The effect of abstract representation and response feedback on serial dependence in numerosity perception

Michele Fornaciai* & Joonkoo Park

University of Massachusetts Amherst, 135 Hicks Way, Amherst, MA, 01003, USA

* Corresponding author

Michele Fornaciai (michele.fornaciai@gmail.com)

Current address:

International School for Advanced Studies (SISSA)

Via Bonomea 265, Trieste (TS), Italy

ABSTRACT

Serial dependence entails an attractive bias based on the recent history of stimulation, making the current stimulus to appear more similar to its preceding one. Although serial dependence is ubiquitous in perception, its nature and mechanisms remain unclear. Here, in two independent experiments, we test the hypothesis that this bias originates from high-level processing stages at the level of abstract information processing (Exp. 1) or at the level of judgment (Exp. 2). In Exp. 1, serial dependence was induced by a task-irrelevant "inducer" stimulus in a numerosity discrimination task, similarly to previous studies. Importantly, in this experiment, the inducers were either arrays of dots similar to the task-relevant stimuli (e.g., twelve dots), or symbolic numbers (e.g., the numeral "12"). Both dots and symbol inducers successfully yielded attractive serial dependence biases, suggesting that abstract information about an image is sufficient to bias the perception of the current stimulus. In Exp. 2, participants received feedback about their responses in each trial of a numerosity estimation task, which was designed to assess whether providing external information about the accuracy of judgments would modulate serial dependence. Providing feedback significantly increased the attractive serial dependence effect, suggesting that external information at the level of judgment may modulate the weight of past perceptual information during the processing of the current image. Overall, our results support the idea that, although serial dependence may operate at a perceptual level, it originates from high-level processing stages at the level of abstract information processing and at the level of judgment.

Keywords. Serial dependence, numerosity perception, symbolic numbers, response feedback.

INTRODUCTION

The content of our moment-by-moment perception is not a static and independent snapshot of the external world. Instead, what we perceive is strongly influenced by the temporal context in which a given visual scene is embedded in, which can sometimes dramatically change the appearance of the external stimuli or events. A classic example of how the temporal context (i.e., the recent history of stimulation) can affect the perception of a sensory stimulus, is the process of adaptation (e.g., see Kohn, 2007 for a review). After prolonged exposure to a given stimulus, what we perceive afterwards is strongly distorted (i.e., perceptual aftereffect). For example, after a long exposure to a set of 10 items, if the subsequent stimulus is more numerous (e.g., 20 items), its numerosity will be strongly overestimated and perceived as even more numerous than it actually is (e.g., Burr & Ross, 2008; Fornaciai et al., 2016). This strong distortion of perception after adaptation has been linked for instance to functional mechanisms increasing sensitivity to change (e.g., Kaliukhovich & Vogels, 2016). While adaptation has a "repulsive" influence on our perception (i.e., the perception of the adapted stimulus is pushed away from the adaptor, making it seemingly more different than it actually is compared to the adaptor), a new class of "attractive" aftereffects – named "serial dependencies" – has recently been discovered (e.g., Corbett et al., 2011; Fischer & Whitney, 2014). Different from adaptation, serial dependence does not arise from prolonged stimulation, and makes what we are currently seeing to appear more similar to what we saw before (Fischer & Whitney, 2014; Fornaciai & Park, 2018b). For instance, in the typical perceptual paradigm used to measure this effect, the stimulus that a participant has to judge appears to be more similar to the stimulus that the participant judged in the previous trial.

Similar to adaptation, serial dependence has also been shown to be ubiquitous in vision, affecting the perception of basic features such as orientation (Fischer & Whitney, 2014; Pascucci et al., 2019) or numerosity (Corbett et al., 2011; Fornaciai & Park, 2018b, 2019a), as well as more complex features such as face identity (Liberman et al., 2014) or attractiveness (Xia et al., 2016). However, while decades of research have provided good accounts of the physiological bases of adaptation (Kohn, 2007), the mechanisms mediating serial dependence are far from being clear. According to one of the major accounts of serial dependence, this attractive effect would be the consequence of brain mechanisms facilitating the stability and continuity of perception (Liberman et al., 2016; Manassi et al., 2017), integrating past and present information across a spatially and temporally extended "continuity field" (Fischer & Whitney, 2014). This framework has been challenged on the ground that serial dependence seems to be more tightly linked to decision-making rather than perception (e.g., Pascucci et al., 2019; Wehrman et al., 2020), suggesting that it occurs at a more "cognitive" rather than perceptual level. Whether serial dependence is a perceptual or decisional effect thus remains highly debated (Bliss et al., 2017; Cicchini et al., 2017; Manassi et al., 2017; Manassi & Whitney, 2022). When considering these different accounts of serial dependence, however, it is important to note that the roles of perception and cognition are not necessarily mutually exclusive. Indeed, the brain locus or processing stage where serial dependence *originates* from needs not correspond to where it *operates* at. For instance, the effect could originate at a high level of the brain processing hierarchy but affect perception at a much lower level.

In a recent series of works from our group, we have provided electrophysiological evidence that a signature of serial dependence can be decoded from brain responses extremely early after the onset of a stimulus (i.e., 50-200 ms), suggesting that early perceptual processing might indeed be involved in establishing the attractive effect (Fornaciai & Park, 2018a, 2020a). However, in psychophysical studies, we have found that serial dependence nevertheless shows the hallmarks of a high-level effect, at least in numerosity perception. Namely, it depends on attention (Fornaciai & Park, 2018b; see also Fischer & Whitney, 2014; Fritsche & de Lange, 2019), it works even if the numerosity of the past and the current stimulus is conveyed by very different stimuli (i.e., a series of flashes versus an array of dots; Fornaciai & Park, 2019b), and depends on the perceived, rather than physical, properties of a stimulus (Fornaciai & Park, 2021). Further evidence has shown that serial dependence emerges only when a past stimulus is consciously perceived, and that this attractive bias disappears when conscious perception is suppressed via backward masking (Fornaciai & Park, 2019a, 2021). Based on this evidence, we have proposed the idea that serial dependence originates at a high-level processing stage, and it operates – effectively modulating perception – via feedback to earlier visual areas (e.g., Fornaciai & Park, 2019a). This framework could thus potentially reconcile different accounts based on perception versus cognition and decision-making.

In the present study, we further test two additional predictions based on the hypothesis that serial dependence in numerosity perception originates at a high-level of the visual processing pathway. First, if this idea holds true, then serial dependence for numerosity perception represented in dot arrays should be induced by the numerical magnitude conveyed by a symbolic number. Previous results indeed show that parietal cortices similarly represent the magnitude of both symbolic and non-symbolic stimuli (Piazza et al., 2007). Additionally, a similar, but distinct effect – numerical priming – has been shown to generalize across symbolic and non-symbolic stimuli (Bahrami et al., 2010). Although the priming effect often shows a more semantic rather than perceptual nature (e.g., Naccache & Dehaene, 2001), this supports the idea that numerical representations gets abstracted from the specific sensory inputs at a high level of the brain processing hierarchy (see also Arrighi et al., 2014). Therefore, if serial dependence originates at such a high-level stage, symbolic numbers should be able to affect the perceived numerosity of an array of dots. Second, if serial dependence arises at a high-level processing stage at the level of judgment, the effect might also be weighted by external information about that judgment. Previous results have shown that serial dependence can be modulated by a subjective evaluation of performance (i.e., confidence; Suarez-Pinilla et al., 2018; Samaha et al., 2019). Here we further predict that serial dependence might also be modulated by external – i.e., objective – information concerning the accuracy of judgments as that information could modulate the underlying perceptual representations.

To test these predictions, we performed two independent experiments. In Exp. 1, we used a numerosity discrimination task, whereby participants compared the numerosity of a constant reference against a variable dot-array probe. In this paradigm, serial dependence is induced by an "inducer" stimulus irrelevant to the discrimination task, presented before the reference. In two separate conditions performed by the same participants, the inducer could either be a dot-array (i.e., an array of either 12 or 24 dots; Fornaciai & Park, 2018b, 2019a), or a symbolic number (i.e., either the numeral "12" or the numeral "24"). If serial dependence arises based on a truly abstract representation of numerical magnitude, then both types of inducer should affect the perceived numerosity of the reference to a similar extent. In Exp. 2, we used a numerosity estimation task, whereby participants estimated the numerosity of an array of dots presented in each trial. In order to actively modulate the apparent reliability of perceptual judgments, we provided a feedback after each response, indicating whether the estimate was too low, about right, or too high. The pattern of serial dependence effects obtained in the presence of response feedback was then compared with the data obtained without feedback. If serial dependence arises from a high-level processing stage, then the effect should be modulated by external information concerning the accuracy of perceptual estimates, and particularly when the response in the previous trial was marked as correct.

METHODS

Participants

A total of 35 participants was tested in Exp. 1, and a total of 20 participants were tested in Exp. 2, for a total of 55 participants tested across the two experiments (39 females; mean age ± SD = 21.1 ± 2.2 years). All the participants had normal or corrected-to-normal vision and audition, had no history of neurological, psychiatric, or developmental disorders, and signed a written informed consent form before taking part in the study. Subjects were compensated for their participation with $8/hour. All the experimental procedures were approved by the internal review board of the University of Massachusetts Amherst, and were in line with the declaration of Helsinki. A total of four participants were excluded from data analysis in Exp. 1 due to poor performance (see *Data analysis* for more information about the exclusion criteria), while two participants were excluded from data analysis in Exp. 2 since they did not complete the entire experimental session.

