Attractive serial dependence between memorized stimuli

Michele Fornaciai[1]* & Joonkoo Park[1,2]

[1] Department of Psychological & Brain Sciences, University of Massachusetts, Amherst, MA, USA.

[2] Commonwealth Honors College, University of Massachusetts, Amherst, MA, USA.

* Correspondence:

Michele Fornaciai

Present address:

International School for Advanced Studies (SISSA)

Via Bonomea 265, 34136 Trieste (TS), Italy

michele.fornaciai@gmail.com

ABSTRACT

Attractive serial dependence – a bias whereby the current stimulus appears more similar to the previous ones – is thought to reflect a stability mechanism integrating past and current visual signals. Prior work suggests that serial dependence originates from both perceptual and cognitive mechanisms, but the conditions under which this attractive bias occurs remain to be studied. In particular, whether serial dependence can occur solely from memory interference remains unclear. Here, we address this question by testing the hypothesis that if memory interference is sufficient to generate serial dependence, it should occur irrespective of the order of stimulus presentation. In Exp. 1, we used a numerosity estimation task in which participants estimated the number of dots of a briefly flashing dot-array comprising 8 to 32 dots. The pattern of serial dependence was found in that numerical estimates of a dot array were biased towards the numerosity of the preceding dot array. In Exp. 2, we presented a series of three such dot arrays, and cued the one to be estimated only after the whole series was presented, making the participants first form a memorized representation of the three dot arrays. The results show a pattern of attractive biases both in the forward (stimulus presented before biases stimulus presented after) and the backward (stimulus presented after biases stimulus presented before) directions. Overall, our results demonstrate that serial dependence can be induced solely from memory interference and that this interference can operate irrespective of the chronological order of the stimulus presentation.

Keywords: Serial dependence; Numerosity perception; Memory interference; Visual stability

1. INTRODUCTION

Visual perception is not constructed by assembling a series of static snapshots of the external world. Instead, how we perceive stimuli in our subjective present is affected by the recent history of stimulation. For instance, one of the most studied contextual or stimulation history effects is perceptual adaptation: after a relatively long stimulation (e.g., varying from tens of milliseconds to several seconds depending on the context), sensory responses get recalibrated so that the response to a subsequent stimulus is repulsed away from the adapting stimulus (e.g., see Kohn, 2007 for a review). Adaptation occurs at many levels across the sensory processing pathways and at many timescales, and thus influences many aspects of perception (e.g., Brown & Masland, 2001; Boynton & Finney, 2003; Montaser-Kouhsari et al., 2007; Kohn & Movshon, 2003; Kohn, 2007; Glasser et al., 2011).

While adaptation has a *repulsive* effect on perception – effectively making successive stimuli to appear more different – *attractive* effects have also been documented. A recent line of research has started to focus on such attractive effects, called *serial dependencies*, and their functional significance and physiological properties. Similarly to adaptation, attractive serial dependence has been documented across a large variety of visual features. Indeed, attractive biases have been observed in domains spanning from basic perceptual attributes such as orientation (Fischer & Whitney, 2014), numerosity (Corbett et al., 2011; Cicchini et al., 2014; Fornaciai & Park, 2018a, 2018b, 2019a), position (Bliss et al., 2018; Manassi et al., 2018), and motion (Alais et al., 2017), to more complex perceptual attributes such as faces (Libermann et al., 2014; Xia et al., 2016; Libermann et al., 2018), stimulus variance (Suarez-Pinilla et al., 2018), and summary statistics (Manassi et al., 2017), and have also been shown to generalize across different stimulus presentation formats (Fornaciai & Park, 2019b).

However, many properties of attractive serial dependence are still unknown. For instance, whether the attractive bias has origins in perception or in memory has been hotly debated. On the one hand, it has been proposed that the serial dependence reflects the outcome of a "continuity field," whereby current and past visual information is integrated to smooth out noise from neural signals in the service of visual stability and continuity (Fischer & Whitney, 2014). Such an account based on visual stability and continuity has been challenged on the ground that the bias may arise at the memory/decision stage (Fritsche et al., 2017; Bliss et al., 2018), thus making it a more "cognitive" rather than "perceptual" effect. Moreover, while the continuity field is based on a relatively low-level mechanism (Fischer & Whitney, 2014), it has been recently proposed that the effect may originate at a relatively high-level, read-out, stage. According to this view, serial dependence would emerge either due to lingering "decisional templates" at the level of perceptual decision (read-out) units (i.e., the set of read-out weights used to form a perceptual representation out of the low-level population activity of sensory neurons; Pascucci et al., 2019), or by means of modulatory feedback signals sent to low-level sensory areas (Fornaciai & Park, 2019a). Crucially, even if the bias occurs at a high-level stage, it could still affect perception directly, effectively biasing the appearance of a stimulus.

Such different interpretations of serial dependence effects stem from the fact that experimental results are currently mixed, with some studies showing a signature of a perceptual effect, and other studies showing a contribution from memory and decision processes. For instance, on the one hand, some behavioral results show that the effect mostly depends on the past stimulus rather than past responses, suggesting that, at least in this context, decision processes play only a little role in the observed effect (Cicchini et al., 2017). It is worth it mentioning however that other studies (e.g., St. John-Saaltink et al., 2017) instead observed a more prominent role of past responses, as opposed to past stimuli. Whether serial dependence

operates based on past stimuli or past responses may thus depend on the stimuli (i.e., their noisiness) and task used. Moreover, electroencephalography results show that brain responses are biased by the previous stimulus at extremely early latencies after stimulus onset, suggesting that serial dependence starts at the earliest levels of perceptual processing (Fornaciai & Park, 2018a; Fornaciai & Park, 2020). Other evidence supports the idea of a perceptual effect, but further suggests a relatively high level origin of this bias. Namely, it has been shown that the effect depends on attention (Fischer & Whitney, 2014; Fornaciai & Park, 2018b), it requires awareness of the stimuli (Fornaciai & Park, 2019a), and even generalizes across stimuli with widely different low-level sensory properties (Fornaciai & Park, 2019b). As mentioned above, recent accounts of serial dependence consistent with these findings concern a bias provided by lingering traces of past decision templates (Pascucci et al., 2019) or high-level modulatory feedback signals affecting early sensory activity (Fornaciai & Park, 2018a; Fornaciai & Park, 2019a). On the other hand, there is evidence that serial dependence has a source in memory processes. The attractive bias increases with increased time between the stimulus presentation and the behavioral response, implicating a modulation of the effect during working memory storage (Fritsche et al., 2017; Bliss et al., 2018; but see Manassi et al., 2018 for results opposing this point). Due to the variability in findings across different studies and paradigms, whether serial dependence is a unitary phenomenon remains unclear. Indeed, widely different mechanisms (i.e., a bias in perceptual processing or an interference between memory traces) may result in similar effects at the behavioral level. Therefore, an important point that needs to be investigated in this context is the conditions under which serial dependence arises, to understand the contexts leading to perceptual or memory effects.

In the present work, we aim to address this question in the context of numerosity perception. In Exp. 1, we first employed a simple numerosity estimation task in which the participant reported the estimated number of a dot array using the number pad, in order to assess whether numerosity estimation is affected by serial dependence. Doing so, we confirmed that numerosity estimation performance is indeed robustly and systematically affected by serial dependence, with the strongest effect provided by the immediately preceding trial. Furthermore, in Exp. 2 we presented a series of three different dot arrays in each trial, and cued which one the participant had to estimate after the sequence. Using such a sequence of multiple potential target stimuli, we have two specific predictions concerning the possible perceptual and memory effects. Namely, while a perceptual effect (i.e., affecting the perceptual representation, and hence the appearance, of a stimulus) should strictly follow the temporal order of the sequence of stimuli, a memory effect would occur irrespective of that – i.e., memory interference could occur after a stimulus has been perceived. More specifically, while in the first case we should observe an effect provided by the earlier-presented stimuli on later-presented stimuli and not vice versa, the interference between different representations held in memory could occur irrespective of which stimulus was presented first. Results

from this experiment show an attractive bias working both in the "forward" (i.e., the preceding stimulus affecting its successor) and in the "backward" direction (i.e., the successor affecting the preceding stimulus), which suggests that serial dependence can occur solely from memory interference.

2. METHODS

2.1. Participants

A total of 66 subjects participated in the study (51 females, mean age = 20.7, SD = 1.8 years). Subjects were compensated with course credits for their participation. All participants had normal or corrected-to-normal vision, and provided a written informed consent prior to taking part in the study. All the experimental procedures were approved by the Institutional Review Board of the University of Massachusetts Amherst and were in line with the Declaration of Helsinki. Six participants were excluded before data analysis due to poor motivation or because they could not complete the experiment.

2.2. Apparatus and stimuli

Visual stimuli were generated using the routines provided by Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) for Matlab (version r2016b; The Mathworks, Inc.). During the experiment, stimuli were presented on a monitor screen encompassing approximately 35×20 degrees of visual angle (from a viewing distance of about 80 cm; resolution = 1920 × 1080 pixel), and running at 144 Hz.

