Sensory readout accounts for adaptation

Timothy C. Sheehan[1*], John T. Serences[1,2,3]

[1]Neurosciences Graduate Program, [2]Department of Psychology, and [3]Kavli Institute for Brain and Mind, University of California, San Diego, California 92093

*corresponding author, tsheehan@health.ucsd.edu

The authors declare no competing financial interests.

Pages: 43+references

Figures: 5

Supplemental Figures (and Tables): 8 (1)

Abstract: 172 words

Introduction: 908 words (including references)

Discussion: 2320 words (including references)

Acknowledgements

Funded by NEI R01-EY025872 to JTS and NIMH Training Grant in Cognitive Neuroscience (T32-MH020002) to TCS. Thanks to Chaipat Chunharas for critical discussions in experimental design and assistance with scanning and to Anika Jollorina and Shuangquan Feng for assistance with behavioral data collection. Thanks to Marcelo Mattar for helpful comments on our model and to Margaret Henderson, Sunyoung Park, and Kirsten Adam for thoughtful comments on earlier versions of the manuscript.

Abstract

Sensory responses and behavior are strongly shaped by stimulus history. For instance, perceptual reports are sometimes biased towards previously viewed stimuli (*serial dependence*). While behavioral studies have pointed to both perceptual and post-perceptual origins of this phenomenon, neural data that could elucidate where these biases emerge is limited. We recorded fMRI responses while human participants (male and female) performed a delayed orientation discrimination task. While behavioral reports were *attracted* to the previous stimulus, response patterns in visual cortex were *repelled*. We reconciled these opposing neural and behavioral biases using a model where both sensory encoding and readout are shaped by stimulus history. First, neural adaptation reduces redundancy at encoding and leads to the repulsive biases that we observed in visual cortex. Second, our modeling work suggest that serial dependence is induced by readout mechanisms that account for adaptation in visual cortex. According to this account, the visual system can simultaneously improve efficiency via adaptation while still optimizing behavior based on the temporal structure of natural stimuli.

Keywords: neural adaptation, sensory decoding, Bayesian inference, serial dependence

Introduction

Natural stimuli are known to have strong statistical dependencies across both space and time, such as a prevalence of vertical and horizontal (cardinal) orientations and a higher probability of small orientation changes in given spatial region over short time intervals (Dong and Atick, 1995; Felsen et al., 2005; Girshick et al., 2011; van Bergen and Jehee, 2019). These regularities can be leveraged to improve the efficiency and accuracy of visual information processing. For example, regularities can yield attenuated neural responses to frequently occurring stimuli in early visual cortex (*adaptation*), reducing metabolic cost and redundancy in neural codes (Dragoi et al., 2001, 2002; Benucci et al., 2013; Patterson et al., 2014; Fritsche et al., 2022). At readout, regularities might support the formation of Bayesian priors that can be used to bias decision-making in favor of higher probability stimuli (Stocker and Simoncelli, 2006; Cicchini et al., 2014; Wei and Stocker, 2015). While the effects of stimulus history on sensory coding and behavior have been studied extensively, it is unclear how changes in sensory coding shape behavior.

Adaptation increases coding efficiency by modulating sensory tuning properties as a function of the recent past. For instance, reducing the gain of neurons tuned to a recently seen adapting stimulus reduces the temporal autocorrelation of activity when similar stimuli are presented sequentially, improving the overall efficiency of sensory codes (Clifford et al., 2000; Dragoi et al., 2000; Durant et al., 2007; Barlow, 2012; Benucci et al., 2013). Importantly, adapted representations early in the processing stream (e.g. in LGN) are inherited by later visual areas meaning the changes in coding properties could in turn shape decision making(Gardner et al., 2005; Dhruv and Carandini, 2014; Patterson et al., 2014). Although adaptation increases coding efficiency, it comes at a cost to perceptual fidelity as adaptation can lead to repulsion away from the adapting stimulus for features such as orientation and motion direction (He and MacLeod, 2001; Moradi et al., 2005; Dekel and Sagi, 2015). For example, after

continuously viewing and adapting to motion in one direction, stationary objects will appear to be moving in the opposite direction (i.e., current perceptual representations are *repelled* away from recent percepts). However, this potentially deleterious aftereffect is accompanied by better discriminability around the adapting stimulus, which may be more important than absolute fidelity from a fitness perspective (Phinney et al., 1997; Clifford et al., 2001; Abbonizio et al., 2002; Durant et al., 2007).

In contrast to the repulsive perceptual biases typically associated with neural adaptation, perceptual reports are sometimes attracted to recently presented items – a phenomenon termed "serial dependence". Studies utilizing low contrast oriented stimuli suggest serial dependence can be perceptual in nature as it operates before a peripheral tilt illusion, impacts the perception of simultaneously presented items, biases perceptual reports even when no probe is presented, and does not require a working memory delay (Fischer and Whitney, 2014; Cicchini et al., 2017, 2021; Manassi et al., 2018; Murai and Whitney, 2021). This perceptual account could arise from activity changes in early visual cortex, consistent with a fMRI study which measured early sensory biases that match 'attractive' behavioral reports (St. John-Saaltink et al., 2016). This neural finding, however, is challenging to interpret as consecutive trials were always the same or orthogonal orientations which, by definition, cannot distinguish attractive from repulsive biases. Related studies decoding past stimuli from EEG activity do not measure how current stimulus representations are biased, precluding a connection to behavioral biases (Fornaciai and Park, 2018; Bae and Luck, 2019; Bae, 2021).

Counter to studies reporting a perceptual locus of serial dependence which utilized brief or low contrast stimuli, other behavioral studies utilizing high contrast spatial stimuli have found that serial dependence does not emerge immediately but instead emerges only, and increases with, a working memory maintenance period (Papadimitriou et al., 2015; Bliss et al., 2017; Stein et al., 2020). This observation suggests that serial dependence could be implemented by a later readout or memory maintenance circuit (Papadimitriou et al., 2015; Bliss and D'Esposito, 2017; Pascucci et al., 2019;

Barbosa et al., 2020). There is evidence that such a readout mechanism is Bayesian, as the influence of the "prior" (the previous stimulus) is larger when sensory representations are less precise due to either external or internal noise (Cicchini et al., 2018; van Bergen and Jehee, 2019). Thus, the existing behavioral evidence suggests that serial dependence can operate both on perceptual and working memory representations (Papadimitriou et al., 2015; Cicchini et al., 2017; Kiyonaga et al., 2017). It is open question how and where past trial information interacts with incoming sensory and memory representations.

To determine what role visual cortex plays in driving serial dependence, we applied multivariate fMRI decoding techniques to data collected while subjects performed a delayed orientation discrimination task (Figure 1A). We replicated classic serial dependence findings where behavioral reports were attracted to the orientation of the previous stimulus. However, this attractive behavioral bias was not accompanied by attractive biases in visual cortex, as predicted by early sensory models of serial dependence. Rather, we observed *repulsive biases* in early visual cortex that were consistent with adaptation. We then examined several possible read-out mechanisms and found that only decoding schemes that account for adaptation can reconcile the neural and behavioral biases found in our data. More generally, these results explain a mechanism where the visual system can reduce energy usage without sacrificing precision by optimizing sensory coding and behavioral readout relative to the temporal structure of natural environments.

Results

Behavior

To probe the behavioral effects of serial dependence, we designed a delayed discrimination task where participants judged whether a bar was tilted clockwise (CW) or counterclockwise (CCW) relative to the orientation of a remembered grating (Figure 1A). We first report the results from a behavior-only study (n=47) followed by an analysis of neural activity for a cohort completing the same task in the fMRI

scanner (n=6). Task difficulty was adjusted for each participant by changing the magnitude of the probe offset ($\delta\theta$) from the remembered grating and was titrated to achieve a mean accuracy of ~70% (accuracy 69.8±0.82%, $\delta\theta$: 4.61±0.27°; all reported values mean ±1SEM unless otherwise noted). Fixing subjects at this intermediate accuracy level helped to avoid floor/ceiling effects and improved our sensitivity to detect perceptual biases while keeping participants motivated.

To quantify the pattern of behavioral responses, we modelled the data as the product of a noisy encoding process described by a Gaussian distribution centered on the presented orientation with standard deviation σ and bias μ. Optimal values for σ and μ were found by maximizing the likelihood of responses for probes of varying rotational offsets from the remembered stimulus, thus converting

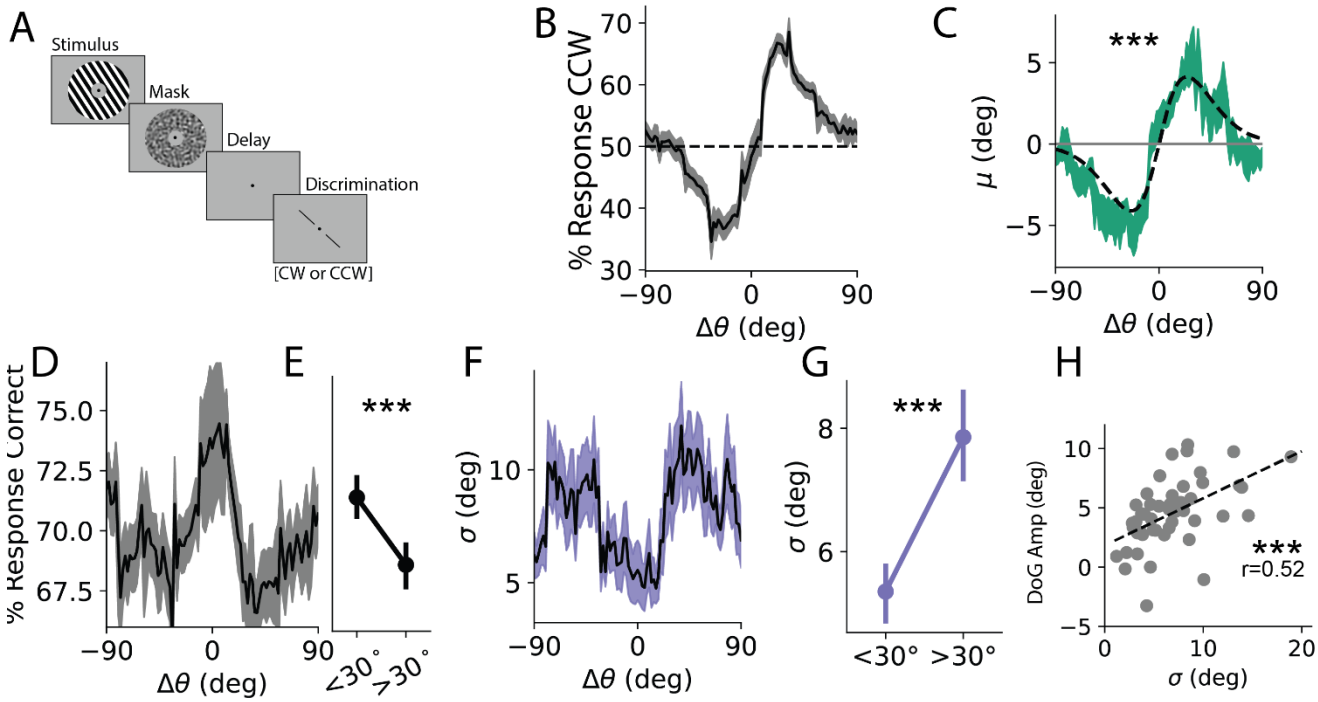

Figure 1 Behavior. **A**: Task Schematic. An orientated stimulus is followed by a probe bar that is rotated <15° from the stimulus. Participants judged whether the bar was CW or CCW relative to the stimulus in a binary discrimination task. **B**: Response bias: % of responses that were CCW as a function of $\Delta\theta=\theta_{n-1}-\theta_n$ (± SEM across participants). **C:** Behavioral bias, green: average model-estimated bias as a function of Δθ (± SEM across participants); gray: average DoG fit to raw participant responses sorted by Δθ (± 1SEM across participants). **D**: Response accuracy as a function of Δθ. **E:** Responses are significantly more accurate for |Δθ|<30°. **F**: Behavioral *σ* as a function of Δθ. **G**: Behavioral variance is significantly less for |Δθ|<30°. Note that in computing variance we 'flip' the sign of errors following CCW inducing trials to avoid conflating bias with variance (see methods) **H:** Bias is positively correlated with variance across participants. ***, p<.001

pooled binary responses into variance and bias measured in degrees (see Response Bias, Figure S1). This allowed us to measure precision for individual participants and also allowed us to measure how responses were biased as a function of the orientation difference between the remembered gratings on consecutive trials $\Delta\theta = \theta_{n-1} - \theta_n$, an assay of serial dependence.

Responses were robustly biased towards the previous stimulus (Figure 1C, green curve), which we quantified by fitting a Derivative-of-Gaussian (DoG) function to the raw response data for each participant (gray curve; amplitude: 4.53°±0.42°, $t(46) = 7.8$, $p = 5.9*10^{-10}$, one sample t-test; full width at half max (FWHM): 42.9°±1.8°, see Serial Dependence). The magnitude and shape of serial dependence is consistent with previous reports (Fischer and Whitney, 2014; Fritsche et al., 2017). This bias is not an artifact of our parameterization as the same pattern is observable in the raw proportion of CCW responses (Figure 1B). Note that as participants are reporting the orientation of the probe relative to the grating stimulus, a greater proportion of reports that the probe was CCW corresponds to a CW shift in the perception of the grating.