The sample size of Exp. 1 was determined a priori, based on a power analysis including the average effect size of serial dependence (considering comparisons tested with t-tests) in previous studies from our group employing a similar methodology. Namely, we considered the effect size observed in Exp. 1 and Exp. 2 of Fornaciai & Park, 2018b, the effect size in Exp. 1 and Exp. 2 of Fornaciai & Park, 2019a, and of Exp. 1 and Exp. 2 of Fornaciai & Park, 2019b. The average effect size (Cohen's d) of serial dependence was 0.65. Based on this effect size, a power of 95%, and a two-tailed distribution, we estimated a minimum required sample size of 33 participants. On the other hand, the sample size of Exp. 2 was based on the effect size of the serial dependence effect induced by the immediately preceding stimulus (n-1) in the numerosity estimation task used in Fornaciai & Park, 2020b. Based on such effect size ($d = 1.64$), a similar power analysis resulted in a minimum sample size of 8 participants. However, this estimate was based on the pure effect of serial dependence against the null hypothesis of zero effect. Since in this context we were instead interested in the difference in serial dependence across two conditions (feedback and no-feedback), we doubled such estimate, aiming to test at least 16 participants.

Apparatus and stimuli

Both Exp. 1 and Exp. 2 were performed in a quiet and dimly lit room, with participants sitting about 80 cm away from a monitor screen. Stimuli were presented on a 1920 x 1080 pixels monitor screen running at 144 Hz, which encompassed 35 x 20 degrees of visual angle (deg) from the viewing distance of 80 cm. All the visual stimuli were generated using the routines of the Psychophysics toolbox (Brainard, 1997; Kleiner M et al., 2007; Pelli, 1997) in Matlab (version r2016b, The Mathworks, Inc.).

In Exp. 1, the main task-relevant stimuli ("reference" and "probe") were arrays of black and white dots (equal proportion of black and white; in case of odd numerosities, the color of the exceeding dot was determined randomly). In addition to the two task-relevant stimuli, we also presented a task-irrelevant "inducer" stimulus, which could either be a dot-array (non-symbolic inducer condition), or a symbolic number (symbolic inducer condition), with these two conditions performed by the same participants in two separate sessions. The symbolic inducer was either a "low" or a "high" number (i.e., lower or higher than the reference, see below). The number presented as inducer could either span from "10" to "14" (in the case of the low inducer), or from "22" to "26" (in the case of the high inducer). The numerals composing the symbolic inducer had a font size of 100 pt. The non-symbolic inducer was instead an array of either 12 (low inducer) or 24 (high inducer) dots. The range of different symbolic inducers was used with the rationale of mimicking the variability in perception of the non-symbolic one.

The task-relevant stimuli were identical in both the symbolic and non-symbolic inducer condition. Namely, the inducer was followed by the presentation of a constant reference dot-array, always containing 16 dots. Finally, we presented a probe stimulus varying in numerosity from trial to trial (8, 10, 13, 16, 20, 25, or 32 dots). The dot-array stimuli were constructed to vary along three orthogonal dimensions (numerosity, size, spacing), following the stimulus design introduced by DeWind et al., 2015 and Park et al., 2016. The dimensions of size and spacing were constructed by scaling (log2) and combining the dimensions of individual and total dot size, and field area (i.e., the virtual circular area containing the array) and sparsity (i.e., the inverse of density), respectively. For more information about this stimulus construction procedure, see DeWind et al. (2015) and Park et al. (2016). The radius of each dot ranged from 7.5 to 15 pixels (0.13-0.26 deg), while the radius of the field area of the array ranged from 150 to 300 pixels (2.59-5.17 deg). Within each array, all the dots had the same size, and their position was randomly determined to fit within the field area, with the only constraint of a minimum inter-dot distance equal to the radius of a dot. A pool of 1,000 arrays for each numerosity level was generated offline according to this procedure, and stimuli were randomly selected from this pool during the experiment. All the stimuli were presented on the screen for about 28 ms (4 screen frames).

In Exp. 2, all the stimuli were arrays of black and white dots, varying in numerosity (8, 9, 11, 12, 14, 16, 18, 21, 24, 28, 32), with only one stimulus presented centrally in each trial. Dot arrays in Exp. 2 were constructed using the same stimulus construction procedure as Exp. 1 and had similar ranges in dot radius

(7.5-15 pixels) and field radius (150-300 pixels). As in Exp. 1, we constructed offline a pool of 1,000 stimuli for each level of numerosity, and randomly selected the stimulus in each trial drawing from these pools. When the feedback was given, the participant's response was followed by a brief message indicating whether the reported numerosity was either correct (“CORRECT” presented in green font), underestimated (“TOO LOW” reported in red font), or overestimated (“TOO HIGH” reported in red font). The feedback appeared in the lower part of the screen in the same position were participants typed the response (font = 30 pt).

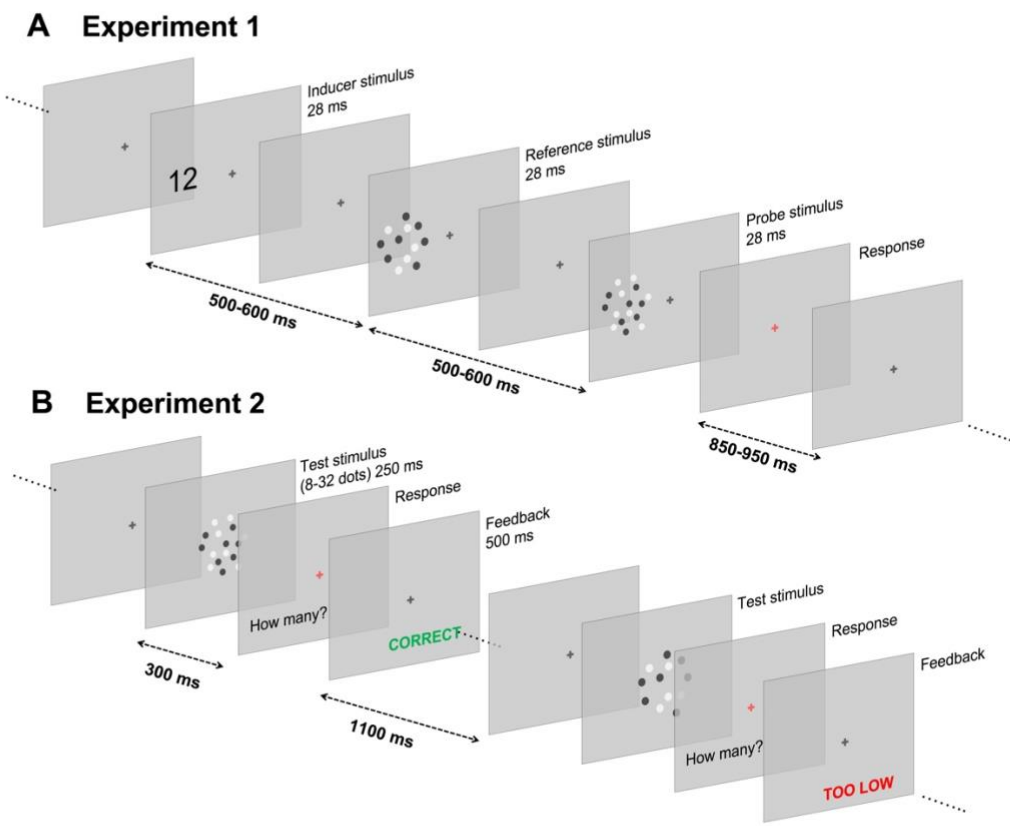

FIGURE 1 – Experimental procedure. *(A) Experimental procedure of Exp. 1, showing the symbolic inducer condition. The stimulus sequence in Exp. 1 involved the presentation of a task-irrelevant inducer stimulus – aimed to induce serial dependencies – followed by a constant reference (16 dots) and a variable probe (8-32 dots). Each stimulus was presented on the screen for 28 ms. The experiment was divided into two different sessions, in which we used two types of inducer. In the symbolic inducer condition (which is*

shown in the figure), the inducer was a symbolic number representing a magnitude either lower ("10"-"14") or higher ("22"-"26") compared to the reference. In the non-symbolic inducer condition (not shown in the figure), the inducer was an array of dots containing either 12 or 24 dots. Every other aspect of the procedure was identical across the two conditions. At the end of each trial, participants indicated whether the reference or the probe seemed to contain more dots. In each of these two conditions, participants performed 4 blocks of 56 trials. (B) Experimental procedure of Exp. 2, showing the feedback condition. In Exp. 2, participants performed a numerosity estimation task, reporting an exact estimate of the numerosity of a dot array presented in each trial for 250 ms. The experiment consisted of 12 blocks of 55 trials, with the first 8 blocks providing feedback (shown in the figure), and the last 4 blocks without feedback (not shown in the figure). In the feedback condition, each response was immediately followed by a feedback informing the participant about whether the response was "too low," "correct," or "too high." The no-feedback condition was identical to the feedback condition, but no information about the response was provided to the participants. Stimuli are not depicted in scale.