All the stimuli were arrays of black and white dots presented on a gray background. Each dot was randomly positioned within a virtual circular area (i.e., field area, see below), with possible positions only constrained by keeping a minimum inter-dot distance equal to at least the radius of one dot. In Exp. 1, one dot-array stimulus was presented in each trial, while in Exp. 2 a sequence of three dot arrays was presented in each trial. Following a technique used in previous studies (e.g., Park et al., 2016; DeWind et al., 2015), all the stimuli were systematically constructed to span equal ranges in three orthogonal dimensions, reflecting *numerosity*, *size*, and *spacing*. Besides numerosity, the two other dimensions orthogonal to it (size and spacing) were obtained by logarithmically scaling and combining the area of the individual items and the total area occupied by all the items in an array, or the area of the circular field containing the dots and the sparsity of the items (i.e., the inverse of the density of the array). The dot-array stimuli across both experiments were modulated across 11 levels of each dimension. For more details about this stimulus construction procedure, see Park et al. (2016) and DeWind et al. (2015). Note,

however, that since the effect of serial dependence on numerosity estimation performance was the main goal of the present work, we collapsed together the different non-numerical dimensions for data analysis.

The stimulus parameters were set as follows. Dot-array stimuli included 8, 9, 11, 12, 14, 16, 18, 21, 24, 28, or 32 dots. Regarding the other non-numerical dimensions, the minimum individual dot area was set to 176 pixel^2 (0.06 deg^2), equal to a diameter of 0.28 deg (15 pixel pixels), while the maximum individual dot area was 707 pixel^2 (0.24 deg^2), equal to a diameter of 0.56 deg (30 pixel). The minimum field area (i.e., the virtual circular area where the dots were drawn) was set to 70,686 pixel^2 (23.9 deg^2), encompassing 5.5 deg (300 pixels), while the maximum field area was 282,743 pixel^2 (95.7 deg^2), encompassing 11 degrees in diameter (600 pixels). In all cases, the individual dot size was kept equal within an array.

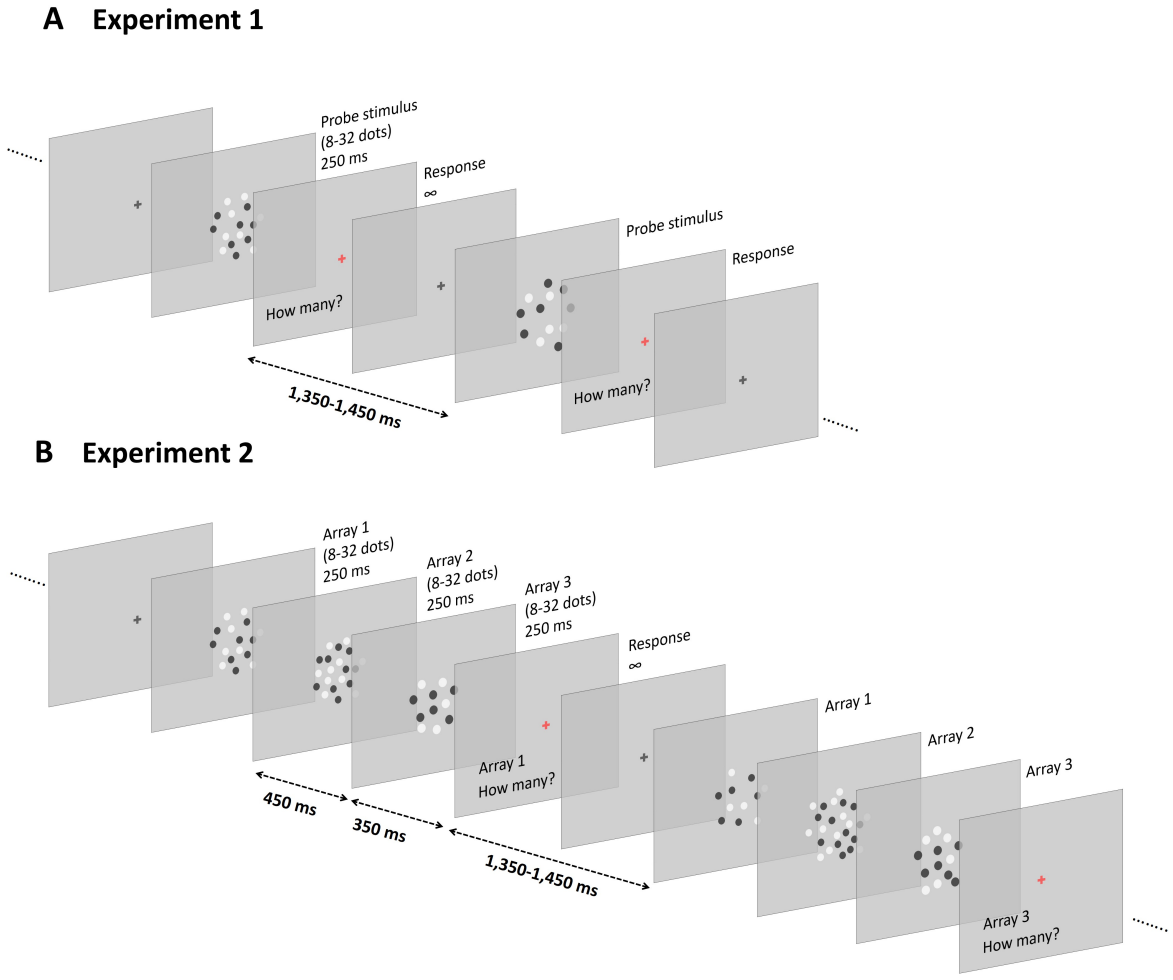


FIGURE 1. Procedure. *(A) Stimulus presentation procedure in Exp. 1. On each trial, a single probe dot-array stimulus (containing 8-32 dots) was presented at the center of the screen for 250 ms. After the offset of the stimulus, participants were asked to estimate the number of dots in the probe stimulus by typing the number on a keyboard. The number appeared on the screen while typing. After the participant confirmed the response by pressing enter, the next trial started automatically after 1,350-1,450 ms. (B) Stimulus presentation procedure in Exp. 2. In the second experiment, a series of three different dot-array stimuli was presented in each trial, each of them presented at the center of the screen for 250 ms, with an inter-stimulus interval of 450 ms. At the end of the sequence, participants had to estimate only the target probe stimulus which was indicated with a cue (1, 2, or 3, corresponding to the first, the second, or the third stimulus in the sequence) appearing on the screen 350 ms after the offset of the last stimulus. After providing the estimate, the next trial started automatically after 1,350-1,450 ms.*

2.3. Procedure

In Exp. 1, participants performed a numerosity estimation task, reporting how many dots they saw in a probe dot-array stimulus presented in each trial. Namely, while participants fixated on a central fixation cross, the probe dot-array was presented for 250 ms at the center of the screen (replacing the fixation cross). After the stimulus presentation, the question "how many dots?" appeared on the screen below the fixation cross, and participants were instructed to report the number of dots by entering the number on the numerical pad of a standard keyboard. Numbers were displayed on the screen, and participants had the possibility to correct their response by pressing backspace and deleting the number. When ready, subjects were instructed to press enter to confirm the response, and the next trial started automatically after a pause of 1,350-1,450 ms. To provide a reference range for performing the estimation task, but without revealing the real numerosity range, participants were told that the stimuli could be from 6 to 40 dots. This strategy was used to introduce some uncertainty about the extreme stimuli in the range (i.e., 8 and 32).

In Exp. 2 the procedure was very similar, except that a sequence of three dot array stimuli was displayed on the screen in each trial. Each stimulus was displayed for 250 ms, with an inter-stimulus interval of 450 ms. After the stimulus presentation, a cue appeared on the screen (after 350 ms from the offset of the last stimulus) indicating which probe stimulus in the sequence the participant had to estimate (i.e., 1, 2, or 3; respectively for the first, second, or third stimulus in the sequence). In this experiment, in order to avoid typos, participants' responses were constrained so that they could not enter a response lower than 6 or higher than 40. Participants completed 8 blocks of 55 trials in Exp. 1, and 10 blocks of 55 trials in Exp. 2 (with the exception of one participants who completed 7 blocks in Exp. 1 and one who completed 9

186 blocks in Exp. 2 due to time constraints). Before starting the experiment, participants were shown a brief
187 tutorial showing several examples of the stimuli with the actual numerosity indicated on the screen, in
188 order to familiarize themselves with the task. An entire experimental session took about 50 minutes, and
189 participants were free to take breaks between blocks.