We next examined how response precision (σ) varied as a function of $\Delta\theta$ and found that responses were more precise around small trial-to-trial orientation changes (Figure 1F), again consistent with previous reports (Cicchini and Burr, 2018). We quantified this difference in precision by splitting trials into 'close' and 'far' bins (greater than or less than 30° separation) and confirmed that responses following 'close' stimuli were more precise ($t(46)=-3.72$, $p=0.0003$, paired 1-tailed t-test, Figure 1G , see Response Precision). Note that the choice of 30° was arbitrary, but all threshold values between 20° and 40° yielded significant ($p<.05$) results. As with bias, this variance result was not an artifact of our parameterization as raw accuracy showed a similar pattern such that responses were more accurate following close stimuli ($t(46)=3.66$, $p=.0003$, Figure 1D-E). We additionally confirmed that our finding of reduced bias around small changes in orientation is not driven by a higher proportion of 'cardinal' orientations (here defined as being ±22.5° of 0 or 90°) as the proportion of cardinal orientations did not

differ between close and far bins of Δθ (mean % cardinal close: 50.6±0.5%, far: 50.2±0.3%, t(46)=0.9, p=0.39, paired t-test).

Previous work has shown that serial dependence is greater when stimulus contrast is lower (Manassi et al., 2018) and when internal representations of orientation are weaker due to stimulus independent fluctuations in encoding fidelity (van Bergen and Jehee, 2019). We tested a Bayesian interpretation of these findings by asking whether less precise individuals are more reliant on prior expectations and therefore more biased. Consistent with this account, we found a positive correlation between DoG amplitude and σ (Figure 1H, r(45)=0.52, p=.0001, 1-tailed Pearson's correlation). This relationship was not an dependent on our response parameterization as we report found similar relationships between DoG amplitude and both accuracy (r=-0.41, Pearson correlation, p<.005) and average task difficulty $\delta\theta$ (r=.44, p<.005).

A subset of participants completed a version of the experiment with inhomogeneities in their stimulus sequences (such that consecutive orientations were more likely to be between ±22.5 and 67.5° from the previous stimulus). We repeated all of the above analyses excluding these participants and found all of our findings were qualitatively unchanged (Figure S2).

Stimulus history effects in visual cortex

To examine the influence of stimulus history on orientation-selective response patterns in early visual cortex, six participants completed between 748 and 884 trials (mean 838.7) of the task in the fMRI scanner over the course of four, two-hour sessions (average accuracy of 67.7% ±0.4% with an average probe offset, $\delta\theta$, of 3.65°). As with the behavior-only cohort, behavioral reports in these participants showed strong attractive serial dependence (Figure 2A, green) that was significantly greater than 0 when parameterized with a DoG function (amplitude=3.50°±0.27°, $t(5)=11.93$, $p=.00004$; FWHM=35.9°±2.34°, Figure 2A black dotted line). This bias was not significantly modulated by inter-trial

interval, delay period, or an interaction between the two factors (all p-values > 0.5, mixed linear model grouping by participant). Similar to the behavioral cohort, we found that variance was generally lower

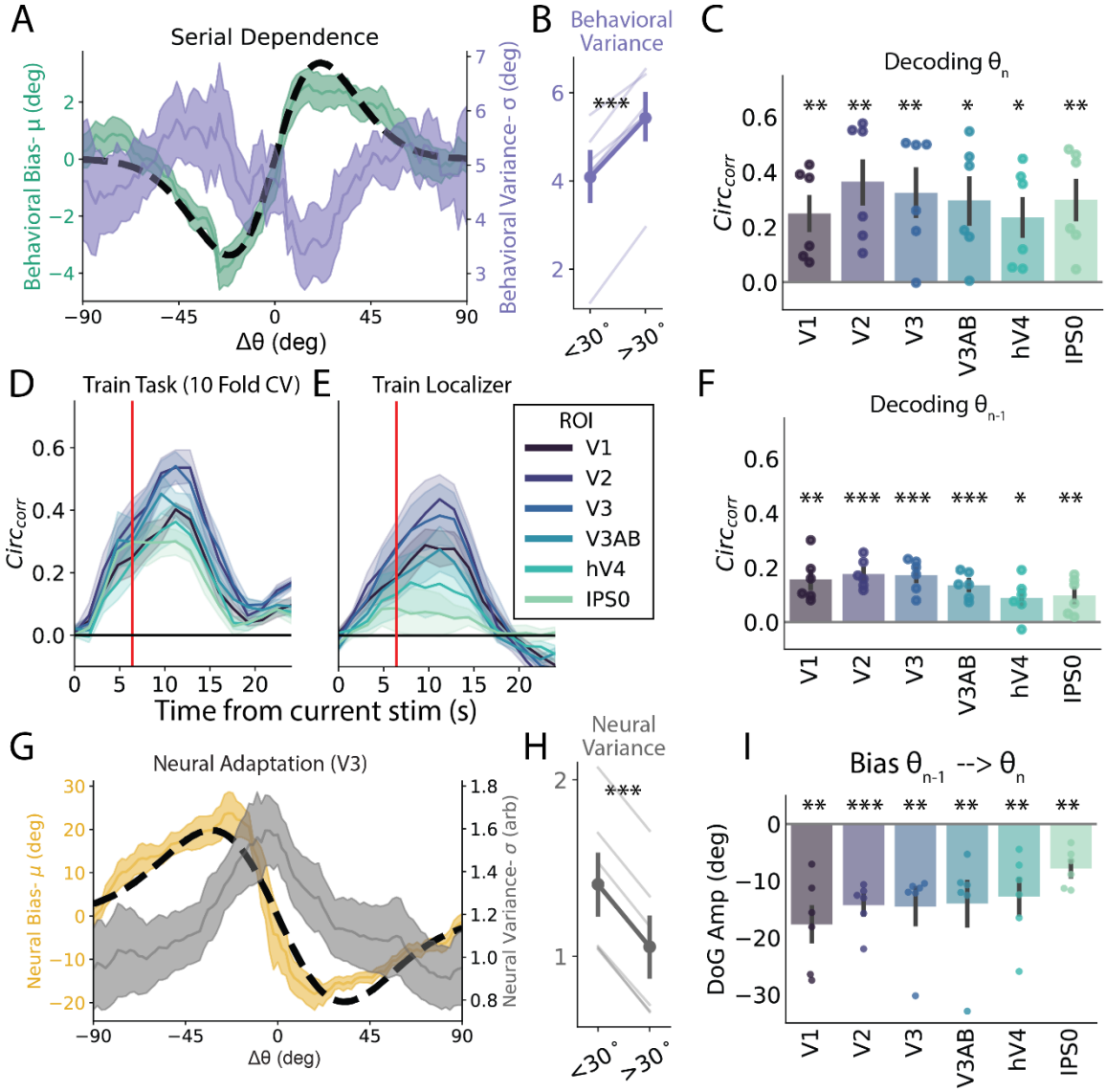

Figure 2 Behavioral and Neural Bias. ***A****: Left-axis, Behavioral serial dependence. Shaded green: average* model-estimated bias as a function of Δθ (± SEM across participants); dotted black line: average DoG fit to raw participant responses sorted by Δθ. Right-axis, variance. Purple shaded line: model-estimated variance as a function of Δθ (± SEM across participants). **B**: Behavioral σ is significantly less for |Δθ|<30°. **C**: Decoded orientation was significantly greater than chance when indexed with circular correlation for all ROIs examined. Error bars indicate ±SEM across participants. Dots show data from individual participants. **D**: Decoding performance across time for a subset of ROIs. Vertical red line indicates time point used in most analysis. **E**: Decoding performance across time for a decoder trained on a separate sensory localization task. **F**: Performance of task decoder trained and tested on identity of previous stimulus across all ROIs. **G**: Left-axis, decoding bias. Shaded yellow line: decoded bias (μ_{circ} of decoding errors) sorted by Δθ (± SEM across participants); dotted black line: average DoG fit to raw decoding errors sorted by Δθ. Right-axis, decoded σ_{circ} . Shaded gray line: average decoding variance (σ_{circ}) as a function of Δθ (± SEM across participants). Note that σ_{circ} can range from [0, inf] and has no units. **H:** Decoded variance is significantly greater for |Δθ|<30° **I:** Decoded errors are significantly repulsive when parameterized with a DoG in all ROIs. *, p<.05; **, p<.01; ***, p<.001.

144 around small values of Δθ. We quantified variance in the same manner as the behavioral cohort (flipping
145 responses to match biases and down-sampling the larger group) and found that responses were more
146 precise following close (<30°) relative to far stimuli (>30°, $t(5) = -9.96$, $p=0.00009$, 1-tailed paired t-test,
147 Figure 2B). This pattern was significant ($p<0.05$) for thresholds between 20° and 40°. A subset of these
148 participants completed some sessions where consecutive stimuli were not strictly independent as they
149 were more likely to be between ±22.5 and 67.5° from the previous stimulus (see Methods, Behavioral
150 Discrimination Task, 4 out of 6 subjects had between 357-408 trials that were non independent
151 accounting for between 40-50% of their trials and 32% of all trials completed). However, we replicated
152 all of our main analysis excluding these sessions and found that our conclusion remained unchanged
153 with the exception that our finding of reduced variance trended in the same direction but no longer
154 reached significance (Figure S3).

155 To characterize activity in early visual areas, independent retinotopic mapping runs were
156 completed by each subject to identify regions of interest (ROIs) consisting of: V1, V2, V3, V3AB, hV4, and
157 intraparietal sulcus area IPS0. In addition, a separate localizer task was used to sub-select the voxels that
158 were most selective for the spatial position and orientation of the stimuli used in our task (see Voxel
159 Selection).

160 To examine how visual representations are affected by stimulus history, we trained a decoder
161 on the orientation of the sample stimulus on each trial based on BOLD activation patterns in each ROI.
162 We used the vector mean of the output of an inverted encoding model (IEM) as a single trial measure of
163 orientation using a leave-one-run-out cross-validation across sets of 68 consecutive trials (4 blocks of 17
164 trials) that had orientations pseudo randomly distributed across all 180° of orientation space (see
165 Orientation Decoding for details). We first quantified single-trial decoding performance using circular
166 correlation (r_{circ}) between the decoder-estimated orientations and the actual presented orientations and
167 found that all ROIs had significant orientation information (Figure 2C). Our ability to decode extended

168 for the duration of the trial, peaking around 12s after stimulus presentation (Figure 2D). This memory
169 signal seems to be largely in a 'sensory code' as a decoder trained on a separate localizer task where
170 participants viewed stimuli without holding them in memory achieved similar performance over a
171 similar timescale (see fMRI Localizer Task, Figure 2E). Thus, visual ROIs showed robust orientation
172 information that could be decoded across the duration of the trial. For all analyses not shown across
173 time, we used the average of four TRs (spanning 4.8-8.0s) following stimulus presentation to minimize
174 the influence of the probe stimulus (which came up ≥6s into the trial and thus should have a negligible
175 influence on activity in the 4.8-8.0s window after accounting for hemodynamic delay, see Figure 5A).

176 We are interested in the how the identity of the previous stimulus influences representations of
177 the current stimulus, akin to previous EEG studies that have demonstrated the ability to decode the
178 previous stimulus during the current trial (Bae and Luck, 2019). We performed a similar analysis by
179 training and testing our task decoder on the identity of the previous stimulus using the same time-points
180 as the current trial decoder. This decoder was able to achieve above chance decoding in all ROIs
181 examined indicating trial history information is present in the activity patterns (Figure 2F). As a control
182 analysis, we attempted but were unable to decode the identity of the next stimulus using the same
183 procedure (Figure S3F). The performance of the memory decoder for the previous stimulus peaked
184 around 6s after stimulus presentation but remained above chance throughout the delay period (Figure
185 S4A).

186 The high SNR of the BOLD decoder additionally allowed us to examine residual errors on
187 individual trials. When measuring the bias (circular mean, μ_{circ}, *see* Neural Bias) of these decoding
188 errors as a function of stimulus history (Δθ), we observed a strong *repulsive* bias reflecting neural
189 adaptation (V3, Figure 2G yellow). This bias was significant when quantified with a DoG (amplitude=-
190 14.5°±2.9°, $t(5)=-3.56$, $p=.0029$; FWHM=52.2°±2.94°, Figure 2G black dotted-line), and all ROIs had a
191 significantly negative amplitude (p<.01, Figure 2I). Critically, this bias was present across all TRs for both

192 the task and localizer decoders and was visible in the bias curve computed for each individual
193 participant (Figure S4). In addition to the model-based analysis of responses in visual cortex, we also
194 performed a model-free assessment of the dimensionality of activation patterns conditioned on the
195 prior stimulus. Consistent with our main analysis, responses following close stimuli have a higher
196 dimensionality than responses following far stimuli. This suggests that changes due to neural adaptation
197 should assist pattern separation regardless of stimulus identity (see Dimensionality Analysis, Figure S5).