Procedure

Experiment 1

In Exp. 1, participants performed a numerosity discrimination task, comparing the numerosity of a constant reference with a variable probe dot-array. To induce serial dependencies, the reference was always preceded by a task-irrelevant "inducer" stimulus, which was either a dot-array (non-symbolic inducer condition), or a symbolic number (symbolic inducer condition), depending on the session. Throughout the experiment, participants were asked to keep their gaze on a central fixation cross. Each trial started with the presentation of the inducer stimulus (28 ms). After an inter-stimulus interval (ISI) of 500-600 ms, the constant reference stimulus was presented on the screen (28 ms) and followed by the variable probe (28 ms) after another ISI of 500-600 ms. All the stimuli were presented either on the left or on the right of the central fixation cross (randomized across trials), with a horizontal eccentricity of 8.75 deg from the center of the screen. After the offset of the probe, the fixation cross turned red, signaling the end of the trial. Participants were instructed to judge whether the reference or the probe contained more dots, providing a response using a standard keyboard. After providing a response, the trial started automatically after an inter-trial interval of 850-950 ms. Before the start of the experiment, participants were told that the inducer (the first stimulus in the sequence) was always irrelevant for the task that they had to carry out, but were encouraged anyway to pay attention to the entire sequence of stimuli to avoid getting distracted. Besides the type of the inducer (symbolic vs. non-symbolic), the procedure employed in the two conditions was identical. In each condition, participants performed 4 blocks of 56 trials, for a total of 16 repetition of each inducer (i.e., low vs. high) and probe magnitude. Fig. 1A shows a depiction of the symbolic inducer condition of Exp. 1. Note that although both reference and probe were presented in the same position as the inducer, the effect is expected to be limited to the reference stimulus. Indeed, serial dependence in numerosity perception shows a limited temporal profile, with an effect only provided by the immediately preceding stimulus (Fornaciai & Park, 2020; see also the results of Exp. 2 below).

Experiment 2

In Exp. 2, participants performed a numerosity estimation task, estimating the exact numerosity of a single dot-array presented in each trial. While participants fixated on a central fixation cross, a dot array appeared on the center of the screen for 250 ms. After 300 ms from the offset of the stimulus, the instruction "How many?" appeared in the lower portion of the screen (about 2 deg below the fixation cross). Participants were then instructed to type the estimated number of dots by using the numerical pad of a standard keyboard. The number appeared on the screen while they typed, and participants could also use the backspace to change response if needed. Finally, they confirmed their estimate by pressing enter. During the response phase, participants were allowed to look away from the fixation cross in order to better confirm their response. The experiment was divided into 12 blocks of 55 trials. In the first 8 blocks of the session, we provided feedback concerning the response, appearing on the screen immediately after the response was confirmed and remaining on the screen for 500 ms. The feedback informed the participants about whether their response was too low (underestimation), correct, or too high (overestimation). Considering the approximate nature of numerosity perception, providing a truly exact estimate of a dot-array is however very difficult. We thus considered a correct response any number within a range of ±1 around the veridical numerosity. For example, if the numerosity was 16 and the participant responded either 15 or 17, the response was considered correct for the purpose of the feedback. After the presentation of the feedback, the next trial started automatically after 1100 ms. In the remaining 4 blocks of the session, no feedback was presented, and the screen was left blank for the same amount of time. The true range of numerosity presented throughout the experiment (8-32) was not revealed to the participants to reduce edge effects, but participants were told that they could not respond less than 6 or more than 40. All the responses outside this range were excluded from data analysis. Finally, at the end of the experimental session, participants were asked whether they found the response feedback actually useful to perform the task. Fig. 1B shows a depiction of the procedure of the feedback condition. Note that we chose to include more blocks in the feedback condition in order to exclude the initial part of the experiment from our main analysis. That is, we considered the first 4 blocks of trials as training, in order for the participants to get used to the response feedback and learn the task. The no-feedback part of the experiment was always performed after the feedback part to ensure a similar level of training, in other words to reduce any difference in performance due to the amount of training.

Data analysis

Data analysis in Exp. 1 was performed by assessing the proportion of "probe more numerous" responses as a function of probe numerosity, separately from the different inducer magnitudes (i.e., low vs. high) and conditions (symbolic vs. non-symbolic). A cumulative Gaussian function (psychometric function) was fitted to the proportion of responses across the probe range, according to the maximum likelihood method (Watson, 1979). As a measure of goodness of fit, we computed the R^2 of the psychometric fit to the data. The R^2 across the group in the symbolic inducer condition spanned from 0.29 to 0.73 (mean ± SD = 0.51 ± 0.13), while in the non-symbolic inducer condition it spanned from 0.37 to 0.75 (mean ± SD = 0.58 ± 0.09). We then defined the point of subjective equality (PSE) as the median of the psychometric fit. The PSE provided a measure of the participants' accuracy in the task, which reflects the perceived numerosity of the reference. To better assess the effect of the inducer in the different conditions, we also computed a serial dependence effect index as follows:

Serial dependence effect = $((PSE_{high} - PSE_{low}) / PSE_{low}) \times 100$;

Where PSE_{high} refers to the PSE obtained in the presence of an inducer with higher numerical magnitude compared to the reference, and PSE_{low} to the PSE obtained with an inducer magnitude lower than the reference. As a measure of precision in the task, we first computed the just noticeable difference (JND) based on the slope of the psychometric fit, and set this value to reflect the difference in probe numerosity between chance level and 75% "probe more numerous" responses. Note that due to the symmetry of the psychometric fit around its median, setting the threshold level above (75%) or below (25%) the median does not change the absolute value of the JND. The JND was then used to set an exclusion criterion reflecting the level of performance in the task. Namely, we excluded all the participants exceeding the upper JND boundary based on the third quartile (Q3) and interquartile range (IQR) of the group (i.e., JND ≥ Q3 + 1.5 × IQR; JND ≥ 7.9 dots). A total of 4 participants was excluded based on this criterion. We also computed the Weber's fraction (WF = JND/PSE) as an additional measure of precision.

In Exp. 2, we first assessed the general performance in the feedback and no-feedback condition by computing the average numerical estimates across the different levels of numerosity, and the precision in the task in terms of WF (in this case, computed as the standard deviation of numerical estimates at each numerosity level divided by the average estimate). Additionally, we assessed other measures of performance such as response times and estimation error (in this case, the average of the absolute difference between the veridical numerosity and the estimation response) for each block of trials, in order to track how the performance changed across different blocks and different parts of the experiment. Serial dependence was examined following the procedure previously used in Fornaciai & Park, 2020b. First, we assessed the extent to which the judgment of the stimulus in the current trial (n) is influenced by the stimulus in the immediately preceding trial (n-1) and/or in trials further back in the past (n-2, n-3, and so on up to n-7). To do so, we computed the estimation error in each trial and sorted it according to the magnitude of the stimulus in the preceding trial. We then fitted a linear function to the data (see Fig. 4), and the slope of the linear fit was taken as an index of the attractive serial dependence effect. A positive slope indicates that the estimation error tends to become more positive (i.e., more towards overestimation) when the preceding magnitude was high, and vice versa (i.e., more negative response errors) when the preceding magnitude was low. A negative slope instead would index an opposite, repulsive effect. In order to ensure that participants had correctly learned the task and familiarized with the response feedback, we excluded the first four blocks of the session. That is, we included in data analysis only the last four blocks of the feedback condition (i.e., blocks 5 to 8) and all the blocks of the no-feedback condition (i.e., blocks 9 to 12). Finally, we assessed the role of the feedback provided in the immediately preceding trial (n-1) on the strength of the serial dependence effect, by sorting the data according to whether the feedback provided in the previous trials was "too low," "correct," or "too high."

In both experiments, frequentist statistics (t-tests, ANOVAs) was complemented by Bayesian tests. The frequentist tests results were thus reported along with the Bayes factor (BF), computed assuming a standard Cauchy prior (scale = 0.707), indicating to what extent the data supports the null or the alternative hypothesis. BF values below 0.33 are interpreted as evidence for the null hypothesis (i.e., no difference between two conditions), while BF values above 3 are considered evidence in favor of the alternative hypothesis (i.e., significant difference between two conditions). Values closer to 1 ($0.33 \leq BF \leq 3$) are instead considered as anecdotal evidence for either the null or the alternative hypothesis, depending on whether the value is lower or higher than 1, respectively. Bayes factors were computed for t-tests, Pearson's correlation tests, and the main effects of ANOVAs.

The data analysis and statistical tests were performed using Matlab (r2021b, The Mathworks, Inc.) and JASP (v0.16; JASP Team, 2021).

Data availability

All the data generated during the experiments described in this manuscript have been uploaded to Open Science Framework, and can be accessed following this link: https://osf.io/veysa/.

RESULTS

Experiment 1

In Exp. 1, we assessed whether serial dependence in numerosity perception could similarly be induced by a symbolic and a non-symbolic magnitude. Participants (N = 31 after applying the exclusion criterion) performed a numerosity discrimination task similar to what we used in previous studies (Fornaciai & Park, 2018b, 2019a, 2019b, 2021), comparing the numerosity of a constant reference (16 dots) with the numerosity of a probe varying from trial to trial (8-32 dots). In two separate conditions, serial dependence was induced by either a dot-array (with numerosity either lower, 12 dots, or higher, 24 dots, compared to the reference) or a numeral (with magnitude either lower, e.g., "12," or higher, e.g., "24," compared to the reference).

