Serial dependence in the perception of visual variance

Marta Suárez-Pinilla
Sackler Center for Consciousness Science and Department of Informatics, University of Sussex, Brighton, UK

Anil K. Seth
Sackler Center for Consciousness Science and Department of Informatics, University of Sussex, Brighton, UK

Warrick Roseboom
Sackler Center for Consciousness Science and Department of Informatics, University of Sussex, Brighton, UK

The recent history of perceptual experience has been shown to influence subsequent perception. Classically, this dependence on perceptual history has been examined in sensory-adaptation paradigms, wherein prolonged exposure to a particular stimulus (e.g., a vertically oriented grating) produces changes in perception of subsequently presented stimuli (e.g., the tilt aftereffect). More recently, several studies have investigated the influence of shorter perceptual exposure with effects, referred to as *serial dependence*, being described for a variety of low- and high-level perceptual dimensions. In this study, we examined serial dependence in the processing of dispersion statistics, namely variance—a key descriptor of the environment and indicative of the precision and reliability of ensemble representations. We found two opposite serial dependences operating at different timescales, and likely originating at different processing levels: A positive, Bayesian-like bias was driven by the most recent exposures, dependent on feature-specific decision making and appearing only when high confidence was placed in that decision; and a longer lasting negative bias—akin to an adaptation aftereffect—becoming manifest as the positive bias declined. Both effects were independent of spatial presentation location and the similarity of other close traits, such as mean direction of the visual variance stimulus. These findings suggest that visual variance processing occurs in high-level areas but is also subject to a combination of multilevel mechanisms balancing perceptual stability and sensitivity, as with many different perceptual dimensions.

Introduction

Considerable evidence indicates that the human visual system is able to extract statistical information from sensory signals supporting the formation of summary or ensemble representations across a variety of dimensions, including low-level features such as orientation or size, as well as higher level (complex or abstract) traits such as facial expressions in a crowd (Alvarez, 2011; Alvarez & Oliva, 2009; Ariely, 2001; Chong & Treisman, 2003; Geisler, 2008; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012). Such information can be used to efficiently encode (Dahmen, Keating, Nodal, Schulz, & King, 2010; Fairhall, Lewen, Blalek, & de Ruyter van Steveninck, 2001) and interpret subsequent sensory inputs, and to make predictions about future events (Summerfield & de Lange, 2014; Summerfield & Egner, 2009).

Many forms of visual input can be summarized in terms of statistical moments such as central tendency (e.g., mean) and variance or dispersion (consider, for example, a random-dot kinematogram, which will have a mean and a variance in the distribution of dot motion). Most studies on ensemble processing have focused on central-tendency statistics (Albrecht & Scholl, 2010; Chong & Treisman, 2005; Corbett & Oriet, 2011; Haberman & Whitney, 2009; Im & Chong, 2014; Sweeny & Whitney, 2014; Wolfe, Kosovicheva, Leib, Wood, & Whitney, 2015), while variance computations have received less attention. However, variance is known to play a crucial role in visual experience, modulating perceptual grouping, ensemble averaging (Brady & Alvarez, 2015; de Gardelle & Mammasian, 2015; de Gardelle & Summerfield, 2011;

Citation: Suárez-Pinilla, M., Seth, A. K., & Roseboom, W. (2018). Serial dependence in the perception of visual variance. *Journal of Vision*, *18*(7):4, 1–24, https://doi.org/10.1167/18.7.4.

https://doi.org/10.1167/18.7.4 Received January 17, 2018; published July 2, 2018 ISSN 1534-7362

Maule & Franklin, 2015; Maule, Witzel, & Franklin, 2014; Zylberberg, Roelfsema, & Sigman, 2014), texture processing (Morgan, Chubb, & Solomon, 2014; Morgan, Mareschal, Chubb, & Solomon, 2012), and comparison between arrays (Fouriezos, Rubenfeld, & Capstick, 2008). Variance is also critical to perceptual prediction, since it provides a measure of the expected range of stimuli (Summerfield & de Lange, 2014) as well as the precision (reliability) of the sensory input (Corbett, Wurnitsch, Schwartz, & Whitney, 2012; Meyniel, Sigman, & Mainen, 2015; Sato & Kording, 2014). As an indication of sensory reliability, variance also affects metacognitive judgments, although evidence is conflicting regarding the extent and direction of this effect (de Gardelle & Mammasian, 2015; Spence, Dux, & Arnold, 2016; Zylberberg et al., 2014).

Notably, most studies involving variance manipulations have examined its impact on perceptual decisions about other features rather than on the perception of variance itself. Comparatively few studies have investigated the mechanisms of variance processing directly. Those that do have addressed mainly three questions: What are the general properties of variance processing (speed, automaticity, attentional demands)? To what extent does variance computation rely on the processing of the individual elements of the ensemble? And does it operate as a specific trait of the ensemble or feature dimension over which it is computed, or rather as an abstract property? So far, these studies have employed heterogeneous designs and reached disparate conclusions. Concerning the general properties of variance processing, a study examining judgments of color diversity (Bronfman, Brezis, Jacobson, & Usher, 2014) suggested a rapid, preattentive mechanism. This is in agreement with another study which reported priming by visual variance, an effect that seems to occur rapidly, automatically, and without need of feature-based attention (Michael, de Gardelle, & Summerfield, 2014); however, this latter study did not examine variance perception directly, but only the priming effect of variance on mean judgments.

Regarding the second question, namely the reliance of variance computation on the processing of individual elements, available evidence (based on highly heterogeneous studies) is conflicting: One study on pattern regularity (positional variance) suggested a very inefficient computation of variance, underwritten by subsampling of a small fraction of the elements of the array (Morgan et al., 2012). By contrast—and surprisingly, given the finding of a rapid, preattentive mechanism—the aforementioned study on color diversity reported that variance processing required a conscious representation of the individual elements (Bronfman et al., 2014). In a similar manner, a study on facial emotions in a crowd determined that variance judgments along this dimension relied on high-level processing of individual faces (Haberman, Lee, & Whitney, 2015).

Finally, regarding the question whether variance, once computed over a certain feature dimension of a visual ensemble, retains its specificity or emerges as an abstract property, the previously reported study on priming suggested that the effect of (implicit) variance on mean perception was feature specific (Michael et al., 2014). In contrast, a study on direct variance perception found negative adaptation aftereffects which generalized across dimensions of visual variance, suggesting a high-level rather than a sensory origin for this effect, and therefore indicating that variance may operate as an independent cognitive property (Payzan-LeNestour, Balleine, Berrada, & Pearson, 2016). In summary, available evidence shows some dissension, but a picture starts to emerge: Variance computations would be relatively rapid, but appear to require high-level processing of the individual elements of the array; however, once computed, variance would become dissociated from the properties of the ensemble and of the perceptual dimension over which it was estimated, and operate as a high-level cognitive trait.

To clarify some these issues, here we examine variance processing as a distinct perceptual feature by investigating the influence of previous variance presentations on judgments about this dimension. It has long been known that perception is affected by previous input. Influences of past perceptual events on current perception fall generally into two different categories. Best known are adaptation aftereffects—repulsive (negative) biases in perception exerted after prolonged exposure to a certain stimulus magnitude—which have been described for many low- and high-level traits, including variance (Campbell & Maffei, 1971; Mather, Verstraten, & Anstis, 1998; Payzan-LeNestour et al., 2016; Roseboom, Linares, & Nishida, 2015), and which are classically employed as an experimental tool for investigating perceptual mechanisms (Kohn, 2007). The second category is observed in relation to shorter presentations, generally consisting of an attractive (positive) perceptual bias toward recent sensory history. These *serial dependences* have been found for several low- and high-level features (Cicchini, Anobile, & Burr, 2014; Fischer & Whitney, 2014; John-Saaltink, Kok, Lau, & de Lange, 2016; Liberman, Fischer, & Whitney, 2014; Xia, Liberman, Leib, & Whitney, 2015). It has been proposed that these two different effects contribute in opposite ways to the tuning of the balance between perceptual sensitivity and stability: While negative adaptation produces a normalization of neural representations in order to maximize sensitivity to changes around the most frequent stimulus intensity, serial dependence contributes to perceptual stability by smoothing out discrete

discontinuities as sensory noise (Fischer & Whitney, 2014).

Our study employs serial dependence in variance judgments as a way to track the dynamics and timescales of the processing of this statistic as a distinct perceptual feature. We conducted three experiments.

Experiment 1 investigated the existence, magnitude, and direction of serial dependence in visual variance perception, as well as its relationship with associated stimulus features such as ensemble mean and spatial location. In addition, we separately explored fovea and periphery, as the compression of visual information into summary statistics is particularly relevant to the much larger, poorly spatially resolved peripheral field (Balas, Nakano, & Rosenholtz, 2009; Freeman & Simoncelli, 2011; Rosenholtz et al., 2012; Ziemba & Simoncelli, 2015). Experiment 2 attempted to identify the specific levels of perceptual decision that give rise to serial dependence in variance: whether it is a bias in low-level perceptual, decision-making, or response processes. In Experiment 3 we investigated the relationship between the reported confidence in perceptual decisions and their influence in subsequent judgments along the same dimension, especially considering Bayesian accounts of confidence as a measure of the precision of neural representations (Meyniel et al., 2015).

Overall, our results indicate that judgments of visual variance are subject to serial-dependence effects as seen for many other sensory dimensions. These effects are independent of basic stimulus features such as spatial location, but do depend on whether the previous judgment made was regarding the same (visual variance) or a different (visual direction) dimension, and on the level of confidence expressed in previous judgments. Together, these results suggest that visual variance is processed as a more abstract feature of perception, though is subject to the same processes of efficient coding and perceptual stability found for many other perceptual dimensions.

Experiment 1: Serial dependences in variance judgments

Figure 1 outlines the experimental paradigm utilized across all three experiments. We employed random-dot kinematograms (RDKs), which allow independent manipulation of mean and variance, and asked participants to score the randomness of the motion of RDKs using a visual analogue scale. Several variations of this basic paradigm were used to characterize the effects of previous history on variance judgments.

In Experiment 1, we investigated the existence of serial dependence in variance judgments and its relationship to basic features of stimulus presentation, including eccentricity, spatial location, and mean RDK motion direction. Thus, Experiment 1 employed the basic task (variance estimation without further requirements), with separate blocks in which the RDK was displayed in fovea and periphery.

Methods

Stimuli

The stimulus consisted of a cluster of random moving dots (RDK) displayed for 500 ms at a certain eccentricity (0° or 20°; see later) over a dark-gray background (3.92 cd/m^2). The cluster spanned 5° of visual angle (°va) along the horizontal and vertical dimensions and comprised 100 light-gray dots (diameter = 0.11°va, luminance = 43.14 cd/m^2) moving along a straight trajectory at a rate of 2 pixels/frame (8.45°va/s). The initial position of each dot was uniformly randomized (excluding overlap with other dots), and its coordinates were updated per frame by a trigonometric calculation based on the individual dot's angular-motion direction, re-entering the cluster from the opposite side if it reached a boundary. Each dot's motion direction was extracted from a circular Gaussian (von Mises) distribution that varied for each stimulus presentation: Its mean could take any random integer value from 0° to 359°, and its standard deviation was pseudorandomized among six possible values—namely 5°, 10°, 20°, 30°, 40°, and 60°. This parameter, standard deviation of RDK motion (StD), is the dimension of interest in this experiment.

Procedure

The experimental session comprised a practice block with 72 trials and eight experimental blocks with 60 trials each. The practice had a double purpose: familiarizing participants with the scoring process and the scale used in the experiment, and training maintenance of centered gaze fixation. A broader range of StD values was presented during the practice block compared to the experimental blocks (1°, 10°, 20°, 36°, 60°, and 90°). In both practice and experimental blocks, participants had to score the randomness (variance) of the motion of the RDK using a visual analogue scale (see Figure 1) by adjusting the position of a sliding bar with the mouse. During the practice block, feedback was provided after each response by showing the correct response (score corresponding to the veridical variance) on an additional scale which appeared below the one employed by the participant, after the participant's response. For simplicity, the scale was a linear translation of the StD numeric values ranging from 0° (left end) to 90° (right end). Within this block,