190

191 *2.4. Data analysis*

192 In Exp. 1, we first analyzed the numerosity estimation performance and excluded trials where the
193 response was either lower than 6 or higher than 40, in order to exclude typos from being included in the
194 analysis. Estimation performance was evaluated by assessing the average reported numerosity for each
195 actual level of numerosity in the range, and computing the coefficient of variation (CoV; the standard
196 deviation of numerical estimates divided by the physical numerosity) as a measure of precision. The
197 extent to which average subjective reports deviate from the veridical magnitude at different numerosity
198 levels was assessed with a series of one-sample t-tests against the actual numerical magnitude. We also
199 assessed how precision (CoV) in the task varied as a function of numerosity with a one-way repeated
200 measure ANOVA with factor "numerosity." To assess the serial dependence effect, the estimation error
201 (response – stimulus veridical numerosity) in the current trial (n) was plotted as a function of the stimulus
202 numerosity in the previous trials (spanning from n-1 to n-7), as well as the stimulus in the immediately
203 successive trial (n+1) as a control. A linear function was fitted to the data arranged in this way,
204 individually for each participant and separately for each condition assessing the influence of different
205 preceding (or successive) trials. The slope of the linear fit (henceforth referred to as "serial dependence
206 effect") was taken as an index of the bias provided by past stimuli on current numerical estimates
207 (Cicchini et al., 2014): a negative slope represents a repulsive effect, and a positive slope indexes an
208 attractive effect. Additionally, the serial dependence effect was assessed in a number-by-number fashion.
209 Namely, we computed the serial dependence effect by pooling all the trials in which a specific number
210 was presented, and again computing the estimation error in the current trial as a function of the stimulus
211 presented in the previous trial (limited to the n-1 case). In both cases, the significance of serial
212 dependence effects was assessed individually using one-sample t-tests against a null hypothesis of zero
213 effect (i.e., either for individual trial-back conditions, or for individual numerosities). To control for
214 multiple comparisons, the significance level of individual t-tests was corrected using a false discovery
215 rate (FDR) correction (Benjamini & Hochberg, 1995), with the critical threshold of the false discovery
216 rate (q) of 0.05. In those cases, we reported the FDR-adjusted p-values in the Results section.
217 Furthermore, a one-way repeated measure ANOVA (either with factor "trial back" or with factor

"numerosity," respectively for the two analyses) was used to assess the overall pattern and compare the different conditions against each other by means of post-hoc tests.

In Exp. 2, we first analyzed the general estimation performance as in Exp. 1. A series of one-sample t-tests against the different veridical numerosities was used to assess whether numerical estimates deviate from the presented numerosity, while a one-way repeated measure ANOVA (with factor "numerosity") was used to assess CoVs. Additionally, both numerical estimates and CoV were compared to the results of Exp. 1 using a two-way ANOVA with factors "numerosity" and "experiment." Regarding serial dependence, a similar analysis compared to Exp. 1 was performed to assess the effect of the three stimuli in the sequence on each other. Namely, we separately assessed the influence (in terms of slope of a linear fit to the response error as a function of previous/successive stimulus numerosity) of the first and second stimulus in cases when the third one was cued, the first and the third on the second one, and the second and the third on the first one. The significance of serial dependence effects across the different conditions was assessed using one-sample t-tests against a null hypothesis of zero effect. Moreover, we also compared the different conditions by using a two-way repeated measure ANOVA. In this context, we entered as factors the "target stimulus" (i.e., first, second, or third selected as the target stimulus), and the "comparison" type, which was coded as 1 and 2 for all pairs of conditions (i.e., effect of the first and second stimulus in the case of the third selected as relevant, and so on) included in each target stimulus condition. In addition, the serial dependence effects *across* different trials were assessed by quantifying the bias induced by each of the three stimuli presented in the previous trial on the stimuli presented in the current trial, separately for the conditions in which different stimuli were cued to be reported in the estimation task. The significance of such effects was first assessed with a two-way repeated measure ANOVA, with factors based on the stimulus in the previous trial and the one in the current trial, followed by a series of one-sample t-tests against zero. In all cases, the significance level of individual t-tests was corrected by means of a FDR procedure, as in Exp. 1.

Finally, we devised a series of simulations to test for the possible role of swap errors in the observed pattern of results (Bays et al., 2009; Bays, 2016; Pratte, 2019). Indeed, any pattern of attractive serial dependence or memory bias could be alternatively explained by occasional mistakes where the previous stimulus (or another stimulus in the sequence in Exp. 2) is reported instead of the correct one. To assess the possible effect of swap errors in Exp. 1, we simulated the performance of a number of participants and trials equal to the real experiment. In each trial, the expected response was based on the numerosity selected, with the addition of an error term based on the average CoV measured in the experiment. In a subset of trials (2%, based on Bays, 2016), the response was instead based on the stimulus selected in the previous virtual trial, to simulate a swap error. We then compared the average estimate of the effect (i.e.,

slope of the linear fit) of the simulated data with the empirical results. At the same time, we evaluated the average R2 of a linear fit to the simulated data and compared it to the average R2 measured with the real data, with the rationale that response distribution in the case of swap error would be mostly centered on the veridical numerosity displayed with a small proportion of "outlier" responses, which would result in a much lower goodness of fit, in contrast to the case of serial dependence which would be represented by a uniform bias across all trials. The simulation results first showed that swap errors were insufficient to generate an effect similar to what we observed in Exp. 1 (i.e., mean slope = 0.068-0.12, compared with the observed slope of 0.23 at trial n-1). Next, we observed a much higher R2 in the case of the real data (mean ± SD = 0.021 ± 0.02, compared with 0.004 ± 0.0002 obtained with the simulation), together making serial dependence a more likely explanation than swap errors in Exp. 1.

In the case of Exp. 2, we performed a similar simulation, with the exception that swap errors were assumed to be more frequent (swap error rate of 5%; again based on Bays, 2016) and dependent on the target stimulus position in the sequence. Namely, when the first stimulus was selected as target, we assumed that most of the swaps (66%) would be made with the immediately successive one (the second), and fewer errors (33%) would be based on the third one (i.e., see for instance Pratte, 2019 for the effect of item distance in determining swap errors). Similarly, in the case of the third stimulus selected as target, the effect would be based most of the times (66%) on the immediately preceding one, and fewer times (33%) on the first stimulus in the sequence. In the case of the second stimulus, instead, swap errors are predicted to be symmetrically distributed between the first and the third (50%/50%). The results showed patterns incompatible with the empirical data. That is, in the case of the second stimulus, simulation results showed a nearly perfectly symmetrical effect of the preceding and successive stimulus (slope = 0.95 in both cases), which are not observed in real data. The simulation results showed an effect of the first stimulus to the third and vice versa (average slope = 0.06 in both cases), which again are not observed in real data. Finally, we again observed much higher average R2 values for the fits to the real data (0.010-0.012 in the simulated dataset, 0.017-0.024 in the real dataset). These results collectively suggest that the current results are unlikely to be explained by swap errors.

3. RESULTS

3.1. Experiment 1

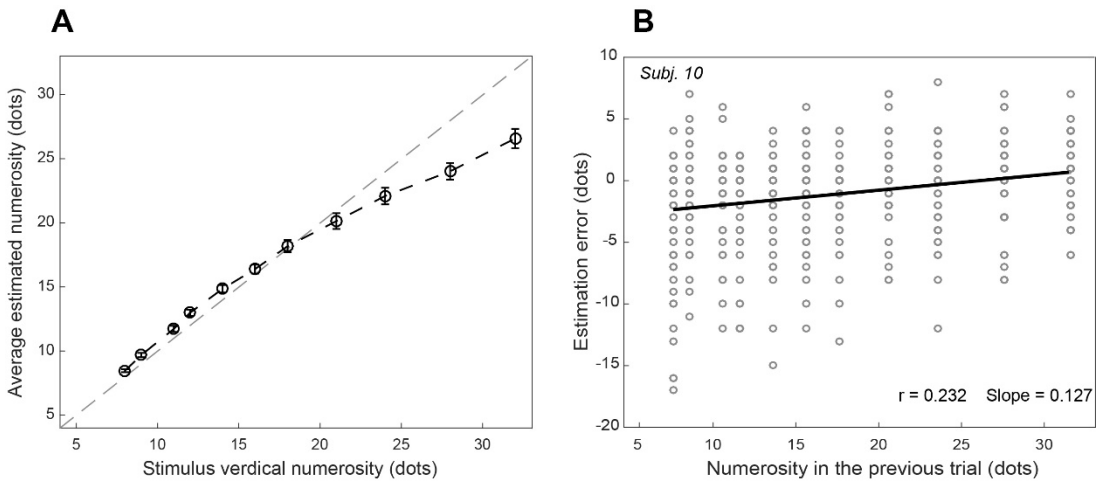


FIGURE 2. General estimation performance results and individual serial dependence effect in Exp. 1. *(A) Average estimated numerosity as a function of stimulus veridical numerosity. Error bars are SEM. (B) Example of how the serial dependence effect was assessed at the individual level, for one representative participant in Exp. 1. The distribution of estimation errors (response – veridical numerosity) was plotted as a function of the numerosity in the previous trial (n-1). The thick line represents a linear fit to all the data. The slope of the linear fit was taken as an index of the serial dependence effect. Dots represent individual trials. Note that such a pattern of serial dependence is unlikely to be explained by trivial swap errors whereby on occasional trials the previous stimulus is mistakenly reported instead of the previous one, as demonstrated by the simulation reported in the Methods section (see Data analysis).*

