

# **SERIAL DEPENDENCE IN VISUAL VARIANCE**

Marta Suárez-Pinilla<sup>1,2</sup>, Anil K. Seth<sup>1,2</sup> and Warrick Roseboom<sup>1,2</sup>

1. Sackler Centre for Consciousness Science, University of Sussex, Brighton, BN1 9QJ, United Kingdom.
2. Department of Informatics, University of Sussex, Brighton, BN1 9QJ, United Kingdom.

Email: [M.Suarez-Pinilla@sussex.ac.uk](mailto:M.Suarez-Pinilla@sussex.ac.uk)

## ABSTRACT

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 dependencies 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 after-effect - 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 multi-level mechanisms balancing perceptual stability and sensitivity, as with many different perceptual dimensions.

**Keywords:** serial dependence, visual variance, ensemble processing, adaptation after-effects

## 1. 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 (1-6). Such information can be used to efficiently encode (7, 8) and interpret subsequent sensory inputs, and to make predictions about future events (9, 10).

Many forms of visual input can be summarised 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 (11-17), while variance computations have received less attention. However, variance is known to play a crucial role in visual experience, modulating perceptual grouping, ensemble averaging (18-23), texture processing (24, 25) and comparison between arrays (26). Variance is also critical to perceptual prediction, since it provides a measure of the expected range of stimuli (10) as well as the precision (reliability) of the sensory input (27-29). As an indication of sensory reliability, variance also affects metacognitive judgments, although evidence is conflicting regarding the extent and direction of this effect (19, 20, 30).

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 whether it operates as a specific trait of the ensemble or the 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 colour diversity (31) suggested a rapid, pre-attentive 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 (32); 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 (24). By contrast, and surprisingly given the finding of a rapid, pre-attentive mechanism, the abovementioned study on colour diversity reported that variance processing required a conscious representation of the individual elements (31); in 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 (33). Finally, regarding the question about whether variance, once computed over a certain feature-dimension of a visual ensemble, retains its specificity or rather emerges as an abstract property, the previously reported study on priming suggested that the effect of (implicit) variance on mean perception was feature-specific (32). In contrast, a study on direct variance perception found negative adaptation after-effects which generalized across dimensions of visual variance, suggesting a high-level, rather than sensory origin for this effect, and therefore, that variance operates as an independent cognitive property (34). 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 of the above 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. Most well-known are adaptation after-effects - repulsive (negative) biases in perception exerted after prolonged exposure to a certain stimulus magnitude - that have been described for many low and high-level traits (including variance) (34-37) and which are classically employed as an experimental tool for investigating perceptual mechanisms (38). The second category is observed in relation to shorter presentations, generally consisting of an attractive (positive) perceptual bias toward recent sensory history. These 'serial dependencies' have been found for several low and high-level features (39-43).

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 (39).

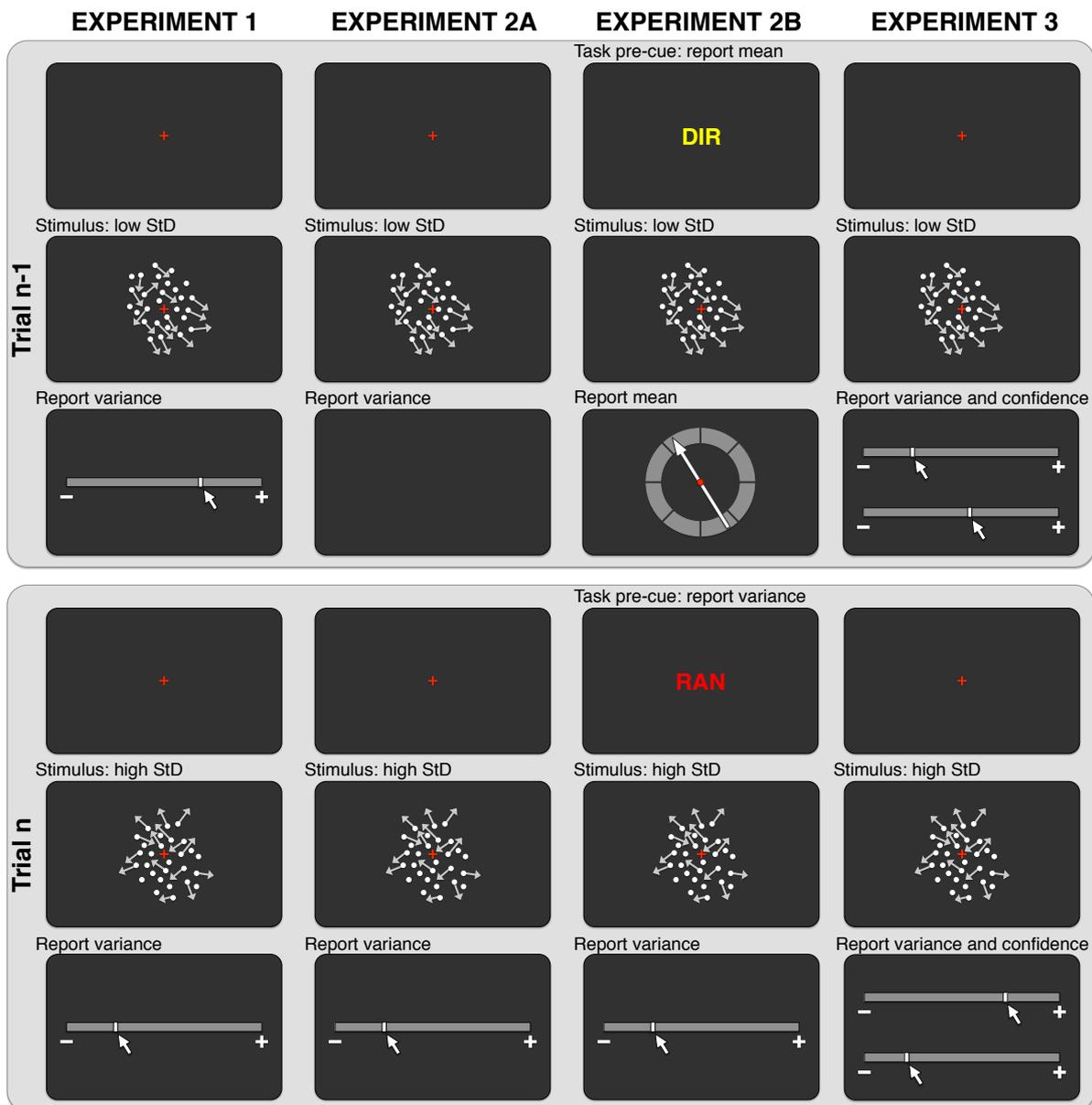
Our study employs serial dependence in variance judgments as a way to track the dynamics and timescales of the processing of this statistic as a distinct perceptual feature. We conducted three experiments:

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 (4, 44-46). Experiment 2 attempted to identify the specific level(s) 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(29).

Overall, our results indicate that judgements 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 judgement made was regarding the same (visual variance) or a different (visual direction) dimension, and on the level of confidence expressed in previous judgements. 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.

## **2. EXPERIMENT 1: SERIAL DEPENDENCES IN VARIANCE JUDGMENTS**

Figure 1 outlines the experimental paradigm utilised 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 characterise the effects of previous history on variance judgments.



**Fig 1: Experiments 1 – 3: structure.** In all experiments, each trial presented an RDK of a certain mean and variance (standard deviation, StD) in the motion trajectories of its component dots. In the example, trial 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 2/3 of trials in which variance reports were required, and 1/3 in which either no response (2A) or mean trajectory estimation (2B) was required. In 2B the trial type was pre-cued, so that the word DIR and 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, by using a similar visual analogue scale.

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.

## **2.1. METHODS**

### **2.1.1. Stimuli**

The stimulus consisted of a cluster of random moving dots (RDK) displayed for 500 milliseconds at a certain eccentricity ( $0^\circ$  or  $20^\circ$ , see below) over a dark grey background ( $3.92 \text{ cd/m}^2$ ). The cluster spanned 5 degrees of visual angle (dva) along the horizontal and vertical dimensions and was comprised of 100 light grey dots (diameter of 0.11 dva, luminance  $43.14 \text{ cd/m}^2$ ) moving along a straight trajectory at a rate of 2 pixels by frame ( $8.45 \text{ dva/s}$ ). The initial position of each dot was uniformly randomized (excluding overlap with other dots) and its 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  $0^\circ$  to  $359^\circ$  and its standard deviation was pseudorandomized among 6 possible values, namely  $5^\circ$ ,  $10^\circ$ ,  $20^\circ$ ,  $30^\circ$ ,  $40^\circ$  and  $60^\circ$ . This parameter (henceforth **StD**) is the dimension of interest in this experiment.

