participants to judge the number of elements, the estimation task, and to identify a specific target (at the centre of a patch of elements), the crowding task.

We manipulated two aspects of the configuration. Because of the known effect of clustering on numerosity, we manipulated the spacing between elements to determine if clustering has the same effect on both estimation and crowding tasks. Moreover, because of the known effect of similarity on crowding (Kooi et al., 1994), we used stimuli with the same colour (black *or* white) and stimuli with mixed colours (black *and* white) to determine if similarity affects crowding and estimation in the same way.

The crowding hypothesis predicts that clustering should impair estimation and crowding; further, dissimilarity between elements should alleviate crowding and hence underestimation. We tested the crowding hypothesis in two complementary experiments. In the first we adapted a standard crowding task to assess both crowding and estimation. Here the elements were presented at a low density (0.12 items/deg^2) to test the hypothesis when estimation is carried out by the visual numerosity sensing mechanism (Anobile et al., 2013). In the second experiment, we tested whether the results of the first experiment generalised to higher densities (0.88 items/deg^2). Numerosity for these stimuli may be estimated based on perceived texture/density mechanism. In this latter experiment, we also used a different stimulus and identification task to assess if the previous results were specific to the procedure used.

2. Experiment 1

We presented participants with a patch of oriented squared thetas and asked them to either estimate their numerosity (relative to a reference patch) or identify the orientation of the sole central T. We manipulated two factors in the patch. To test the effect of clustering, we varied the minimal spacing between the elements (near or far). Second, to test the effect of similarity, we varied the contrast polarity of the elements: all of them had the same colour (black or white) or roughly half of them were black and the other half were white. We assessed if these two factors affected the two tasks in the same way. We also included a standard crowding task with a target T flanked by four squared thetas (e.g., Scolari, Kohnen, Barton, & Awh, 2007; Tripathy & Cavanagh, 2002). Since there was a single T in the entire display, the participants simply had to report the orientation of this element. We tested two variants of this experiment (1A and 1B), which differed in the range of numerosities tested.

2.1. Method

2.1.1. Participants

Twenty-three observers (15 Female; Age Mean ± SD: 20.9 ± 2.2 years) participated in Experiment 1A and 20 (17 Female, Age: 23.8 ± 7.5 years) in Experiment 1B. All participants reported normal or corrected-to-normal vision and provided written informed consent. The study was approved by the Psychology Ethics committee at the University of Aberdeen. This study examines visual crowding and numerosity estimation, both of which are robust phenomena observable even in individual participants (with hundreds to thousands of trials per participant). Hence, the predictions of the crowding hypothesis should be testable even with a handful of participants. Nevertheless, we have used 2–4 times the usual sample size included in such studies.

2.1.2. Materials and stimuli

Stimuli were generated in MATLAB using Psychophysics toolbox extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007) and displayed on a 21-inch Sony CRT screen (Sony Trinitron GDM−F520, Sony, Tokyo, Japan; 100 Hz; 1024 by 728 pixels). Viewing distance was set to 57 cm by the use of a chinrest. The elements of the patch were squared thetas of size 0.8° × 0.8°, oriented vertically or horizontally. The target element at the centre of the probe patch was a T (0.8° × 0.8°), presented in one of four orientations (upright, left, inverted, right). Fig. 2 provides examples of stimuli and illustrates the procedure.

2.1.2.1. Experiment 1A. Elements were presented within a circular region (diameter 16°) centred at 12° eccentricity along the horizontal axis. In the numerosity experiment, two such patches were presented simultaneously. The *reference* patch had 24 elements (Fig. 2B, Reference, first column, top row). Numerosity of the *probe* patch varied between 16 and 32 in steps of 4.

The probe patch was constructed as follows: The T was placed at the centre of the patch. This element was surrounded by four flanking elements (squared-thetas), one in each cardinal direction (top, bottom, left, right). The remaining elements (11–27) were distributed according to an algorithm. First, six *seed* elements were placed at equidistant

A: Experimental protocol
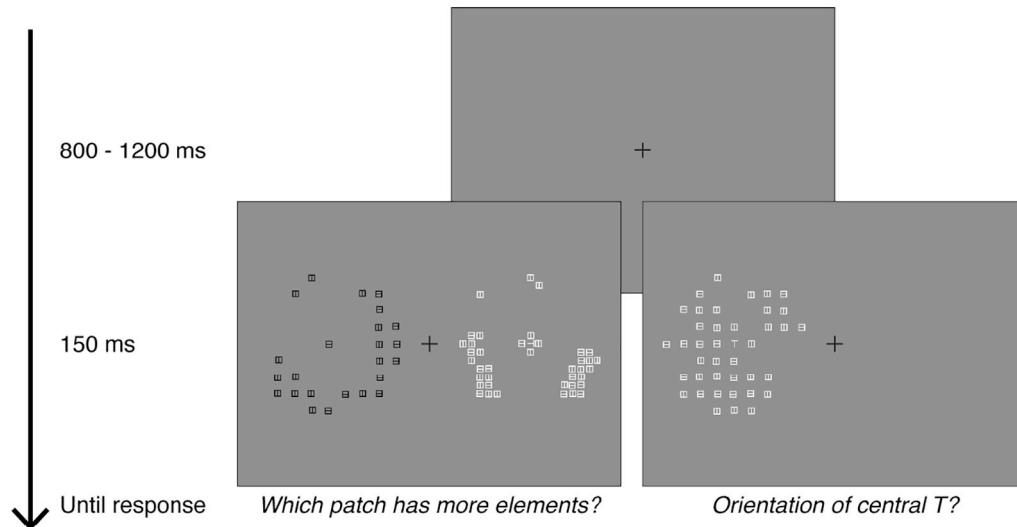
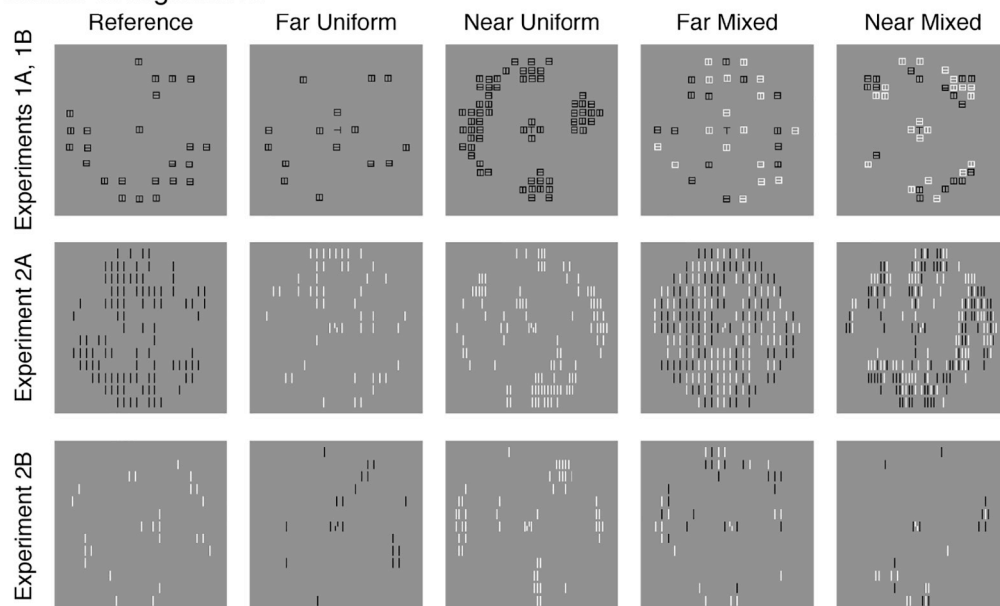## B: Stimulus configurations

Fig. 2. Experimental procedure and stimuli. A) Each trial began with fixation of variable duration (800–1200 ms). The stimuli were subsequently presented briefly (150 ms). In separate blocks, either the estimation task (left) or the identification task (right) was tested. Participants were asked to report the patch with more elements in the estimation task, and the orientation of the central T in the identification task. B) Example stimuli used in each of the four experiments. Stimuli used in Experiments 1A and 1B are depicted in the top row, Experiment 2A in the middle row and 2B in the bottom row. The leftmost column shows the reference patch used in the estimation task of the respective experiments. The other four columns illustrate the probe patches in the four test conditions.

points along the edge of the patch (8° from the centre of the patch). This ensured that the area of the patch was comparable across all numerosities and conditions. Then, each of the remaining elements was added to the patch in turn. Iteratively, each element was added by randomly choosing one of the occupied locations (except the central T and its four flanking elements). A new element was then placed at a specified distance but in a random direction from this chosen element such that a) no two elements were closer than the specified distance (hence, there was no overlap between elements) and b) no element was placed beyond the edge of the patch.

Two variables were manipulated in the construction of the probe (Fig. 2B, top row). First, the elements could be either *near* or *far*: the minimal spacing between elements was either 1° or 2°, respectively. In these conditions, the central T and its flankers were near each other (crowded) or far (less crowded), respectively. Second, all elements

could share the same contrast polarity (*uniform* polarity condition; all white or all black elements) or could be of *mixed* polarity: the central T had a different polarity relative to its four surrounding flankers (white T surrounded by black flankers, or vice versa). Each of the remaining elements was randomly assigned white or black colour. The orientation of the central T was chosen from among 0°, 90°, 180° and 270°. The orientation of each squared theta was randomly assigned between 0° and 90°. The reference patch was constructed in the same way except for the following differences: a) the minimal spacing between elements was set to 2°, and b) the contrast polarity was the same for all elements.

In the identification task, only the probe patch was presented. To examine the nature of crowding at each probe numerosity, we included conditions where only four flankers were presented. Further, a no-flanker condition was also included to test the amount of crowding in the flanked conditions.