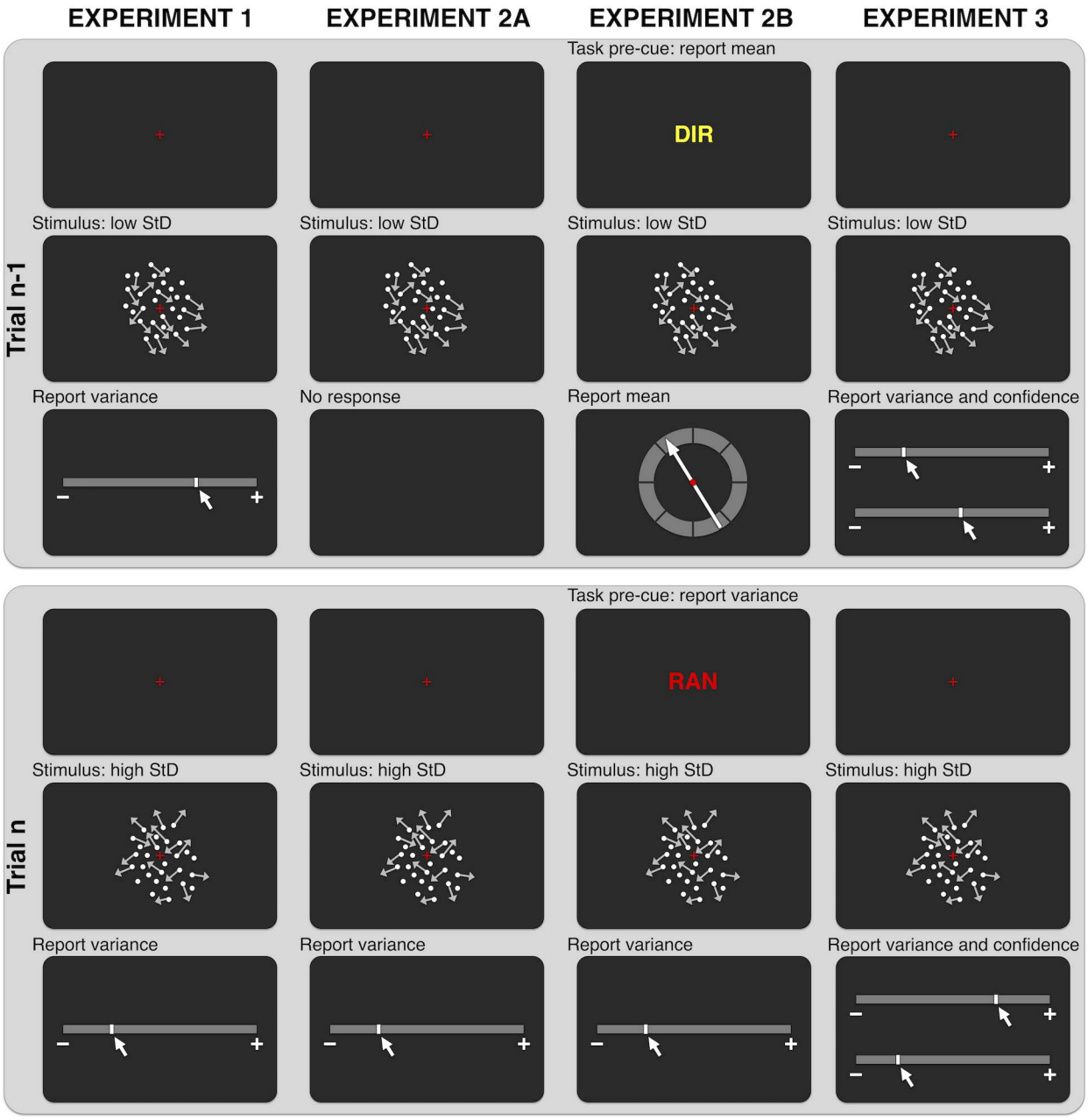

Figure 1. Experiments 1–3: Structure. In all experiments, each trial presented a random-dot kinematogram of a certain mean and variance (standard deviation, StD) in the motion trajectories of its component dots. In the example, Trials $n-1$ and n have low and high StD values, respectively. Experiment 1 required variance (StD) reports for each trial, using a visual analogue scale. Experiment 2 interleaved two thirds of trials in which variance reports were required and one third in which either no response (Experiment 2A) or mean trajectory estimation (Experiment 2B) was required. In Experiment 2B the trial type was precued, so that the label "DIR" or "RAN" at the beginning of each trial indicated whether a mean or variance judgment was required for that trial. Experiment 3 required subjective confidence ratings following variance reports, using a similar visual analogue scale.

the first 36 trials were foveal (the stimulus was presented at 0°va eccentricity) and the remaining 36 were peripheral (20°va).

The eight experimental blocks employed the narrower set of StD values detailed under Stimuli (5°, 10°, 20°, 30°, 40°, and 60°) and did not have feedback. Half of the eight blocks were foveal (stimulus presentation at 0°va eccentricity for all trials) and half peripheral (presentation at 20°va along the horizontal axis), equally frequent in the right and left hemifields. The sequence of foveal and peripheral blocks was pseudorandomized for each participant.

Eye tracking was performed during the entire experimental session. Calibration of the eye-tracking

system was performed at the beginning of each block (practice and experimental) using a standard five-point grid, allowing for a maximal average error of 0.5°va.

At the beginning of each trial a red fixation cross appeared at the center of the screen, spanning 1.1°va (horizontally and vertically). Participants were instructed to maintain their gaze on the fixation cross. The RDK stimulus appeared after 1,000 ms, and both the stimulus and the fixation cross disappeared simultaneously at 1,500 ms from trial onset. Immediately after, the response scale and sliding bar were displayed on the screen. The initial position of the bar was randomized for each trial along the whole length of the scale, to exclude the possibility that participants simply repeated the same (response) action on each trial. If the participant failed to respond within 5 s, the next trial started automatically. The intertrial interval was randomized between 250 and 1,000 ms.

On each trial, participants were allowed to correct their gaze position during the first 700 ms if they noticed that their gaze had deviated from the central fixation cross. However, if a deviation (of more than 5°va) occurred between 700 and 1,000 ms, the trial was aborted and restarted. About a third of participants (9/30) were tested with a slightly different procedure, in which a trial abortion led directly to the start of the next trial (after the intertrial interval). This procedure led to the exclusion of more trials from analysis, since poorly fixated trials were not restarted. Importantly, in both cases trials retained for analysis were those in which fixation was maintained during stimulus presentation (1,000–1,500 ms), and no trial was aborted or restarted after stimulus onset at 1,000 ms.

Participants

Participants were recruited through online advertisement and among members of the laboratory. All were over 18 years old and reported normal or corrected-to-normal vision. Every participant signed an informed consent form before taking part and was either awarded 10 course credits or paid £10 for participation. The study was granted ethical approval by the Research Ethics Committee of the University of Sussex.

Apparatus

Experiments were programmed in MATLAB 2012b (MathWorks, Natick, MA) with Psychtoolbox 3.0.10 and displayed on a LaCie Electron 22BLUE II 22-in. monitor with screen resolution of 1,024 × 768 pixels and refresh rate of 60 Hz. Eye tracking was performed with the EyeLink 1000 Plus (SR Research, Mississauga, Ontario, Canada) at a sampling rate of 1,000 Hz, using a level desktop camera mount. Head position was stabilized at 43 cm from the screen with a chin and forehead rest.

Statistical analysis

Statistical analyses (detailed in the Results section) were performed in MATLAB 2016a, R 3.4.2 (http://www.r-project.org), and JASP (Version 0.8.3.1).

Results

Thirty participants (25 female, five male; mean age = 19.0 years, standard deviation = 1.35) participated in this experiment. Except for two members of the laboratory, they were first-year psychology students who volunteered for course credits.

To ensure the validity of foveal and peripheral conditions, trials without centered gaze fixation during stimulus presentation were removed from the analysis: A trial was deemed valid if the participant maintained fixation within 5°va of the center of the screen for over 80% of the stimulus-presentation period (1,000–1,500 ms from trial onset). Invalid trials were removed, as well as all data regarding trial history that involved at least one of these trials; for instance, if Trial n was valid but Trial $n-3$ was not, Trial n was not included in analyses regarding serial dependence associated to position $n-3$ or further backwards. A total of 12,480 trials entered the analysis.

Overview of responses

Figure 2A shows the distribution of responses (R_n) for each StD value and visual eccentricity. Showing that participants were able to perceive the different levels of variance presented in the experiment, reports were positively correlated and monotonically increased with stimulus StD for both foveal and peripheral presentations.

To examine the general pattern of variance judgments, we conducted a repeated-measures analysis of variance (ANOVA) on the influence of two within-subject factors—StD in the current trial (StD_n) and eccentricity—on participant's responses. Both main effects and their interaction were significant (sphericity correction was applied by the Greenhouse–Geisser method). For StD_n, the main effect yielded $F(1.825, 45.621) = 473.80$, $p < 0.001$, $\eta_p^2 = 0.950$, in relation to higher reports for larger stimulus StD. For eccentricity the main effect was $F(1,25) = 33.32$, $p < 0.001$, $\eta_p^2 = 0.571$: Peripheral presentation was associated with lower variance reports, with a mean difference of 6.798 (fovea minus periphery), $t(25) = 8.237$, $p < 0.001$, Cohen's $d = 1.615$. The $StD_n \times$ Eccentricity interaction was also significant, $F(2.715, 67.882) = 20.06$, $p <$

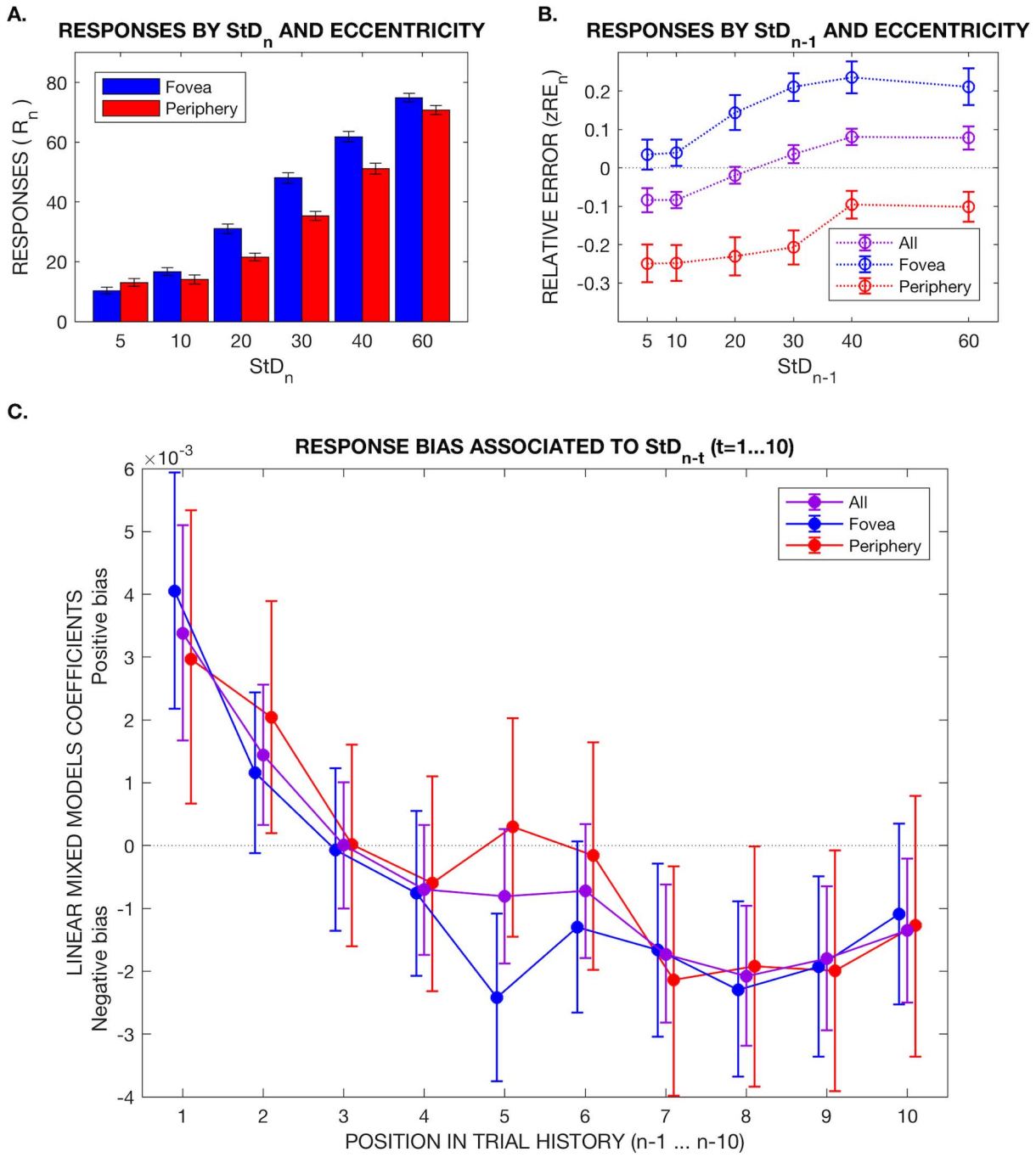

Figure 2. Experiment 1. (A) Distribution of responses by StD_n and eccentricity. The height of the bars represents the mean, and the error bars the between-subjects standard error. (B) Normalized relative error in current response (zRE_n) as a function of the StD presented in the previous trial (StD_{n-1}). The relative error, defined as $RE_n = (R_n - StD_n)/StD_n$, has been normalized by the distribution of errors provided by each subject for the current StD_n; thus, a positive zRE_n means a larger report in that trial than the participant's average for that stimulus level, and conversely a negative zRE_n indicates a lower-than-average score—that is, sign is not necessarily related to comparison with veridical StD_n, if the participant exhibits a systematic bias for that StD_n. Consequently, plotting zRE_n reports by StD_{n-1} allows examination of any possible bias in relation to previous-trial StD_{n-1}, beyond any unrelated source of bias. The error bars represent the between-subjects standard error. The ascending slope of the plots indicates a positive bias associated with StD_{n-1}, for both foveal and peripheral presentations; relative overestimation occurs for larger StD_{n-1}. (C) Response bias associated with StD presented in recent history. Each data point represents the fixed-effects coefficient estimate (B) in a Bayesian linear mixed-effects model for the association between the StD presented in Trials $n - 1$ to $n - 10$ (StD_{n-t}, $t = 1, \ldots, 10$) and the normalized response error in the current trial. The value of the B coefficient represents the linear slope between the past StD at a certain trial position (StD_{n-t}) and the normalized response error provided in the current trial—that is, the variation (in z scores) observed on the current response (regardless of the presented StD), when StD_{n-t} was increased by 1°. A positive B represents an attractive bias (ascending slope), and a negative B a repulsive bias (descending slope). The error bars depict the 95% credible intervals for the value of the B coefficient.

0.001, $\eta_p^2 = 0.445$, indicating that the difference between foveal and peripheral responses increased for large StD_n values, as shown in Figure 2A. These results were confirmed in a Bayesian repeated-measures ANOVA with the same variables: The full model (both main effects and interaction) was the most explanatory according to the Bayes factor, outperforming the second best (only the two main effects) by a factor of $BF_{full/main} = 1.075 \times 10^6$. These findings (lower responses in periphery than in fovea, especially for large StD_n) seem to relate to a greater regression to the mean exhibited in responses about peripheral stimuli (likely due to worse discrimination between stimulus levels), combined with the fact that the range of the response scale allows for larger errors by overestimation than underestimation.