### **2.1.2. 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: (i) familiarizing participants with the scoring process and the scale used in the experiment, and (ii) training maintenance of centered gaze fixation. A broader range of StD values was presented during the practice block compared to the experimental blocks ( $1^\circ$ ,  $10^\circ$ ,  $20^\circ$ ,  $36^\circ$ ,  $60^\circ$  and  $90^\circ$ ). Feedback was provided after each response. Within this block, the first 36 trials were 'foveal' (the stimulus was presented at 0 dva eccentricity) and the remaining 36 were 'peripheral' (20 dva).

Regarding the eight experimental blocks, they employed the narrower set of StD values detailed in the previous 'Stimuli' section ( $5^\circ$ ,  $10^\circ$ ,  $20^\circ$ ,  $30^\circ$ ,  $40^\circ$  and  $60^\circ$ ) and did not have feedback. Half of the eight blocks were 'foveal' (stimulus presentation at 0 dva eccentricity for all trials) and half 'peripheral', with presentation at 20 dva 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 5-point grid, allowing for a maximal average error of 0.5 dva.

At the beginning of each trial a red fixation cross appeared on the centre of the screen, spanning 1.1 dva (horizontally and vertically). Participants were instructed to maintain their gaze on the fixation cross. The RDK stimulus appeared after 1000 ms, and both the stimulus and the fixation cross disappeared simultaneously at 1500 ms from trial onset. Participants were instructed to score the 'randomness' (variance) of the motion of the RDK using a visual analogue scale (see Fig. 1), adjusting the position of a sliding bar with the mouse. 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. For simplicity, the scale was a linear translation of the StD numeric values ranging from

0° (left end) to 90° (right end of the scale). If the participant failed to respond within 5 seconds, the next trial started automatically. The inter-trial interval was randomized between 250-1000ms.

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 dva) occurred between 700 - 1000 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 inter-trial 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 (1000 ms – 1500 ms), and no trial was aborted or restarted after stimulus onset at 1000 ms.

### **2.1.3. 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 their participation. The study was granted ethical approval by the Research Ethics Committee of the University of Sussex.

### **2.1.4. Apparatus**

Experiments were programmed in MATLAB 2012b (MathWorks Inc., Natick, US-MA) with Psychtoolbox 3.0.10 and displayed on a LaCie Electron 22BLUE II 22" with screen resolution of 1024 x 768 pixels and refresh rate of 60 Hz. Eye tracking was performed with Eyelink 1000 Plus (SR Research, Mississauga, Ontario, Canada) at a sampling rate of 1000 Hz, using a level desktop camera mount. Head position was stabilized at 43 cm from the screen with a chin and forehead rest.

### **2.1.5. Statistical analysis**

Statistical analyses (detailed in the Results section) were performed on Matlab 2016a (MathWorks Inc., Natick, US-MA), R 3.4.2 (The R Foundation for Statistical Computing, <http://www.R-project.org>) and JASP (JASP Team (2017). JASP (Version 0.8.3.1, Mac OS X – El Capitan (10.11)).

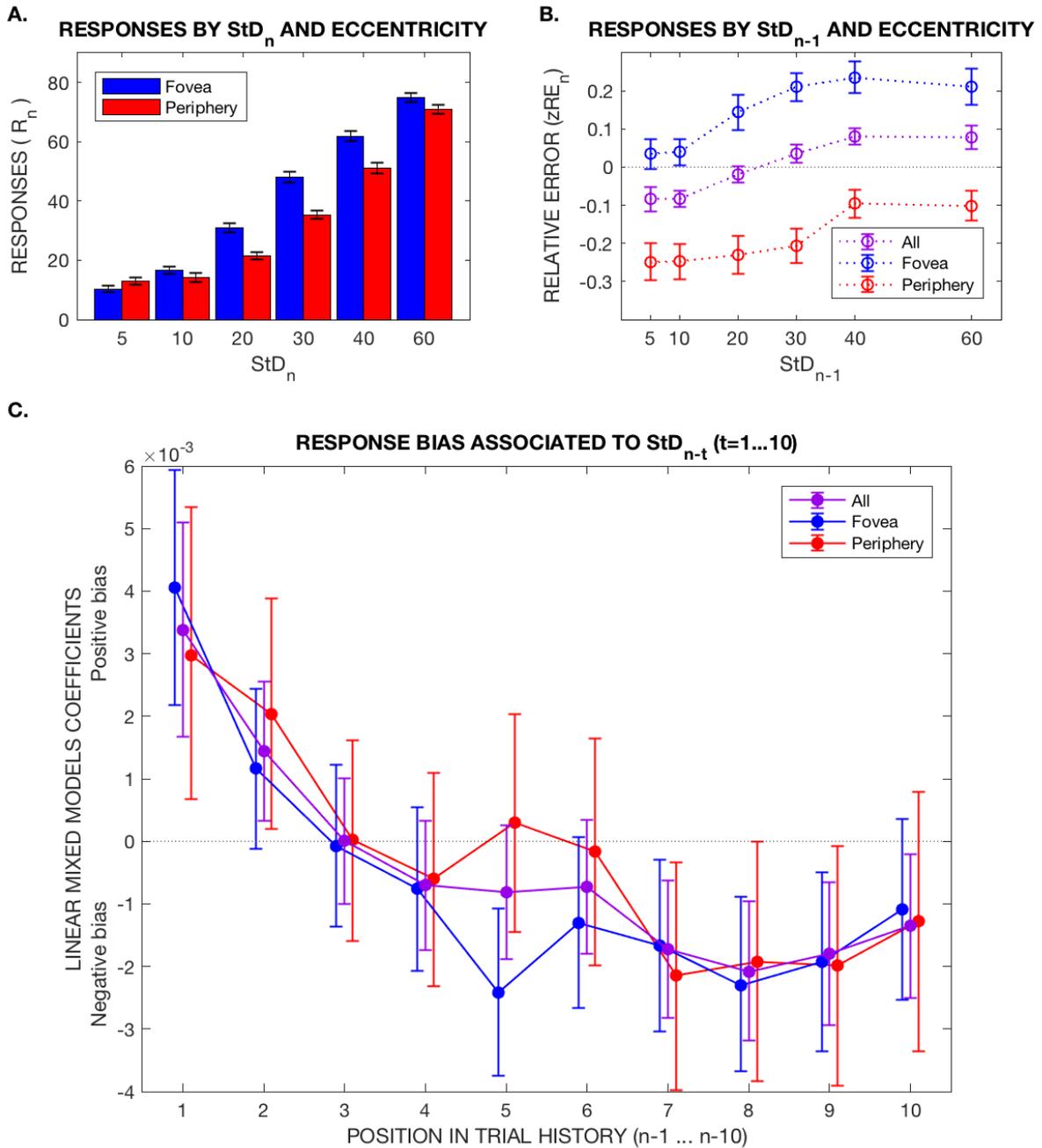
## **2.2. *RESULTS***

Thirty participants (25 female, mean age 19.0 yr., standard deviation 1.35) participated in this experiment. Except for two members of the laboratory, the rest were first-year Psychology students who volunteered for course credits.

To ensure the validity of foveal and peripheral conditions, trials without centred gaze fixation during stimulus presentation were removed from the analysis: a trial was deemed valid if the participant maintained fixation within 5 dva from the centre of the screen for over 80% of the stimulus presentation period (1000-1500 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 12480 trials entered the analysis.

### **2.2.1. 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.



**Fig. 2. Experiment 1. 2a.** Distribution of responses by  $StD_n$  and eccentricity. The height of the bars represents the mean and the error bars the between-subject standard error. **2b.** 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 i.e., 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-subject 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  $StD_{n-1}$ . **2c.** 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 (LMM) for the association between the  $StD$  presented in trials  $n-1$  to  $n-10$  ( $StD_{n-t}$ ,  $t=1...10$ ) and the normalized response error in the current trial.