First, we assessed the general estimation performance in the task. Fig. 2A shows the average estimated numerosity as a function of stimulus veridical numerosity. Overall, responses were noticeably compressed, with a slight overestimation at the lower end of the range and underestimation at the higher end of the range, in line with previous studies (e.g., Arrighi et al., 2014; Cicchini et al., 2014). Indeed, running a series of one-sample t-tests (against veridical numerosity) shows that average estimates significantly deviate from the stimulus veridical numerosity, in terms of overestimation, for stimuli ranging from 8 to 14 (one-sample t-tests, 8: $t(31) = 4.01$, $p < 0.001$, Cohen's $d = 0.71$; 9: $t(31) = 4.03$, $p < 0.001$, $d = 0.72$; 11: $t(31) = 2.96$, $p = 0.006$, $d = 0.52$; 12: $t(31) = 3.84$, $p < 0.001$, $d = 0.68$; 14: $t(31) = 2.79$, $p = 0.009$, $d = 0.49$), while a significant underestimation was observed for stimuli from 24 to 32 (24: $t(31) = -2.93$, $p = 0.006$, $d = 0.52$; 28: $t(31) = -5.97$, $p < 0.001$, $d = 1.05$; 32: $t(31) = -7.13$, $p < 0.001$, $d = 1.26$). No significant deviation of average subjective reports was observed in the range spanning from 16 to 21 (16: $t(31) = 1.07$, $p = 0.29$; 18: $t(31) = 0.36$, $p = 0.72$; $t(31) = -1.39$, $p = 0.17$).

To assess how estimation precision (coefficient of variation, CoV; data not shown) varied as a function of numerosity, we used a one-way repeated measure ANOVA on CoV with the factor "numerosity." While precision measures appeared to be lower (i.e., higher precision) at the extremes of the range (average CoV = 0.176 ± 0.018, 0.182 ± 0.009, and 0.160 ± 0.007, respectively for 8, 28, and 32), compared to the middle range (9-24; average CoV spanning from 0.212 to 0.256 at 12, which showed the lowest precision), the test results did not show any statistically significant difference ($F(10,31) = 1.41$, $p = 0.172$). The average CoV (± SD) across the range was 0.217 ± 0.031.

Besides these general measures of performance, the main results concern the serial dependence effect. Fig. 2B shows an example of how we defined the serial dependence effect in the context of numerosity estimation, at the individual level. As shown in the figure, a positive slope of the linear fit indexes an attractive effect: when a low number was presented in the previous trial, participants tended to underestimate the numerosity of the current trial (i.e., negative estimation error); when a high number was presented in the previous trial, participants tended to overestimate the numerosity of the current trial (i.e., positive estimation error). First, this analysis was performed individually across all the trials, to assess the magnitude of serial dependence effect at the group level, considering a wide range of past trials (from the previous trial, n-1, to seven trials back, n-7), as well as the successive, future, trial (n+1) as a control (Fig. 3A).

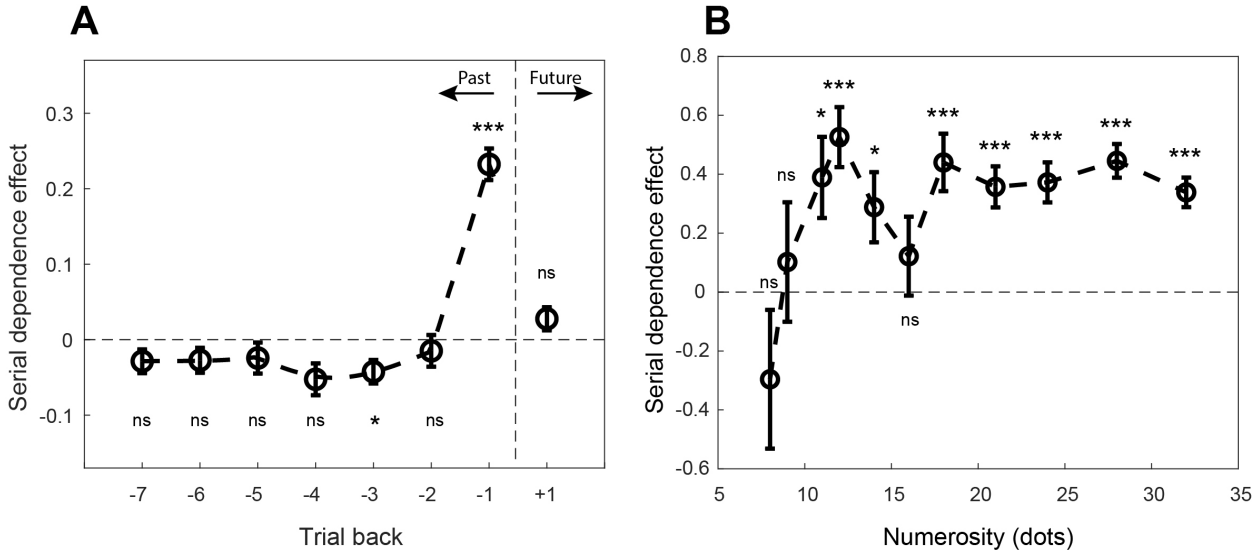


FIGURE 3. Serial dependence effects in Exp. 1. *(A) Serial dependence effects provided by the numerosity presented in previous trials, ranging from the immediately preceding trial, to seven trials back. The following trial was also assessed as a control, as no effect is expected to arise as a function of a future trial. (B) Serial dependence effects as a function of numerosity in the current trial. This analysis*

*shows how different numerosities are susceptible to serial dependence effects. Error bars are SEM. The significance levels indicated in the figure refer to FDR-adjusted p-values, ns = not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

Fig. 3A shows the serial dependence effect across a range of past trials. The strongest effect was provided by the immediately preceding trial (n-1; average effect = 0.232 ± 0.025; one-sample t-test against zero, $t(31) = 9.08$, $p < 0.001$, $d = 1.64$), while at trials further back in time all the effects are close to zero or slightly negative (i.e., repulsive). At n-2, no significant effect was observed ($t(31) = -0.71$, $p = 0.486$), while at trial n-3 a small but significant negative effect emerged (average effect = -0.042 ± 0.015; $t(31) = -2.71$, $p = 0.043$, $d = 0.48$). At trials further back in the past, no significant effect was observed (n-4 to n-7, average effect spanning from -0.024 to -0.052; all p-values > 0.05). As a control for our data, we also assessed the effect at n+1 – that is, the effect provided by the immediately *successive* trial. As expected, no significant effect was observed (average effect = 0.027 ± 0.015; $t(31) = 1.80$, $p = 0.13$). This shows that while the numerosity in the immediately preceding trial exerts a systematic attractive effect on numerical estimates in the current trial, stimuli further back in the past have mostly a negligible influence, or even provide a small repulsive effect. Furthermore, we used a one-way ANOVA with factor "trial back" to assess the overall pattern of effects. Doing so, we found a significant main effect of trial back ($F(7,31) = 24.74$, $p < 0.001$, $\eta_p^2 = 0.44$). With a series of post-hoc tests, we also found that such an effect was exclusively driven by the n-1 condition, which resulted to be significantly different from all the other conditions ($t(31)$ ranging from 7.71 to 10.72, all p-values < 0.001). All the other comparisons did not show any significant difference. Also the conditions showing a repulsive effect (n-3, n-4) did not result to be significantly different from other conditions showing no effect in the previous analysis ($t(31)$ ranging from 0.52 to 1.42, all p-values > 0.924).

Furthermore, we also assessed how individual numbers are susceptible to serial dependence effects. Fig. 3B shows the average serial dependence effect as a function of numerosity in the current trial. The effect appears to be very small and not significant at lower numbers such as 8 (average effect = -0.296 ± 0.239; one-sample t-test, $t(31) = -1.23$, $p = 0.275$) and 9 (average effect = 0.102 ± 0.206; $t(31) = 0.49$, $p = 0.624$). A significant effect was instead observed across most of the higher numerosities in the range (11: $t(31) = 2.78$, $p = 0.014$, $d = 0.49$; 12: $t(31) = 5.08$, $p < 0.001$, $d = 0.90$; 14: $t(31) = 2.38$, $p = 0.032$, $d = 0.42$; 18: $t(31) = 4.43$, $p < 0.001$, $d = 0.78$; 21: $t(31) = 5.04$, $p < 0.001$, $d = 0.89$; 24: $t(31) = 5.40$, $p < 0.001$, $d = 0.95$; 28: $t(31) = 7.65$, $p < 0.001$, $d = 1.35$; 18: $t(31) = 6.62$, $p < 0.001$, $d = 1.17$), with the highest effect at numerosity 12 (0.526 ± 0.103). Interestingly, no significant effect was observed at 16 ($t(31) = 0.89$, $p = 0.415$), which is the central value of the range. These results show that while relatively

low numbers such as 8 and 9 are more resistant to serial dependence effect, higher numbers are more easily affected by the previous trial. Also in this context, we used a one-way repeated measure ANOVA, with factor "numerosity," to assess the overall pattern of effects. The results showed a main effect of numerosity on the serial dependence effect ($F(10,31) = 3.59$, $p < 0.001$, $\eta_p^2 = 0.10$). A series of post-hoc tests further showed that the effect at numerosity 8 is significantly different from most of the other numerosities (11, 12, 14, 18, 21, 24, 28, 32; $t(31)$ ranging from 3.39 to 4.78, all p-values < 0.037), with the exception of 9 ($t(31) = 2.31$, $p = 0.61$) and 16 ($t(31) = 2.42$, $p = 0.52$). All the other pairwise comparisons did not show any significant difference (all p-values > 0.49).