2.1.2.2. Experiment 1B. There were two differences in Experiment 1B relative to 1A: a) the numerosities tested were 20, 24, 32, 40, 48, 64 in the near spacing condition and 16, 20, 24, 32, 40, 48 in the far spacing condition, and b) the number of seed locations at the edge of the patch was increased to eight to accommodate the larger numerosities tested. The numerosities were chosen based on the results of Experiment 1A, where we could not fit psychometric curves in the near spacing condition. Here, we used a wider range and higher numerosities to cover the full range of performance.

2.1.3. Procedure

Each participant completed both identification and estimation tasks (Fig. 2A). Each task was performed in separate blocks. The order of blocks was randomised.

2.1.3.1. Estimation. Each probe numerosity was tested on four conditions (2 spacing: near or far × 2 polarity: uniform or mixed). Observers were trained on one block of 56 practice trials before the main experiment. Each block began with a key press. Two patches appeared in the periphery for 150 ms. Participants were asked to report, via an appropriate key, which patch had more elements. No feedback was provided. The locations of the probe and reference patches were randomised. The next trial began 800–1200 ms after the response.

2.1.3.2. Crowding. The same range of numerosities used in the estimation experiment was tested. The sequence was the same, except that a single patch was presented either to the left or the right of fixation. Observers reported the orientation (4AFC) of the target T with a key press.

2.2. Results

2.2.1. Experiment 1A

Data from three participants were excluded from analysis (final $n = 20$), because the linear fits to their estimation data (see below) were poor and yielded parameter estimates that were either not meaningful (negative PSE) or far outside the tested range of numerosities (> 70 elements).

2.2.1.1. Estimation. Fig. 3A plots the proportion of cases in which the probe patch was reported to have higher numerosity than the reference patch, as a function of numerosity. As expected, this *proportion more* response increased with probe numerosity. Originally, we had intended to fit psychometric curves to these data to estimate the Point of Subjective Equality (PSE), that is, the number of elements needed in the probe to be perceptually equivalent to the reference. However, performance in the near spacing conditions did not reach halfway (0.5) of the response range (Fig. 3A, filled symbols). Since the ascending part of a typical psychometric curve is approximately linear, we fit a straight line to the *proportion more* reports as a function of numerosity. From these fits, we extracted the PSEs.

The PSEs (Fig. 3A–B) were roughly 50% higher in the *near* spacing condition (mean ± sem = 37 ± 2.2 elements) compared to the *far* spacing condition (24.8 ± 0.3; $F(1,19) = 33.1$, $p < 0.0001$, $_p\eta^2 = 0.64$). That is, numerosity was underestimated in the clustered conditions but was perceived veridically in the far spacing condition, with PSEs close to 24.

Inter-element similarity also modulated PSEs ($F(1,19) = 4.6$, $p = 0.045$, $_p\eta^2 = 0.19$). This effect was mild, with PSEs ~10% higher in the mixed polarity condition (32.7 ± 1.8) than in the same polarity condition (29.2 ± 0.8). Note that the direction of this difference was contrary to the crowding hypothesis' prediction. According to the crowding hypothesis, underestimation should be less in the mixed polarity condition than in the uniform polarity condition, since crowding should be weaker in the former. Here, unexpectedly, underestimation was higher in the mixed polarity condition than in the same polarity

condition. Finally, there was an interaction between spacing and polarity ($F(1,19) = 4.3$, $p = 0.052$, $_p\eta^2 = 0.18$). This interaction was driven by a stronger effect of polarity (again, in the wrong direction) in the near spacing condition (Mixed-Uniform PSE = 6.4 ± 3 elements) than in the far spacing condition (0.6 ± 0.4).

The strong effect of spacing and a weaker of effect of similarity on numerosity judgment can also be observed in participants' median reaction times (Supplementary Materials S2). Overall, participants were slower in the near flanker conditions than in the far flanker conditions, at larger probe numerosities. This reflects their hesitation in deciding which patch had more elements when the elements in the probe patch were clustered. Similarly, for close elements, the slowest responses were observed at a numerosity higher than the reference patch numerosity, suggesting that participants perceived a patch with a higher number of closely spaced elements as equally matched with the reference patch.
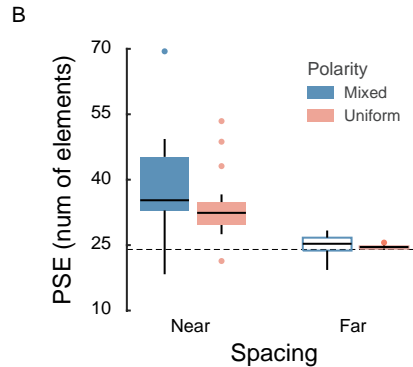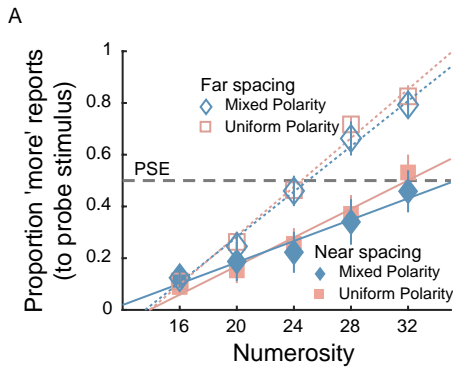
2.2.1.2. Crowding. Fig. 3C plots accuracy (proportion correct) in identifying the orientation of the target T in each of the four conditions as a function of numerosity. Fig. 3D summarises the data by collapsing across numerosities (averaged identification accuracy across all the higher numerosities (16–32), for each condition). We separately analysed identification accuracies in the presence of four-flankers (standard crowding paradigm) and many-flankers (our estimation task configuration). We also compared identification performance across numerosities and present the results in the Supplementary materials (Supp. Materials S3). Here, we note that adding flankers substantially increases crowding. That is, performance in the four-flanker condition was better than in the many flanker conditions.

In the four-flanker condition, increasing inter-element spacing increased performance by 20 percentage points (near = 0.45 ± 0.03; far = 0.65 ± 0.03; $F(1,19) = 38.6$, $p < 0.0001$, $_p\eta^2 = 0.67$). Performance in the mixed polarity condition was higher than that in the uniform polarity condition by 8 percentage points (uniform = 0.51 ± 0.03; mixed = 0.59 ± 0.03; $F(1,19) = 17$, $p = 0.001$, $_p\eta^2 = 0.47$). There was no interaction between these two factors ($F(1,19) = 1.7$, $p = 0.21$, $_p\eta^2 = 0.08$).
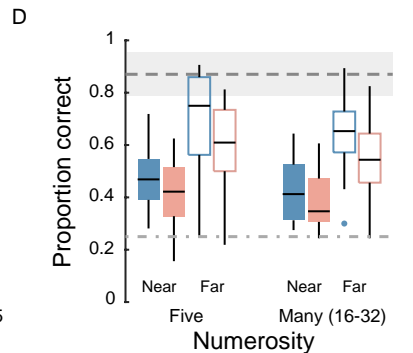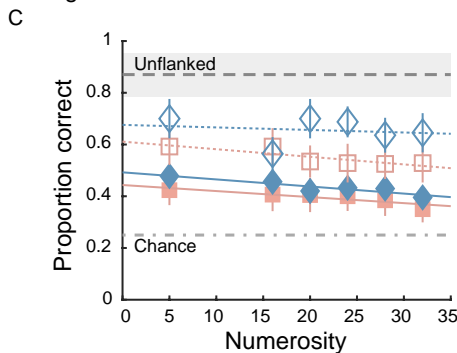
In the many-flankers condition, there was, once again, an effect of spacing with performance higher, by ~18%, in the far condition (0.59 ± 0.03) than in the near condition (0.41 ± 0.03; $F(1,19) = 34.2$, $p < 0.0001$, $_p\eta^2 = 0.64$). Performance was likewise higher in the mixed polarity (0.54 ± 0.02) condition relative to the uniform polarity condition by ~7% (0.47 ± 0.02; $F(1,19) = 57.1$, $p < 0.0001$, $_p\eta^2 = 0.75$). However, there was an interaction between these two factors ($F(1,19) = 9.2$, $p = 0.007$, $_p\eta^2 = 0.33$): the effect of polarity (similarity) was weak in the near condition (Mixed-Uniform = 0.04 ± 0.01) and stronger in the far condition (0.11 ± 0.02). This might be because, in dense arrays, flankers at the near spacing interfered with the target so strongly that polarity did not modulate performance, whereas the effect of polarity was manifest more clearly at the larger spacing. It is evident that flanker spacing and similarity affects identification performance in both the standard configuration and in stimuli with a large number of flankers. It is also interesting to note that the magnitude of the effects of spacing (~20%) and similarity (~7%) are comparable in both configurations, with and without a large number of flankers, even though the overall performance was lower in the presence of a large number of flankers (see Supplementary Materials S3 for statistical tests).