To examine the general pattern of variance judgments, we conducted a repeated-measures 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 method of Greenhouse-Geisser). For  $StD_n$ , the main effect yielded  $F(1.825, 45.621)=473.80$ ,  $p<0.001$ ,  $\eta^2_p=0.950$ , in relation with higher reports for larger stimulus StD. For eccentricity the main effect was  $F(1, 25)=33.32$ ,  $p<0.001$ ,  $\eta^2_p=0.571$ : peripheral presentation was associated with lower variance reports, with a mean difference of 6.798 (fovea - periphery),  $t(25)=8.237$ ,  $p<0.001$ , Cohen's  $d=1.615$ . The interaction effect  $StD_n$ \*eccentricity was also significant,  $F(2.715, 67.882)=20.06$ ,  $p<0.001$ ,  $\eta^2_p=0.445$ , 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 RM 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  $BF_{full/main\ effects}=1.075*10^6$ . 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 ( $\sigma_R$ ), 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 a  $F(2.994, 74.840)=58.426$ ,  $p<0.001$ ,  $\eta^2_p=0.700$ , in relation with greater response dispersion for large StD levels, as is common in magnitude estimation tasks:  $\sigma_R$  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  $F(1, 25)=4.165$ ,  $p=0.052$ ,  $\eta^2_p=0.143$ , suggesting a trend towards greater response dispersion in peripheral presentations: mean difference -0.658 (fovea - periphery),  $t(25)=-1.738$ ,  $p=0.086$ , Cohen's  $d=-0.339$ . Last, the effect of the interaction term  $StD_n$ \* eccentricity was  $F(3.530, 88.244)=4.757$ ,  $p=0.002$ ,  $\eta^2_p=0.160$ , due to the larger response dispersion in periphery occurring mainly for large StD values. In a Bayesian RM 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 towards greater response dispersion for peripheral presentations, especially at large StD, suggesting a slightly worse performance at 20 dva eccentricity compared to 0 dva, in agreement with the previous finding of a greater regression to the mean in peripheral responses.

### **2.2.2. 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 backwards 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, so that their absolute value is independent of the current StD<sub>n</sub> level and of each participant's global scoring biases.

Figure 2b presents the average  $zRE_n$  as a function of the previous stimulus (StD <sub>$n-1$</sub> ), plotted separately by eccentricity. Regardless of generally lower reports at larger eccentricity, a trend towards larger  $zRE_n$  for higher StD <sub>$n-1$</sub>  values is evident for all trials pooled as well as for both foveal and peripheral presentations. 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 <sub>$n-1$</sub>  level (as within-subject factor) on current variance reports ( $zRE_n$ ). The effect of StD <sub>$n-1$</sub>  was statistically significant ( $F(3.231,93.697)=7.221$ ,  $p<0.001$ ,  $\eta^2_p=0.199$ , Greenhouse-Geisser correction applied). The Bayes factor for the inclusion of StD <sub>$n-1$</sub>  compared to the null model (both of them included participant as grouping variable) was  $BF_{inclusion}=56187.91$ , indicating *extreme* (47) evidence for a 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 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 (39). Moreover, if the function of (positive) serial dependence is to promote perceptual continuity (39), 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 <sub>$n-1$</sub>  and each of the features of interest, separately: eccentricity, retinal location and similarity of means.

For eccentricity, both main effects were statistically significant ( $F_{eccentricity}(1,25)=31.004$ ,  $p<0.001$ ,  $\eta^2_p_{eccentricity}=0.554$ ;  $F_{StD_{n-1}}(2.662, 66.556)=7.029$ ,  $p<0.001$ ,  $\eta^2_p_{StD_{n-1}}=0.219$ ; sphericity, correction by Greenhouse-Geisser), 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 <sub>$n-1$</sub> . To formally test this hypothesis, we turned to Bayesian repeated-measures ANOVA. Table 1a summarizes the comparison between all competing models. The largest Bayes

factor corresponds to the model including both main effects but not the interaction ( $BF_{10}=3.432*10^{29}$ ), which outperforms the model that also includes the interaction term  $StD_{n-1}*eccentricity$  by a factor of  $BF_{main\ effects/full}= 17.645$  – strong (47) evidence against its inclusion and supporting 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 analysed 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: i.e. same presentation location versus a separation of 40 dva 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 interaction term  $StD_{n-1}* hemifield$  ( $BF_{10}=0.120$ ). 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, i.e. the one with  $StD_{n-1}$  only ( $BF_{full/StD_{n-1}}=0.058$ ). These results support the hypothesis of serial dependence being unaffected by the spatial location of consecutive stimuli.

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 5 categories:  $\leq 36^\circ$ ,  $37^\circ-72^\circ$ ,  $73^\circ-108^\circ$ ,  $109^\circ-144^\circ$ ,  $145^\circ-180^\circ$ . 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  $StD_{n-1}$  ( $BF_{10}=3.210*10^5$ ), whereas the model including both main effects and its interaction was the second worst after the one with mean difference only, with a  $BF_{10}=0.281$ . The Bayes factor for inclusion of the interaction term indicated extreme evidence against it ( $BF_{inclusion}=3.491*10^{-6}$ ); this was also the case if the comparison was made between the full model and the model lacking only the interaction ( $BF_{main\ effects/full} = 1789.55$ ). This lack of association of mean similarity and serial dependence in variance was further confirmed in a further experiment (detailed in Supplementary Materials, section 1) that used a limited range of mean trajectories, allowing for only four between-trial differences ( $0^\circ$ ,  $35^\circ$ ,  $55^\circ$  and  $90^\circ$ ).

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.

**TABLE 1. Serial dependence and stimulus properties - Model comparison**

**1a. Eccentricity**

Models	P(M)	P(M data)	$BF_M$	$BF_{10}$	error %
Null model (incl. subject)	0.200	2.747e -30	1.099e -29	1.000	

### 1a. Eccentricity

Models	P(M)	P(M data)	BF <sub>M</sub>	BF <sub>10</sub>	error %
StDn-1	0.200	1.229e -29	4.918e -29	4.476	0.626
Eccentricity	0.200	0.004	0.015	1.385e +27	1.303
StDn-1 + Eccentricity	0.200	0.943	65.900	3.432e +29	3.951
StDn-1 + Eccentricity + StDn-1 * Eccentricity	0.200	0.053	0.226	1.945e +28	2.392

### 1b. Spatial location (hemifield)

Models	P(M)	P(M data)	BF <sub>M</sub>	BF <sub>10</sub>	error %
Null model (incl. subject)	0.200	0.232	1.208	1.000	
StDn-1	0.200	0.481	3.702	2.073	0.464
Hemifield	0.200	0.079	0.343	0.340	1.690
StDn-1 + Hemifield	0.200	0.181	0.883	0.780	3.545
StDn-1 + Hemifield + StDn-1 * Hemifield	0.200	0.028	0.114	0.120	1.387

### 1c. Mean difference (n-1, n)

Models	P(M)	P(M data)	BF <sub>M</sub>	BF <sub>10</sub>	error %
Null model (incl. subject)	0.200	3.111e -6	1.244e -5	1.000	
StDn-1	0.200	0.998	2550.135	320978.693	0.570
Mean difference	0.200	4.784e -9	1.914e -8	0.002	0.726
StDn-1 + Mean difference	0.200	0.002	0.006	502.187	0.815
StDn-1 + Mean difference + StDn-1 * Mean difference	0.200	8.729e -7	3.491e -6	0.281	1.091

Note. All models include subject.

**Table 1. Experiment 1.** Serial dependence (associated with trial n-1) and stimulus properties. Each table 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 than previous stimulus) and difference in the mean trajectories of the RDKs presented in consecutive trials. 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).

### Positive serial dependence in variance extends up to the latest 2 trials

Investigations of serial dependence have typically focussed 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 modelling the fixed-effects size of serial dependence while allowing for between-subject variability, building ten varying intercept, varying slope Bayesian linear mixed-effects models (LMM) with  $zRE_n$  as dependent variable, and  $StD_{n-t}$  ( $t=1...10$ , respectively) as 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 a LKJ prior with shape parameter  $\eta = 2.0$  for the random-effect correlation matrices. Unless stated otherwise, analogous priors were established for other Bayesian LMMs reported in this paper. Fixed-effects coefficient estimates were largely insensitive to prior selection, as can be seen in the example presented in the Supplementary Materials (section 2).

