Sensory prediction errors increase coding efficiency in mouse visual cortex through gain amplification

Matthew F. Tang[1,2,3 #*], Ehsan Kheradpezhouh[1,2 #], Conrad C.Y. Lee[1,2,3], J. Edwin Dickinson[4], Jason B. Mattingley[2,3,5,6], and Ehsan Arabzadeh[1,2]

1. Eccles Institute of Neuroscience, John Curtin School of Medical Research, The Australian National University, Canberra, ACT, Australia
2. Australian Research Council Centre of Excellence for Integrative Brain Function, Victoria, Australia
3. Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia
4. School of Psychological Sciences, The University of Western Australia, Perth, WA, Australia
5. School of Psychology, The University of Queensland, Brisbane, QLD, Australia
6. Canadian Institute for Advanced Research (CIFAR), Toronto, Canada

co-first authors

* Corresponding author

Corresponding author
Matthew F. Tang
Eccles Institute of Neuroscience
John Curtin School of Medical Research
The Australian National University
Acton, ACT, 2601, Australia
Email: matthew.tang@anu.edu.au

Abstract

The efficiency of sensory coding is affected both by past events (adaptation) and by expectation of future events (prediction). Here we employed a novel visual stimulus paradigm to determine whether expectation influences orientation selectivity in the primary visual cortex. We used two-photon calcium imaging (GCaMP6f) in awake mice viewing visual stimuli with different levels of predictability. The stimuli consisted of sequences of grating stimuli that randomly shifted in orientation or systematically rotated with occasionally unexpected rotations. At the single neuron and population level, there was significantly enhanced orientation-selective response to unexpected visual stimuli through a boost in gain, which was prominent in awake mice but also present to a lesser extent under anesthesia. We implemented a computational model to demonstrate how neuronal responses were best characterized when adaptation and expectation parameters were combined. Our results demonstrated that adaptation and prediction have unique signatures on activity of V1 neurons.

Introduction

There is often more information in the sensory environment than the brain has the capacity to fully process. To cope with this information overload, and to enhance the efficiency of sensory processing, neuronal circuits use strategies such as adaptation[1,2] attention[3,4], and prediction[5,6]. Sensory adaptation is known to increase neuronal coding efficiency by taking into account the statistics of past sensory input[1,7,8]. Likewise, selective attention enhances neural responses to task-relevant features[3,9]. Similarly, predictions about specific future stimuli, formed using Bayesian inference, could enhance information processing[5,6]. Based on this *predictive coding* view of sensory processing, the mammalian cortex is constructed as a predictive machine which iteratively generates an internal model of its external environment based on the statistical regularities of incoming sensory inputs.

Predictive coding provides a simple theoretical view of perception which is supported by a body of work in human neuroimaging and behavioral studies[10,11]. The classic mismatch negativity effect has become a hallmark of this literature[12,13]. When encountering an unexpected stimulus, the brain generates a significantly larger M/EEG evoked response compared with the response following an expected stimulus[11]. Decoding of EEG activity in humans has revealed expectation shapes the representation of visual information in the response[14–18]. However, non-invasive neuroimaging techniques reflect overall population activity and it is unclear how sensory representations are affected by expectation at the single-neuron level. Recent work suggests that prediction affects single neuron response across a number of sensory modalities[19–24]. Computational modelling suggests that predictions are inhibitory copies of the expected stimulus passed down the cortical hierarchy to the earliest

sensory regions[6], where they are integrated with incoming sensory inputs. If a stimulus is expected, the inhibitory copy should minimize the neuronal response, allowing the stimulus to be efficiently encoded. By contrast, any mismatch between the expected and presented stimulus should result in a prominent response.

Here, we tested key elements of predictive coding theory at the neuronal level in mouse primary visual cortex (V1). We used two-photon calcium imaging (GCaMP6*f*) in awake mice that were exposed to sequences of visual stimuli at different levels of predictability, to directly test how expectation affects the responses of individual V1 neurons. We characterized how prediction shapes orientation selectivity in V1 neurons and how changes in orientation tuning affect the amount of information individual neurons and neuronal populations carry about the sensory input. We demonstrate that unexpected stimuli produce a higher gain in neuronal tuning to the preferred stimulus, resulting in increased information at the single-neuron and population level. This enhanced coding of unexpected stimuli was prominent in awake mice but also present to a lesser extent in anesthetized mice. Finally, we used a computational model to quantify the contribution of adaptation and expectation to neuronal responses at the single trial level.

Results

We combined experimental and modelling approaches to determine how prediction affects neuronal responses in V1 in mice. We asked whether orientation selectivity of individual neurons changes with the expectations about subsequent stimuli. To do this, we presented sequences of oriented gratings with different levels of predictability to awake mice ($N = 5$, 23 sessions, 1693 neurons) while imaging Layer 2/3 V1 activity using two-photon excitation

microscopy (Figure 1ABC, Movie 1). The stimulus sequence was adapted from the Allen Brain Institute's Brain Observatory paradigm[25]. This consisted of a sequence of full-screen gratings (0.0034 c/deg, 50% contrast) oriented between 0° and 150° in 30° steps, presented at 4 Hz with no inter-stimulus interval. In the *Random* condition (Figure 1B), the orientation of each grating was uncorrelated with the preceding gratings (i.e., white noise stimulation). To establish a prediction about grating orientation, in the *Rotating* condition the grating rotated either clockwise or anti-clockwise for 5 to 9 presentations (in 30° step), before jumping to an unexpected random orientation. In this condition, *Expected events* occurred during the rotating sequence whereas the *Unexpected events* occurred when the sequence made a random jump to an unpredicted orientation. For unexpected events the jump from the predicted orientation was to a random orientation, identical to the correlation statistics for the stimulus sequence embedded in the *Random* condition.