198 We also examined how the precision of neural representations changed as a function of
199 stimulus history. In sharp contrast to behavior, σ_{circ} exhibited a monotonic trend such that neural
200 decoding was *least* precise when the previous stimulus was similar (Figure 2G, gray curve, see Neural
201 Variance). We quantified this difference in sensory uncertainty in a similar manner to the behavioral
202 data and found that variance in the sensory representations was significantly greater following a similar
203 stimulus (<30°, $t(5)=72.4$, $p=4.8*10^{-9}$, paired 1-tailed t-test, V3, Figure 2H). This pattern was significant
204 ($p<.05$) in all ROIs except IPS0 (Figure S6A). The results did not change qualitatively when we utilized
205 vector length as a proxy for decoding precision derived directly from our channel estimates (Figure S6C-
206 D), or when we used other thresholds between 20° and 40°. The repulsion of sensory representations
207 and the corresponding reduction in decoding precision around the previous orientation is consistent
208 with neural adaptation where recently active units are attenuated, thus leading to lower SNR responses
209 in visual cortex.

210 Accounting for the Timecourse of the Hemodynamic Response Function

211 We considered whether the repulsive adaptation we observed in visual cortex could be
212 explained by residual undershoot of the hemodynamic response function (HRF) from the previous
213 stimulus. To address this concern, we directly modeled the evoked response in each voxel to the
214 stimulus and probe using a deconvolution approach and used a parameterization of the resulting filter
215 (double gamma function) to model the stimulus evoked response on each trial (see kernel based

216 decoding). Notably the stimulus response has an undershoot that extends up to 25s following stimulus
217 presentation (see Figure 3A for an example voxel and parameterization). Estimating responses using this
218 filter on individual trials and using the resulting weights to train a decoder removes the linear
219 contribution of previous stimulus/ probe presentations (Dale, 1999; Glover, 1999). Any bias in the
220 resulting decoder should thus be due to changes in BOLD activity driven by neuronal activity rather than
221 a hemodynamic artifact. We repeated all analyses after correcting for the shape of the HRF, and while
222 the resulting decoder was less precise than one trained on the time course data (eg. V3 r_{circ}=0.19±.07
223 versus 0.32±0.08 with time course decoder), it was still significantly predictive across all visual ROIs
224 (ps<.05) except IPS0. Despite the noisier decoding, we still observed a significant repulsive bias in all
225 visual ROIs that matched the pattern found when decoding the raw BOLD timecourse (Figure 3B).

226 To further understand whether the time course of our task could lead to artifacts, we also
227 simulated responses to our task using tuned voxels that were modeled after the task sequence and
228 estimated HRFs observed in our experiment (see supplementary modeling section, Figure S7). These
229 simulations show that repulsive biases like the ones we observed with both our time course and
230 deconvolution-based decoders are only possible when the underlying tuning of voxels is adapted by past
231 stimuli/responses.

232 We additionally examined the time-course of the bias. Significant repulsive biases were
233 observable through the duration of the trial, in all early visual ROIs (Figure S4). As the undershoot
234 portion of the HRF extended to ~25s, we examined the bias relative to the time of the presentation of
235 the previous stimulus. We included only trials with an inter-stimulus interval (ISI) greater than the
236 median of 17.5s and plotted bias as a function of the minimum time from the previous stimulus (Figure
237 3C). Notably bias was still significantly repulsive for 30s following the previous stimulus presentation in
238 all early visual ROIs, further shrinking the possibility that our biases are driven by the slow timecourse of

239 the HRF (Figure 3C, last time point). Finally, we examined how far back previous stimuli shape early

240 visual representations. We examined the influence of not just the N-1 stimulus, but N-2 and N-3 stimuli

241 as well, corresponding to median ISIs of 35.1 and 52.5s respectively (Figure 3D-E). As any influence of

242 these more distant stimuli should be diminished relative to N-1, we maximized our sensitivity by taking

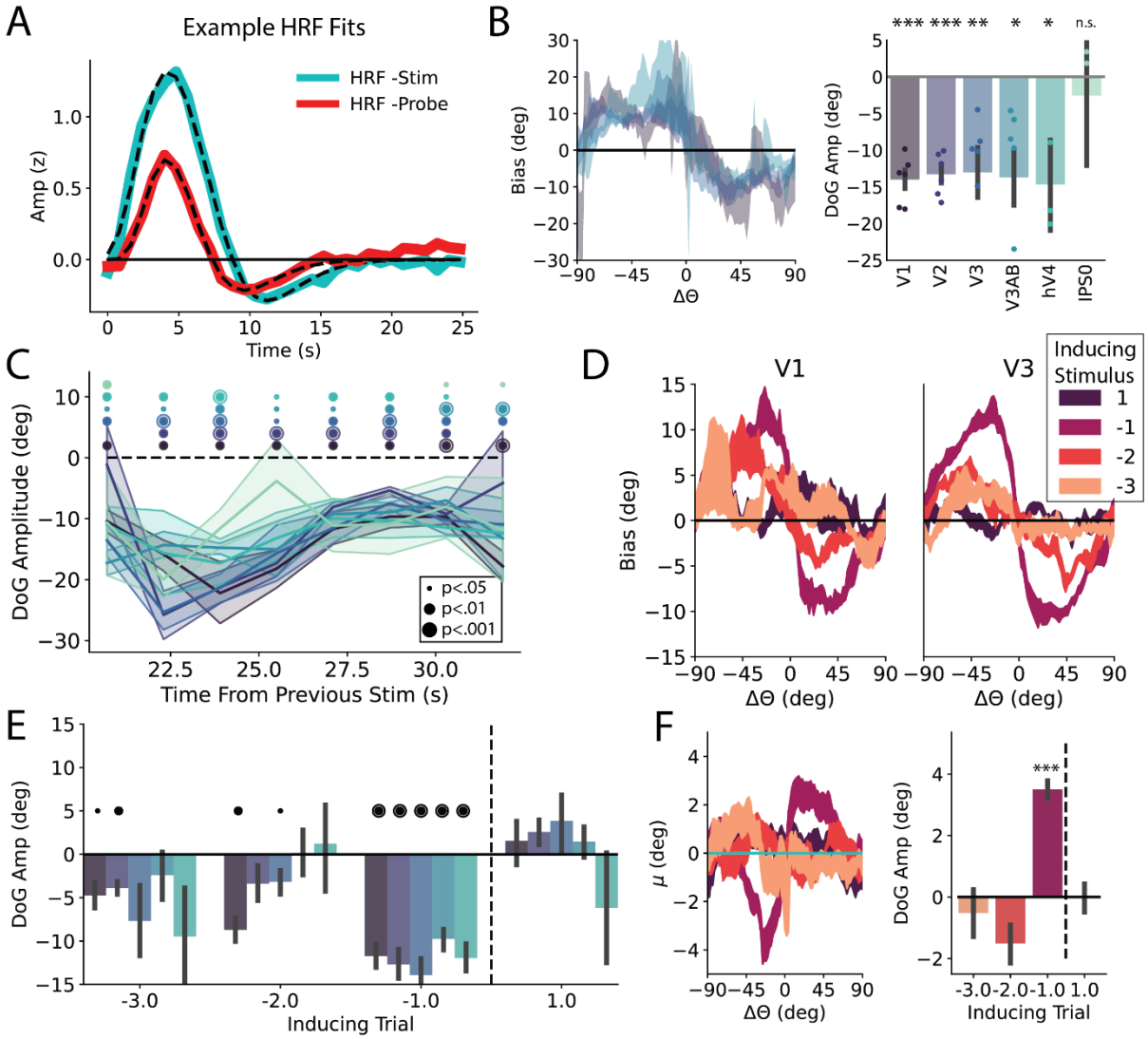

Figure 3 Influence of BOLD specific biases on repulsive bias. **A:** average V1 HRF through deconvolution for stimulus and probe. Average best fit double gamma function overlaid in dotted lines. **B:** (left) Bias curves from decoder trained on response patterns from deconvolved double-gamma functions (± SEM across participants). Here excluding hV4 and IPS0 for clarity. (right) bias quantified with a DoG function across ROIs. **C**: Bias across time including only trials with an ISI of at least 17.5s. X-axis reflects minimum time from previous stimulus. Repulsion significant in all ROIs at 32s. **D**: Bias as a function of various relative orientations for V1 and V3 (± SEM across participants). **E:** Bias across early visual ROIs for N-1, -2, -3. Color scheme same as C. N+1 control analysis to ensure effects not driven by some unknown structure in stimulus sequence. **F**: Behavioral bias for various relative orientations. N-1 data same as data presented in Figure 2. *, p<.05, **, p<.01, ***, p<.001.

243 the average decoded representation from 4-12s. While the control N+1 stimulus showed no impact on
244 decoded orientation as expected, we continued to see biases that are significantly repulsive through the
245 N-3 stimulus in V1 and V2 (Figure 3E). These neural biases were surprisingly persistent and are in line
246 with recent studies which have found adaptation signatures extending 22s in mouse visual cortex spiking
247 activity (Fritsche et al., 2022). It is not clear why our effects persist even longer, but it is likely driven in
248 part by the long ISIs, resulting in fewer intervening stimuli compared to the paradigm utilized in Fritsche
249 et al., 2022. We separately extended our analysis of behavioral biases and found no significant effect of
250 trials except for N-1, although biases were trending towards being repulsive for N-2 and N-3 reflecting
251 the pattern reported in (Fritsche et al., 2020) (Figure 3F). Together these analyses suggest that our
252 observed biases are driven by adaptation in the underlying neural population and provide additional
253 evidence that behavior is not directly linked to early visual representations.

254 Encoder-Decoder Model

255 We observed an attractive bias and low variability around the current stimulus feature in
256 behavior, and a repulsive bias and high variability around the current feature in the fMRI decoding data.
257 Thus, the patterns of bias and variability observed in the behavioral data are opposite to the patterns of
258 bias and variability observed in visual cortex. To better understand these opposing effects, we reasoned
259 that representations in early visual cortex do not directly drive behavior but instead are read out by later
260 cortical regions that determine the correct response given the task (Crick and Koch, 1995; Grunewald et
261 al., 2002; Gold and Shadlen, 2007; Siegle et al., 2021). In this construction, the decoded orientations
262 from visual cortex represent only the beginning of a complex information processing stream that, in our
263 task, culminates with the participant making a speeded button press response. Thus, we devised a two-