We applied these models to foveal and peripheral blocks separately as well as overall dataset. Figure 2c presents the fixed-effect 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 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  $1^\circ$  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 outcome variable (adding the current  $StD_n$  and the interaction with  $StD_{n-1}$  to the models). Here, the increase is of 0.0586 (0.0272 - 0.0892) units per unit of  $StD_{n-1}$ , or an attractive effect of 5.9% towards the previous stimulus, whereas for  $StD_{n-2}$  the effect size is 0.0242 (0.0006 - 0.0483) (2.4%).

Thus, variance judgments at one specific trial ( $n$ ) are attracted a small but meaningful amount towards the variance presented in the previous trial ( $n-1$ ) and to a lesser extent the trial before ( $n-2$ ). 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.

### **2.2.3. Variance reports are subject to a negative bias driven by less recent trial history**

Looking past the previous 2 trials, as shown in figure 2c, a reversal from positive (trial  $n-1$  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 after-effects (34, 38). 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 faded afterwards, largely declining from trial  $n-10$  onwards. Similar effect sizes and timescales are observed for foveal and peripheral presentations (see figure 2).

## **3. 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 dependencies exert their influence? The non-local nature and independence from inter-trial 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 (39, 43, 48). To address this issue, in this experiment we applied several manipulations to the task to disambiguate the contributions of low-level sensory processes, perceptual decisions and responses to serial dependence of variance judgements.

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 pre-cued, meaning that 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 (pre-cued) task switching design to disentangle the contribution of perception and decision processes.

### **3.1. METHODS**

The methods of these experiments were similar to those of Experiment 1 with the following exceptions:

#### **3.1.1. Stimuli**

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

#### **3.1.1. Procedure**

Experiment 2A had a practice block with 72 trials and ten experimental blocks with 60 trials each; same for 2B except that the practice block was longer (90 trials), due to the additional demands of its task-switching design (see below). In Experiment 2B 9 participants (out of 16) performed a session twice as long (10 blocks), due to differential availability of different participants. For both experiments, in each block, 2/3 of the trials required 'randomness' scores as described for Experiment 1. In Experiment 2A, the remaining 1/3 were no-response trials: after stimulus presentation, instead of the response bar only a blank screen appeared for a randomized interval between 1000-3000 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 pre-cued in any way. In Experiment 2B, 1/3 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 pre-cued at the beginning of the trial: one three-letter word, either 'RAN' ('randomness' report required) or 'DIR' (mean direction report required) was displayed for one second 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).

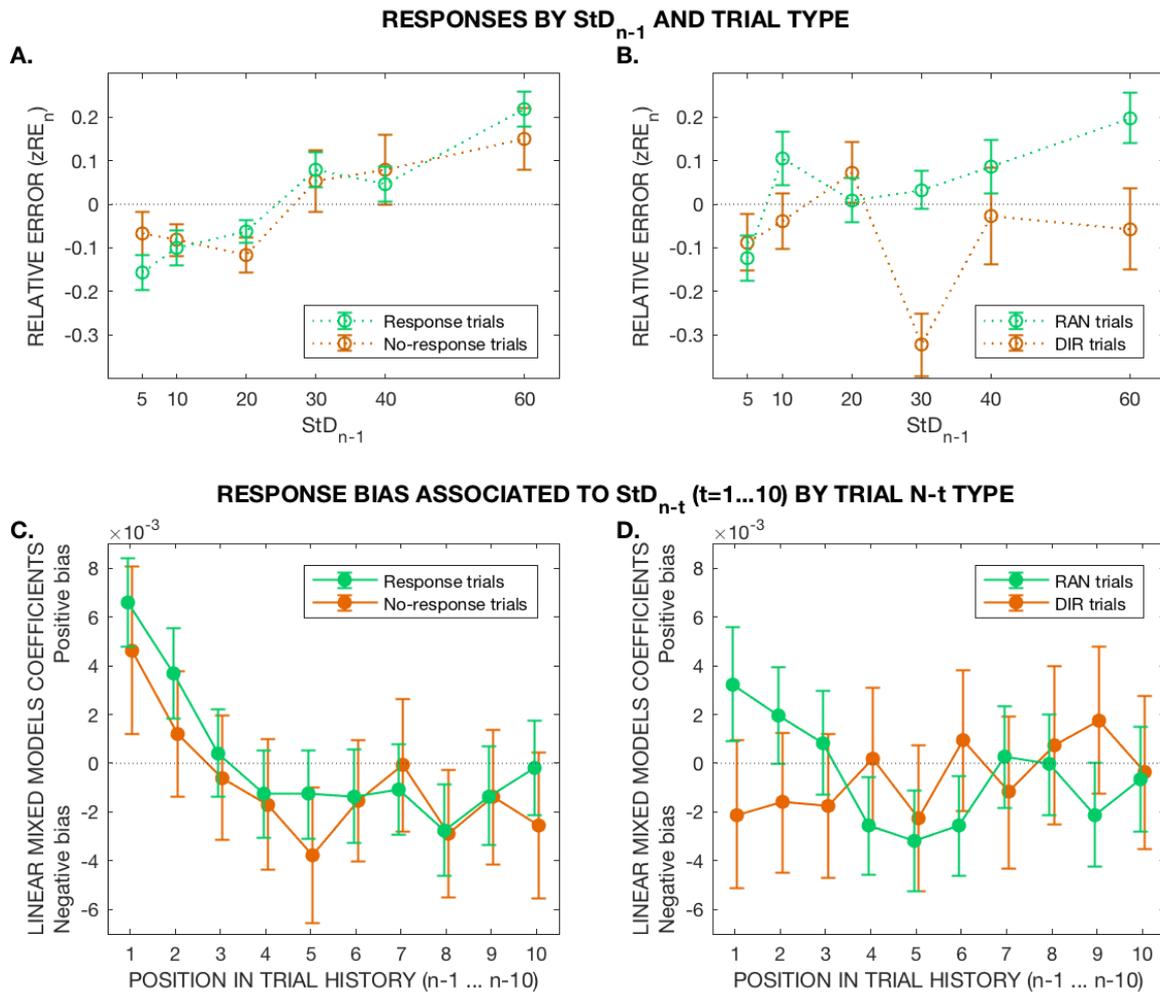
## **3.2. RESULTS**

### **3.2.1. Experiment 2A: effect of response execution on serial dependence in variance reports**

Fifteen Psychology students (13 female, mean age 20.4, 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 9000, out of which 3000 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 StD ( $StD_{n-1}$ ) and type, i.e. whether  $n-1$  had been 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  $StD_{n-1}$  ( $BF_{10}=2.386*10^6$ ). There was strong evidence against the inclusion of the interaction term  $StD_{n-1}*trial\ n-1\ type$ :  $BF_{inclusion}=0.051$ . In a direct comparison between the main-effects model and the full model the ratio was given by  $BF_{main\ effects/full}=10.75$ . This lack of interaction confirmed that  $StD_{n-1}$  effect on current report was independent of response execution.



**Fig 3. Experiment 2. 3a - 3b.** Normalized relative error in current response ( $zRE_n$ ) as a function of the  $StD_{n-1}$ , plotted separately by trial  $n-1$  type: response versus no-response in **3a** (experiment 2A), RAN versus DIR in **3b** (experiment 2B). The error bars represent the between-subject 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 3a), whereas in figure 3b, only RAN trials elicit such positive serial dependence. **3c - 3d.** Fixed-effects coefficient estimates in 20 Bayesian LMMs with  $StD_{n-t}$  ( $t=1...10$ ) as predictor of current response ( $zRE_n$ ), modelled separately by trial  $n-t$  type: in **3c**, response versus no-response trials (experiment 2A); in **3d**, RAN versus DIR trials (experiment 2B). Since the dependent variable is the current variance ('randomness') judgment, trial  $n$  is always a response (3c) or a RAN (3d) trial. The error bars represent the 95% credible intervals for the true value of the coefficient.