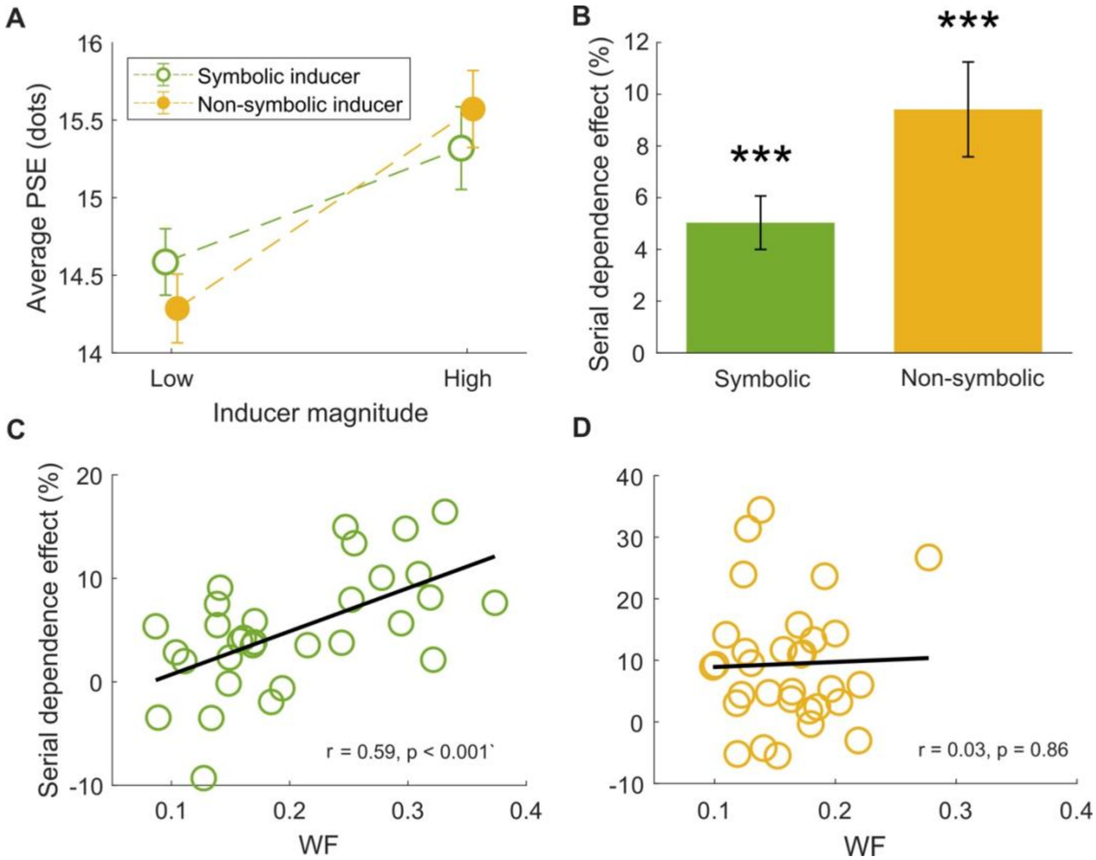

FIGURE 2 – Results of Experiment 1. *(A) Average point of subjective equality (PSE) as a function of the magnitude of the inducer, for the symbolic and non-symbolic inducer conditions. (B) Average serial dependence effect index in the two conditions. (C) Individual measures of the serial dependence effect index plotted as a function of the precision in the task (Weber fraction, WF), in the symbolic inducer condition. (D) Individual measures of the serial dependence effect index plotted as a function of WF in the non-symbolic inducer condition. The black lines are linear fits to the data. Error bars are SEM. *** $p < 0.001$.*

The results of Exp. 1 are shown in Fig. 2. First, we assessed the average point of subjective equality (PSE), which provides a measure of the perceived numerosity of the reference stimulus, as a function of the inducer magnitude and separately for the symbolic and non-symbolic inducer condition (Fig. 2A). As shown in Fig. 2A, we observed a robust relative under- and over-estimation of the reference numerosity congruent with the inducer magnitude. Namely, when the inducer magnitude was lower than the reference, the reference itself was perceived as relatively less numerous (PSE = 14.58 ± 1.19 and 14.29 ± 1.23 dots, respectively for the symbolic and non-symbolic inducer) compared to when the inducer had a higher numerical magnitude (PSE = 15.32 ± 1.48 and 15.57 ± 1.38 dots). To better assess the strength of the effect induced by symbolic and non-symbolic inducer stimuli, we computed a serial dependence effect index based on the normalized difference in PSE between the two inducer magnitudes (Fig. 2B). The effect turned out to be robust and significant across the two conditions (one-sample t-test against zero, $t(30) = 4.86$, $p < 0.001$, Cohen's $d = 0.87$, Bayes factor, $BF = 675$, and $t(30) = 5.14$, $p < 0.001$, $d = 0.92$, $BF = 1390$, respectively for the symbolic and non-symbolic condition), and no significant difference was observed across them (9.41% ± 10.20% vs. 5.03% ± 5.75%; paired t-test, $t(30) = 2.04$, $p = 0.051$, $BF = 1.17$).

In terms of precision in the task, we analyzed the Weber's fraction (WF), and found a significantly better performance (i.e., lower WF) in the non-symbolic inducer condition (0.16 ± 0.04 vs. 0.20 ± 0.08, respectively for the non-symbolic and symbolic inducer condition; paired t-test, $t(30) = 3.15$, $p = 0.004$, $d = 0.57$, $BF = 10.49$). Namely, despite the task-relevant stimuli were effectively identical across the two conditions, participants showed significantly more consistent (i.e., less variable) responses in the non-symbolic condition. We then assessed whether the individual level of precision in the task might modulate the strength of the serial dependence effect. Indeed, it has often been reported that the weight of past information determining serial dependence depends on the perceptual uncertainty entailed by the stimuli or task (Cicchini et al., 2018; Fornaciai & Park, 2019b). In the symbolic inducer condition, we observed a significant correlation between the strength of serial dependence and the precision in the task ($r = 0.59$, $p < 0.001$, $BF = 75.67$), showing that the higher the WF (i.e., the lower the precision), the stronger the effect. However, no such a correlation was observed in the non-symbolic inducer condition ($r = 0.03$, $p = 0.86$, $BF = 0.23$).

Overall, although with some differences, the results from Exp. 1 show that attractive serial dependence for non-symbolic numerosity perception could be induced even by a symbolic inducer, with a similar effect size compared to a non-symbolic inducer (i.e., 0.87 and 0.92, respectively), demonstrating that serial dependence can originate from abstract representations of numerical magnitude.

Experiment 2

In Exp. 2, in line with our overarching hypothesis of serial dependence originating at a high-level processing stage, we asked whether the effect could be modulated by external information about the accuracy of perceptual judgements. To address this question, we used a numerosity estimation task, whereby the participants (N = 18) estimated the numerosity of a dot-array ("test" stimulus) presented in each trial. The experiment was composed of 12 blocks of 55 trials, with the initial 8 blocks including a response feedback (i.e., "too low," "correct," or "too high") after each response, and the final 4 blocks without such feedback. This task structure was specifically chosen in order to be able to remove the first four blocks of each session, which were considered as a training aimed to ensure that participants familiarized with the task and with the response feedback. Indeed, as the level of performance in some cases correlates with the strength of serial dependence (as shown in Exp. 1), this alone might provide a difference in serial dependence independently from the feedback provided to participants. In our analysis, we thus excluded the first four blocks of trials.

First, we assessed the general performance in the numerosity estimation task. Fig. 3A and B show the average estimated numerosity at each level of the range (8-32). The pattern of average responses across different numerosities was very similar across the two conditions, suggesting that the overall performance did not differ much in the presence of response feedback. A two-way repeated measures ANOVA on estimated numerosity with factors "numerosity" (i.e., the different levels of the numerosity range) and "feedback" (i.e., feedback vs. no-feedback) showed only the main effect of numerosity ($F(10,170) = 267.04$, $p < 0.001$, $\eta^2_p = 0.94$, BF > 2000), with no effects of feedback ($F(1,17) = 0.029$, $p = 0.87$, BF = 0.11) or the interaction ($F(10,170) = 0.61$, $p = 0.80$). We then assessed the average Weber fraction (WF) across the different levels of numerosity, computed as the standard deviation of responses at each level divided by the average response. As shown in Fig. 3C, the level of precision in the task was very similar irrespective of feedback. No significant difference was observed in the average WFs (paired t-test, $t(17) = 1.79$, $p = 0.09$, $BF = 0.91$).

