The dynamics of grouping-induced biases in apparent numerosity revealed by a continuous tracking technique

Pierfrancesco Ambrosi

Department of Neuroscience, Psychology, Pharmacology, and Child Health, University of Florence, Florence, Italy ✉

Antonella Pomè

Department of Neuroscience, Psychology, Pharmacology, and Child Health, University of Florence, Florence, Italy ✉

David Charles Burr

Department of Neuroscience, Psychology, Pharmacology, and Child Health, University of Florence, Florence, Italy
School of Psychology, University of Sydney, Sydney, NSW, Australia ✉

Connecting pairs of items causes robust underestimation of the numerosity of an ensemble, presumably by invoking grouping mechanisms. Here we asked whether this underestimation in numerosity judgments could be revealed and further explored by continuous tracking, a newly developed technique that allows for fast and efficient data acquisition and monitors the dynamics of the responses. Participants continuously reproduced the perceived numerosity of a cloud of dots by moving a cursor along a number line, while the number of dots and the proportion connected by lines varied over time following two independent random walks. The technique was robust and efficient, and correlated well with results obtained with a standard psychophysics task. Connecting objects with lines caused an underestimation of approximately 15% during tracking, agreeing with previous studies. The response to the lines was slower than the response to the physical numerosity, with a delay of approximately 150 ms, suggesting that this extra time is necessary for processing the grouping effect.

Introduction

Humans, together with many species of animals, can estimate without counting the numerosity of a group of items, rapidly and efficiently (Dehaene, 2011; Nieder, 2019). Like many other visual features, numerosity perception has been shown to be susceptible to adaptation (Arrighi, Togoli, & Burr, 2014; Burr & Ross, 2008), suggesting that it is a fundamental visual attribute with a dedicated perceptual mechanism (but see Gebuis, Cohen Kadosh, & Gevers, 2016 and Leibovich, Katzin, Harel, & Henik, 2017 for a different

point of view). This approximate number system has been found to be largely independent of other perceptual features, such as item size, overall occupied area, and item density (Anobile, Cicchini, & Burr, 2014; Cicchini, Anobile, & Burr, 2016), and has also been related to more complex numerical and mathematical abilities (Anobile, Stievano, & Burr, 2013; Halberda, Mazzocco, & Feigenson, 2008; Piazza et al., 2010).

The existence of this seemingly innate ability in humans and many animals has been associated with the evolutionary advantage of rapidly evaluating food resources and social situations. Dedicated mechanisms for numerosity perception have been shown to arise spontaneously in neural networks not specifically trained for numerosity discrimination (DeWind, Park, Woldorff, & Brannon, 2019; Stoianov & Zorzi, 2012), suggesting that the approximate number system may originate in any complex visual system evolved for object recognition.

An interesting hallmark of numerosity perception is its susceptibility to perceptual grouping. When the items in a scene are connected into pairs, either by lines or more subtle forms of connection, people tend to underestimate their numerosity (Adriano, Rinaldi, & Girelli, 2021; Franconeri, Bemis, & Alvarez, 2009; He, Zhang, Zhou, & Chen, 2009; Kirjakovski & Matsumoto, 2016). The grouping by connectedness effect has been consistently reported in many studies and provides evidence that the approximate number system operates on segmented objects rather than individual local elements. For densely packed items, this connectivity effect is greatly decreased (Anobile, Cicchini, Pomè, & Burr, 2017), showing that it is limited to the estimation range of numerosity, not texture density. It also affects functional magnetic resonance

imaging response to number (He, Zhou, Zhou, He, & Chen, 2015), adaptation to number (Fornaciai, Cicchini, & Burr, 2016), and pupillometry (Castaldi, Pomè, Cicchini, Burr, & Binda, 2021). Furthermore, it has been demonstrated recently that the magnitude of the effect varies according to the perceptual styles of the participants, being weaker for those having a more local–detailed perceptual style, compared with those preferring a more global configuration, as evident from the Autistic Scores of participants (Pomè, Caponi, &. Burr, 2021a), and it requires attention to be directly focused to the stimulus (Pomè, Caponi, & Burr, 2021b).

A recent electroencephalographic and functional magnetic resonance imaging study (Fornaciai & Park, 2018) showed that the effect of grouping is not immediate, but arises after 150 ms after stimulus onset: earlier neural activity (<100 ms) was modulated only by physical numerosity, possibly reflecting the sensory representation of a dot array before perceptual segmentation. This finding provides evidence for multiple stages of processing in the formation of numerosity representations (Park, DeWind, Woldorff, & Brannon, 2016; Roggeman, Santens, Fias, & Verguts, 2011), in line with several computational models of numerosity encoding (Dehaene & Changeux, 1993; Stoianov & Zorzi, 2012).

Here, we investigate the effect of grouping on numerosity perception using the novel technique of continuous tracking (Bonnen, Burge, Yates, Pillow, & Cormack, 2015). In continuous tracking, participants are presented with a stimulus that varies sequentially in some property and are asked to reproduce or counteract changes in the stimulus. The similarity between changes in the stimulus and responses from participants is measured by cross-correlation. This technique has been used mostly in object tracking (Bonnen et al., 2015; Bonnen, Huk, & Cormack, 2017; Mulligan, Stevenson, & Cormack, 2013; Mulligan, 2002), where a stimulus moves across the screen in a random walk, has also provided reliable information in other perceptual domains, including motion (Bhat, Cicchini, & Burr, 2018) and numerosity perception (Ambrosi, Burr, & Cicchini, 2021). Continuous tracking has the advantage of producing large amounts of data in short acquisition times, as well as investigating dynamic properties of perceptual processes.

In the present study, we asked participants to track the numerosity of a dynamic cloud of items by continuously moving a cursor on a number line. The numerosity of the cloud was manipulated so that some of the disks were connected by thin lines, causing an underestimation of numerosity via the grouping effect. Testing different rates of changes in the stimulus provided information on the dynamics of participants' responses that were compared with a standard two-alternative forced choice (2AFC) experiment. The results revealed that participant responses were

modulated by both physical numerosity and grouping, and that the grouping requires additional encoding time (about 150 ms) to affect numerosity perception.