3.2. Experiment 2

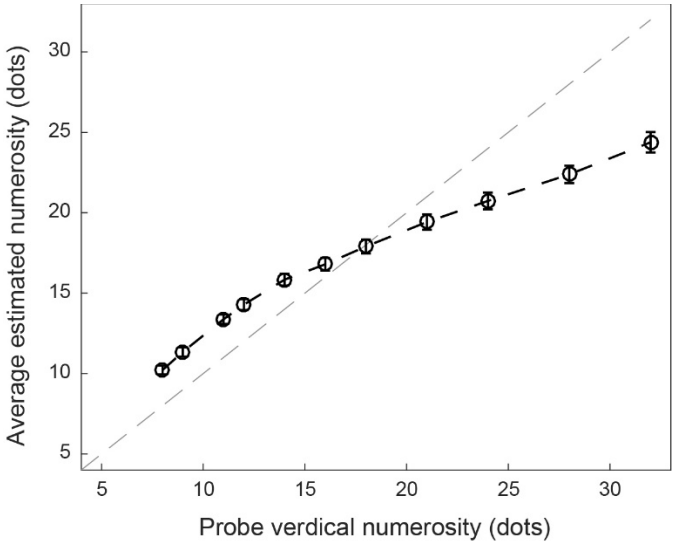
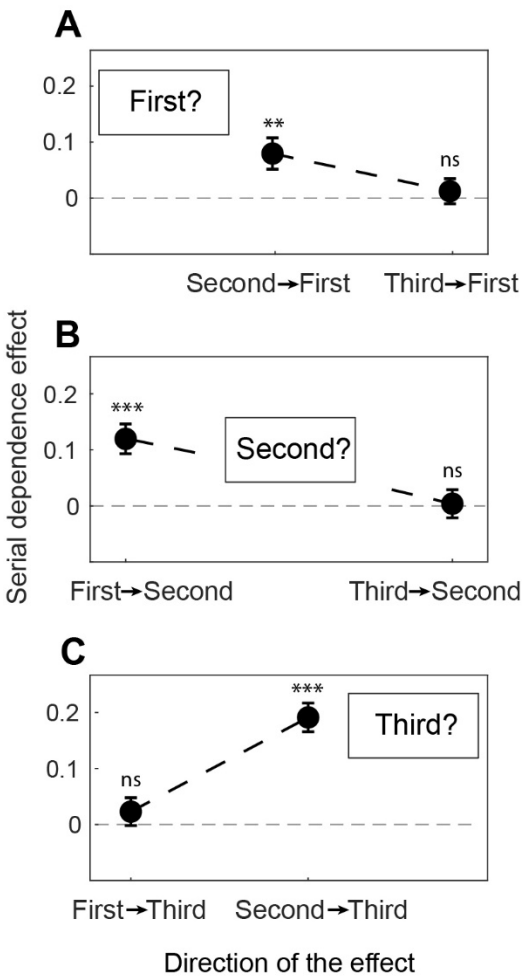


FIGURE 4. Estimation performance in Exp. 2. *Average estimated numerosity of the target probe stimulus as a function of stimulus veridical numerosity. Error bars are SEM.*

In Exp. 2, again, we first assessed the general estimation performance. As shown in Fig. 4, the average estimates across the range were noticeably compressed, with overestimation at lower values, and underestimation at higher values. With a series of one-sample t-tests we indeed confirmed that responses for most of the numerosities across the range significantly deviated from the veridical stimulus numerosity, in the overestimation direction for lower numbers (i.e., 8-14; $t(33)$ ranging from 4.89 to 8.17, all p-values < 0.001, d ranging from 0.90 to 1.44), and in the underestimation direction for higher numbers (i.e., 21-32; $t(33)$ ranging from -3.22 to -11.76, all p-values < 0.003, d ranging from 0.49 to

1.95). At intermediate values (16 and 18), instead, the average estimates were not significantly different from the veridical value ($t(33) = 1.96$, $p = 0.058$, and $t(33) = -0.23$, $p = 0.81$, respectively). Using a two-way ANOVA with factors "numerosity" and "experiment," we further compared accuracy measures across the range and across Exp. 1 and Exp. 2. The results show a main effect of numerosity ($F(1,10) = 264.56$, $p < 0.001$, $\eta_p^2 = 0.79$), but no effect of experiment ($F(1,10) = 0.56$, $p = 0.455$). However, we also found a significant interaction between the two factors ($F(1,10) = 5.09$, $p < 0.001$, $\eta_p^2 = 0.07$). A series of post-hoc tests confirmed that numerical estimates deviated to a larger extent from veridical numerosity in Exp. 2 compared to Exp. 1, but only at the extremes of the curve (Exp. 1 vs Exp. 2, 8: $t(64) = 2.72$, $p = 0.007$; 9: $t(64) = 2.58$, $p = 0.010$; 11: $t(64) = 2.56$, $p = 0.011$; 12: $t(64) = 2.03$, $p = 0.043$; 24: $t(64) = 2.11$, $p = 0.036$; 28: $t(64) = 2.54$, $p = 0.011$; 32: $t(64) = 3.41$, $p < 0.001$). For the numerosities in the middle range (14-18), no difference was observed between Exp. 1 and Exp. 2 (all p-values > 0.125). Overall, this pattern shows that the increased difficulty of the task in Exp. 2, due to the presence of multiple stimuli, led to significantly less accurate judgements, especially for numerosities at the lower and higher ends of the range.

Regarding the precision in the task of Exp. 2, we first performed a one-way repeated measure ANOVA with factor "numerosity," to assess whether precision varied as a function of the target numerosity. Differently from Exp. 1, here we found a significant effect of numerosity on CoV measures ($F(10,33) = 48.54$, $p < 0.001$, $\eta_p^2 = 0.59$), showing that precision did significantly vary across the range. In this context, the highest precision was obtained with 32 (0.175 ± 0.007), while, surprisingly, 8 showed the lowest precision (0.411 ± 0.032). On average, CoV tended to be higher in Exp. 2 compared to Exp. 1 (0.254 ± 0.075, compared to 0.217 ± 0.031 in Exp. 1). To directly compare the precision in the task across the two experiments, we performed a two-way ANOVA with factors "numerosity" and "experiment." We observed a main effect of both numerosity ($F(1,10) = 6.80$, $p < 0.001$, $\eta_p^2 = 0.09$) and experiment ($F(1,10) = 11.41$, $p < 0.001$, $\eta_p^2 = 0.02$) on precision measures, but also a significant interaction between the two factors ($F(1,10) = 5.81$, $p < 0.001$, $\eta_p^2 = 0.08$). A series of post-hoc tests further showed that the main difference in precision between the two experiments concerns the lower part of the range (Exp. 2 vs Exp. 1, 8: $t(64) = 7.07$, $p < 0.001$; 9: $t(64) = 2.97$, $p = 0.003$; 11: $t(64) = 2.38$, $p = 0.017$), while CoV across all the other numbers were not significantly different (all p-values > 0.224). This suggests that the higher difficulty of the task in Exp. 2, differently from accuracy measures, mostly impacted estimation performance for relatively low numerosities, while for higher numerosities estimation precision remained similar.