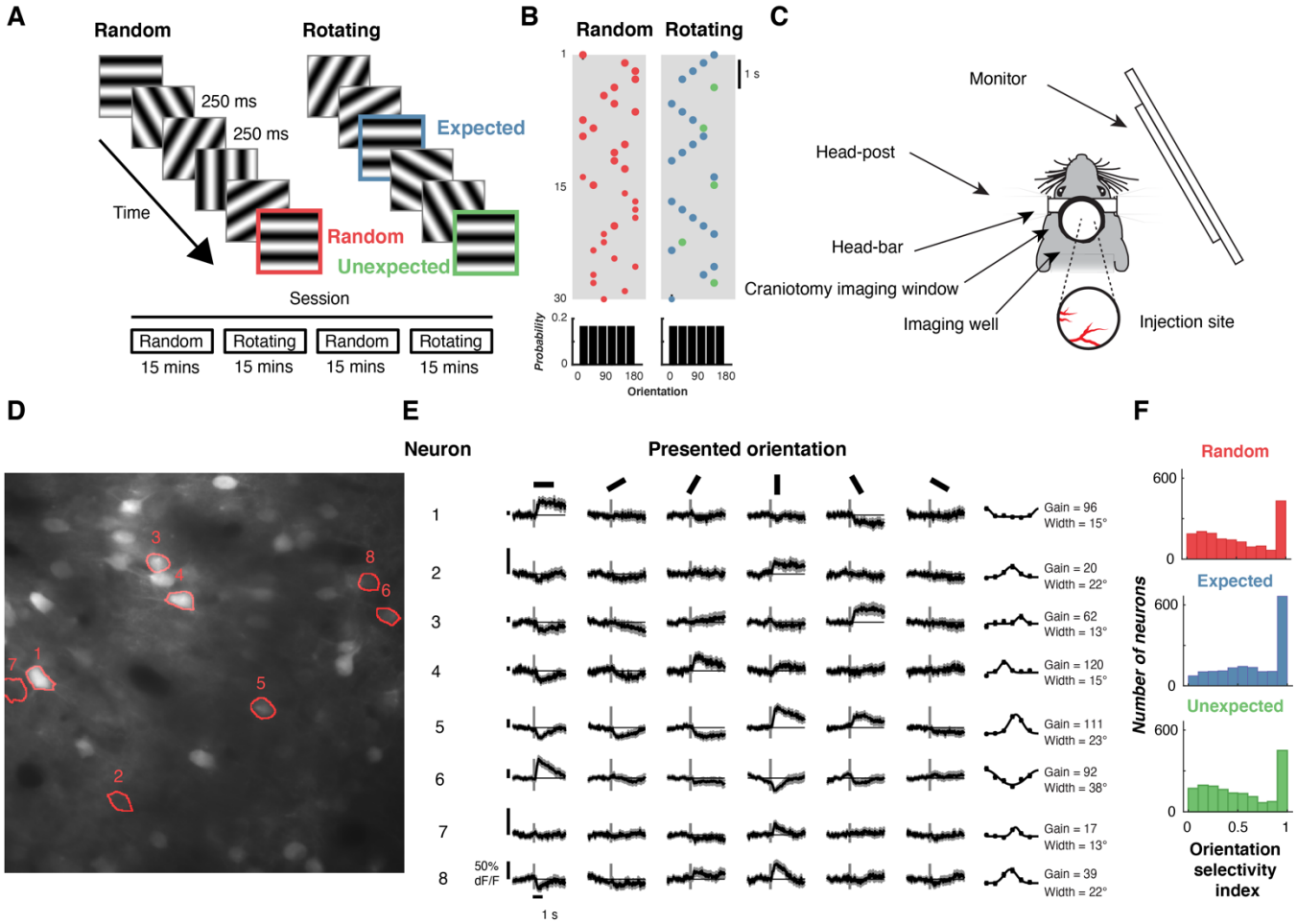


Figure 1. *Experimental procedure for testing the predictive coding account of visual perception in mouse V1 neurons.* ***(A)*** *Schematic of the Random and Rotating sequences of oriented gratings.* ***(B)*** *In the Random condition, the orientation of each stimulus was drawn from a pseudo-random distribution (from 0 to 150° in 30° steps). In the Rotating condition, the gratings rotated clockwise (i.e., 0° -> 30° -> 60°) or anti-clockwise (0° -> 150° -> 120°) for 5-9 presentations (black dots) before randomly jumping to an unexpected orientation (indicated by the red dots).* ***(C)*** *The experimental apparatus for using two-photon calcium imaging in combination with visual stimulation.* ***(D)*** *A mean motion-corrected two-photon image from a single session, with individual neurons highlighted in red.* ***(E).*** *Time course of activity in the corresponding neurons highlighted in D in response to different grating orientations. The right panels show the average response from 0 - 1000 ms after stimulus presentation. Points are fitted with a circular Gaussian with a baseline offset. The key parameters of the fits are given as the gain (height) and width of the Gaussians for each neuron. Shading and error bars show ∓1 standard error over trials.* ***(F).*** *Distribution of orientation selectivity index (see Methods) for all neurons in the three stimulus conditions.*

Movie 1. *Example sequence of gratings in the Rotating condition. The grating rotates in one direction for 4-9 presentations before jumping to a random orientation.*

Prediction affects single neuron activity

In line with previous work, a large proportion (462/1693) of the imaged neurons showed orientation selectivity for this spatial frequency (one-way ANOVA $p < 0.05$). We first examined how orientation selectivity of individual neurons was affected by stimulus predictability (Figure 2). The three example neurons shown in Figure 2A all exhibit orientation selectivity, which becomes evident from ~85-100 ms after stimulus onset. The first neuron responded maximally to gratings at 0° (horizontal), with slight suppression for the orthogonal orientations. During presentation of the *Expected* sequence, modulation of neuronal activity began (0 ms). This pre-stimulus modulation was most likely due to the rotating nature of the sequence: the stimulus presented at -500 ms was orthogonal to that presented at 0 ms, meaning that in the 0° condition (the anti-preferred), the preferred stimulus was presented at -500 ms. The rotating nature of the stimuli during the *Expected* sequence thus produced a specific temporal profile in neuronal response. For this reason, here we focus on the *Random* and *Unexpected* stimuli where the stimuli presented immediately before 0 ms were uncorrelated with the current stimulus.