Methods

Participants

Eleven voluntary participants (aged 26–33 years, 6 female) were recruited, all with normal or corrected-to-normal vision. All participants had prior experience in psychophysical experiments, but only one had prior knowledge about the details of the experiment (author P.A.). All were right handed and used their right hand for tracking. All participants gave written informed consent, and experimental procedures were approved by the local ethics committee (Commissione per l'Etica della Ricerca, University of Florence, July 7, 2020, n. 111) and are in line with the Declaration of Helsinki.

Apparatus

Stimuli were displayed on a 70 × 40 cm Display++ LCD Monitor (Cambridge Research Systems, Rochester, UK) with a resolution of 1920 × 1080 pixels and a refresh rate of 120 Hz. A Thrustmaster 2960623 USB joystick for PC was used to collect responses in the tracking experiment, and a regular USB keyboard for the 2AFC experiment. In all experiments, participants were placed 70 cm from the screen.

Tracking experiment

Stimuli

Stimuli were generated by the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). The target was a cloud of black and white dots of random sizes (from 0.3° to 0.4° diameter) on a uniform gray background, comprised within a circular area (diameter 11°). Each dot was attached to a line segment (length 0.55°) of the same color, resulting in a cloud of lollipop-shaped objects. Lollipop sizes and positions were controlled to avoid overlapping lines. Each lollipop had a maximum presentation time (lifetime) of 100 ms (12 frames), then appeared in a different random position. This limited lifetime paradigm avoided participants from considering only new or disappearing items. At the beginning of each session, there were always 36 lollipops, with 16 of them (approximately 40%) connected to make dumbbell-shaped objects. During the experiment, both the numerosity of the cloud and the number of connected objects varied
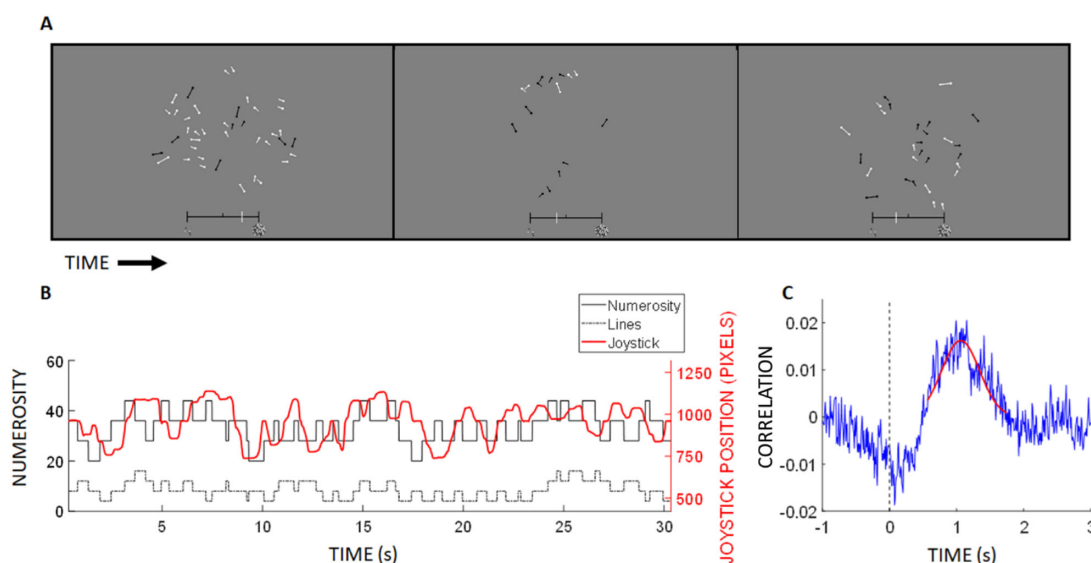
Figure 1. (A) Example of stimulus sequence. The stimulus is a cloud of randomly spaced dots with short lines attached, either in isolation ("lollipops") or connected to form "dumbbells." Both the number of items, and the proportion connected, changed independently over time. Participants tracked the perceived numerosity by moving with a joystick a cursor along a number line below the stimulus. (B) Example data from a 30-second trial, showing numerosity, number of lines and cursor position as a function of time. The black solid line is the number of dots on screen, both isolated and connected; the black dashed line is the number of connected pairs; the red line is the cursor position through time. (C) Example of temporal similarity measured through cross correlation between changes in numerosity and cursor velocity. The red line is the result of a gaussian fit around the peak of the cross-correlogram.

randomly, following a random walk constrained within the range of 20 to 54 dots, with 0 to 32 connected. A number line (length 8°) was displayed below the stimuli, and participants moved a cursor with a joystick to record continuously their estimate of perceived numerosity. At the extremes of the number line, two clouds of dots of numerosity 5 and 60 were presented as references. Figure 1A and Movie 1 show examples of the stimuli.

Experimental procedure

The numerosity of the stimulus changed at fixed time intervals, with an additional temporal jitter between 8 and 50 ms to minimize temporal autocorrelations. The number of changes in the positive and negative directions was equal in every 30-second session. The task of the participants was to move the joystick left or right to track the numerosity on screen. Participants were not required to keep fixation and were free to inspect the stimulus. Each change on screen consisted in ±8 lollipops and ±4 connecting lines. Three different experimental conditions with different rates of changes were tested: 800 ms, 500 ms, and 200 ms. The 500-ms condition comprised 30 blocks of 30 seconds each (15 minutes), and both the 800-ms and 200-ms 15 blocks of 30 seconds (7.5 minutes), collected in separate sessions. The total of the 30-minute testing session was

performed in two rounds by each participant, both preceded by a 2-minute training session.