To further characterize perception of variance throughout the different presented StD levels and confirm the apparently worse performance in the periphery, we examined the dispersion of the responses per StD (σ_R), defined as the standard deviation of the distribution of responses per stimulus level. We conducted a repeated-measures ANOVA on the influence of StD_n level and eccentricity on response dispersion. The main effect for StD_n yielded $F(2.994, 74.840) = 58.426$, $p < 0.001$, $\eta_p^2 = 0.700$, in relation to greater response dispersion for large StD levels, as is common in magnitude-estimation tasks: The value of σ_R was lowest at 9.87 for StD = 5 and steadily increased with StD value until it peaked at 21.03 for StD = 30, remained almost equal ($\sigma_R = 20.51$) for StD = 40, and decreased moderately for StD = 60 ($\sigma_R = 15.98$), probably due to a ceiling effect. As for the main effect of eccentricity on response dispersion, it was $F(1, 25) = 4.165$, $p = 0.052$, $\eta_p^2 = 0.143$, suggesting a trend toward greater response dispersion in peripheral presentations: mean difference = −0.658 (fovea minus periphery), $t(25) = -1.738$, $p = 0.086$, Cohen's $d = -0.339$. Last, the effect of the $StD_n \times$ Eccentricity interaction was $F(3.530, 88.244) = 4.757$, $p = 0.002$, $\eta_p^2 = 0.160$, due to the larger response dispersion in periphery occurring mainly for large StD values. In a Bayesian repeated-measures ANOVA with the same variables as in the frequentist counterpart, the best model was the full model (StD_n, eccentricity, and interaction), which outperformed the second best (with only StD_n) by a factor of $BF_{full/StDn} = 8.747$. In summary, response dispersion increased with stimulus (StD) level and there was a (nearly significant) trend toward greater response dispersion for peripheral presentations, especially at large StDs, suggesting a slightly worse performance at 20°va eccentricity compared to 0°va, in agreement with the previous finding of a greater regression to the mean in peripheral responses.

Variance reports are subject to a positive bias driven by very recent trial history

To characterize the existence of serial dependences in variance reports, we tracked whether the response errors provided by each participant for each StD level were different as a function of the StD level presented in the previous trial (serial dependence in relation to Trial $n-1$) or at positions further backward in trial history (Trial $n-t$). Thus, the response variable in our analyses of serial dependence, unless stated otherwise, is the normalized response error relative to the current stimulus (zRE_n). Response errors, defined as $RE_n = (R_n - StD_n)/StD_n$, are normalized by the distribution of reports provided by each individual for the level of StD presented in the current trial. Thus, zRE_n sums to zero across all trials for a given participant and StD_n level: A negative zRE_n indicates that the participant provided a below-average response in that trial compared to their responses for other physically identical stimuli, while a positive zRE_n indicates an above-average response. Therefore, normalization ensures that the value of the response variable zRE_n is independent of the current StD_n level and of each participant's global scoring biases.

Figure 2B presents the average zRE_n as a function of the previous stimulus (StD_{n-1}), plotted separately by eccentricity. Regardless of generally lower reports at larger eccentricity, a trend toward larger zRE_n for higher StD_{n-1} values is evident for all trials pooled as well as for both foveal and peripheral presentations, as shown by the ascending slope of the three plots (Fovea, Periphery, All). In other words, there was a relative overestimation of the current stimulus when the previous stimulus had a large StD, and a relative underestimation when the previous StD was small, compared to other trials in identical conditions of eccentricity. This indicates a positive (attractive, Bayesian-like) bias driven by Trial $n-1$: Current responses resemble the previous stimulus—serial dependence for visual variance.

To verify this observation, we ran a repeated-measures ANOVA on the effect of StD_{n-1} level (as a within-subject factor) on current variance reports (zRE_n). The effect of StD_{n-1} was statistically significant, $F(3.231, 93.697) = 7.221$, $p < 0.001$, $\eta_p^2 = 0.199$, Greenhouse–Geisser correction applied. The Bayes factor for the inclusion of StD_{n-1} compared to the null model (both of them included participant as grouping variable) was $BF_{inclusion} = 56{,}187.91$, indicating *extreme* (Wagenmakers et al., 2017) evidence for superior explanatory ability of the model that included this term.

Serial dependence in variance does not depend on other stimulus properties (visual eccentricity, spatial location, or ensemble mean): Having established the existence of a positive serial dependence exerted by the previous trial, we sought to ascertain which properties of the

Model	$P(M)$	$P(M\|data)$	BF_M	BF_{10}	Error %
(a) Eccentricity					
Null model	0.200	2.747×10^{-30}	1.099×10^{-29}	1.000	
StD_{n-1}	0.200	1.229×10^{-29}	4.918×10^{-29}	4.476	0.626
Eccentricity	0.200	0.004	0.015	1.385×10^{27}	1.303
StD_{n-1} + Eccentricity	0.200	0.943	65.900	3.432×10^{29}	3.951
StD_{n-1} + Eccentricity + StD_{n-1} × Eccentricity	0.200	0.053	0.226	1.945×10^{28}	2.392
(b) Spatial location (hemifield)					
Null model	0.200	0.232	1.208	1.000	
StD_{n-1}	0.200	0.481	3.702	2.073	0.464
Hemifield	0.200	0.079	0.343	0.340	1.690
StD_{n-1} + Hemifield	0.200	0.181	0.883	0.780	3.545
StD_{n-1} + Hemifield + StD_{n-1} × Hemifield	0.200	0.028	0.114	0.120	1.387
(c) Mean difference ($n - 1$, n)					
Null model	0.200	3.111×10^{-6}	1.244×10^{-5}	1.000	
StD_{n-1}	0.200	0.998	2,550.135	320,978.693	0.570
Mean difference	0.200	4.784×10^{-9}	1.914×10^{-8}	0.002	0.726
StD_{n-1} + Mean difference	0.200	0.002	0.006	502.187	0.815
StD_{n-1} + Mean difference + StD_{n-1} × Mean difference	0.200	8.729×10^{-7}	3.491×10^{-6}	0.281	1.091

Table 1. Serial dependence and stimulus properties—model comparison: Experiment 1, serial dependence (associated with Trial $n - 1$) and stimulus properties. Each panel presents the results of a Bayesian repeated-measures ANOVA on zRE_n, with two within-subject factors: StD_{n-1} and one property of interest—eccentricity, spatial location (peripheral blocks only: same or opposite hemifield relative to previous stimulus), and difference in the mean trajectories of the random-dot kinematograms presented in consecutive trials. *Notes*: $P(M)$ = prior probability of each model, assumed to be equal for all; $P(M|data)$ = posterior probability of the model (given the data); BF_M = Bayes factor for the model; BF_{10} = Bayes factor for the alternative hypothesis relative to a null model (expressed by each model). All models include subject.

stimulus presentation might modulate such bias. Previous studies on serial dependence have observed that it appears in the fovea as well as the periphery, and its strength is tuned by spatiotemporal proximity (Fischer & Whitney, 2014). Moreover, if the function of (positive) serial dependence is to promote perceptual continuity (Fischer & Whitney, 2014), it seems reasonable to expect that similarity of other attributes of consecutive stimuli would lead to a stronger influence of the studied feature dimension, especially for two attributes as closely related as ensemble mean and variance.

To test the influence of these properties, we conducted a repeated-measures ANOVA on zRE_n (as dependent variable) with two within-subject factors: StD_{n-1} and each of the features of interest separately (eccentricity, retinal location, and similarity of means).

For eccentricity, both main effects were statistically significant—$F_{\text{eccentricity}}(1, 25) = 31.004$, $p < 0.001$, $\eta^2_{\text{p eccentricity}} = 0.554$; $F_{\text{StD}_{n-1}}(2.662, 66.556) = 7.029$, $p < 0.001$, $\eta^2_{\text{p StD}_{n-1}} = 0.219$ (Greenhouse–Geisser sphericity correction)—while the interaction was not, $F(3.789, 94.722) = 1.710$, $p = 0.157$, $\eta^2_p = 0.064$. This result, as indicated by the roughly parallel plots for fovea and periphery in Figure 2B, suggests that while eccentricity influences the absolute value of the current StD response, it does not modulate the serial dependence exerted by StD_{n-1}. To formally test this hypothesis, we turned to Bayesian repeated-measures ANOVA. Table 1a summarizes the comparisons between all competing models. The largest Bayes factor corresponds to the model including both main effects but not the interaction ($BF_{10} = 3.432 \times 10^{29}$), which outperforms the model that also includes the StD_{n-1} × Eccentricity interaction by a factor of $BF_{\text{main/full}} = 17.645$—strong evidence (Wagenmakers et al., 2017) against its inclusion and in favor of the conclusion that while there is an overall difference in reports, there is no difference in serial dependence across eccentricity.

Regarding the influence of spatial location, we analyzed only peripheral presentation blocks, classifying trials according to whether the previous stimulus had been presented on the same or the opposite hemifield as the current one—that is, same presentation location versus a separation of 40°va between consecutive presentations. Results for the model comparison given by a Bayesian repeated-measures ANOVA are presented in Table 1b: The best model in terms of evidence includes only StD_{n-1} ($BF_{10} = 2.073$), while the worst model also includes the hemifield and the StD_{n-1} × Hemifield interaction ($BF_{10} = 0.120$). This indicates moderate evidence against the full model (including interaction) compared to the null, and strong evidence against it when compared to the most explanatory model—that is, the one with StD_{n-1} only ($BF_{\text{full/StD}_{n-1}} = 0.058$). These results support the hypothesis of serial

dependence being unaffected by the spatial location of consecutive stimuli. To confirm the absence of tuning by spatial proximity, we further assessed the strength of serial dependence separately for trials with repeated versus opposite hemifield location with respect to the previous stimulus. Results for these analyses are presented in the Supplementary Materials, section 1 (see also Supplementary Figure S1). While the data of Experiment 1 suggest a nonsignificant trend toward a stronger serial-dependence effect for same presentation locations, these results are not confirmed by the data of other experiments; for Experiment 3, the trend goes in the opposite direction.

Last, we examined the influence of mean RDK direction on serial dependence of variance—specifically, whether the magnitude of the serial-dependence effect (in variance) depended on the successive presentations containing a similar mean direction. With this aim, we binned the absolute difference between the mean RDK directions in the previous and current trial into five categories: ≤36°, 37°–72°, 73°–108°, 109°–144°, and 145°–180°. As before, we conducted a Bayesian repeated-measures ANOVA with two within-subject factors (StD_{n-1} and mean difference). As shown in Table 1c, the best model included only StD_{n-1} ($BF_{10} = 3.210 \times 10^5$), whereas the model including both main effects and their interaction was the second worst (after the one with mean difference only), with $BF_{10} = 0.281$. The Bayes factor for inclusion of the interaction term indicated extreme evidence against it ($BF_{inclusion} = 3.491 \times 10^{-6}$); this was also the case if the comparison was made between the full model and the model lacking only the interaction ($BF_{main/full} = 1789.55$). This lack of association of mean similarity and serial dependence in variance was further confirmed in a different experiment (detailed in Supplementary Materials, section 2) that used a limited range of mean trajectories, allowing for only four between-trials differences (0°, 35°, 55°, and 90°).

In summary, serial dependence in variance reports is not modulated by low-level properties of the stimulus, including visual eccentricity or spatial location, or associated features such as mean, suggesting that visual variance (operationalized as variance of motion direction) is processed as a feature dimension independent from these properties, at least at the level of perceptual decision making that gives rise to serial dependence.

Positive serial dependence in variance extends up to the latest two trials: Investigations of serial dependence have typically focused on the influence of very recent trial history, examining only the effect of the immediately previous and penultimate trials on reports. We examined serial dependence through trial history by modeling the fixed-effects size of serial dependence while allowing for between-subjects variability, building 10 varying-intercept, varying-slope Bayesian linear mixed-effects models (LMMs) with zRE_n as a dependent variable and StD_{n-t} ($t = 1, \ldots, 10$) as an independent predictor, with random effects grouped by participant. We chose a uniform prior distribution over the real numbers for the fixed-effects coefficient and for the standard deviation of the by-subject varying intercepts and slopes, and an LKJ prior with shape parameter $\eta = 2.0$ for the random-effects correlation matrices. Unless stated otherwise, analogous priors were established for other Bayesian LMMs reported in this article. Fixed-effects coefficient estimates were largely insensitive to prior selection, as can be seen in the example presented in the Supplementary Materials (section 3, Supplementary Figure S2).

We applied these models to foveal and peripheral blocks separately, as well as to the overall data set. Figure 2C presents the fixed-effects coefficients and 95% credible intervals for the association between past StD (up to Trial $n - 10$) and current report for all trials, as well as per eccentricity. The value of the LMM fixed-effects coefficient estimate for the effect of StD_{n-t} on zRE_n represents the linear slope for the relationship between the StD presented in Trial $n - t$ and the normalized response error provided in the current trial: in other words, the variation (in z scores) in zRE_n when StD_{n-t} increases by 1°. Therefore, a positive B coefficient represents an attractive bias: A larger StD in a past trial drives a larger response in the present one, regardless of the current stimulus. Conversely, a negative B coefficient represents a repulsive bias.

The fixed-effects B coefficient estimates for the effect of StD_{n-1} and StD_{n-2} on zRE_n are positive, indicating an attractive bias. For StD_{n-1} (all trials pooled), $B = 0.0034$ [0.0017, 0.0051] suggesting that regardless of the value of StD_n, participants' judgments of visual variance increased by a magnitude of 0.0034 (z score) per 1° increase in previous-trial StD (StD_{n-1}). The effect of StD_{n-2} is weaker but still present: $B = 0.0014$ [0.0003, 0.0026]. To make clear the size of these effects, we can consider absolute responses as an outcome variable (adding the current StD_n and the interaction with StD_{n-1} to the models). Here, the increase is 0.0586 (0.0272–0.0892) units per unit of StD_{n-1}, or an attractive effect of 5.9% toward the previous stimulus, whereas for StD_{n-2} the effect size is 0.0242 (0.0006–0.0483), or 2.4%.