Besides these main analyses, we also assessed how different measures of performance evolved across different blocks (i.e., blocks 5 to 8 for the feedback condition, and blocks 9 to 12 for the no-feedback condition). First, we considered the average absolute estimation error across the different blocks of trials, which is shown in Fig. 3D. A two-way repeated measures ANOVA on this absolute estimation error with factors “block” and “feedback” showed a significant main effect of block ($F(3,51) = 4.34$, $p = 0.008$, $\eta^2_p = 0.20$, $BF = 3.11$), no main effect of feedback ($F(1,17) = 0.55$, $p = 0.47$, $BF = 0.28$), and no interaction ($F(3,51) = 1.22$, $p = 0.31$). As the most interesting difference might occur at the transition between the feedback and no-feedback condition (block 8 and 9), we directly compared the error rates across these two successive blocks. However, no significant difference was observed (paired t-test, $t(17) = 1.39$, $p = 0.21$, $BF = 0.55$). Moreover, we also assessed the WF across the different blocks. The results showed no significant main effect of block number ($F(3,51) = 1.16$, $p = 0.33$, $BF = 0.13$), no significant main effect of feedback ($F(1,17) = 2.55$, $p = 0.13$, $BF = 1.15$), and no interaction ($F(3,51) = 0.017$, $p = 0.99$). No significant difference was also observed at the transition between the two conditions ($t(17) = 1.64$, $p = 0.12$, $BF = 0.75$). Finally, we looked at the average response times in the different blocks. Response times appeared to show a gradual decline over the last blocks of the feedback condition (i.e., participants becoming faster in performing the task), and a slight increase at the transition to the no-feedback blocks. A two-way repeated measures ANOVA showed a significant main effect of block number ($F(3,51) = 4.74$, $p = 0.005$, $\eta^2_p = 0.22$, $BF = 1.36$), but no main effect of feedback ($F(1,17) = 0.65$, $p = 0.43$, $BF = 0.27$) or interaction ($F(3,51) = 0.40$, $p = 0.76$). At the transition between conditions, however, response times appeared to significantly increase at the first no-feedback block ($t(17) = 2.54$, $p = 0.02$, $d = 0.31$, $BF = 2.86$), although the effect size of this difference is low.

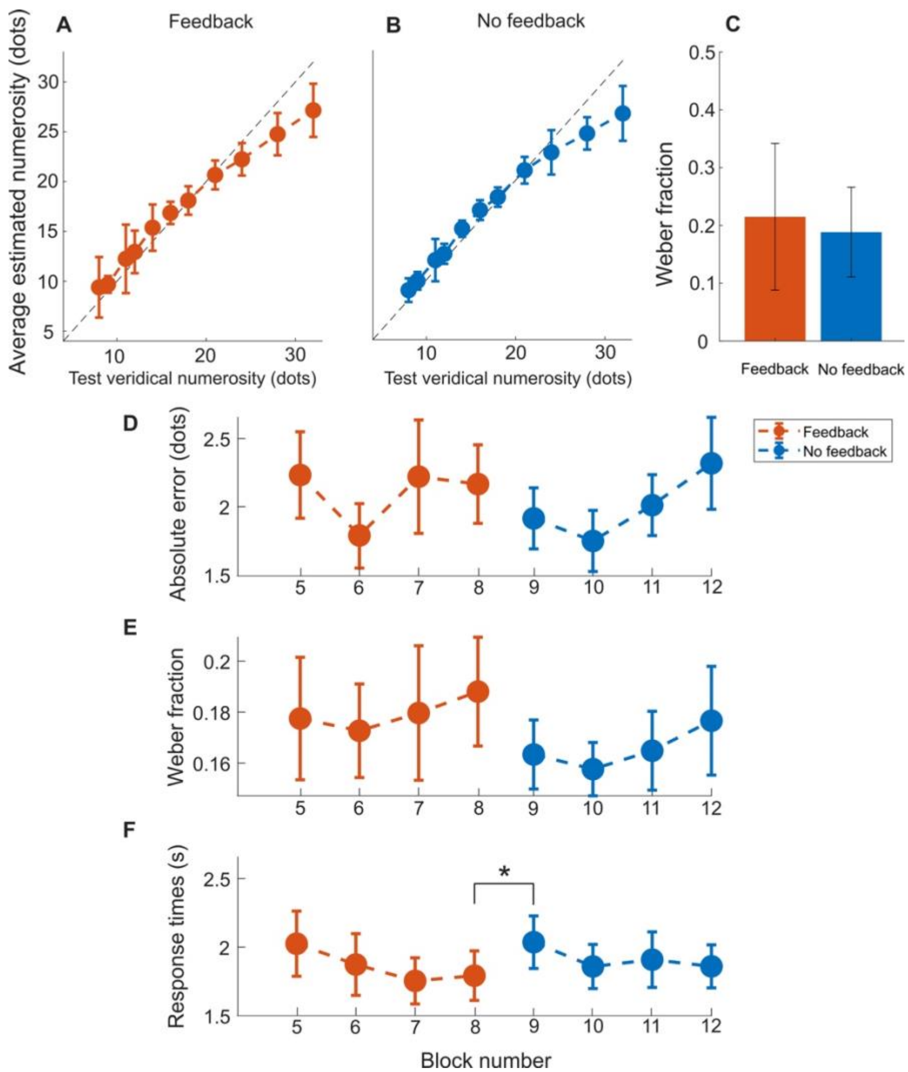

FIGURE 3 – General performance measures in Exp. 2. *(A) Average numerical estimates at each level of the numerosity range, in the feedback condition of Exp. 2. (B) Average numerical estimates in the no-feedback condition. (C) Average Weber's fraction (WF) across the two conditions. (D) Average absolute estimation error computed in each individual block included in the analysis. (E) Average WF in each individual block. (F) Average response times in each individual block. Error bars are SEM. * $p < 0.05$.*

After assessing the basic measures of participants' performance in the task, we then addressed the serial dependence effect across the two conditions. Serial dependence was assessed individually for each participant and condition, arranging the estimation error measured in each trial (trial n) as a function of the numerosity of the stimulus presented in the immediately preceding trial (n-1) or other trials further back in the past (n-2, n-3, and so on up until n-7). We then fitted a linear model to the data arranged in this fashion, and the slope of the fit was taken as a measure of serial dependence (as we did in Fornaciai & Park, 2020). Fig. 4 shows an example of single-subject data from a representative participant.

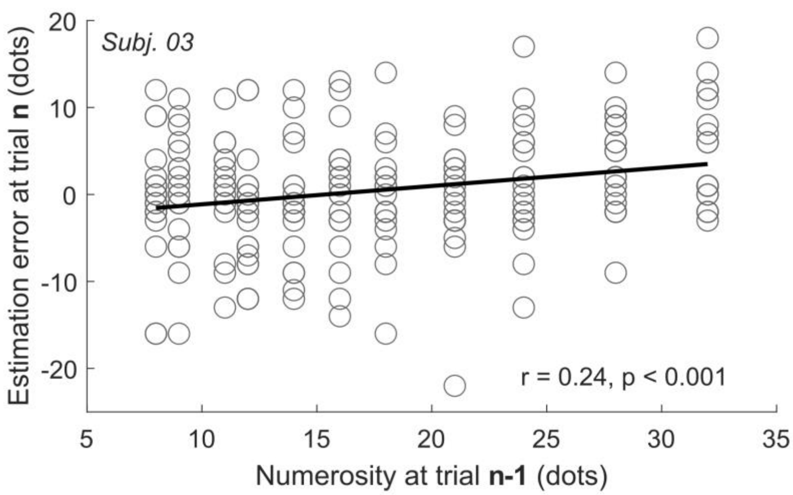

***FIGURE 4 – Example of the serial dependence effect**. The figure shows data from a representative subject in the feedback condition. To assess the serial dependence effect in this context, we arranged the data according to the response error in the current trial (trial n) as a function of numerosity in the previous trial (n-1 in this example, but the procedure was identical to the different n-back cases). We then fitted a linear model to the data arranged in this fashion, and the slope of the fit was taken as a measure of the strength of serial dependence. A positive slope indicated that the estimation error tends to be more positive (overestimation of the current stimulus) when a higher numerosity was presented in the previous trial, and more negative (underestimation) when a lower numerosity was presented in the previous trial. Note that before data analysis, all the responses clearly outside of the range (i.e., less than 6 or more than 40) were excluded from data analysis.*

The average serial dependence effects induced by the immediately preceding stimulus (n-1) and stimuli further back in the past are shown in Fig. 5. For comparison purposes, the figure also shows the results from our previous study (Fornaciai & Park, 2020b), which involved a similar methodology but without any feedback provided to participants. Importantly, in that previous study participants performed the task without any response feedback right from the beginning of the session. These data thus provide a useful comparison to assess whether having the no-feedback blocks always preceded by the feedback blocks affects the pattern of results.

In general, the most prominent attractive serial dependence effect was provided by the immediately preceding stimulus (n-1), while the effect of stimuli further back in the past approached zero or even turned slightly negative (i.e., potentially indexing a repulsive effect). We first performed a series of one-sample t-tests against zero, corrected for multiple comparisons using the false discovery rate (FDR) procedure ($q = 0.05$). In the feedback condition, we observed a significant serial dependence effect at n-1 ($t(17) = 7.59$, adjusted-$p < 0.001$, $d = 1.81$, $BF > 2000$). Additionally, we observed a significant repulsive effect at n-4 ($t(17) = -2.86$, adj-$p = 0.037$, $d = 0.69$, $BF = 4.94$), which is in line with our previous results (Fornaciai & Park, 2020). No other effect reached significance after the FDR correction (all adj-$p > 0.05$, max $BF = 0.46$). In the no-feedback condition, we only observed a statistically significant effect at n-1 ($t(17) = 6.46$, adj-$p < 0.001$, $d = 1.5$, $BF > 2000$), and no other significant influence from stimuli further back in the past (all adj-$p > 0.05$, max $BF = 0.84$). To assess the difference between the feedback and no-feedback conditions at n-1, we performed a paired t-test, which showed a statistically significant difference ($t(17) = 2.86$, $p = 0.01$, $d = 0.80$, $BF = 4.94$).