2AFC Experiment

We also measured perceived numerosity with standard 2AFC technique. Stimuli were presented sequentially to mimic the sequential presentation of the tracking experiment. Two conditions were tested in the connected condition, where the probe stimulus comprised 36 lollipops with 14 connected by lines (40% connected). The tests stimulus consisted in isolated lollipops with numerosity spanning from 10 to 60. In the isolated condition, both the probe and the test comprised isolated lollipops. Participants had to indicate which stimulus was more numerous, by pressing the corresponding arrows on a keyboard. Stimuli were identical to the ones used in the tracking experiment, with a presentation time of 500 ms. Participants were allowed 2 seconds to answer, before the presentation of a new trial. Each participant performed 400 trials.

Data analysis

Psychophysics

Data from each participant in the 2AFC experiments were analyzed separately, plotting individual psychometric functions, proportion of trials where test

appeared of higher numerosity than the probe as a function of probe numerosity. Data were fit with a cumulative Gaussian distribution whose mean defines the point of subjective equality (PSE). The width of the underlying Gaussian divided by its mean gives the participant's Weber fraction. From the PSEs of each participant, we computed the bias index expressed as:

$$BIAS\,(\%) = 100 * \left(\frac{PSE}{N} - 1 \right)$$

where $N = 36$ is the probe numerosity.

Tracking

We measured the normalized cross-correlation between changes in cursor position (output) and computer-driven changes in the stimuli (input). The value of the maximum correlation (peak) was taken as a measure of participants' performance in each condition, where higher peaks of correlation mean a greater similarity between changes in the stimulus and tracks reproduced by participants. The width of the cross-correlogram is also related to perceptual abilities (Ambrosi et al., 2021; Bonnen et al., 2015), and can be regarded as an indication of how long the response to a stimulus persists. In contrast, the lag of the cross-correlogram is only mildly related to perceptual properties and more to the promptness of the decision-making process: lower lags are generally found when more visible changes in the stimulus are used, but they do not correlate with perceptual thresholds, as do peaks and widths (Ambrosi et al., 2021; Bonnen et al., 2015). Figure 1C shows an example of the resulting cross-correlation between numerosity changes and joystick movements. The red curve is the best fitting Gaussian to the peak correlation.

To estimate the effect of grouping, the on-screen numerosity was recalculated as

$$N' = N - w_G N_{Lines} \quad (1)$$

where the recalculated numerosity N' is the tracking estimate of the PSE, N is the physical number of lollipops on screen, N_{Lines} is the number of lines connecting the lollipops and w_G is the magnitude of the grouping effect. The optimal w_G was estimated for each subject as the weighting factor that maximizes the peak correlation, testing the range of $-1 < w_G < 2$. Positive values of w_G correspond with underestimation caused by grouping, where $w_G = 1$ means two disks appear as a single object (theoretical maximum effect), and $w_G = 0$ means no effect at all. To compare the grouping effect

w_G with standard psychophysical measures, we express the grouping bias index:

$$BIAS\,(\%) = 100 * \left(\frac{N'}{N} - 1 \right)$$

Substituting N' from the previous equation gives

$$BIAS\,(\%) = -100 * w_G \frac{N_{Lines}}{N}$$

In the 2AFC experiment, $N_{Lines} = N/5$ (40% connected), so we express the bias index as grouping effect w_G:

$$w_G = -5 * \left(\frac{BIAS\,(\%)}{100} \right)$$

To address the temporal properties of grouping, we analyzed data from the aggregate participant (all participants combined), simulating the delayed effect of linking items, from 0 ms to 500 ms, while independently varying the weighting factor w_G, as described elsewhere in this article. The optimal delay and weight were again estimated to maximize the peak correlation between stimulus changes and joystick movements.

All the analyses were implemented on Matlab 2019b and on JASP (Version 0.14.1; jasp-stats.org). Correlation analyses and analyses of variance were complemented with Bayes Factors estimations. They correspond to inclusion Bayes factors resulting from the analysis of the effects across "all matched models." Bayes factors are reported in logarithmic base 10 units ($\log_{10}$BF) and their absolute values should be interpreted as providing anecdotal (0–0.5), substantial (0.5–1), strong (1–1.5), or very strong (>1.5) evidence, in favor of the alternative hypothesis if positive or the null hypothesis if negative (van Doorn et al., 2021).

Results

Tracking

As described in the Methods, participants were required to continuously estimate the numerosity of a cloud of dots and reproduce it on a number line using a joystick. Both the number of dots and the proportion connected by lines varied over time, following two independent random walks (Figure 1B). To study the temporal dynamics of participants' responses, three different rates of changes in the stimulus were tested (800, 500, and 200 ms).

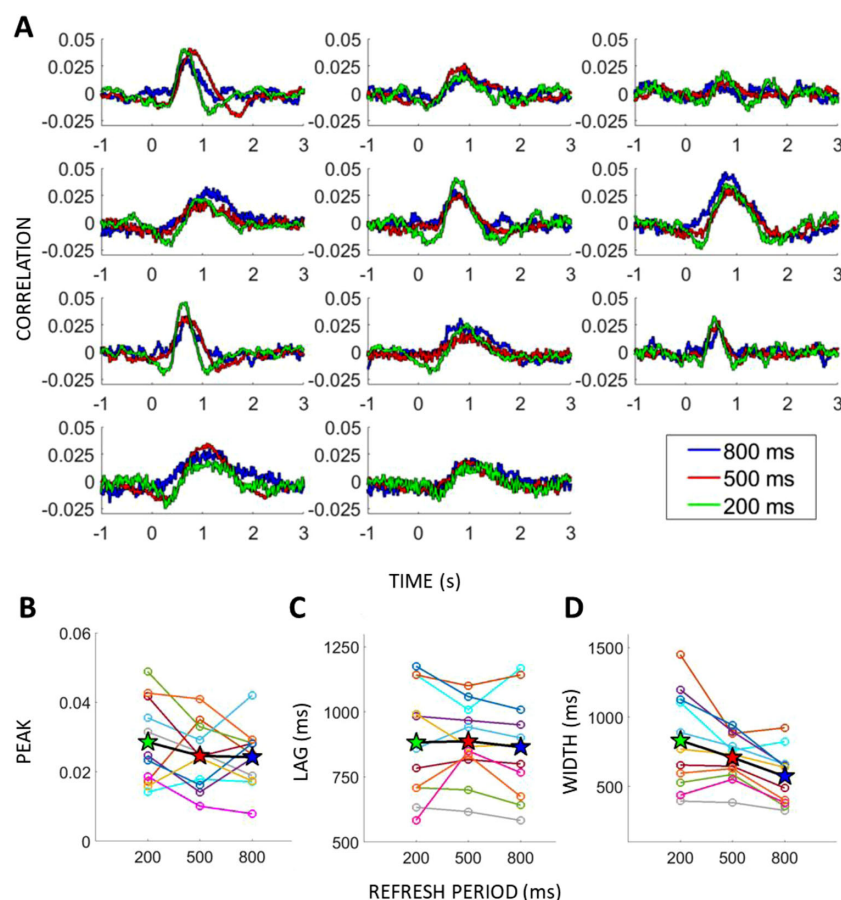The main analyses were performed on the data of individual participants. To characterize the strength of

