The Mechanisms and Neural Signature of Time-averaged Numerosity Perception

Irene Togoli[1,2], Olivier Collignon[1,3], Domenica Bueti[2], and Michele Fornaciai[1,2]

Abstract

■ The animal brain is endowed with an innate sense of number allowing to intuitively perceive the approximate quantity of items in a scene, or "numerosity." This ability is not limited to items distributed in space, but also to events unfolding in time and to the average numerosity of dynamic scenes. How the brain computes and represents the average numerosity over time, however, remains unclear. Here, we investigate the mechanisms and EEG signature of the perception of average numerosity over time. To do so, we used stimuli composed of a variable number (3–12) of briefly presented dot arrays (50 msec each) and asked participants to judge the average numerosity of the sequence. We first show that the weight of different portions of the stimuli in determining the judgment depends on how many arrays are included in the sequence itself: the longer the sequence, the lower the weight of the latest arrays. Second, we show systematic adaptation effects across stimuli in consecutive trials. Importantly, the EEG results highlight two processing stages whereby the amplitude of occipital ERPs reflects the adaptation effect (~300 msec after stimulus onset) and the accuracy and precision of average numerosity judgments (~450–700 msec). These two stages are consistent with processes involved with the representation of perceived average numerosity and with perceptual decision-making, respectively. Overall, our findings provide new evidence showing how the visual system computes the average numerosity of dynamic visual stimuli, and support the existence of a dedicated, relatively low-level perceptual mechanism mediating this process. ■

INTRODUCTION

Humans and other animals have an innate ability to rapidly estimate the number—or *numerosity*—of objects in a visual scene (e.g., Feigenson, Dehaene, & Spelke, 2004; see also the recent reviews by Visibelli, Porru, Lucangeli, Butterworth, & Benavides-Varela, 2024; Visibelli, Vigna, Nascimben, & Benavides-Varela, 2024). This ability is independent of counting and produces an approximate estimation prone to errors proportional to the number of items being estimated (e.g., Anobile, Cicchini, & Burr, 2016; but see Testolin & McClelland, 2021). Because of the properties of numerosity perception, like, for instance, it being subject to perceptual adaptation effects, numerosity has been proposed to represent a "primary" perceptual attribute (Anobile, Chicchini, et al., 2016; Burr & Ross, 2008; but see Leibovich, Katzin, Harel, & Henik, 2017, for a different account), that is, one of the fundamental building blocks of our perceptual experience. Research into numerosity perception focused especially on the judgment of items presented simultaneously in space, like arrays of dots. Numerosity, however, can be computed from several different types of stimuli. For instance, rather than objects distributed in space, numerosity can be extracted from series of events (e.g., brief flashes) presented over time (e.g., Arrighi,

Togoli, & Burr, 2014). Research in this context has shown that different types of numerical stimuli can affect each other via the process of perceptual adaptation (Anobile, Arrighi, Togoli, & Burr, 2016; Arrighi et al., 2014) and the serial dependence effect (Fornaciai & Park, 2019a), suggesting the existence of an abstract "number sense" (Anobile, Arrighi, et al., 2016; Arrighi et al., 2014). In terms of neural correlates, a numerosity-sensitive brain activity has been observed throughout the visual dorsal stream starting from early visual areas, in terms of localization (Castaldi, Piazza, Dehaene, Vignaud, & Eger, 2019; DeWind, Park, Woldorff, & Brannon, 2019; Harvey, Klein, Petridou, & Dumoulin, 2013; Roggeman, Santens, Fias, & Verguts, 2011), and from very early processing stages, in terms of timing (Fornaciai & Park, 2018; Fornaciai, Brannon, Woldorff, & Park, 2017; Park, DeWind, Woldorff, & Brannon, 2016; Temple & Posner, 1998).

Although the use of static dot-array stimuli probably remains the most common practice in numerosity perception research, the external environment and the stimuli that our sensory organs receive are rarely static. Because of the intrinsically dynamic nature of the external world, numerosity perception is likely to involve dynamic processes based on information received over a relatively extended period, rather than being based on static snapshots of the external world. Imagine, for instance, watching people walking down a street: The number of people varies from moment to moment, and whether the street

[1]Université catholique de Louvain, Louvain-la-Neuve, Belgium, [2]International School for Advanced Studies (SISSA), Trieste, Italy, [3]HES-SO Valais-Walis, Lausanne and Sion, Switzerland

would appear crowded or not would depend on how many people we see on average in a given period. In many cases, our experience of numerosity could thus be defined by the average number of items observed in a given period rather than a static snapshot of a visual scene.

An intriguing question is thus how the visual system computes and processes the average numerosity of dynamic visual events over time (or "time-averaged" numerosity), involving both the spatial and the temporal dimension. Previous research in this context shows that when presented with dynamic stimuli modulated over time (Togoli, Fornaciai, & Bueti, 2021) or series of discrete stimuli (Katzin, Rosenbaum, & Usher, 2021), humans are able to judge their average numerosity with good accuracy and precision. In line with the concept of "summary statistics" (e.g., Whitney & Yamanashi Leib, 2018; McDermott, Schemitsch, & Simoncelli, 2013), many studies indeed show that the visual system can easily extract or compute the average value of a feature modulated over time (e.g., Robitaille & Harris, 2011; de Fockert & Wolfenstein, 2009; Chong & Treisman, 2005). In terms of the properties of average numerosity perception, it has been shown that the judgment precision of average numerosity tends to increase with the sequence length (i.e., the number of individual dot arrays included in the sequence), leading to better averaging performance when more information is provided (Katzin et al., 2021). In addition, Katzin and colleagues' (2021) results provided some indications of recency effects, whereby more recent information in the sequence has a larger weight on perceptual decisions, although this effect was observed only in some participants. However, such results were obtained with sequences of relatively long (i.e., 500 msec) discrete stimuli, which likely engage mnemonic rather than perceptual processes.

In the present study, we thus aim to further address the perceptual mechanisms of average numerosity perception and the neural signature of this process. Could average numerosity be considered a perceptual feature explicitly processed and represented by the visual system? To address this possibility, we employed a classification task of the average numerosity of dynamic dot-array stimuli (Togoli et al., 2021). Namely, in each trial, the participants observed a fast sequence of multiple (3–12) individual arrays varying in numerosity, presented at 20 Hz, and were asked to judge whether their average numerosity was higher or lower compared with a reference shown before the task. The stimuli were specifically designed to give the impression of a continuous stimulus varying over time, rather than a sequence of individual, static arrays. To understand how average numerosity is computed and represented, we first assessed the weight of different arrays in the sequence (first, middle, last) in determining the judgment. This was done separately according to how many individual arrays were included in the sequence, to test whether increasing the amount of information may affect the weight of different portions of the sequence.

Moreover, we assessed trial-history effects across different stimuli and, in particular, the presence of perceptual adaptation effects. Indeed, if average numerosity is explicitly represented by the visual system, we would expect it to be susceptible to perceptual adaptation, similarly to what has been demonstrated with other types of stimuli like static dot arrays (e.g., Burr & Ross, 2008) or sequential stimuli (e.g., Arrighi et al., 2014). Adaptation is usually induced by exposure to a relatively long, sustained stimulus and entails a "repulsive" change in the perceived magnitude of a stimulus, increasing the difference between "adaptor" and "adapted" stimulus (e.g., see Kohn, 2007, for a review about the mechanisms of perceptual adaptation). For instance, after exposure to an array with a low number of dots, the perceived numerosity of a stimulus increases, and vice versa. In numerosity perception, this process has been shown to arise not only with long exposures but also with multiple briefer exposures (Aagten-Murphy & Burr, 2016). With EEG, we measured the neural signature of average numerosity processing and the adaptation effect, as well as the relationship between behavioral and neural measures, to better understand the brain processing stages linked to the representation of average numerosity. Specifically, if average numerosity is explicitly represented by the visual system, we expect EEG signals to highlight specific processing stages sensitive to the average numerosity of the stimuli. In addition, we would expect such a correlate of average numerosity to be able to predict the behavioral performance in the task. Finally, a genuine correlate of average numerosity should reflect not only the numerosity of the sequences but also the distortion of perceived average numerosity, as induced by perceptual adaptation effects across different stimuli. We thus also addressed the relationship between brain responses as a function of the stimulus in the preceding trial and the adaptation effect measured behaviorally.

METHODS

Participants

The sample tested in this study included 22 adult volunteers (mean age = 23 years, SD = 2.82 years; age range = 18–31 years; 1 male). All participants provided written informed consent before testing and received monetary compensation for their time (10€/hr). All the participants had normal or corrected-to-normal vision, were naive to the purpose of the experiment, and reported no history of neurological, attentional, or psychiatric disorders. The research protocol was approved by the ethics committee of the International School for Advanced Studies (Protocol 10035-III/13) and was in line with the Declaration of Helsinki. One participant was excluded from the data analysis because of equipment failure (i.e., missing EEG data), leaving 21 participants included in the final sample of the study. Because we did not have a specific a priori hypothesis concerning the expected effect size in terms of the

mechanisms of average numerosity, the sample size estimation was based on the secondary aim of the study to assess the mutual interference effects across numerosity and duration. The sample size of the study was thus computed with a power analysis based on two previous studies addressing the interference effects across magnitudes (Togoli et al., 2021, 2022). In the power analysis, we took an average effect size (Cohen's d) computed from previous results of $d = 0.82$. Considering a two-tailed distribution and a power of 95%, the power analysis indicated a sample size of 22 participants. Note that although the majority of participants recruited were female participants, numerosity perception is not expected to differ based on the sex or gender of the participants (Kersey, Braham, Csumitta, Libertus, & Cantlon, 2018).