Interestingly, the strength of serial dependence (at n-1) in the no-feedback condition appeared to be very similar to our previous study (average effect = 0.23 ± 0.14; Fornaciai & Park, 2020b). In fact, a quantitative comparison between the current no-feedback results and the results from our previous study showed no difference in their serial dependence effects (independent-sample t-test, $t(48) = 0.19$, $p = 0.85$, $BF = 0.29$), while the comparison between the current feedback results and the results from our previous study showed a significant difference (t(48) = 2.87, p = 0.006, d = 0.80, BF = 7.13).

Finally, as a control for our results, we also assessed the influence provided by the stimulus in the *future* trial (n+1) on the current one, which is not expected to provide any effect. In both the feedback and no-feedback condition, the effect at n+1 was not significantly higher than zero (t(17) = -0.78, p = 0.44, BF = 0.31, and t(17) = 1.31, p = 0.21, BF = 0.50, respectively). Overall, this set of results shows that in the presence of response feedback, the serial dependence effect provided by the immediately preceding stimulus on the current one is strongly amplified.

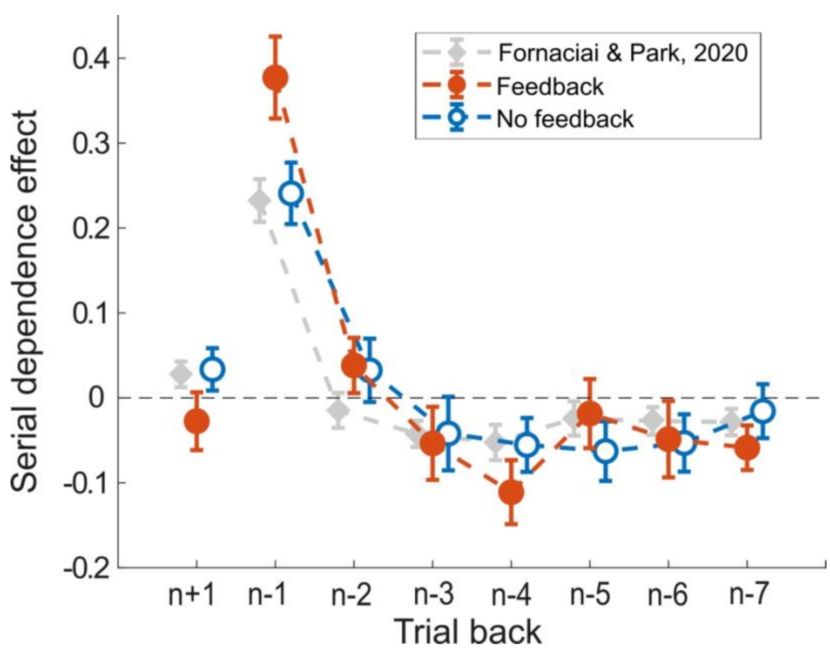

FIGURE 5 – Serial dependence effects in Exp. 2*. Serial dependence effects induced by the immediately preceding stimulus (n-1) and stimuli further back in the past (n-2, n-3, and so on), as well as the future trial (n+1) as a control. The data points in grey (diamonds) show the results of a previously published study (Fornaciai & Park, 2020) using the same methodology but without response feedback. Error bars are SEM.*

These results show that providing feedback to participants generally increases the serial dependence effect induced by the immediately preceding stimulus. However, if response feedback acts by increasing the weight of past information during perceptual judgements, then the type of feedback received in the previous trial might further modulate the strength of the serial dependence effect. Namely, when participants are told that they have responded correctly, then the increased weight of past information should increase the influence that it exerts on the successive stimulus. When a stimulus is instead not judged correctly, the weight of the perceptual information driving such judgment should be lower, leading to a reduced effect. To address this possibility, we thus assessed the serial dependence effect according to the feedback received in the previous trials (i.e., either "too low," "correct," or "too high"). Fig. 6 shows the results of this analysis.

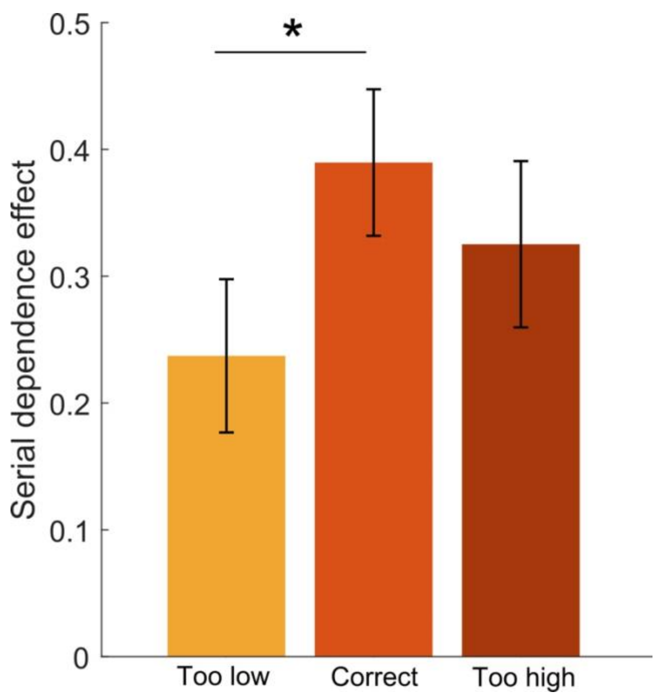

FIGURE 6 – Serial dependence effect as a function of response feedback type*. Serial dependence effects provided by the immediately preceding stimulus (n-1) on the current one, according to the type of feedback received in the previous trial. Error bars are SEM.*

As shown in Fig. 6, we indeed observed differences in the effect according to the type of feedback received by participants. As predicted based on the idea that external information could boost the weight of perceptual information in serial dependence, the strongest effect was observed when the response in the previous trial was marked as correct. The effect in the case of a correct response at n-1 was significantly higher compared to when the response was marked as "too low" (i.e., underestimation of the stimulus numerosity; $t(17) = 2.56$, $p = 0.02$, $d = 0.78$, $BF = 2.96$). On the other hand, no significant difference was observed between correct and "too high" responses (i.e., overestimation; $t(17) = 0.67$, $p = 0.51$, $BF = 0.30$), and between too high and too low responses ($t(17) = 1.84$, $p = 0.08$, $BF = 0.98$). This suggests that the type of feedback provided to participants could have a different impact on the serial dependence effect, depending on whether the response was correct and on the direction of the estimation error in the previous trial.

Finally, although it is only anecdotal, we also asked participants at the end of the experiment about whether the response feedback was actually useful to perform the task. The majority of participants (15 out of 18) reported that the feedback was indeed useful, and the task was easier with feedback. Two participant reported instead that the feedback was not very useful, and one was not sure. This, coupled with the lack of an actual improvement in accuracy or precision in the task, and a difference in response times (see Fig. 3) at the transition between the feedback and no-feedback condition (e.g., see for instance Maldonado Moscoso et al., 2020; Pereira et al., 2020) might further suggest that the response feedback served to boost the participants' subjective confidence in their judgments (see the *Discussion* below), but with only a small impact on the objective measures of performance.

DISCUSSION

In the present study, we addressed the nature of serial dependence in numerosity perception by testing two different predictions based on the hypothesis that this effect originates from high-level visual processing stages. Serial dependence has been shown to be nearly ubiquitous in vision, affecting virtually all aspects of perception (Fischer & Whitney, 2014; Fornaciai & Park, 2018b; Liberman et al., 2014; Suárez-Pinilla et al., 2018; Togoli et al., 2021). The nature of this effect, however, is still unclear, and whether serial dependence actually entails a genuine perceptual bias (e.g., Fornaciai & Park, 2018a; Manassi et al., 2018) or a decisional or mnemonic bias (Bliss et al., 2017; Fritsche et al., 2017; Pascucci et al., 2019) has been hotly debated in recent years. Evidence so far suggests that the serial dependence effect is unlikely to be a "low-level" effect such as adaptation (Kohn, 2007) which in many perceptual domains originates in relatively early visual areas (Boynton & Finney, 2003). Indeed, serial dependence shows the hallmarks of a "high-level" effect, such as a relatively broad spatial selectivity (Collins, 2019; Fischer & Whitney, 2014; Fornaciai & Park, 2018b), dependence on attention (Fischer & Whitney, 2014; Fornaciai & Park, 2018b) and task-relevance (Pascucci et al., 2019; Togoli et al., 2021), and dependence on the conscious perception of the stimuli (Fornaciai & Park, 2019a, 2021).

However, the level at which serial dependence *originates* does not necessarily coincides with the level at which it *operates*. Serial dependence may indeed originate at a high-level in the visual processing hierarchy – explaining several properties of this effect – but still operate at a lower level, thus involving a genuine perceptual bias altering the phenomenological appearance of a stimulus (and not just how we judge or remember it). We recently proposed that serial dependence may operate via feedback (i.e., top-down) signals from high-level to low-level brain areas (Fornaciai & Park, 2019a, 2021). This idea is based on the observation that serial dependence is disrupted by visual backward marking, which is known to make a stimulus invisible by suppressing the feedback processing it needs to reach consciousness (Boehler et al., 2008; Fahrenfort et al., 2007). Importantly, this idea could potentially reconcile mixed findings demonstrating the high-level nature of serial dependence measured behaviorally and the electrophysiological findings showing that its signature emerges very early after the onset of the current stimulus (i.e., suggesting the involvement of early visual areas; Fornaciai & Park, 2019a, 2020a).