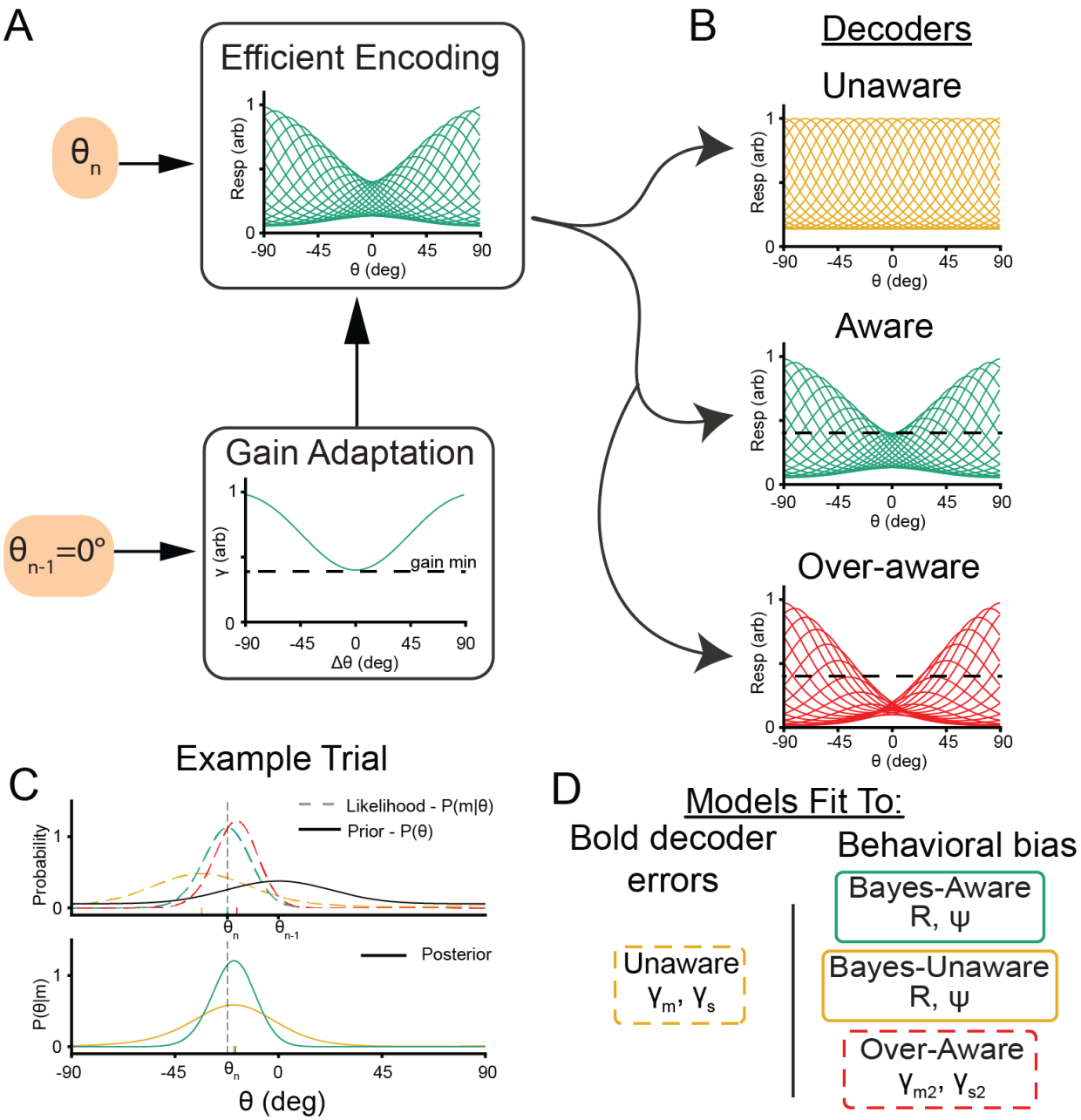

Figure 4 Encoder-Decoder model schematic. **A**: Encoding. Units with von Mises tuning curves encodes incoming stimuli. The gain of individual units undergoes adaptation such that their activity is reduced as a function of their distance from the previous stimulus. **B**: Decoding. This activity is then read out using a scheme that assumes one of three adaptation profiles. The unaware decoder assumes no adaptation has taken place, the aware decoder assumes the true amount of adaptation while the over-aware decoder over-estimates the amount of adaptation (note center tuning curves dip lower than the minimum gain line from encoding). **C**: Example stimulus decoding. Top: The resulting likelihood function for the unaware readout (dotted yellow line) has its representation for the current trial (θ_n=-30°) biased away from the previous stimulus (θ_{n-1}=0°). The aware readout (dotted green line) is not biased, while the over-aware readout is biased towards the previous stimulus. These likelihood functions can be multiplied by a prior of stimulus contiguity (solid black line) to get a Bayesian posterior (bottom) where Bayes-unaware and Bayes-aware representations are shifted towards the previous stimulus. Tick marks indicate maximum likelihood or decoded orientation. **D**: Summary of models and free parameters being fit to both BOLD decoder errors and behavioral bias.

stage encoder-decoder model to describe observations in both early visual cortex and in behavior (see

265 modeling).

266 The encoding stage consists of cells with uniformly spaced von Mises tuning curves who's
267 amplitude is adapted by the identity of the previous stimulus (θ_{n-1}, Figure 4A). The decoding stage
268 reads out this activity using one of three strategies (Figure 4B). The *unaware* decoder assumes no
269 adaptation has taken place and results in stimulus likelihoods $p(m|\theta)$ that are repelled from the
270 previous stimulus (Figure 4C, yellow, where m is the population activity at the encoding stage). This
271 adaptation-naïve decoder is a previously hypothesized mechanism for behavioral adaptation (Seriès et
272 al., 2009) and likely captures the process that gives rise to the repulsive bias we observe in visual cortex
273 using a fMRI decoder that is agnostic to stimulus history (Figure 2G). Alternatively, the *aware* decoder
274 (Figure 4C, green) has perfect knowledge of the current state of adaptation and can thus account for
275 and 'un-do' biases introduced during encoding. Finally, the *over-aware* decoder knows the identity of
276 the previous stimulus but over-estimates the amount of gain modulation that takes place, resulting in a
277 net attraction to the previous stimulus (Figure 4C, red). We additionally built off of previous work
278 showing stimuli are generally stable across time by implementing a prior of temporal contiguity (van
279 Bergen and Jehee, 2019). In our implementation, a Bayesian prior centered on the previous stimulus
280 (Figure 4C, black) is multiplied by the decoded likelihood to get a Bayesian posterior (Figure 4C, bottom).
281 We applied this prior of temporal contiguity to both the *aware* decoder as well as the *unaware* decoder
282 to test the importance of awareness at decoding. We did not apply a prior to the *over-aware* model to
283 balance the number of free parameters between the various decoders and to see if the *over-aware*
284 model could achieve attractive serial dependence without a Bayesian prior (Figure 4, Supplementary
285 Table 1).

286 For each participant, we fit the encoder-decoder model in two steps (Figure 4D). All model
287 fitting was performed using the same cross-validation groups as our BOLD decoder and each stage had
288 two free parameters that were fit using grid-search and gradient descent techniques. We first report

289 results from the encoding stage of the model. The gain applied at encoding was adjusted to minimize

290 the residual sum of squared errors (RSS) between the output of the *unaware* decoder and the residual

291 errors of our BOLD decoder. The *unaware* readout of the adapted encoding process (Figure 5A, yellow)

292 provided a good fit to the average decoding errors obtained with the BOLD decoder (Figure 5A, black

293 outline, ρ=0.99) and across individual participants (Figure S8A, ranges: ρ= [0.84,0.98]). The *unaware*

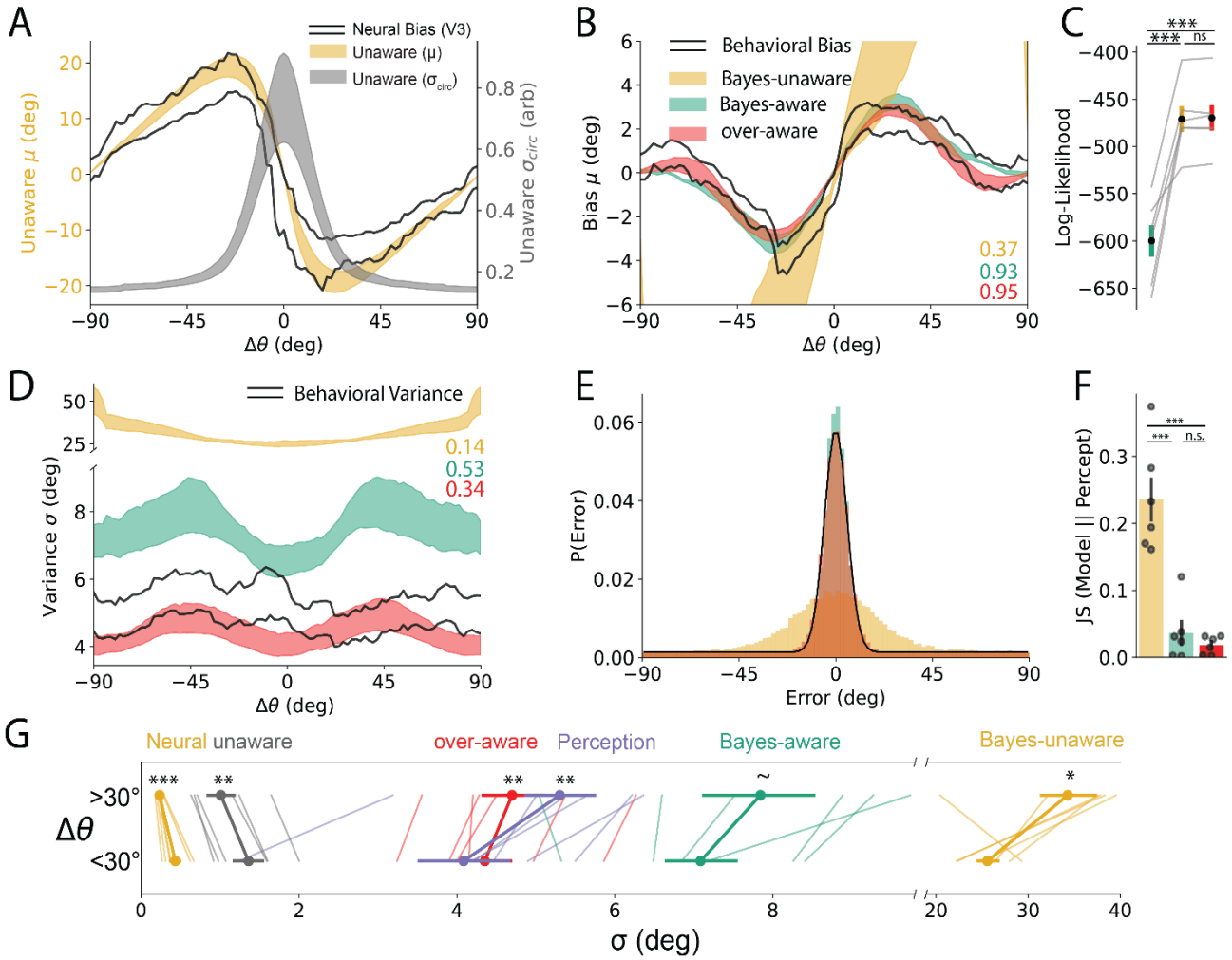

Figure 5 Model performance bias. **A-C** Neural/Behavioral Bias, **D-G** Neural/Behavioral Variance. **A**: Unaware decoder (yellow) provides a good fit to neural bias (black outline). Decoded variance decreases monotonically with distance from previous stimulus. (± SEM across participants). **B**: Perceptual bias (black outline) was well fit by the Bayes-aware and over-aware models but not the Bayes-Unaware model (± SEM across participants). **C:** Participant responses were significantly more likely under aware models. **D**: Behavioral variance had a similar shape and magnitude to Bayes-aware and over-aware model fits. Bayes-unaware model output was much less precise and had a different form. **E:** Distribution of empirically predicted response errors (black line) and simulated model fits for an example participant. **F:** The unaware model's error distribution had significantly higher Jenson-Shannon Divergence from BOLD decoder than either aware model. **F:** Visualization of all uncertainties split as a function of close and far stimuli. Note that the Bayes-unaware model had an average uncertainty that was on average 6x that of perception. *, p<.05; **, p<.01; ***, p<.001

readout provided a better fit to the outputs of our neural decoder than the null alternative of the presented orientation ($t(5)=3.41$, $p=.01$) because it captured a significant proportion of the variance in decoding errors as a function of $\Delta\theta$ ($t(5)=7.5$, $p=.0007$). This analysis demonstrates that our adaptation model does a reasonable job of recovering our empirical decoding data (both of which use a decoder unaware of sensory history).

We next considered three readout schemes of this adapted population to maximize the likelihood of our behavioral responses (Figure 5B). The *Bayes-aware* decoder is consistent with previous Bayesian accounts of serial dependence (van Bergen and Jehee, 2019), but additionally asserts that Bayesian inference occurs after encoding and that readout must account for adaptation. Alternatively, the *Bayes-unaware* decoder tests whether this awareness is necessary to achieve attractive serial dependence. Both aware models achieved biases that were significantly more likely than the unaware model ($t(5)=6.53$, $p=.001$, Bayes-aware; $t(5)=6.6$, $p=.001$, over-aware, t-test on log-likelihood, Figure 5C) but were indistinguishable from each other ($p=.36$). Thus, both aware models were able to explain the response biases while the unaware model did a relatively poor job, suggesting some awareness of the adapted state is necessary.

Finally, we examined the variance of our decoders to see if this mapped onto our empirically observed variance. As model coefficients were fit independent of observed variance, correspondence between model performance and BOLD/behavioral data would provide convergent support for the best model. While the models were trained using noiseless activity at encoding, we simulated responses using Poisson rates to induce response variability. We simulated 1000 trials from each cross-validated fit and pooled the model outputs. We first confirmed that the variance of the *unaware* decoder was highest following small changes of $\Delta\theta$ (Figure 5A, gray; Figure 5G $t(5)=3.93$, $p=.005$, paired 1-tailed t-test $<30°$ vs $>30°$) matching the output of our neural decoder (Figure 2G) and providing additional support for gain adaptation causing the observed repulsion in the fMRI data. Next, we compared the different

decoders and found that, matching real behavioral responses, all three decoders were more precise following small values of Δθ (Figure 5G, Bayes-unaware, $t(5)=2.25$, $p=.037$; Bayes-aware $t(5)=1.90$, $p=.058$; and over-aware $t(5)=5.43$, $p=.001$). While the pattern of the Bayes-unaware variance matched behavior, it's overall variance was much higher than our behavioral data such that it diverged from the behavioral data significantly more than either of the aware models (Figure 5E-G; ps<.005, paired t-test comparing Jenson-Shannon divergence of error distributions). Together, the variance data provides additional evidence in favor of adaptation driving the repulsive biases that were observed in the BOLD data and awareness of the current state of adaptation being a requisite condition for the observed attractive serial dependence. More generally, this model has notable advantages that can lead to enhanced discrimination, reduced energy usage, and improved discrimination in naturalistic conditions over a static labeled-line representation.