Stimuli

The visual stimuli were generated using the routines of the Psychophysics Toolbox (v.3; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) in MATLAB (r2021b, MathWorks, Inc.). During the experiment, the stimuli were displayed on a 1920 × 1080 LCD monitor running at 120 Hz, encompassing a visual angle of about 48° × 30° from a distance of 57 cm. Each participant performed the experiment in a closed booth equipped with a Faraday cage, where the only source of light was the monitor screen. The stimulus design was based on Togoli and colleagues (2021) and consisted of dynamically modulated arrays of dots. Specifically, the dynamic stimuli involved a sequence of multiple briefly flashed (50 msec each; 20 Hz frequency) dot arrays modulated in average numerosity (i.e., the average amount of dots displayed across all the arrays included in a sequence). The number of dots in each array varied around the mean numerosity of the sequence selected in each trial (±50%). The numerosity of each array was computed before the presentation of the stimulus in order for the sequence to result in a specific average numerosity, while keeping the overall variance of the sequence roughly equal irrespective of the sequence length. This stimulus construction procedure and presentation frequency (20 Hz) were chosen in order for the sequence to be perceived as a fast dynamic stimulus, while avoiding too much perceptual "overlap" between consecutive arrays. With increasing presentation frequency, it is indeed likely that consecutive arrays could be increasingly "fused" together. In other words, we aimed to have stimuli that give the impression of a continuous modulation of numerosity (rather than a series of discrete, static arrays), while retaining the ability to assess the influence of different portions of the sequence on the global judgment of average numerosity. The positions of the dots were computed to avoid overlapping, considering a minimum interdot distance of 2.5 times the radius of an individual dot. The minimum interdot distance was set to scale with the radius of the dots to keep the same border-to-border minimum distance irrespective of dot size. Dot sizes and the radius of

the area encompassing them were systematically varied in a trial-by-trial fashion, in line with the procedure used in previous studies (Fornaciai et al., 2017; Park et al., 2016; DeWind, Adams, Platt, & Brannon, 2015). The radius of the dots ranged from 6 to 10 pixels (0.14°–0.24° of visual angle), whereas the radius of the area of the stimulus spanned from 200 to 400 pixels (4.76°–9.52° of visual angle). Each array in a sequence had the same area and the same dot size. The dots in each array were 50% black and 50% white (MATLAB RGB color values: black = [0 0 0], gray = [0.5 0.5 0.5], white = [1 1 1]; International Commission on Illumination, CIE L*a*b color values: black = [0 0 0], gray = [53.39 0 0], black = [0 0 0]), to keep the global contrast of the stimuli similar to the gray background. In the case of odd numerosities, the color of the exceeding dot was determined randomly (i.e., either black or white). The dots were presented at maximum contrast with the background. The luminance of the white dots was 98.17 cd/m^2, the luminance of the black dots was 0.24 cd/m^2, and the luminance of the gray background was 46.79 cd/m^2. In case of odd numerosities, the color of the exceeding dot was determined randomly. The average numerosity of each stimulus could be either 15, 21, 30, 42, or 60 dots. The number of arrays in the sequence could be 3, 4, 6, 9, or 12 arrays, corresponding to a total duration of the stimulus of 150, 200, 300, 450, and 600 msec. Both the average numerosity range and the number of arrays in the sequence were devised to be spaced in a Log2 scale (i.e., ±1 logarithmic unit around the middle value). The average numerosity range and the number of arrays were combined, resulting in 25 different stimulus types. Before the beginning of the session and before each block, we presented a reference stimulus that the participants had to memorize and use as a comparison to provide a judgment. The reference stimulus had the intermediate values of the average numerosity and number of arrays, that is, it had an average numerosity of 30 dots and was composed of six arrays (duration = 300 msec). Note that although having a mask at the end of the sequence might have reduced the potential visual persistence of the last array in the sequence, we chose not to present a mask to avoid interfering with effects across stimuli in different trials and with the visibility of the last array. A movie of the stimuli, including all the 25 combinations of average numerosity and sequence length, is available on Open Science Framework at https://osf.io/fjrnp.

Procedure

The experiment was conducted in a dark, sound-attenuated room, with each participant sitting in front of the computer screen at a distance of about 57 cm. In the testing room, the screen was the only source of light, to avoid distractions. The study involved a classification task of the average numerosity of the dynamic stimuli. EEG was also recorded throughout the session to measure the brain responses to the stimuli. At the beginning of the

session, participants were shown the reference stimulus (average numerosity of 30 dots, six arrays) that they were instructed to use to classify the stimuli in the main task sequence. The reference was displayed 10 times. During the task, participants kept their gaze on a central fixation point, and the dynamic stimuli were presented at the center of the screen. Following the offset of each stimulus, there was a 600-msec interval after which the fixation cross turned red, signaling to the participant to provide a response. The participant was then asked to judge, by pressing the left or the right arrow on the keyboard, whether the average numerosity was lower or higher compared with the memorized reference stimulus (respectively). The time available to provide a response was limited to 1200 msec. If they could not respond within this interval, the next trial started automatically. The intertrial interval (ITI) was 1100–1300 msec. The trials in which participants were not able to provide a response were excluded from the data analysis (1.1% ± 1.2%). This response deadline was added to limit the length and variability of the interval between stimuli in consecutive trials. Because one of the main goals of the study was to assess effects across different stimuli (i.e., perceptual adaptation effects), this limit was thus added to reduce the potential decay of such effects over time. Participants received no feedback about their response. The reference stimulus was presented again to the participants before the beginning of each block (displayed five times). Each participant completed 10 blocks of 100 trials, for a total of 1000 trials and 40 repetitions of each combination of average numerosity and number of arrays. Before the start of the session, subjects were familiarized with the task with ~10 practice trials.

Behavioral Data Analysis

To assess the performance in the task, we first focused on the point of subjective equality (PSE), reflecting the accuracy of numerical estimates, and the just noticeable difference (JND) and Weber's fraction (WF), reflecting the precision in the task. To compute these values, a cumulative Gaussian (psychometric) function was fitted to the proportion of "more numerous" responses as a function of the different levels of average numerosity, collapsing together the different numbers of arrays. The psychometric fitting was performed following the maximum likelihood method described by Watson (1979). From the psychometric fit, we computed the PSE as the average numerosity corresponding to chance-level responses, reflecting the perceptual match with the memorized reference. The JND was computed from the slope of the fit. As an additional measure of precision in the task, we computed the WF, which is the ratio of the JND and the PSE. This additional measure allows to assess the precision in the task while accounting for changes in the perceived magnitude of the stimuli. In terms of exclusion criteria based on behavioral performance, we set a broad cutoff

of WF ≥1. Indeed, the task was designed to be challenging, and we wanted to make sure to only exclude participants unable or unwilling to do the task. No participant was, however, excluded based on behavioral performance, as the highest WF observed in the task was ~0.57. In addition to computing the general measures of performance, we also computed the accuracy (PSE) and precision (WF) of average numerosity judgments as a function of the number of arrays included in each sequence. To do so, we performed the psychometric fit separately for the trials in which the number of arrays was 3, 4, 6, 9, or 12. The PSE and WF were computed from these fits as explained above. To assess the biases in perceived average numerosity as a function of sequence length, we used a linear mixed-effect (LME) model test on the PSEs, entering the sequence length as predictor and the subject as the random effect.

To assess the weights of different arrays in the sequence in driving the judgment of average numerosity, we employed a nonlinear regression analysis. A nonlinear analysis was chosen in this case because the weights were assessed based on the raw, binary response provided in each trial. Because of the nonlinear nature of the binary response, a nonlinear regression analysis is thus necessary in this context. The analysis was performed separately according to the number of arrays in the sequences, to further assess whether the amount of information provided affects the weighting profile of different arrays in the sequence. In the analysis, the binary response of the classification task was entered as the dependent variable, and the numerosity of the arrays along the sequence as the predictors. To have the same number of parameters across the different tests (i.e., to make the results more easily comparable across different tests), the predictors included only the first array in the sequence, the last array, and either the middle array (in the case of three-array sequences) or an average of two intermediate positions in the sequence (second and third, third and fourth, fourth and sixth, and sixth and seventh, respectively, for sequences of 4, 6, 9, and 12 arrays). The resulting beta values were first analyzed using an LME model including serial position and sequence length as predictors (and subjects as the random effect) to assess the difference across the temporal weighting profiles as a function of the number of arrays in the sequence. Then, follow-up LME tests were performed within each level of sequence length.

Finally, we assessed the perceptual adaptation effects across successive trials, measuring how the perceived average numerosity of the stimuli is affected by the numerosity presented in the preceding trial. We employed again a psychometric fitting procedure, performed separately according to the average numerosity of the preceding stimulus. To better characterize the effect of the preceding stimulus, we performed two versions of this analysis. The first one was performed by pooling together all the different lengths of the preceding stimulus. Thus, the PSE was computed only based on the average numerosity of the preceding stimulus. Because an effect like perceptual

adaptation is also expected to be modulated by the duration (or number of arrays) of the previous stimulus, we also repeated the analysis by separating the trials according to whether the preceding stimulus was short (3–4 arrays; 150–200 msec) or long (9–12 arrays; 450–600 msec). Thus, by doing so, we obtained two sets of PSE values according to the average numerosity of the preceding sequence and according to whether the preceding sequence was short or long. The PSEs obtained in this analysis were then used to compute an adaptation effect index, based on the difference in PSE between trials in which the preceding stimulus had a numerosity of 30 (the same as the reference; "PSE_{30}"), and either lower (15, 21 dots) or higher (45, 60 dots) average numerosities ("PSE_j"), according to the following formula:

$$\text{Adaptation index} = -1 \times \left(\left(PSE_J - PSE_{30} \right) / PSE_{30} \right) \times 100$$

The sign was switched to make the interpretation of the index more intuitive, that is, a negative adaptation index indicates a relative underestimation of the stimulus, whereas a positive index indicates an overestimation. The adaptation indices were then analyzed using LME models to assess the extent to which the perceived average numerosity of the current stimulus is affected by the stimulus in the previous trial.

EEG Recording and Preprocessing

Throughout the experimental session, we recorded the EEG to address the neural signature of average numerosity processing and the signature of the adaptation effect. The EEG was recorded using the Biosemi ActiveTwo system (2048-Hz sampling rate) and a 64-channel cap based on the 10–20 system layout. To monitor artifacts because of eye movements and blinks, the EOG was measured via an additional electrode attached below the left eye of the participant. The electrode offset values across the channels were usually kept below 20 μV, but occasional values up to 30 μV were tolerated.

The data preprocessing was performed offline in MATLAB (Version R2021b), using the functions of the EEGLAB (Delorme & Makeig, 2004) and ERPlab (Lopez-Calderon & Luck, 2014) toolbox. First, EEG signals were resampled to a sampling rate of 1000 Hz. Then, each combination of average numerosity and number of arrays was binned individually, for a total of 25 bins. In addition, we added bins corresponding to the combination of different numerosities and different numbers of arrays of the stimuli in the previous trial, and different numerosities in the current trial, to assess the signature of adaptation effects (125 unique combinations). The continuous EEG data were then epoched, time-locking the signal to the onset of each stimulus (i.e., the onset of the first array in each stimulus sequence). The epochs spanned from −300 to 1200 msec around the stimulus onset. The prestimulus interval (−300:0 msec) was used for baseline correction. The EEG signal was band-pass filtered with cutoffs at 0.1 and 40 Hz. To reduce artifactual activity in the data, we used an independent component analysis, aimed at removing identifiable artifacts such as eye movements and blinks. We additionally employed a step-like artifact rejection procedure (amplitude threshold = 40 μV, window = 400 msec, step = 20 msec) to further remove any remaining large artifact from the signal, leading to the exclusion of 2.9% ± 2.8% of the trials, on average (±SD). Finally, the ERPs were computed by averaging EEG epochs within each bin. ERPs were further low-pass filtered with a cutoff at 30 Hz, and smoothed with a sliding-window average with a width of 20 msec and a step of 5 msec.