Figure 2. (A) Cross correlations in the three tested conditions (800 ms in blue, 500 ms in red, and 200 ms in green) for the 11 participants, with each panel showing the results from a single participant. The cross-correlograms are qualitatively similar in all conditions. (B) Peak of the cross-correlogram for each participant and each condition. (C) Lag of the peak in milliseconds of the cross-correlogram for each participant and each condition. (D) Width of the cross-correlogram, estimated as the Full width at half maximum of the gaussian fit around the peak. For B, C, and D, circles are single participants values; stars are average values, color coded as in A.

the cross correlation, we fitted the positive lobe with a Gaussian, free to vary in height, width, and position (Figure 1C). Figure 2 shows the results of the cross-correlation analysis for all conditions and participants. For each participant, shown in the different panels of Figure 2A, cross-correlations are very similar to each other, with largely overlapping profiles. This outcome indicates that tracking performance was not degraded by the faster refresh rates. In fact, all three conditions (200, 500, and 800 ms) result in quantitatively similar performances. No statistically significant changes were detected in peak of correlation (Figure 2B) or lag of the peak response (Figure 2C): all Bonferroni corrected $p_{\text{Bonf}} > 0.05$ in post hoc comparisons on Gaussian fits of the cross-correlograms with $\log_{10}\text{BF} = -0.5$ for peaks and $\log_{10}\text{BF} = -0.6$ for lags. The widths of the cross-correlogram (Figure 2D) were statistically different, but only when comparing the 800 and 200 ms conditions ($p_{\text{Bonf}} = 0.04$, $\log_{10}\text{BF} = 0.26$). This difference can be interpreted as follows: with faster

refresh rates, participants' responses are less spread in time, because more rapid movements are made when the stimulus is changing faster, but because the peaks of correlation are not significantly different from each other, this factor does not result in lower tracking performance. Actually, narrower widths of the cross-correlation are generally associated with better performance (Ambrosi et al., 2021; Bonnen et al., 2015; Bonnen et al., 2017). The results, therefore, suggest a minor trend favoring the conditions with faster refresh rate, with slightly higher peaks of correlation (Figure 2B) and narrower widths (Figure 2D).

Figure 3 shows the method for calculating the effect of connectivity on apparent numerosity. We assumed that connecting the dots decreased the apparent numerosity in direct proportion to the number of connected dots, with a weighting factor w_G, where $w_G = 0$ means zero effect and $w_G = 1$ means that each connected dumbbell was perceived as a single item (Equation 1). We determined the grouping effect
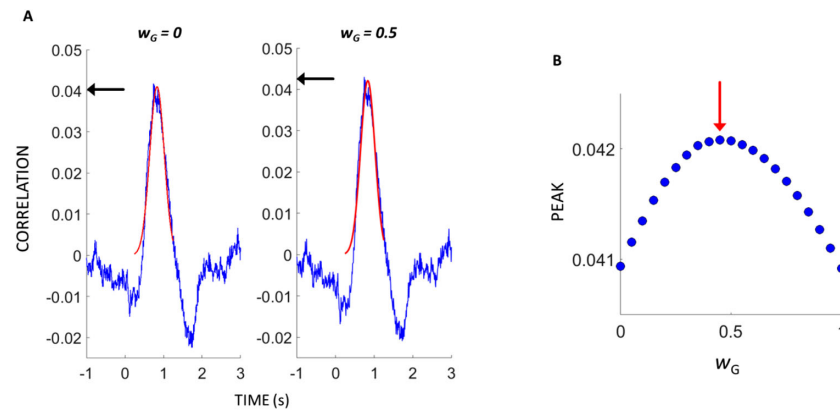
Figure 3. (A) Example of the resulting cross-correlograms when varying the weighting factor w_G of the grouping effect (from left to right, respectively: 0 and 0.5), as explained in the Methods. Weighting differently the lines results in a modulation of the peak correlation. Red lines are results of gaussian fits around the maximum value. Black arrows highlight the peak of the fitted cross-correlogram. (B) Example of maximum peak correlation values as a function of the weighting factor w_G. The red arrow indicates the w_G that maximizes the peak correlation.
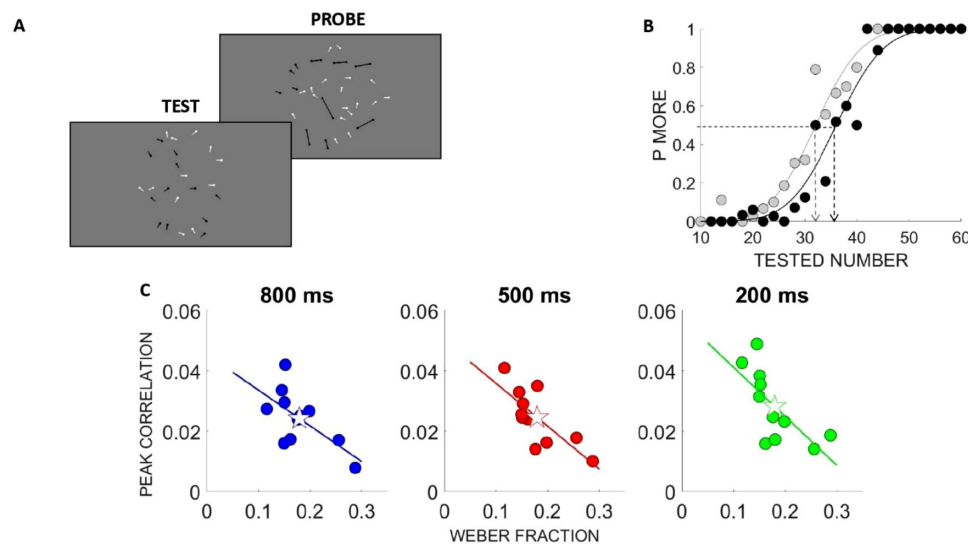

