Disambiguating serial effects of multiple timescales

Nikos Gekas

Laboratoire des Systèmes Perceptifs, Département d'études cognitives, École normale supérieure, PSL University, Paris, France
School of Psychology, University of Nottingham, Nottingham, UK ✉

Kyle C. McDermott

Laboratoire des Systèmes Perceptifs, Département d'études cognitives, École normale supérieure, PSL University, Paris, France
Vizzario, Inc., CA, USA ✉

Pascal Mamassian

Laboratoire des Systèmes Perceptifs, Département d'études cognitives, École normale supérieure, PSL University, Paris, France ✉

What has been previously experienced can systematically affect human perception in the present. We designed a novel psychophysical experiment to measure the perceptual effects of adapting to dynamically changing stimulus statistics. Observers are presented with a series of oriented Gabor patches and are asked occasionally to judge the orientation of highly ambiguous test patches. We developed a computational model to quantify the influence of past stimuli presentations on the observers' perception of test stimuli over multiple timescales and to show that this influence is distinguishable from simple response biases. The experimental results reveal that perception is attracted toward the very recent past and simultaneously repulsed from stimuli presented at short to medium timescales and attracted to presentations further in the past. All effects differ significantly both on their relative strength and their respective duration. Our model provides a structured way of quantifying serial effects in psychophysical experiments, and it could help experimenters in identifying such effects in their data and distinguish them from less interesting response biases.

Introduction

Human perception is affected by what has been previously experienced. However, there is ongoing debate with regards to the exact nature of the correlation (positive or negative), the timescales in-volved (from the recent to the distant past), and the mechanisms responsible (one vs. multiple mechanisms). Visual *adaptation*, for example, produces a plethora of visual aftereffects, from motion (Mather, Verstraten, & Anstis, 1998) to color (Webster & Mollon, 1991) and orientation (Jin, Dragoi, Sur, & Seung, 2005). Consistently, these aftereffects reveal a negative correlation between the current percept and the adaptor (Thompson & Burr, 2009). For example, after adaptation to a leftwards oriented grating, the perceived orientation of a vertical grating is biased rightwards, opposite of the adaptor.

Contrary to these classical negative aftereffects, many studies have reported that there is a positive correlation between visual features of the current percept, such as orientation, numerosity, or facial attributes, with those of the immediate past (Cicchini, Anobile, & Burr, 2014; Fischer & Whitney, 2014; Liberman, Fischer, & Whitney, 2014). The argument for this *serial dependence* is that the physical world is usually stable and continuous over time, so the recent past can be a good predictor of the present. It is counterintuitive that the same mechanisms can be responsible for two diametrically opposite effects. Recently, Fritsche, Mostert, and de Lange (2017) suggested that perception is repelled away from previous stimuli, while postperceptual decisions are biased toward previous stimuli. In their paradigm, the positive bias increased for longer delays between perception and decision, suggesting that working memory representations might be the source of this

bias. These findings suggest that two different mechanisms may be involved in perceptual and decisional biases. Even so, Cicchini, Mikellidou, and Burr (2017) found that it is possible to observe strong serial dependencies within both perceptual and decisional processes.

The features of the stimulus itself as well as the timescale of presentation largely determine the nature of the visual aftereffect. Adaptation usually results from prolonged exposure to the same strong, salient stimulus, whereas serial dependence is present after brief exposure to a less salient stimulus and appears to be highly dependent on attention (Kiyonaga, Scimeca, Bliss, & Whitney, 2017). It is important to acknowledge that serial dependence may be present in adaptation paradigms but it may be too weak to be measured in the data. This implies also a significant difference in the magnitude between the two effects. However, opposite effects can be seen even for the same stimulus. In a face features discrimination task, Taubert, Alais, and Burr (2016) found strong positive dependencies for gender features but negative dependencies for expressions. In this case, the biases appear to depend on the permanence of the features, or, in other words, the expected timescale of change of these features (people change their expression faster than their gender).

Most visual aftereffects last for a limited amount of time. Serial dependence appears to progressively decay and disappear after several seconds (Fischer & Whitney, 2014). Adaptation can last longer after extensive and continuous exposure or even for days in special cases like the McCollough effect (McCollough, 1965; Jones & Holding, 1975) but generally also decays and disappears quickly. An interesting question that has been investigated less often is how adaptation affects perception after a longer period of time, or, inversely, how stimuli further in the past affect the current percept. Chopin and Mamassian (2012) found that adaptation produces a negative correlation between the current percept and visual events presented just before and a positive correlation with events presented further in the past. This finding cannot be explained by most theories of adaptation that posit that the visual system attempts to recalibrate itself relative to the recent past (Kohn & Movshon, 2003) or corrects the activity in the sensory channels by comparing it to a fixed distribution (Dodwell & Humphrey, 1990). A potential explanation is that adaptation is predictive. The visual system may use the distant past as an estimate of the world's statistics, which is then combined with the more recent history to predict the next percept (Chopin & Mamassian, 2012).

In this study, we address the question of how the current percept is affected by the stimulus history from the immediately preceding stimulus to hundreds of stimuli further in the past. We do this through two novel experimental paradigms in which we measured the perceptual effects of adapting to changing stimulus statistics. We presented human observers with a series of Gabor patches and monitored how their perception of the orientation of highly ambiguous stimuli were changing in conjunction with the abrupt (Experiment 1) or gradual (Experiment 2) change in stimulus statistics. We hypothesized that a negative tilt aftereffect for short timescales gradually changes into a positive effect, though relatively weaker, for trials further in the past. We quantify the relation between the presented stimuli and the observers' responses using a series of computational models, and we show that the positive correlation with stimuli further in the past is a genuine effect of the stimulus statistics on perception and not an artefact of the observers' own responses.

Experiment 1

Methods

Thirty-eight human participants took part in Experiment 1. They all had normal or corrected-to-normal vision. All were naive with regard to the purpose of the study, and they gave informed written consent in accordance with the local ethics committee and the Declaration of Helsinki.