Figure 3c shows the fixed-effects coefficient estimates and 95% credible intervals for 20 Bayesian LMMs for  $zRE_n$ , with  $StD_{n-t}$  ( $t=1...10$ ) as putative predictor, split by trial  $n-t$  type and modelled 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 than 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 had also been a response trial) and classified the dataset according to whether the intermediate trial (n-1) had been 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 interaction  $StD_{n-2} * \text{trial n-1 type}$ ), the latter outperformed the former by only a factor of  $BF_{\text{main effects/full}} = 1.98$ . Overall, the Bayes-factor for inclusion of the interaction term indicated moderate evidence against it ( $BF_{\text{inclusion}}=0.261$ ), suggesting that the attractive bias related to previous trials is not disrupted (nor boosted) by the participant providing a response on the intermediate trials.

**TABLE 2. Serial dependence and task requirements - Model comparison**

**2a. Response (Experiment 2A)**

Models	P(M)	P(M data)	BF <sub>M</sub>	BF <sub>10</sub>	error %
Null model (incl. subject)	0.200	3.575e -7	1.430e -6	1.000	
StDn-1	0.200	0.853	23.192	2.386e +6	0.393
Trial n-1 type (response vs. no-response)	0.200	5.551e -8	2.220e -7	0.155	0.772
StDn-1 + Trial n-1 type	0.200	0.135	0.622	376281.756	1.178
StDn-1 + Trial n-1 type + StDn-1 * Trial n-1 type	0.200	0.013	0.051	35151.882	2.835

**2b. Dimension-specific judgment (Experiment 2B)**

Models	P(M)	P(M data)	BF <sub>M</sub>	BF <sub>10</sub>	error %
Null model (incl. subject)	0.200	0.012	0.048	1.000	
StDn-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
StDn-1 + Trial n-1 type	0.200	0.283	1.578	23.921	2.994
StDn-1 + Trial n-1 type + StDn-1 * Trial n-1 type	0.200	0.573	5.371	48.459	2.705

Note. All models include subject.

**Table 2. Experiment 2.** Serial dependence (associated with trial n-1) and trial n-1 type (task requirements in that trial). Each table section presents the model performance on Experiment 2A and 2B datasets, 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. 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).

**2.2. 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 pre-cued 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 pre-cued 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, mean age 21.4 yr, standard deviation 8.8) participated in this experiment in exchange for course credits, under the conditions described above. In total they performed 7200 trials, out of which 2400 were DIR-trials (alternative task).

### **Serial dependence is related to dimension-specific decision-making**

We analyzed the data in 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 -i.e. 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 favour compared to the model with only the main effects was anecdotal ( $BF_{full/main\ effects} = 2.026$ ). However, evidence in favour of the interaction term was larger when taking in consideration all possible models:  $BF_{inclusion}=5.371$  –moderate evidence for inclusion. 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 had been 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 (48-50). 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 observed difference in serial dependence by  $StD_{n-1}$  when n-1 had been a RAN compared to a DIR trial: i) trial n-1 type (RAN/DIR) or ii) inter-stimulus time being the cause of this difference. Thus, we compared the explanatory power of a model with  $StD_{n-1}$ , trial n-1 type (RAN/DIR) 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 favour. Overall, analysis of each separate effect indicated extreme evidence against inclusion of the interaction term  $StD_{n-1} * time_{n-1,n}$  ( $BF_{inclusion}=6.668 * 10^{-4}$ ). 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 inter-trial 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  $StD_{n-t}$  ( $t=1...10$ ) and  $zRE_n$ , after splitting the dataset according to the trial type at each position: thus, the influence of RAN and DIR trials is modelled separately by 20 Bayesian LMMs. As expected from the previous analysis, the positive effect related to  $StD_{n-1}$  is only present 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 after-effect.

If we ask why is there 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  $n-2$  had also been RAN type, and examined the association between  $StD_{n-2}$  and current report in relation to the intermediate trial ( $n-1$ ) task. An explanatory role for the interaction term  $StD_{n-2} * \text{trial } n-1$  response type would indicate that the intermediate trial type influenced  $n-2$ -related serial dependence. 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  $StD_{n-2}$ -related serial dependence.

#### **4. 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 (29).

##### **4.1. METHODS**

#### **4.1.1. Stimuli**

Stimulus presentation was identical to Experiment 1; as in this case, we employed again 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.

#### **4.1.2. 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.

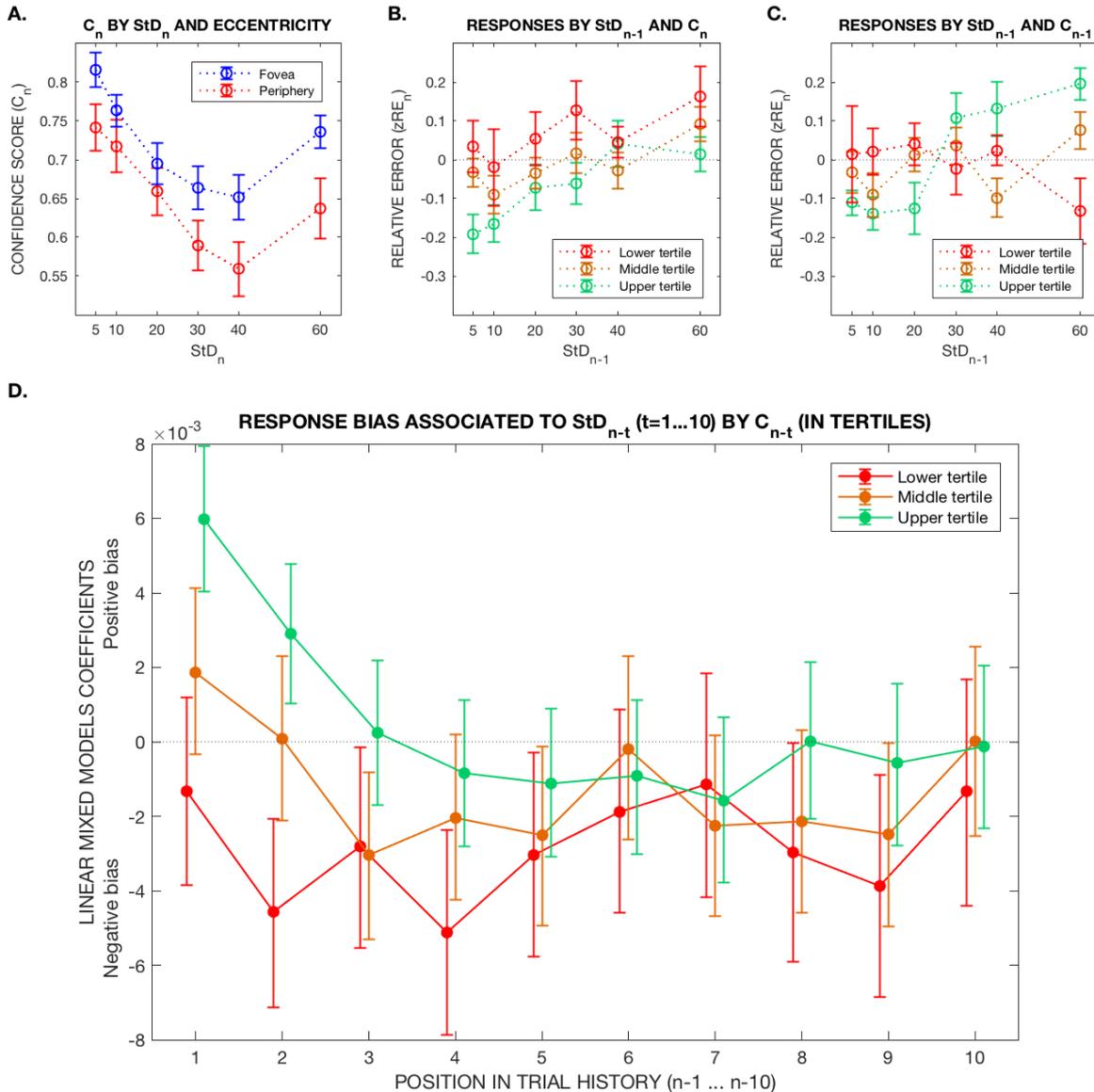
During the response phase of each trial, two identical visual analogue scales were displayed on the screen: the upper one for scoring 'randomness' (variance) and the lower one for confidence (see Figure 1). The initial position of each sliding bar was randomized separately, and the time allowed for responding to both items was 6 seconds. 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.