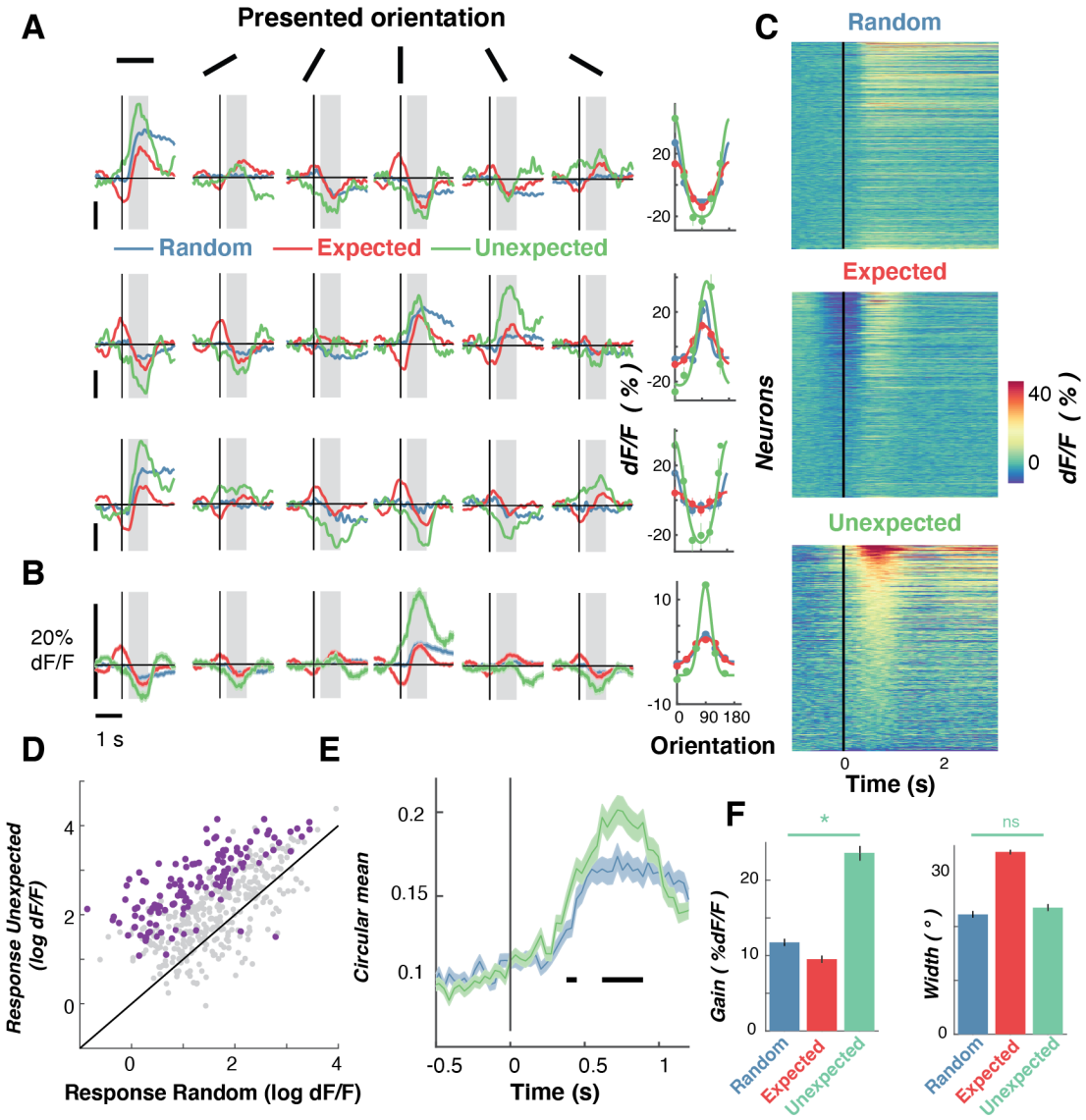


Figure 2. *Expectation affects orientation-selective responses of individual V1 neurons.* ***(A)*** *The time course of three example neurons (each neuron is a row) in response to oriented grating stimuli in the expected, unexpected and random conditions. The right panels in each row show orientation tuning curves given by the averaged response from 250 to 1000ms after stimulus onset for that neuron (grey shading in other panels). The solid line is a fitted Gaussian function with a constant offset.* ***(B)*** *Same as A, but shows activity for all orientation-selective neurons (N= 463) aligned to their preferred orientation to allow averaging. Right panel: Same as B but showing the population response.* ***(C)*** *Response to the preferred orientation across the three conditions for all orientation-selective neurons. The time courses are smoothed with a Gaussian with a 3.33 ms kernel for presentation. The color surrounding each box matches the colors in A.* ***(D).*** *Difference in response to the Unexpected minus Random condition at the preferred orientation for all neurons. Green dots show neurons (N=133) significantly modulated by expectation; grey dots are non-modulated neurons.* ***(E)*** *Time-course of orientation-selectivity (circular mean) for the neurons significantly modulated by expectation (N = 263).*

Black horizontal line indicates conditions are statistically different determined using non-parametric cluster-corrected procedures (see Methods). ***(F)****. Summary statistics for fitted Gaussian parameters across the population for the different sequence types. * indicates p < 0.05. Across all panels error bars and shading represent ± 1 standard error of mean.*

The main effect of prediction is evident from the example neurons illustrated in Figure 2. There was a systematic increase in neuronal responses to the preferred orientation, and a decrease to the anti-preferred orientation, in the *Unexpected* compared with the *Random* condition. This response profile is consistent with a positive gain modulation for unexpected gratings. The overall population response (aligned to the preferred orientation) showed the same pattern of results (Figure 2C), with an increased response to the preferred stimulus in the *Unexpected* versus *Random* condition. The response of 226 neurons (11%) was significantly modulated in the *Unexpected* condition relative to the *Random* condition (one-way ANOVA, $p < 0.05$). Of these, all but two showed a larger response in the *Unexpected* condition (Figure 2D), and this increase in selectivity emerged shortly after stimulus presentation (Figure 2E).

We next determined how orientation selectivity was affected by prediction. To do this, we fitted circular Gaussian tuning functions to separately extract amplitude and width parameters of orientation selectivity for each neuron (Figure 2E). The amplitude (or gain) of the tuning curve was significantly greater in the *Unexpected* condition than in the *Random* condition ($t(961) = 34.01$ $p < 0.001$). By contrast, there was no difference in the width between these two conditions, ($t(961) = 0.45$, $p = 0.65$). These results are consistent with our recent work examining how prediction affects orientation selectivity measured non-invasively in

178 humans[14,15]. A control condition showed these effects were not due to the systematic rotations
179 that followed *Unexpected* gratings (Supplementary Figure 1).

180 ***Prediction affects population coding of orientation***

181 In our initial set of analyses, we found that expectation affected orientation selectivity of
182 single V1 neurons. We next examined how enhanced orientation selectivity for unexpected
183 stimuli observed at the single-neuron level in turn shaped the information contained within the
184 population response. Previous human neuroimaging studies using multivariate pattern analysis
185 have shown that expectation affects classification accuracy[14–17,26]. To determine how these
186 findings generalize across species, we applied a similar multivariate pattern analysis to the
187 neuronal population data. We used all imaged neurons (N = 1954; 23 imaging sessions),
188 including both orientation-selective and non-orientation selective neurons to decode the
189 presented orientation using inverted/forward encoding modelling (see multivariate analysis
190 section in Methods for details). We used a forward (or inverted) encoding approach to
191 determine the amount of orientation-selective information contained in the population activity
192 on a trial-by-trial basis (Figure 3A-D). In line with the human work[14–17,26], this method uses an
193 encoding model to estimate neuronal selectivity to each orientation, and in a second step
194 inverts these weights to reconstruct the stimulus representation from the population response
195 on each trial.

196 We first examined the effect of different population sizes of neurons on decoding
197 accuracy. To do this, we selected groups of neurons and used a 20-fold cross-validation
198 procedure to train and test the classifier at each time point around stimulus presentation. This
199 procedure was repeated 24 times with different subsets of neurons selected. The same

decoding procedure was used as in the previous analysis, except that it was performed on the average neuronal responses from 250 to 1000 ms (i.e., the post-stimulus epoch over which decoding accuracy was best), and different sized pools of neurons were selected (2 to 1600, in 2 neuron steps). This analysis again showed that the presented orientation was decoded significantly better in the *Unexpected* versus the *Random* condition. This effect emerged in relatively few neurons (~20). The *Expected condition* also increased decoding accuracy relative to the *Random* condition, but there was a smaller increase than in the *Unexpected* condition and this did not emerge until a population of ~100 neurons was included.