The experiment was completed in one session that lasted approximately 1 hr and 15 min. The session was divided into two phases. During Phase 1 (Figure 1A), participants saw three high contrast (80%) Gabors (spatial frequency: 2 cycles/degree) in succession at fixation. Two Gabors, that we label A and B, had orientations that were 25° and 65° right of vertical and were presented multiple times in a randomized order. The third Gabor, that we label X, had an orientation between these two extremes. Whenever a test stimulus X was presented, participants were asked to report in a two-alternative forced-choice (2AFC) task whether its orientation was closer to A or B. The orientation of the test stimulus was determined by four interleaved adaptive staircases. Feedback was provided in the form of sound to familiarize participants with the task. Afterwards, the orientations of all stimuli were rotated by 90° counterclockwise and participants performed four interleaved adaptive staircases without feedback. The process was repeated and a full psychometric function was generated from the eight staircases. The psychometric function is a logistic function defined as $\psi(\theta) = 1/\left(1 + e^{-k(\theta - c)}\right)$, where θ is the orientation of test stimulus X, k is the maximum slope, and c is the point of subjective equality (PSE). For each participant, the PSE c and the threshold k', which is defined as the distance in degrees between 0.5 and 0.75 of the
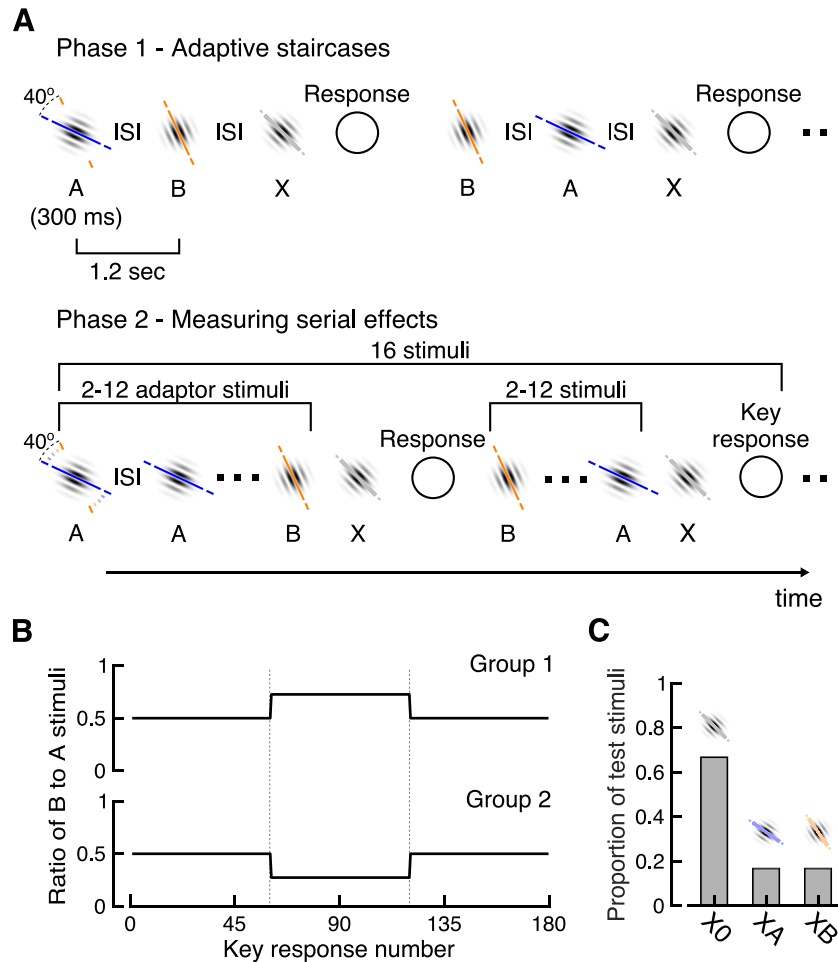
Figure 1. Experiment 1 paradigm. (A) Experimental procedure. During Phase 1, participants saw three Gabors in succession at fixation. The first two Gabors were "A" and "B" Gabors in a randomized order. The third Gabor had an orientation between these two extremes. Participants were asked to report whether its orientation was closer to A or B. They repeated this task for a total of eight adaptive staircases. During Phase 2, participants were presented with a series of high-contrast Gabors in succession. They were asked to attend to the orientations of the Gabors, and, when the response cue appeared, to report whether the orientation of the last Gabor before the cue was closer to the A or B stimulus. A response cue was shown every 16 stimuli (key response) with an additional cue randomly inserted between these stimuli. (B–C) Experimental stimuli. (B) The proportion of adaptor stimuli (A and B) that participants did not respond to before each key response (14 stimuli in total) was manipulated over time. In the first and last thirds of the session, A and B Gabors were seen in equal proportions. In the middle third of the session the ratio of A to B stimuli was 3:11, constituting a clockwise bias in the distribution of orientations for participants of Group 1 and vice versa for Group 2. (C) Participants were tested on three distinct orientations: X0, which was at the participants' initial PSE and XA and XB, which were halfway between the PSE and each of A and B. Over the whole session, the X0 orientation was shown 2 times as often than the XA and XB orientations combined.

psychometric function, were fitted to the data. These were used to generate the stimuli for the second phase of the experiment. The relationship between k' and k is $k' = 2 \cdot \ln(9)/k$.

During Phase 2 (Figure 1A), a series of Gabors were presented to the participants, who were asked to attend to their orientations and wait for a circular cue to appear after a random number of stimuli. Each Gabor was presented for 300 ms and the interstimulus interval was 900 ms. When the cue appeared, participants were asked to report whether the last shown stimulus (the X

stimulus) was closer to the A or B stimulus (65° and 25° left of vertical, respectively). The next stimulus was presented after the participant's response. The response cue appeared consistently every 16 stimuli (the "key response") and once more in between these 16 stimuli at random. The proportion of A and B stimuli (14 stimuli in total) before each key response was manipulated over time. For the first and last third of the experiment, there was an equal number of A and B stimuli (ratio A to B equal to 7:7) before each key response. During the middle third of the experiment, the proportion was

skewed (Figure 1B); for Group 1 (22 participants), there were 11 B stimuli to 3 A (ratio A to B equal to 3:11), and vice versa for Group 2 (ratio 11:3; 16 participants). The orientations of the test Gabors X were chosen based on each participant's psychometric function. Two thirds of the stimuli had orientations at the PSE (X0), namely the most ambiguous orientations. Even though these X0 stimuli were the most informative because they were the ones most likely to be biased by previous stimuli and responses, we avoided only presenting these stimuli because this could have led participants to always respond the same thing if they noticed it. The remaining test Gabors were split between the XA orientation (equidistant from the PSE and A) and XB orientation (equidistant from the PSE and B; Figure 1C). Participants were presented with 2,880 stimuli in total and were asked to respond 360 times. Supplementary Figure S1A shows the stimuli orientations (blue dots) shown to a representative observer (Participant 3) of Group 1 and the responses (orange dots) provided by that participant. The blue and orange lines show the moving average of the previous 320 stimulus orientations and 40 responses, respectively (320 stimuli and 40 responses correspond to the same duration in seconds).

All stimuli were generated using the MATLAB (MathWorks, Natick, MA) programming language with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and displayed on a CRT monitor with a resolution of 1280 × 960 pixels at 100 Hz. Participants viewed the display in a darkened room at a viewing distance of 60 cm, and a chin rest was used to maintain a constant head location and viewing distance.

Data analysis

Using the responses to the test stimuli in the first third of the experiment, where the proportion of A and B stimuli was balanced, we calculated a new psychometric function for each participant. The extracted threshold k' values were used to evaluate whether participants performed adequately in the task. For five participants of Group 1 and one participant of Group 2, threshold values were larger than 100°, indicating an inability to accurately discriminate between the test stimuli and correctly perform the task. Therefore, these participants were removed from further analysis and the results presented below correspond to 17 participants in Group 1 and 15 participants in Group 2.

Computational model

We developed a computational model to quantify the effect of past stimuli and past responses on participants' behavior in the experiment (Figure 2A).

We are interested in the probability p_s of responding "closer to B" for a stimulus s with physical orientation θ_s. We take this probability to be simply the value of the psychometric function with midpoint c_s at θ_s, i.e.:

$$p_s = \frac{1}{1 + e^{-k(\theta_s - c_s)}} \quad (1)$$

So, for example, if the physical orientation of a stimulus is at the midpoint ($\theta_s = c_s$), the probability of responding 'closer to B' is 0.5 (Figure 2B).