Here we thus tested two additional predictions based on the idea that serial dependence originates at a high-level processing stage. Our first prediction, tested in Exp. 1, is that serial dependence in numerosity perception should work according to the magnitude of the stimuli irrespective of how magnitude is conveyed. In a previous study from our group (Fornaciai & Park, 2019b), we have shown that the effect works even across completely different numerosity formats – i.e., the numerosity of a sequence of visual events affects the perceived numerosity of an array of dots. However, if serial dependence truly originates at a stage where numerical magnitude is abstracted from the properties of the stimulus conveying it, it should work even when there is no actual numerosity information at the sensory level, and magnitude is conveyed by the *meaning* of the stimulus. Our results show that this is indeed the case: presenting the number "12" on the screen makes a 16-dot array stimulus to be slightly underestimated, compared to presenting the number "24." This effect appears to be slightly weaker (although not significantly different) compared to the effect induced by a dot-array, but in our results the effect size (Cohen's d) in the two cases was actually very similar (0.87 and 0.92, respectively for the symbolic and non-symbolic inducer). One difference between the two conditions is in the correlation between the strength of the effect and the level of precision in the task, which was significant only in the case of the symbolic inducer. A possibility is thus that symbolic magnitude information may be more flexibly used according to the uncertainty of perceptual judgments, while non-symbolic magnitude may provide a more "automatic" effect, less dependent on uncertainty (see also Fornaciai & Park, 2019b). However, this difference may be more parsimoniously explained by the overall lower and, more importantly, less variable WFs observed in the non-symbolic inducer condition, which may have prevented us from capturing a potential correlation in such a condition. Interestingly, participants were significantly more precise in the presence of a non-symbolic inducer, even if the two task-relevant stimuli were identical across conditions. This might for instance be due to a pre-activation of the approximate number system making the processing of non-symbolic numerosity more efficient when the task-relevant stimuli are preceded by the dot-array inducer.

Considering the similarity of the paradigm used in Exp. 1 compared a typical priming paradigm, is it possible that the effect shown here represents an instance of numerical priming rather than serial dependence? Numerical priming has been indeed shown to similarly generalize across symbolic and non-symbolic stimuli (Bahrami et al., 2010). However, considering the difference between serial dependence and priming, our results are unlikely to be characterized as priming. First, the effect of numerical priming is usually conceptualized as an interference with performance, in terms for instance of speeded up or slowed down reaction times to a primed stimulus (e.g., Koechlin et al., 1999). Serial dependence, instead, concerns a bias in the perception or judgement of a stimulus, provided by the immediately preceding stimulus in the sequence (Fischer & Whitney, 2014), which is not reliably accompanied by changes in other measures of performance (Fornaciai & Park, 2019b). Moreover, while priming can similarly work with conscious or unconscious stimuli (Bahrami et al., 2010), serial dependence is suppressed by masking, rather making a repulsive effect to emerge (Fornaciai & Park, 2019b; Glasser et al., 2011). Finally, priming shows a more semantic nature (e.g., Naccache & Dehaene, 2001), and can transfer from audition to vision (Kouider & Dehaene, 2009), while serial dependence does not (Fornaciai & Park, 2019b). Overall, although the two effects are measured with similar paradigms and may even share some underlying brain process, their widely different nature and properties make it difficult to explain our results as a numerical priming effect.

Our second prediction, tested in Exp. 2, concerns the influence of external information in modulating the strength of serial dependence. Previously, it has been shown (Suarez-Pinilla et al., 2018; Samaha et al., 2019) that the strength of serial dependence is influenced by subjective confidence in one's own perception and perceptual judgments. This finding has been linked to serial dependence being weighted by a subjective estimate of perceptual uncertainty not necessarily reflecting the objective uncertainty of a visual stimulus. In line with the idea of a high-level effect, the serial dependence effect should also be weighted based on external information about the accuracy of one's own perceptual judgements. Our results from Exp. 2 show that in the presence of response feedbacks, the serial dependence effect provided by the immediately preceding stimulus is significantly stronger compared to the effect obtained in the absence of such feedback.

Even more interesting, when the response in the previous trial was marked as correct, the effect was stronger compared to when the response was wrong. This suggests that external information concerning the perceptual judgment of a stimulus was able to boost the attractive bias toward it. The boosting effect of "correct" response feedback seems, however, more pronounced in comparison to when participants underestimated the numerosity in the previous trial, but not when they overestimated it. This further suggests that the direction of the estimation error in the previous trial also modulates serial dependence. We speculate that this difference in the direction may be driven by the fact that what determines the correct discrimination of two stimuli is usually their ratio rather than their absolute difference (e.g., Feigenson et al., 2004; Xu & Spelke, 2000). For example, the numerosity 16 is equally distinguishable from 12 (an absolute difference of 4 items) as it is from 24 (a difference of 8 items). Considering this, an overestimation of 2 or 3 items (for instance) in absolute terms is thus more likely a much less severe error compared to underestimating a stimulus to a similar extent. Nevertheless, the fact that our response feedback did not discriminate between moderate and larger errors leaves this point speculative, and a dedicated experiment using a different feedback procedure is probably needed to more conclusively address this difference.

An interesting point in this context is: what is the underlying factor boosting the serial dependence effect? On the one hand, external information about previous perceptual judgments may simply weigh current perceptual representations in favor of the previous stimulus. In other words, the availability of additional information concerning the accuracy of previous judgments could increase the weight of past information on current percepts. However, while this hypothesis predicts an increased effect when the previous judgment was correct – as we observed in our results – it would also predict a much reduced effect (i.e., less than what we observed in the absence of feedback) when the previous judgment was incorrect. An incorrect response should indeed decrease the weight of past information, as the visual system should "trust" it to a lesser extent. On the other hand, the response feedback may act by boosting the confidence in the perceptual history. In this scenario, confidence could operate at two different levels: first, in a trial-by-trial fashion, boosting the effect when the previous stimulus was perceived and judged correctly (similarly to the previous hypothesis), but also on a more general level, via an overall increased confidence in one's own judgments due to the presence of feedback (i.e., as suggested by the task seeming overall easier in this case). For instance, the discontinuity in the response times – which are often associated with confidence (e.g., Maldonado Moscoso et al., 2020; Pereira et al., 2020; Zylberberg et al., 2016) – observed at the transition between the feedback and no-feedback condition provides some support for a role of confidence in our experiment. This in turn would be in line with previous studies showing the influence of confidence on the serial dependence effect (Suarez-Pinilla et al., 2018; Samaha et al., 2019), and would also suggest that serial dependence is sensitive to confidence driven not only by a subjective evaluation of performance, but also by objective, external information. These hypotheses are however not mutually exclusive, since the effect may be determined by a combination of the specific information received in each trial (i.e., the feedback) and a more general effect of increased confidence driven by the presence of response feedback. However, since we did not directly measure confidence, this interpretation remains speculative, and more evidence is needed to better assess the potential role of confidence in this context.

On a different note, our experimental design had a limitation that warrants further discussion. Indeed, the no-feedback condition was always performed after the feedback blocks. We preferred such a fixed order since we planned to exclude the first four blocks of the experiment, which were considered as training. This, in turn, was aimed at ensuring that participants reached a good level of training and to familiarize them with the response feedback, while keeping the proportion of feedback and no-feedback blocks identical for all participants. If instead we had randomized the order of the two parts, that would have involved having a different number of feedback/no-feedback blocks across different participants, which we preferred to avoid. In any case, the major drawback of having the no-feedback condition always at the end is that the influence of response feedback may "leak" to the no-feedback blocks, reducing the difference and making our results only more conservative. Despite this potential issue, the comparison of the current data with the results obtained in a previous study (Fornaciai & Park, 2020b) shows that the fixed order of conditions is unlikely to have affected the results. Indeed, the magnitude of the serial dependence effect measured in the current no-feedback blocks is virtually indistinguishable from the effect measured in our previous study, where participants performed the task without feedback right from the beginning of the session. On the other hand, the effect obtained with response feedback resulted to be significantly higher also compared to our previous data.

Overall, while our results are neutral when it comes to showing where serial dependence operates, they support the idea that this effect originates at a high level of the visual processing hierarchy. First, our results converge with previous studies in showing that, in the context of numerosity perception, the effect relies on an abstract representation of numerical magnitude (Fornaciai & Park, 2019b). Such a representation is based on the perceived, rather than physical, numerosity of an array (Fornaciai & Park, 2021), and does not discriminate between stimuli with widely different low-level sensory properties – at least within the visual modality (Fornaciai & Park, 2019b). In other perceptual domains such as orientation perception, serial dependence has been however shown to be more sensitive to the features of the stimuli (Fischer et al., 2020), suggesting that the level of abstraction may differ in different perceptual domains. In terms of brain regions where serial dependence may originate from, the posterior parietal cortex represents a good candidate in the context of numerosity perception. Indeed, parietal areas such as the intraparietal sulcus have been shown to encode numerosity in an abstract fashion (see Roitman et al., 2012 for a review), to similarly encode non-symbolic and symbolic stimuli (Piazza et al., 2004, 2007), and to be sensitive to attentional modulation (Castaldi et al., 2019). Second, our results show that serial dependence can be modulated by external information. Providing a response feedback significantly increased the serial dependence effect, especially when the feedback in the previous trial confirmed that the response was correct. This suggests that serial dependence originates at a processing level sufficiently high to allow external information unrelated to the stimulus itself to play a role in determining the effect. Subjective confidence has been indeed already shown to modulate serial dependence, and a possibility (although speculative) is that the response feedback could have affected the confidence in the participants' responses.