411

412 ***FIGURE 5. Serial dependence effects across the three stimuli presented in Exp. 2.*** *(A) Serial*
413 *dependence effects concerning the first stimulus in the sequence. In this case, we assessed the influence of*
414 *the successive second and third stimulus on the first one. (B) Serial dependence effects at the second*
415 *stimulus in the sequence, provided by the preceding (first) and successive (third) stimulus. (C) Serial*
416 *dependence effects at the third stimulus, provided by the preceding two stimuli. The labels in the boxes*
417 *within each panel indicate the target stimulus chosen. Labels on the x-axis indicate the direction of the*
418 *effect – i.e., the effect of each stimulus in the sequence on the one chosen for the estimation task. Note that*
419 *this pattern of attractive effects is unlikely to be driven by swap errors between different stimuli in the*
420 *sequence, as demonstrated by the simulation analysis reported in the Methods section (see Data*
421 *analysis). Error bars are SEM. ns = not significant, ** $p < 0.01$, *** $p < 0.001$.*

422

423 Regarding the serial dependence effect, we assessed the influence of the three stimuli presented within
424 each trial on each other, individually for the cases where each of the stimuli was selected as the target

one. According to our predictions, we expected two possible effects. On the one hand, a memory interference effect should operate independently from the order of the stimuli. Namely, as the response was provided at the end of the sequence, the three representations held in memory could interact and influence each other independently from their presentation order. On the other hand, a strictly perceptual effect (i.e., for instance in the form of a persistent read-out template as proposed by Pascucci et al., 2019) should operate according to the order of the stimuli: a stimulus should be affected by its preceding one, and not by the successive one. Fig. 5 shows the effects across the sequence of stimuli. As shown in the figure, both kinds of effect seemed to emerge from our stimulation procedure. When the first stimulus was selected as the target one (Fig. 5A), we found a significant attractive influence provided by its immediately successive stimulus (i.e., the second one; one-sample t-test against zero; $t(33) = 2.78$, $p = 0.016$, $d = 0.48$). The third stimulus (Fig. 2C), instead, did not provide any effect ($t(33) = 0.54$, $p = 0.70$). A similar but opposite pattern was observed when the third stimulus was chosen as the relevant one: the immediately preceding one exerted a significant attractive effect ($t(33) = 7.35$, $p < 0.001$, $d = 1.26$), while the first stimulus, further back in the past, did not provide any significant effect ($t(33) = 0.91$, $p = 0.55$). Finally, in the case of the second stimulus – which could be influenced by both its preceding (first) and successive (third) stimulus – we however only observed a forward attractive effect that is provided by the preceding stimulus ($t(33) = 4.43$, $p < 0.001$, $d = 0.76$), and not by the successive one ($t(33) = 0.16$, $p = 0.87$).

We further assessed the overall pattern of results by using a two-way repeated measure ANOVA with factors “target stimulus” (i.e., first, second, third), and “comparison” (coded as 1 and 2 for all pairs of conditions included in each target stimulus condition). With this analysis, we observed a significant main effect of target stimulus ($F(1,2) = 3.25$, $p = 0.045$, $\eta_p^2 = 0.11$), but no effect of comparison type ($F(1,2) = 0.05$, $p = 0.82$). We also observed a significant interaction between the two factors ($F(1,2) = 23.85$, $p < 0.001$, $\eta_p^2 = 0.42$). We then ran a series of post-hoc tests to further characterize the pattern of results. First, we observed a difference within each of the three target stimulus conditions. Namely, in the case of the first stimulus selected as target, the effect provided by the second one was significantly higher that the effect provided by the third stimulus ($t(33) = 2.00$, $p = 0.048$). At the level of the second stimulus, the effect provided by the preceding first stimulus was significantly higher compared to the third one ($t(33) = 3.44$, $p < 0.001$). Finally, in case of the third stimulus selected as target, the effect provided by the preceding second stimulus was significantly higher than the effect provided by the first stimulus in the sequence ($t(33) = 5.01$, $p < 0.001$). Furthermore, we compared the significant effects (as observed with the one-sample t-tests) against each other. The results show that the effect provided by the second stimulus on the third one (Fig. 5C) is significantly higher compared to both the effect of the first on the second one ($t(33) = 2.14$, $p = 0.04$), and the effect of the second one on the first stimulus ($t(33) = 2.95$, p

459 = 0.006). Instead, the effect of the second on the first one (Fig. 5A) and the effect of the first on the
460 second one (Fig. 5B) were not significantly different ($t(33) = 1.21$, $p = 0.23$).

461 Besides the effect within each sequence, we also assessed whether and to what extent the stimuli in the
462 previous trial could affect the estimated numerosity of the stimulus reported in the current trial (see
463 Methods; figure not shown). Namely, we assessed the influence of each stimulus in the previous trial
464 (irrespective of whether it was selected or not as the one to be reported) on the stimulus selected in the
465 current trial (separately for the cases in which the first, the second, or the third was selected). We first
466 performed a two-way repeated measure ANOVA with factor "stimulus in the previous trial" and
467 "stimulus selected in the current trial." The results show no main effect of either stimuli in the previous
468 trial ($F(2,66) = 0.30$, $p = 0.75$) or in the current trial ($F(2,66) = 1.9$, $p = 0.16$), and no interaction between
469 the two factors ($F(4,132) = 2.22$, $p = 0.07$). We also performed a series of one-sample t-tests (against zero
470 effect), but did not observe any significant effect of serial dependence across trials (one-sample t-tests, all
471 adjusted $p > 0.332$). As pooling together previous stimuli irrespective of whether they were selected or
472 not may have somehow masked the effect, we additionally considered the effect of only the stimulus
473 selected in the previous trial on the one selected in the current trial. Running an ANOVA as in the
474 previous test again did not show any main effect of stimuli in the previous ($F(2,66) = 0.63$, $p = 0.54$) and
475 in the current trial ($F(2,66) = 1.67$, $p = 0.20$), and no interaction ($F(4,132) = 2.01$, $p = 0.10$). A series of
476 one-sample t-tests similarly show no significant serial dependence effect (all $p > 0.07$).

477

478 ## 4. DISCUSSION

479 Attractive serial dependence biases appear to be ubiquitous in vision, affecting how we perceive a large
480 variety of visual attributes. The nature of serial dependence, however, is subject to debate. Multiple
481 accounts have been proposed, based on a continuity field supporting visual stability and continuity (e.g.,
482 Fischer & Whitney, 2014), memory and decision biases (Fritsche et al., 2017), high-level modulatory
483 feedback to low-level sensory areas (Fornaciai & Park, 2018b, 2019a, 2019b), or lingering perceptual
484 decision templates at a read-out (high-order) level (Pascucci et al., 2019).

485 In the present study, we aimed to investigate the conditions under which serial dependence arises in
486 numerosity perception. While previous research from our group shows that attractive effects in this
487 context are likely perceptual in nature (Fornaciai & Park, 2018a, 2018b, 2019a, 2019b), biases of different
488 nature may contribute to the result observed at the behavioral level. For instance, memory interference
489 between representations held in memory has been shown to contribute to biases in magnitude perception,
490 across different dimensions like space and time (Cai & Connell, 2015, 2016; Cai et al., 2018). More

specifically, representations of the same format like magnitudes being concurrently held in memory could interact with each other in an attractive way (Cai et al., 2018). Here, we further questioned whether such an effect of memory interference can elicit serial dependence independent from the order of the stimuli. In Exp. 1, we first characterized the serial dependence effect in numerosity estimation by using a simple task, whereby participants had to estimate the number of dots in a single dot-array (i.e., one stimulus with varying numerosity presented in each trial). Crucially, in Exp. 2, we presented a sequence of three stimuli on each trial, and cued the relevant one only after the presentation. This required the participants to hold all the stimuli in memory until the end of the trial, allowing us to test whether memory interference effects can make the temporally latter stimulus to bias the temporally former stimulus.

First, regarding the estimation performance, we observed relatively compressed responses, resembling a logarithmic mapping of increasing numerosity. This pattern is consistent with previous research concerning numerosity perception, showing that numerical estimates across a relatively large range of numerosity often tend to be nonlinearly distributed (e.g., Dehaene, 2003; Anobile et al., 2012a; Anobile et al., 2012b; Arrighi et al., 2014; Cicchini et al., 2014). Moreover, such a logarithmic compression was more severe in the numerical estimates in Exp. 2, likely due to increased difficulty and attentional demands. Indeed, previous studies have shown that the logarithmic-like compression increases under conditions of attentional load (Anobile et al., 2012a; Anobile et al., 2012b), explaining the pattern of accuracy measures in the present study. Interestingly, Cicchini et al. (2014) linked this compressive representation, in the context of number-line mapping, to serial dependence, proposing that an attractive bias could explain the non-linearity in numerical representation. However, while based on such results we may expect a link between the strength of serial dependence and the compressive non-linearity, our data do not provide hints in that direction. Indeed, in our Exp. 2 where the more difficult task resulted in a more severe compression, we observed a somewhat weaker effect (see below). This pattern may suggest that a link between serial dependence and compressed numerical performance may be limited to the number-line mapping task, which may entail different processes compared to numerosity estimation (e.g., see for instance Reinert et al., 2019 for a comparison between different tasks).

We instead observed a different pattern on precision measures (CoV). While accuracy measures were worse across the board in the more difficult task of Exp. 2, CoVs across the different numerosities were more selectively affected, limited to the lower part of the numerical range. Task difficulty and attentional/memory load thus seem to have a greater influence on the estimation performance for low numerosities, while performance in the higher part of the range did not significantly differ across the two experiments. Interestingly, this pattern of results resembles previous observations about the difference in the effects of attentional load on subitizing (i.e., 1-4) and higher (5+) numerosities (Burr et al., 2010).