Thus, variance judgments at one specific trial (n) are attracted by a small but meaningful amount toward the variance presented in the previous trial ($n - 1$) and, to a lesser extent, the trial before that ($n - 2$). Note that since the initial position of the response bar is randomized for each trial, simple motor routines involved in response execution cannot explain this serial dependence.

Variance reports are subject to a negative bias driven by less recent trial history

Looking past the previous two trials, as shown in Figure 2C, a reversal from positive (Trials $n - 1$ and $n -$

2) to negative B coefficient values is observed for less recent presentations, indicative of a negative (i.e., repulsive, anti-Bayesian) bias: Current responses were *less* similar to the StD presented in those trials, in a manner akin to sensory-adaptation aftereffects (Kohn, 2007; Payzan-LeNestour et al., 2016). This effect started at Trial $n - 4$, peaked at Trials $n - 7$ to $n - 9$ (StD_{n-8}: $B = -0.0021$ [−0.0032, −0.0010]), and started to decline afterward. Similar effect sizes and timescales are observed for foveal and peripheral presentations (see Figure 2C).

In the Supplementary Materials (section 4, Supplementary Figure S3) we present a complete exploration of the evolution of the negative effect in relation to more remote positions in trial history. Evidence for the negative effect starts to fade after Trial $n - 9$ but persists to some extent until Trial $n - 20$.

To confirm that the observed serial effects were truly dependent on trial history, we conducted extensive control analyses exploring potential serial dependences in relation to future presentations (StD_{n+1}) and to shuffled data (see Supplementary Materials, section 5, Supplementary Figure S4). These analyses confirm that only in the true trial history is there evidence for the obtained negative and positive aftereffects, supporting the conclusion that these effects are not simply due to statistical artifacts.

Experiment 2: Processing stages involved in serial dependence in variance reports

In the previous experiment we found evidence for serial dependence in judgments about the variance of RDK stimuli. Specifically, we found two opposite types of bias at different timescales: an attractive, Bayesian-like bias related to the StD of the very recent ($n - 1$ and $n - 2$) trials and a repulsive, negative bias which operates on a longer timescale.

At what level of processing do these serial dependences exert their influence? The nonlocal nature and independence from intertrial similarity in RDK direction mean suggest that attractive serial dependence may not be driven by low-level, sensory processes. However, the specific stages of variance processing at which it arises are yet to be determined (Fischer & Whitney, 2014; Fritsche, Mostert, & de Lange, 2017; John-Saaltink et al., 2016). To address this issue, in this experiment we applied several manipulations to the task to disambiguate the contributions of low-level sensory processes, perceptual decision, and responses to serial dependence of variance judgments.

Experiment 2A aimed to isolate the contribution of response to the serial dependence effect by introducing no-response trials, in order to exclude the influence of physically making a response. However, in this experiment no-response trials were not precued, meaning that a potential contribution of decision making and response preparation during stimulus presentation could not be ruled out. For this reason, in Experiment 2B we employed a (precued) task-switching design to disentangle the contribution of perception and decision processes.

Methods

The methods of these experiments were similar to those of Experiment 1, with exceptions as follows.

Stimuli

All stimuli were presented on the center of the screen (where a fixation cross was displayed), and eye tracking was not performed, as visual eccentricity was not under examination.

Procedure

Experiment 2A had a practice block with 72 trials and 10 experimental blocks with 60 trials each; the same was true for Experiment 2B, except that the practice block was longer (90 trials), due to the additional demands of the task-switching design (see later). In Experiment 2B nine participants (out of 16) performed a session twice as long (10 blocks), due to differential availability of different participants. For both experiments, two thirds of the trials in each block required randomness scores as described for Experiment 1. In Experiment 2A, the remaining one third were no-response trials: After stimulus presentation, instead of the response bar only a blank screen appeared, for a randomized interval between 1,000 and 3,000 ms, after which the next trial started. Participants were told in advance that they should expect a certain number of no-response trials, but they did not know the proportion and these trials were not precued in any way. In Experiment 2B, one third of trials required participants to report the mean direction of the motion of the RDK, by adjusting a rotating arrow with the mouse (see Figure 1). The required task was precued at the beginning of the trial: One three-letter string, either "RAN" (randomness report required) or "DIR" (mean direction report required), was displayed for 1 s before the appearance of the fixation cross. The rest of the trial structure was the same as in Experiment 1 (only the response scale differed in RAN and DIR trials).

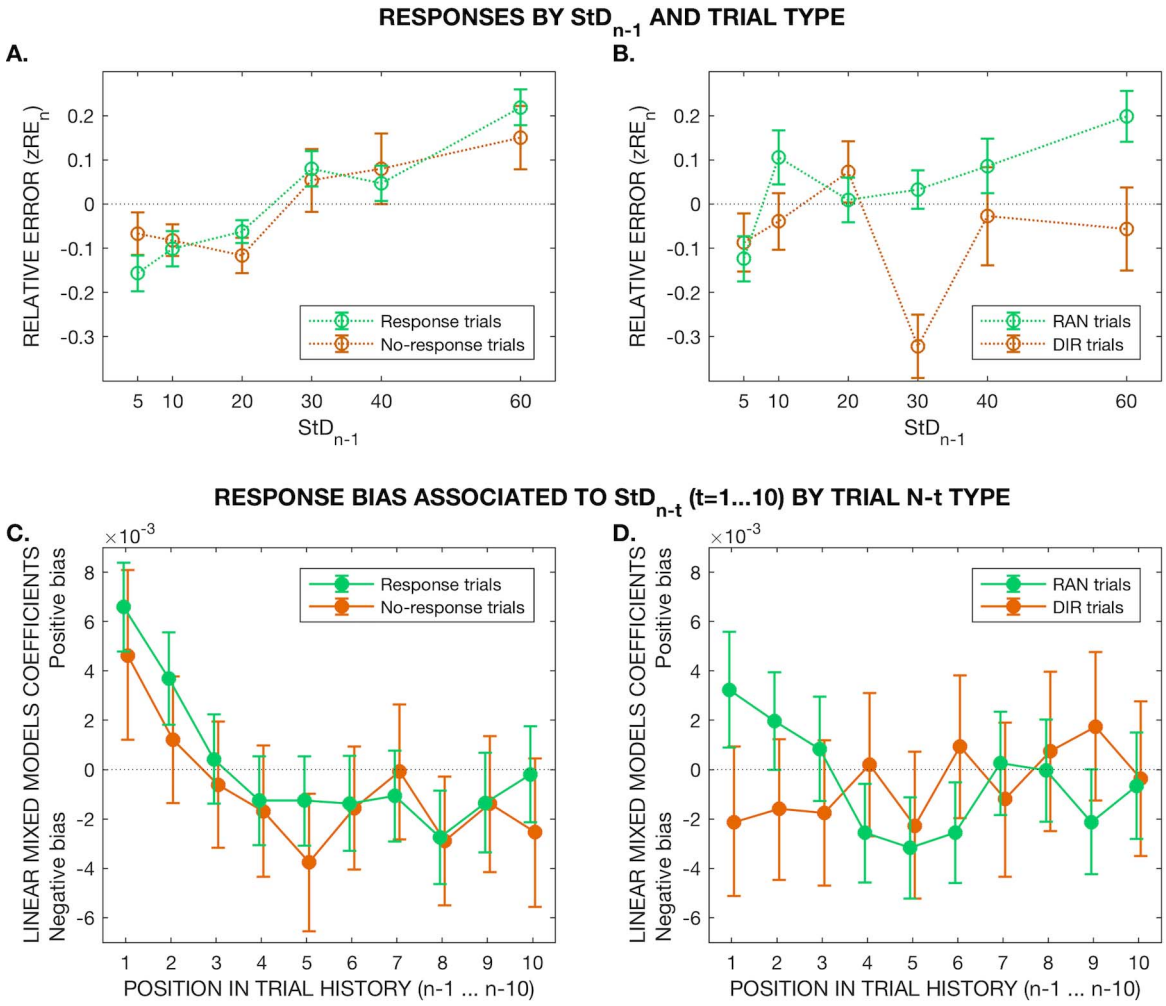

Figure 3. Experiment 2. (A, B) Normalized relative error in current response (zRE_n) as a function of the StD presented in the previous trial (StD_{n-1}), plotted separately by Trial $n-1$ type: (A) response versus no-response (Experiment 2A); (B) RAN versus DIR (Experiment 2B). The error bars represent the between-subjects standard error. Both response and no-response trials are associated with a positive bias by Trial $n-1$, as suggested by the ascending plot lines in (A), whereas in (B), only RAN trials elicit such positive serial dependence. (C, D) Fixed-effects coefficient estimates in 20 Bayesian linear mixed-effects models with StD_{n-t} ($t = 1, \ldots, 10$) as predictor of current response (zRE_n), modeled separately by Trial $n-t$ type: (C) response versus no-response (Experiment 2A); (D) RAN versus DIR (Experiment 2B). Since the dependent variable is the current variance (randomness) judgment, Trial n is always a response (C) or RAN (D) trial. The error bars represent the 95% credible intervals for the true value of the coefficient.

Results

Experiment 2A: Effect of response execution on serial dependence in variance reports

Fifteen psychology students (13 female, two male; mean age = 20.4 years, standard deviation = 5.3) volunteered in exchange for course credits, under the conditions described previously. The total number of trials collected across all participants was 9,000, out of which 3,000 were no-response trials.

Serial dependence of previous StD is not affected by response processes: Figure 3A shows the distribution of normalized variance reports (zRE_n) as a function of the previous trial's StD (StD_{n-1}) and type—that is, whether $n-1$ was a response or a no-response trial. The ascending and roughly parallel plots for each trial ($n-1$) type suggest that serial dependence in relation to StD_{n-1} was similar in magnitude and sign (i.e., attractive effect) regardless of whether Trial $n-1$ was a response or a no-response trial. To formally test this observation, we conducted a Bayesian repeated-measures ANOVA on the effect of StD_{n-1} and Trial $n-1$ type (as within-subject factors) on zRE_n. A comparison of all possible models based on the results of this analysis is shown in Table 2A. The best model includes only StD_{n-1} ($BF_{10} = 2.386 \times 10^6$). There was strong evidence against the inclusion of the $StD_{n-1} \times$ Trial $n-1$ type interaction: $BF_{inclusion} = 0.051$. In a direct comparison between the main-effects model and the full model, the ratio was given by $BF_{main/full} = 10.75$.

Models	$P(M)$	$P(M\|data)$	BF_M	BF_{10}	Error %
(a) Response (Experiment 2A)					
Null model	0.200	3.575×10^{-7}	1.430×10^{-6}	1.000	
StD_{n-1}	0.200	0.853	23.192	2.386×10^{6}	0.393
Trial $n-1$ type (response vs. no-response)	0.200	5.551×10^{-8}	2.220×10^{-7}	0.155	0.772
StD_{n-1} + Trial $n-1$ type	0.200	0.135	0.622	376,281.756	1.178
StD_{n-1} + Trial $n-1$ type + $StD_{n-1} \times$ Trial $n-1$ type	0.200	0.013	0.051	35,151.882	2.835
(b) Dimension-specific judgment (Experiment 2B)					
Null model	0.200	0.012	0.048	1.000	
StD_{n-1}	0.200	0.023	0.095	1.964	0.360
Trial $n-1$ type (variance vs. mean estimation)	0.200	0.109	0.489	9.202	1.348
StD_{n-1} + Trial $n-1$ type	0.200	0.283	1.578	23.921	2.994
StD_{n-1} + Trial $n-1$ type + $StD_{n-1} \times$ Trial $n-1$ type	0.200	0.573	5.371	48.459	2.705

Table 2. Serial dependence and task requirements—model comparison: Experiment 2, serial dependence (associated with Trial $n-1$) and Trial $n-1$ type (task requirements in that trial). The panels present model performance on Experiment 2A and 2B data sets, respectively, according to the results of a Bayesian repeated-measures ANOVA on zRE_n, with two within-subject factors: StD_{n-1} and Trial $n-1$ type (response/no-response in Experiment 2A, RAN/DIR in Experiment 2B). *Notes*: $P(M)$ = prior probability of each model, assumed to be equal for all; $P(M|data)$ = posterior probability of the model (given the data); BF_M = Bayes factor for the model; BF_{10} = Bayes factor for the alternative hypothesis relative to a null model (expressed by each model). All models include subject.

This lack of interaction confirmed that the effect of StD_{n-1} on current reports was independent of response execution.

Figure 3C shows the fixed-effects coefficient estimates and 95% credible intervals for 20 Bayesian LMMs for zRE_n, with StD_{n-t} ($t = 1, \ldots, 10$) as putative predictor, split by Trial $n-t$ type and modeled separately. A similar pattern in terms of effect size and direction can be seen regardless of whether previous trials required response or not: an attractive bias in relation to the latest two trials (weaker for $n-2$), a roughly zero effect of Trial $n-3$, and a reversal toward a negative effect peaking around Trials $n-5$ to $n-9$, with a similar magnitude and timescale as for Experiment 1.