ERPs Analysis

The analysis of ERPs was performed by first selecting a set of channels of interest, based on previous studies. Namely, we selected a series of four occipital channels, including O1, O2, Oz, and Iz, based on previous studies on numerosity perception (Fornaciai & Park, 2018; Fornaciai et al., 2017) and trial-history effects in magnitude perception (Fornaciai, Togoli, & Bueti, 2023; Tonoyan, Fornaciai, Parsons, & Bueti, 2022). First, we assessed numerosity-sensitive brain responses by sorting the ERPs according to the average numerosity of the stimuli, collapsing together the different durations. To assess the modulation of ERPs as a function of numerosity, we computed the linear contrast of the brainwaves (weights = [−2 −1 0 1 2], corresponding to the different levels of average numerosity) throughout the epoch. We then performed a series of one-sample t tests against zero, corrected for multiple comparisons with a false discovery rate (FDR) procedure (FDR = 0.05). Moreover, we assessed the relationship between ERPs, in terms of the linear contrast amplitude, and behavior in terms of the accuracy (PSE) and precision (JND) of average numerosity perception. To do so, we employed an LME model entering the contrast amplitude as the dependent variable, PSE and JND as predictors, and the subject as the random effect. To control for multiple comparisons, in this case, we used a nonparametric cluster-based test. Namely, we repeated the analysis across the clusters of consecutive significant time windows observed in the actual LME test, randomly shuffling the vectors of PSE and JND values at each iteration. This procedure was repeated 10,000 times, and we measured how many times we could observe similar clusters of consecutive significant time points in this simulation. The significance threshold of each test in this control analysis was set based on the lowest t value observed in the corresponding cluster of the actual analysis.

To assess the impact of adaptation effects on ERPs, we further sorted the data according to the average numerosity and number of arrays of the preceding stimulus, taking only the trials in which the middle numerosity (30 dots) was presented in the current trial. The modulation of ERPs was then tested by performing a series of LME tests across a series of small windows throughout the epoch, in a sliding-window fashion (width = 50 msec, step = 5 msec)

to increase the signal-to-noise ratio. The tests included the ERP amplitude as the dependent variable, and the average numerosity and number of arrays of the preceding stimulus (as well as their interaction) as predictors (fixed effects). The subjects were added as the random effect. Because adaptation effects are expected to be modulated by the number of arrays/duration of the preceding stimulus, we specifically looked for an interaction between the two factors as evidence for a signature of adaptation. To assess the nature of such an interaction, we further computed the effect of the preceding numerosity on ERP amplitude (i.e., the difference in amplitude corresponding to trials in which the preceding stimulus had 60 dots and 15 dots), as a function of the different number of arrays of the previous stimulus. This analysis was limited to the average ERPs within the latency window showing a significant interaction between average numerosity and number of arrays of the preceding stimulus. Again, to control for multiple comparisons, we used a nonparametric cluster-based analysis (see above).

Finally, we assessed the relationship between the bias in perceived numerosity induced by adaptation, and the extent to which adaptation modulates the ERP amplitude. To do so, we computed two corresponding measures of the adaptation effect on behavior (ΔPSE) and on ERPs (ΔERP). The ΔPSE was computed as the difference in PSE between the cases where the preceding stimulus had a numerosity either lower (15, 21 dots) or higher (45, 60 dots) than the reference (30 dots), and the case where the numerosity of the preceding stimulus was equal to the reference. The ΔERP measure was computed in a similar fashion, as a difference in ERP amplitude at each time point in the cases where the preceding stimulus

had either a lower or higher numerosity than the reference, and the case where the preceding stimulus had the same numerosity as the reference. To compute this index, we considered only the trials in which the stimulus in the current trial had 30 dots. We then used an LME model to assess the relationship between ΔPSE and ΔERP (ΔPSE $\sim$ ΔERP + (1|subj)). This analysis was performed across the latency windows where the effect of adaptation on ERPs showed a significant interaction between the average numerosity and the number of arrays of the previous stimulus. This was done to limit the tests to the most significant latency windows in terms of adaptation effect, as we predicted that genuine perceptual adaptation effects should depend on both the numerosity and the duration of the preceding stimulus.

RESULTS

In the present study, we addressed the mechanisms of time-averaged numerosity perception and its neural signature, using fast dynamic dot-array stimuli. See Figure 1 for a depiction of the experimental procedure.

First, we assessed the general performance of average numerosity judgments, computing measures of accuracy (PSE) and precision (JND and WF). Figure 2A shows the general measures of performance. On average, the perceived numerosity of the stimuli was quite accurate and precise, showing a PSE ($\pm SD$) of 31.42 $\pm$ 4.41 dots, a JND of 7.70 $\pm$ 2.34 dots, and a WF of 0.25 $\pm$ 0.11. These results show that participants are able to judge an average numerosity fairly accurately (i.e., considering the average PSE of ~31 dots compared with the reference numerosity of 30 dots) and precisely (i.e., considering the average
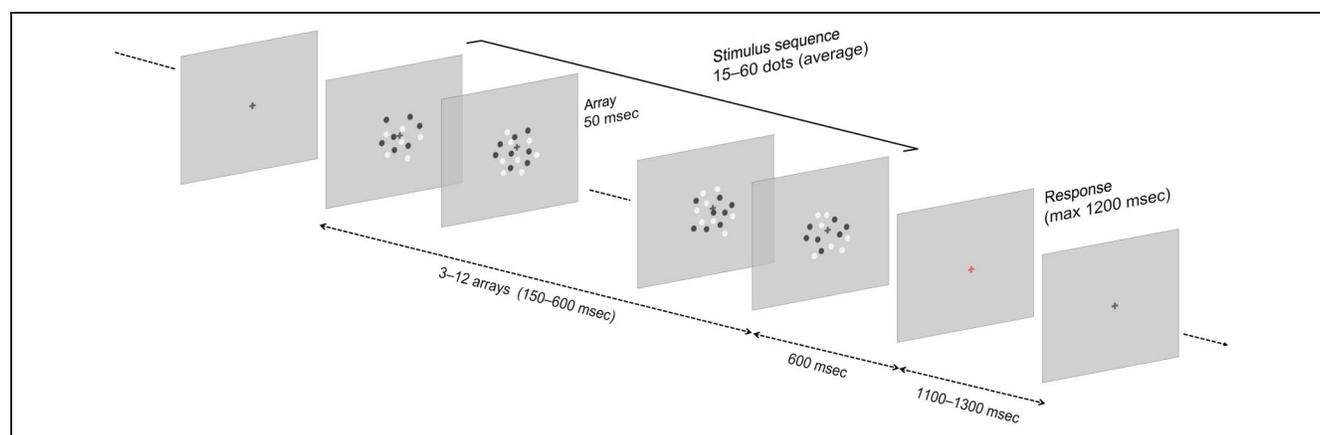
Figure 1. Stimulation procedure. The classification task involved participants watching a series of dynamic stimuli modulated in average numerosity and in the number of arrays presented, and determining whether the average numerosity in each trial was higher or lower compared with a memorized reference stimulus. The reference was presented at the beginning of the session and at the beginning of each block. Each array in the stimulus sequence included a set of black-and-white dots drawn within a circular area, presented for 50 msec. The number of arrays presented in each sequence varied from 3 (150 msec) to 12 (600 msec). The stimuli were designed to facilitate the impression of a continuous, dynamic modulation of numerosity rather than a sequence of static dot arrays. The offset of the stimulus sequence was followed by a 600-msec blank interval. After the interval, the fixation cross became red, signaling to the participant to provide a response by pressing the appropriate key on a standard keyboard. The maximum time allowed to provide a response was 1200 msec, after which the next trial started automatically. The proportion of trials in which participants did not provide a response was mean = 1.1%, SD = 1.2%. The ITI was 1100–1300 msec. The stimuli are not depicted in the scale.
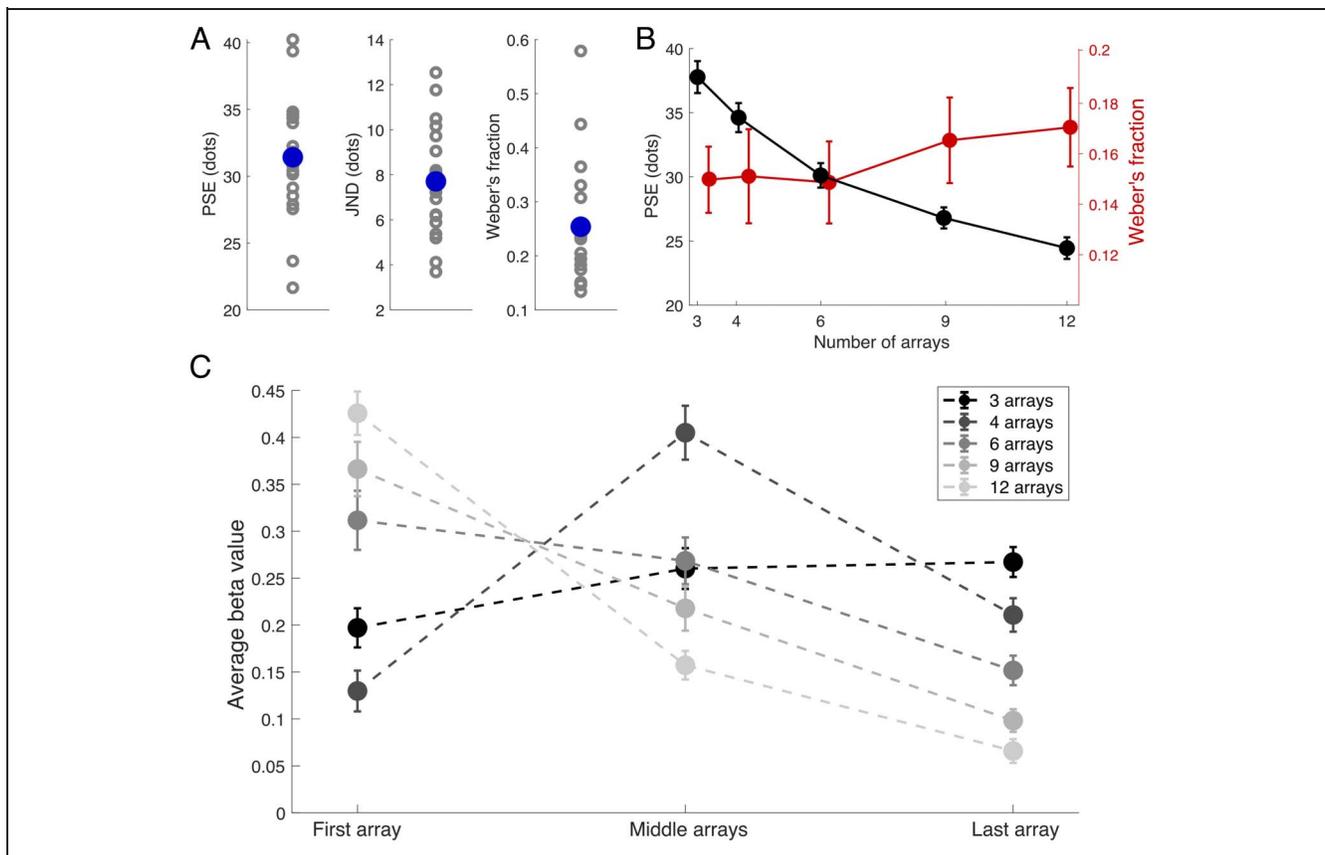
Figure 2. Behavioral results. (A) General measures of performance, including the PSE as a measure of accuracy, and the JND and WF as a measure of precision. (B) Biases in perceived average numerosity as a function of the number of arrays presented, and modulation of the precision of judgments (WF). (C) Temporal weighting profile, reflecting the weights of the first, middle, and last array(s) in the sequence in driving the overall judgment of average numerosity, separately for the different numbers of arrays composing the sequence. Error bars are *SEM*.