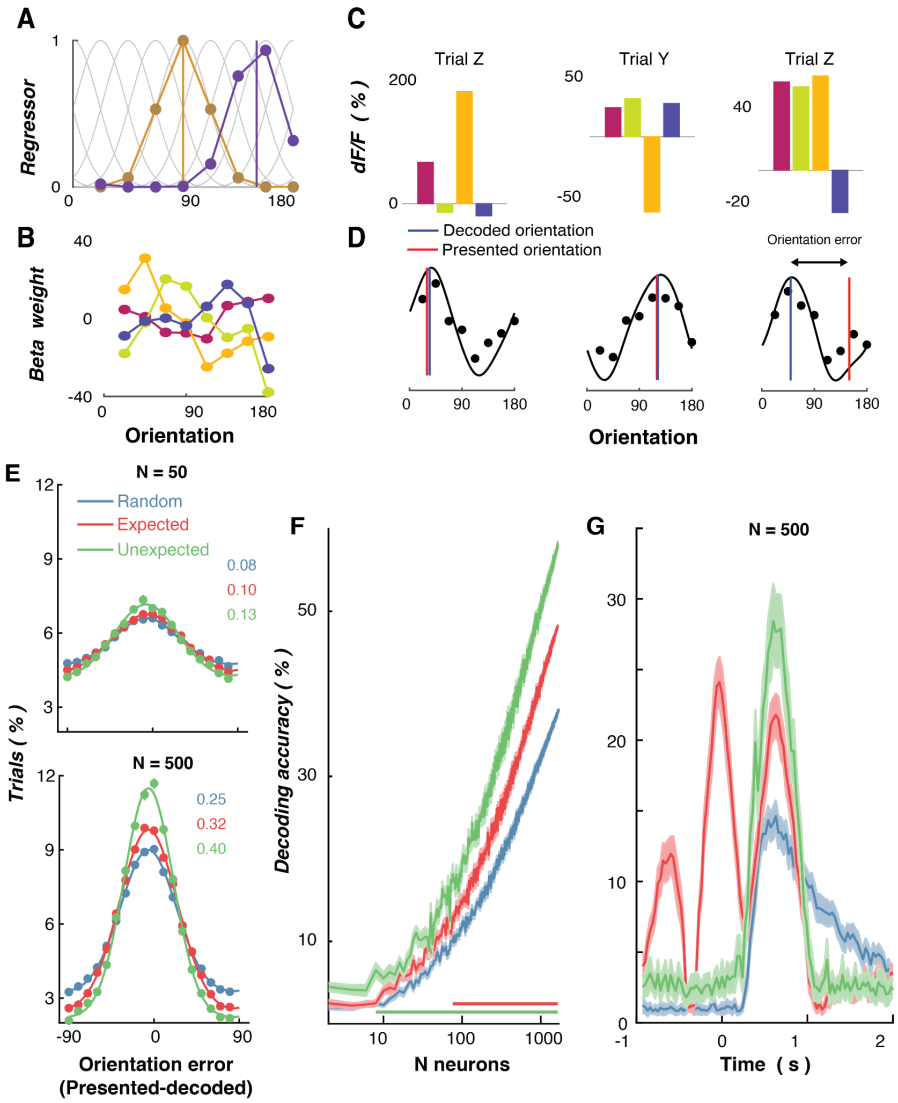


Figure 3. *Effects of prediction on information carried by multivariate population activity using forward encoding modelling. Encoding was performed separately on groups of 50 neurons and 500 neurons) at a time (with 24 permutations of different neuronal combinations).* ***(A).*** *The basis function (grey lines) in response to two orientations (events) which produce the regressor weights.* ***(B).*** *Beta weights for four example neurons (each color is a neuron) for each of the regressors found from a training set of data.* ***(C).*** *Activity for the four neurons (color-matched with those in B) in three test trials.* ***(D).*** *Inverting the Beta weights (B) and multiplying them with the test data from the four neurons produces the predicted orientation response. The difference between the predicted and presented orientation for a given stimulus is the orientation error.* ***(E).*** *Histograms of decoded orientations for the three conditions from the forward encoding results. The vector sum of these histograms was taken as the decoding accuracy for each condition. The colored numbers show the vector sum for the corresponding curves.* ***(F).*** *Decoding accuracy scales with the number of included neurons. The classifier was trained and tested on the average response from 250 to 1000 ms following stimulus presentation, with different numbers of neurons included (24 permutations of different*

neurons). The colored horizontal lines indicate the sign-flipped cluster permutation (N=2000, cluster p < 0.05, alpha p < 0.05) comparing Random vs. Unexpected (green line) and Random vs. Expected (red line). ***(G).*** *Time-resolved classification from forward encoding modelling (N= 500 neurons) with 24 permutations of different groups of neurons. In all panels, shading/error bars indicates ±1 standard error of the mean across permutations.*

Predictive coding under anesthesia

Previous work on expectation violations in humans has reported larger neural responses to unexpected than to expected stimuli during sleep[27,28], in different attention states[15,29], when individuals are in a coma[30] or vegetative state[31–33] or under anesthesia[34]. These findings suggest that the prediction errors can be generated across various states of consciousness but are modulated by global brain state. We therefore asked whether the prediction error effects we found here for individual neurons in area V1 were also present under anesthesia. The same stimulus paradigm was used but under isoflurane anesthesia (N = 4 animals). Consistent with previous studies[35], a lower proportion of neurons (273/ 577) showed orientation-selective responses in any of the experimental conditions relative to the awake experiments. Despite this, however, prediction errors had a similar effect on orientation selectivity when animals were anesthetized (Figure 4AB). The amplitude of orientation selectivity was significantly enhanced in the *Unexpected* relative to the *Random* condition ($t(272) = 14.10$, $p < 0.0001$). Again, as in the awake animals, there was a small but non-significant decrease in the width of the tuning curve in the *Unexpected* condition relative to the *Random* condition (Figure 3D, $t(272) = 1.67$, $p = 0.10$).