Having established that serial dependence does not arise from response itself, we questioned whether intermediate responses (i.e., responses made in past trials between the current one, n, and a Trial $n-t$ whose serial effect is considered) could affect the degree to which the effect of further trials carried through. For simplicity, we considered only the case of serial dependence related to Trial $n-2$ (for the sake of homogeneity, we limited the analysis to those response trials wherein Trial $n-2$ was also a response trial) and classified the data set according to whether the intermediate trial ($n-1$) was a response or a no-response trial. We ran a Bayesian repeated-measures ANOVA on the effect of StD_{n-2} and Trial $n-1$ type (as within-subject factors) on zRE_n. The best model contained only StD_{n-2} ($BF_{10} = 30.045$), outperforming the full model (two factors and interaction) by a factor of 12.87. However, when the comparison was made between the full model and the equivalent model stripped of the effect of interest (i.e., the $StD_{n-2} \times$ Trial $n-1$ type interaction), the latter outperformed the former by a factor of only $BF_{main/full} = 1.98$. Overall, the Bayes factor for inclusion of the interaction term indicated moderate evidence against it ($BF_{inclusion} = 0.261$), suggesting that the attractive bias related to previous trials is not disrupted (nor boosted) by the participant providing a response on the intermediate trials.

Experiment 2B: Effect of decision on serial dependence in variance reports

Experiment 2A demonstrated that serial dependence in visual variance is not due to response execution; however, as trials were not precued as to whether a response would be required, these results do not disambiguate between perception and decision making (response preparation). Therefore, in Experiment 2B we deployed a precued task-switching design in which participants needed to prepare and respond to two different perceptual tasks: reporting the variance (RAN trials) or the mean (DIR trials) of the motion of the RDK.

Fifteen first-year psychology students (13 female, two male; mean age = 21.4 years, standard deviation = 8.8) participated in this experiment in exchange for course credits, under the conditions already described. In total they performed 7,200 trials, out of which 2,400 were DIR-trials (alternative task).

Serial dependence is related to dimension-specific decision making: We analyzed the data in a similar manner to Experiment 2A, ascertaining the influence of trial type in the observed serial dependence on variance judgments. Figure 3B presents the distribution of variance reports (zRE_n) as a function of StD_{n-1} and Trial $n-1$ type—that is, whether it required a decision

about variance (RAN) or mean (DIR). Only when successive decisions were both regarding variance do we see an ascending slope in relation to increasing StD_{n-1}, suggesting that the attractive bias associated with StD_{n-1} was only exerted if a decision on that dimension had been made.

Table 2b presents a Bayesian repeated-measures ANOVA with StD_{n-1} and Trial $n-1$ type (RAN/DIR) as within-subject factors. The most explanatory was the full model including both main effects and their interaction ($BF_{10} = 48.459$), although the evidence in its favor compared to the model with only the main effects was anecdotal ($BF_{full/main} = 2.026$). However, evidence in favor of the interaction term was larger when taking into consideration all possible models: $BF_{inclusion} = 5.371$, which is moderate evidence. Thus, results point to serial dependence by StD_{n-1} being dependent on which dimension participants had to judge in the previous trial.

We noted that the average time between onsets of consecutive trials was longer if the first was a RAN trial (4.63 vs. 4.48 s, Bayesian pair-samples t test: $BF_{10} = 29.63$). We therefore wondered whether time could be confounding the interaction between StD_{n-1} and Trial $n-1$ response type, since it has been shown to influence serial dependence in previous studies (Bliss, Sun, & D'Esposito, 2017; Fritsche et al., 2017; Kanai & Verstraten, 2005). To test this possibility we defined $time_{n-1,n}$ as the interval between consecutive stimulus onsets, binned into two levels, either below or above the participant's median. This variable was added as a third within-subject factor to the Bayesian repeated-measures ANOVA described in the previous paragraph. We sought to directly compare two explanatory hypotheses for the cause of the observed difference in serial dependence by StD_{n-1} when $n-1$ was a RAN compared to a DIR trial: Trial $n-1$ type or interstimulus time. Thus, we compared the explanatory power of a model with StD_{n-1}, Trial $n-1$ type, and their interaction against a model with StD_{n-1}, $time_{n-1,n}$, and their interaction. The former outperformed the latter by a factor of 105.37, indicating extreme evidence in its favor. Overall, analysis of each separate effect indicated extreme evidence against inclusion of the $StD_{n-1} \times time_{n-1,n}$ interaction ($BF_{inclusion} = 6.668 \times 10^{-4}$). This indicated that the difference between serial dependence driven by RAN compared to DIR trials was better explained by the trial type itself rather than by the intertrial time. There was no support for an independent contribution of time to the observed difference between RAN and DIR trials.

As in previous experiments, we also examined serial dependence within a broader span of trial history. Figure 3D presents the fixed-effects coefficient estimates and 95% credible intervals for the association between StD_{n-t} ($t = 1, \ldots, 10$) and zRE_n, after splitting the data set according to the trial type at each position; thus, the influence of RAN and DIR trials is modeled separately by 20 Bayesian LMMs. As expected from the previous analysis, the positive effect related to StD_{n-1} is present only when those trials required participants to report variance; this is also the case for StD_{n-2}. As for the negative effect appearing at longer timescales, it is clearly present in RAN trials, while for DIR trials, although the credible intervals for the coefficient contain zero at all trial positions (likely due to the smaller number of DIR trials), the negative effect seems to appear as early as Trial $n-1$ ($B = -0.0021$ [–0.0051, 0.0009]), peak at Trial $n-5$ ($B = -0.0023$ [–0.0052, 0.0007]), and decrease afterward. The appearance of a negative serial dependence regardless of the task suggests that it may be sensory in origin—an adaptation aftereffect.

If we ask why there is a serial-dependence effect at $n-1$, we should also ask why there is no such effect at $n-3$. Thus, having established that positive serial dependence arises from feature-specific decision making, we investigated the inverse question: What is the contribution of feature-specific decision making to the fading of positive serial dependence for trials located further away in history? Is this decline affected in a different way by subsequent decisions made on the same, compared to a different, feature dimension? Like for Experiment 2A, we considered all those RAN trials for which Trial $n-2$ was also of type RAN, and examined the association between StD_{n-2} and current report in relation to the intermediate trial's ($n-1$) task. An explanatory role for the $StD_{n-2} \times$ Trial $n-1$ response type interaction would indicate that the intermediate trial type influenced serial dependence related to $n-2$. In a Bayesian two-factor repeated-measures ANOVA, the best model included only Trial $n-1$ response type ($BF_{10} = 41.799$), suggesting that there was no interaction with serial dependence related to StD_{n-2}.

Experiment 3: Influence of confidence in serial dependence

Results of Experiments 1 and 2 indicate that positive serial dependence in visual variance involves mid- to high-level processes, namely decision making about the same feature dimension. In light of this, we questioned how confidence in those decisions modulates serial dependence. We were especially interested in the modulation of the positive (Bayesian-like) bias exerted by very recent trials, in the light of Bayesian accounts of confidence as a measure of the precision of neural representations (Meyniel et al., 2015).

Methods

Stimuli

Stimulus presentation was identical to Experiment 1; we again used foveal and peripheral (20°) presentations, as we considered that the interplay between decision making, confidence, and serial dependence might vary at different degrees of sensory precision.

Procedure

Experiment 3, like Experiment 1, had a 72-trial practice block and eight 60-trial experimental blocks, half of which were foveal and half peripheral. Eye tracking was performed in the same manner as in Experiment 1.

During the response phase of each trial, two identical visual analogue scales were displayed on the screen: the upper one for scoring randomness (variance) and the lower one for confidence (see Figure 1). The initial position of each sliding bar was randomized separately, and the time allowed for responding to both items was 6 s. For data analysis, we obtained the numerical scores as a linear translation from the selected positions; for confidence, the score was expressed as a 0-to-1 proportion of the overall length of the line.

Results

Twenty-two participants (17 female, five male; mean age = 19.6 years, standard deviation = 2.42) volunteered for this experiment. All except for three members of the laboratory were first-year psychology students. As in Experiment 1, trials without valid fixation during stimulus presentation were removed from the analysis, as well as data about trial history of valid trials involving any invalid trial. In total, 8,880 trials were included in the analyses.

Confidence reports correlate with the accuracy and precision of variance judgments

Figure 4A presents the distribution of confidence scores (C_n) plotted by current-stimulus StD (StD_n) and eccentricity. For both foveal and peripheral trials, a trend toward decreasing C_n for larger StD_n is observed, except for the maximal StD (60°). For each StD value, confidence scores are lower in the periphery. To test these observations, we conducted a Bayesian repeated-measures ANOVA on the effect of StD_n and eccentricity (as within-subject factors) on C_n. The best model was the one including both main effects only ($BF_{10} = 6.657 \times 10^{26}$), outperforming the full model with the $StD_n \times$ Eccentricity interaction by a factor of $BF_{main/full} = 9.615$. This indicates that despite the overall lower confidence scores in peripheral blocks, the relationship between different stimulus levels and confidence is the same regardless of eccentricity.

Subsequently we explored whether confidence reports were differentially shaped by response accuracy or precision, and considered the role of eccentricity. Regarding accuracy, we defined error size as the absolute value of the difference between real and reported StD: $E_n = |StD_n - R_n|$. In a Bayesian LMM with C_n as dependent variable and E_n, StD_n, and their interaction as independent variables, C_n reports are inversely associated with error size ($B = -0.0083$, 95% credible interval [−0.0103, −0.0062]) and StD_n ($B = -0.0056$ [−0.0071, −0.0040]) and positively associated with the interaction between both ($B = 0.0003$ [0.0002, 0.0003]). The inverse association between error size and C_n suggests that participants' reports of confidence are, at least in part, grounded in task accuracy. Furthermore, the positive sign of the coefficient estimate for the $E_n \times StD_n$ interaction suggests that confidence tracks relative rather than absolute error: The inverse association between error size (defined as an absolute value) and confidence is weighted down for large StD values. Considering both error size and eccentricity, the negative association with error size remains ($B_{error} = -0.0078$ [−0.0102, −0.0055]), whereas foveal presentations are associated with higher confidence reports independent of task accuracy ($B_{eccentricity} = 0.0510$ [0.0080, 0.0908]). However, the interaction term does not show evidence of a different evaluation of increases in error size in low compared to high eccentricities ($B_{error \times eccentricity} = -0.0013$ [−0.0040, 0.0016]).

As for precision, we calculated the standard deviation of each participant's responses per StD value (σ_R) as a measure of response dispersion. We subsequently modeled confidence by σ_R, StD, and their interaction. As expected, response dispersion shows a negative correlation with confidence: $B = -0.0101$ (95% credible interval [−0.0160, −0.0045]). When we add eccentricity to this model, the main effect for σ_R is close in value ($B = -0.0105$ [−0.0162, −0.0050]), whereas the $\sigma_R \times$ Eccentricity interaction ($B = -0.0003$ [−0.0067, 0.0062]) suggests that the interaction between response dispersion and confidence is similar in fovea and periphery. In summary, our results indicate that confidence is a measure of response precision, and to the extent to which the latter can be considered a proxy for perceptual precision, they are in agreement with Bayesian accounts of metacognition (Meyniel et al., 2015).

Interestingly, we observed a very strong serial dependence for confidence reports. Modeling reported confidence (by a Bayesian LMM) as a function of the report provided in the previous trial (C_{n-1}), we find that the coefficient for the latter is $B = 0.1874$ [0.1445, 0.2307], with $R^2 = 0.3188$. Importantly, if we add the