Figure 4. (A) Example of stimuli in the two-alternative forced choice (2AFC) experiment. The probe comprised a cloud of 36 lollipops, with 40% connected by lines in the connected condition. The test was always a cloud of isolated lollipops of various numerosities. Test and probe were presented sequentially, in random order, with a presentation time of 500 ms. (B) Example of psychometric curve obtained in the isolated (black) and 40% connected (gray) 2AFC experiments by plotting the probability of reporting the test numerosity as more numerous than the probe. The gray curve is clearly shifted to the left, showing that in the connected condition the probe numerosity is underestimated. The vertical dashed lines show the estimates of the PSE, given by the mean of the fitted cumulative Gaussian functions. (C) Peaks of cross-correlograms as a function of precision (Weber fraction) in the 2AFC experiment, for the three conditions separately. Circles represent single subjects; stars show the mean across subjects. Thick lines show the linear fit through the data.

w_G empirically by varying it over a wide range and choosing the value that gave the maximum peak in cross-correlation. Figure 3A illustrates the procedure for an example participant for two different weights: the highest peak was for $w_G = 0.5$. Figure 3B shows how the correlation varied smoothly with w_G over the range, with a maximum at 0.5, indicated by a red arrow. The best fitting weights for all participants for the three refresh rates are plotted on the ordinate of Figure 5, as discussed elsewhere in this article.
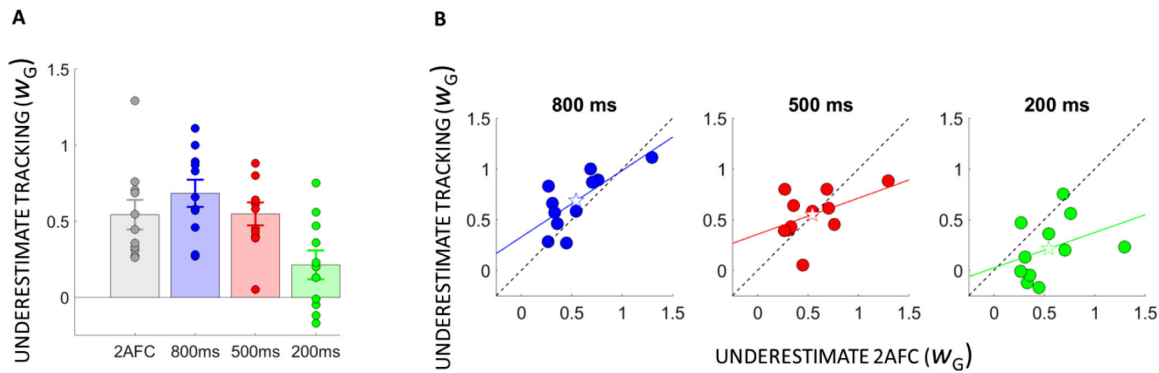
Figure 5. (A) Bias indexes for each participant in the two-alternative forced choice (2AFC) and in each condition of the tracking experiment. Dots are estimates for each participant separately, bars are mean values across participants, error bars are standard error of the mean. (B) Underestimation bias (%) in the 2AFC experiment as a function of the bias in the tracking experiment for the three conditions separately (800 in blue, 500 in red, and 200 in green). Circles represent single subjects; stars show the mean across subjects. Thick lines show the linear fit through the data. The black dashed line is the equality line.

Psychophysics

After completing the first tracking experiment, participants were asked to judge which of two sequentially presented arrays seemed to include more dots. One was the test (spanning 5–60 isolated dots), the other the probe (with 36 dots connected by 7 lines; see the Methods for experimental details) (Figure 4A)

Figure 4B shows the psychometric curves obtained from plotting the probability of reporting the test as higher numerosity for the two experimental conditions (isolated in black and connected in gray) for a typical participant. Data were fit with a cumulative Gaussian distribution, shown as solid lines in the figure, whose mean defines the PSE. The width of the underlying Gaussian divided by its mean gives the participant's Weber fraction. The PSE (0.5 response) yields the numerosity of the test that matched the probe. The dashed arrows highlight the PSEs in the two conditions: as expected in the isolated condition, PSEs are close to the numerosity of the probe (mean PSE $\pm$ standard error of the mean = 37.1 $\pm$ 0.2), whereas connecting the 40% of dots yields to an underestimation of approximately 11% (mean $\pm$ standard error of the mean PSE = 32.1 $\pm$ 0.7). Also, the Weber fractions in the connected and unconnected condition were significantly correlated ($r = 0.73$; $p = 0.01$; $\log_{10}$BF = 0.8) and nonsignificantly different in a pairwise t test.