Discussion

In this study, we sought to understand the neural underpinning of attractive serial dependence, and how changes in tuning properties at encoding shape behavior. Based on previous behavioral and neural studies, we expected to observe attractive biases in line with observed behavior and decoding from early visual areas (St. John-Saaltink et al., 2016). Instead, we found that representations were significantly repelled from the previous stimulus starting in primary visual cortex and continuing through IPS0 (Figure 2I). This repulsion is consistent with bottom up adaptation beginning either at or before V1 and cascading up the visual hierarchy (Dhruv and Carandini, 2014; Patterson et al., 2014; Fritsche et al., 2022). As repulsive biases are in the opposite direction as behavioral biases, we built a model to link these conflicting patterns. The critical new insight revealed by the model is that only readout schemes that account for adaptation can explain the attractive behavioral bias observed in our paradigm. More generally, our BOLD data argue against an early sensory origin of serial dependence for orientation and

instead suggest that serial dependence is driven by post-perceptual or mnemonic circuits (Pascucci et al., 2019; Barbosa et al., 2020). However, because we used a paradigm that required working memory, our results may not generalize to other situations in which serial dependence is observed even in the absence of a memory delay (Fischer and Whitney, 2014; Cicchini et al., 2017; Collins, 2020; Murai and Whitney, 2021). Thus, future work is needed to better understand the role of sensory representations in paradigms with low contrast stimuli, that do not require a memory delay period, and that utilize other features besides orientation.

There have been many prior studies arguing for either a perceptual or post-perceptual origin of serial dependence. Some behavioral studies have found that serial dependence emerges almost immediately after the offset of a stimulus, pointing to an early perceptual origin of the effect (Fischer and Whitney, 2014; Cicchini et al., 2017, 2018, 2021). One study additionally demonstrated that attraction to the previous stimulus seems to occur before the 'tilt-illusion' driven by concurrently presented flanking stimuli (Cicchini et al., 2021). If history biases indeed operate before spatial context, this could point to a distinct assimilative mechanism for serial dependence in early visual processing which may only emerge under low stimulus drive. As our experiment always utilizes a working memory delay, it is unclear if the bias towards past stimuli is driven by a change in their perception of the stimulus itself or instead somehow biases their comparison with the probe stimulus only after the working memory maintenance period.

Others have found that serial dependence is repulsive at very short delays and only becomes attractive when items are held for an extended time in working memory (Papadimitriou et al., 2015; Bliss et al., 2017). This apparent discrepancy was reconciled by (Manassi et al., 2018), who showed that attractive biases disappear without a working memory delay, unless the stimuli are rendered at a very low contrast. This observation suggests that serial dependence may emerge immediately when high

364 sensory uncertainty is induced by low contrast stimuli, and it may emerge later if high sensory
365 uncertainty is induced by extended working memory delay periods. It is curious that unlike some spatial
366 working memory studies (Papadimitriou et al., 2015; Bliss et al., 2017; Stein et al., 2020), we did not find
367 that behavioral biases increased with delay time. One possible explanation is that this phenomenon is
368 actually unique to spatial working memory due to either a more consistent increase in sensory
369 uncertainty of spatial location due to eye movements or a separate mechanism of memory maintenance
370 that becomes more susceptible to proactive interference relative to orientation memories.

371 Evidence for an early sensory origin of serial dependence comes from an fMRI study with low
372 contrast stimuli and a short (500 ms) delay period which reported that both behavioral responses and
373 V1 representations were more precise following a matching stimulus (St. John-Saaltink et al., 2016). This
374 departure from our own finding could be driven by the stimuli that were rendered to have a very high
375 uncertainty. Past work studying adaptation in non-human primates found repulsive patterns following
376 long (4s and 40s) but attractive patterns following short (0.4s) stimulus presentations suggesting
377 stimulus duration may have a large influence on how past stimuli shape future sensory processing
378 (Patterson et al., 2013). That said, the stimuli used in the fMRI study of St. John-Saaltink et al (2016)
379 were always one of two orthogonal orientations, which, given a circular feature space like orientation,
380 precludes an assessment of attraction or repulsion. Furthermore, correct motor responses were directly
381 yoked to the stimulus so any behavioral tendency to report seeing the same stimulus on successive trials
382 could be due to motor priming rather than stimulus based serial dependence (e.g. a “stay” bias). Related
383 work has shown the ability to decode the previous stimulus from EEG activity patterns (Fornaciai and
384 Park, 2018; Bae and Luck, 2019; Bae, 2021) but it is important to note that our study also showed robust
385 decoding of the previous stimulus that did not also correspond with an attractive bias in the neural
386 representation of the current stimulus (Figure 2F-G, S4). This is because the representations of current
387 and past stimuli are not necessarily stored using the same code. Thus, while previous neural studies

388 have argued that serial dependence emerges in visual cortex, no study has demonstrated an attraction
389 towards the previous stimulus dependent on feature similarity consistent with behavioral biases.
390 Further work examining neural biases using low contrast stimuli will shed further light on a potential
391 role of coding changes in sensory cortex driving serial dependence.

392 In contrast to studies favoring an early sensory account – and more in line with the paradigm
393 and findings reported in this manuscript – a single unit recording study in non-human primates used
394 high contrast stimuli and a longer working memory delay (1.4-5.6s) (Papadimitriou et al., 2016). Under
395 these conditions, neural responses in the Frontal Eye Fields (FEF) were *repelled* from the previously
396 remembered location even though saccades were attracted to the previously remembered location.
397 Given the tight link between the FEF and attentional control (Moore and Fallah, 2001; Schall, 2004;
398 Moore and Zirnsak, 2017), the authors speculated that the observed neural repulsion was due to
399 residual attentional shifts carrying over from the previous trial. However, our observation of repulsive
400 biases starting in V1 and persisting across later visual areas suggests that bottom-up adaptation may be
401 a viable alternative explanation (which the authors also acknowledged). Further support for this account
402 comes from a recent MEG study showing that representations were repelled from past stimuli both
403 within the current trial and from the previous trial (Hajonides et al., 2021). As in our study, this neural
404 repulsion contrasts with attractive behavioral biases to the previous stimulus, suggesting sensory
405 representations do not directly shape behavior even in simple sensory paradigms (Siegle et al., 2021).
406 Behavioral studies using similar high contrast orientation stimuli to our own have also shown that
407 responses are attracted to past decisions and repelled from past stimuli, further suggesting these
408 attractive biases do not emerge in early sensory areas (Pascucci et al., 2019; Sadil et al., 2021). Several
409 modeling studies additionally suggest that serial biases are mediated by later readout circuits due to
410 synaptic changes arising from persistent bump attractor dynamics as opposed to early sensory
411 processing (Bliss and D'Esposito, 2017; Barbosa et al., 2020). Thus, in line with our findings: behavioral,

412 neuronal, and modeling studies utilizing high contrast stimuli in working memory paradigms consistently
413 point to attractive effects emerging in either memory or decision-making circuits and not early sensory
414 areas.

415 In line with classic accounts, adaptation in visual cortex should lead to a reduction in energy
416 usage during encoding (Clifford et al., 2000). However the main advantage of adaptation may be to
417 decorrelate inputs, thus enhancing the discriminability of incoming stimuli (Clifford et al., 2000, 2007)
418 and even acting as a form of short term memory (Hu et al., 2021). An optimal processing stream may
419 emphasize differences at encoding and only favor stability once a stimulus has been selected by
420 attention for more extensive post-perceptual processing (Pascucci et al., 2019). This motif of pattern
421 separation followed by pattern completion would not be unique to adaptive visual processing. Similar
422 mechanisms have been proposed as a critical component of long term memory processing in the
423 hippocampus and associative memory formation in the fly mushroom body (Cayco-Gajic and Silver,
424 2019). Thus, the biases introduced by adaptation may be beneficial in part because they expand the
425 dimensionality of the representational space as we found in our recordings (Figure S5).

426 We did not explicitly define how awareness of adaptation is implemented, but it is clear that
427 both attention to and conscious awareness of the previous stimulus are necessary for serial dependence
428 to occur (Fischer and Whitney, 2014; Kim et al., 2020). This is consistent with our model, and it suggests
429 that some representation of information about stimulus history should be a minimum requirement for
430 an aware decoding scheme. The identity of the previous stimulus for spatial position and angle has been
431 shown to be decodable from the spiking activity of single units in the frontal eye field (FEF) and large-
432 scale activity patterns in human EEG and MEG (Papadimitriou et al., 2016; Fornaciai and Park, 2018; Bae
433 and Luck, 2019; Bae, 2021; Hajonides et al., 2021). We additionally demonstrate that information about
434 the previous trial is encoded in patterns of fMRI activity in human visual cortex (Figure 2F), but not in a
435 sensory-like code (Figure S4AB). These signals could potentially be represented concurrently with

representations of the current stimulus in the same populations of sensory neurons but in orthogonal codes analogous to what has been found for sequentially encoded items in primate prefrontal cortex and human EEG (Wan et al., 2020; Xie et al., 2022). An alternate account holds that representations of stimulus history are maintained outside of early visual areas, consistent with findings from mouse parietal and primate prefrontal cortex (Akrami et al., 2018; Barbosa et al., 2020). This anatomical segregation could disambiguate incoming sensory drive from representations of stimulus history. Critically, optogenetically suppressing non-sensory representations of stimulus history eliminated history effects, thus providing strong support for some form of an aware readout mechanism (Akrami et al., 2018).

For the decoding stage of our model, we established that only readout schemes that are aware of adaptation can explain attractive serial dependence. The *Bayes-aware* model is an extension of previously proposed models that employ an explicit prior but that did not consider effects of adaptation at encoding (van Bergen and Jehee, 2019). In contrast, the *over-aware* model is a novel account that can achieve similar performance without needing an explicit prior based on stimulus history. While model fit metrics did not readily distinguish one of these two models as superior, the *over-aware* model may prove to be more flexible. For instance, one of our fMRI participants showed significant repulsion from far stimuli, an observation also reported by others (Bliss et al., 2017; Fritsche et al., 2017). While the *over-aware* model can fit this repulsive regime, the *Bayes-aware* model is incapable of generating repulsive patterns (compare models fits for subj #3, Figure 9). This limitation of a purely Bayesian account of serial dependence is also observable in prior work (Figure 6B in van Bergen and Jehee, 2019).

The *over-aware* model proposed in our study may instead be a special condition of a decoder with “fixed awareness” that is based on temporal transition probabilities in natural scenes that are steeply peaked around 0 (no change) over short time scales (Dong and Atick, 1995; Felsen et al., 2005; van Bergen and Jehee, 2019). Such a readout would correct for the most encountered levels of

adaptation by accounting for the transition probabilities of stimuli while being 'fixed', or inflexible, when stimuli violate these expectations. This decoder could account for additional phenomena not directly assessed in the present study such as the tilt after-effect (TAE). The TAE and other forms of (repulsive) behavioral adaptation are often ascribed to an unaware decoder (Seriès et al., 2009; Benucci et al., 2013) but might instead reflect levels of adaptation that exceed the fixed level of adaptation expected by a "fixed-aware" decoder due to long presentations or high contrast stimuli. This is supported by an apparent disconnect in the magnitude of repulsive biases between behavior and neural representations (Dragoi et al., 2001; He and MacLeod, 2001). In contrast, the fixed awareness decoder would lead to attractive biases (serial dependence) when stimuli create less bottom-up drive than expected (e.g., through brief presentations or low contrast items). This 'fixed-aware' decoder is consistent with previous findings of attractive biases disappearing or switching to repulsion when stimulus contrast or duration is increased (Fischer and Whitney, 2014; Manassi et al., 2018). This scheme could extend to spatial adaptation such as the tilt-illusion where the joint probability of center and surround orientations being perfectly distinct would be vanishingly rare in natural scenes (Howe and Purves, 2005; Schwartz et al., 2007, 2009).

In this study, we extended previous descriptions of serial dependence by quantifying how both bias and variance are shaped by stimulus history. We report a robust pattern of perception being most precise following small changes in successive stimulus features (Figure 1F-G, 2A-B). This relationship violates a proposed perceptual 'law' that bias is inversely proportional to the derivative of discrimination thresholds (Wei and Stocker, 2017). This account would assert that our attractive bias should come with a less precise representation following small changes (or a repulsive bias to account for our enhanced precision). We argue that serial dependence is not violating this law, but rather believe this is further evidence for delay dependent serial dependence being a post-sensory phenomenon. Neural representations exhibit repulsive biases, expanding the perceptual space and allowing greater

discriminability (Figure S5). When these representations are read out by an aware decoder, the bias is undone but the enhanced discriminability remains (Figure 5D,G).

Methods

Participants

Behavioral study: 56 participants (male and female) were drawn from a subject pool of primarily undergraduate students at UC San Diego. All participants gave written consent to participate in the study in accordance with the UC San Diego IRB (approval number 180067) and were compensated either monetarily or with class credit. Of these 56 participants, 9 were removed from further analysis for completing less than 200 trials (2) or getting less than 60% of trials correct (7). We included the remaining 47 participants who completed on average 421 trials, range: [204, 988], in our lab over the course of 1 to 3 sessions.

fMRI study: 6 participants (3 female, mean age 24.6 ±0.92) participated in four, 2-hour scanning sessions. Each subject completed between 748 and 884 trials (mean 838.7). For two participants, one session had to be repeated due to technical difficulties that arose during scanning.