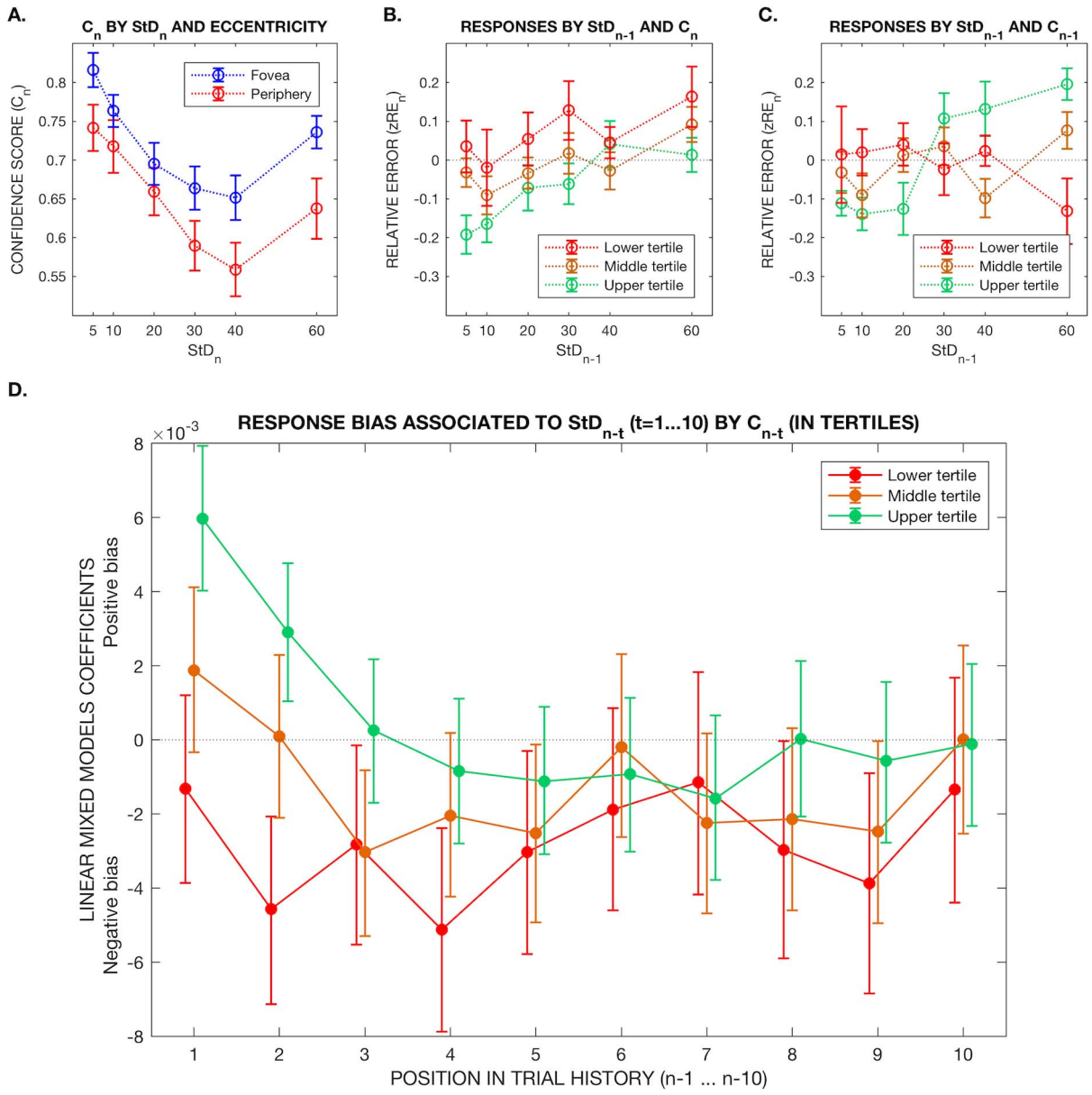

Figure 4. Experiment 3. (A) Confidence scores (C_n) by StD_n plotted separately by eccentricity. (B, C) Normalized relative error in current response (zRE_n) as a function of the StD presented in the previous trial (StD_{n-1}), plotted separately by confidence reported in (B) the current or (C) the previous trial. Confidence scores have been binned into tertiles according to each participant's distribution of reports. The error bars represent the between-subjects standard error. The plots in (B) are all ascending and roughly parallel, indicating that current confidence does not modulate serial dependence by previous-trial StD. Conversely, when considering confidence reported in the previous ($n-1$) trial (C), we observe drastically different slopes: While the high-confidence plot (upper tertile) has a clear ascending slope indicative of a positive bias, the middle-tertile plot is only mildly positive, and the lower-tertile plot is slightly descending, suggesting a negative bias away from low-confidence $n-1$ trials. (D) Fixed-effects coefficient estimates in 30 Bayesian linear mixed-effects models with StD_{n-t} ($t = 1, \ldots, 10$) as predictor of current response (zRE_n), modeled separately by confidence reported in Trial $n-t$ (C_{n-t}), binned into tertiles. The error bars represent the 95% credible intervals for the true value of the coefficient. As suggested for Trial $n-1$ in (C), the size and direction of the bias associated with each trial position depends on the confidence reported in that position, so that the bias will be more negative (or less positive) the lower the reported confidence, within the general trend of an increasingly negative (less positive) bias as we move backward in history.

error size of the previous trial (E_{n-1}) to the model, as well as the $E_{n-1} \times C_{n-1}$ interaction, the coefficient estimate for C_{n-1} has a similar (even larger) value: $B = 0.2197$ [0.1698, 0.2720]. This is also the case when StD_{n-1} is included in the model, suggesting that the serial dependence in confidence scores is due not only to accuracy or attention fluctuating at timescales of several trials, nor to the direct influence of the StD in

Model	$P(M)$	$P(M\|data)$	BF_M	BF_{10}	Error %
(a) Current-trial confidence					
Null model	0.200	0.002	0.010	1.000	
StD_{n-1}	0.200	0.011	0.043	4.437	0.393
C_n	0.200	0.137	0.637	57.150	0.642
$StD_{n-1} + Cn$	0.200	0.841	21.088	349.668	1.213
$StD_{n-1} + Cn + StD_{n-1} \times Cn$	0.200	0.009	0.036	3.738	5.208
(b) Previous-trial ($n-1$) confidence					
Null model	0.200	0.923	47.768	1.000	
StD_{n-1}	0.200	0.025	0.102	0.027	0.535
C_{n-1}	0.200	0.030	0.123	0.032	1.097
$StD_{n-1} + C_{n-1}$	0.200	8.080×10^{-4}	0.003	8.757×10^{-4}	1.439
$StD_{n-1} + C_{n-1} + StD_{n-1} \times C_{n-1}$	0.200	0.022	0.089	0.024	1.099

Table 3. Serial dependence and reported confidence—model comparison: Experiment 3, serial dependence (associated with Trial $n-1$) and confidence reported in the current and previous trial. The panels report the results of a Bayesian repeated-measures ANOVA on zRE_n, with two within-subject factors: StD_{n-1} and current- or previous-trial confidence. *Notes*: $P(M)$ = prior probability of each model, assumed to be equal for all; $P(M|data)$ = posterior probability of the model (given the data); BF_M = Bayes factor for the model; BF_{10} = Bayes factor for the alternative hypothesis (expressed by each model). All models include subject.

the previous stimulus, but rather may be an expression of response inertia or "confidence leak" as described by Rahnev, Koizumi, McCurdy, D'Esposito, and Lau (2015).

Confidence in a past trial determines the direction of serial dependence in variance reports

According to Bayesian accounts of perceptual decision making, reliance on prior information is greater when the current sensory input is noisy or imprecise, or when the prior itself is highly precise (Cicchini et al., 2014; Petzschner & Glasauer, 2011; Summerfield & de Lange, 2014). Within this framework, confidence is often regarded as a measure of the precision of the sensory signal (Meyniel et al., 2015), a consideration that is in agreement with our data. Thus, we hypothesized that high reported confidence in the current trial (C_n) would decrease any attractive pull toward previous history (with respect to variance judgments), whereas confidence in past trials (C_{n-t}) would have the opposite effect. We further reasoned that such effect of confidence in the past trials would apply mostly to very recent trials, whose information represents a more important contribution when priors are iteratively updated. Indeed, this second hypothesis is in agreement with our observation of a positive bias in variance judgments exerted by only the most recent trials (see Figure 2C for an example).

Figure 4B and 4C depicts zRE_n as a function of StD_{n-1}, plotted separately by confidence in the current (4b) and previous (4c) trial. Confidence scores have been binned into tertiles on a per-participant basis. In Figure 4B, all three plots present an ascending, roughly parallel slope: It appears that serial dependence exerted by Trial $n-1$ takes place independently of the confidence placed in the current judgment, contrary to our initial hypothesis. However, when we consider the influence of confidence in the previous response, we do see a striking interaction, in line with what would be expected within a Bayesian framework: Low-confidence $n-1$ judgments do not exert any positive serial dependence—quite the opposite, the plot has a slightly descending slope, pointing toward a negative bias in relation to StD_{n-1}. This slope is mildly ascending for medium confidence and neatly positive only for high-confidence past decisions.

In order to validate these observations, we first performed a Bayesian repeated-measures ANOVA on the effect of StD_{n-1} and C_n (confidence score in the current trial, binned into tertiles) on zRE_n. Results are presented in Table 3a. The best model contains both main effects but not the interaction ($BF_{10} = 349.668$), outperforming the model with the interaction term by a factor of $BF_{main/full} = 93.544$. This provides very strong evidence against the inclusion of the interaction term and indicates that confidence in the current judgment does not modulate serial dependence from the previous trial.

Subsequently we performed an analogous analysis, but with StD_{n-1} and C_{n-1} as within-subject factors. Table 3b presents the results of this analysis. Evidence is in favor of the null model by a large margin (31.25 times more explanatory than the second best, which includes only C_{n-1}). Nevertheless, when we consider the term of interest for our hypothesis, namely the $StD_{n-1} \times C_{n-1}$ interaction, there is strong evidence in favor of its inclusion compared to the model stripped of that effect (including only the two main factors): $BF_{full/main} = 26.989$. Still, because neither competing model was superior to the null model, this result must be taken with caution.

We next asked to what degree confidence in trials located further back in history, up to $n - 10$, influenced serial dependence of variance judgments. We split the data set according to the confidence scores reported in each past position (C_{n-t}, discretized into tertiles within each participant's scores), and ran three Bayesian LMMs per position (30 models in total) for the association between StD_{n-t} and zRE_n at each level of past confidence. Figure 4D presents the B coefficient estimates and 95% credible intervals for each Trial $n - t$ ($t = 1, \ldots, 10$). A marked influence of past confidence on the size and direction of serial dependence is observed, such that when high confidence was reported in very recent trials ($n - 1$, $n - 2$), an attractive pull toward recent StD values is manifest, although this bias fades rapidly, being absent by Trial $n - 3$ and thereafter. Note that trials with highest confidence (upper tertile) do not exert a clear, unambiguous negative bias at any point of trial history, although some traces seem to be present from Trial $n - 4$ onward. The largest negative bias is driven by low-confidence trials, for which it seems to appear as recently as Trial $n - 1$ (although the credible intervals contain zero), becomes unambiguous at $n - 2$, and peaks at $n - 4$, decreasing afterward—in contrast with the slower buildup of the negative bias seen for past trials with intermediate confidence. Thus, the reversal from positive to negative bias seen in this and previous experiments seems related to the rapid decay of the positive bias of high-confidence trials. As for the negative effect, it seems to appear as early as whenever such competing (positive) bias is not manifest, but fades more slowly. Results were similar when we considered foveal and peripheral blocks separately.

At first glance, the early appearance of the negative effect (after exposure to a single subsecond presentation) and its association with low confidence could suggest that it is at least in part of decisional origin rather than exclusively a product of sensory adaptation. However, some amount of negative bias was observed in relation with past DIR trials in Experiment 2B (trials in which participants were not making a decision on variance). Thus, it seems more likely that the apparent relationship between the negative effect and confidence is due to concealment of the effect in the presence of the positive bias, the latter being associated with high-confidence decision making.

On average, response times for variance reports in low-, medium-, and high-confidence trials were 1.59, 1.46, and 1.30 s, respectively (Bayesian repeated-measures ANOVA: $\mathrm{BF}_{10} = 22{,}288$, extreme evidence for the alternative hypothesis), presumably related to subjective trial difficulty. Therefore, we sought to rule out the possibility that the effect of past confidence on serial dependence was related only to the difference in response times, and consequently in interstimulus times. For each trial up to $n - 10$, we performed a three-way Bayesian repeated-measures ANOVA for zRE_n (as dependent variable) with three within-subject factors: StD_{n-t}, C_{n-t} (in tertiles), and $\mathrm{time}_{n,n-t}$ (time between stimulus onset of Trials $n - t$ and n, binned in two levels with respect to the median). In all cases, the evidence for inclusion of the $\mathrm{StD}_{n-t} \times \mathrm{time}_{n,n-t}$ interaction was extremely low—that is, the Bayes factor for this specific effect was always below 0.01. This suggests that time was not confounding the reported interaction between confidence and serial dependence.

Time and the additional confidence report might promote an earlier reversal toward negative serial dependence in variance judgments

Experiment 3 had an identical design to Experiment 1 except for the requirement of an additional report (about confidence) per trial. Consequently, an additional difference was introduced: The intertrial time was longer in Experiment 3 that in Experiment 1 (5.06 vs. 3.69 s, Bayesian independent samples t test: $\mathrm{BF}_{10} > 6.690 \times 10^7$). As previous work has strongly implicated time between successive stimuli or stimuli and response as critical contributors to serial dependence (Bliss et al., 2017; Fritsche et al., 2017; Kanai & Verstraten, 2005), we sought to take advantage of this circumstance to inquire (post hoc) about the factors that drive the decrease and eventual shift toward negative of the serial-dependence effect as we move backward in trial history.

Figure 5A presents the Bayesian LMM coefficients and 95% credible intervals for the effect of StD_{n-t} ($t = 1, \ldots, 10$) in current variance report as found for Experiments 1 and 3. An extension of this comparison for more distant trial positions is presented and discussed in the Supplementary Materials, section 4 (see Supplementary Figure S3). While the positive bias exerted by StD_{n-1} is similar in magnitude in both experiments ($B = 0.0034$ [0.0017, 0.0051] in Experiment 1, $B = 0.0030$ [0.0018, 0.0042] in Experiment 3), such attraction is still present at StD_{n-2} in Experiment 1 ($B = 0.0014$ [0.0003, 0.0026]) but has virtually disappeared in Experiment 3 ($B = 0.0003$ [−0.0009, 0.0015]). Thus, in Experiment 3 the reversal to negative bias occurs as early as Trial $n - 3$ and peaks at $n - 5$ ($B = -0.0023$ [−0.0036, −0.0010]), with a similar effect size as the maximum negative bias in Experiment 1, which is seen at $n - 8$ ($B = -0.0021$ [−0.0032, −0.0010]). As shown in the Supplementary Materials, negative serial dependences also decline and disappear earlier than in Experiment 1. This earlier buildup of the negative bias could be related to the longer interstimulus intervals in the present experiment: Time might, hypothetically, drive the reversal to repulsive serial effects and posterior fading. Results of Experiment 2B (concerning