This in turn could have boosted the reliability of perceptual information, increasing the weight of past sensory information during the processing of the current stimulus.

To conclude, our results provide novel evidence suggesting that serial dependence originates at a high level in the visual processing hierarchy, entailing an abstract representation of the stimuli and being sensitive to external information modulating confidence. These findings, combined with previous results showing that serial dependence involves a genuine perceptual bias (e.g., Manassi et al., 2018) and activity in early visual areas (e.g., Fornaciai & Park, 2018a), support the idea that this effect operates by propagating from high-level to low-level brain areas. Our results are also in line with the idea that serial dependence may reflect a perceptual mechanism facilitating the stability and continuity of visual processing over time (Fischer & Whitney, 2014), but further highlight the flexibility of such a mechanism in exploiting all the available information, not limited to the purely sensory aspects of the external stimuli.

Acknowledgments. This study was supported by the National Science Foundation CAREER Award BCS1654089 to J. P.

Declarations

Funding. This study was supported by the National Science Foundation CAREER Award BCS1654089 to J. P.

Conflicts of interest/Competing interests. The authors declare no competing interest.

Ethics approval. All the experimental procedures were approved by the internal review board of the University of Massachusetts Amherst.

Consent to participate. Participants signed a written informed consent form prior to participating in the study.

Consent for publication. Not applicable (data is fully anonymized).

Availability of data and materials. All the data generated during the experiments described in this manuscript have been uploaded to Open Science Framework, and can be accessed following this link: https://osf.io/veysa/.

Code availability. Experimental code will be made available upon request to the corresponding author.

Authors' contributions. M.F. and J.P. devised the study. M.F. collected the data. M.F. and J.P. analyzed the data, interpreted the results, wrote and revised the manuscript.

Open Practices Statement

All the data generated during the experiments described in this manuscript have been uploaded to Open Science Framework, and can be accessed following this link: https://osf.io/veysa/. The study was not preregistered.

REFERENCES

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

Bahrami, B., Vetter, P., Spolaore, E., Pagano, S., Butterworth, B., & Rees, G. (2010). Unconscious Numerical Priming Despite Interocular Suppression. *Psychological Science*, *21*(2), 224–233. https://doi.org/10.1177/0956797609360664

Bliss, D. P., Sun, J. J., & D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific Reports*, *7*(1), 14739. https://doi.org/10.1038/s41598-017-15199-7

Boehler, C. N., Schoenfeld, M. A., Heinze, H.-J., & Hopf, J.-M. (2008). Rapid recurrent processing gates awareness in primary visual cortex. *Proceedings of the National Academy of Sciences*, *105*(25), 8742–8747. https://doi.org/10.1073/pnas.0801999105

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

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

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

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*. https://doi.org/10.7554/eLife.45160

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

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

Collins, T. (2019). The perceptual continuity field is retinotopic. *Scientific Reports*, *9*(1), 18841. https://doi.org/10.1038/s41598-019-55134-6

Corbett, J. E., Fischer, J., & Whitney, D. (2011). Facilitating Stable Representations: Serial Dependence in Vision. *PLoS ONE*, *6*(1), e16701. https://doi.org/10.1371/journal.pone.0016701

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

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking Disrupts Reentrant Processing in Human Visual Cortex. *Journal of Cognitive Neuroscience*, *19*(9), 1488–1497. https://doi.org/10.1162/jocn.2007.19.9.1488

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

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., Cicchini, G. M., & Burr, D. C. (2016). Adaptation to number operates on perceived rather than physical numerosity. *Cognition*, *151*. https://doi.org/10.1016/j.cognition.2016.03.006

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

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

Fornaciai, M., & Park, J. (2020b). Attractive serial dependence between memorized stimuli. *Cognition*, *200*, 104250. https://doi.org/10.1016/j.cognition.2020.104250

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

Fritsche, M., & de Lange, F. P. (2019). The role of feature-based attention in visual serial dependence. *Journal of Vision*, *19*(13), 21. https://doi.org/10.1167/19.13.21

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

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*, *108*(45). https://doi.org/10.1073/pnas.1101141108

Kaliukhovich, D. A., & Vogels, R. (2016). Divisive Normalization Predicts Adaptation-Induced Response Changes in Macaque Inferior Temporal Cortex. *The Journal of Neuroscience*, *36*(22), 6116–6128. https://doi.org/10.1523/JNEUROSCI.2011-15.2016

Kleiner M, Brainard D, & Pelli D. (2007). What's new in Psychtoolbox-3? *Perception 36 ECVP Abstract Supplement*.

Koechlin, E., Naccache, L., Block, E., & Dehaene, S. (1999). Primed numbers: Exploring the modularity of numerical representations with masked and unmasked semantic priming. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1882–1905. https://doi.org/10.1037/0096-1523.25.6.1882

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

Kouider, S., & Dehaene, S. (2009). Subliminal Number Priming Within and Across the Visual and Auditory Modalities. *Experimental Psychology*, *56*(6), 418–433. https://doi.org/10.1027/1618-3169.56.6.418

Liberman, A., Fischer, J., & Whitney, D. (2014). Serial Dependence in the Perception of Faces. *Current Biology*, *24*(21), 2569–2574. https://doi.org/10.1016/j.cub.2014.09.025

Liberman, A., Zhang, K., & Whitney, D. (2016). Serial dependence promotes object stability during occlusion. *Journal of Vision*, *16*(15), 16. https://doi.org/10.1167/16.15.16

Maldonado Moscoso, P. A., Cicchini, G. M., Arrighi, R., & Burr, D. C. (2020). Adaptation to hand-tapping affects sensory processing of numerosity directly: evidence from reaction times and confidence. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1927), 20200801. https://doi.org/10.1098/rspb.2020.0801

Manassi, M., Liberman, A., Chaney, W., & Whitney, D. (2017). The perceived stability of scenes: serial dependence in ensemble representations. *Scientific Reports*, *7*(1), 1971. https://doi.org/10.1038/s41598-017-02201-5

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

Manassi, M., & Whitney, D. (2022). Illusion of visual stability through active perceptual serial dependence. *Science Advances*, *8*(2). https://doi.org/10.1126/sciadv.abk2480

Naccache, L., & Dehaene, S. (2001). Unconscious semantic priming extends to novel unseen stimuli. *Cognition*, *80*(3), 215–229. https://doi.org/10.1016/S0010-0277(00)00139-6

Park, J., DeWind, N. K., Woldorff, M. G., & Brannon, E. M. (2016). Rapid and Direct Encoding of Numerosity in the Visual Stream. *Cerebral Cortex*, bhv017. https://doi.org/10.1093/cercor/bhv017

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

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

Pereira, M., Faivre, N., Iturrate, I., Wirthlin, M., Serafini, L., Martin, S., Desvachez, A., Blanke, O., van de Ville, D., & Millán, J. del R. (2020). Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. *Proceedings of the National Academy of Sciences*, *117*(15), 8382–8390. https://doi.org/10.1073/pnas.1918335117

Piazza, M., Izard, V., Pinel, P., le Bihan, D., & Dehaene, S. (2004). Tuning Curves for Approximate Numerosity in the Human Intraparietal Sulcus. *Neuron*, *44*(3), 547–555. https://doi.org/10.1016/j.neuron.2004.10.014

Piazza, M., Pinel, P., le Bihan, D., & Dehaene, S. (2007). A Magnitude Code Common to Numerosities and Number Symbols in Human Intraparietal Cortex. *Neuron*, *53*(2), 293–305. https://doi.org/10.1016/j.neuron.2006.11.022

Roitman, J. D., Brannon, E. M., & Platt, M. L. (2012). Representation of numerosity in posterior parietal cortex. *Frontiers in Integrative Neuroscience*, *6*. https://doi.org/10.3389/fnint.2012.00025

Samaha, J., Switzky, M., & Postle, B. R. (2019). Confidence boosts serial dependence in orientation estimation. *Journal of Vision*, *19*(4), 25. https://doi.org/10.1167/19.4.25

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

Togoli, I., Fedele, M., Fornaciai, M., & Bueti, D. (2021). Serial dependence in time and numerosity perception is dimension-specific. *Journal of Vision*, *21*(5), 6. https://doi.org/10.1167/jov.21.5.6

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

Wehrman, J. J., Wearden, J., & Sowman, P. (2020). Decisional carryover effects in interval timing: Evidence of a generalized response bias. *Attention, Perception, & Psychophysics*, *82*(4), 2147–2164. https://doi.org/10.3758/s13414-019-01922-1

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

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11. https://doi.org/10.1016/S0010-0277(99)00066-9

Zylberberg, A., Fetsch, C. R., & Shadlen, M. N. (2016). The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *ELife*, *5*. https://doi.org/10.7554/eLife.17688