2.2.1.3. Comparison. The crowding hypothesis predicts that crowding leads to underestimation of numerosity. Although mechanisms can vary at different densities for numerosity estimation, we take the general crowding hypothesis to imply that the extent of crowding is directly related to the extent of underestimation. It follows that factors affecting crowding in one direction should affect underestimation in the same direction.
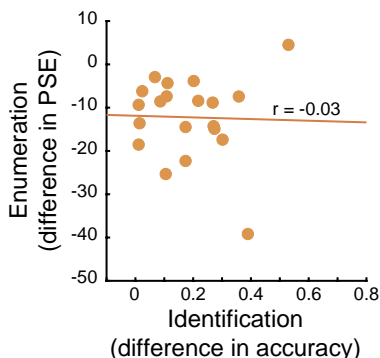
Estimation

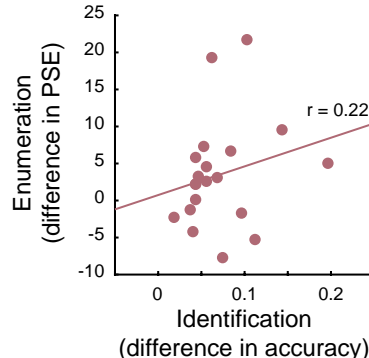**Fig. 3.** Results of Experiment 1A. A) Proportion more responses to the probe patch in the estimation task as a function of numerosity. Straight line fits are also shown. Error bars represent 95% CI. B) Box plots of PSEs estimated from the linear fits in each of the four conditions. Filled box plots represent data from the near conditions and open box plots represent data from the far conditions. C) Identification performance in the crowding task as a function of numerosity. Error bars represent 95% CI. Linear fits are also plotted. The dashed line and the grey region around it depict performance in the unflanked condition with its 95% CI. The dot-dashed line is chance performance. D) Box plots of performance in each of the four conditions in the four-flankers and the many-flanker conditions (collapsed over numerosities 16–32). E & F) Comparing performance across the two tasks, identification and enumeration. E) The effect of spacing on the two tasks for each participant. Its effect on identification is computed as the average accuracy in the far conditions minus the average accuracy in the near conditions. Its effect on enumeration is computed as the average PSE in the far conditions minus the average PSE in the near conditions. F) The effect of polarity on the two tasks. Here the differences are computed between mixed polarity and uniform polarity conditions. The best fitting straight lines along with the correlation coefficient are also plotted. Please see online for the colour version of this and subsequent figures.

We tested this prediction by comparing the effect of spacing and polarity on both identification and estimation across participants. Fig. 3E shows that the strength of crowding, indexed as the improvement in identification performance when flankers are farther apart, does not correlate with the improvement in numerosity judgment, indexed as the change in PSE when flankers are spaced farther (r (18) = −0.03, p = 0.91).

This shows that while spacing affects both tasks, it affects them in different ways. Similarly, Fig. 3F shows that there is weak to no correlation between the effects of polarity on identification and enumeration (r(18) = 0.22, p = 0.34). Even if we consider that there is a weak link between the two, the direction of this link is opposite to that predicted by the crowding hypothesis. While configurations with mixed polarity lead to better identification performance than configurations with same polarity, they lead to worse performance for numerosity estimation. That is, the PSE is higher, and hence there is more underestimation, in the mixed polarity condition. A manipulation that

reduces crowding (reducing similarity between elements) worsens estimation. It is important to note, however, that the absence of the expected correlation does not definitively demonstrate that the two tasks are based on distinct mechanisms. It is, nevertheless, additional evidence against a common mechanism operating in the two tasks.

To determine if these results are specific to the stimuli we tested, we conducted an experiment with oriented rectangles as elements and a 2AFC identification task (so that both enumeration and identification tasks have the same number of possible responses). We found nearly identical results (Supplementary Experiment 1; also see Experiment 2B below).

2.2.2. Experiment 1B

We extended the range of numerosities tested in Experiment 1B in order to fit psychometric curves to the estimation data. One participant was excluded from data analysis (final n = 19), since the psychometric curve fit to their data was poor (mean r-square averaged across four

conditions < 0.7[1]).

2.2.2.1. Estimation.
Fig. 4A plots participants' estimation performance as a function of numerosities. We fit cumulative Gaussians (Eq. (1)) to these data separately for the four conditions (2 spacing × 2 polarity conditions).

$$y = \frac{1}{2} erfc\left(\frac{-\sigma(x - \mu)}{\sqrt{2}}\right)$$

(1)

where y is proportion more reports, x is numerosity, σ is the slope of the psychometric curve, μ is the midpoint of the curve, and erfc is the complementary error function. The cumulative Gaussian spans performance from 0 (lower asymptote) to 1 (upper asymptote). The parameter μ is the estimate of the PSE.

As in Experiment 1A, spacing substantially affected numerosity judgment ($F(1,18) = 61.8$, $p < 0.0001$, $_p\eta^2 = 0.77$). PSEs (Fig. 4B) in the near condition (32.8 ± 1.4) were about a third higher than in the far condition (24 ± 0.4), where numerosity perception was veridical (PSE ~24). However, PSEs in the uniform polarity condition (28.2 ± 0.6) were comparable with that in the mixed polarity condition (28.6 ± 1.3; $F(1,18) = 0.2$, $p = 0.67$, $_p\eta^2 = 0.01$). That is, there was no effect of similarity. Further, there was no interaction between the two factors ($F(1,18) = 2$, $p = 0.17$, $_p\eta^2 = 0.1$).

2.2.2.2. Crowding.
Fig. 4C plots identification performance in the four conditions at various numerosities. To analyse this data, we averaged identification performance across all larger numerosities (16 and above) for each of the four conditions separately. We analysed averaged performance (Fig. 4D) separately in the standard crowding paradigm and in the many-flankers condition (our estimation task configuration). In the standard crowding condition, increasing inter-element spacing increased performance ($F(1,18) = 22.7$, $p < 0.001$, $_p\eta^2 = 0.56$): Accuracy in the far spacing condition (0.66 ± 0.05) was 20 percentage points higher than in the near spacing condition (0.45 ± 0.03). Similarly, performance in the mixed polarity condition (0.62 ± 0.04) was higher than performance in the uniform polarity condition (0.5 ± 0.04) by ~12% ($F(1,18) = 26$, $p < 0.0001$, $_p\eta^2 = 0.59$). There was no interaction between these two factors ($F(1,18) = 0.18$, $p = 0.68$, $_p\eta^2 = 0.01$).

In the many-flankers condition, there was, once again, an effect of spacing with performance higher (by ~27%) in the far condition (0.64 ± 0.04) than in the near condition (0.37 ± 0.02; $F(1,18) = 39.4$, $p < 0.0001$, $_p\eta^2 = 0.69$). Performance was likewise higher in the mixed polarity (0.54 ± 0.03) condition relative to the uniform polarity (0.47 ± 0.02) condition by ~7% ($F(1,18) = 25.7$, $p < 0.0001$, $_p\eta^2 = 0.59$). There was no interaction between these two factors ($F(1,18) = 1.6$, $p = 0.23$, $_p\eta^2 = 0.08$).

Overall, these results replicate those of Experiment 1A. Performance in the many flanker condition was quite similar to that in the standard crowding configuration, except that it was worse. Of interest is the finding that both spacing and similarity affected identification performance, which was not the case in the estimation task, where only spacing modulated PSEs but not similarity.

2.2.2.3. Comparison.
As in Experiment 1, we tested whether the effect of the variables we manipulated, spacing and polarity, was similar in the two tasks, identification and estimation. The crowding hypothesis predicts that any variable that strengthens (or weakens) crowding should correspondingly increase (or decrease, respectively) the underestimation of numerosity. Fig. 4E shows that there was no relationship between the effect of spacing on crowding and its effect on estimation ($r(18) = -0.02$, $p = 0.86$). This once again shows that while spacing affects both tasks, it affects them in different ways.

Similarly, Fig. 4F shows that there is a weak to no correlation between the effects of polarity on identification and enumeration ($r(18) = -0.22$, $p = 0.35$).

3. Experiment 2

Results from Experiment 1 indicated that spacing between elements affects both estimation and identification (crowding) tasks. However, the strength of this effect does not seem to correlate across the two tasks. More importantly, the similarity between elements modulates the two tasks in very different ways. Similar objects crowd each other more, whereas they either have no effect on estimation or they cause less underestimation than dissimilar objects. These findings suggest that the two tasks might have different underlying mechanisms, and that crowding does not underlie numerosity underestimation for clustered elements. One caveat for this conclusion is that we tested sparsely distributed elements (reference patch density 0.12 items/deg^2; probe patch density 0.08–0.24 items/deg^2). It has been argued that at this low density, elements are estimated directly by the visual system. At higher densities, however, the visual system has to rely on texture or density estimation mechanisms. Since the numerosity estimation mechanisms are different at different densities, it is possible that crowding between elements might lead to underestimation at higher densities, which wasn't tested in the previous experiment. Experiment 2 was designed to test this possibility. We also used this opportunity to examine if the results of the previous experiments generalise to other kinds of stimuli and identification tasks. Instead of testing orientation discrimination of a target T, we asked participants to complete a Vernier discrimination task, a protocol used successfully in several crowding studies (e.g., Manassi et al., 2012). This stimulus also has the advantage that it allows us to pack multiple elements close together to achieve higher presentation densities, unlike extended objects such as letters.

3.1. Experiments 2A and 2B

The two experiments (2A and 2B) differed only in the range of numerosities tested, and hence the density of elements within the patches. 2A tested high density, substantially above the limit where estimation can be computed directly by the visual system and 2B tested a density within the direct sensing range, and was comparable to Experiments 1A and 1B.

3.2. Method

3.2.1. Participants
Twenty-one participants (11 Female; Age Mean ± SD = 22 ± 2 years) took part in Experiment 2A and another 22 (13 Female; 23.6 ± 8.5 years) in Experiment 2B. All reported normal or corrected to normal vision and provided written informed consent.

3.2.2. Material and stimuli
The materials were the same as in Experiment 1. The stimuli were modified to allow the patches to have higher densities. The patch size was reduced to 12 deg. diameter. It was centred at 8 deg. eccentricity to the left or right of fixation on the horizontal axis. The target was a Vernier made up of a pair of vertical lines 0.4 deg. in height, misaligned by 0.16 deg. The top half was always placed in the centre of the patch. The lower one was displaced either to the left or right of the top half. The flankers were vertical lines 0.8 deg. in height. The horizontal distance between adjacent lines was a minimum of either 0.25 deg. (close spacing) or 0.5 deg. (far spacing). The vertical distance between adjacent lines was a minimum of 1 deg. The colours of the line (white or black) were chosen as in Experiment 1. Hence, there were two similarity conditions: same polarity lines and mixed polarity lines.