### **4.2. RESULTS**

Twenty-two participants (17 female, mean age 19.6 yr., 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, 8880 trials were included in the analyses.

#### **4.2.1. 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 towards 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 ( $BF_{10}=6.657*10^{26}$ ), outperforming the full model with the interaction term  $StD_n*eccentricity$  by a factor of  $BF_{main\ effects/full} = 9.615$ . 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.



**Fig 4. Experiment 3.** **4a.** Confidence scores ( $C_n$ ) by  $StD_n$  plotted separately by eccentricity. **4b – 4c.** 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 the current (**4b**) or previous (**4c**) trial. Confidence scores have been binned into tertiles according to each participant’s distribution of reports. The error bars represent the between-subject standard error. The plots in **4b** 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 (**4c**), 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 for the lower-tertile is slightly descending, suggesting a negative bias away from low-confidence  $n-1$  trials. **4d.** Fixed-effects coefficient estimates in 30 Bayesian LMMs with  $StD_{n-t}$  ( $t=1...10$ ) as predictor of current response ( $zRE_n$ ), modelled 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 figure **4c**, 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 backwards in history.

Subsequently we explored whether confidence reports were differentially shaped by

response accuracy or precision – and considered also 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 interaction term  $E_n * StD_n$  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. When 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 independently 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 ( $B_{error * eccentricity} = -0.0013$  ( $-0.0040 - 0.0016$ )).

As for precision, we calculated the standard deviation of each participant’s responses per StD value ( $\sigma_R$ ) as a measure of response dispersion. Subsequently we modelled confidence by  $\sigma_R$ , 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 adding eccentricity to this model, the main effect for  $\sigma_R$  is close in value ( $B = -0.0105$  ( $-0.0162 - -0.0050$ )), whereas the interaction term  $\sigma_R * eccentricity$  ( $B = -0.0003$  ( $-0.0067 - 0.0062$ )) 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 (29).

Interestingly, we observed a very strong serial dependence for confidence reports. Modelling reported confidence (by a Bayesian LMM) as a function of the report provided in the previous trial ( $C_{n-1}$ ), the coefficient for the latter is  $B = 0.1874$  (95% credible interval  $0.1445 - 0.2307$ ), with an  $R^2 = 0.3188$ . Importantly, if we add the error size of the previous trial ( $E_{n-1}$ ) to the model, as well as the interaction  $E_{n-1} * C_{n-1}$ , 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 not only due to accuracy/attention fluctuating at timescales of several trials, nor to the direct influence of the StD in the previous stimulus, but rather may be an expression of response inertia and/or ‘confidence leak’ as described in (51).

#### **4.2.2. 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 (10, 42, 52). Within this framework, confidence is often regarded as a

measure of the precision of the sensory signal (29), 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 only by the most recent trials (see figure 2c as example).

Figures 4b and 4c depict  $zRE_n$  as a function of  $StD_{n-1}$ , plotted separately by current (4b) and previous (4c) trial confidence. Confidence scores have been binned into tertiles on a per-participant basis. In 4c, 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, first we 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 of a Bayesian repeated measures ANOVA are presented in Table 3a. The best model contains both main effects ( $StD_{n-1}$  and  $C_n$ ) but not the interaction ( $BF_{10}=349.668$ ), outperforming the model with the interaction term by a factor of  $BF_{main\ effects/full} = 93.544$ . 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  $C_{n-1}$  (confidence score in the previous trial, by tertiles) as within-subject factors. Table 3b presents the results of this analysis. Evidence is in favour 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 considering the term of interest for our hypothesis, namely the interaction  $StD_{n-1} * C_{n-1}$ , there is a strong evidence in favour of its inclusion compared to the model stripped of that effect (including only the two main factors):  $BF_{full/main\ effects}=26.989$ . Still, because none of both competing models were superior to the null model, this result must be taken with caution.

**TABLE 3. Serial dependence and reported confidence- Model comparison**

**3a. Current trial confidence**

Models	P(M)	P(M data)	BF <sub>M</sub>	BF <sub>10</sub>	error %
Null model (incl. subject)	0.200	0.002	0.010	1.000	
StDn-1	0.200	0.011	0.043	4.437	0.393

### 3a. Current trial confidence

Models	P(M)	P(M data)	BF <sub>M</sub>	BF <sub>10</sub>	error %
Cn	0.200	0.137	0.637	57.150	0.642
StDn-1 + Cn	0.200	0.841	21.088	349.668	1.213
StDn-1 + Cn + StDn-1 * Cn	0.200	0.009	0.036	3.738	5.208

### 3b. Previous trial (n-1) confidence

Models	P(M)	P(M data)	BF <sub>M</sub>	BF <sub>10</sub>	error %
Null model (incl. subject)	0.200	0.923	47.768	1.000	
StDn-1	0.200	0.025	0.102	0.027	0.535
Cn-1	0.200	0.030	0.123	0.032	1.097
StDn-1 + Cn-1	0.200	8.080e-4	0.003	8.757e-4	1.439
StDn-1 + Cn-1 + StDn-1 * Cn-1	0.200	0.022	0.089	0.024	1.099

Note. All models include subject.

**Table 3. Experiment 3.** Serial dependence (associated with trial n-1) and confidence reported in current and previous trial. Bayesian repeated-measures ANOVA on  $zRE_n$ , with two within-subject factors:  $StD_{n-1}$  and current (**3a**) or previous (**3b**) trial confidence. 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<sub>M</sub>: Bayes factor for the model. BF<sub>10</sub>: Bayes factor for the alternative hypothesis (expressed by each model).

We next asked the degree to which confidence in trials located further back in history, up to n-10, influenced serial dependence of variance judgements. We split the dataset 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 position n-t (t=1...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 onwards. The largest negative bias is driven by low-confidence trials, for which it seems to appear as recently as in trial n-1 (although the credible intervals contain zero), becomes unambiguous at n-2 and peaks at trial n-4, decreasing afterwards, in contrast with the slower build-up 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 than the former. Results were similar when considering foveal and peripheral blocks separately.

At first glance, the early appearance of the negative effect (after exposure to a single sub-second 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 presence of the positive bias, the latter being associated with high-confidence decision-making.

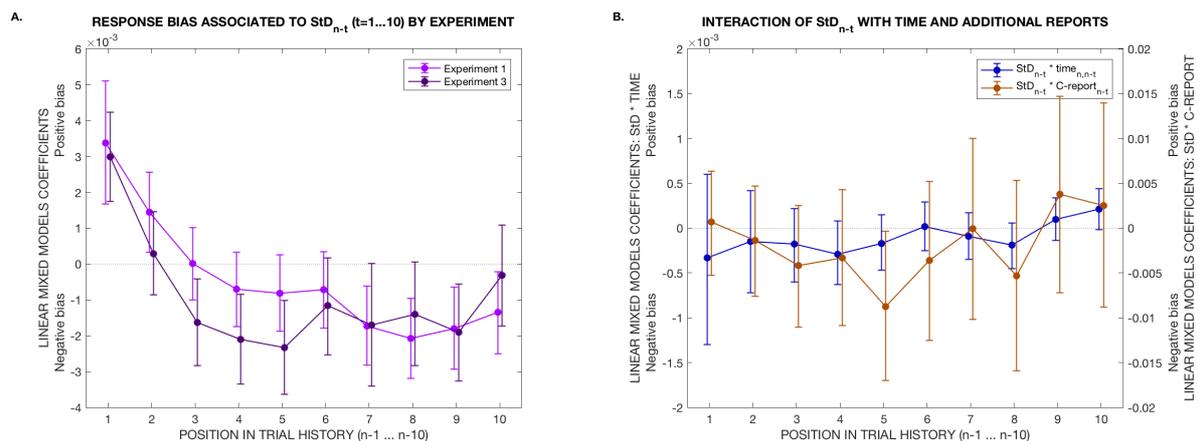
On average, response times for variance reports in low, medium and high confidence trials were 1.59, 1.46 and 1.30 seconds, respectively (Bayesian RM ANOVA:  $BF_{10}=22288$ , extreme evidence for the alternative hypothesis), presumably related to subjective trial difficulty. Therefore, we sought to rule out the possibility that the effect of past confidence on serial dependence was only related to the difference in response times, and consequently in inter-stimulus times. For each trial position 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  $time_{n,n-t}$  (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 interaction term  $StD_{n-t} * time_{n,n-t}$  was 'extremely' low, i.e. the Bayes factor for this specific effect was always below  $1/100$ . This suggested that time was not confounding the reported interaction between confidence and serial dependence.