The model assumes that the history of past stimuli and responses shifts the original midpoint of the psychometric function c_0 such that the new midpoint of the function c_s is:

$$c_s = c_0 + shift_{stimuli} + shift_{responses} \quad (2)$$

We assume that the influence of past stimuli is stronger for stimuli more distant from the PSE. Our assumption is based on the well documented pattern of the tilt aftereffect (Clifford, Wenderoth, & Spehar, 2000; Gibson & Radner, 1937), where the direction and magnitude of the orientation shift depend on the relative orientation difference between the adapting and test stimuli. Orientation differences up to around 50° lead observers to perceive the test pattern as oriented opposite to that of the adapting pattern with peak strength between 20° and 30°. It is plausible that adapting patterns very far from the test pattern will have a negligible adapting effect, but our experiments did not include these stimuli. We tested more complex relations between the magnitude of the effect and the relative orientation difference between the stimulus and the PSE, and we think that a linear relation is a good first approximation without the need for the inclusion of more free parameters. So, for example, for the stimulus presented just before the current one, this past stimulus will create a shift equal to:

$$shift_1 = (c_{s-1} - \theta_{s-1}) \cdot F(1)$$

where θ_{s-1} is the orientation of the previous stimulus, c_{s-1} is the PSE of the psychometric function when that previous stimulus was presented, and $F(1)$ is a weight parameter. Generalizing for multiple past stimuli, the overall effect $shift_{stimuli}$ is calculated by multiplying the distance between the orientation of each past stimulus $\theta_{s-1} \ldots \theta_{s-n}$ and the midpoint of the function at that time $c_{s-1} \ldots c_{s-n}$ with a function F of the influence each past stimulus has on the shift:

$$shift_{stimuli} = k' \sum_{i=1}^{S} (c_{s-i} - \theta_{s-i}) \cdot F(i) \quad (3)$$

where S is the number of all stimuli presented before stimulus s. We add the threshold k' in Equation 3 to normalize the magnitude of the function across participants, independently of the participant's sensi-
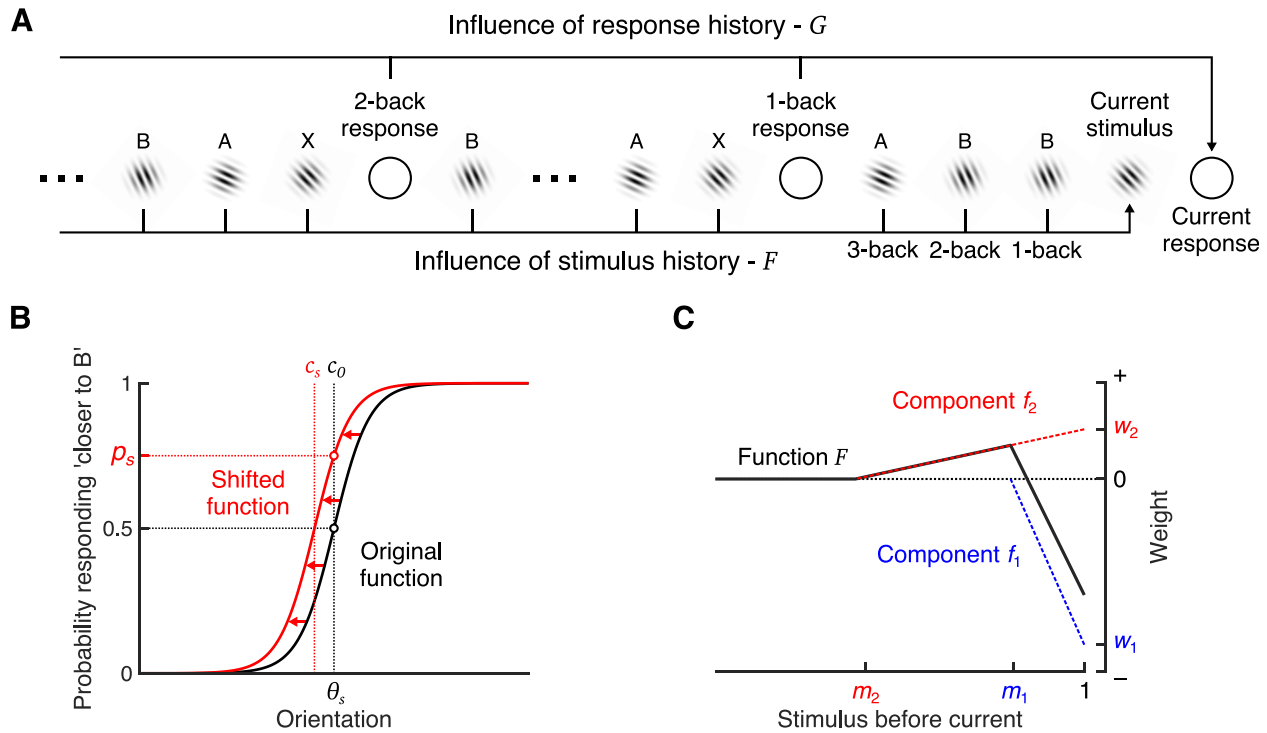
Figure 2. Computational model of influence of past stimuli and responses. (A) The perceived orientation of the current stimulus is affected by the perceived orientations of past stimuli according to function F, and by past responses according to function G. (B) The probability of responding "closer to B" p_s for a test stimulus s with physical orientation θ_s calculated from the psychometric function individually for each participant. The midpoint of the function c_s is shifted from the original midpoint c_0 by the history of past stimuli and/or responses and a new probability of responding p_s is calculated from the shifted psychometric function. (C) Each component f of the influence function F is a linear function determined by the initial weight w with the one-back stimulus (or response for G) and the number of stimuli (or responses) m until the weight reaches zero. The influence function can be composed of an increasing number of components that sum up to a piecewise linear function. The model fits the parameters of the function by maximizing the log likelihood of obtaining the participants' responses.

tivity. The larger the distance of each past stimulus orientation from the midpoint, the larger the shift. Without the influence of past stimuli, the shift regresses to zero.

The function F has as many degrees of freedom as there are past stimuli. However, we can assume that this function is smooth, and to a first approximation, we assume that it is a linear summation of simple linear functions $F = f_1 + f_2 + f_3 + \ldots$. Each linear function $f_1, f_2, \ldots$ is defined by two parameters (Figure 2C): (a) the initial weight w of the one-back stimulus, and (b) the number of stimuli m it takes for that weight to reach zero. The value of the function for each preceding stimulus i is calculated as follows:

$$f(i) = \begin{cases} w + \dfrac{-w \cdot (i-1)}{m}, & 1 \leq i \leq m \\ 0, & i > m \end{cases} \quad (4)$$

The complexity of function F can increase by adding more linear components and summing them. The resulting function is a piecewise linear function with an even and moderate number of free parameters. We also considered a function F as a superposition of expo-

nential functions. These are similarly described by two free parameters; $f(i) = w \cdot e^{-i/\tau}$, where w is the initial weight of the one-back stimulus and τ is the decay rate. We did not find any qualitative differences between the models. We selected a piecewise linear function because we believe that it represents a good approximation of the underlying function without making a strong assumption about it. In addition, the piecewise linear function is able to reach 0 at a specific past stimulus. This has two advantages; first, it is reasonable to expect that there is a cutting point after which the effect of past stimuli is negligible instead of continuing into a very long tail. Second, it is computationally less expensive during the fitting process as the function does not have to take into account all past stimuli unless the optimization process dictates so.

The effect of the history of past responses $shift_{responses}$ is quantified in a similar way. A different function G is multiplied with the responses given in the past $r_{s-1} \ldots r_{s-n}$ and the sum corresponds to the shift in the midpoint of the psychometric function:

$$shift_{responses} = \sum_{i=1}^{R'} r_{s-i} \cdot G(i) \quad (5)$$

where R' is the number of all past responses before the current stimulus, and responses r are coded as -1 for A and $+1$ for B. Similar to F, the function G is a linear summation of simple linear functions $G = g_1 + g_2 + g_3 + \ldots$. Each function g is defined by two parameters: (a) the initial weight w of the one-back response, and (b) the number of responses m it takes for that weight to reach zero (Equation 4).