---

[1] R-square of the Gaussian fit was < 0.3 in at least one of the four conditions.

Estimation

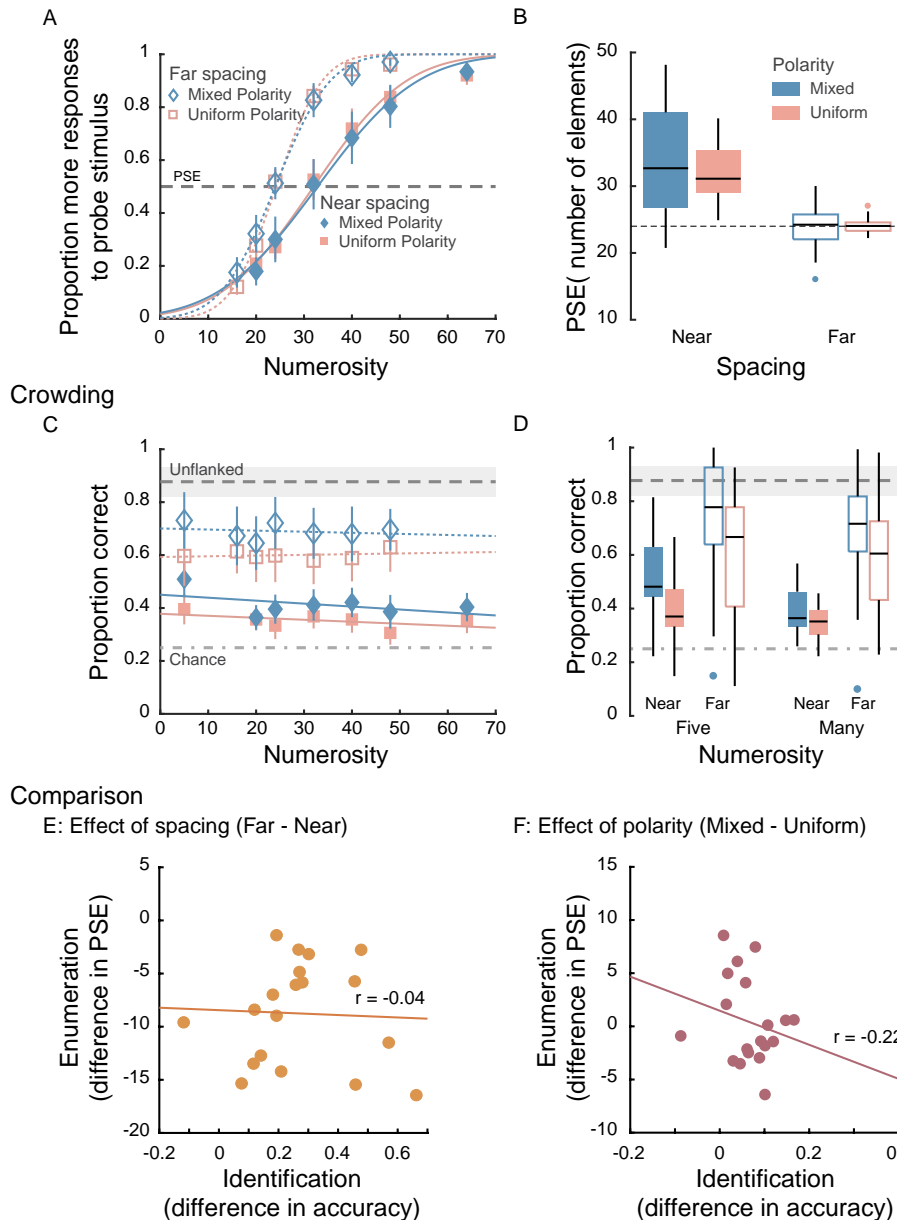## Crowding

Comparison

Fig. 4. Results of Experiment 1B. A) Proportion more responses to the probe patch in the estimation task as a function of numerosity. Cumulative Gaussian fits are also shown. Error bars represent 95% CI. B) Box plots of PSEs estimated from the fits in the numerosity task. Dashed line indicates veridical estimation (PSE = 24). C) Identification performance in the crowding task as a function of numerosity along with straight line fits. Error bars represent 95% CI. D) Box plots of performance in each of the four conditions in the four-flankers and the many-flanker conditions. The horizontal dashed line and grey shaded region in panels C and D represent accuracy of identifying an unflanked element with 95% CI. The dot-dashed line represents chance performance (0.25). E & F) Comparing performance across the two tasks, identification and enumeration. E) The effect of spacing on the two tasks for each participant. F) The effect of polarity on the two tasks. The solid lines in both panels are the best fitting straight lines.

3.2.2.1. Experiment 2A. To test the role of crowding in numerosity underestimation at higher density, we presented 100 lines in the reference patch (density = 0.88 items/deg²; roughly seven times higher than in Experiment 1; Fig. 1B, middle row). This numerosity was tested against probe numerosities of 75, 100, 125, 150, 200, and 250 in the near spacing condition, and 50, 75, 100, 125, 150, and 200 in the far spacing condition (at the highest numerosity, density was 2.2 items/deg²). To construct these patches, we placed the Vernier target at the centre and two flankers one on either side of this target. We then placed 15 lines as seeds at the edge of the patch. We also scattered 15 further lines within the patch area as seeds. The remaining lines were placed as per the algorithm described in Experiment 1A. The only difference was that each new line was displaced either horizontally or vertically from a chosen seed and not in any possible direction (0–360°) as was done in Experiment 1.

3.2.2.2. Experiment 2B. Experiment 2B differed from 2A only in the range or numerosities tested. The reference numerosity was 25 (Fig. 1B, bottom row), one fourth the density of that tested in Experiment 2A and

roughly twice that in Experiment 1 (density = 0.22 items/deg²). The probe numerosities were 20, 25, 30, 40, 50, and 60 in the near spacing condition and 15, 20, 25, 30, 40, and 50 in the far spacing condition. Here, we used 6 lines at the edge of the patch as seeds and a further 4 scattered within the patch.

3.2.3. Procedure

The procedure was the same as in Experiment 1, except that the identification task was to report if the lower half of the Vernier was displaced to the left or right of the top half.

3.3. Results

3.3.1. Experiment 2A

3.3.1.1. Estimation. We estimated the point of subject equality (PSE) for each of the four conditions by fitting cumulative Gaussians to the 'proportion more' reports to the probe stimulus data as a function of numerosity (Fig. 5A and B). Reducing the inter-element spacing led to substantial underestimation ($F(1,20) = 151.4$, $p < 0.0001$,
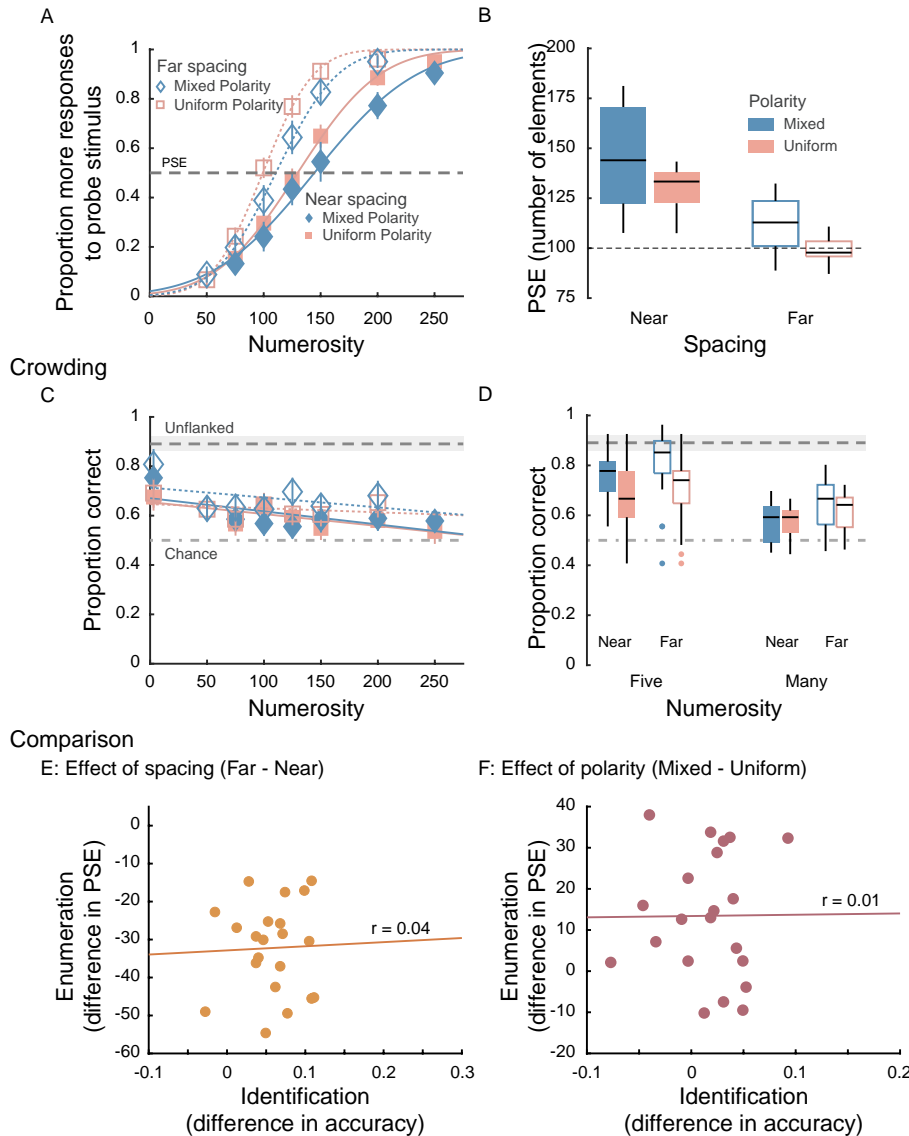
Estimation

Fig. 5. Results of Experiment 2A. A) Proportion more responses to the probe patch in the estimation task as a function of numerosity. Cumulative Gaussian fits are also shown. Error bars represent 95% CI. B) Box plots of PSEs estimated from the fits in the numerosity task. Dashed line indicates veridical estimation (PSE = 100). C) Identification performance in the crowding task as a function of numerosity along with straight line fits. Error bars represent 95% CI. D) Box plots of performance in each of the four conditions in the four-flankers and the many-flanker conditions. The horizontal dashed line and grey shaded region in panels C and D represent accuracy of identifying an unflanked element with 95% CI. The dot-dashed line represents chance performance (0.5). E & F) Comparing performance across the two tasks, identification and enumeration. E) The effect of spacing on the two tasks for each participant. F) The effect of polarity on the two tasks. The solid lines in both panels are the best fitting straight lines.