What has been observed in this context is that attentional load (in the form of a secondary task) strongly affects the precision of numerical estimates for very low numerosities, while it yields only a modest cost for performance with higher numerosities. Although our numerosities are all well beyond the subitizing range, the increased attentional load of Exp. 2 may have more severely affected estimation in the lower part of the range in a similar fashion.

Regarding the serial dependence effect, Exp. 1 shows that numerical estimates in each trial are systematically affected by the recent history of stimulation. In line with Cicchini et al. (2014), but differently from Fischer & Whitney (2014), such an attractive effect was limited to the immediately preceding stimulus, while stimuli further back in the past did not affect current estimates. These results suggest that serial dependence for different features like for instance orientation and numerosity may have different temporal properties (as also suggested by Taubert et al., 2016), potentially suggesting the involvement of (at least partially) independent, domain-specific, mechanisms, rather than a domain-general mechanism. This is however in line with a perceptual account of serial dependence, as different sensory/perceptual pathways dedicated to different attributes are characterized by different physiological and functional properties. Additionally, we also observed a small repulsive effect provided by stimuli at three trials back in the past. To a smaller extent, a repulsive effect seems to emerge also from Cicchini et al.'s (2014) results, at three trials back similarly to the current study (see Fig. 3C in Cicchini et al., 2014), although the effect was not significant. This may indicate that current percepts tend to be slightly repulsed away from more remote stimuli. However, the repulsive effects observed here and in previous studies are very small, and most likely represent only a negligible influence on the actual behavioral performance. Additionally, another difference from previous research – especially in the context of orientation perception (e.g., Fischer & Whitney, 2014; Fritsche et al., 2017; Pascucci et al., 2019) – is the tuned nature of the effect. While several previous studies observed a clear tuning of the effect, with the magnitude of the bias increasing with increasing difference between the current and previous stimuli, peaking at a certain distance in the stimulus space, and then decreasing again with larger differences, here we observed a linear effect (e.g., see Fig. 2B). Namely, our results show that the magnitude of the bias increases as a function of the difference between the two numerosities, peaking at the largest differences (i.e., the smaller the previous numerosity, the stronger the underestimation of current ones, and vice versa). This feature of serial dependence may again be related to the specific structure of the dimension under analysis. While circular dimensions like orientation (or, for instance, the circular positional space in Manassi et al., 2018) show a more precise tuning, magnitude dimensions such as numerosity instead present a linear relation between the effect and the magnitude of previous stimuli (but see also Xia et al., 2016 for a similarly linear effect in the context of face attractiveness).

On the other hand, looking at the extent to which each numerosity in our range is susceptible to serial dependence, we found a peculiar pattern. Indeed, our results first show a repulsive, although not significant, trend at lower numerosity, showing that stimuli like 8 and 9 dots are mostly not affected by the numerosity in the previous trial. This is consistent for instance with Cicchini et al. (2014) and Cicchini et al. (2018) in showing that less noisily-perceived stimuli are more likely to show reduced serial dependence effects, since such a low numerosities at the lower bound of the range are likely encoded with a lower level of noise. Interestingly, 16 – the middle numerosity in the range – was also not affected by serial dependence. This peculiar feature may be related to the long-term summary statistics of the stimulation procedure, as the central value of the range may be represented in a more robust fashion compared to extreme numerosities, making it less prone to distortions. Testing the relation between serial dependence and other processes like central tendency (e.g., Jazayeri & Shadlen, 2010), which may be responsible for this finding, is an interesting possibility for future studies. Furthermore, all numerosities above 16 seem similarly prone to attractive biases. This is in contrast with the sharp reduction in the effect at the higher stimulation range observed by Cicchini et al. (2014). However, the noisier representation of higher numerosities might have favored the effect also at the higher end of the range. Note however that analyzing the effect at the level of single numerosities involved using much smaller subsets of data compared to the overall effect at different trial-back conditions. Due to this, it is difficult to draw strong conclusions about the sensitivity of individual stimuli to the serial dependence effect.

In general, it is interesting to note that despite the fact that we included the entire range of numerosities in the analysis, we did not observe any significant "edge effect", neither on CoVs nor on the serial dependence effect (i.e., a sharp increase in precision at the edges of the range, or a symmetric reduction in the serial dependence effect as in Cicchini et al., 2014) which would instead be expected in this context. The fact that we did not observe such an edge effect may be attributed to the fact that we did not reveal the actual stimulation range to the participants. In fact, the participants were told that they are given a larger range (6 to 40 dots) than it actually was (8 to 32 dots). The increased uncertainty concerning the extreme numerosities might have thus resulted in weaker edge effects.

In Exp. 2, we used a novel paradigm involving a rapid sequence of stimuli, with only one of them – cued after the entire presentation – actually relevant for the task, which creates a condition favoring a mnemonic serial dependence effect. This paradigm differs from the paradigm more often used in serial dependence research (i.e., involving only one stimulus presented in each trial followed by a response, as we did in Exp. 1) in that it requires the participant to keep all the stimuli in memory to perform the task later. While presenting only one stimulus at a time makes serial dependence to be defined as the effect of a previous, irrelevant stimulus on the current one, this novel paradigm allows us to measure the mutual

influence of relevant stimuli on each other. Doing so, we aimed to assess whether the attractive bias still operates in the forward direction (from previous to successive stimuli) or also in the reverse direction (from successive to previous stimuli). This latter case would indeed highlight an exclusively mnemonic serial dependence effect and would demonstrate that it occurs irrespective of the chronological order of stimulus presentation.

Result from Exp. 2 show a pattern of attractive biases among the three stimuli in the sequence working irrespective of their temporal order. That is, in some cases the bias works in the "forward" direction (i.e., a preceding stimulus affecting a successive one) which is in line with the perceptual account, although it is not possible to rule out the concurrent effect of memory bias. Critically, the bias also works in the opposite, "backward," direction, with the chronologically successive stimulus affecting its predecessor. Such a backward effect could only occur at a post-perceptual processing stage, such as memory storage, as the bias takes place only after both stimuli are processed, represented, and stored. Overall, these results demonstrate that attractive biases generated by memory interference operate independently from the order of stimulus presentation.

While the representation of the visual content is stored in memory in order to perform a task at a later time, different representations could be nudged by each other, resulting in an attractive bias (e.g., Cai et al., 2018). Such a bias, at least in the context of the present work, seems limited to the first item in the sequence (i.e., the one farthest back in the past), possibly indicating that a deterioration of stimulus representation with time may induce or facilitate such an interference. However, the fact that we did not observe an effect of the third stimulus on the second or the first one may be attributed to the relatively short interval between its presentation and the onset of the cue (350 ms). There may not have been enough time for the memory interference from the third stimulus to occur, because the memorized representations of the dot arrays are likely to be discarded as soon as the cue is given. The question of how long it takes for a stimulus to affect other stimuli held in memory at the same time thus remains an interesting open question for future studies.

While a parsimonious account of these results is that all the biases observed in Exp. 2 are produced at a memory storage stage, an interesting possibility is that perceptual and memory biases might coexist within the same sequence of stimuli. Indeed, previous research from our group shows that the serial dependence bias is measurable from neural signals starting very early after stimulus onset, suggesting a biased perceptual representation (Fornaciai & Park, 2018a) which then gets stored and maintained for a relatively long retention period after perception (Fornaciai & Park, 2020). In this scenario, the "forward" attractive effect could thus be determined before the stimulus is stored in memory (i.e., the representation stored in memory is already biased), while the "backward" effect is determined at a later stage during

memory encoding. However, behavioral data alone is not enough to conclusively disentangle these two possibilities, and assessing whether biases of different nature (with the same behavioral outcome) could be intertwined within the same sequence of stimuli remains an open question for future studies. In any case, one certain conclusion that our results reach is that serial dependence driven by memory bias occurs even in reverse chronological order in that the representation of one stimulus is affected by the one appearing after that.

While we observe several attractive effects between items included in the same sequence, we did not observe any influence across different trials. A possibility for this lack of influences extending from the previous to the current trial might be due to having in general numerous intervening stimuli between the two trials. Nevertheless, while this easily explains the lack of influence from the previous trial on the second and third stimulus in the current sequence, we may still expect to see an effect from the last stimulus in the previous sequence to the first one in the current trial (i.e., as there is no intervening stimulus between them). However, the lack of across-trial effect even in this case additionally suggests that the memory effect occurring within the current sequence (i.e., the backward effect from the second to the first stimulus) suppresses any influence from the previous trial.

To conclude, our results first show that serial dependence in numerosity perception generalizes to the case in which the approximate visual representation has to be mapped (or re-coded) in a symbolic format. This shows that the serial dependence bias in this context is extremely robust and is present irrespective of the specific task used. More crucially, we show that serial dependence induced by memory interference can operate irrespective of the chronological order of the stimulus presentation. Such a peculiar characteristic of serial dependence opens new doors for understanding the mechanism of this ubiquitous phenomenon.