In order to fit the model's parameters to the data, we calculate the log likelihood of obtaining each participant's responses and the model's parameters are the ones that maximize the log likelihood L:

$$\log L(p) = \sum_{i=1}^{R} r_i \cdot \log(p_i)$$
$$+ \sum_{i=1}^{R} (1 - r_i) \cdot \log(1 - p_i) \quad (6)$$

where R is the number of all responses given by the participant during the experiment, and responses r are normalized to 0 for A and $+1$ for B. We tested increasingly complex models, with various number of components quantifying the effect of past stimuli and responses, and we compared the fitness of each model by calculating the Akaike Information Criterion (AIC) value of each one. The AIC is defined (Cavanaugh, 1997) as $2\kappa - 2\ln(L) + 2\kappa(\kappa + 1)/(n - \kappa - 1)$, where L is the likelihood of generating the experimental data from the model, κ is the number of parameters in the model, and n is the number of data points available. Further, we calculated the Akaike weights (Wagenmakers & Farrell, 2004) of each model defined as $w_i(AIC) = \exp(-0.5\Delta_i(AIC))/$ $\sum_{j=1}^{M} \exp(-0.5\Delta_j(AIC))$, where $\Delta(AIC)$ is the difference in AIC value between each model and the best candidate model (the model with the smallest AIC) for all M models tested. Akaike weights can be considered as conditional probabilities for each model and are more intuitive than raw $\Delta(AIC)$ values. All models were fitted to all responses from all participants, so each model's parameters were fitted to a total of 32 (participants) $\times$ 360 (responses) $= 11{,}520$ data points.

Results

Participants viewed serially presented oriented gratings selected from five stimulus orientations: A, XA, X0, XB, and B. Upon the appearance of a cue, participants were asked to indicate whether the last shown stimulus's orientation (X0, XA, or XB) was

closer to the A or B stimulus. The distribution of A and B orientations was balanced evenly for the first and last third of the experiment, and biased toward one orientation for the middle third. The black lines in Figure 3A show the moving average for Group 1 (top) and Group 2 (bottom) of the previous 40 responses. The yellow lines show the moving average of the adaptor pattern for visual comparison.

A first point to notice is that there is a clear differentiation between the mean responses in the balanced and unbalanced parts of the experiment. The mean responses during the middle third of the experiment appear to be biased away from the more frequently presented orientation (B for Group 1 and A for Group 2). Moreover, there are differences between the two balanced parts of the experiment. While in the first part, the mean responses appear mostly stationary over time in agreement with the balanced distribution, there is a deviation at the beginning of the last third of the experiment where the mean response is attracted toward the biased orientation, and this attraction appears to last even longer for Group 2. According to our hypothesis, this might be a lingering effect of exposure to the unbalanced distribution even after exposure to hundreds of stimuli with a balanced distribution.

We used a series of models to quantify the effect of past stimuli and responses from the recent to the more distant past (see Methods). By design, the test stimuli are highly ambiguous so it is expected that responses should be noisy. Even with a high number of participants (32), there was a large variability between participants as indicated by the confidence intervals (CIs) in Figure 3A. Thus, we fitted the same parameters to all participants in order to increase the reliability of the fitting process and avoid overfitting noise. Each model is fitted to each participant separately and not to averaged data, as each participant has presumably a distinct stimulus and response history. During the fitting procedure, the models were free to include any number of preceding stimuli or responses from one to all and any positive or negative value for the initial weight. Thus, the shape of the function was defined strictly by the data. We tested seven models in total with various combinations of influence functions; Stimulus 1, Stimulus 2, and Stimulus 3, where the stimulus influence function had one, two, and three components, respectively; Response 1 and Response 2, where the response influence function had one and two components, respectively; Stimulus Response, where there was both an effect of the stimulus and the response history (functions of one component each); and No History, where there were no serial effects at all. An overview of the models and their Akaike weights can be seen in Supplementary Table S1, and their $\Delta(AIC)$ values in Supplementary Figure S2A. The AIC comparison of the models suggests that the Stimulus 3 model is overall the most probable out of the
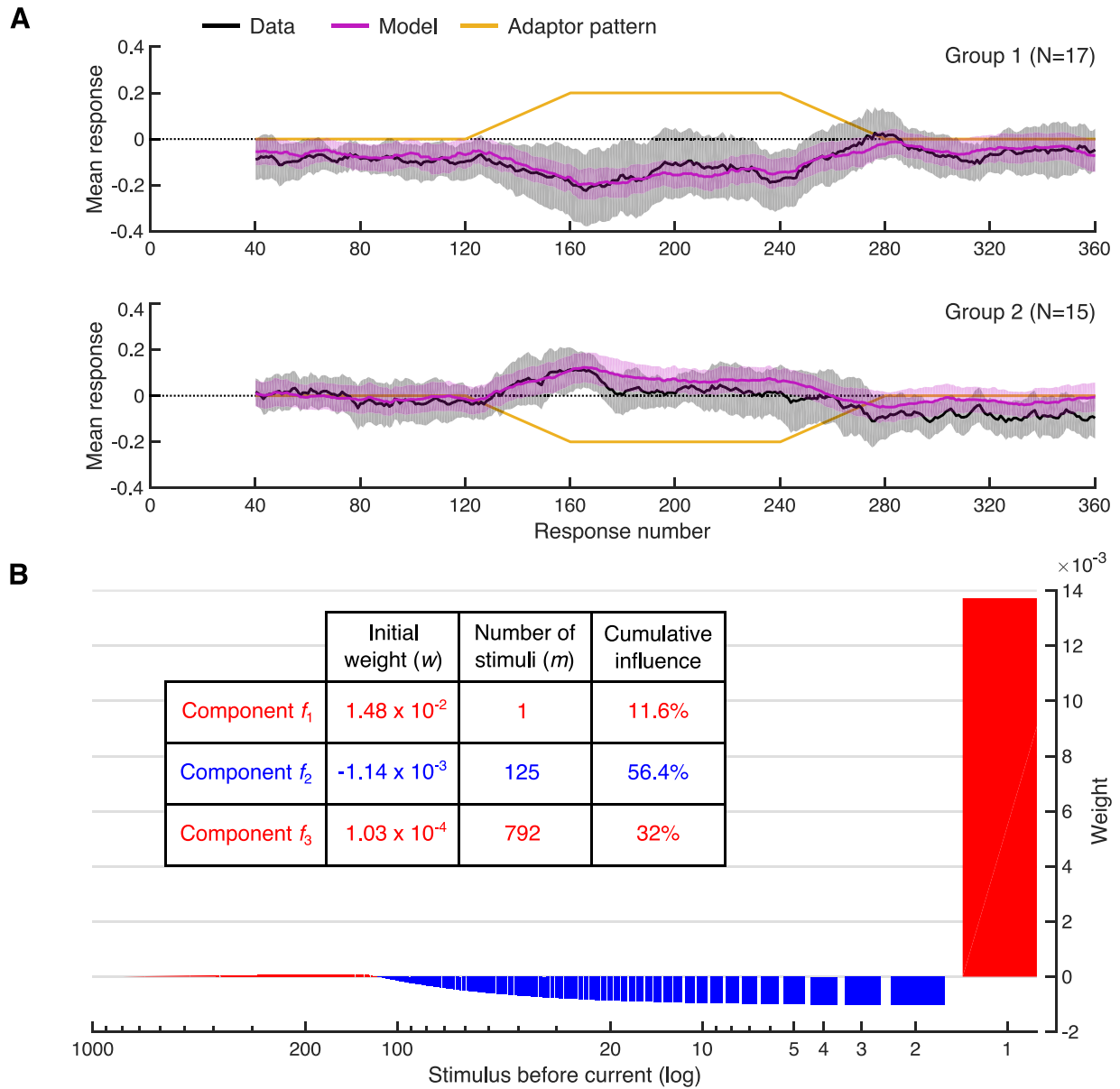
Figure 3. Experiment 1 data and model comparison. (A) Participants' mean responses and model fits. Lines show moving averages of responses (black) for all participants of Group 1 (top) and Group 2 (bottom), probability of responding as generated from the best fitting Stimulus 3 model (purple), and the pattern of adaptor stimuli (yellow). Each point of the line is the average of the current response and 39 previous responses. Gray and purple areas indicate 95% CI for the data and the model respectively. (B) Model weights. The weight of each past stimulus before the current stimulus is plotted against the stimulus's position in time, according to the best fitting Stimulus 3 model. Red bars indicate positive weights and blue bars negative. The x-axis is in log scale because of the differences in timescale between the three components of the model. The inset table shows for each component of the model the initial weight (w), the number of stimuli with a nonzero weight value (m), and the cumulative influence of the component on the current stimulus.

tested models (Akaike weight of 0.99 out of a maximum of 1). This model assumes that only the stimulus history has an effect on participants' responses and that the history is best described by three components. The models that assume only an influence of past responses or a combination of past stimuli and responses perform badly, indicating that the stimulus statistics drive participant behavior.