$_p\eta^2 = 0.88$). PSEs in the near spacing condition (137.5 ± 3.6) were about 40% higher than the reference numerosity (100), and about 30% higher than in the far spacing condition (105.3 ± 1.6). Similarity also modulated estimation, but once again, in a direction opposite of that predicted by the crowding hypothesis of underestimation. Dissimilar elements (128.1 ± 4) led to more underestimation than similar elements (114.7 ± 1.4; F(1,20) = 16.1, $p = 0.001$, $_p\eta^2 = 0.45$). There was no interaction between the two factors (F(1,20) = 1.1, $p = 0.3$, $_p\eta^2 = 0.05$).

3.3.1.2. Crowding. Fig. 5C plots performance in the identification task at various numerosities in each of the four conditions. Fig. 5D summarises these for the standard crowding and multiple-flanker configurations separately. In the standard crowding (two-flanker) setup, performance in the far spacing condition (0.75 ± 0.03) was, surprisingly, not statistically different from that in the near spacing condition (0.71 ± 0.02; F(1,20) = 3, $p = 0.1$, $_p\eta^2 = 0.13$), however, it was somewhat higher in the former. On the other hand, dissimilar flankers (0.78 ± 0.02) did improve performance relative to similar flankers (0.68 ± 0.02; F(1,20) = 23.2, $p = 0.0001$, $_p\eta^2 = 0.54$). There was no interaction between the two factors (F(1,20) = 1.2, $p = 0.29$, $_p\eta^2 = 0.05$).

In the presence of a large number of flankers, performance in the far spacing condition (0.63 ± 0.02) was better than in the near spacing condition (0.58 ± 0.01; F(1,20) = 46.7, $p < 0.0001$, $_p\eta^2 = 0.7$). Similarity, however, did not modulate performance (F(1,20) = 2.9, $p = 0.1$, $_p\eta^2 = 0.13$): performance in the uniform polarity condition (0.6 ± 0.01) was comparable with that in the mixed polarity condition (0.61 ± 0.02). Further, there was no interaction between spacing and similarity (F(1,20) = 2.9, $p = 0.1$, $_p\eta^2 = 0.13$). As mentioned above, when a large number of elements are present, similarity appears to no longer be a strong modulatory factor, as documented by previous studies (Manassi et al., 2012; Rosen & Pelli, 2015).

3.3.1.3. Comparison. We compared the effect of spacing and polarity on the two tasks across participants. Once again, there was no correlation between the effect of spacing on crowding and its effect on estimation (Fig. 5E; r(20) = 0.04, $p = 0.88$). Similarly, there was no correlation between the effects of polarity on identification and enumeration (Fig. 5F; r(20) = 0.01, $p = 0.97$).

3.3.2. Experiment 2B

Experiment 2B examined estimation with a reference numerosity of 25 (low density) and hence served as a replication for Experiment 1 (A

& B), and it provided a way to compare the role of density in underestimation using the same stimulus as in Experiment 2A. Data from five participants were excluded (final $n = 17$), because of poor psychometric fits to the estimation data (mean r-square averaged across four conditions < 0.7[2]).

3.3.2.1. Estimation. Inter-element spacing affected numerosity perception, with clustered elements (PSE = 28.1 ± 1.3) being underestimated more than distant elements (25 ± 0.5; $F_{(1,16)} = 12.9$, $p = 0.002$, $_p\eta^2 = 0.45$). However, similarity did not affect numerosity judgment ($F_{(1,16)} = 0.61$, $p = 0.45$, $_p\eta^2 = 0.04$) and performance in the uniform polarity condition (26.1 ± 0.7) was not different from that in the mixed polarity condition (27 ± 1.3). We also did not observe any interaction between the two factors ($F_{(1,16)} = 0.9$, $p = 0.37$, $_p\eta^2 = 0.05$). These results (Fig. 6A and B) match that in Experiment 1B, which tested the same range of numerosities but with a different set of elements and at a lower density.

3.3.2.2. Crowding. In the standard crowding configuration (Fig. 6D, left boxes), both spacing ($F_{(1,16)} = 12.6$, $p = 0.003$, $_p\eta^2 = 0.44$) and similarity ($F_{(1,16)} = 10.8$, $p = 0.005$, $_p\eta^2 = 0.4$) modulated performance. Far flankers (0.72 ± 0.03) interfered less than near flankers (0.63 ± 0.03). Dissimilar flankers (0.71 ± 0.03) interfered less than similar flankers 0.64 ± 0.03. There was also an interaction between these two ($F_{(1,16)} = 10.1$, $p = 0.006$, $_p\eta^2 = 0.39$), mainly because far flankers (0.78 ± 0.03) of the opposite polarity caused the least deterioration in performance. In other words, as in Experiment 1A, the effect of similarity was stronger in the far spacing condition (Mixed – Uniform = 0.13 ± 0.02) than in the near spacing condition (0.01 ± 0.03).

In the presence of many flankers (Fig. 6D, right boxes), dissimilar flankers (0.61 ± 0.02) led to less crowding than similar flankers (0.55 ± 0.02; $F_{(1,16)} = 20.2$, $p < 0.001$, $_p\eta^2 = 0.56$). However, there was only a marginal effect of spacing ($F_{(1,16)} = 3.9$, $p = 0.064$, $_p\eta^2 = 0.2$). There was a much smaller difference in performance between the near (0.57 ± 0.02) and far (0.59 ± 0.02) flanker conditions. There was no interaction ($F_{(1,16)} = 1.6$, $p = 0.22$, $_p\eta^2 = 0.09$) between the two factors. That is, surprisingly, we found, at best, a weak effect of spacing but a stronger effect of similarity. However, this pattern of results is quite distinct from that found in the estimation task, where there was a strong effect of spacing but not that of similarity.

3.3.2.3. Comparison. There was no correlation between the effect of spacing on crowding and its effect on estimation (Fig. 6E; $r = 0.19$, $p = 0.46$). However, there was a link, although not statistically significant, between the effects of polarity on identification and enumeration (Fig. 6F; $r = -0.36$, $p = 0.16$). This relationship was in the expected direction – a reduction in crowding for dissimilar elements was accompanied by an improvement in estimation (reduced PSE).

4. General discussion

The current study was aimed at understanding the relationship between two known phenomena: crowding and numerosity estimation. This comparison allowed us to test the hypothesis that crowding underlies the underestimation of numerosity for stimuli with clustered elements. We tested participants on estimation and crowding tasks on the same stimulus configurations. We manipulated factors that are known to affect crowding and assessed if they affected estimation in the same way. We did this for sparse displays (Experiments 1A, 1B and 2B) and for dense displays (Experiment 2A), and with two types of stimuli and identification tasks. The pattern of results was different in the two tasks, across all these experiments (see Table 1 for a summary). As

expected, closer elements caused stronger crowding than far flankers. In line with this, numerosity judgment too was substantially affected when elements were close to each other, leading to underestimation. On the other hand, target identification was either a) modulated by contrast polarity (Experiment 1) or b) not affected by polarity (Experiment 2). This would have led us to expect that polarity would modulate numerosity estimation in Experiment 1 but would not do so in Experiment 2. However, this was not the case. Contrary to the predictions of the crowding hypothesis, dissimilarity between elements increased underestimation (Experiments 1A and 2A) or did not modulate numerosity judgment (Experiment 1B and 2B). Further, the performance in the two tasks did not correlate with each other. These findings suggest that there is a dissociation between numerosity estimation and crowding performance, implying that crowding cannot explain cluster-induced underestimation.

It is interesting to note that both spacing and similarity manipulations produced the expected effects with the standard crowding configuration (a target surrounded by two or four flankers). However, the effect of similarity was modulated by the stimulus/task in the presence of a large number of flankers. Identifying a target T surrounded by several flankers was affected by similarity in the expected manner but identifying a Vernier offset was not. Similarity of the nearest flankers did not seem to affect performance in the latter case. This might be in line with recent findings by van der Burg, Olivers, and Cass (2017) where the best identification performance in dense displays was observed when the nearest *tangential* flankers were dissimilar to the target (albeit in orientation). In Experiment 1, here, both tangential and radial flankers were dissimilar, whereas in Experiment 2, only the radial flankers were guaranteed to be dissimilar. Given the density of displays it could be that this dissimilarity is not sufficient to lead to improved performance.

Additionally, in Experiment 1, the flankers were relatively farther away from the target (1 or 2 deg. away from a target presented at an eccentricity of 12 deg., giving a spacing-to-target-eccentricity ratio of 0.08 and 0.17, respectively), whereas in Experiment 2, the radial flankers were much closer (0.25 or 0.5 deg. away from the target at an eccentricity of 8 deg., giving a ratio of 0.03 or 0.06). At very close spacing, the interference from nearby flankers might be strong enough to swamp the effect of similarity. For example, Kooi et al. (1994) reported that the largest difference between similarity conditions were observable at moderate distances, when crowding was moderate. That is, the flankers crowd the target to such an extent that the effect of their similarity is not observable. This can explain the lack of an observable effect of similarity in Experiment 2 and for the near spacing in Experiment 1.