Behavioral Discrimination Task

Participants in the behavior-only study completed the task on a desktop computer in a sound attenuated room. Subjects were seated with a chin rest to stabilize viewing 50 cm from a 39 by 29 cm CRT monitor (1600x1200 px) with a visual angle of 42.6° (screen width). Each trial consisted of a full-field oriented grating (1000 ms) which had to be remembered across a delay period (3,500 ms) before a test. At test, the participant judged whether a line was slightly clockwise (CW) or counter-clockwise (CCW) relative to the remembered orientation (max response time window: 3,000ms, Figure 1A). The oriented grating consisted of a sine wave grating (spatial frequency 1.73 cycles/°, 0.8 Michelson contrast) multiplied by a 'donut' mask (outer diameter Ø=24.3°, inner Ø=1.73°). The stimulus was then convolved

with a 2D Gaussian filter (1.16° kernel, SD = 0.58°) to minimize edge artifacts (Roth et al., 2018). Phase and orientation were randomized across trials, and the stimulus was phase-reversed every 250ms. After the offset of the oriented grating, a mask of filtered noise was presented for 500ms. The mask was generated by band passing white noise [low 0.22, high 0.87 cycles/°], multiplying by the same donut mask, and convolving with a 2D Gaussian filter (0.27° kernel, SD = 0.11°). The mask was phase reversed once after 250 ms. A black fixation point (diameter .578°) was displayed throughout the extent of the block and turned white for 500 ms prior to stimulus onset on each trial. The probe was a white line (width 0.03°, length 24.3°) masked by the same donut. Subjects indicated whether the probe line was CW or CCW from the remembered orientation by pressing one of two buttons ('Q', 'P') with their left and right pointer fingers. The next trial started after a 1000ms inter trial interval (ITI). For some behavioral participants (n=9) delay and ITI were varied between 0.5-7.5s without notable effects on performance.

First, subjects completed a training block to ensure that they understood the task. Next, they completed a block of trials where difficulty was adjusted by changing the probe offset ($\delta\theta$) between the stimulus and probe to achieve 70% accuracy. This $\delta\theta$ was used in subsequent blocks and was adjusted on a per-block basis to keep performance at approximately 70%. Participants completed an average of 5.76±0.24 blocks [min = 3, max = 9]. Some participants completed the task with slight variations in the distribution and sequence of orientations presented. For completeness we include those details here. Note, however, we additionally report a set of control analyses in which we repeat all of our main analyses excluding blocks with binned stimuli and find no relevant difference in behavior. For most participants, stimuli were pseudo-randomly distributed across the entire 180° space such that they were uniformly distributed across blocks of 64 trials (n=25). However, some participants saw stimuli that were binned (with some jitter) every 22.5° to purposefully avoid cardinal and oblique orientations (11.25°, 33.75°, 56.25°, etc.) and the trial sequence was ordered so that a near oblique orientation was always

531 followed by a near cardinal orientation (n=7). This was implemented to maximize our ability to observe
532 serial dependencies in our binary response data as it is typically strongest around orientation changes of
533 20° and is more pronounced around oblique orientations (Cicchini and Burr, 2018). The remaining
534 participants completed both blocks with uniform and blocks with binned stimuli (n=14). All participants
535 were interviewed after the study and reported that stimuli were non-predictable and that all
536 orientations felt equally likely. For our main analysis we include all trials from all participants,
537 irrespective of whether they participated in uniform blocks, binned blocks, or both.

538 ## fMRI Discrimination Task

539 In the scanner, participants completed the behavioral task outlined above with slight
540 modifications. fMRI participants completed the task using a fiber-optic button box while viewing stimuli
541 through a mirror projected onto a screen mounted inside of the bore. The screen was 24 by 18 cm and
542 was viewed at a distance of 47 cm (width: 28.6° visual angle; 1024x768 px native resolution). The
543 stimulus timing was the same except that the sample-to-probe delay period was either 5, 7 or 9 s and
544 the ITIs were uniformly spaced between 5s and 9s and shuffled pseudo-randomly on each run of 17
545 trials. The oriented gratings had a spatial frequency of 1.27 cycles/°, outer Ø=21.2°, inner Ø=2.37° and
546 were smoothed by a Gaussian filter (0.79° kernel, sd=0.79°). The noise patch (SF low 0.16, high 0.63
547 cycles/°) was also smoothed by a Gaussian filter (0.29° kernel, sd=0.11°). The probe stimulus was a white
548 line (width = 0.03°).

549 fMRI participants completed 44-52 blocks of 17 trials spread across 4, two-hour scanning
550 sessions for a total of 748-884 trials. As in the behavior-only task described above, 4 out of 6 fMRI
551 subjects had some blocks of trials where the stimuli were binned in 22.5° increments and ordered in a
552 non-independent manner (21-24 blocks/participant). However, all of the fMRI subjects also participated
553 in blocks with a uniform distribution of orientations across the entire 180° space (24-52

554 blocks/participant). For our main analysis we include all trials from all participants. However, as with the
555 behavioral analyses, we also report control analyses in which we repeat all of our main analyses
556 excluding blocks with non-random stimuli.

557 ## fMRI Localizer Task

558 Interleaved between the main task blocks, participants completed an independent localizer task
559 used for voxel selection where they were presented with a sequence of grating stimuli at different
560 orientations. Stimuli had a pseudo-randomly determined orientation that either matched the spatial
561 location occupied by the *donut* stimuli used in our main task (outer diameter Ø=21.2°, inner diameter
562 Ø=2.37°) or were a smaller foveal oriented Gabor corresponding to the 'hole' in the *donut* stimuli
563 (diameter Ø=2.37°). Participants were instructed to attend to one of three features orthogonal to
564 orientation depending on the block: detect a contrast change across the entire stimulus, detect a small
565 grey blob appearing over part of the stimulus, or detect a small change in contrast at the fixation point.
566 Each stimulus was presented for 6000 ms and was separated by an ITI ranging from 3-8s.

567 ## Response Bias

568 Each trial consisted of a stimulus and a probe separated by a probe offset ($\delta\theta$) that was either
569 positive (probe is CW of stimulus) or negative. We report degrees in a compass-based coordinate system
570 such that 0° is vertical and orientation values increase moving CW (eg. 30° would point towards 1 o-
571 clock). Participants judged whether the probe was CW or CCW relative to the remembered orientation
572 by making a binary response. To quantify the precision and the response bias, we fit participant
573 responses with a Gaussian cumulative density function with parameters μ and σ corresponding to the
574 *bias (mean)* and *standard deviation* of the distribution. The likelihood of a given distribution was
575 determined by the area under the curve (AUC) of the distribution of CW (CCW) offsets between the
576 stimulus and the probe ($\delta\theta$) on trials where the participant responded CW (CCW). In extreme cases, a

very low standard deviation (σ) value with no bias would mean that all $\delta\theta$ would lie outside the distribution and the participant would get every trial correct. A high negative bias (μ) value would mean that $\delta\theta$ would always lie CW relative to the distribution and the participant would respond CW on every trial. The best fitting parameters were found using a bounded minimization algorithm (limited memory BFGS) on the negative log likelihood of the resulting responses (excluded the small number of trials without a response) given the generated distribution (SciPy 1.0 Contributors et al., 2020). We included a constant 25% guess rate in all model fits to ensure the likelihood of any response could never be 0 (critical for later modelling). While this was critical to fitting our model to raw data, the specific choice had no qualitative effect on our behavioral findings besides making the σ values smaller compared to having a 0% guess rate. By having a constant guess rate rather than varying it as a free parameter we were able to directly compare σ values across participants as a measure of performance.

Serial Dependence

To quantify the dependence of responses on previous stimuli, we analyzed response bias and variance as a function of the difference in orientation between the previous and current orientation ($\Delta\theta = \theta_{n-1} - \theta_n$). We performed this analysis using a sliding window of 32°, such that a bias centered on 16° would include all trials with a $\Delta\theta$ in the range [0°, 32°].

We additionally fit a Derivative of Gaussian (DoG) function to parameterize the bias of participant responses. The DoG function is parameterized with an amplitude A and width w

$$y = xAwce^{-(wx)^2} \quad [1]$$

where $c = \sqrt{2e}$ is a normalization constant. For the purpose of fitting to our participant responses, x is $\Delta\theta$ and y corresponds to μ in our response model. For each participant we adjusted three parameters: A, w, and σ to maximize the likelihood of participant responses. We report the magnitude of our fits as well as the resulting full width at half max (FWHM) estimated numerically.

Response Precision

In addition to quantifying how responses were biased as a function of stimulus history, we also estimated how precise responses were depending on their unsigned distance from the previous stimulus ($|\Delta\theta|$). When quantifying variance difference between close and far trials, we 'folded' trials with $\Delta\theta < 0$ so that the bias would generally point in the same direction and not artificially inflate our variance measure. Values from the bin with more samples (typically 'far') were resampled (31 repetitions) without replacement with the number of samples in the smaller bin and the median chosen to control for sample number differences.

Scanning

fMRI task images were acquired over the course of four 2-hour sessions for each participant in a General Electric Discovery MR750 3.0T scanner at the UC San Diego Keck Center for Functional Magnetic Resonance Imaging. Functional echo-planar imaging (EPI) data were acquired using a Nova Medical 32-channel head coil (NMSC075-32- 3GE-MR750) and the Stanford Simultaneous Multi-Slice (SMS) EPI sequence (MUX EPI), with a multiband factor of 8 and 9 axial slices per band (total slices 72; 2-mm3 isotropic; 0-mm gap; matrix 104 x 104; field of view 20.8 cm; TR/TE 800/35 ms; flip angle 52°; in-plane acceleration 1). Image reconstruction and un-aliasing was performed on cloud-based servers using reconstruction code from the Center for Neural Imaging at Stanford. The initial 16 repetition times (TRs) collected at sequence onset served as reference images required for the transformation from k-space to the image space. Two 17s runs traversing k-space using forward and reverse phase-encoding directions were collected in the middle of each scanning session and were used to correct for distortions in EPI sequences using FSL top-up (FMRIB Software Library) for all runs in that session (Andersson et al. 2013, Jenkinson et al. 2012). Reconstructed data was motion corrected and aligned to a common image. Voxel data from each run was de-trended (8TR filter) and z-scored.

We also acquired one additional high-resolution anatomical scan for each subject (1 x 1 x 1-mm3 voxel size; TR 8,136 ms; TE 3,172 ms; flip angle 8°; 172 slices; 1-mm slice gap; 256x192-cm matrix size) during a separate retinotopic mapping session using an Invivo eight-channel head coil. This scan produced higher quality contrast between gray and white matter and was used for segmentation, flattening, and visualizing retinotopic mapping data. The functional retinotopic mapping scanning was collected using the 32-channel coil described above and featured runs where participants viewed checkerboard gratings while responding to an orthogonal feature (transient contrast changes). Separate runs featured alternating vertical and horizontal bowtie stimuli; rotating wedges; and an expanding donut to generate retinotopic maps of the visual meridian, polar angle, and eccentricity respectively (see Sprague and Serences, 2013). These images were processed using FreeSurfer and FSL functions and visual regions of interest (ROI) were manually drawn on surface reconstructions (for areas: V1-V3, V3AB, hV4, and IPS0).

Voxel Selection

To include only voxels that showed selectivity for the location of the oriented grating stimulus used in our main experimental task, we used responses evoked during the independent *localizer* task (see fMRI Localizer Task). For all analyses we used TRs 5-11 (4-8.8s) following stimulus onset. First, voxels were selected based on their response to the spatial location of the grating stimulus by performing a t-test on the responses of each voxel evoked by the donut and the donut-hole stimuli, selecting the 50% of the voxels most selective to the donut for a given ROI. Of the voxels that passed this cutoff, we then performed an ANOVA across 10° orientation bins and selected the 50% of voxels with the largest F-score thus retaining ~25% of the initial voxel pool. These selected voxels were used in all main analyses.

Orientation Decoding

We performed orientation decoding by training an inverted encoding model (IEM) (Brouwer and Heeger, 2009) on BOLD activation patterns using a sliding temporal window of 4 TRs. For most analyses we focused on a 3.2s (4 TR) window centered 6.4 s after stimulus presentation. We first designed an encoding model which assumes voxels are composed of populations of neurons with tuning functions centered on one of 8 orientations evenly tiling the 180° space. The response of population i to stimulus θ is given by:

$$c_i(\theta) = \max\,(0, cos^5(\theta - \omega_i)\,) \qquad [2]$$

where ω_i is the center of the tuning function. The response of voxel j is defined as a weighted sum of these hypothetical populations:

$$B_j = \sum_{i}^{8} c_i w_i \qquad [3]$$

Or in matrix notation,

$$B = CW \qquad [4]$$

Where B *(trial x voxel)* is the resulting BOLD activity, C *(trial x channel)* is the hypothetical population response, and W *(channel x voxel)* is the weight matrix. The weight matrix W is estimated as:

$$\widehat{W} = C^{-1}B \qquad [5]$$

where C^{-1} (*channel x trial*) is the pseudo-inverse of C (implemented using the NumPy pinv function). We then estimated channel responses using the inverse of our estimated weight matrix:

$$\hat{C} = B\widehat{W}^{-1} \qquad [6]$$

This channel response corresponds to a representation of orientation activity. To decode orientation, we took the inner product with a vector of the tuning curve centers in polar coordinates.