The purple lines in Figure 3A show the predicted mean responses of the best fitting model for each group. The model can track the long-term trends of the data adequately, especially the positive correlation at the start of the last third of the session. Figure 3B shows the influence function of past stimuli fitted by the Stimulus 3 model. The bar heights represent the weight of each past stimulus on the current percept, and the color represents

the sign of the correlation: red for positive and blue for negative. The function is shown in log space over the *x*-axis because of the significant difference in the timescales involved. The inset table shows the fitted parameters of each of the components of the model: initial weight *w* and the number of stimuli with a nonzero weight value *m*. The weight values suggest distinct effects at three different timescales. There is a strong positive correlation with the immediate past one-back stimulus. Then, there is a weaker negative correlation with preceding stimuli that lasts up to 115 stimuli in the past, followed by an even weaker positive correlation that lasts up to 792 stimuli in the past. Looking at the timescales involved, the negative correlation lasts up to 3 min, while the positive correlation lasts up to 17 min. Figure 3B illustrates the significant difference between the magnitudes of each of the components. We calculated the cumulative influence of each component on the current stimulus by comparing the summed weights of each component $\sum_{i=1}^{m} f(i)$ to the summed weights of all components $\sum_{j=1}^{3} \sum_{i=1}^{m} f_j(i)$. Interestingly, while the weight of the positive one-back stimulus component is the strongest, when factoring the cumulative influence of all stimuli up to 792 stimuli in the past, it accounts for only 11.6% of the total influence. The second (negative) component accounts for the majority of the influence with 56.4%, whereas the third (positive) component for 32% of the influence. Overall, the modeling of participants' behavior suggests that the effect of past stimuli may be more complex than a simple negative or positive correlation and changes both in sign and strength for stimuli further in the past.

Experiment 2

In Experiment 1 the statistics of the stimuli are stable for extended parts of the session. For Experiment 2, we wanted to investigate how participants are able to track gradual changes in the statistics over time. To accomplish that, the experimental paradigm of Experiment 2 was modified in two major ways. First, the orientations of adaptor stimuli changed gradually over the whole session and were generated randomly from a range of possible orientations instead of being restricted to only two possible orientations. Second, we now asked participants to compare the orientation of the test stimulus with a static reference because a comparison to the two extreme orientations was not possible anymore.

Methods

Twelve human participants took part in Experiment 2. They all had normal or corrected-to-normal vision.

All but one were naive with regards to the purpose of the study, and they did not take part in Experiment 1. The results did not differ when the nonnaive participant was removed from the analysis. All participants gave informed written consent in accordance with the local ethics committee and the Declaration of Helsinki.

The experiment was completed in one session that lasted approximately 1 hr and 20 min. The experimental paradigm was very similar to that of Experiment 1 but with some key differences in the design. We first estimated the orientation discrimination accuracy and precision of each observer. During Phase 1 (Figure 4A), participants did an orientation discrimination task where they compared the orientation of a high contrast (80%) Gabor with a static reference that was placed outside the envelope of the Gabor. In a 2AFC task, they indicated whether the orientation of the Gabor was clockwise or counterclockwise of the reference. As in Experiment 1, a full psychometric function was generated from two sets of four interleaved adaptive staircases. The PSE and threshold were calculated for each participant and they were used to generate the stimuli for the remainder of the experiment.

During Phase 2 (Figure 4A), a series of Gabors was presented to participants who were asked to attend to their orientations and wait for the reference to appear occasionally after a random number of stimuli. The orientations of these stimuli were randomly drawn from a Gaussian distribution (Figure 4B) of which the mean followed a hard-to-predict oscillating pattern (Figure 4C) and the variance matched the variance of the measured psychometric function of each participant. The frequency of the oscillating pattern was steadily increasing for participants of Group 1 (six participants), whereas it was decreasing for Group 2 (six participants). For Group 1, the mean of the Gaussian before each response was calculated according to $mean(r) = PSE_0 + 10 \cdot \sin((1 + \lambda(r - 1))r)$, where *r* is the response number (1 to 360) and $\lambda = 4/359$ (angles are in degrees). For Group 2, the equation was $mean(r) = PSE_0 + 10 \cdot \sin((3 + \lambda(r - 1))(361 - r))$, where $\lambda = -2/359$. Similar to Experiment 1, one key response was every 16 stimuli, with one more response randomly in between. When the reference appeared, participants were asked to report whether the orientation of the last shown Gabor before the appearance of the reference was clockwise or counterclockwise of the reference. Unbeknownst to the participants, the orientation of the stimuli presented before the reference was always at the PSE_0 (Figure 4D). Unlike Experiment 1 where we were concerned not to always present the same test stimulus, we reasoned that in Experiment 2 this was not a problem because of the rich variability of orientations of all the other stimuli. Supplementary Figure S1B shows the stimuli orientations presented to
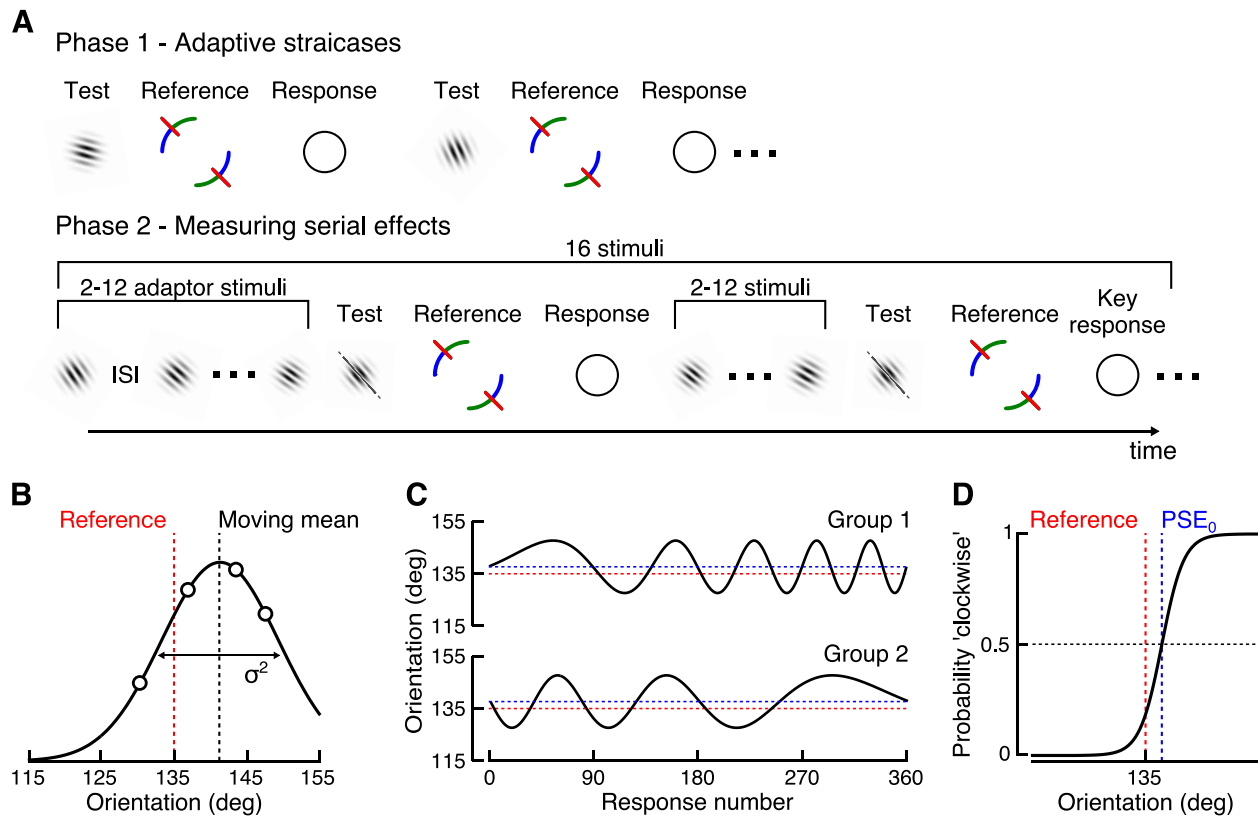
Figure 4. Experiment 2 paradigm. (A) Experimental procedure. During Phase 1, participants saw a Gabor followed by a static reference. They were asked to do an orientation discrimination task where they compared the orientation of the Gabor with that of the reference and indicate whether the orientation of the Gabor was clockwise or counterclockwise of the reference. They repeated this task for a total of eight adaptive staircases. During Phase 2, participants were presented with a series of high-contrast Gabors in succession. They were asked to attend to the orientations of the Gabors and, when the reference appeared, to compare the orientation of the last shown Gabor with the reference. (B–D) Experimental stimuli. The stimuli orientations of adaptor stimuli to which participants did not respond to were randomly drawn from a Gaussian distribution (B) of which the mean followed a complex oscillating pattern (C) over time. The mean was updated after every response. The oscillating pattern was contracting for participants of Group 1 and expanding for Group 2. (D) Test stimuli orientations were always at the initial PSE of each participant as measured at the beginning of the experimental session.