Another factor that could affect performance in the two experiments is that the amount of clustering and grouping in Experiment 2 is different from that in Experiment 1. It has been shown that target-flanker similarity in the form of polarity or colour differences is not effective in modulating target identification if several elements are present in the display (Manassi et al., 2012; Rosen & Pelli, 2015). Supporting these results, we found that similarity affected identification when the elements were presented in the standard crowding configuration, but not in extended multi-element displays. Further, the closer spacing in Experiment 2A and 2B along with the higher density in Experiment 2A might have enhanced the influence of these extra elements, explaining the lack of an effect of similarity. These findings can be attributed to grouping between or texture formation across the large set of elements, irrespective of their polarity, which reduces the ability to individuate the target element. That is, heterogeneity among elements offers no protection from texturization.

4.1. Crowding and enumeration of elements at high and low densities

Burr et al. (2018) have suggested that crowding-like interactions play a role in numerosity estimation by enabling the texturization of the

---

[2] R-square of the Gaussian fit was < 0.3 in at least one of the four conditions
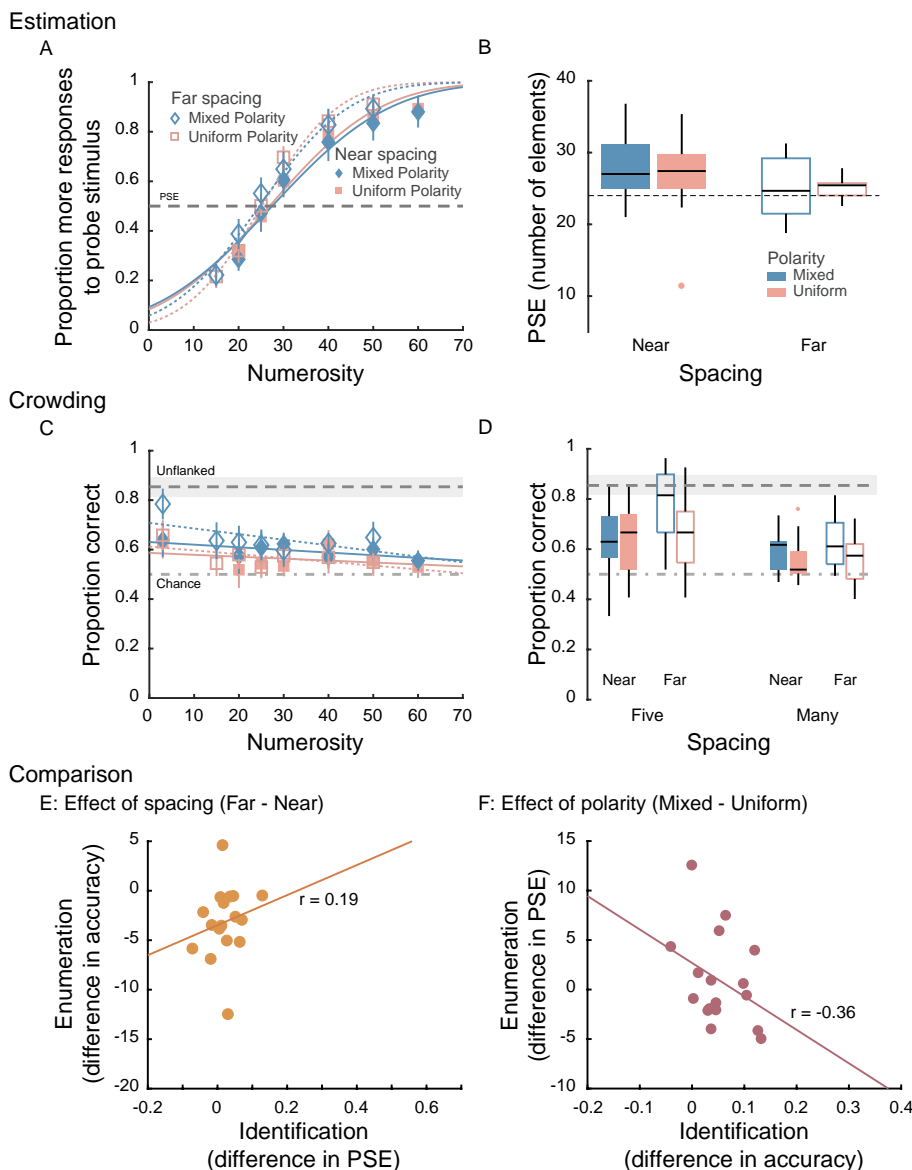
Estimation

Fig. 6. Results of Experiment 2B. A) Proportion more responses to the probe patch in the estimation task as a function of numerosity. Cumulative Gaussian fits are also shown. Error bars represent 95% CI. B) Box plots of PSEs estimated from the fits in the numerosity task. Dashed line indicates veridical estimation (PSE = 25). C) Identification performance in the crowding task as a function of numerosity along with straight line fits. Error bars represent 95% CI. D) Box plots of performance in each of the four conditions in the four-flankers and the many-flanker conditions. The horizontal dashed line and grey shaded region in panels C and D represent accuracy of identifying an unflanked element with 95% CI. The dot-dashed line represents chance performance (0.5). E & F) Comparing performance across the two tasks, identification and enumeration. E) The effect of spacing on the two tasks for each participant. F) The effect of polarity on the two tasks. The solid lines in both panels are the best fitting straight lines.

elements, but only when the density of elements is high. At lower densities, numerosity is argued to be perceived directly, without the necessity of invoking texture-based processing, whereas at high densities numerosity is sensed indirectly via spatial frequency and density/texture processing (Anobile et al., 2015; Zimmermann, 2018; Zimmermann & Fink, 2016). We tested the crowding hypothesis that cluster-induced underestimation of elements can be attributed to crowding between closely spaced elements at both low (Experiments 1A, 1B, and 2B) and high (Experiment 2A) display densities. However, we obtained the same dissociation between estimation and crowding tasks at both densities, indicating that crowding does not seem to play a role in cluster induced underestimation at any density. Nevertheless, it would be useful to distinguish here the claim that crowding plays a role in numerosity estimation from the claim that crowding causes the underestimation due to clustering, the latter of which is the question of the current study. It could be the case that texture perception, and hence crowding, plays a role in numerosity estimation at high densities, but yet does not underlie underestimation due to clustering. Interestingly, it was recently shown that a texture-based model of crowding was successful in predicting underestimation of peripherally viewed objects, as expected from a texture-based mechanism for numerosity estimation (Balas, 2016). It, nonetheless, was unsuccessful in

accounting for the underestimation of clustered dots. Thus, crowding due to excessive pooling might lead to texturization of objects or a perception of the scene's 'summary statistics', but this mechanism cannot explain the underestimation observed when elements are close to each other, supporting our findings. That is, crowding does not seem to provide a general mechanism to explain underestimation.

Nevertheless, the pattern of our results is somewhat different at the two densities when similar tasks and stimuli were utilised (compare Experiments 2A and 2B). In particular, the effect of similarity on estimation differed at the two densities: At low densities, there was no effect of similarity, whereas at high densities dissimilar elements led to higher underestimation. This result appears to contradict earlier findings that, at high densities, contrast polarity does not affect estimation (Dakin et al., 2011; Tibber et al., 2012). However, in those studies, polarity was not manipulated exclusively in the probe patch while the parameters of the reference patch were kept constant but was manipulated in both (probe and reference) patches, which might have led to the observance of a lack of an effect of polarity on estimation. It would not have been possible to discern an effect even if there was one.

Our finding that polarity has differential effects on estimation at different densities lends support to the argument that estimation mechanisms are different at the two densities. The direct sensing

Table 1

Summary of results from all 4 experiments. The top half summarises the effects of reducing spacing and the bottom half documents the effects of reducing similarity (creating a pop-out effect). The effects of spacing are congruent with the crowding hypothesis – reducing spacing increases crowding and underestimation. However, reducing similarity reduces crowding but increases underestimation contradicting the crowding hypothesis. Green shading indicates agreement with the crowding hypothesis, whereas red and orange indicate disagreement.

		Experiment 1A (24 shapes)	Experiment 1B (24 shapes)	Experiment 2A (100 lines)	Experiment 2B (25 lines)
→ Spacing	Estimation	↑ underestimation	↑ underestimation	↑ underestimation	↑ underestimation
	Crowding (multiple flankers)	↑ crowding	↑ crowding	↑ crowding	↑ crowding
	Crowding (standard configuration)	↑ crowding	↑ crowding	↔ No effect	↑ crowding
→ Similarity	Estimation	↑ underestimation	↔ No effect	↑ underestimation	↔ No effect
	Crowding (multiple flankers)	↓ crowding	↓ crowding	↔ No effect	↔ No effect
	Crowding (standard configuration)	↓ crowding	↓ crowding	↓ crowding	↓ crowding

mechanism is insensitive to polarity differences between elements whereas the indirect computation mechanism is susceptible to it.

It is important to note that the way we designed the stimuli could have contributed to the observed differences at the two densities. In the probe patches, the central element was flanked by two distracters. The rest of the elements were placed according to the algorithm described in the Methods section, where elements were placed near 'seed' elements. These seed elements were distributed around the circumference and the interior of the patch. The immediate flankers of the central element were never used as seeds. Hence, the nearest flankers themselves would have been unlikely to be flanked by other elements. That is, the set of the central element and its immediate flankers would remain relatively isolated. This does not affect the testing of the crowding hypothesis, but it is possible that at higher numerosities (Experiment 2A) other elements would be sufficiently near these flankers such that all elements formed a texture, whereas at low numerosities (Experiment 2B), they might not have been integrated to the same extent. This difference might explain the differences in results across these two experiments.