Figure 4C plots the values of peak correlation for the tracking experiment against Weber Fractions (the ratio between the mean and standard deviation of the best-fitting Gaussians for the psychometric functions) for the three different refresh rates. For all the three conditions, the two variables are significantly negatively correlated across participants (800 ms: $r = -0.64$, $p = 0.03$, $\log_{10}$BF = 0.33; 500 ms: $r = -0.76$, $p < 0.01$,

$\log_{10}$BF = 0.92; 200 ms: $r = -0.70$, $p = 0.02$, $\log_{10}$BF = 0.58), meaning that the lower the Weber fraction, the higher the peak correlation.

Figure 5 shows the comparison of bias indexes estimated from the 2AFC experiments and from tracking. Figure 5A displays the bias for each participant estimated in the 2AFC experiment (gray), together with its average across participants and the averages bias estimated with tracking (blue for the 800-ms condition, red for the 500-ms condition, and green for the 200-ms condition). The mean bias across subjects in the 2AFC experiment (mean $\pm$ standard deviation = $-11\% \pm 2\% = 0.55 \pm 0.10$ in terms of the weighting factor w_G) agrees with the mean bias in the 800-ms and 500-ms tracking experiments (0.7 $\pm$ 0.1 and 0.55 $\pm$ 0.05 respectively), but not in the 200-ms condition (0.2 $\pm$ 0.1).

Interestingly, the bias estimated from the 2AFC experiment significantly correlated with bias from tracking in the 800 ms condition ($r = 0.72$, $p = 0.01$, $\log_{10}$BF = 0.70), but not with the bias in the other two conditions (500 ms: $r = 0.46$, $p = 0.1$, $\log_{10}$BF = -0.20; 200 ms: $r = 0.36$, $p = 0.3$, $\log_{10}$BF = -0.39), as shown in Figure 5B.

This result indicates a dependency of the grouping effect on the time available to respond, approaching the 2AFC bias for smaller refresh rates. To explore further this idea, we used an aggregate participant analysis, pooling across all participants, and studied the temporal properties of the response to the lines connecting the discs. We varied independently the grouping weighting factor and the delay between changes in numerosity and changes in the number of lines connecting the items on screen and searched for the maximum cross-correlation peak over the two dimensions. We simulated a delay of 0 to 500 ms for the effectiveness of grouping, while
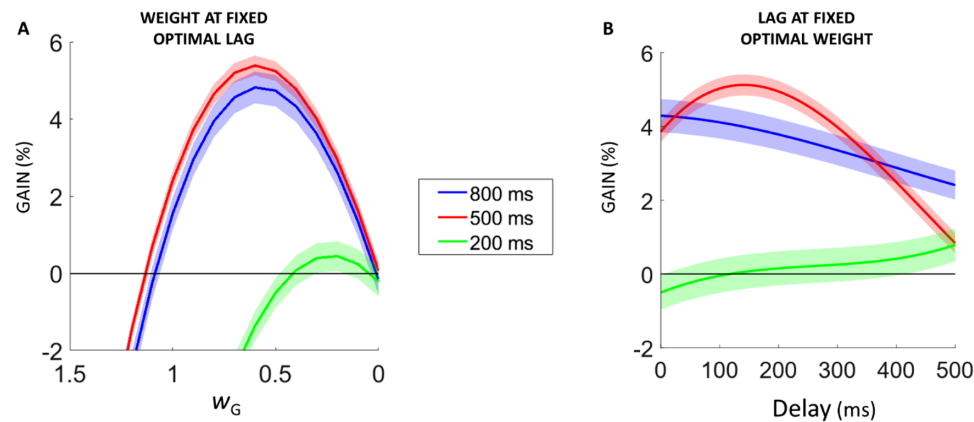
Figure 6. (A) Gain in peak correlation as a function of the grouping weighting factor, with delay between numerosity and number of lines changes fixed at the optimal value found in the two-dimensional grid search analysis. Shaded areas represent the ±1 standard error of the mean estimated via bootstrapping. The gain was calculated as the relative increment of the peak correlation as a function of the grouping effect w_G and of the delay, in respect to the peak found for $w_G = 0$ and null delay. (B) Gain in peak correlation as a function of the delay between changes in numerosity and changes in the number of lines. Shaded areas represent ± 1 standard error of the mean.

simultaneously varying the weight for the effect of the lines.

The results of this analysis are shown in Figure 6. Figures 6A and B show the trend of the gain in correlation with respect to the response to the physical numerosity, that is, the relative increase in the peak correlation with respect to the value obtained for a w_G of 0 and a null delay. To ease visualization, data were plotted showing the variations owing to only one of the two parameters, fixing the other to the optimal value found in the grid search analysis. The displayed trends were obtained fitting the results with a polynomial function, and the optimal degree of the polynomial was found as the best trade-off between the goodness of fit measured through root mean squared error and model complexity. The 200-ms condition has very low gain (0.88%) with respect to the response to the physical numerosity when compared with the 500- and 800-ms conditions (4.8% and 4.4%, respectively). Also, the dependency from the delay between changes in numerosity and in the number of lines connecting the items is close to null, with a maximum gain for a delay of 458 ms, which may be considered spurious, because the corresponding gain is very low. Note that the gain at delays of up to approximately 100 ms are negative because the gain is evaluated with respect to the response to the physical numerosity (delay = 0 ms and $w_G = 0$).

The 800- and 500-ms conditions show a stronger dependency on the delay (Figure 6B): in the 800-ms condition, the dependency from the delay is broad, with a global peak of 4.4% at 40 ms. In this condition, gains over the range of 0 to 140 ms are not significantly different from the highest gain at 40 ms (two-sample *t* tests on peaks of correlation obtained

through bootstrapping and comparing the resulting distributions for each delay with the distribution corresponding to the optimal delay). Interestingly, the gain in the 500-ms condition has a peak at 150 ms, with a gain of almost 5%. Gains in the 50- and 230-ms ranges are not significantly different from the peak at 150 ms. Distributions were considered not significantly different if $\log_{10}\mathrm{BF} < 0$.