a representative observer (Participant 2 of Group 2) and responses provided by that participant.

Computational model

As the test stimulus orientation was always at the PSE, it is impossible to recalculate participants' psychometric functions, so the threshold k' values were generated from the original psychometric function measured from the adaptive staircases. We used the same models and model comparison metrics as in Experiment 1. Likewise, the models were fitted to all responses from all participants, so each model's parameters were fitted to a total of 12 (participants) × 360 (responses) = 4,320 data points.

Results

We presented a series of oriented Gabors to participants and we measured the perceived orientation of an ambiguous Gabor in comparison to a static reference. Adaptor orientations were drawn from a Gaussian distribution with a moving mean following a hard-to-predict oscillating pattern and fixed variance, while the test stimuli were always at the initial PSE_0 measured at the beginning of the experiment. The black lines in Figure 5A show the moving average of the previous 40 responses for Group 1 (top) and Group 2 (bottom), and the yellow lines show the moving average of the adaptor pattern for comparison. Participants' responses are strongly anticorrelated with the adaptor pattern and participants are more consistent in their
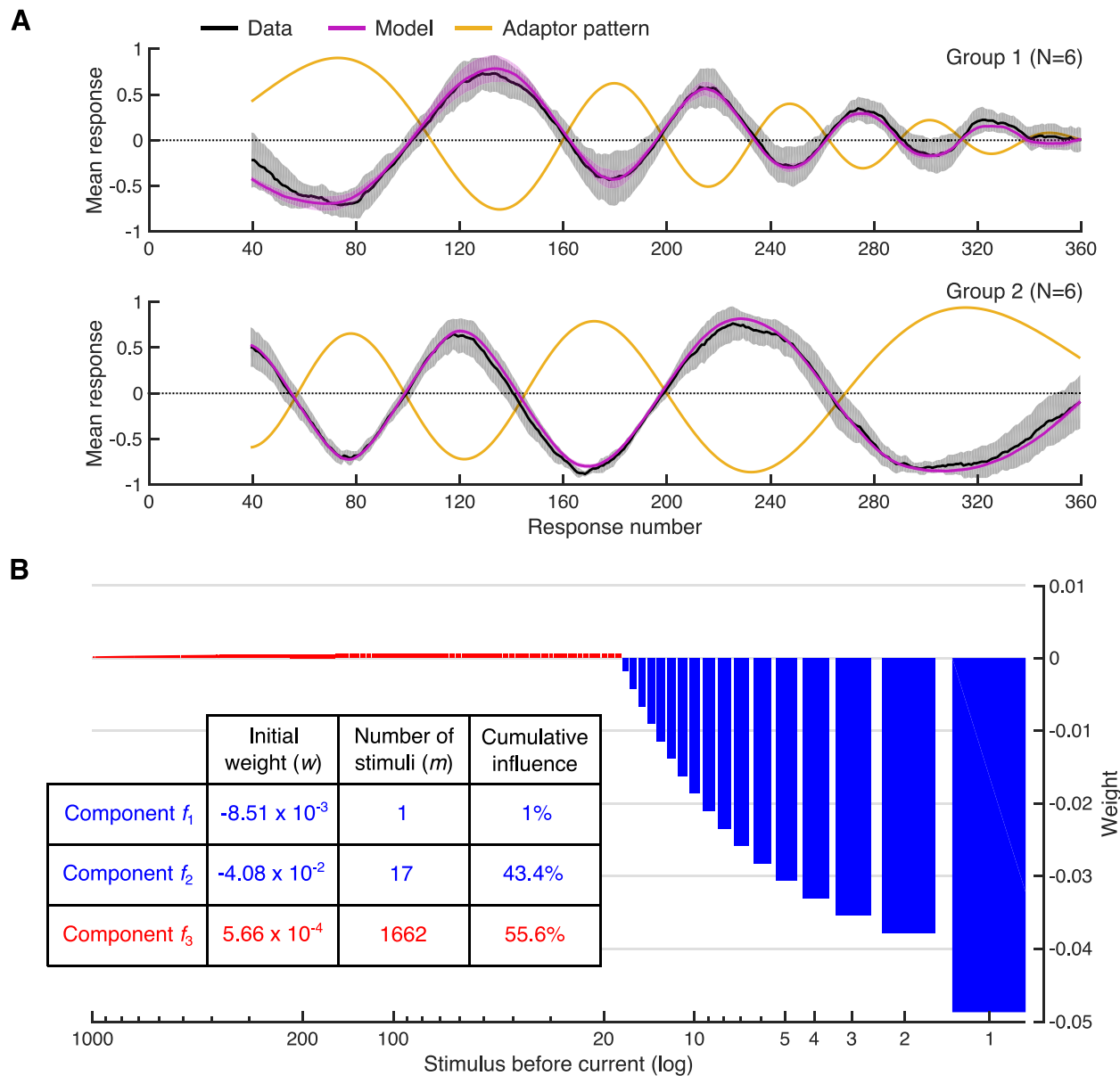
Figure 5. Experiment 2 data and model comparison. (A) Participants' mean responses and model fits. Lines show moving averages of responses (black) for all participants of Group 1 (top) and Group 2 (bottom), probability of responding as generated from the best fitting Stimulus 3 model (purple), and the pattern of adaptor stimuli (yellow). Gray and purple areas indicate 95% CI for the data and the model respectively. (B) Model weights. The weight of each past stimulus before the current stimulus is plotted against the stimulus' position in time, according to the best fitting Stimulus 3 model. The inset table shows for each component of the model the initial weight (w), the number of stimuli with a nonzero weight value (m), and the cumulative influence of the component on the current stimulus.

behavior than those in Experiment 1. On the other hand, it is now difficult to visually identify from the mean responses the effect of stimuli further in the past because the stimulus distribution is changing gradually and not abruptly as in Experiment 1. The computational model analysis is required to see if there is indeed a positive influence of past stimuli or responses.