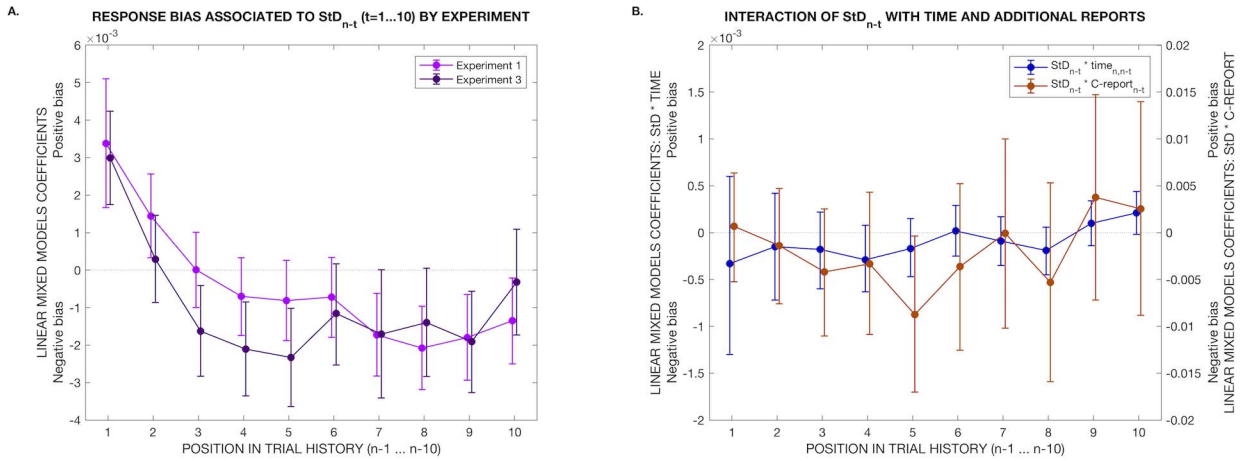

Figure 5. Comparison between Experiments 1 and 3. Both experiments have the same design except for the requirement of a confidence report (in addition to a variance report) per trial in Experiment 3. This also makes the interstimulus time longer, on average, for Experiment 3 compared to Experiment 1. (A) Fixed-effects coefficient estimates in 20 Bayesian linear mixed-effects models with StD_{n-t} ($t = 1, \ldots, 10$) as predictor of current response (zRE_n), with the data of Experiments 1 and 3 modeled separately. The error bars represent the 95% credible intervals for the true value of the coefficient. The shift toward negative coefficient estimates takes place at earlier trial positions in Experiment 3. (B) Fixed-effects coefficient estimates for the $StD_{n-t} \times time_{n,n-t}$ and $StD_{n-t} \times C\text{-report}_{n-t}$ interactions in 10 Bayesian linear mixed-effects models for prediction of zRE_n, with StD_{n-t}, $time_{n,n-t}$, $C\text{-report}_{n-t}$, and all interactions as putative predictors. The variable $time_{n,n-t}$ reflects the time between onsets of the stimuli in Trials $n - t$ ($t = 1, \ldots, 10$) and *n*. $C\text{-report}_{n-t}$ is a binary factor indicating whether confidence reports were made in all trials between $n - t$ and n or in none, regardless of the content of the reports (i.e., the amount of confidence). A negative interaction term with StD_{n-t} indicates a less positive (more negative) serial-dependence effect in relation with longer time or the requirement of an additional confidence report per trial. While credible intervals contain zero in most instances, there is a predominance of negative estimates up to $n - 5$, which could suggest a causal role for both time and the additional confidence report in terms of promoting an earlier reversal of the bias in Experiment 3 compared to Experiment 1.

the effect of DIR trials) and on low-confidence trials in Experiment 3 seem to suggest that the negative bias appears as early as whenever the conditions for the arising of a positive bias are not met. If, hypothetically, positive serial dependence declines with time, the negative effect could become evident in an earlier trial in relation to the longer interstimulus times observed in Experiment 3. Another explanation for the earlier shift toward negative in Experiment 3 would be a disruption of the positive bias caused by the additional confidence report—especially if such Bayesian-like pull is caused by decision processes or depends upon memory to some extent.

To further examine this issue, we pooled all valid trials from Experiments 1 and 3. To ascertain the influence of an additional decision made per trial beside the variance judgment, we defined a binary variable, named $C\text{-report}_{n-t}$, indicating whether or not all intermediate trials between n and $n - t$ had a confidence report in addition to a variance report. Note that the content of the reports (i.e., the amount of confidence) did not affect this definition. When participants missed at least one confidence report in the considered historical span of a certain trial, that trial was excluded from the model, in order to make the comparison unambiguous. Subsequently we built 10 Bayesian LMMs for zRE_n (as dependent variable) in relation with three variables defined at each considered point of trial history, namely StD_{n-t}, $time_{n,n-t}$, and $C\text{-report}_{n-t}$, and all interactions. The fixed-effects B coefficients of the $StD_{n-t} \times time_{n,n-t}$ and $StD_{n-t} \times C\text{-report}_{n-t}$ interactions are plotted in Figure 5B, for Trials $n - 1$ to $n - 10$ as predictors of current variance judgment. A negative interaction coefficient would indicate a comparatively less positive (more negative) serial-dependence effect at that position in relation to longer time or the extra report, respectively.

At all positions, credible intervals for both interaction terms contain zero (except for $StD_{n-t} \times C\text{-report}_{n-t}$ at $n - 5$). However, there is a predominance of negative values for both interaction terms within the recent half of the considered span of trial history, up to Trial $n - 5$. Thus, although results are inconclusive regarding the causes of the different patterns of serial dependence in Experiments 1 and 3, the mostly negative $StD_{n-t} \times time_{n,n-t}$ and $StD_{n-t} \times C\text{-report}_{n-t}$ interactions suggest that both time and the additional confidence report might promote a less positive (more negative) serial dependence in variance and thus contribute to the observed earlier reversal in the direction of the bias. An interesting possibility would be that the dimension-specific, decision-based positive serial dependence is

subject to memory decay as well as a decision-capacity bottleneck. The presented data do not conclusively support a particular interpretation, so future experiments are required to elucidate the relative contribution of time itself and additional judgments in shaping the effects of trial history.

Discussion

The examination of serial dependence provides a valuable window into perceptual processing. In a series of experiments, we applied this approach to visual statistics rather than to individual perceptual features—specifically, to variance, a basic trait in the interpretation of noisy information about complex visual scenes. We found evidence for two opposite serial-dependence effects operating on different timescales: an attractive (positive) bias associated with very recent variance presentations, which is exerted only when a judgment about that dimension was made in the most recent one or two trials and high confidence was placed in that decision, and a repulsive (negative) bias which appears even for the most recent trial history for low-confidence variance presentations but generally becomes manifest several trials into history and persists for at least 10 trials.

Several studies on serial dependence have found a positive (attractive) bias toward recent perceptual history, which is modulated by attention, is enhanced by spatial proximity yet not specific to retinal location, takes place in the fovea as well as the periphery, and fades after 5–15 s but does not require explicit memory (Fischer & Whitney, 2014). While control experiments support the proposition that this effect does not require a motor response, there is an ongoing debate about whether its basis is perceptual or postperceptual: The results of a two-alternative forced-choice discrimination task (with a sample size of three participants; Fischer & Whitney, 2014), a recent behavioral study (Cicchini, Mikellidou, & Burr, 2017), and a V1-based fMRI study (John-Saaltink et al., 2016) have been used in support of a perceptual origin, while another study employing a combination of appearance and performance tasks has made the case for a postperceptual (decisional) source (Fritsche et al., 2017). All four studies examining the mechanistic basis of serial dependence have used a low-level feature like orientation; nevertheless, serial dependence has also been described for high-level features, including facial appearance (Liberman et al., 2014; Xia et al., 2015), relative timing (Roseboom, 2017), and statistical properties such as numerosity (Cicchini et al., 2014) and ensemble mean (Manassi, Liberman, Chaney, & Whitney, 2017).

In our experiments on visual variance (a high-order visual statistic), we found a positive bias that shares many of these characteristics but differs in others: In terms of similarities, it operates on a similar timescale (temporal tuning seems to be slightly shorter for high-level domains, as shown in our data and a study with face perception; Liberman et al., 2014), occurs similarly across presentation eccentricities, and is not related to response execution. It also exhibits other characteristics suggesting that for visual variance, the bias depends on decisional rather than perceptual processes. First, it is entirely independent of retinal location, appearing with similar magnitude for successive stimuli displayed at the same position or at an angular distance of 40°, as shown in the peripheral trials in Experiment 1. Second, it is independent of the closely related statistical property of mean direction (previous studies have highlighted a strong relationship between mean and variance, showing that variance plays an important role in the accuracy and confidence of mean judgments; Fouriezos et al., 2008; Maule & Franklin, 2015). Together, these properties make a low-level, perceptual origin very unlikely. Note that priming of mean judgments by visual variance, as described by Michael et al. (2014), is also independent of the similarity of means and retinal location.

The most compelling argument in favor of a decisional origin for the positive serial dependence in our results is that in a task-switching design, the bias disappears entirely when participants are engaged in a decision about a different feature dimension than variance. This is shown in our Experiment 2B, where participants made decisions about either the variance or the mean direction of the RDK stimuli. This is particularly notable, since mean judgments are strongly dependent on ensemble variance (Fouriezos et al., 2008; Maule & Franklin, 2015), and the stimulus is identical for both tasks. Even so, the possibility remains that the absence of serial dependence in this alternative task may be related to a withdrawal of feature-specific attention (withdrawal of attending to the feature of variance), since serial dependence is enhanced by attention (Fischer & Whitney, 2014). Characterizing the precise effect of attention on serial dependence of variance judgments, and its interaction with task, remains an opportunity for future studies.

From a predictive perception perspective, the fact that only high-confidence trials drive the positive serial dependence may be considered supportive of both perceptual and decisional origin, as a more precise prior would give rise to a stronger reliance on sensory or decisional history (Meyniel et al., 2015). However, an interpretation based on sensory precision might also predict two associations that are not found in our data: an inverse association of positive serial dependence with current-trial confidence, and an inverse associa-

tion with eccentricity, given lower sensory precision in the peripheral field. To the contrary, our experiments strongly support a lack of association of serial dependence with these two factors. In broader terms, serial dependence in variance judgments could be regarded as part of a generic strategy of mirroring or transferring trusted decisions. This explanation could also encompass the negative serial dependence associated with low confidence (as a repulsion away from judgments deemed unreliable); however, the different timescales over which the positive and negative biases operate suggest that they are independent mechanisms rather than two aspects of a confidence-based strategy (Alais, Ho, & Han, 2017).

Finally, several pieces of evidence in our experiments suggest that the positive serial dependence is disrupted by additional decision making, regardless of the domain on which the subsequent decisions operate (variance, mean, confidence). In other words: In all our experiments, the attractive effect of StD_{n-2} stimulus on the current response (which arises only when a high-confidence judgment about variance was made in Trial $n-2$) is much weaker, on average, than that of StD_{n-1}. When inquiring into the factors (interposed between StD_{n-2} and the current response) that might explain this decline, we failed to find any difference based on the *type* of decision that was made in the following trial ($n-1$): In Experiment 2B, the magnitude of the effect of StD_{n-2} (when a variance judgment was made at that point) did not appear to depend upon whether a decision in Trial $n-1$ was made about the variance or the mean of the stimulus. However, if the *number* of interposing decisions was increased, and an additional decision (about confidence) was required in Trial $n-1$, the positive effect of Trial $n-2$ was greatly diminished. This apparent relationship with quantity but not quality of subsequent decisions (made after the one that exerts the bias and before the one that is biased) suggests that serial dependence may be limited by an amodal decision-capacity bottleneck. The apparent fading of the effect with time also points to some sort of memory limitation. Note, however, that these considerations arise from post hoc analyses which revealed only suggestive trends, although the evidence was not conclusive in any case. The factors contributing to the disruption or fading of positive serial dependence in relation with more remote presentations are deserving of further research.

In summary, it is likely that variance-related positive serial dependence is driven by high-level perceptual decision-making processes. In this respect, our findings are in agreement with those of Fritsche et al. (2017), who assert the same for orientation judgments. Those authors propose that working-memory representations are biased toward previous (dimension-specific and task-specific) decisions, a hypothesis that is supported by the potentiation of the bias when several seconds are allowed between stimulus offset and response. A recent study by Bliss et al. (2017) provides converging evidence, reporting that serial dependence is absent at the moment of perception but increases in visual working memory, reaching a maximum when a 6-s delay between stimulus offset and response is imposed (a similar study, however, has reported evidence for serial dependence at the time of perception; Wittmann, Simmons, Aron, & Paulus, 2010). Interestingly, Kanai and Verstraten (2005) have also found a decision-based positive bias on the reported direction of ambiguous motion, appearing only when the stimulus was presented several seconds after the adaptor; they called this effect *perceptual sensitization*.

In our study, participants could respond immediately after the stimulus offset, but due to the relatively long duration of stimulus presentation (500 ms), it is likely that they made an initial decision before that time—as suggested by the results of Experiment 2, wherein the bias exerted by the previous trial was unaffected if a response had not been required (Experiment 2A), but disrupted if a different decision had been indicated by a precue (Experiment 2B). Thus, it is likely that at the moment of response, the representational content produced for the current decision had been already distorted by the memory of previous decisions. More broadly, our results suggest that memory representations of not only the current but also previous perceptual decisions may be subject to similar limitations related to time and informational capacity. While the specific mnemonic processes involved are unclear—our methodology was not designed to operationalize specific instances of memory (such as working memory)—Kiyonaga, Scimeca, Bliss, and Whitney (2017) have noted, in line with our observations, the similarities between serial-dependence effects and well-studied disruptions related to working-memory limitations (such as proactive interference). They suggest that the latter might be a maladaptive aspect of a generally beneficial and widespread brain mechanism for stabilizing internal representations at different levels of processing, including perception, attention, and memory.

Our conclusion of a high-level mechanism of variance processing is also in line with the conclusions of Payzan-LeNestour et al. (2016) regarding variance-driven adaptation aftereffects, which suggest that variance is an abstract property that works independently from its sensory origin and generalizes across domains. Michael et al. (2014) have also proposed variance as an independent property from ensemble average, but suggest that, regarding priming, it operates through feature-specific channels. In our experiments we used a single formalization of variance—dispersion of a dot-motion cloud—so the degree

to which our results will generalize to other variance-related serial dependences requires further investigation.

What are the perceptual or neural mechanisms underlying the observed positive serial dependence? Although this is still uncertain, previous works have proposed exposure-related gain changes or shifts in the neural tuning (Fischer & Whitney, 2014). Furthermore, its behavior resembles that implied by Bayesian frameworks of information processing, in which judgments about a certain dimension are attracted toward prior information. Several studies have recognized that the observed systematic errors in magnitude-estimation tasks, across diverse dimensions, can be well accounted for by assuming an iteratively updated prior, in which recent information is given more weight compared to the overall statistical properties of the environment (Cicchini et al., 2014; Luca & Rhodes, 2016; Petzschner & Glasauer, 2011; Roach, McGraw, Whitaker, & Heron, 2017). Variance-related positive serial dependence indeed shares many characteristics with recursive Bayesian dynamics, including the greater weight of more recent information and the association with high confidence in past trials. Positive serial dependence is probably Bayesian-like in many aspects, but there are some nuances to perceptual decision making that demand further investigation.