#### **4.2.3. 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 inter-trial time was longer in Experiment 3 than in Experiment 1 (5.06 versus 3.69 seconds, Bayesian independent samples t-test:  $BF_{10}>6.690*10^7$ ). As previous work has strongly implicated time between successive stimuli or stimuli and response as critical contributors to serial dependence (48-50), we sought to take advantage of this circumstance to enquire (post-hoc) about the factors that drive the decrease and eventual shift towards negative of the serial dependence effect as we move backwards in trial history.

Figure 5a presents the Bayesian LMM coefficients and 95% credible intervals for the effect of  $StD_{n-t}$  ( $t=1...10$ ) in current variance report as found for Experiments 1 and 3. 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 for Experiment 3 ( $B=0.0003$  (-0.0009 - 0.0015)). Thus, in Experiment 3 the reversal to negative bias occurs as early as in 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)). This earlier build-up of the negative bias could be related to the longer inter-stimulus intervals in the present experiment: time might

drive, hypothetically, the reversal to repulsive serial effects and its posterior fading. Results of Experiment 2B (concerning 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 inter-stimulus times observed in the present experiment. Another explanation for the earlier shift towards 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.



**Fig 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 inter-stimulus time longer, on average, for Experiment 3 compared to Experiment 1. **5a.** Fixed-effects coefficient estimates in 20 Bayesian LMMs with  $StD_{n-t}$  ( $t=1...10$ ) as predictor of current response ( $zRE_n$ ), with the data of Experiment 1 and 3 modelled 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. **5b.** Fixed-effects coefficient estimates for the interaction terms  $StD_{n-t} * time_{n,n-t}$  and  $StD_{n-t} * C-report_{n-t}$  in 10 Bayesian LMMs for prediction of  $zRE_n$ , with  $StD_{n-t}$ ,  $time_{n,n-t}$ ,  $C-report_{n-t}$  and all interactions as putative predictors. The variable  $time_{n,n-t}$  reflects the time between onsets of the stimuli in trials  $n-t$  ( $t=1...10$ ) and  $n$ .  $C-report_{n-t}$  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 1.

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  $C-report_{n-t}$ , 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 had 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 ten Bayesian LMM 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  $C-report_{n-t}$ , and all interactions. The fixed-effects B coefficients of the interaction terms  $StD_{n-t} * time_{n,n-t}$  and  $StD_{n-t} * C-report_{n-t}$  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  $\text{StD}_{n-t} * \text{C-report}_{n-t}$  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 for the different patterns of serial dependence in Experiments 1 and 3, the ‘mostly-negative’ interaction terms  $\text{StD}_{n-t} * \text{time}_{n,n-t}$  and  $\text{StD}_{n-t} * \text{C-report}_{n-t}$  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 to 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 judgements in shaping the effects of trial history.

## 5. DISCUSSION

The examination of serial dependence provides a valuable window on perceptual processing. In a series of studies, 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 1-2 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 which generally becomes manifest several trials into history and remains constant for up to ten trials.

Several studies on serial dependence have found a positive (attractive) bias towards recent perceptual history, which is modulated by attention, 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 seconds but does not require explicit memory (39). While control experiments support that this effect does not require a motor response, there is an ongoing debate about whether its basis is perceptual or post-perceptual: the results of a two-alternative forced-choice discrimination task (39) (with a sample size of three participants), a recent behavioural study (53) and a V1-based fMRI study (43) 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 post-perceptual (decisional) source (48). 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 (40, 41), relative timing (54) and numerosity (42).

In our experiments on visual variance (a high-order visual statistic), we found a positive bias that shares many of the characteristics listed above, 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 a study with face perception (40) and also in our data), occurs similarly across presentation eccentricities, and is not related to response execution. It also exhibits other characteristics that suggest 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  $40^\circ$  – as shown in the peripheral trials in Experiment 1. Second, it is independent of a closely related statistical property - the 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)(22, 26). Together, these properties make a low-level, perceptual origin very unlikely. Note that priming of mean judgments by visual variance, as described in (32), is also independent of the similarity of means and retinal location.

The most compelling argument in favour 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 make 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 (22, 26), 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 (39). Characterising the precise effect of attention on serial dependence of variance judgements, 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/decisional history (29). However, an interpretation based on sensory precision might also predict two associations that are not found in our data: and i) an inverse association of positive serial dependence with current-trial confidence, and ii) an inverse association with eccentricity, given lower sensory precision in the peripheral field. Rather, 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 (55).

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 only arises when a high-confidence judgment about variance was made in trial  $n-2$ ) is much weaker, on average, than that of the  $StD_{n-1}$ . When enquiring about 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 had been 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  $n-1$  (in addition to the decision about variance), 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 Fritsche et al (48), who assert the same for orientation judgments. These 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 and colleagues (49) 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-second delay between stimulus offset and response is placed. Interestingly, in 2005 Kanai and colleagues 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' (50). 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 in this study, 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 pre-cue (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.

Our conclusion of a high-level mechanism of variance processing is also in line with the conclusions of Payzan-Le Nestour et al (2016) regarding variance-driven adaptation after-effects, which suggest that variance is an abstract property that works independently from its sensory origin and generalizes across domains (34). Michael and colleagues (32) also propose 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 generalise to other variance-related serial dependencies requires further investigation.

What are the perceptual/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 (39)). Furthermore, its behaviour resembles that implied by Bayesian frameworks of information processing, in which judgments about a certain dimension are attracted towards prior information. Several studies have recognised 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 (42, 52, 56, 57). 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 after-effects, like the variance adaptation described in (34). The fact that the negative effect is observed in relation with individual presentations lasting only 500 ms, appears as early as for the following trial, and remains even for trial n-9 could seem unusual for a 'sensory' after-effect. However, negative after-effects in response to sub-second stimuli have been described previously ((50), (48)), and in (48) it also lasts for several seconds. In (48), the authors propose that it is not the stimulus itself, but a memory trace that causes the negative after-effect 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 after-effect strictly sensory in the first place (58). As stated in the previous section, 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 post-perceptual after-effects may be blurred concerning statistical properties (34, 58).

Some previous studies on different features, both low-level (namely motion (50) and orientation (48)) as well as high-level (such as face attributes (59)) have reported concomitant positive and negative biases exerted by the same stimulus. In Kanai et al (50), a negative rapid motion after-effect of sensory origin (rMAE) was elicited by a short, sub-second sine-wave luminance grating presented immediately before. However, when the inter-stimulus interval (ISI) was long enough (>3 seconds), 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). In Fritsche et al (45), opposite effects of recent history on orientation judgments were exerted by perception (negative bias) and decision (positive

serial dependence), very much in line with our findings. The authors propose 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 suggest the same duality in their study of serial dependences in face attributes (59), 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 after-effects maximizing sensitivity would predominate in changeable attributes such as facial expression.

In summary, our study on visual variance reveals two opposite inter-trial dependences that operate at different timescales and which 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 modelled within an iterative Bayesian framework.

**Acknowledgements.** All authors are grateful to the Dr. Mortimer and Theresa Sackler Foundation which supports the Sackler Centre for Consciousness Science and supports a PhD scholarship for MSP. MSP 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.

**Author contributions:** MSP, AKS and WR designed research; MSP performed research; MSP analysed the data; MSP, AKS and WR wrote the manuscript.

## REFERENCES

1. Alvarez GA. Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*. 2011;15(3):122-31.
2. Geisler WS. Visual Perception and the Statistical Properties of Natural Scenes. *Annual Review of Psychology*. 2008;59:167-92.
3. Ariely D. Seeing Sets: Representation by Statistical Properties. *Psychological Science*. 2001;12(2):157-62.
4. Rosenholtz R, Huang J, Raj A, Balas BJ, Ilie L. A Summary Statistic Representation in Peripheral Vision Explains Visual Search. *Journal of Vision*. 2012;12(4):1-29.
5. Chong SC, Treisman A. Representation of statistical properties. *Vision Research*. 2003;43(4):393-404.
6. Alvarez GA, Oliva A. Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences*. 2009;106(18):7345-50.
7. Fairhall AL, Lewen GD, Blalek W, Steveninck RRdRv. Efficiency and ambiguity in an adaptive neural code. *Nature*. 2001;412(6849):787-92.