660 The angle of the resulting vector was taken as the estimated orientation ($\hat{\theta}$) while the vector length was

661 taken as a proxy for model certainty ($\hat{R}$).

$$\hat{\theta} = angle(\hat{C}e^{i\omega}) \quad [7]$$

$$\hat{R} = \|\hat{C}e^{i\omega}\| \quad [8]$$

662 The weight matrix of our model was estimated from a subset of our data and used to estimate

663 orientation representations on a held-out portion of the task data. We used leave-1-block-out cross-

664 validation where each block was a set of 4 consecutive runs (64 trials). These blocks had orientations

665 that were linearly spaced across the entire 180°, with a random phase offset for each block, to ensure a

666 balanced training set. We performed an additional analysis training a model on all data from the

667 localizer task and testing on the memory task. This model had lower SNR than models trained on the

668 task but showed qualitatively similar results as our task trained neural decoder.

669 Kernel Based Decoding

670 *Estimating average voxel HRFs through deconvolution*

671

672 Because we are measuring the effects of previous stimuli on responses to the current stimulus,

673 we did an additional analysis to quantify any influence of overlapping Hemodynamic Response Functions

674 (HRF) that last for 20-30s (e.g. the "undershoot" that happens approximately 8-18s post-stimulus; see

675 Figure 3A). To account for overlapping HRFs, we used deconvolution to estimate the average univariate

676 response separately in each voxel in each ROI by modeling the responses to both the stimulus and probe

677 for 30 TRs (24s) post-stimulus (Dale, 1999; Glover, 1999). We created a design matrix (rows x columns =

678 total number of TRs x 30) with the first column containing ones corresponding to the onset TR of each

679 stimulus (and zeros elsewhere). Subsequent columns were the same vector shifted forward in time by

680 one TR. Following the same procedure, another design matrix was defined for the probe onset times.

681 These matrices were stacked with a column of ones added for each run as a constant term, yielding a

682 final design matrix X of dimensions (total number of TRs x (60+number of blocks)). We created a related

683 matrix of voxel activity Y (total number of TRs x number of voxels) by concatenating responses in each
684 voxel across blocks. We then estimated the HRF by performing least squares regression using the normal
685 equation:

$$h = (X^T X)^{-1}(X^T Y). \quad [9]$$

686 The resulting weights corresponded to the average timecourse of the HRF evoked separately by the
687 stimulus and the probe across all trials. We note that this HRF is estimated independent of the
688 orientation of the presented stimuli as we wanted to use these estimates to then decode orientation
689 dependent changes in activation patterns. For each voxel we then parameterized the HRF using a 6-
690 parameter double gamma function using scipy.optimize.minimize so that we could use the voxel-specific
691 HRF model in a GLM to estimate the response magnitude in each voxel on each trial. We excluded the
692 11% of voxels which failed to converge on a solution.

693 *Estimating trial-by-trial responses using parameterized voxel HRFs*
694

695 For each voxel, we then created a design matrix Xv (rows x columns = total number of TRs x
696 (number of trials * 2 +number of blocks)) with each column a delta function centered at the onset of the
697 stimulus (or probe). We then regressed this matrix onto the (total # of TRs) vector Yv of voxel activity
698 using equation 9. This resulted in a simultaneous estimation of the trial-by-trial magnitude of responses
699 to each stimulus grating and each probe which was repeated for each voxel to allow voxel specific HRFs
700 to be utilized in the creation of Xv. The resulting activity pattern associated with each stimulus was used
701 in the same manner as the raw time course of the BOLD response to train and test an IEM, and the
702 resulting estimates should be largely independent of linear contributions of previous stimuli (Dale,
703 1999).

Neural Bias

To quantify how BOLD representations were biased by sensory history we computed the circular mean of decoding errors (θ_{error} = wrap($\theta_{decode} - \theta_{stim}$)):

$$\mu_{circ} = angle(\vec{R}), \quad [10]$$

$$\vec{R} = \frac{1}{nTrials}\sum_{k=0}^{nTrials} e^{i\theta_{error}^{k}}. \quad [11]$$

We estimated this bias using the same 32° sliding window as a function of $\Delta\theta$ used for visualizing response bias from participant responses. We additionally quantified the magnitude of the bias in decoding errors by fitting a DoG function to the raw decoding errors by minimizing the residual sum of squares (RSS) and reporting the amplitude term.

Neural Variance

To quantify the variance of decoded orientations from visual areas, we computed the circular standard deviation on binned decoding errors:

$$\sigma_{circ} = \sqrt{-2\ln|\vec{R}|}. \quad [12]$$

This was visualized using the same sliding window analysis as well as in reference to whether it was close or far from the previous stimulus.

Dimensionality Analysis

To quantify how stimulus history shaped the structure of neural responses independent of neural tuning we utilized principal component analysis (PCA). For a given set of neural responses R (number of trials x number of voxels) we mean centered and performed eigenvalue decomposition on the (number of voxels x number of voxels) covariance matrix. Eigenvalues were sorted in descending order and our response matrix was projected into PCA space (for visualization purposes) by multiplying by the sorted eigenvectors.

To compare dimensionality across conditions, we sub-set our data into trials following close (<30°) or far (>60°) trials and randomly sub selected trials from the larger group (without replacement) to equate trial numbers. We then performed PCA separately for each group and compared the relative proportion of total variance explained as the magnitude of the sorted eigenvalues. We quantified both the minimum number of components to reach at least 90% of the variance explained and also recorded the mean (AUC) of the variance curve.

Modeling

We sought to develop a model that could explain both neural and behavioral biases as a function of stimulus history. For the fMRI data, we focused on explaining changes in encoding that could lead to the observed biases in the output of the BOLD decoder that was specifically designed to be '*unaware'* of stimulus history. To explain the behavioral data, we assumed that a decoder would receive inputs from the same population of sensory neurons that we measured with fMRI and that the decoder would read out this information in a manner that gives rise to attractive serial dependence. We considered readout models that were either *unaware*, *aware*, or *over-aware* of adaptation and additionally applied a Bayesian inference stage, which integrates prior expectations of temporal stability, to the *unaware* and *aware* decoders (van Bergen and Jehee, 2019). We then compared performance between these competing models to see which could best explain our behavioral data.

Our full models consisted of two stages: an encoding stage where the gain of artificial neurons was changed as a function of the previous stimulus (adaptation) and a decoding stage where the readout from this adapted population was modified. The encoding population consisted of 100 neurons with von Mises tuning curves evenly tiling the 180° space. The expected unadapted population response is:

$$Resp_N(\theta_n) = R\,\gamma_N e^{\kappa\cos(\Phi-\theta_n)-1} \qquad [13]$$

Where γ_N is the scalar 1 for constant gain without adaptation, Φ is the vector of tuning curve centers, θ_n is the orientation of the current stimulus, κ=1.0 is a constant controlling tuning width, and R is a general gain factor driving the average firing rate. We implemented sensory adaptation by adjusting the gain of tuning curves relative to the identity of the previous stimulus, θ_{n-1} (Figure 4A, *Gain Adaptation*):

$$\gamma_A(\theta_{n-1}) = \gamma_N - rect(\gamma_m \cos^3(\gamma_s(\Phi - \theta_{n-1}))) \quad [14]$$

Where γ_m is the magnitude of adaptation, γ_s scales the width of adaptation, and *rect* is the half-wave rectifying function. The responses of the adapted population thus depend on both the current and previous stimulus (Figure 4A, *Efficient Encoding*):

$$Resp_A(\theta_n, \theta_{n-1}) = R\ \gamma_A e^{\kappa \cos(\Phi - \theta n) - 1} \quad [15]$$

Unaware decoder: We first considered a model in which an adapted orientation-encoding representation is being decoded by an *unaware* readout mechanism (Figure 7B). The likelihood of each orientation giving rise to the observed response profile across *N* neurons was estimated assuming activity was governed by a Poisson process:

$$P_{unaware}(\ Resp_A|\theta) = \exp\left(\sum_{i=1}^{N} \log P_{Poisson}\left(Resp_A^i(\theta); Resp_N^i(\theta)\right)\right) \quad [16]$$

$$P_{Poisson}(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad [17]$$

Where $Resp_N^i(\theta)$ is the expected response of the unadapted neuron i to stimulus θ and $P_{Poisson}(k;\lambda)$ is the probability of observing k spikes given an expected firing rate of λ. The decoded orientation is then the θ giving rise to the maximum likelihood (MLE).

Aware decoder: In addition to the unaware decoder, we also evaluated the ability of a decoder that was aware of the current state of adaptation to explain behavior. The *aware* decoder differs from the

unaware decoder in that its assumed activity level for each unit is modulated as a function of stimulus history:

$$P_{aware}(\,Resp_A|\theta_n;\theta_{n-1}) = \exp\left(\sum_{i=1}^{N}\log P_{Poisson}\left(Resp_A^i(\theta_n,\theta_{n-1}),Resp_A^i(\theta_n,\theta_{n-1})\right)\right) \quad [18]$$

Note that here the rate parameter $k \equiv \lambda \equiv Resp_A$ such that the observed and expected values perfectly align with the presented orientation. $P_{aware}(\,Resp_A|\theta_n;\theta_{n-1})$ is dependent on sensory history and is non-biased.

Over-Aware decoder: Our final decoding scheme we call the *over-aware decoder*. This model can test whether serial dependence can be achieved without an explicit stage of Bayesian inference introduced in the next section. The decoder has an assumed adaptation defined by a unique set of free parameters γ_{m2} *and* γ_{s2} which shapes a separate gain adaptation:

$$\gamma_{OA}(\theta_{n-1}) = \gamma_N - rect(\gamma_{m2}\cos^3(\gamma_{s2}(\Phi - \theta_{n-1}))) \quad [19]$$

which in turn shapes the response profile of $Resp_{OA}$ in the same manner as $Resp_A$. The likelihood profile is then defined as:

$$P_{over-aware}(\,Resp_A|\theta) = \exp\left(\sum_{i=1}^{N}\log P_{Poisson}\left(Resp_A^i(\theta);Resp_{OA}^i(\theta,\theta_{n-1})\right)\right) \quad [20]$$

where our expected (assumed) rate λ is designated by $Resp_{OA}$. By having a larger assumed adaptation than implemented at encoding (through either $\gamma_{m2} > \gamma_m$ or $\gamma_{s2} > \gamma_s$) the net effect of the over-aware decoder should be behavioral attraction.

Bayesian Inference: In addition, we explored the effect of applying an explicit Bayesian prior based on temporal contiguity to the likelihood functions derived from these different readout schemes. This type

779 of prior has been previously used to explain behavioral biases without considering how encoding might
780 also be affected by stimulus history (van Bergen and Jehee, 2019). Specifically, the prior is defined by
781 the transition probability between consecutive stimuli and is defined as a mixture model of a circular
782 Gaussian and a uniform distribution:

$$P_T(\theta_n|\theta_{n-1}) = \frac{1}{Z} e^{-\frac{angle(\theta,\theta_{n-1})^2}{2\psi^2}} \quad [21]$$

$$P_{Bayesian}(\theta_n|\theta_{n-1}) = P_{SAME} P_T(\theta|\theta_{n-1}) + \frac{1}{2\pi}(1 - P_{SAME}) \quad [22]$$

783 with P_{SAME} set to 0.64 (as found empirically in (van Bergen and Jehee, 2019)), Z as a normalization
784 constant so P_T integrates to 1, and ψ is a free parameter describing the variance of the transition
785 distribution. This prior (Figure 4C, black line) is multiplied by the *unaware* likelihood (Figure 4C, yellow
786 dashed-line): to get the posterior estimate of our *Bayesian-unaware* decoder (Figure 4C, yellow solid-
787 line):

$$P_{Bayesian-unaware}(\theta_n|Resp_A;\ \theta_{n-1}) = P_{Bayesian}(\theta|\theta_{n-1}) P_{unaware}(Resp_A|\theta_n) \quad [23]$$

788 We can additionally examine a *Bayesian-aware* decoder by substituting its respective likelihood
789 function. We did not examine a *Bayesian-over-aware* model so that all decoding models would have the
790 same number of free parameters and so that we could directly evaluate the need for an explicit prior.

791 *Model Fitting:* The *encoding* stage of the model has two free parameters and for each subject these
792 parameters were optimized to minimize the residual sum of squares (RSS) between our measured fMRI
793 decoding errors and the decoding errors of our *unaware* decoder. For simplicity we only fit our model to
794 decoding errors from V3 as it had the highest SNR, but other early visual ROIs showed similar results.
795 After fitting the *encoding* stage of the model, we then separately fit the three competing *decoding*
796 models to best account for the behavioral data: *Bayes-unaware, Bayes-aware,* and *over-aware* (two free
797 parameters each). The output of this readout stage was treated as the behavioral bias (μ) and the free

parameters were optimized to maximize the likelihood of the observed responses (assuming constant standard deviation σ estimated empirically for each participant). For the purposes of fitting the model, the firing rates of the modelled neurons were deterministic (no noise process). Having noiseless activity had no effect on the expected bias (verified with additional simulations) and served to make model fitting more reliable and less computationally intensive. Both stages of the model were fit using the same cross-validation groups as our neural decoder. To ensure all models had a sufficient chance of achieving a good fit to behavioral data, we implemented a grid search sampling 30 values along the range of each variable explored (900 locations total) followed by a local search algorithm (Nelder-Mead) around the most successful grid point. We found dense sampling of the initial parameter space was especially important for our *Bayes-unaware* model.