The basis of the longer-lasting negative bias is less conclusive, but may be related to adaptation aftereffects, like the variance adaptation described by Payzan-LeNestour et al. (2016). The facts that the negative effect is observed in relation with individual presentations lasting only 500 ms, appears as early as the following trial, and remains even for Trial $n - 9$ could seem unusual for a sensory aftereffect. However, negative aftereffects in response to subsecond stimuli have been described previously (Fritsche et al., 2017; Kanai & Verstraten, 2005), and sometimes lasting for several seconds (Fritsche et al., 2017). Fritsche et al. (2017) have proposed that it is not the stimulus itself but a memory trace that causes the negative aftereffect on orientation. It is likely that the observed relationship between the current trial and a specific trial in history (e.g., $n - 5$) is actually driven by a broader, averaged contextual representation and not by the individual stimuli several trials removed from the present. In our case, as we dealt with a more abstract dimension, we might not consider this high-level aftereffect strictly sensory in the first place (Storrs, 2015). As stated previously, some aspects of this negative bias could point to a decisional component, including its independence of retinal location, predominance in low-confidence trials, and seemingly smaller size when a different decision was required in the past (DIR trials in Experiment 2B; note, however, that the interaction with trial type was not significant). In any case, the line between perceptual and postperceptual aftereffects may be blurred with respect to statistical properties (Payzan-LeNestour et al., 2016; Storrs, 2015).

Some previous studies on different features—both low level (namely motion, Kanai & Verstraten, 2005; and orientation, Fritsche et al., 2017) and high level (such as face attributes; Taubert, Alais, & Burr, 2016)—have reported concomitant positive and negative biases exerted by the same stimulus. Kanai and Verstraten (2005) elicited a negative rapid motion aftereffect of sensory origin by a short, subsecond sine-wave luminance grating presented immediately before. However, when the interstimulus interval was long enough (>3 s), a positive bias was elicited instead, in response to the percept and not the low-level sensory signal (as proven by the use of ambiguous motion adaptors). Fritsche et al. (2017) found opposite effects of recent history on orientation judgments exerted by perception (negative bias) and decision (positive serial dependence), very much in line with our findings. They have proposed that each of these effects has a different biological function, namely increasing sensitivity to changes within the current sensory context and promoting perceptual stability. Taubert et al. (2016) have suggested the same duality in their study of serial dependences in face attributes, although in their case positive and negative biases are exerted concomitantly by different high-level features of the same visual stimulus (faces): Stable traits such as gender would be subject to positive biases in order to smooth away noise, whereas negative aftereffects maximizing sensitivity would predominate in changeable attributes such as facial expresion.

In summary, our study on visual variance reveals two opposite intertrial dependences that operate at different timescales and likely arise at different levels of perceptual decision making: a positive serial dependence in relation to high-confidence, dimension-specific decisions, and a longer lasting negative bias of likely sensory origin. Further investigations are needed to elucidate the precise mechanistic basis of variance-related serial dependence, whether it generalizes to other instances of variance, its relationship to other instances of serial dependence, and the extent to which its properties can be modeled within an iterative Bayesian framework.

Keywords: serial dependence, visual variance, ensemble processing, adaptation aftereffects

Acknowledgments

All authors are grateful to the Dr. Mortimer and Theresa Sackler Foundation, which supports the Sackler Center for Consciousness Science and a PhD

scholarship for MS-P. MS-P is additionally supported by the School of Engineering and Informatics at the University of Sussex, AKS by the Canadian Institute for Advanced Research (Azrieli Programme in Mind, Brain, and Consciousness) and the Wellcome Trust (Engagement Fellowship), and AKS and WR by ERC H2020 project TIMESTORM. MS-P, AKS, and WR designed the research; MS-P performed the research and analyzed the data; and MS-P, AKS, and WR wrote the manuscript.

Commercial relationships: none.
Corresponding author: Marta Suárez-Pinilla.
Email: M.Suarez-Pinilla@sussex.ac.uk.
Address: Sackler Center for Consciousness Science and Department of Informatics, University of Sussex, Brighton, UK.

References

Alais, D., Ho, T., & Han, S. (2017). A matched comparison across three different sensory pairs of cross-modal temporal recalibration from sustained and transient adaptation. *i-Perception*, *8*(4), 1–18, https://doi.org/10.1177/2041669517718697.

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science*, *21*(4), 560–567, https://doi.org/10.1177/0956797610363543.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131, https://doi.org/10.1016/j.tics.2011.01.003.

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences, USA*, *106*(18), 7345–7350, https://doi.org/10.1073/pnas.0808981106.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12):13, 1–8, https://doi.org/10.1167/9.12.13. [PubMed] [Article]

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.

Brady, T. F., & Alvarez, G. A. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision*, *15*(15):6, 1–24, https://doi.org/10.1167/15.15.6. [PubMed] [Article]

Bronfman, Z. Z., Brezis, N., Jacobson, H., & Usher, M. (2014). We see more than we can report: "Cost free" color phenomenality outside focal attention. *Psychological Science*, *25*(7), 1394–1403, https://doi.org/10.1177/0956797614532656.

Campbell, F. W., & Maffei, L. (1971). The tilt after-effect: A fresh look. *Vision Research*, *11*, 833–840.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.

Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*, 891–900.

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

Cicchini, G. M., Mikellidou, K., & Burr, D. (2017). Serial dependencies act directly on perception. *Journal of Vision*, *17*(14):6, 1–9, https://doi.org/10.1167/17.14.6. [PubMed] [Article]

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica*, *138*(2), 289–301.

Corbett, J. E., Wurnitsch, N., Schwartz, A., & Whitney, D. (2012). An aftereffect of adaptation to mean size. *Visual Cognition*, *20*(2), 211–231, https://doi.org/10.1080/13506285.2012.657261.

Dahmen, J. C., Keating, P., Nodal, F. R., Schulz, A. L., & King, A. J. (2010). Adaptation to stimulus statistics in the perception and neural representation of auditory space. *Neuron*, *66*(6), 937–948, https://doi.org/10.1016/j.neuron.2010.05.018.

de Gardelle, V., & Mammasian, P. (2015). Weighting mean and variability during confidence judgments. *PLoS One*, *10*(3), e0120870, https://doi.org/10.1371/journal.pone.0120870.

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences, USA*, *108*(32), 13341–13346, https://doi.org/10.1073/pnas.1104517108.

Fairhall, A. L., Lewen, G. D., Blalek, W., & de Ruyter van Steveninck, R. R. (2001, August 23). Efficiency

and ambiguity in an adaptive neural code. *Nature*, *412*(6849), 787–792.

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

Fouriezos, G., Rubenfeld, S., & Capstick, G. (2008). Visual statistical decisions. *Perception & Psychophysics*, *70*(3), 456–464.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201, https://doi.org/10.1038/nn.2889.

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite effects of recent history on perception and decision. *Current Biology*, *27*, 1–6, https://doi.org/10.1016/j.cub.2017.01.006.

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, *59*, 167–192, https://doi.org/10.1146/annurev.psych.58.110405.085632.

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision*, *15*(4):16, 1–11, https://doi.org/10.1167/15.4.16. [PubMed] [Article]

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble procesing for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–734, https://doi.org/10.1037/a0013899.

Im, H. Y., & Chong, S. C. (2014). Mean size as a unit of visual working memory. *Perception*, *43*(7), 663–676, https://doi.org/10.1068/p7719.

John-Saaltink, E. S., Kok, P., Lau, H., & de Lange, F. P. (2016). Serial dependence in perceptual decisions is reflected in activity patterns in primary visual cortex. *The Journal of Neuroscience*, *36*(23), 6186–6192.

Kanai, R., & Verstraten, F. A. J. (2005). Perceptual manifestations of fast neural plasticity: Motion priming, rapid motion aftereffect and perceptual sensitization. *Vision Research*, *45*(2005), 3109–3116, https://doi.org/10.1016/j.visres.2005.05.014.

Kiyonaga, A., Scimeca, J. M., Bliss, D. P., & Whitney, D. (2017). Serial dependence across perception, attention, and memory. *Trends in Cognitive Sciences*, *21*(7), 493–497.

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

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.

Luca, M. D., & Rhodes, D. (2016). Optimal perceived timing: Integrating sensory information with dynamically updated expectations. *Scientific Reports*, *6*: 28563, https://doi.org/10.1038/srep28563.

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.

Mather, G., Verstraten, F., & Anstis, S. (1998). *The motion aftereffect: A modern perspective*. Cambridge, MA: MIT Press.

Maule, J., & Franklin, A. (2015). Effects of ensemble complexity and perceptual similarity on rapid averaging of hue. *Journal of Vision*, *15*(4):6, 1–18, https://doi.org/10.1167/15.4.6. [PubMed] [Article]

Maule, J., Witzel, C., & Franklin, A. (2014). Getting the gist of multiple hues: Metric and categorical effects on ensemble perception of hue. *Journal of the Optical Society of America A: Optics, Image Science and Vision*, *31*(4), A93–A102, https://doi.org/10.1364/JOSAA.31.000A93.

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian probability: From neural origins to behavior. *Neuron*, *88*, 78–92, https://doi.org/10.1016/j.neuron.2015.09.039.

Michael, E., de Gardelle, V., & Summerfield, C. (2014). Priming by the variability of visual information. *Proceedings of the National Academy of Sciences, USA*, *111*(21), 7873–7878, https://doi.org/10.1073/pnas.1308674111.

Morgan, M. J., Chubb, C., & Solomon, J. A. (2014). A 'dipper' function for texture discrimination based on orientation variance. *Journal of Vision*, *8*(11):9, 1–8, https://doi.org/10.1167/8.11.9. [PubMed] [Article]

Morgan, M. J., Mareschal, I., Chubb, C., & Solomon, J. A. (2012). Perceived pattern regularity computed as a summary statistic: Implications for camouflage. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *279*, 2754–2760, https://doi.org/10.1098/rspb.2011.2645.

Payzan-LeNestour, E., Balleine, B. W., Berrada, T., & Pearson, J. (2016). Variance after-effects distort risk perception in humans. *Current Biology*, *26*, 1–5, https://doi.org/10.1016/j.cub.2016.04.023.

Petzschner, F. H., & Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: A study on human path integration. *The Journal of Neuroscience*, *31*(47), 17220–17229.

Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., & Lau, H. (2015). Confidence leak in perceptual decision making. *Psychological Science*, *26*(11), 1664–1680, https://doi.org/10.1177/0956797615595037.

Roach, N. W., McGraw, P. V., Whitaker, D. J., & Heron, J. (2017). Generalization of prior information for rapid Bayesian time estimation. *Proceedings of the National Academy of Sciences, USA*, *114*(2), 412–417, https://doi.org/10.1073/pnas.1610706114.

Roseboom, W. (2017). Serial dependence across multisensory relative timing tasks. *PsyArxiv*, https://doi.org/10.17605/OSF.IO/6BKDA.

Roseboom, W., Linares, D., & Nishida, S. (2015). Sensory adaptation for timing perception. *Proceedings of the Royal Society B*, *282*, 20142833, https://doi.org/10.1098/rspb.2014.2833.

Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, *12*(4):14, 1–17, https://doi.org/10.1167/12.4.14. [PubMed] [Article]

Sato, Y., & Kording, K. P. (2014). How much to trust the senses: Likelihood learning. *Journal of Vision*, *14*(13):13, 1–13, https://doi.org/10.1167/14.13.13. [PubMed] [Article]

Spence, M., Dux, P., & Arnold, D. (2016). Computations underlying confidence in visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(5), 671–682, https://doi.org/10.1037/xhp0000179.

Storrs, K. R. (2015). Are high-level aftereffects perceptual? *Frontiers in Psychology*, *6*:157, 151–154, https://doi.org/10.3389/fpsyg.2015.00157.

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision-making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*, 745–756, https://doi.org/10.1038/nrn3838.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409, https://doi.org/10.1016/j.tics.2009.06.003.

Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention: Ensemble perception of a crowd's gaze. *Psychological Science*, *25*(10), 1903–1913, https://doi.org/10.1177/0956797614544510.

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

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76, https://doi.org/10.3758/s13423-017-1323-7.

Wittmann, M., Simmons, A. N., Aron, J. L., & Paulus, M. P. (2010). Accumulation of neural activity in the posterior insula encodes the passage of time. *Neuropsychologia*, *48*(10), 3110–3120, https://doi.org/10.1016/j.neuropsychologia.2010.06.023.

Wolfe, B. A., Kosovicheva, A. A., Leib, A. Y., Wood, K., & Whitney, D. (2015). Foveal input is not required for perception of crowd facial expression. *Journal of Vision*, *15*(4):11, 1–13, https://doi.org/10.1167/15.4.11. [PubMed] [Article]

Xia, Y., Liberman, A., Leib, A. Y., & Whitney, D. (2015). Serial dependence in the perception of attractiveness. *Journal of Vision*, *15*(12): 1219, https://doi.org/10.1167/15.12.1219. [Abstract]

Ziemba, C., & Simoncelli, E. (2015). Opposing effects of summary statistics on peripheral discrimination. *Journal of Vision*, *15*(12): 770, https://doi.org/10.1167/15.12.770. [Abstract]

Zylberberg, A., Roelfsema, P. R., & Sigman, M. (2014). Variance misperception explains illusions of confidence in simple perceptual decisions. *Consciousness & Cognition*, *27*, 246–253, https://doi.org/10.1016/j.concog.2014.05.012.