WF), in line with previous studies (Katzin et al., 2021; Togoli et al., 2021).

Second, we assessed whether the amount of information provided, that is, the number of arrays presented in the sequence, affects the accuracy and/or the precision of judgments. To address this question, we computed the PSE and WF separately as a function of the number of arrays of the sequences (Figure 2B). The results showed substantial biases in perceived numerosity according to the length of the sequence, with a pattern of both under- and overestimation. Indeed, when the duration/number of arrays was smaller than 300 msec/6 arrays (i.e., the middle values of the range, corresponding to the reference), PSEs were higher (37.78 ± 1.24 and 34.62 ± 1.13, respectively, for three and four arrays), showing an underestimation of perceived average numerosity (i.e., a higher number of dots is necessary to perceptually match the reference). Conversely, PSEs were lower when the sequence was longer than the 300 msec/6 arrays (26.80 ± 0.82 and 24.44 ± 0.84, respectively, for 9 and 12 arrays), showing an overestimation. An LME regression model (PSE ~ number of arrays + (1|subj)) showed a significant modulation of PSE as a function of sequence length (adjusted R^2 = .85, β = −1.43, t = −17.69, p < .001). On the other hand, the WF (shown in red in

Figure 2B) showed a small but significant increase (i.e., worsening of performance) with sequence length (LME test on WFs; adjusted R^2 = .76, β = 0.002, t = −2.27, p = .025).

Furthermore, we addressed the temporal weighting profile of the dynamic stimuli—that is, the extent to which arrays in different positions along the sequence contribute to the perception and judgment of average numerosity. To do so, we employed a nonlinear regression analysis, entering the binary classification response as the dependent variable and the numerosity of arrays along the sequence as predictors. This analysis was performed separately as a function of the number of arrays in the sequence, to further assess whether the amount of information presented affected the weight of different portions of the sequence. To ensure that the results concerning different sequences are comparable, we included the same number of parameters (array position) in the nonlinear regression: the first and the last array, and either the middle array (in the case of three arrays) or the average of two intermediate arrays. The beta values obtained with this analysis reflect the extent to which arrays in different positions contribute to the classification judgment (i.e., the higher the beta value, the higher the weight on perceptual decisions).

The results, shown in Figure 2C, show a clear difference in the temporal weighting profile as a function of the sequence length. An LME test on the beta values, with factors position and number of arrays, showed indeed a significant interaction between the two factors (adjusted $R^2 = .37$, $\beta = -0.025$, $t = -10.93$, $p < .001$). Interestingly, considering the pattern across the three sequence positions (first, middle, and last; Figure 2C), the change in beta values showed different directions according to the number of arrays presented. Follow-up LME tests on beta values performed individually within each level of sequence length (i.e., 3–12 arrays) first showed a significant difference in the three-array sequence, in the positive direction, that is, higher beta values at the end of the sequence ("recency" effect; adjusted $R^2 = .08$, $\beta = 0.035$, $t = 2.55$, $p = .013$). With four arrays, the middle position showed a higher weight, but the overall difference was not significant (adjusted $R^2 = .03$, $\beta = 0.040$, $t = 1.73$, $p = .09$). Conversely, with sequences longer than four arrays, the tests showed significant differences in the negative direction, reflecting a higher weight of the first array in the sequence compared with the last ("primacy" effect; adjusted $R^2 = .24$, $\beta = -0.080$, $t = -4.59$, $p < .001$, adjusted $R^2 = .53$, $\beta = -0.134$, $t = -8.52$, $p <$

.001, and adjusted $R^2 = .73$, $\beta = -0.180$, $t = -13.09$, $p < .001$, respectively, for 6, 9, and 12 arrays).

Then, we assessed the presence of trial-history effects across different stimuli and, in particular, perceptual adaptation effects (Aagten-Murphy & Burr, 2016). To address this possibility, we computed the perceived average numerosity of the stimuli as a function of the average numerosity preceding them. The results of this analysis are shown in Figure 3. First, we observed a robust modulation of the perceived numerosity of the stimuli (PSE) as a function of the preceding numerosity (Figure 3A), in line with an adaptation effect. To better assess the bias induced by the previous stimuli, we computed an adaptation effect index reflecting how much the perception of the stimuli changes when the previous stimulus had either a lower or higher numerosity than the intermediate, reference numerosity (Figure 3B). What we observed was a relative overestimation when the preceding stimulus had fewer dots than the middle numerosity level (15, 21 dots), and a relative underestimation when the preceding stimulus had more dots (45, 60 dots). The biases ranged from 6% to about −8%. An LME test showed a significant difference in the adaptation effect as a function of the preceding numerosity (adjusted $R^2 = .61$, $\beta = -0.306$, $t = -6.37$,

Figure 3. Perceptual adaptation effects. (A) PSE values as a function of the preceding numerosity. (B) Average adaptation effect indices as a function of the preceding numerosity. (C) PSEs as a function of the preceding numerosity, separately for the cases where the previous stimulus had a short (3, 4 arrays) or long (9, 12 arrays) sequence. (D) Adaptation effect indices as a function of preceding numerosity and sequence length (short vs. long). Error bars are *SEM*.
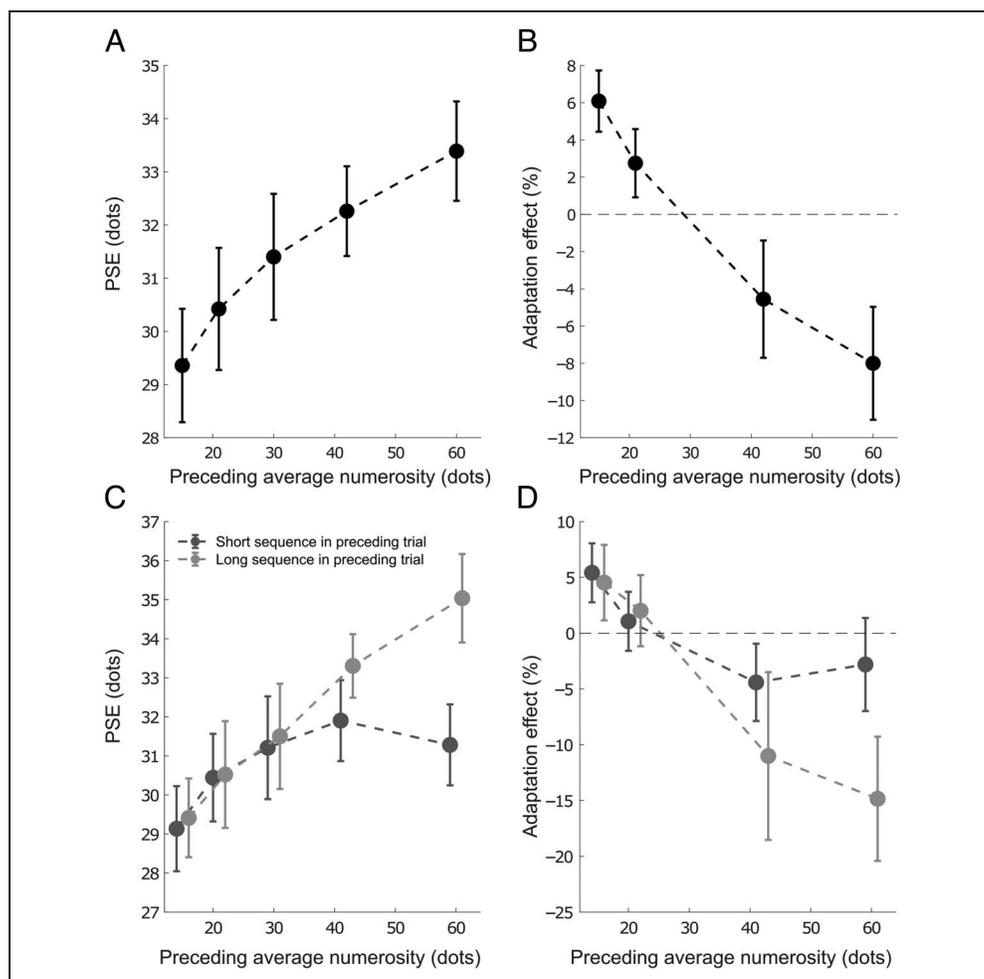
$p < .001$). In addition, we further tested whether the number of arrays/duration of the previous stimulus could modulate the effect. The PSEs shown in Figure 3C indeed suggest that when the preceding stimulus was longer (9, 12 arrays), the bias in perceived numerosity was stronger compared with a shorter sequence (3, 4 arrays), although mostly at larger numerosities. To assess the pattern of adaptation effects computed as a function of the number of arrays of the preceding stimulus (Figure 3D), we used an LME model test with factors numerosity and number of arrays of the preceding stimulus. The results showed a significant interaction between the two factors (adjusted $R^2 = .60$, $\beta = -0.271$, $t = -2.44$, $p = .016$), suggesting a stronger effect when the preceding sequence was longer, especially at larger numerosities. Further LME tests, performed individually for the two sequence lengths, showed, however, that in both cases, the effect is statistically significant (adjusted $R^2 = .71$, $\beta = -0.179$, $t = -3.51$, $p < .001$, and adjusted $R^2 = .70$, $\beta = -0.450$, $t = -5.53$, $p < .001$, respectively, for short and long sequences).

After characterizing the properties of average numerosity perception at the behavioral level, we addressed its neural signature. First, we assessed the brain responses sensitive to average numerosity, which are shown in Figure 4A. To do so, we sorted the ERPs according to corresponding average numerosity and computed the linear contrast of ERPs as a measure of sensitivity to numerical information. The linear contrast was then tested with a series of one-sample t tests against zero, corrected for multiple comparisons with an FDR procedure (FDR = 0.05). The results showed a significant numerosity-sensitive activity in four distinct latency windows, starting very early after the onset of the sequence: 20–55 msec, $t(20) \geq 3.10$, adjusted $p \leq .039$; 115–155 msec, $t(20) \leq -3.35$, adjusted $p \leq .026$; 190–230 msec, $t(20) \geq 2.99$, adjusted $p \leq .049$; and 430–595 msec, $t(20) \leq -2.98$, adjusted $p \leq .049$. Topographic plots of the average contrast amplitude within these significant windows show patterns of activity mostly at central occipital scalp locations (rightmost columns in Figure 4A).