However, one argument against this possible confound is that similarity did not affect identification (Vernier discrimination) performance at both densities. If there was a lack of 'texturisation' of the central elements at the lower density, we should have observed the standard effect of similarity on crowding, which we did not. This was despite finding an effect of similarity in the standard crowding configuration with the same stimuli.

4.2. Mechanisms of cluster-induced underestimation

Our findings suggest that crowding does not underlie cluster-induced underestimation. If so, what mechanism might account for it? One of the earliest explanations for the effect of underestimation was the occupancy model (Allïk & Tuulmets, 1991). In this model, each element has a region of influence, and these regions can overlap. The total region under the influence of the elements is used to compute numerosity. When elements are close to each other, the regions overlap and hence their contribution is reduced. This leads to underestimation. However, this model cannot explain other observations of underestimation where density, and therefore the extent of overlap between regions of influence, is not altered. For example, linking even a few elements within a large set (by connecting them with lines or encircling

them within a common region) reduces perceived numerosity (Franconeri et al., 2009; He et al., 2015; Yu et al., 2019). Further, Anobile et al. (2017) found that this effect is limited to low density displays. It is possible that the occupancy model explains cluster-induced underestimation, while additional processes are at play when other manipulations (e.g., connecting elements) are implemented.

4.2.1. Grouping

A mechanism that might account for all of these findings, including cluster-induced underestimation, is grouping. Grouping binds some of the elements together, which might then be considered as parts of a single object, which leads to underestimation (Franconeri et al., 2009; He et al., 2015; Yu et al., 2019). A similar grouping model of underestimation has been proposed by Im, Zhong, and Halberda (2016). According to this model, the visual system segregates nearby elements into perceptual groups even before numerosity extraction occurs. This grouping is said to occur within windows of around 4° diameter. Visual elements within groups tend to be perceived as units and are bound to each other. Numerosity is extracted from such groups; that is, the number of such perceived groups drives estimates of numerosity. When elements are close to each other, they fall within the same group; hence, only a few perceptual groups are segmented from the stimulus, leading to underestimation. When elements are farther apart, they can potentially join several distinct perceptual groups. The number of perceived groups is therefore higher and there is no underestimation. Thus, grouping might serve as a mechanism for underestimation of configurations with clustered elements.

It has been argued that grouping takes at least two forms (Roelfsema, 2006; Roelfsema & Houtkamp, 2011): 'base' grouping and 'incremental' grouping. Base grouping is automatic, fast, occurs in parallel across the visual field and is thought to be encapsulated by the Gestalt principles of grouping. On the other hand, incremental grouping is slow, serial, requires attention and feedback processing. The grouping that might modulate estimation is likely to be the base kind, where rapid computations regarding the layout of the stimuli are made. When elements in a display are clustered, they are grouped into units that lead to underestimation of the numerosity. Similarly, when dissimilar elements are interspersed, there might be a further cue of grouping based on similarity. Hence, heterogenous displays might be more underestimated thanhomogenous ones. The elements in a homogenous

display are all similar to each other, but as noted above, it might not be clear which element belongs to which cluster and hence there might be several clusters claiming a given element as its member. On the other hand, randomly intermixed elements might incidentally be close to elements of the same type and provide a cue for grouping, allowing exclusive assignment of elements to clusters. Hence dissimilarity can lead to increased underestimation. As can be surmised, such a cue would not be strong; hence the relatively weak effect of similarity on underestimation that we observe. If this is the case, one testable prediction would be that if elements of the same colour or contrast polarity were clustered together, that is, if black elements were close to other black elements and white elements were close to white ones in patches within a large display of elements, we would expect even more underestimation than what we observed here. One piece of evidence that supports this prediction is that it is more inefficient to enumerate the number of colours in a display when the coloured elements are intermixed than when they are segregated or clustered (Watson, Maylor, & Bruce, 2005).

Interestingly, there is evidence that grouping plays an important role in crowding as well. Grouping modulates feature and object binding that in turn affects identification (Herzog, Sayim, Chicherov, & Manassi, 2015). It has long been evident the Gestalt factors such as proximity, similarity and common fate modulate crowding (Bex & Dakin, 2005; Bouma, 1970; Kooi et al., 1994; Scolari et al., 2007). However, recent studies have demonstrated a deeper relationship between crowding and grouping. In general, flankers that group with the target increase crowding (Chakravarthi & Pelli, 2011; Saarela, Sayim, Westheimer, & Herzog, 2009; Saarela, Westheimer, & Herzog, 2010) whereas those that group among themselves without involving the target reduce crowding (e.g., Livne & Sagi, 2007; Manassi et al., 2012; Manassi, Sayim, & Herzog, 2013). Remarkably, flankers can affect target identification through grouping even if they are far away, well outside what has traditionally been considered the extent of crowding. Such results have led some researchers to reject the bottom-up pooling hypothesis of crowding and instead propose that grouping of elements across the entire visual scene first takes place which then dictates how the visual system performs a given task, such as target identification (Herzog & Manassi, 2015).

Nevertheless, it is interesting to note that the role of grouping is different in the two tasks, identification and estimation. For example, similarity plays a substantial role in crowding wherein dissimilar flankers do not impair target identification (current study; Kooi et al., 1994; Scolari et al., 2007). By contrast, dissimilarity *impairs* estimation. Recent results have shown that manipulating Gestalt principles such as proximity, connectivity and common region affected estimation but not (colour) similarity (He et al., 2015; Yu et al., 2019). It has been argued that when there are elements of different colours, there is no underestimation because attention cannot select different colours at the same time (Yu et al., 2019) or colour segregation cannot lead to separate topological units (He et al., 2015), which are needed for underestimation. Further, even though proximity affects both tasks, we found that its influence on one task is not predicted by its influence on the other, indicating that proximity acts on the two processes in distinct ways.

4.2.2. Alternative explanations

One alternative to the grouping hypothesis is to posit that our stimulus setup measured crowding only in a local region. That is, the key finding that similarity affected crowding but not estimation applies only to the region surrounding the target, since that is the only area where an element is similar or dissimilar to its flankers (e.g., a black T surrounded by either black or white thetas). In other parts of the patch, polarity is randomly assigned. Hence, there is no guarantee that this result about the effect of similarity on crowding holds in the rest of the patch. In summary, this proposal argues that we have not or could not have tested how the manipulations affect crowding in other regions of

the patch. However, we are confident that spacing would affect crowding in the same way in all parts of the patch, as crowding has been demonstrated at several eccentricities with a variety of stimuli. On the other hand, one might argue that crowding across the patch should be comparable across the two polarity conditions. It has been shown that alternating polarity across flankers impairs identification as much as when all stimuli share the same polarity (Manassi et al., 2012; Rosen & Pelli, 2015). Even if our stimuli are not exactly alternating, they could form a texture in the same way as in those experiments leading to strong crowding. If this is the case in our patch as well, the crowding hypothesis would predict that estimation will be affected by spacing but not polarity. At best, estimation should be slightly worse in the uniform polarity condition, since there should be slightly more crowding in the uniform condition than in the mixed polarity condition. This is because even if the two polarity conditions produce the same amount of crowding over the entire patch, there would be slightly more crowding for the central element in the same polarity condition (e.g., Experiment 1). Thus the average amount of crowding should be greater in the uniform polarity condition than in the mixed polarity condition. Instead, we found a small benefit for uniform over mixed polarity conditions in both estimation experiments. Hence, we do not think that the effect of similarity is restricted to the central region of the patch or that we have not appropriately measured crowding across the etnire patch. These results suggest that crowding and estimation are dissociable.

Another possibility is that the task requirements and attentional demands are different in the two tasks, identification and estimation. For identification, observers are expected to filter out most of the elements and attend only to the central element, whereas to estimate the number of elements, they have to spread their attention to cover the entire region of the display patch. Hence, it is plausible that the two factors we manipulated, spacing and similarity, might have differing effects on the two tasks. We first note that task-based differences will exist by definition. It would be difficult, if not impossible to equate all factors across the two tasks. For example, it would be challenging for attention to be held diffusely in an identification task. This is particularly true in our setting since it is not possible to identify a large array of objects at once, one of which is then randomly probed. That is, crowding cannot be tested over the entire patch simultaneously. Similarly, estimation with focussed attention would be impossible. Hence, our hypothesis would be confounded with the task requirements.

Nevertheless, it is not clear how this difference (narrow or diffuse attention) can explain the current findings. Manipulations such as proximity and similarity should act on the display in substantially different ways under conditions of narrow and diffuse attention to explain the observed results. The argument might go like this: the manipulations affect the two tasks differently because of the differing size of the focus. For example, spacing and similarity affect identification, since the target object pops-out (e.g., Scolari et al., 2007; Soo, Chakravarthi, & Andersen, 2018), but only spacing affects estimation whereas similarity does not, since salience is not relevant to estimation. However, this proposal is another way of arguing that crowding is due to a limitation of attentional resolution, which has been strongly debated (e.g., Dakin, Bex, Cass, & Watt, 2009; Freeman & Pelli, 2007; Harrison & Bex, 2015). It also leaves unanswered the question of why salience does not modulate the estimation process. Similarly, it doesn't explain why spacing should matter when attention is diffuse, without invoking other processes such as crowding. Further, while the identification task expects the participant to possibly narrow their focus of attention, there is substantial evidence that objects and elements placed considerably far from the target influence the ability to identify the target (Herzog & Manassi, 2015). This suggests that although it might not be optimal to utilise a diffuse focus of attention in a crowding task, the visual system nevertheless operates over a large region of space for *both* tasks (Herzog & Manassi, 2015; Soo et al., 2018). Hence it is not obvious that task requirements per se drive the results. But we acknowledge that we

cannot rule out the possibility that task and attentional requirement differences play a role, and they need to be studied further.