From the model comparison (Supplementary Figure S2B), the Stimulus 3 model outperformed all other models overall. However, looking at the Akaike

weights, the Stimulus 3 model obtained a value of 0.7 and the Stimulus 2 model 0.3. That indicates that three components are not as probable as they were in Experiment 1. The Response models again were outperformed by the Stimulus models, indicating that the effect of previous responses is minimal. The purple lines in Figure 5A show the moving average of predicted response probability obtained from the Stimulus 3 model for each group. The model is able to capture the participants' behavior very accurately.

Figure 5B shows the influence function of past stimuli fitted by the Stimulus 3 model. There is a strong negative influence that lasts up to 17 stimuli followed by a weak positive influence that lasts up to thousands of stimuli. There are two major differences with the results of Experiment 1. There is no strong positive effect from the one-back stimulus. Instead, there is an additional negative influence of the one-back stimulus. However, looking at the cumulative influences, the influence of that component is minimal, as it accounts only for 1% of the influence. The longer negative component accounts for the 43.4% of the influence, whereas the positive component for the majority of the influence with 55.6%. Looking at the Stimulus 2 model, it assumes a short negative and a long positive component. This suggests that the long-term positive component is required to explain the data in Experiment 2 in both most likely models (Stimulus 2 and Stimulus 3).

It is important to note here that the stimulus statistics of Experiment 2 are significantly more autocorrelated than the statistics of Experiment 1. This is because the stimulus orientations are drawn from a concentrated distribution whose mean is varying continuously. So, the statistics of the recent past could sometimes match the statistics of the distant past. This could explain the increased influence of the long-term positive component in comparison to Experiment 1. In order to mitigate this effect, the frequency of the oscillating pattern that controls the properties of the adaptor stimuli changed over time, so that the autocorrelations also changed over time. Moreover, the characteristics of the pattern differ greatly between the two groups of participants. Therefore, it is unlikely that the strong effect is only due to autocorrelations. Overall, the experimental and modeling results suggest that apart from the expected negative correlation of recent stimuli with participants' responses, there is a relatively weaker but significantly longer lasting positive correlation with stimuli further in the past.

Discussion

Our experiments show that there are serial effects at different timescales that affect the perception of visual features. A positive correlation with the immediate past is followed by a negative correlation with presentations of the recent past and by a positive correlation with presentations of the more distant past. Crucially, the presence of these correlations can only be revealed and measured if the experimental conditions are suitable. In Experiment 1, participants were asked to compare the most recently presented stimulus with two distinct stimuli oriented at two extremes. The presentation of

one of the two alternatives before the test stimulus had a strong attractive influence on the perception of the current stimulus. In Experiment 2, where participants compared always with a static reference and the one-back stimulus was oriented pseudorandomly, we did not observe a strong attraction to that stimulus. Instead, we observed a slightly increased negative influence. This does not imply that the attractive effect is not present but that the effect may not be strong enough to be distinguishable. Due to the high number of repetitions of similar orientations over a longer period of time, the adaptation effect is stronger in Experiment 2 than in Experiment 1. The positive effect of the immediate past may then be masked by the strong adaptation. Similarly, the attractive effect of stimuli further in the past was observed in both experiments but it is stronger in Experiment 2 because the design of the experiment facilitated the extraction of that effect from the data (e.g., by removing the strong attraction to the one-back stimulus).

Our experimental paradigm differs from most serial dependence studies (e.g., Fischer & Whitney, 2014; Fritsche et al., 2017) in two crucial ways. First, participants were asked to attend to each stimulus, but they did not have to make an explicit judgment on each one. Second, they were placed in a 2AFC task instead of manually reproducing the stimulus orientation. Our design could minimize potential memory, motor, or response biases that affect responses to sequential decisions (Abrahamyan, Silva, Dakin, Carandini, & Gardner, 2016; Bliss, Sun, & D'Esposito, 2017). Moreover, the presentation of hundreds of stimuli has a two-faceted effect on observers. It induces visual adaptation due to repeated presentations of highly salient stimuli and it provides extensive information about the stimulus statistics, allowing the observer to build a strong hypothesis of the generative model that produces these stimuli. Thus, it facilitates the generation of multiple effects at multiple timescales that are well balanced, as indicated by the cumulative influence of each of the model components on the perception of the current stimulus. In comparison, Pinchuk-Yacobi, Dekel, and Sagi (2016) looked at the effect of expectations on the tilt aftereffect and found that the influences of adaptors shown more than 4 s before the test have negligible effect on the perception of the test stimulus. Likewise, Suárez-Pinilla, Seth, and Roseboom (2018) showed that, in judgments of visual variance, there may be two opposite serial effects at two distinct timescales: one positive with the immediate past and one negative up to 10 trials in the past (approximately 1 min). The current study shows unequivocally that these long-term effects are distinct from effects of shorter timescales, and it quantifies not only their relative timescale (as in Chopin & Mamassian, 2012) but also for the first time their relative strength.

The three distinct serial effects suggest that the visual system is processing information differently at three distinct timescales. As Taubert et al. (2016) suggested, the sign of the serial effect may depend on the expected timescale of change of the features in question (e.g., expression vs. gender). Likewise, the positive correlation with the immediate past may correspond to the tendency of the visual environment to remain mostly stable in the millisecond to second range where observers typically maintain eye fixation on an object. The negative correlation might instead reflect changes in the visual image due to eye and head movements, as well as small objects displacements. Finally, the long-term positive correlation might reflect long-term stability of environments in the minutes and longer timescales. The relative magnitude of these effects may also correspond to the weight that the visual system gives each time in the past. As the immediate past can be the best predictor of the present, it is weighted significantly more than the distant past. However, if the environment is shifting quickly, the negative effect is going to overwhelm the immediate positive effect. It is unclear from our data whether the strength of the effects is solely a function of time or a function of time and number of presented stimuli. Chopin and Mamassian (2012) suggested that adaptation is more dependent on discrete events than duration, and Aagten-Murphy and Burr (2016) found that to be true for numerosity judgments. We speculate that there is a complex relationship between number, duration, and attributes (e.g., orientation difference from the present stimulus and the PSE) of past events that can be uncovered by careful investigation of serial effects.

We presented a computational model that is able to distinguish between the effects of stimulus history and response history. Fründ, Wichmann, and Macke (2014) modeled history biases as a combination of stimulus and response weights and reported that observers are influenced by previous choices even when trials are independent. Our modeling results suggest that previous responses did not affect participants' behavior. The distance in time between response prompts (from 3.6 to 14.4 s) may have reduced any memory or motor biases. Our model can be easily modified and applied to different experimental paradigms. For example, in a motion-direction discrimination task, the probability that a random dot stimulus is moving leftwards or rightwards may depend on the stimulus' coherency. The only part of the model that needs to change is the calculation of the perceptual shift. In this case, the perceived motion direction of the current stimulus is affected by the motion directions of past stimuli. The model can also be extended to more complex psychometric functions where the shift can be applied to more than one parameter of the function. A potential weakness of the model is the neglect of a possible

change of the slope of the psychometric function in addition to the modeled change of the PSE. This is common to models that describe serial dependencies as response biases (Abrahamyan et al., 2016; Fründ et al., 2014; Raviv, Ahissar, & Loewenstein, 2012). From our experimental design we expect that the slope changes more slowly than the PSE, so any changes to the sensitivity would be minimal in comparison with the changes in the center of the function. Additionally, we chose to test participants around the PSE of the psychometric function so that changes in sensitivity would not solely explain the systematic biases observed in the data. Nevertheless, we think that it would be worthwhile to investigate the timescales of the change in the slope in the pursuit of a more complete model of serial dependencies.