What are the processing stages contributing to the judgment of average numerosity? To answer this questions, we focused on the relationship between brain activity and behavioral performance, in terms of the accuracy (PSE) and precision (JND) of numerical judgments. We thus performed LME tests on the contrast amplitude throughout the poststimulus interval, entering PSE and JND as predictors and the subjects as the random effect. The results first showed that the amplitude of numerosity-sensitive responses can be predicted by the PSE at two latency windows: 580–625 msec (adjusted $R^2 = .62$–.65, $\beta \leq -0.022$, $t \leq -2.11$, $p \leq .049$), and 635–710 msec (adjusted $R^2 = .61$–.87, $\beta \leq -0.026$, $t \leq -2.13$, $p \leq .047$). Both these windows show a negative relationship between amplitude and PSE, suggesting that higher amplitudes are associated with lower PSE. Second, the results also showed a significant relationship between the contrast amplitude and JND at

three latency windows: 410–455 msec (adjusted $R^2 = .62$–.65, $\beta \geq 0.045$, $t \geq 2.22$, $p \leq .039$), 470–500 msec (adjusted $R^2 = .63$–.67, $\beta \geq 0.051$, $t \geq 2.36$, $p \leq .030$), and 650–715 msec (adjusted $R^2 = .66$–.86, $\beta \leq -0.045$, $t \leq -2.25$, $p \leq .037$). The two earlier windows show a positive relationship between amplitude and JND, suggesting that higher amplitudes are associated with poorer precision. Instead, at the later window (650–715 msec), the negative relationship suggests that the higher the sensitivity of responses to numerosity, the higher the precision (i.e., the lower the JND). To control for multiple comparisons, the significant windows observed in the LME tests were assessed with a nonparametric cluster-based test (see the Methods section). All the cluster p values resulted to be $<.001$.

Furthermore, we assessed the EEG signature of the perceptual adaptation effect across stimuli in different trials (Figure 4B). In this context, we took the ERPs corresponding to the presentation of 30 dots in the current trial and sorted them according to the average numerosity and sequence length of the preceding stimuli. Figure 4B shows a representative set of the ERPs reflecting the responses to the "current" (i.e., average numerosity of 30 dots) stimulus as a function of the preceding stimulus. In the analysis, however, we considered the full set of 25 unique combinations of average numerosity and number of arrays of the preceding stimulus. We then performed LME tests on the ERP amplitude throughout the poststimulus interval, adding the numerosity and the sequence length of the preceding stimulus as predictors, and the subjects as the random effect. Because the effect of adaptation is expected to increase with the duration (or number of arrays) of the previous stimulus, a neural signature of the effect is expected to show a similar interaction between the two factors. The results first show a significant effect of numerosity on the ERP amplitude at two latency windows: 245–330 msec (adjusted $R^2 = .55$–.66, $\beta \leq -0.025$, $t \leq -2.33$, $p \leq .020$) and 670–725 msec (adjusted $R^2 = .14$–.15, $\beta \leq -0.021$, $t \leq -1.97$, $p \leq .049$). The topography of activity within these two windows (contrast amplitude based on the different levels of average numerosity of the previous stimulus; top two plots in the rightmost part of Figure 4B) showed a less unitary distribution compared with the effect of average numerosity shown in Figure 4A. The activity in this case encompassed both occipital and occipito-parietal scalp locations, especially in the earlier window (245–330 msec). Moreover, the effect of sequence length was observed at two similar latency windows: 240–350 msec (adjusted $R^2 = .51$–.67, $\beta \leq -0.002$, $t \leq -1.97$, $p \leq .048$) and 665–735 msec (adjusted $R^2 = .12$–.16, $\beta \leq -0.002$, $t \leq -2.07$, $p \leq .039$). The topographic distribution of activity corresponding to the effect of the previous stimulus length (bottom two images in the rightmost part of Figure 4B) showed again peaks at both occipital and occipito-parietal scalp locations, especially at the earlier window (240–350 msec). Additional significant effects of the sequence length were observed at other latencies
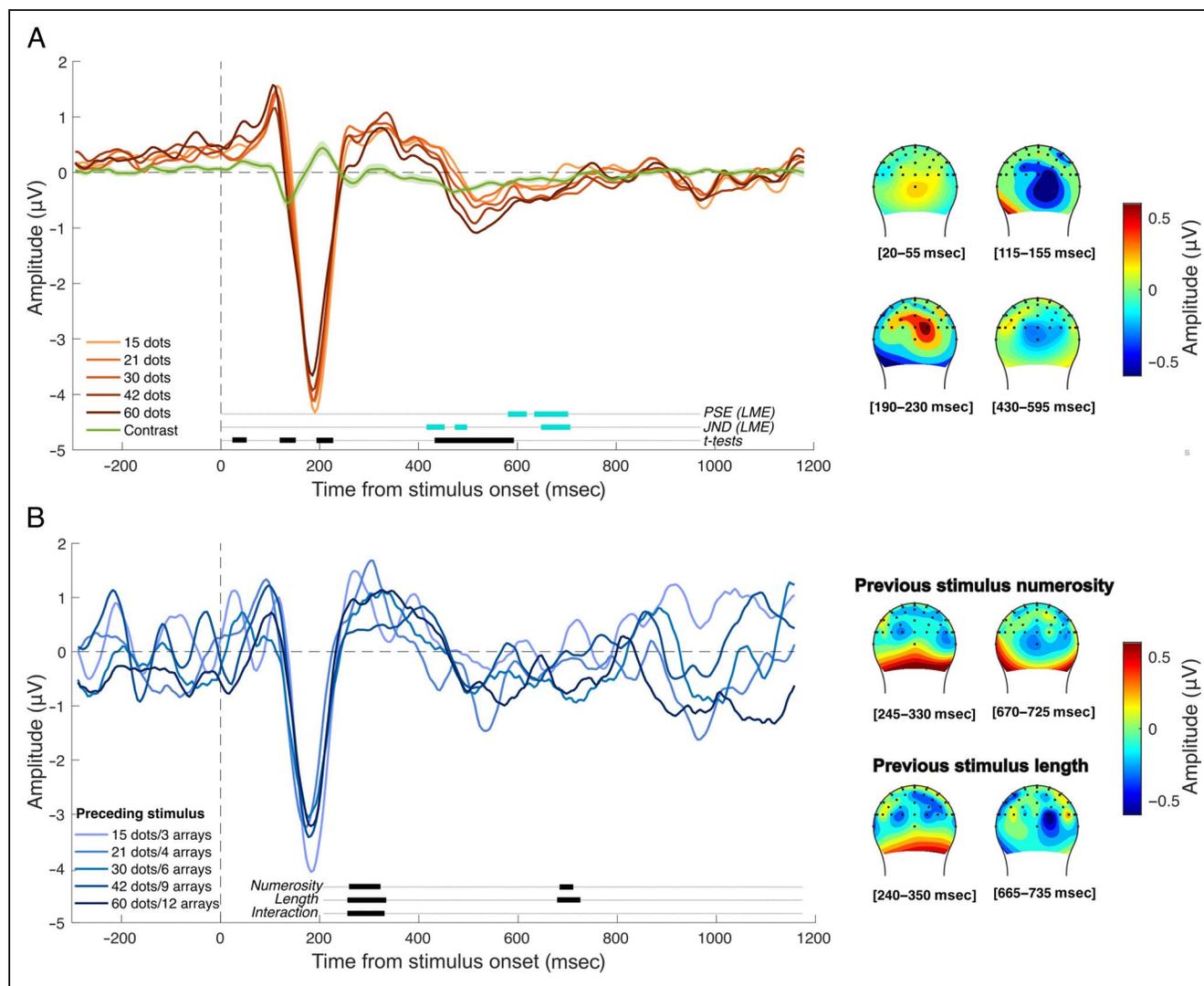
Figure 4. ERPs reflecting average numerosity and adaptation effects. (A) ERPs sorted according to average numerosity. The green wave represents the linear contrast of ERPs, reflecting the extent to which the ERPs are modulated by numerosity. The green shaded area around the contrast wave represents the *SEM*. Lines at the bottom of the plot show the significance of different tests. The thick black lines indicate the latency windows where the linear contrast is significantly different from zero (FDR-corrected one-sample *t* tests). The cyan lines instead indicate the results of the LME test relating the ERPs to the behavioral performance. Namely, the lines show the latency windows for which we observed a significant relationship between the amplitude of the linear contrast and either the PSE or the JND. The topographic plots in the rightmost part of the figure show the distribution of activity at posterior scalp locations, averaged across the four latency windows where we observed significant numerosity-sensitive responses (i.e., significant one-sample *t* tests). (B) Representative ERPs reflecting the effect of the preceding stimulus on the responses to the intermediate average numerosity in the current trial (30 dots). For the sake of clarity, only the pairwise corresponding combinations of average numerosity and number of arrays (length) are shown in the plot. However, the analysis was performed on the full set of 25 unique combinations of average numerosity and length of the preceding stimulus. The lines at the bottom of the plot show the results of the LME tests, reflecting the significance of the average numerosity, length, and their interaction in driving the ERPs. The topographic plots show the distribution of the linear contrast amplitude considering ERPs sorted according to either the average numerosity of the previous stimulus (top) or the length of the previous sequence (bottom). The linear contrast was computed as the average across the two latency windows showing a significant effect of numerosity and length of the previous stimulus.

where no other effect was found. Because an effect of the preceding sequence length in isolation (i.e., not coupled with an effect of numerosity or an interaction) is difficult to interpret in this context, we did not consider such latency windows for further analysis. More importantly, we observed a significant interaction between numerosity and sequence length at 240–345 msec (adjusted R^2 = .51–.67, β ≥ 0.001, $t ≥ 2.04$, $p ≤ .042$). These significant latency windows were further assessed with nonparametric

cluster-based tests to control for multiple comparisons. All the cluster *p* values were <.001.