These results suggest that grouping requires additional time for processing, resulting in responses to the number of lines connecting the dots delayed with respect to the response to the physical numerosity on screen.

Discussion

We used continuous tracking to study the effect of grouping on numerosity perception. In a typical tracking experiment, participants are asked to follow or counteract changes in the stimulus, using a cursor or eye movements. Here, participants tracked a property of the ensemble—its numerosity—continuously by sliding a joystick. Both the physical numerosity of a cloud of dots, and the apparent numerosity (by connecting a proportion of the dots with lines), varied following independent random walks. The results show that continuous tracking is a valid technique to study numerosity perception and the effect of grouping induced biases, verified by standard 2AFC techniques. More important, continuous tracking provided information on temporal properties of numerosity perception, suggesting that grouping requires about 150 ms extra time to influence the perception of numerosity.

As in other studies on tracking, we measured the peaks, lags, and widths of cross-correlation between stimulus and joystick position. Like previous studies (Ambrosi et al., 2021; Bonnen et al., 2015; Bonnen et al., 2017), the value of peak correlation was the best predictor of performance and strongly correlated with precision measures using standard 2AFC techniques. The correlation peaks were not significantly different for the three refresh rates, implying that task difficulty was similar. However, the widths of the correlations were significantly different between the 800 and 200 ms conditions, with slightly narrower cross-correlograms for faster refresh rates. Previous studies (Ambrosi et al., 2021; Bonnen et al., 2015) report that narrower cross-correlograms are generally associated with better tracking performances, so this difference excludes the possibility that the 200-ms refresh rate condition was more difficult.

It is known that, when visual items such as circles or squares are grouped together with connecting lines, they seem to be less numerous (Anobile et al., 2017; Franconeri et al., 2009; He et al., 2009; Pomè et al., 2021a), evidence that numerosity operates on segmented objects, defined by grouping, rather than individual local elements. Here, we confirmed that connecting items causes underestimation, both by a classical discrimination task and by the tracking experiment. In the 800- and 500-ms conditions, the underestimation bias in the tracking experiment approached that of the 2AFC (approximately 0.70 and 0.55, respectively, with a gain in correlation of almost 5% on average), whereas the very fast refresh rate (200 ms) showed much lower dependency on the lines, with an average gain in correlation of less than 1%, associated with an average grouping effect of approximately 0.2. More important, the bias indexes estimated in the 2AFC and in the tracking experiments were significantly correlated in the very slow changing rate (800 ms). In contrast, the intermediate (500 ms) and very fast (200 ms) refresh rates resulted in lower grouping effects, with almost no effect in the fastest refresh rate condition. This finding suggested that the underestimation of numerosity caused by grouping is time dependent.

Previous studies have reported no difference in the bias index associated with different presentation times in the 50 to 1000 ms range (Franconeri et al., 2009), so it is possible that that the reduction of the grouping effect was related to the shorter time available to respond. This point was also suggested by the difference between widths of the cross-correlations, which indicates that participants' responses were significantly longer-lasting in the 800-ms condition than in the 200-ms condition.

When simulating changes in the number of lines connecting the items delayed with respect to changes in numerosity, we found that, in the 500-ms condition, the correlation was maximized for delays in the 100- to 200-ms range, implicating a longer processing time

for the grouping effect. The 800-ms condition also showed an increase in correlation when simulating a delayed change in the number of lines, but this increase was in the 0- to 140-ms range. This difference may be because, in this condition, changes in the stimulus were less frequent, so participants were less pressured when performing the task. This factor could explain responses to the lines with slightly lesser delays with respect to the response to the physical numerosity, and a weaker dependency of the gain in correlation as a function of the delay. Moreover, the 200-ms condition showed almost no dependency from delay, with only a small gain in correlation, consistent with the fact that the connectivity effect takes time to cut in, and this refresh rate was too fast.

Taken together, these results suggest that grouping requires additional encoding time to affect numerosity perception, and that this additional encoding time has measurable consequences on behavior. This conclusion is in line with other recent studies on the effects of grouping on numerosity. Fornaciai and Park (2018) showed that neural activity encoding numerosity in early visual cortex precedes neural activity related to the effect of grouping by 50 to 100 ms. They first explained their result as a reentrant feedback process from midlevel visual areas in the dorsal and ventral stream related to the segmentation of a visual scene into perceptual objects. However, later studies from the same group showed that feedback from higher order cortical areas was not strictly necessary for perceptual segmentation, suggesting that feedforward activity carrying high- and low-frequency information are sufficient for numerosity representation (Fornaciai & Park, 2021). Although further experiments are needed to link the neural signature observed in the early visual cortex and the behavioral signature of the grouping process reported in this article, it is possible that the slower processing shown at the neural level might be related to the slower influence of grouping on participants' responses.

Results comparable with ours were also obtained using a speed-reaching paradigm (Milne et al., 2013), which studied the effect of grouping on speeded-reaching movements. Their results show that fast-reaching movements, initiated in less than 325 ms, were not affected by grouping, but only by the physical numerosity, regardless of the proportion of connected items. This result can be compared with our tracking experiment with a 200-ms refresh rate, where only a negligible effect of grouping was present. Also, computational models of numerosity encoding (Dehaene & Changeux, 1993; Stoianov & Zorzi, 2012) predict that numerosity representations arise from multiple stages of processing (Park et al., 2016; Roggeman et al., 2011).

We conclude that the tracking paradigm is a useful tool to study numerosity, including the underestimation

effect caused by grouping, with results compatible with standard psychophysical experiments. The continuous nature of tracking provided dynamic information that suggested, in line with previous behavioral (Milne et al., 2013), neuroimaging (Fornaciai & Park, 2018), and computational (DeWind, 2019; Stoianov & Zorzi, 2012) studies, that numerosity representation arises through multiple stages of processing. From a broader point of view, the large amount of data in reasonable acquisition times produced in tracking experiments might be useful to test children and clinical populations, which may not be able to undergo long testing sessions. Also, the continuous nature of this technique might allow to test more ecologically relevant stimulus conditions, that is, to use more realistic stimuli.