5. Conclusion

We tested the crowding hypothesis of underestimation in cluttered conditions by assessing identification and estimation in two experiments. We compared two tasks (one about the identity of a target, and one about relative numerosity) using the same stimulus configuration, and we manipulated spacing and similarity between elements. These manipulations affected identification and estimation differently, indicating that crowding could not be the cause of underestimation of clustered elements. We discussed how grouping between elements, rather than crowding, could account for the findings.

CRediT authorship contribution statement

Ramakrishna Chakravarthi: Conceptualization, Methodology, Formal analysis, Writing - review & editing, Data curation, Investigation, Project administration, Resources, Software, Writing - original draft. **Marco Bertamini:** Conceptualization, Methodology, Formal analysis, Writing - review & editing.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2020.104195.

References

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

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., Turi, M., Cicchini, G. M., & Burr, D. (2015). Mechanisms for perception of numerosity or texture-density are governed by crowding-like effects. *Journal of Vision, 15*(5), 4. https://doi.org/10.1167/15.5.4.

Balas, B. (2016). Seeing number using texture: How summary statistics account for reductions in perceived numerosity in the visual periphery. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-016-1204-6.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision, 9*(12), 1–18. https://doi.org/10.1167/9.12.13 13.

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.

Bex, P. J., & Dakin, S. C. (2005). Spatial interference among moving targets. *Vision Research, 45*(11), 1385–1398. https://doi.org/10.1016/j.visres.2004.12.001.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*, 177–178. https://doi.org/10.1038/226177a0.

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

Burg, E. V. der, Olivers, C. N. L., & Cass, J. (2017). Evolving the keys to visual crowding. Journal of Experimental Psychology: Human Perception and Performance, 43(4), 690–699. doi:https://doi.org/10.1037/xhp0000337.

Burr, D., Anobile, G., & Arrighi, R. (2018). Psychophysical evidence for the number sense. *Philosophical Transactions of the Royal Society B, 373*(1740), 20170045.

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., & Cavanagh, P. (2007). Temporal properties of the polarity advantage effect in crowding. *Journal of Vision, 7*(2), 11. https://doi.org/10.1167/7.2.11.

Chakravarthi, R., & Pelli, D. G. (2011). The same binding in contour integration and crowding. *Journal of Vision, 11*(8), 1–12. https://doi.org/10.1167/11.8.10 10.

Dakin, S. C., Bex, P. J., Cass, J. R., & Watt, R. J. (2009). Dissociable effects of attention and crowding on orientation averaging. *Journal of Vision, 9*(11), 1–16. https://doi.org/10.1167/9.11.28 28.

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.

Durgin, F. H. (1995). Texture density adaptation and the perceived numerosity and distribution of texture. *Journal of Experimental Psychology: Human Perception and Performance, 21*(1), 149.

Durgin, F. H. (2008). Texture density adaptation and visual number revisited. *Current Biology, 18*(18), R855–R856. https://doi.org/10.1016/j.cub.2008.07.053.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002.

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

Freeman, J., & Pelli, D. G. (2007). An escape from crowding. *Journal of Vision, 7*(2), 22. https://doi.org/10.1167/7.2.22.

Gebuis, T., Cohen Kadosh, R., & Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: A critical review. *Acta Psychologica, 171*, 17–35. https://doi.org/10.1016/j.actpsy.2016.09.003.

Gebuis, T., & Reynvoet, B. (2012). 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.

Ginsburg, N. (1980). The regular-random numerosity illusion: Rectangular patterns. *The Journal of General Psychology, 103*(2), 211–216. https://doi.org/10.1080/00221309.1980.9921000.

Ginsburg, N. (1991). Numerosity estimation as a function of stimulus organization. *Perception, 20*(5), 681–686. https://doi.org/10.1068/p200681.

Ginsburg, N., & Goldstein, S. R. (1987). Measurement of visual cluster. *The American Journal of Psychology, 100*(2), 193. https://doi.org/10.2307/1422403.

Ginsburg, N., & Nicholls, A. (1988). Perceived numerosity as a function of item size. *Perceptual and Motor Skills, 67*(2), 656–658.

Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2010). Crowding changes appearance. *Current Biology, 20*(6), 496–501. https://doi.org/10.1016/j.cub.2010.01.023.

Harrison, W. J., & Bex, P. J. (2015). A unifying model of orientation crowding in peripheral vision. *Current Biology, 25*(24), 3213–3219. https://doi.org/10.1016/j.cub.2015.10.052.

Harvey, B. M., & Dumoulin, S. O. (2017). Can responses to basic non-numerical visual features explain neural numerosity responses? *NeuroImage, 149*, 200–209. https://doi.org/10.1016/j.neuroimage.2017.02.012.

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.

Herzog, M. H., & Manassi, M. (2015). Uncorking the bottleneck of crowding: A fresh look at object recognition. *Current Opinion in Behavioral Sciences, 1*, 86–93. https://doi.org/10.1016/j.cobeha.2014.10.006.

Herzog, M. H., Sayim, B., Chicherov, V., & Manassi, M. (2015). Crowding, grouping, and object recognition: A matter of appearance. *Journal of Vision, 15*(6), 5. https://doi.org/10.1167/15.6.5.

Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences, 103*(51), 19599–19604. https://doi.org/10.1073/pnas.0609485103.

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.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43*(3), 171–216. https://doi.org/10.1006/cogp.2001.0755.

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.

Kennedy, G. J., & Whitaker, D. (2010). The chromatic selectivity of visual crowding. *Journal of Vision, 10*(6), 1–13. https://doi.org/10.1167/10.6.15 15.

Kleiner, M., Brainard, D., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception, 36*(14), 1.

Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision, 8*(2), 255–279.

Livne, T., & Sagi, D. (2007). Configuration influence on crowding. *Journal of Vision, 7*(2), 1–12 4.

Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision, 12*(10), 1–14. https://doi.org/10.1167/12.10.13 13.

Manassi, M., Sayim, B., & Herzog, M. H. (2013). When crowding of crowding leads to uncrowding. *Journal of Vision, 13*(13), 1–10. https://doi.org/10.1167/13.13.10 10.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*(7), 739–744.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*(12), 12. https://doi.org/10.1167/4.12.12.

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.

Rosen, S., Chakravarthi, R., & Pelli, D. G. (2014). The Bouma law of crowding, revised:

Critical spacing is equal across parts, not objects. *Journal of Vision, 14*(6), 1–15. https://doi.org/10.1167/14.6.10 10.

Rosen, S., & Pelli, D. G. (2015). Crowding by a repeating pattern. *Journal of Vision, 15*(6), 10. https://doi.org/10.1167/15.6.10.

Saarela, T. P., Sayim, B., Westheimer, G., & Herzog, M. H. (2009). Global stimulus configuration modulates crowding. *Journal of Vision, 9*(2), 1–11. https://doi.org/10.1167/9.2.5 5.

Saarela, T. P., Westheimer, G., & Herzog, M. H. (2010). The effect of spacing regularity on visual crowding. *Journal of Vision, 10*(10), 1–7. https://doi.org/10.1167/10.10.17 17.

Sayim, B., & Wagemans, J. (2017). Appearance changes and error characteristics in crowding revealed by drawings. *Journal of Vision, 17*(11), 8. https://doi.org/10.1167/17.11.8.

Scolari, M., Kohnen, A., Barton, B., & Awh, E. (2007). Spatial attention, preview, and popout: Which factors influence critical spacing in crowded displays? *Journal of Vision, 7*(2), 1–23 7.

Soo, L., Chakravarthi, R., & Andersen, S. K. (2018). Critical resolution: A superior measure of crowding. *Vision Research, 153*, 13–23. https://doi.org/10.1016/j.visres.2018.08.005.

Strasburger, H. (2005). Unfocussed spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision, 5*(11), 1024–1037.

Strasburger, H., Harvey, L. O., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics, 49*(6), 495–508.

Stuart, J. A., & Burian, H. M. (1962). A study of separation difficulty*: Its relationship to visual acuity in normal and amblyopic eyes. *American Journal of Ophthalmology,* *53*(3), 471–477.

Tibber, M. S., Greenwood, J. A., & Dakin, S. C. (2012). Number and density discrimination rely on a common metric: Similar psychophysical effects of size, contrast, and divided attention. *Journal of Vision, 12*(6), 1–19. https://doi.org/10.1167/12.6.8 8.

Tripathy, S. P., & Cavanagh, P. (2002). The extent of crowding in peripheral vision does not scale with target size. *Vision Research, 42*(20), 2357–2369.

Tyler, C. W., & Likova, L. T. (2007). Crowding: A neuroanalytic approach. *Journal of Vision, 7*(2), 1–9 16.

Valsecchi, M., Toscani, M., & Gegenfurtner, K. R. (2013). Perceived numerosity is reduced in peripheral vision. *Journal of Vision, 13*(13), 1–16. https://doi.org/10.1167/13.13.7 7.

Vos, P. G., Van Oeffelen, M. P., Tibosch, H. J., & Allik, J. (1988). Interactions between area and numerosity. *Psychological Research, 50*(3), 148–154.

Watson, D. G., Maylor, E. A., & Bruce, L. A. M. (2005). The efficiency of feature-based subitization and counting. *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1449–1462. https://doi.org/10.1037/0096-1523.31.6.1449.

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.

Zimmermann, E. (2018). Small numbers are sensed directly, high numbers constructed from size and density. *Cognition, 173*, 1–7. https://doi.org/10.1016/j.cognition.2017.12.003.

Zimmermann, E., & Fink, G. R. (2016). Numerosity perception after size adaptation. *Scientific Reports, 6*, 32810.