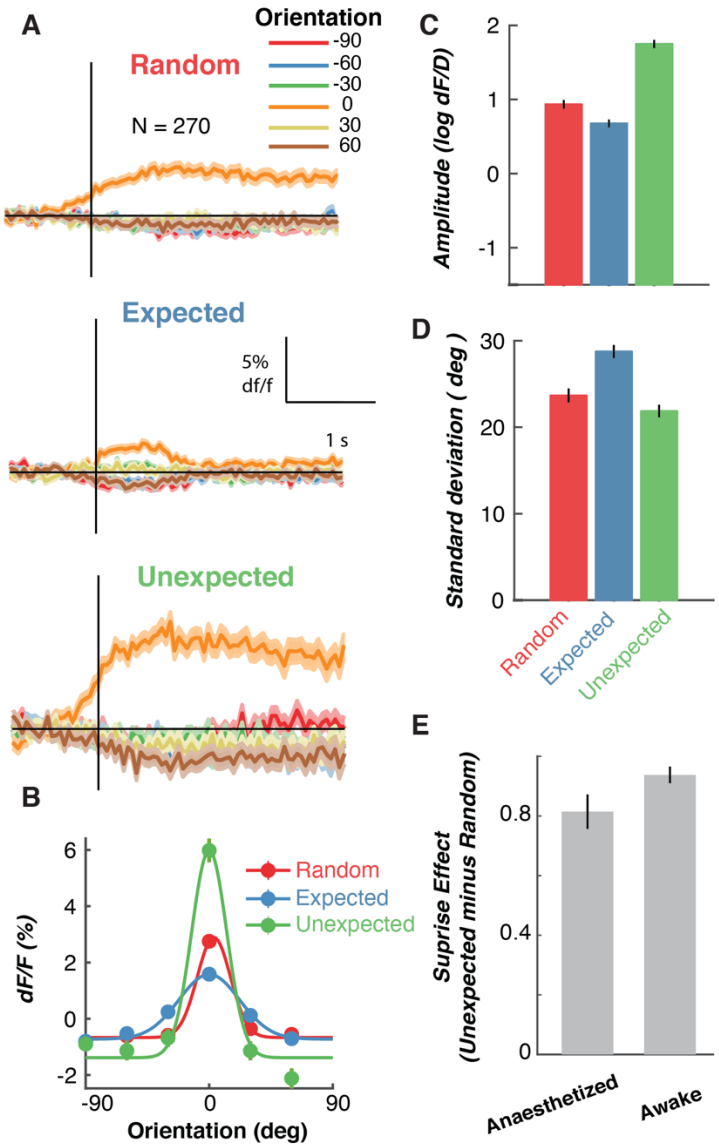


Figure 4. *Expectations affect the gain of orientation-selective V1 neurons under anesthesia.* ***(A)*** *The time course of all orientation-selective neurons (N = 273) aligned to their preferred orientation to allow averaging. Shading indicates ∓ 1 standard error of the mean across neurons.* ***(B)*** *Orientation tuning curve of the population from the three expectation conditions, averaged across an epoch from 0 to 1,000 ms after stimulus presentation. The solid line is a fitted Gaussian function with a constant offset. Summary statistics for fitted Gaussian amplitude* ***(C)*** *and width* ***(D)*** *in the different conditions.* ***(E)*** *Comparison of the "surprise" effect (Unexpected events minus Random events) comparison between awake and anesthetized animals. Panels B-C error bars indicate ± 1 standard error of the mean across neurons.*

In humans, anesthesia reduces the magnitude of the effect of prediction error on neural responses[30]. We therefore calculated the "surprise" effect by subtracting the amplitude of the Gaussian tuning curve for the *Unexpected* condition from that of the Random condition for each neuron (Figure 5E). A value larger than 0 indicates that the neuron's orientation selectivity was enhanced in the *Unexpected* condition. There was a slightly, but non-statistically significant, larger Surprise effect in awake animals than in those that had been anesthetized ($t(733)=1.35$, $p = 0.18$) . This suggests individual V1 neurons' orientation-selective responses to violations of expectation are modulated by conscious state.

A computational model quantifies the relative contribution of adaptation and prediction on coding efficiency

The models of predictive coding state that high-level cortical areas pass predictions, which are inverse copies of the expected stimulus, to the lower-level areas[5,6]. According to this framework, only a small neuronal response is required for representation if a stimulus matches the expectation. This is reminiscent of the effect of adaptation on neuronal representation[1,8]. Both adaptation and prediction rely on the statistics of the sensory input. While adaptation relies on the statistics of the recent history of stimulation, prediction is thought to use the statistics to extract future patterns. We created a computational model of orientation processing to quantify how adaptation and prediction interact to determine neural coding efficiency. The model is based on a number of tuned orientation-selective neurons maximally sensitive to different orientations (Figure 5A). The neurons respond proportionally based on their sensitivity to the presented orientation. We incorporated two sources of inhibition: adaptation (in response to the previously-presented stimulus) and expectation (in response to

the predicted future stimulus). Similar to previous work[36–39], adaptation provides a gain modulation to neuronal orientation selectivity based on the response to the preceding stimulus (green dots Figure 5A). Prediction, on the other hand, affects neuronal responses by producing an inverse copy of the expected orientation (red line). The modulation of stimulus sensitivity is consistent with previous work which found that (i) uncommon stimuli result in stimulus-specific adaptation in auditory cortex[40] and (ii) V1 population response adapts to high level stimulus statistics in a homeostatic manner[41]. To account for commonly seen long-lasting effects of gain modulation on orientation sensitivity[42,43], the model allows sensitivity to recover gradually over a number of trials, and thus to predict serial dependency effects (Supplementary Figure 2).