Keywords: numerosity perception, continuous tracking, perceptual grouping

Acknowledgments

Commercial relationships: none.
Corresponding author: Pierfrancesco Ambrosi.
Email: pfa2804@gmail.com.
Address: Department of Neuroscience, Psychology, Pharmacology, and Child Health, University of Florence, Padiglione 26, Via di San Salvi, 26, 50135, Florence, Italy.

References

Adriano, A., Rinaldi, L., & Girelli, L. (2021). Visual illusions as a tool to hijack numerical perception: Disentangling nonsymbolic number from its continuous visual properties. *Journal of Experimental Psychology. Human Perception and Performance, 47*(3), 423–441.

Ambrosi, P., Burr, D. C., & Cicchini, G. M. (2021). Ideal observer analysis for continuous tracking of numerosity. *Journal of Vision, submitted*.

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

Anobile, G., Cicchini, G. M., Pomè,, & 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.

Anobile, G., Stievano, P., & Burr, D. C. (2013). Visual sustained attention and numerosity sensitivity correlate with math achievement in children. *Journal of Experimental Child Psychology, 116* (2), 380–391.

Arrighi, R., Togoli, I., & Burr, D. C. (2014). A generalized sense of number. *Proceedings. Biological Science, 281*(1797), 20141791.

Bhat, A., Cicchini, G. M., & Burr, D. C. (2018). Inhibitory surrounds of motion mechanisms revealed by continuous tracking. *Journal of Vision, 18*(13), 7.

Bonnen, K., Burge, J., Yates, J., Pillow, J., & Cormack, L. K. (2015). Continuous psychophysics: Target-tracking to measure visual sensitivity. *Journal of Vision, 15*(3), 14.

Bonnen, K., Huk, A. C., & Cormack, L. K. (2017). Dynamic mechanisms of visually guided 3D motion tracking. *Journal of Neurophysiology, 118*(3), 1515–1531.

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

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

Castaldi, E., Pomè, A., Cicchini, G. M., Burr, D. C., & Binda, P. (2021). Pupil size automatically encodes numerosity. *Nature Communications, under review*.

Cicchini, G. M., Anobile, G., & Burr, D. C. (2016). Spontaneous perception of numerosity in humans. *Nature Communications, 7*, 12536.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics, Rev. and updated ed*. New York: Oxford University Press.

Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience, 5*(4), 390–407.

DeWind, N. K. (2019). The number sense is an emergent property of a deep convolutional neural network trained for object recognition. *bioRxiv*, 609347.

DeWind, N. K., Park, J., Woldorff, M. G., & Brannon, E. M. (2019). Numerical encoding in early visual cortex. *Cortex, 114*, 76–89.

Fornaciai, M., Cicchini, G. M., & Burr, D. C. (2016). Adaptation to number operates on perceived rather than physical numerosity. *Cognition, 151*, 63–67.

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.

Fornaciai, M., & Park, J. (2021). Disentangling feedforward versus feedback processing in numerosity representation. *Cortex, 135*, 255–267.

Franconeri, S. L., Bemis, D. K., & Alvarez, G. A. (2009). Number estimation relies on a set of segmented objects. *Cognition, 113*(1), 1–13.

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

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*(7213), 665–668.

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.

He, L., Zhou, K., Zhou, T., He, S., & Chen, L. (2015). Topology-defined units in numerosity perception. *Proc Natl Acad Sci USA, 112*(41), E5647–5655.

Kirjakovski, A., & Matsumoto, E. (2016). Numerosity underestimation in sets with illusory contours. *Vision Research, 122*, 34–42.

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

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From "sense of number" to "sense of magnitude": The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences, 40*, e164.

Milne, J. L., Chapman, C. S., Gallivan, J. P., Wood, D. K., Culham, J. C., & Goodale, M. A. (2013). Connecting the dots: Object connectedness deceives perception but not movement planning. *Psychological Science, 24*(8), 1456–1465.

Mulligan, J., Stevenson, S., & Cormack, L. (2013). Reflexive and voluntary control of smooth eye movements. *Proceedings of SPIE - The International Society for Optical Engineering, 8651*.

Mulligan, J. B. (2002). Sensory processing delays measured with the eye-movement correlogram.

Annals of the New York Academy of Science, 956, 476–478.

Nieder, A. (2019). *A brain for numbers. The biology of the number instinct*. Cambridge, MA: MIT Press.

Park, J., DeWind, N. K., Woldorff, M. G., & Brannon, E. M. (2016). Rapid and direct encoding of numerosity in the visual stream. *Cereb Cortex, 26*(2), 748–763.

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

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., & Lucangeli, D. et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116*(1), 33–41.

Pomè, A., Caponi, C., & Burr, D. C. (2021a). Grouping-induced numerosity biases vary with autistic-like personality traits. *Journal of Autism and Developmental Disorders*, Apr 28, doi:10.1007/s10803-021-05029-1. Online ahead of print.

Pomè, A., Caponi, C., & Burr, D. C. (2021b). The grouping-induced numerosity illusion is attention-dependent. *Frontiers in Human Neuroscience - Cognitive Neuroscience,* Oct 6;*15*:745188, doi:10.3389/fnhum.2021.745188. eCollection 2021.

Roggeman, C., Santens, S., Fias, W., & Verguts, T. (2011). Stages of nonsymbolic number processing in occipitoparietal cortex disentangled by fMRI adaptation. *Journal of Neuroscience, 31*(19), 7168–7173.

Stoianov, I., & Zorzi, M. (2012). Emergence of a 'visual number sense' in hierarchical generative models. *Nature. Neuroscience, 15*(2), 194–196.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., & Draws, T. et al. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review, 28*(3), 813–826.