8. Dahmen JC, Keating P, Nodal FR, Schulz AL, King AJ. Adaptation to Stimulus Statistics in the Perception and Neural Representation of Auditory Space. *Neuron*. 2010;66(6):937-48.
9. Summerfield C, Egnér T. Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*. 2009;13(9):403-9.
10. Summerfield C, Lange FPd. Expectation in perceptual decision-making: neural and computational mechanisms. *Nature Reviews Neuroscience*. 2014;15:745-56.
11. Chong SC, Treisman A. Statistical processing: computing the average size in perceptual groups. *Vision Research*. 2005;45:891-900.
12. Corbett JE, Oriet C. The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica*. 2011;138(2):289-301.
13. Albrecht AR, Scholl BJ. Perceptually Averaging in a Continuous Visual World: Extracting Statistical Summary Representations Over Time. *Psychological Science*. 2010;21(4):560-7.
14. Im HY, Chong SC. Mean size as a unit of visual working memory. *Perception*. 2014;43(7):663-76.
15. Haberman J, Whitney D. Seeing the mean: Ensemble processing for sets of faces. *Journal of Experimental Psychology Human Perception and Performance*. 2009;35(3):718-34.
16. Sweeny TD, Whitney D. Perceiving Crowd Attention: Ensemble Perception of a Crowd's Gaze. *Psychological Science*. 2014;25(10):1903-13.
17. Wolfe BA, Kosovicheva AA, Leib AY, Wood K, Whitney D. Foveal input is not required for perception of crowd facial expression. *Journal of Vision*. 2015;15(4):11:1-3.
18. Gardelle Vd, Summerfield C. Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences*. 2011;108(32):13341-6.
19. Gardelle Vd, Mammassian P. Weighting Mean and Variability during Confidence Judgments. *PLOS ONE*. 2015;10(3):e0120870.
20. Zylberberg A, Roelfsema PR, Sigman M. Variance misperception explains illusions of confidence in simple perceptual decisions. *Consciousness & Cognition*. 2014;27:246-53.
21. Maule J, Witzel C, Franklin A. Getting the gist of multiple hues: metric and categorical effects on ensemble perception of hue. *Journal of the Optics Society of America A: Optics, Image Science and Vision*. 2014;31(4):A93-102.
22. Maule J, Franklin A. Effects of ensemble complexity and perceptual similarity on rapid averaging of hue. *Journal of Vision*. 2015;15(4):1-18.
23. Brady TF, Alvarez GA. Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision*. 2015;15(15):6, 1-4.
24. Morgan MJ, Mareschal I, Chubb C, Solomon JA. Perceived pattern regularity computed as a summary statistic: implications for camouflage. *Proceedings of the Royal Society of London Series B, Biological Sciences*. 2012;279:2754-60.
25. Morgan MJ, Chubb C, Solomon JA. A 'dipper' function for texture discrimination based on orientation variance. *Journal of Vision*. 2014;8(11):9.1-9.8.
26. Fouriez G, Rubinfeld S, Capstick G. Visual statistical decisions. *Perception & Psychophysics*. 2008;70(3):456-64.
27. Sato Y, Kording KP. How much to trust the senses: Likelihood learning. *Journal of Vision*. 2014;14(13):13, 1-.
28. Corbett JE, Wurnitsch N, Schwartz A, Whitney D. An aftereffect of adaptation to mean size. *Visual Cognition*. 2012;20(2):211-31.

29. Meyniel F, Sigman M, Mainen ZF. Confidence as Bayesian Probability: From Neural Origins to Behavior. *Neuron*. 2015;88:78-92.
30. Spence M, Dux P, Arnold D. Computations Underlying Confidence in Visual Perception. *Journal of Experimental Psychology: Human Perception and Performance*. 2016;42(5):671-82.
31. Bronfman ZZ, Brezis N, Jacobson H, Usher M. We See More Than We Can Report: "Cost Free" Color Phenomenality Outside Focal Attention. *Psychological Science*. 2014;25(7):1394-403.
32. Michael E, Gardelle Vd, Summerfield C. Priming by the variability of visual information. *Proceedings of the National Academy of Sciences*. 2014;111(21):7873-8.
33. Haberman J, Lee P, Whitney D. Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision*. 2015;15(4):16, 1-1.
34. Payzan-LeNestour E, Balleine BW, Berrada T, Pearson J. Variance After-Effects Distort Risk Perception in Humans. *Current Biology*. 2016;26:1-5.
35. Campbell FW, Maffei L. The tilt after-effect: a fresh look. *Vision Research*. 1971;11:833-40.
36. Mather G, Verstraten F, Anstis S. *The Motion Aftereffect: A Modern Perspective*: The MIT Press; 1998.
37. Roseboom W, Linares D, Nishida Sy. Sensory adaptation for timing perception. *Proceedings of the Royal Society B*. 2015;282:20142833.
38. Kohn A. Visual Adaptation: Physiology, Mechanisms, and Functional Benefits. *Journal of Neurophysiology*. 2007;97:3155-64.
39. Fischer J, Whitney D. Serial dependence in visual perception. *Nature Neuroscience*. 2014;17:738-43.
40. Liberman A, Fischer J, Whitney D. Serial Dependence in the Perception of Faces. *Current Biology*. 2014;24(21):2569-74.
41. Xia Y, Liberman A, Leib AY, Whitney D. Serial Dependence in the perception of attractiveness. *Journal of Vision*. 2015;15(12):1219.
42. Cicchini GM, Anobile G, Burr DC. Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences*. 2014;111(21):7867-72.
43. John-Saaltink ES, Kok P, Lau H, Lange FPD. Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. *The Journal of Neuroscience*. 2016;36(23):6186-92.
44. Balas B, Nakano L, Rosenholtz R. A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*. 2010;9(12):1301-18.
45. Ziemba C, Simoncelli E. Opposing effects of summary statistics on peripheral discrimination. *Journal of Vision*. 2015;15(12):770.
46. Freeman J, Simoncelli EP. Metamers of the ventral stream. *Nature Neuroscience*. 2011;14(9):1195-201.
47. Wagenmakers E-J, Love J, Marsman M, Jamil T, Ly A, Verhagen J, et al. Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*. 2017:1-19.
48. Fritsche M, Mostert P, Lange FPD. Opposite Effects of Recent History on Perception and Decision. *Current Biology*. 2017;27:1-6.
49. Bliss DP, Sun JJ, D'Esposito M. Serial dependence is absent at the time of perception but increases in visual working memory. *BiorXiv*. 2017.

50. Kanai R, Verstraten FAJ. Perceptual manifestations of fast neural plasticity: Motion priming, rapid motion aftereffect and perceptual sensitization. *Vision Research*. 2005;45(2005):3109-16.
51. Rahnev D, Koizumi A, McCurdy LY, D'Esposito M, Lau H. Confidence Leak in Perceptual Decision Making. *Psychological Science*. 2015;26(11):1664-80.
52. Petzschner FH, Glasauer S. Iterative Bayesian Estimation as an Explanation for Range and Regression Effects: A Study on Human Path Integration. *The Journal of Neuroscience*. 2011;31(47):17220-9.
53. Cicchini GM, Mikellidou K, Burr D. Serial dependencies act directly on perception. *Journal of Vision*. 2017;17(14):6, 1-9.
54. Roseboom W. Serial dependence across multisensory relative timing tasks. *PsyArxiv*. 2017.
55. Alais D, Ho T, Han Se. A Matched Comparison Across Three Different Sensory Pairs of Cross-Modal Temporal Recalibration From Sustained and Transient Adaptation. *i-Perception*. 2017:1-18.
56. Luca MD, Rhodes D. Optimal Perceived Timing: Integrating Sensory Information with Dynamically Updated Expectations. *Scientific Reports*. 2016.
57. Roach NW, McGraw PV, Whitaker DJ, Heron J. Generalization of prior information for rapid Bayesian time estimation. *PNAS*. 2017;114(2):412-7.
58. Storrs KR. Are high-level aftereffects perceptual? *Frontiers in Psychology*. 2015;6:157:1-4.
59. Taubert J, Alais D, Burr D. Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports*. 2016.