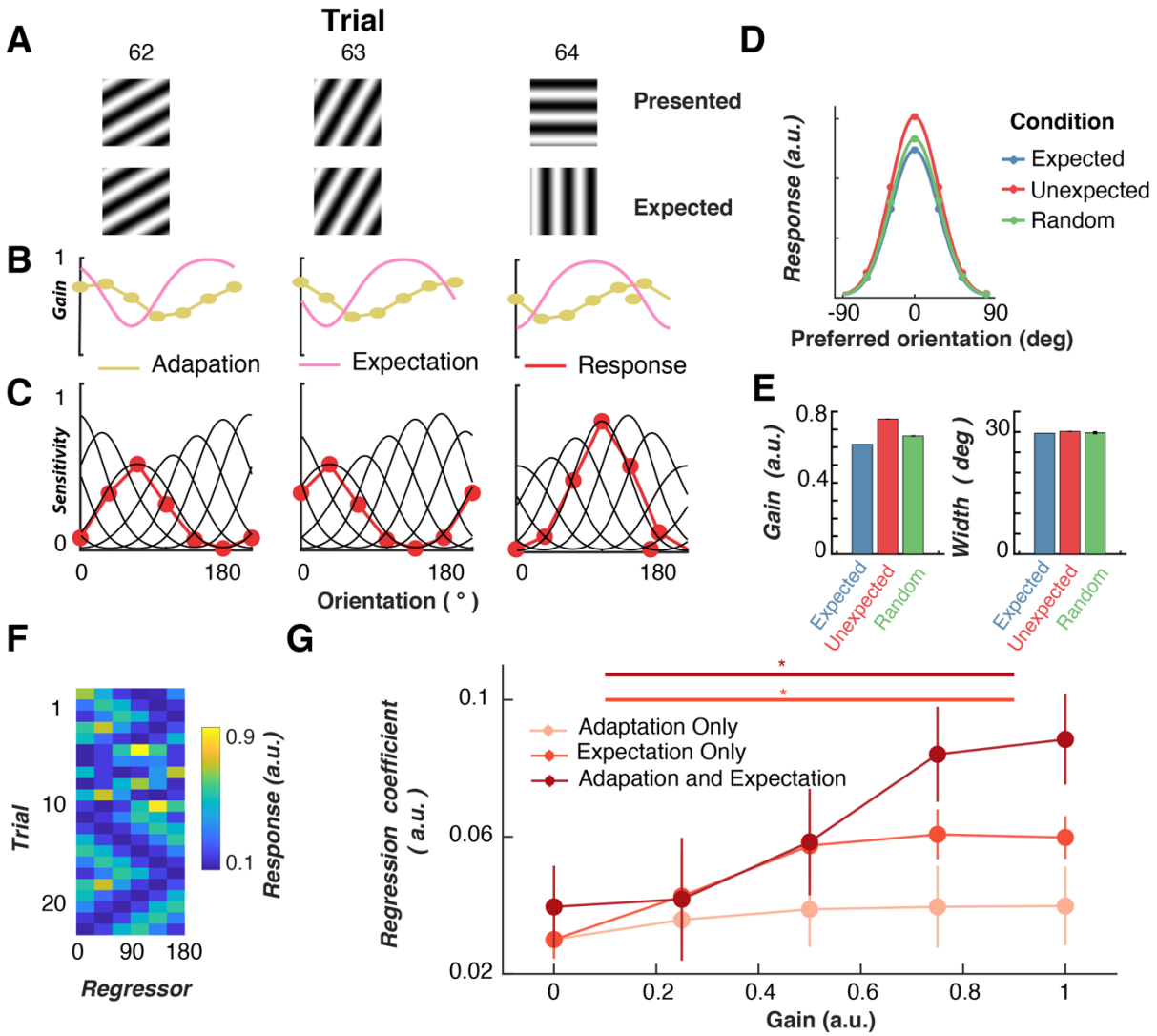


Figure 5. *Computational model for explaining variance in the neuronal activity by incorporating gain modulation from prediction and adaptation effects. The model consists of a bank of six neurons maximally sensitive to different orientations. (A) The four panels show four sequential trials (top to bottom) in the Rotating condition. The model's sensitivity profiles (C, black curves) determine their response (red dots) to the presented orientation (black vertical line). The response to the previous stimulus is multiplied by the adaptation gain factor and modulates the sensitivity of the channels on the next trial (yellow line). The expectation gain (pink line) is given by the inverse copy of the response to the expected orientation which is given by a circular Gaussian function. (**D**) The model's predicted response to orientations presented around their preferred orientation in the three stimulus conditions. All neurons were aligned to their preferred orientation and collapsed together. Dots indicate the responses of the neurons and the curves are fitted Gaussian functions. (**E**) Fitted Gaussian values to the model's responses for the different stimulus conditions. (**F**) An example of regressors (generated by the model's response to the oriented stimulus in one session). The response is determined by the*

expectation and adaptation states as well as the presented orientation. The first 10 trials come from the Random condition, and the following trials come from the Rotating condition. ***(G)*** *Ridge regression results from when the model was used to predict response to the stimulus in the Expected sequence with different levels of modulation from prediction. The regressor (orientation) with the highest beta weight was chosen for each neuron (N= 228 which showed a response modulated by condition (Figure 2))****.***

We presented sequences of orientations in both *Random* and *Rotating* conditions to the model to determine whether the model accurately predicts the observed changes in orientation selectivity with expectation. Because there are two sources of gain (adaptation and expectation) the model assumes an equilibrium of gain modulation is available to the system to allow it to maintain population homeostasis[41]. To this end, in the initial model we implemented 0.5 a.u. of gain available, which was varied in the two stimulus conditions. In the *Random* condition, the expectation gain was set to 0 a.u and the adaptation gain was set to 0.5 a.u. because the stimulus sequence did not contain any predictability. Whereas, in the *Rotating* condition, the gain for both expectation and adaptation were set to 0.25 a.u. We re-aligned neurons to their preferred orientation and determined their response to stimuli under different conditions (Figure 5B) by fitting the same Gaussian to the results (Figure 5C). Consistent with the data, in the model the gain of orientation selectivity increased in the *Unexpected* condition relative to the *Expected* and *Random* conditions. The *Unexpected* trials resulted in greater orientation selectivity compared with *Unexpected* trials as the sensitivity to the stimulus was reduced for a different orientation (the predicted grating orientation) than the one that was actually presented. There was no difference in the width of the representations. The model, therefore, produced a qualitative fit consistent with the effects of expectation on V1 orientation selectivity.

We next determined whether the model provided a quantitative fit to the neuronal activity (Figure 5E). This analysis assessed whether the model accurately predicted the trial-by-trial variability in the observed neuronal response and determined the relative value of the expectation and adaptation gain parameters. To do this, we used the model to generate predictions about neuronal responses which we regressed against the actual data for each neuron. Specifically, for each experimental session for the awake mice, we presented the model with the same orientation sequence the mouse viewed and this in turn generated a predicted response for each simulated neuron on every trial.

We found that more variance in the trial-by-trial activity of neurons could be explained when the model incorporated inhibition from expectation. We presented the orientation sequences from the *Rotating* condition to the model with three different gain responses for expectation. With no gain only the presented stimulus determined the response of the model, and as we increased the gain to 0.25 and 0.75 a.u. more inhibition from expectation increased the model's fit with the data. We used ridge regression to determine beta weights for each of the six regressors (orientations) for the three different gain settings for each neuron. We focused our analysis on the neurons identified previously which were significantly modulated by the prediction condition (Figure 2DF).

The adaptation model showed no increase in ability to explain neuronal activity with increasing gain (one-way ANOVA, $F(4,900) = 0.52$, $p = 0.72$). However, explanatory power of the expectation-only model greatly increased with increasing levels of gain (one-way ANOVA, $F(4,900) = 6.18$, $p < 0.001$). Furthermore, the model that incorporated a moderate amount of adaptation (0.25) with varying degrees of expectation gain best predicted the neuronal

response. A 3 (Model type; Adaptation, Expectation, Combined model) × 5 (Gain level; 0,0.2,0.4,0.6,0.8,1.0) repeated measures ANOVA confirmed this observation showing that the type of model significantly affected variance explained ($F(2,450) = 4.22$, $p = 0.02$) in addition to the amount of Gain ($F(4,900) = 11.55$, $p < 0.001$). These factors significantly interacted ($F(8,1800) = 2.03$, $p = 0.04$) showing that the difference in explanatory power between the models increased with increasing gain. Follow-up tests showed while the expectation model did not explain significantly more variance than the adaptation model across all gain levels ($p = 0.13$), the combined model did ($p = 0.004$).