To better understand the nature of this interaction, we assessed how the adaptation effect is modulated by the length of the preceding sequence, within the 240- to 345-msec latency window (i.e., the latency window showing a significant interaction between average numerosity and sequence length). We thus computed a measure of the effect of numerosity in the preceding trial on ERPs
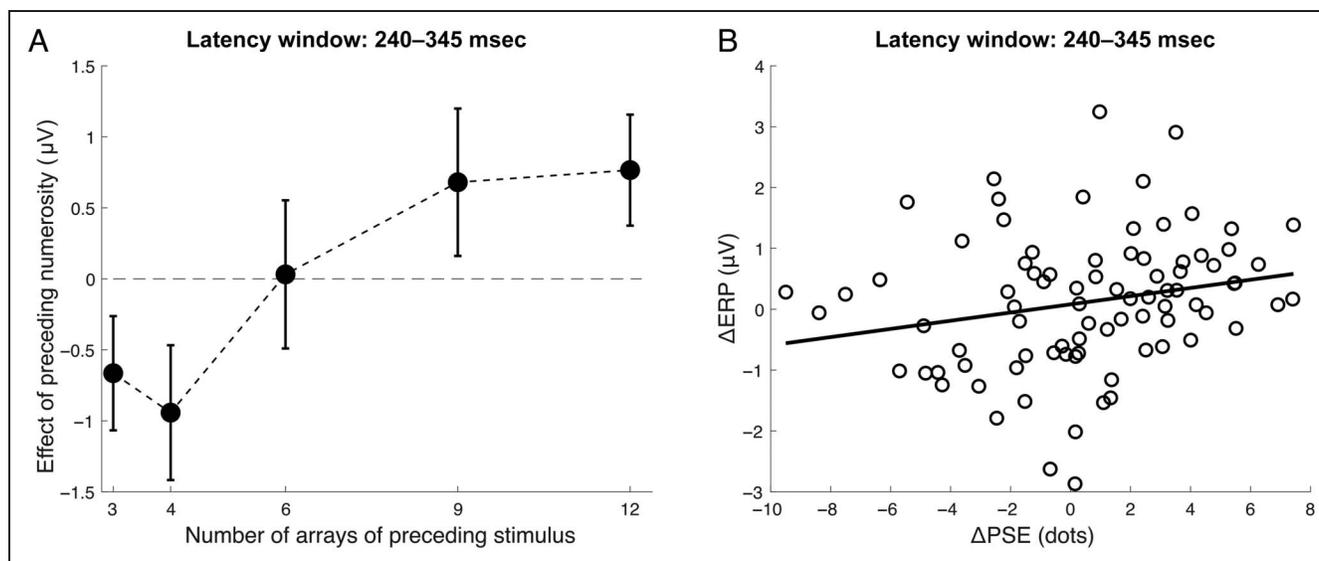
Figure 5. Interaction between numerosity and length and relationship between the behavioral and EEG effect of adaptation. (A) Effect of the preceding numerosity as a function of the length of the preceding sequence. Error bars are *SEM*. (B) Effect of adaptation measured behaviorally (ΔPSE) as a function of the effect measured with EEG (ΔERP). The ΔPSE and ΔERP measures were computed as the difference in PSE or ERP amplitude between either lower (15, 21 dots) or higher (45, 60 dots) numerosities of the preceding stimulus and the intermediate numerosity of the range (30 dots). The line represents a linear fit to the data, to show the general trend. The data were, however, analyzed with an LME model. Both columns report data averaged within the latency window where we observed a significant interaction between the numerosity and length of the preceding stimulus (240–345 msec after stimulus onset).

(i.e., amplitude of responses to 60 dots minus responses to 15 dots) and plotted it as a function of the length of the preceding stimulus (Figure 5A). The results show indeed a clear modulation, with a significant effect of the preceding stimulus length in modulating the impact that numerosity has on ERPs (LME test; adjusted R^2 = .08, β = 0.193, t = 3.14, p = .002). This test shows, however, a low R^2, which may be because of the difficulty of the linear model to capture a potentially nonlinear effect (see the pattern in Figure 5A).

Finally, we assessed the relationship between the behavioral and EEG signature of the adaptation effect. To do so, we computed two corresponding measures: ΔPSE and ΔERP. These measures were computed as the difference in either PSE or ERP amplitude between the lower (15, 21 dots) and higher (45, 60 dots) numerosity levels of the previous stimulus and the intermediate level (30 dots). The ΔERP was computed in a similar fashion, taking all the trials in which the current stimulus had an average numerosity of 30 dots and computing the difference in ERP amplitude as a function of the preceding stimulus (15–21 vs. 45–60 dots). This measure was computed as the average amplitude within the latency window showing an interaction between numerosity and length of the preceding stimulus (240–345 msec). Figure 5B shows the general trend of these data. To assess the relationship between the two measures, we performed an LME test, entering the ΔPSE as the dependent variable, ΔERP as the predictor, and the subject as the random effect. The results showed a significant relationship, whereby the ΔPSE, indexing the behavioral effect of adaptation, can

be predicted by the ΔERP, which indexes the changes in ERP amplitude because of the preceding stimulus (adjusted R^2 = .56, β = 0.813, t = 2.26, p = .026). The relationship is positive, suggesting that the higher the change in ERP amplitude within the 240- to 345-msec window, the higher the behavioral effect of adaptation.

DISCUSSION

In the present study, we addressed the mechanisms mediating the perception of average numerosity over time using fast dynamic stimuli, and the neural signature of this process. Although numerosity perception is most often investigated using static stimuli, the external environment is rarely static. Numerosity information in the environment indeed can vary over time, like, for instance, when observing a scene in which humans or animals move in and out of our visual field. For example, whether a street is perceived as crowded with people or not would depend on how many people we see on average, in a given period, and not on individual "snapshots" of the scene. In many cases, the average numerosity across different points in time or different samples of a visual scenes might thus be more important than "local" information present in a given location or at a given time. Despite its importance, how the brain computes and represents the approximate average of numerosity over time remains mostly unclear.

So far, only a few studies addressed this process, providing initial evidence for the existence of dedicated brain mechanisms supporting the averaging of numerosity information over time (Katzin et al., 2021; Togoli et al.,

2021). Our results provide new evidence showing that (1) the weighting profile of information along the sequence depends on the total amount of information provided, (2) average numerosity is subject to perceptual adaptation effects across trials, and (3) average numerosity and the adaptation effect show robust neural (EEG) signatures, with activity at specific latency windows predicting the behavioral performance and effects.

First, in terms of general performance, our data show that despite the challenging nature of the task, the participants were able to judge average numerosity with good accuracy and precision. This is important as we did not control for how well the reference presented before each block was memorized. However, even if the reference was not memorized, the stimulation range was structured so that the middle levels of the stimuli (both average numerosity and sequence length) corresponded to the reference. Thus, even if participants judged the stimuli based on an internal reference constructed according to the stimuli, they would still be able to classify the average numerosity correctly. The results indeed show that irrespective of whether the initial reference was memorized or not, participants were able to classify the stimuli with good accuracy.

Moreover, we observed a few differences compared with previous results. Namely, whereas Katzin and colleagues (2021) reported that the averaging precision increases as a function of the sequence length, here, we observed a small but significant reduction in precision (i.e., WF). In addition, we also observed biases in accuracy depending on the sequence length, with both under- and overestimation of average numerosity according to whether the sequence was shorter or longer than 300 msec (i.e., the middle duration of the range or the reference duration), respectively. Although previous studies often show only a weak influence of duration on other nontemporal magnitude dimensions like space and numerosity (e.g., Dormal, Seron, & Pesenti, 2006; Droit-Volet, Clément, & Fayol, 2003; but see Javadi & Aichelburg, 2012), these biases seem in line with an effect of duration on perceived numerosity (e.g., Lambrechts, Walsh, & van Wassenhove, 2013; Javadi & Aichelburg, 2012; Walsh, 2003). In this context, one key factor determining the effects across magnitude dimensions is indeed the nature of the stimuli used. When using static stimuli (i.e., a single dot array presented for a given time), duration is usually the feature most easily affected by other dimensions. Conversely, we have recently demonstrated that using dynamic stimuli instead reduces this asymmetry (Togoli, Bueti, & Fornaciai, 2022), making duration much more effective in biasing perceived numerosity (see also Lambrechts et al., 2013). This difference might be because of the processing time-course of different dimensions. Namely, with static stimuli, although numerosity information is available from stimulus onset and likely processed in a relatively short time (~250 msec; Fornaciai et al., 2017; Park et al., 2016), duration has to fully unfold before it could be represented

by the brain. This likely makes it difficult for a duration representation to affect numerosity. With dynamic stimuli based on average numerosity, instead, both numerosity and time information similarly unfold throughout the interval, thus making it easier for duration to affect numerosity before it is represented in a more stable fashion. The biases observed here as a function of the sequence length are thus consistent with our previous results showing an effect of duration on numerosity with dynamic stimuli (Togoli et al., 2022). These biases additionally suggest that our dynamic stimuli were perceived as a continuous sequence rather than a series of discrete stimuli. Indeed, it was the entire duration of the sequence—not the duration of individual arrays (which was constant)—that affected the overall perceived average numerosity. When it comes to the difference between our results and those of Katzin and colleagues (2021), in terms of precision as a function of sequence length, a possible explanation is the difference in the stimuli used. Indeed, while they used sequence of relatively long stimuli (500 msec), which are likely perceived as different, discrete arrays, we purposely used a fast modulation of numerosity (20 Hz) that is likely perceived as a single continuous stimulus. The difference in how the sequence length modulates accuracy and precision may thus suggest that computing the average of continuous versus discrete sequences may engage different mechanisms, more heavily relying for instance on perceptual versus mnemonic resources.

Despite the relatively short duration of the stimulus sequences, we observed systematic perceptual adaptation effects. Numerosity adaptation effects are usually observed with much longer exposures, around a few seconds (e.g., Grasso, Petrizzo, Caponi, Anobile, & Arrighi, 2022; Arrighi et al., 2014; Burr & Ross, 2008), whereas shorter stimuli usually lead to attractive "serial dependence" effects (e.g., Fornaciai & Park, 2019b). However, there is also evidence that repulsive adaptation can be observed even with very short exposures, at least in the case of ambiguous or masked stimuli (Fornaciai & Park, 2019b, 2021; Glasser, Tsui, Pack, & Tadin, 2011). In addition, numerosity adaptation effects have been observed after repeated presentations of brief arrays of dots (Aagten-Murphy & Burr, 2016), making our effect in line with previous results. The emergence of perceptual adaptation effects in this context supports the existence of a perceptual mechanism specifically dedicated to the representation of average numerosity.

Importantly, we observed different temporal weighting profiles according to the sequence length. Although sequences up to four arrays (or 200 msec) showed either a recency effect (i.e., higher weight given to later information) or a flat profile, longer sequences showed primacy effects (i.e., higher weight given to earlier information). This pattern is particularly interesting, as it suggests a capacity limit for the computation of average numerosity. Namely, with a relatively short sequence (i.e., 150 msec), the visual system is able to use the information provided

with similar weights, giving, however, more relevance to more recent information. When the sequence gets longer (i.e., 6–12 arrays, 300–600 msec), the visual system weights further information increasingly less compared with the early information. Comparing these temporal weighting profiles with results obtained in other feature dimensions, the primacy effect observed with long sequences is similar to the one observed by Hubert-Wallander and Boynton (2015) on the average spatial position. Hubert-Wallander and Boynton (2015) also tested and compared different dimensions, such as size, face identity, and motion, which mostly showed recency effects. Our results thus add to those previous observations in suggesting the existence of different, dimension-specific mechanisms for the extraction of summary statistics. Something to take into account when interpreting the behavioral results is that we did not present a mask at the end of the sequences. Thus, the last arrays in the sequence might have perceptually persisted after the off-set of the sequence and throughout the response phase, potentially biasing the judgment. However, we chose not to include a mask at the end of the sequences to avoid interfering with effects across stimuli in different trials and with the visibility of the last array in the sequence. Although a prolonged visual persistence of the last array is a possibility that we cannot exclude, the temporal weighting profile shows that the last array did not necessarily play a more pronounced role in the decision; conversely, it had instead a diminishing weight as a function of the length of the sequence. This shows that, even if visual persistence is possible, it did not affect the judgment of average numerosity.