Model Evaluation: For bias of neural and behavioral responses, we evaluated the performance of the two stages of our model separately. These stages must be evaluated in a qualitatively different manner as the neural data gives us an orientation estimate for each trial while the behavioral data consists of binary responses. For the encoding stage, we quantified how well the output of our *unaware* decoder predicted the raw errors of our BOLD decoder using circular correlation. The performance of this model was contrasted with the true presented orientation which is analogous to the representation of an unadapted population. We additionally computed the variance of the neural decoding errors explained by the model bias (R^2). For the decoding stage of our model, we compared the log-likelihood of observed responses for each model.

We additionally estimated the variance of our models using neurons with rates generated by a Poisson process. The average bias was unaffected by allowing random fluctuations in activity, but the trial-to-trial variance increased. To get a stable estimate, we simulated 1000 trials for each set of parameters estimated for a cross-validation loop for each participant and pooled these outputs. We compared the overall variance of our models to our single parameter estimate of participant precision

using Jensen-Shannon divergence. We additionally examined relative precision of our model for close and far trials in the same manner as participant responses and decoding errors (Response Precision).

Data/Code Availability (upon acceptance for publication)

Code for processing raw data as well as for analyzing decoded representations can be found here (GITHUB). This includes all processing performed on our BOLD data and our implementation of the Bayesian decoder in Python code for running our model as well as the data used to fit models can be found here (GITHUB). Data used in this study will be posted on the first author's Open Science Framework repository (REPOSITORY LINK).

Works Cited

Abbonizio G, Clifford C, Langley K (2002) Contrast adaptation may enhance contrast discrimination. Spat Vis 16:45–58.

Akrami A, Kopec CD, Diamond ME, Brody CD (2018) Posterior parietal cortex represents sensory history and mediates its effects on behaviour. Nature 554:368–372.

Bae G-Y (2021) The Time Course of Face Representations during Perception and Working Memory Maintenance. Cereb Cortex Commun 2:tgaa093.

Bae G-Y, Luck SJ (2019) Reactivation of Previous Experiences in a Working Memory Task. Psychol Sci 30:587–595.

Barbosa J, Stein H, Martinez RL, Galan-Gadea A, Li S, Dalmau J, Adam KCS, Valls-Solé J, Constantinidis C, Compte A (2020) Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory. Nat Neurosci 23:1016–1024.

Barlow HB (2012) Possible Principles Underlying the Transformations of Sensory Messages. In: Sensory Communication (Rosenblith WA, ed), pp 216–234. The MIT Press. Available at: http://mitpress.universitypressscholarship.com/view/10.7551/mitpress/9780262518420.001.0001/upso-9780262518420-chapter-13 [Accessed August 29, 2020].

Benucci A, Saleem AB, Carandini M (2013) Adaptation maintains population homeostasis in primary visual cortex. Nat Neurosci 16:724–729.

Bliss DP, D'Esposito M (2017) Synaptic augmentation in a cortical circuit model reproduces serial dependence in visual working memory Bazhenov M, ed. PLOS ONE 12:e0188927.

Bliss DP, Sun JJ, D'Esposito M (2017) Serial dependence is absent at the time of perception but increases in visual working memory. Sci Rep 7:14739.

Brouwer GJ, Heeger DJ (2009) Decoding and Reconstructing Color from Responses in Human Visual Cortex. J Neurosci 29:13992–14003.

Cayco-Gajic NA, Silver RA (2019) Re-evaluating Circuit Mechanisms Underlying Pattern Separation. Neuron 101:584–602.

Cicchini GM, Anobile G, Burr DC (2014) Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. Proc Natl Acad Sci 111:7867–7872.

Cicchini GM, Benedetto A, Burr DC (2021) Perceptual history propagates down to early levels of sensory analysis. Curr Biol 31:1245-1250.e2.

Cicchini GM, Burr DC (2018) Serial effects are optimal. Behav Brain Sci 41:e229.

Cicchini GM, Mikellidou K, Burr D (2017) Serial dependencies act directly on perception. J Vis 17:6.

Cicchini GM, Mikellidou K, Burr DC (2018) The functional role of serial dependence. Proc R Soc B Biol Sci 285:20181722.

Clifford CWG, Webster MA, Stanley GB, Stocker AA, Kohn A, Sharpee TO, Schwartz O (2007) Visual adaptation: Neural, psychological and computational aspects. Vision Res 47:3125–3131.

Clifford CWG, Wenderoth P, Spehar B (2000) A functional angle on some after-effects in cortical vision. Proc R Soc Lond B Biol Sci 267:1705–1710.

Clifford CWG, Wyatt AM, Arnold DH, Smith ST, Wenderoth P (2001) Orthogonal adaptation improves orientation discrimination. Vision Res 41:151–159.

Collins T (2020) Serial dependence alters perceived object appearance. J Vis 20:9.

Crick F, Koch C (1995) Are we aware of neural activity in primary visual cortex? Nature 375:121–123.

Dale AM (1999) Optimal experimental design for event-related fMRI. Hum Brain Mapp 8:109–114.

Dekel R, Sagi D (2015) Tilt aftereffect due to adaptation to natural stimuli. Vision Res 117:91–99.

Dhruv NT, Carandini M (2014) Cascaded Effects of Spatial Adaptation in the Early Visual System. Neuron 81:529–535.

Dong DW, Atick JJ (1995) Statistics of natural time-varying images. Netw Comput Neural Syst 6:345–358.

Dragoi V, Rivadulla C, Sur M (2001) Foci of orientation plasticity in visual cortex. Nature 411:80–86.

Dragoi V, Sharma J, Miller EK, Sur M (2002) Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. Nat Neurosci 5:883–891.

Dragoi V, Sharma J, Sur M (2000) Adaptation-Induced Plasticity of Orientation Tuning in Adult Visual Cortex. Neuron 28:287–298.

Durant S, Clifford CWG, Crowder NA, Price NSC, Ibbotson MR (2007) Characterizing contrast adaptation in a population of cat primary visual cortical neurons using Fisher information. J Opt Soc Am A 24:1529.

Felsen G, Touryan J, Dan Y (2005) Contextual modulation of orientation tuning contributes to efficient processing of natural stimuli. Netw Comput Neural Syst 16:139–149.

Fischer J, Whitney D (2014) Serial dependence in visual perception. Nat Neurosci 17:738–743.

Fornaciai M, Park J (2018) Attractive Serial Dependence in the Absence of an Explicit Task. Psychol Sci 29:437–446.

Fritsche M, Mostert P, de Lange FP (2017) Opposite Effects of Recent History on Perception and Decision. Curr Biol 27:590–595.

Fritsche M, Solomon SG, de Lange FP (2022) Brief stimuli cast a persistent long-term trace in visual cortex. J Neurosci:JN-RM-1350-21.

Fritsche M, Spaak E, de Lange FP (2020) A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. Neuroscience. Available at: http://biorxiv.org/lookup/doi/10.1101/2020.01.22.915553 [Accessed March 2, 2020].

Gardner JL, Sun P, Waggoner RA, Ueno K, Tanaka K, Cheng K (2005) Contrast Adaptation and Representation in Human Early Visual Cortex. :14.

Girshick AR, Landy MS, Simoncelli EP (2011) Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. Nat Neurosci 14:926–932.

Glover GH (1999) Deconvolution of Impulse Response in Event-Related BOLD fMRI1. NeuroImage 9:416–429.

Gold JI, Shadlen MN (2007) The Neural Basis of Decision Making. Annu Rev Neurosci 30:535–574.

Grunewald A, Bradley DC, Andersen RA (2002) Neural Correlates of Structure-from-Motion Perception in Macaque V1 and MT. J Neurosci 22:6195–6207.

Hajonides JE, Ede F van, Stokes MG, Nobre AC, Myers NE (2021) Multiple and Dissociable Effects of Sensory History on Working-Memory Performance. :2021.10.31.466639 Available at: https://www.biorxiv.org/content/10.1101/2021.10.31.466639v1 [Accessed March 2, 2022].

He S, MacLeod DIA (2001) Orientation-selective adaptation and tilt after-effect from invisible patterns. Nature 411:473–476.

Hu B, Garrett ME, Groblewski PA, Ollerenshaw DR, Shang J, Roll K, Manavi S, Koch C, Olsen SR, Mihalas S (2021) Adaptation supports short-term memory in a visual change detection task. PLOS Comput Biol 17:e1009246.

Kim S, Burr D, Cicchini GM, Alais D (2020) Serial dependence in perception requires conscious awareness. Curr Biol 30:R257–R258.

Kiyonaga A, Scimeca JM, Bliss DP, Whitney D (2017) Serial Dependence across Perception, Attention, and Memory. Trends Cogn Sci 21:493–497.

Manassi M, Liberman A, Kosovicheva A, Zhang K, Whitney D (2018) Serial dependence in position occurs at the time of perception. Psychon Bull Rev 25:2245–2253.

Moore T, Fallah M (2001) Control of eye movements and spatial attention. Proc Natl Acad Sci 98:1273–1276.

Moore T, Zirnsak M (2017) Neural Mechanisms of Selective Visual Attention. Annu Rev Psychol 68:47–72.

Moradi F, Koch C, Shimojo S (2005) Face Adaptation Depends on Seeing the Face. Neuron 45:169–175.

Murai Y, Whitney D (2021) Serial dependence revealed in history-dependent perceptual templates. Curr Biol 31:3185-3191.e3.

Papadimitriou C, Ferdoash A, Snyder LH (2015) Ghosts in the machine: memory interference from the previous trial. J Neurophysiol 113:567–577.

Papadimitriou C, White RL, Snyder LH (2016) Ghosts in the Machine II: Neural Correlates of Memory Interference from the Previous Trial. Cereb Cortex:bhw106.

Pascucci D, Mancuso G, Santandrea E, Della Libera C, Plomp G, Chelazzi L (2019) Laws of concatenated perception: Vision goes for novelty, decisions for perseverance Tong F, ed. PLOS Biol 17:e3000144.

Patterson CA, Duijnhouwer J, Wissig SC, Krekelberg B, Kohn A (2014) Similar adaptation effects in primary visual cortex and area MT of the macaque monkey under matched stimulus conditions. J Neurophysiol 111:1203–1213.

Patterson CA, Wissig SC, Kohn A (2013) Distinct Effects of Brief and Prolonged Adaptation on Orientation Tuning in Primary Visual Cortex. J Neurosci 33:532–543.

Phinney RE, Bowd C, Patterson R (1997) Direction-selective Coding of Stereoscopic (Cyclopean) Motion. Vision Res 37:865–869.

Roth ZN, Heeger DJ, Merriam EP (2018) Stimulus vignetting and orientation selectivity in human visual cortex. eLife 7:e37241.

Sadil P, Cowell R, Huber DE (2021) The Yin-yang of Serial Dependence Effects: Every Response is both an Attraction to the Prior Response and a Repulsion from the Prior Stimulus. Available at: https://psyarxiv.com/f52yz/ [Accessed March 14, 2022].

Schall JD (2004) On the role of frontal eye field in guiding attention and saccades. Vision Res 44:1453–1467.

SciPy 1.0 Contributors et al. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17:261–272.

Seriès P, Stocker AA, Simoncelli EP (2009) Is the Homunculus “Aware” of Sensory Adaptation? Neural Comput 21:3271–3304.

Siegle JH et al. (2021) Survey of spiking in the mouse visual system reveals functional hierarchy. Nature 592:86–92.

Sprague TC, Serences JT (2013) Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. Nat Neurosci 16:1879–1887.

St. John-Saaltink E, Kok P, Lau HC, de Lange FP (2016) Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. J Neurosci 36:6186–6192.

Stein H, Barbosa J, Rosa-Justicia M, Prades L, Morató A, Galan-Gadea A, Ariño H, Martinez-Hernandez E, Castro-Fornieles J, Dalmau J, Compte A (2020) Reduced serial dependence suggests deficits in synaptic potentiation in anti-NMDAR encephalitis and schizophrenia. Nat Commun 11:4250.

Stocker AA, Simoncelli EP (2006) Noise characteristics and prior expectations in human visual speed perception. Nat Neurosci 9:578–585.

van Bergen RS, Jehee JFM (2019) Probabilistic Representation in Human Visual Cortex Reflects Uncertainty in Serial Decisions. J Neurosci 39:8164–8176.

Wan Q, Cai Y, Samaha J, Postle BR (2020) Tracking stimulus representation across a 2-back visual working memory task. R Soc Open Sci 7:190228.

Wei X-X, Stocker AA (2015) A Bayesian observer model constrained by efficient coding can explain “anti-Bayesian” percepts. Nat Neurosci 18:1509–1517.

Wei X-X, Stocker AA (2017) Lawful relation between perceptual bias and discriminability. Proc Natl Acad Sci 114:10244–10249.

Xie Y, Hu P, Li J, Chen J, Song W, Wang X-J, Yang T, Dehaene S, Tang S, Min B, Wang L (2022) Geometry of sequence working memory in macaque prefrontal cortex. Science 375:632–639.