Predictions repels perception away from expected orientation

Finally, we examined a unique prediction of the model; namely that the population representation of orientation should be biased away from the expected orientation. We reasoned that similar to the adaptation aftereffect[37,44], predictions should inhibit the gain of the expected orientations and in turn produce a bias in the participant's judgment of orientation. These gain modulation models predict the tilt-aftereffect, in which perception is biased away from a previously presented orientation. Furthermore, because the gain reductions are orientation-selective and the channels are 30 degrees, the bias should be largest when the expected orientation is approximately 30 degrees from the presented orientation. To determine whether this prediction holds, we separated the forward encoding results (Figure 3) by the difference from the expected orientation (Figure 6A) and determined whether the population encoding of orientation is biased away from the expected orientation. Consistent with the model, we found the representation of the presented orientation was indeed biased away from expectation with the largest effect when there was a 30-degree difference. This result was

accurately predicted by the model (Figure 6B). Finally, we sought to determine whether there was a corresponding change in human perception with expectation, analogous to the tilt-aftereffect with adaptation[38,44]. For this, we had observers view sequences of 4 to 7 gratings from the rotating condition and asked them to judge the perceived orientation of the final Gabor (Figure 6C). Across trials, we presented conditions where the expected orientation varied from -90 to 60 degrees in 30° steps (see Supplementary Figure 3). We found that perception of orientation was indeed biased away from the orientation that was expected (Figure 6D). All subjects showed significant orientation tuned biases away from the expected orientation (all *ps* < 0.05). The effect was in line with the neuronal population response and consistent with the model's prediction. Critically, unlike the classical tilt aftereffect, here perception is biased away from the expected orientation, rather than away from the previously-presented orientation.

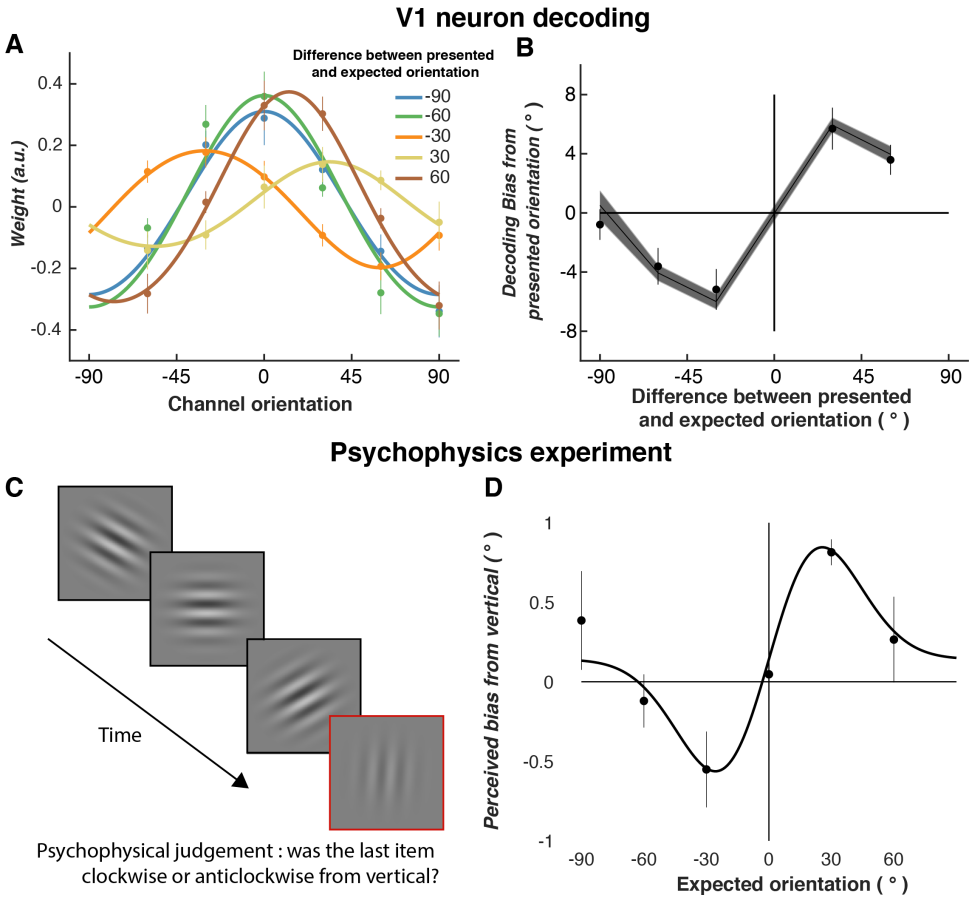


Figure 6. *Population decoding and human perception are both biased away from the expected orientation. For the rotating sequence, we found the difference from the expected orientation for each unexpected stimulus. For instance, if 0 was presented but 60 was expected, the difference is 60.* ***(A)*** *We used forward encoding modelling to estimate the population encoding of orientation) and separated by the difference from the expected orientation.* ***(B)*** *Results from the forward encoding modelling. The population response is biased away from the expected orientation (dots) with the largest bias at 30 degrees, as predicted by the model (shading).* ***(C)*** *The psychophysics experiment used to determine whether human subjects' (N = 3) perception of orientation is biased away from the expected orientation. Each subject viewed a sequence (4-7 Gabors) using the rotating paradigm from mice. They were asked to judge whether the final item in the sequence was clockwise or anti-clockwise from vertical. The test orientation was varied across trials from -6 to +6 degrees using a method of constant stimuli to measure the change in perceived orientation. Cumulative Gaussian functions were fit to responses to determine whether perception was biased from vertical.* ***(D)*** *Bias results from vertical for the human subjects. The line is the fitted first derivative of a Gaussian function used to quantify whether the observer's bias results were significantly greater than chance and orientation tuned. Across all panels error bars indicate ± 1 standard error.*

Discussion

We provide an experimental test of how neuronal representation of visual information is affected by prediction in a primary sensory area. We presented a stream of gratings to awake mice in two conditions; (i) in Random condition one where there was no relationship between the current and previous trials (Random trials) and (ii) Rotating condition where the orientation could be predicted from the past sequence (Expected trials) but occasionally included a random orientation (Unexpected trials). We found consistent evidence that expectations modulate the gain of orientation responses in V1 activity, both at the single-neuron and population level. The increase in gain of orientation selectivity for unexpected trials appears to occur through an increase in the reliability of responses. While the magnitude of prediction error response was decreased under anesthesia, it was still present showing these effects were not caused by top-down arousal-related effects and instead reflect fundamental changes in sensory representation. Finally, we provide a computational implementation of a predictive coding model in V1 to better understand the interaction between adaptation and prediction. By varying the parameters of the model, we found that the best explanation of the neuronal activity needed both inhibition from adaptation in response to immediately preceding stimulus events, and expectations about future stimulus features. Finally, we describe a previously unreported phenomenon in human perception based on characterization of neuronal response in V1 of mice and their modelling. Both neuronal representation and human perception of orientation exhibited similar biases away from the *expected* orientation.