What is the mechanism underlying the peculiar temporal weighting profiles observed in average numerosity? The weighting profile may be because of two possible types of capacity limits, either cognitive (i.e., mnemonic) or perceptual. First, the capacity limitation leading to different weighting profiles may be because of the limits of working memory (WM) encoding, usually considered to be around three to four items (Awh, Barton, & Vogel, 2007; Alvarez & Cavanagh, 2004; Luck & Vogel, 1997). This limit of WM encoding might be consistent with our results, showing that information exceeding the third or fourth array in the sequence is given increasingly less weight. However, one might expect that with longer sequences more recent information would be encoded in WM replacing older information, leading to a recency effect (as found by Katzin et al., 2021). A recency effect was, however, observed only at the shorter sequence. As mentioned above, differently from previous studies (Katzin et al., 2021), our sequences were designed to appear like continuous dynamic stimuli rather than a series of discrete stimuli. Considering the more continuous nature of our stimuli, it is thus unlikely that the results could be explained by the capacity limits of WM in terms of the storing of discrete items. Therefore, a second and, perhaps, more plausible explanation for the weighting profiles is a

perceptual capacity limitation, based on the limits of temporal integration of the visual system. Indeed, the ability of the visual system to integrate information during the presentation of a stimulus is limited by several factors, including for instance rapid adaptation of responses and the correlation in response fluctuations (e.g., see, for instance, Goris, Ziemba, Movshon, & Simoncelli, 2018). Because of these constraints, the limit of temporal integration in visual cortex has been estimated to be around 150–300 msec (e.g., Goris et al., 2018; Burr & Santoro, 2001). After this time, the temporal integration (or summation) of signals—and its benefits in visual perception—tend to plateau. The timing of our weighting profiles seems consistent with this time course, as the turning point after which the weights start to decrease is around 200 msec (four arrays). A possibility is thus that the summary statistics of a relatively brief, continuous sequences may rely of the ability of the early visual system to track and integrate information over time, rather than by the encoding of information in visual WM. However, whether the duration of the sequence or the number of arrays presented was the critical limiting factor responsible for the weighting profile remains unclear. Indeed, addressing this point would require testing different presentation frequencies and different sequence durations, so that similar amounts of information are provided in a shorter or longer interval. Considering the fast presentation rate of the sequences and their lack of clearly defined, discrete arrays (from a perceptual standpoint), we believe that duration is more likely the limiting factor in this context. This interpretation is, however, speculative, and addressing this point thus remains an open question for future studies. Finally, additional possible factors limiting the weight of the latest arrays in the sequences are the motivation of participants to attend the full duration of the stimuli and the possibility of predictions based on the first few arrays of the sequences. First, participants could have lacked the motivation to attend the full stimulus sequence, selecting a response based on the first few arrays and disregarding further information. Because of the relatively brief duration of the stimuli and their fast presentation rate, this is, however, less likely. Indeed, a decrease in the weight of the last array can be observed even with a 300-msec stimulus. At such short durations, it is difficult to assume that participants were not motivated enough to attend the stimulus, especially considering the good accuracy and precision of the judgments. In addition, the bias provided by duration on average numerosity (see Figure 2B) supports the idea that participants attended the whole sequence. Otherwise, no difference in PSE would be expected for durations longer than 300 msec. Second, the weights of later arrays might be reduced by implicit predictions about their numerosity, based on statistical learning. However, although possible, predictions based on the first few arrays would likely be unreliable, because of the overlap between the numerosities of individual arrays. Indeed, within each level of average numerosity,

each individual array could fluctuate ±50% around the average. This means, for example, that an array of 21 dots could be presented within sequences resulting in an average numerosity of 15, 21, 30, or 42 dots—so almost the entire range. Because of this overlap, accurate predictions could be made only based on the most extreme values of numerosity, which in turn would predict a much more variable performance, at odds with what we observed.

Beyond the behavioral results mentioned above, the EEG results provide novel neural insights into the mechanisms involved in the computation of average numerosity. First, we observed numerosity-sensitive brain responses at occipital channels (Iz, Oz, O1, O2) starting as early as 20–50 msec after stimulus onset and continuing throughout the presentation of the stimulus, with the more sustained activity emerging around the offset of the longer stimulus sequences (430–595 msec). The initial onset of such responses is extremely early in terms of latency, but it may be consistent with early responses to numerosity observed in previous studies (e.g., ~75 msec in Park et al., 2016, 50–80 msec in Fornaciai et al., 2017; in both cases with signals measured at Oz). Such an early onset may thus be consistent with the responses of early visual areas such as V1, V2, and V3 (e.g., Fornaciai et al., 2017; Foxe & Simpson, 2002). Because of the nature of our analysis, there is, however, the possibility that such very early responses might be driven by other dimensions of the stimuli correlating with numerosity, such as the density of the items in the array. In addition, because of the dynamic nature of our stimuli, some of the observed brain responses might be driven by phenomena like apparent motion of dots in consecutive arrays. For instance, less numerous arrays might have favored the emergence of stronger apparent motion effects because of their (on average) sparser distribution of dots (i.e., an individual dot might have appeared as "displaced" to a greater distance across successive arrays, increasing the implied speed). The negative deflection observed at around 200 msec (N2 component) shows, for instance, a larger negative amplitude for smaller numerosities. Because the N2 component has been previously associated to motion processing (Tonoyan et al., 2022; Hoffmann, Unsöld, & Bach, 2001; Hoffmann, Dorn, & Bach, 1999), such responses might be consistent with apparent motion. Considering the random nature of the stimuli and their different numerosities, however, it is difficult to quantify the role that apparent motion might have played in driving brain responses. A dedicated paradigm would thus be needed to disentangle the potential role of the different dimensions of our dynamic dot-array stimuli, including apparent motion, in driving such early brain responses. Besides this, it is interesting to note that the ERPs did not show any clear modulation at 20 Hz, reflecting responses evoked by the individual arrays in the sequences. One possibility explaining the lack of a 20-Hz modulation is the overall broad temporal frequency of the stimuli. Indeed, the sequences likely entailed different

local temporal frequencies, varying depending on the relationships between dots in consecutive arrays. The lack of responses to the individual arrays also supports the idea that our sequences were perceived and processed as unitary dynamic stimuli, rather than series of individual stimuli.

To better understand how the brain responses evoked by the stimuli are related to the perception and judgment of average numerosity, we also assessed the relationship between the amplitude of ERPs and the behavioral measures of accuracy (PSE) and precision (JND). The results show a series of latency windows whereby the behavioral performance can predict the amplitude of ERPs, clustered around the offset of the longer stimulus sequences (~400–700 msec). This suggests that the processes occurring in this large latency window may be related either to the computation of average numerosity based on responses integrated during the stimulus presentation or, alternatively, to perceptual decision-making involved with the judgment of average numerosity.

Moreover, the adaptation effect shows instead an earlier signature compared with the relationship with PSE and JND, with occipital responses at ~240–340 msec reflecting the interaction between the average numerosity and the length of the preceding sequence. First, such a localized window suggests that adaptation did not affect the general visual responses to the stimuli or the perceived numerosity of the individual arrays in the sequence, but more likely the computation of average numerosity from the unfolding stimulus sequence. Second, activity within this latency window is related to the adaptation effect measured behaviorally. Thus, these aspects of the adaptation effect suggest that the observed latency window may indeed reflect the computation of perceived average numerosity. Previous results concerning numerosity adaptation (Grasso et al., 2022) observed a correlate of the effect at the P2p component, an ERP component usually associated with numerosity processing (Fornaciai & Park, 2018; Fornaciai et al., 2017; Park et al., 2016; Libertus, Woldorff, & Brannon, 2007). The latency window observed here is not far off from the typical timing of the P2p (200–250 msec), but because of the different nature of the stimuli, it may represent a partially different computational stage. A possibility is that the intermediate stage (~300 msec after stimulus onset) in which we observed a correlate of the adaptation effect might be more genuinely involved with the computation and representation of average numerosity compared with later latency windows showing a relationship with accuracy and precision (~400+ msec). Such later stages, as mentioned above, might instead reflect processes more related to perceptual decision-making. Note, however, that all these different ERP analyses were performed considering the same set of occipital electrodes (Iz, Oz, O1, O2). Although the spatial resolution of EEG is notoriously poor, these results are more consistent with perceptual processing occurring in visual cortex rather than cognitive processes occurring in higher level

brain areas. The topographic plots concerning the effect of adaptation (Figure 4B) nevertheless show peaks extending to occipito-parietal channels, suggesting that signals reflecting adaptation also emerged at scalp locations different from our target channels. Despite this, we were still able to capture robust adaptation effects at our central-occipital target channels. Because of our narrow a priori selection of target channels, we also cannot exclude that other aspects of average numerosity processing might be better captured by other sets of electrodes, reflecting for instance activity in parietal or frontal brain regions. For example, higher level processes related to the judgment of average numerosity might be captured more accurately based on activity from such higher level cortices. Our study was, however, focused on the perceptual aspects of average numerosity processing. Exploring the nature of brain activity at different scalp locations and its potential link to higher level functions thus remains another interesting goal for future studies.

To conclude, our results provide new insights into the computational properties and neural signature of time-averaged numerosity. Our results overall suggest the existence of specific perceptual mechanisms dedicated to the computation of average numerosity over time, subject to the limits of temporal integration of the visual system. The neural signature of average numerosity and the adaptation effect further show two crucial processing stages, at intermediate (~300 msec) and late (~400–700 msec) latencies. These stages are potentially consistent with the initial representation of average numerosity and a subsequent perceptual decision-making stage, respectively.

Corresponding author: Michele Fornaciai, Institut de recherche en sciences psychologiques (IPSY) et en Neurosciences (IoNS), Universite catholique de Louvain, Place Cardinal Mercier 10, Louvain-la-Neuve, Belgium, 1348, e-mail: michele .fornaciai@gmail.com.

Data Availability Statement

All the data generated during the experiment described in this article is available on Open Science Framework, following this link: https://osf.io/9rqs7/.

Author Contributions

Irene Togoli: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Visualization; Writing—Original draft; Writing—review & editing. Olivier Collignon: Conceptualization; Resources; Writing—Original draft; Writing—review & editing. Domenica Bueti: Funding acquisition; Resources; Writing—review & editing. Michele Fornaciai: Conceptualization; Formal analysis; Funding acquisition; Methodology; Project administration; Software; Supervision; Visualization; Writing—Original draft; Writing—review & editing.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this paper report its proportions of citations by gender category to be: M/M = .405; W/M = .19; M/W = .19; W/W = .214.