In summary, our results suggest three distinct serial effects on perception at three distinct timescales: a positive correlation with the immediate past (few seconds), a negative correlation with the recent past (few minutes), and a positive correlation with the more distant past (tens of minutes). The positive correlation with the immediate past has been attributed to postperceptual working memory (Bliss et al., 2017; Fritsche et al., 2017) and attention (Fischer & Whitney, 2014), but it can be observed even in the absence of working memory demands (Cicchini et al., 2017; Manassi, Liberman, Kosovicheva, Zhang, & Whitney, 2018; and our study). Therefore, this positive effect may correspond to a process of very short-term integration of information across multiple levels of cortical processing (Kiyonaga et al., 2017) working independently from longer term effects and being present even when it reduces sensitivity to dynamic stimuli (Alais, Leung, & Van der Burg, 2017). The negative correlation with the recent past may indicate a shift in the perceptual space so that perceptual sensitivity is increased and the whole response range can be used. Finally, the positive correlation with the more distant past may correspond to dynamically updating the observer's prior expectations about the stimulus in a Bayesian framework (e.g., Chalk, Seitz, & Seriès, 2010; Gekas, Seitz, & Seriès, 2015), which combines with the likelihood to form the posterior percept. The model proposed here can be seen as an implementation of the suggestion that adaptation is predictive. The predictive nature of adaptation can be conceptualized such that future percepts are biased to make the statistics of recent perceptual history more similar to that of older history (Chopin & Mamassian, 2012). The contrast between recent and past history is captured in the biphasic nature of the function F in Equation 3 (ignoring here the effect of the immediate past). An outstanding issue is whether the neutral point of the perceptual space (i.e., the midpoint of the psychometric function c_0 in Equation 2) is fixed or

instead depends on the long-term perceptual experience. These two alternatives could be potentially distinguished by measuring this new midpoint after long exposure to biased stimulus statistics but in absence of recent stimulus history.

A direct prediction of our model is that for a less salient test stimulus, which is associated to a flatter psychometric function, the effect of past stimuli should have a smaller effect on the current stimulus because the perceptual shift would be smaller. Serial effects are thought to be more pronounced for weak stimuli (Mattar, Carter, Zebrowitz, Thompson-Schill, & Aguirre, 2018), but it is unclear how these effects might depend on the saliency of the test stimuli. If the two opposing effects (short-term repulsion and long-term attraction) at distinct timescales are due to distinct neural mechanisms, the repulsive effect will be weakened for less salient test stimuli (as predicted by the current model which accounts for the shift in perceptual space), whereas the attractive effect will be strengthened (due to the wider likelihood as predicted by Bayesian models). A systematic manipulation of the saliency of both adaptor and test stimuli is a worthwhile future investigative direction, and it should help us understand how past history affects perception.

Keywords: adaptation, serial dependence, orientation perception, timescale

References

Aagten-Murphy, D., & Burr, D. (2016). Adaptation to numerosity requires only brief exposures, and is determined by number of events, not exposure duration. *Journal of Vision*, *16*(10):22, 1–14, https://doi.org/10.1167/16.10.22. [PubMed] [Article]

Abrahamyan, A., Silva, L. L., Dakin, S. C., Carandini, M., & Gardner, J. L. (2016). Adaptable history biases in human perceptual decisions. *Proceedings of the National Academy of Sciences*, *113*(25), E3548–E3557.

Alais, D., Leung, J., & Van der Burg, E. (2017). Linear summation of repulsive and attractive serial dependencies: Orientation and motion dependencies sum in motion perception. *Journal of Neuroscience*, *37*(16), 4381–4390.

Bliss, D. P., Sun, J. J., & D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific Reports*, *7*(1), 14739.

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

Cavanaugh, J. E. (1997). Unifying the derivations for the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters*, *33*(2), 201–208.

Chalk, M., Seitz, A. R., & Seriès, P. (2010). Rapidly learned stimulus expectations alter perception of motion. *Journal of Vision*, *10*(8):2, 1–18, https://doi.org/10.1167/10.8.2. [PubMed] [Article]

Chopin, A., & Mamassian, P. (2012). Predictive properties of visual adaptation. *Current Biology*, *22*, 622–626.

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

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

Clifford, C. W., Wenderoth, P., & Spehar, B. (2000). A functional angle on some after-effects in cortical vision. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *267*(1454), 1705–1710.

Dodwell, P. C., & Humphrey, G. K. (1990). A functional theory of the McCollough effect. *Psychological Review*, *97*, 78–89.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*, 738–743.

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite effects of recent history on perception and decision. *Current Biology*, *27*, 590–595.

Fründ, I., Wichmann, F. A., & Macke, J. H. (2014). Quantifying the effect of intertrial dependence on perceptual decisions. *Journal of Vision*, *14*(7):9, 1–

16, https://doi.org/10.1167/14.7.9. [PubMed] [Article]

Gekas, N., Seitz, A. R., & Seriès, P. (2015). Expectations developed over multiple timescales facilitate visual search performance. *Journal of Vision*, *15*(9): 10, 1–22, https://doi.org/10.1167/15.9.10. [PubMed] [Article]

Gibson, J. J., & Radner, M. (1937). Adaptation, aftereffect and contrast in the perception of tilted lines. I. Quantitative studies. *Journal of Experimental Psychology*, *20*(5), 453.

Jin, D. Z., Dragoi, V., Sur, M., & Seung, H. S. (2005). Tilt aftereffect and adaptation-induced changes in orientation tuning in visual cortex. *Journal of Neurophysiology*, *94*, 4038–4050.

Jones, P. D., & Holding, D. H. (1975). Extremely long-term persistence of the McCollough effect. *Journal of Experimental Psychology: Human Perception & Performance*, *1*, 323–327.

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

Kohn, A., & Movshon, J. A. (2003). Neuronal adaptation to visual motion in area MT of the macaque. *Neuron*, *39*, 681–691.

Liberman, A., Fischer, J., & Whitney, D. (2014). Serial dependence in the perception of faces. *Current Biology*, *24*, 2569–2574.

Manassi, M., Liberman, A., Kosovicheva, A., Zhang, K., & Whitney, D. (2018). Serial dependence in position occurs at the time of perception. *Psychonomic Bulletin & Review*, *25*(6), 2245–2253.

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

Mattar, M. G., Carter, M. V., Zebrowitz, M. S., Thompson-Schill, S. L., & Aguirre, G. K. (2018). Individual differences in response precision correlate with adaptation bias. *Journal of Vision*, *18*(13): 18, 1–12, https://doi.org/10.1167/18.13.18. [PubMed] [Article]

McCollough, C. (1965, September 3). Color adaptation of edge-detectors in the human visual system. *Science*, *149*, 1115–1116.

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

Pinchuk-Yacobi, N., Dekel, R., & Sagi, D. (2016). Expectations and visual aftereffects. *Journal of Vision*, *16*(15):19, 1–12, https://doi.org/10.1167/16.15.19. [PubMed] [Article]

Raviv, O., Ahissar, M., & Loewenstein, Y. (2012). How recent history affects perception: The normative approach and its heuristic approximation. *PLoS Computational Biology*, *8*(10), e1002731.

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

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

Thompson, P., & Burr, D. (2009). Visual aftereffects. *Current Biology*, *19*(1), R11–R14.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196.

Webster, M. A., & Mollon, J. D. (1991, January 17). Changes in colour appearance following post-receptoral adaptation. *Nature*, *349*, 235–238.