ACKNOWLEDGMENTS

This study was supported by the National Science Foundation (NSF) CAREER Award (#1654089) to J. P.

DECLARATION OF INTEREST

The authors declare no competing interest.

AUTHOR CONTRIBUTIONS

M.F. and J.P. devised the experiments. M.F. collected and analyzed the data. M.F. and J.P. interpreted the results and wrote the manuscript.

REFERENCES

Alais, D., Leung, J., & Van der Burg, E. (2017). Linear Summation of Repulsive and Attractive Serial Dependencies: Orientation and Motion Dependencies Sum in Motion Perception. *The Journal of Neuroscience*, *37*(16), 4381–4390. https://doi.org/10.1523/JNEUROSCI.4601-15.2017

Anobile, G., Cicchini, G. M., & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition*, *122*(3), 454–459. https://doi.org/10.1016/j.cognition.2011.11.006

Anobile, G., Turi, M., Cicchini, G. M., & Burr, D. C. (2012). The effects of cross-sensory attentional demand on subitizing and on mapping number onto space. *Vision Research*, *74*, 102–109. https://doi.org/10.1016/j.visres.2012.06.005

Arrighi, R., Togoli, I., & Burr, D. C. (2014). A generalized sense of number. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1797), 20141791. https://doi.org/10.1098/rspb.2014.1791

Bays, P. M. (2016). Evaluating and excluding swap errors in analogue tests of working memory. Scientific Reports, 6(1), 19203. https://doi.org/10.1038/srep19203

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. Journal of Vision, 9(10), 7–7. https://doi.org/10.1167/9.10.7

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Boynton, G. M., & Finney, E. M. (2003). Orientation-Specific Adaptation in Human Visual Cortex. *The Journal of Neuroscience*, *23*(25), 8781–8787. https://doi.org/10.1523/JNEUROSCI.23-25-08781.2003

678 Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.
679 https://doi.org/10.1163/156856897X00357

680 Brown, S. P., & Masland, R. H. (2001). Spatial scale and cellular substrate of contrast adaptation by
681 retinal ganglion cells. *Nature Neuroscience*, *4*(1), 44–51. https://doi.org/10.1038/82888

682 Burr, D. C., Turi, M., & Anobile, G. (2010). Subitizing but not estimation of numerosity requires
683 attentional resources. *Journal of Vision*, 10(6), 20–20. https://doi.org/10.1167/10.6.20

684 Cai, Z. G., & Connell, L. (2016). On magnitudes in memory: An internal clock account of space–time
685 interaction. *Acta Psychologica*, *168*, 1–11. https://doi.org/10.1016/j.actpsy.2016.04.003

686 Cai, Z. G., Wang, R., Shen, M., & Speekenbrink, M. (2018). Cross-dimensional magnitude interactions
687 arise from memory interference. *Cognitive Psychology*, *106*, 21–42.
688 https://doi.org/10.1016/j.cogpsych.2018.08.001

689 Cai, Z. G., & Connell, L. (2015). Space–time interdependence: Evidence against asymmetric mapping
690 between time and space. *Cognition*, *136*, 268–281. https://doi.org/10.1016/j.cognition.2014.11.039

691 Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space reflects
692 dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National*
693 *Academy of Sciences*, *111*(21), 7867–7872. https://doi.org/10.1073/pnas.1402785111

694 Cicchini, G. M., Mikellidou, K., & Burr, D. (2017). Serial dependencies act directly on perception.
695 *Journal of Vision*, *17*(14), 6. https://doi.org/10.1167/17.14.6

696 Cicchini, G. M., Mikellidou, K., & Burr, D. C. (2018). The functional role of serial dependence.
697 *Proceedings. Biological Sciences*, *285*(1890), 20181722. https://doi.org/10.1098/rspb.2018.1722

698 Corbett, J. E., Fischer, J., & Whitney, D. (2011). Facilitating stable representations: Serial dependence in
699 vision. *PLoS ONE*, *6*(1). https://doi.org/10.1371/journal.pone.0016701

700 Dehaene, S. (2003). The neural basis of the Weber-Fechner law: a logarithmic mental number line.
701 *Trends in Cognitive Sciences*, *7*(4), 145–147.

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

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743. https://doi.org/10.1038/nn.3689

Fornaciai, M., & Park, J. (2018b). Serial dependence in numerosity perception. *Journal of Vision*, *18*(9), 15. https://doi.org/10.1167/18.9.15

Fornaciai, M., & Park, J. (2019b). 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

Fornaciai, M., & Park, J. (2019a). Spontaneous repulsive adaptation in the absence of attractive serial dependence. *Journal of Vision*, *19*(5), 21. https://doi.org/10.1167/19.5.21

Fornaciai, M., & Park, J. (2018a). Attractive Serial Dependence in the Absence of an Explicit Task. *Psychological Science*, *29*(3), 437–446. https://doi.org/10.1177/0956797617737385

Fornaciai, M., & Park, J. (2020). Neural Dynamics of Serial Dependence in Numerosity Perception. Journal of Cognitive Neuroscience, 32(1), 141–154. https://doi.org/10.1162/jocn_a_01474

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite Effects of Recent History on Perception and Decision. *Current Biology*, *27*(4), 590–595. https://doi.org/10.1016/j.cub.2017.01.006

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception ECVP 2007 Abstract Supplement*, *36*(14), 1–16. https://doi.org/10.1068/v070821

Kohn, A. (2007). Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of Neurophysiology*, *97*(5), 3155–3164. https://doi.org/10.1152/jn.00086.2007

Kohn, A., & Movshon, J. A. (2003). Neuronal Adaptation to Visual Motion in Area MT of the Macaque. *Neuron*, *39*(4), 681–691. https://doi.org/10.1016/S0896-6273(03)00438-0

727 Liberman, A., Fischer, J., & Whitney, D. (2014). Serial dependence in the perception of faces. *Current*
728 *Biology : CB*, *24*(21), 2569–2574. https://doi.org/10.1016/j.cub.2014.09.025

729 Liberman, A., Manassi, M., & Whitney, D. (2018). Serial dependence promotes the stability of perceived
730 emotional expression depending on face similarity. *Attention, Perception, & Psychophysics*, *80*(6),
731 1461–1473. https://doi.org/10.3758/s13414-018-1533-8

732 Manassi, M., Liberman, A., Chaney, W., & Whitney, D. (2017). The perceived stability of scenes: Serial
733 dependence in ensemble representations. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-
734 017-02201-5

735 Manassi, M., Liberman, A., Kosovicheva, A., Zhang, K., & Whitney, D. (2018). Serial dependence in
736 position occurs at the time of perception. *Psychonomic Bulletin & Review*, *25*(6), 2245–2253.
737 https://doi.org/10.3758/s13423-018-1454-5

738 Montaser-Kouhsari, L., Landy, M. S., Heeger, D. J., & Larsson, J. (2007). Orientation-Selective
739 Adaptation to Illusory Contours in Human Visual Cortex. *Journal of Neuroscience*, *27*(9), 2186–
740 2195. https://doi.org/10.1523/JNEUROSCI.4173-06.2007

741 Park, J., Dewind, N. K., Woldorff, M. G., & Brannon, E. M. (2016). Rapid and Direct Encoding of
742 Numerosity in the Visual Stream. *Cerebral Cortex*, *26*(2), 748–763.
743 https://doi.org/10.1093/cercor/bhv017

744 Pascucci, D., Mancuso, G., Santandrea, E., Della Libera, C., Plomp, G., & Chelazzi, L. (2019). Laws of
745 concatenated perception: Vision goes for novelty, decisions for perseverance. *PLOS Biology*, *17*(3),
746 e3000144. https://doi.org/10.1371/journal.pbio.3000144

747 Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into
748 movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

749 Pratte, M. S. (2019). Swap errors in spatial working memory are guesses. Psychonomic Bulletin &
750 Review, 26(3), 958–966. https://doi.org/10.3758/s13423-018-1524-8

751 St. John-Saaltink, E., Kok, P., Lau, H. C., & de Lange, F. P. (2016). Serial Dependence in Perceptual
752 Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. Journal of Neuroscience,
753 36(23), 6186–6192. https://doi.org/10.1523/JNEUROSCI.4390-15.2016

Suárez-Pinilla, M., Seth, A. K., & Roseboom, W. (2018). Serial dependence in the perception of visual variance. *Journal of Vision*, *18*(7), 4. https://doi.org/10.1167/18.7.4

Taubert, J., Alais, D., & Burr, D. (2016). Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports*, 6(1), 32239. https://doi.org/10.1038/srep32239

Watson, A. B. (1979). Probability summation over time. *Vision Research*, *19*(5), 515–522. https://doi.org/10.1016/0042-6989(79)90136-6

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313. https://doi.org/10.3758/BF03194544

Xia, Y., Leib, A. Y., & Whitney, D. (2016). Serial dependence in the perception of attractiveness. *Journal of Vision*, *16*(15), 28. https://doi.org/10.1167/16.15.28