REFERENCES

Aagten-Murphy, D., & Burr, D. (2016). Adaptation to numerosity requires only brief exposures, and is determined by number of events, not exposure duration. *Journal of Vision*, 16, 22. https://doi.org/10.1167/16.10.22, PubMed: 27580042

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, 15, 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x, PubMed: 14738517

Anobile, G., Arrighi, R., Togoli, I., & Burr, D. C. (2016). A shared numerical representation for action and perception. *eLife*, 5, e16161. https://doi.org/10.7554/eLife.16161, PubMed: 27504969

Anobile, G., Cicchini, G. M., & Burr, D. C. (2016). Number as a primary perceptual attribute: A review. *Perception*, 45, 5–31. https://doi.org/10.1177/0301006615602599, PubMed: 26562858

Arrighi, R., Togoli, I., & Burr, D. C. (2014). A generalized sense of number. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *281*, 20141791. https://doi.org/10.1098/rspb.2014.1791, PubMed: 25377454

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*, 622–628. https://doi.org/10.1111/j.1467-9280.2007.01949.x, PubMed: 17614871

Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology*, *18*, 425–428. https://doi.org/10.1016/j.cub.2008.02.052, PubMed: 18342507

Burr, D. C., & Santoro, L. (2001). Temporal integration of optic flow, measured by contrast and coherence thresholds. *Vision Research*, *41*, 1891–1899. https://doi.org/10.1016/S0042-6989(01)00072-4, PubMed: 11412882

Castaldi, E., Piazza, M., Dehaene, S., Vignaud, A., & Eger, E. (2019). Attentional amplification of neural codes for number independent of other quantities along the dorsal visual stream. *eLife*, *8*, e45160. https://doi.org/10.7554/eLife.45160, PubMed: 31339490

Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*, 1–13. https://doi.org/10.3758/BF03195009, PubMed: 15912869

de Fockert, J., & Wolfenstein, C. (2009). Short article: Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology*, *62*, 1716–1722. https://doi.org/10.1080/17470210902811249

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009, PubMed: 15102499

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition*, *142*, 247–265. https://doi.org/10.1016/j.cognition.2015.05.016, PubMed: 26056747

DeWind, N. K., Park, J., Woldorff, M. G., & Brannon, E. M. (2019). Numerical encoding in early visual cortex. *Cortex*, *114*, 76–89. https://doi.org/10.1016/j.cortex.2018.03.027, PubMed: 29983159

Dormal, V., Seron, X., & Pesenti, M. (2006). Numerosity-duration interference: A Stroop experiment. *Acta Psychologica*, *121*, 109–124. https://doi.org/10.1016/j.actpsy.2005.06.003, PubMed: 16095549

Droit-Volet, S., Clément, A., & Fayol, M. (2003). Time and number discrimination in a bisection task with a sequence of stimuli: A developmental approach. *Journal of Experimental Child Psychology*, *84*(1), 63–76. https://doi.org/10.1016/S0022-0965(02)00180-7, PubMed: 12553918

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*, 307–314. https://doi.org/10.1016/j.tics.2004.05.002, PubMed: 15242690

Fornaciai, M., Brannon, E. M., Woldorff, M. G., & Park, J. (2017). Numerosity processing in early visual cortex. *Neuroimage*, *157*, 429–438. https://doi.org/10.1016/j.neuroimage.2017.05.069, PubMed: 28583882

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*, 1788–1802. https://doi.org/10.1162/jocn_a_01320, PubMed: 30063175

Fornaciai, M., & Park, J. (2019a). Serial dependence generalizes across different stimulus formats, but not different sensory modalities. *Vision Research*, *160*, 108–115. https://doi.org/10.1016/j.visres.2019.04.011, PubMed: 31078663

Fornaciai, M., & Park, J. (2019b). Spontaneous repulsive adaptation in the absence of attractive serial dependence. *Journal of Vision*, *19*, 21. https://doi.org/10.1167/19.5.21, PubMed: 31112999

Fornaciai, M., & Park, J. (2021). Disentangling feedforward versus feedback processing in numerosity representation. *Cortex*, *135*, 255–267. https://doi.org/10.1016/j.cortex.2020.11.013, PubMed: 33412370

Fornaciai, M., Togoli, I., & Bueti, D. (2023). Perceptual history biases are predicted by early visual-evoked activity. *Journal of Neuroscience*, *43*, 3860–3875. https://doi.org/10.1523/JNEUROSCI.1451-22.2023, PubMed: 37085319

Foxe, J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans. *Experimental Brain Research*, *142*, 139–150. https://doi.org/10.1007/s00221-001-0906-7, PubMed: 11797091

Glasser, D. M., Tsui, J. M. G., Pack, C. C., & Tadin, D. (2011). Perceptual and neural consequences of rapid motion adaptation. *Proceedings of the National Academy of Sciences, U.S.A.*, *108*, E1080–E1088. https://doi.org/10.1073/pnas.1101141108, PubMed: 21709221

Goris, R. L. T., Ziemba, C. M., Movshon, J. A., & Simoncelli, E. P. (2018). Slow gain fluctuations limit benefits of temporal integration in visual cortex. *Journal of Vision*, *18*, 8. https://doi.org/10.1167/18.8.8, PubMed: 30140890

Grasso, P. A., Petrizzo, I., Caponi, C., Anobile, G., & Arrighi, R. (2022). Visual P2p component responds to perceived numerosity. *Frontiers in Human Neuroscience*, *16*, 1014703. https://doi.org/10.3389/fnhum.2022.1014703, PubMed: 36393989

Harvey, B. M., Klein, B. P., Petridou, N., & Dumoulin, S. O. (2013). Topographic representation of numerosity in the human parietal cortex. *Science*, *341*, 1123–1126. https://doi.org/10.1126/science.1239052, PubMed: 24009396

Hoffmann, M., Dorn, T. J., & Bach, M. (1999). Time course of motion adaptation: Motion-onset visual evoked potentials and subjective estimates. *Vision Research*, *39*, 437–444. https://doi.org/10.1016/S0042-6989(98)00186-2, PubMed: 10341975

Hoffmann, M. B., Unsöld, A. S., & Bach, M. (2001). Directional tuning of human motion adaptation as reflected by the motion VEP. *Vision Research*, *41*, 2187–2194. https://doi.org/10.1016/S0042-6989(01)00112-2, PubMed: 11448711

Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision*, *15*, 5. https://doi.org/10.1167/15.4.5, PubMed: 26053144

Javadi, A. H., & Aichelburg, C. (2012). When time and numerosity interfere: The longer the more, and the more the longer. *PLoS One*, *7*, e41496. https://doi.org/10.1371/journal.pone.0041496, PubMed: 22911801

Katzin, N., Rosenbaum, D., & Usher, M. (2021). The averaging of numerosities: A psychometric investigation of the mental line. *Attention, Perception, & Psychophysics*, *83*, 1152–1168. https://doi.org/10.3758/s13414-020-02140-w, PubMed: 33078378

Kersey, A. J., Braham, E. J., Csumitta, K. D., Libertus, M. E., & Cantlon, J. F. (2018). No intrinsic gender differences in children's earliest numerical abilities. *NPJ Science of Learning*, *3*, 12. https://doi.org/10.1038/s41539-018-0028-7, PubMed: 30631473

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, 1–16.

Kohn, A. (2007). Visual adaptation: Physiology, mechanisms, and functional benefits. *Journal of Neurophysiology*, *97*, 3155–3164. https://doi.org/10.1152/jn.00086.2007, PubMed: 17344377

Lambrechts, A., Walsh, V., & van Wassenhove, V. (2013). Evidence accumulation in the magnitude system. *PLoS One*, *8*, e82122. https://doi.org/10.1371/journal.pone.0082122, PubMed: 24339998

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. https://doi.org/10.1017/S0140525X16000960, PubMed: 27530053

Libertus, M. E., Woldorff, M. G., & Brannon, E. M. (2007). Electrophysiological evidence for notation independence in numerical processing. *Behavioral and Brain Functions*, *3*, 1. https://doi.org/10.1186/1744-9081-3-1, PubMed: 17214890

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213. https://doi.org/10.3389/fnhum.2014.00213, PubMed: 24782741

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. https://doi.org/10.1038/36846, PubMed: 9384378

McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature Neuroscience*, *16*, 493–498. https://doi.org/10.1038/nn.3347, PubMed: 23434915

Park, J., DeWind, N. K., Woldorff, M. G., & Brannon, E. M. (2016). Rapid and direct encoding of numerosity in the visual stream. *Cerebral Cortex*, *26*, 748–763. https://doi.org/10.1093/cercor/bhv017, PubMed: 25715283

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. https://doi.org/10.1163/156856897X00366, PubMed: 9176953

Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision*, *11*, 18. https://doi.org/10.1167/11.12.18, PubMed: 22031908

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*, 7168–7173. https://doi.org/10.1523/JNEUROSCI.4503-10.2011, PubMed: 21562280

Temple, E., & Posner, M. I. (1998). Brain mechanisms of quantity are similar in 5-year-old children and adults. *Proceedings of the National Academy of Sciences, U.S.A.*, *95*, 7836–7841. https://doi.org/10.1073/pnas.95.13.7836, PubMed: 9636237

Testolin, A., & McClelland, J. L. (2021). Do estimates of numerosity really adhere to Weber's law? A reexamination of two case studies. *Psychonomic Bulletin & Review*, *28*, 158–168. https://doi.org/10.3758/s13423-020-01801-z, PubMed: 32949010

Togoli, I., Bueti, D., & Fornaciai, M. (2022). The nature of magnitude integration: Contextual interference versus active magnitude binding. *Journal of Vision*, *22*, 11. https://doi.org/10.1167/jov.22.11.11, PubMed: 36259675

Togoli, I., Fornaciai, M., & Bueti, D. (2021). The specious interaction of time and numerosity perception. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *288*, 20211577. https://doi.org/10.1098/rspb.2021.1577, PubMed: 34547911

Tonoyan, Y., Fornaciai, M., Parsons, B., & Bueti, D. (2022). Subjective time is predicted by local and early visual processing. *Neuroimage*, *264*, 119707. https://doi.org/10.1016/j.neuroimage.2022.119707, PubMed: 36341952

Visibelli, E., Porru, A., Lucangeli, D., Butterworth, B., & Benavides-Varela, S. (2024). Neural indicators of numerical abilities in the infant human brain: A systematic review. *Developmental Review*, *74*, 101150. https://doi.org/10.1016/j.dr.2024.101150

Visibelli, E., Vigna, G., Nascimben, C., & Benavides-Varela, S. (2024). Neurobiology of numerical learning. *Neuroscience & Biobehavioral Reviews*, *158*, 105545. https://doi.org/10.1016/j.neubiorev.2024.105545, PubMed: 38220032

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*, 483–488. https://doi.org/10.1016/j.tics.2003.09.002, PubMed: 14585444

Watson, A. B. (1979). Probability summation over time. *Vision Research*, *19*, 515–522. https://doi.org/10.1016/0042-6989(79)90136-6, PubMed: 483579

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, *69*, 105–129. https://doi.org/10.1146/annurev-psych-010416-044232, PubMed: 28892638