While the notion that predictions about the future affect perception was first proposed by Helmholtz[45], there has only been limited evidence showing the neuronal encoding of sensory

429 information of this top-down signal. A number of more recent theoretical models[5,6] have
430 proposed a 'predictive coding' framework with the common idea that the brain inhibits the
431 sensory representation of the expected stimulus to increase coding efficiency. While there has
432 been extensive evidence that predictions affect the magnitude of neural activity measured with
433 neuroimaging[10,46], few studies have determined how neuronal responses are affected even
434 though this is a critical component of these models. The current results fit well with our
435 previous findings using multivariate patterns of neuroimaging activity measured in humans to
436 measure how orientation selectivity changes with expectation[14–16,26]. Similarly to the current
437 work, forward encoding modelling of EEG activity revealed an increase in the gain, but not the
438 width, of orientation tuning in human observers[14].

439 Our results add significantly to the understanding of how expectations affect the
440 representation of sensory information. Previous work[19,22,23], but see[47] for a different
441 interpretation, has suggested that the locomotion-induced increase in activity in primary visual
442 cortex in mouse[47,48]. Under the predictive coding framework, the increased activity caused by
443 locomotion creates an expectation that the stimulus should change size. A prediction error is
444 generated if the stimulus remains static, as is typical when measuring orientation selectivity, or
445 moves in an inconsistent direction. There is significantly less locomotion-induced increase in
446 response if the stimulus is made to move as the animal moves. Our results are consistent with
447 these but show the increase is due to a gain increase with a larger response to the neuron's
448 preferred stimulus.

449 In the human literature, expectation appears to affect sensory response through
450 different neural oscillatory frequency bands[49,50]. Recordings in macaques suggest visual

information is fed forward through high-frequency gamma (60-80 Hz) oscillations while feedback occurs through slow theta-band (14-18 Hz) activity[51]. As the recording was conducted using two-photon imaging with a relatively slow sampling rate we were unable to determine the role of different frequency bands in expectation. Future work, potentially simultaneously recording multiple areas using electrophysiology, could use the current task to determine the role of top-down and bottom-up processing for these gain modulations by expectation.

Methods

Mouse information

A total of 8 wild type mice (C57BL) were used in this experiment. All methods were performed in accordance with the protocol approved by the Animal Experimentation and Ethics Committee of the Australian National University (AEEC 2012/64; 2015/74). Mice were housed in a ventilated and air filtered climate-controlled environment with a 12-hour light–dark cycle. Animals had access to food and water ad libitum. No statistical methods were used to calculate the sample size, but these were consistent with other studies in the field.

Expression of Ca^{2+} indicator (GCaMP6f)

Mice were briefly anesthetized with isoflurane (~2% by volume in O_2) in a chamber and moved to a thermal blanket (37°C, Physitemp Instruments) before the head was secured in a stereotaxic frame (Stoelting, IL). Thereafter, the anesthetic gas (isoflurane, ~2% by volume in O_2) was passively applied through the nose mask at a flow rate of 0.6-0.8 L/min. The level of anesthesia was monitored by the respiratory rate, and hind paw and corneal reflexes. The eyes were covered with a thin layer of Viscotears liquid gel (Alcon, UK). The scalp was opened

473 with ~5 mm rostrocaudal incision at the midline using scissors and the periosteum was gently
474 removed. A circular craniotomy was made over the right visual cortex (3mm diameter;
475 centered 2mm lateral and 4.5mm posterior to Bregma) with the dura left intact. A glass pipette
476 (15-25μm diameter at tip) containing GCaMP6f (AAV1.Syn.GCaMP6f.WPRE.SV40, Penn
477 Vector Core, The University of Pennsylvania, USA) was inserted into the cortex at a depth of
478 230-250 μm below the dura using a micromanipulator (MPC-200, Sutter Instruments, Novato,
479 CA, USA). GCaMP6f was injected at 4-6 sites (with 4 32nL injections per site separated by 2-5
480 mins; rate 92 nLs^{-1}) using a glass pipette. Injections were controlled using a Nanoject II injector
481 (Drumont scientific, PA). After virus injection, the craniotomy was covered with a 3mm
482 diameter cover-glass (0.1 mm thickness, Warner Instruments, CT). This was glued to the bone
483 surrounding the craniotomy. Custom made head bars were fixed to the skull or Bregma using a
484 thin layer of cyanoacrylate adhesive and dental acrylic. A small well was built surrounding the
485 craniotomy window using dental acrylic to accommodate distilled water required for the
486 immersion lens of the 2-photon microscope.

487 Ca^{2+} imaging was performed using a two-photon microscope (Thorlabs Inc., Newton,
488 NJ, USA) controlled by ThorImage OCT software. The cortex was illuminated with a
489 Ti:Sapphire fs-pulsed laser (Chameleon, Coherent Inc., Santa Clara, CA, USA) tuned at 920
490 nm. The laser was focused onto L2/3 cortex through a 16x water-immersion objective lens
491 (0.8NA, Nikon), and Ca^{2+} transients were obtained from neuronal populations at a resolution of
492 512 × 512 pixels (sampling rate, ~30 Hz). To abolish the effect of visual stimuli on the calcium
493 signals, we secured the objective by filling the gap between the objective and the well with
494 removable adhesive (Blu-Tack).

The images were processed using the Suite2p toolbox (https://github.com/cortex-lab/Suite2P) for motion correction and segmentation. The surrounding neuropil signal was subtracted for each neuron's calcium traces. These corrected traces were high-pass filtered before the median response for each neuron was subtracted to determine dF/F.

Visual stimulus

The stimuli were displayed on a 22-inch LED monitor (resolution 1920 x 1080 pixels, refresh rate 60 Hz) using the PsychToolbox presentation software for MATLAB. The mouse was placed next to a monitor, which subtended 76.8° x 43.2° (one pixel = 2.4' x 2.4') orientated 30° from their midline. The visual stimulus sequence was based on the Allen Brain Institute Brain Observatory paradigm used to measure orientation selectivity in mice using two-photon imaging. The stimuli were full-screen gratings (0.0034 c/°, 50% contrast) displayed for 250 ms with no inter-stimulus interval giving a 4 Hz presentation rate with no inter-stimulus blank interval. The spatial frequency was chosen to be close to optimal sensitivity of neurons in V1[25]. The orientations of the gratings were equally spaced between 0 to 150° in 30° steps so we could map each neuron's orientation selective profile.

The predictability of the orientations of the gratings was varied in the two stimulus conditions. In the *Random* condition, the orientations of the gratings were drawn from a pseudo-random distribution with no relationship between the current orientation and the previous orientation. The *Rotating control* condition was introduced to determine whether the stimulus presented after the unexpected jump was affecting orientation selectivity. In this condition, after the unexpected orientation the stimulus made another jump to a random orientation before starting to rotate in the opposite direction as the previous rotation. The

number of events was increased from 7200 in each block to 8400 to have the same number of unexpected trials as the original *Rotating* condition, while all other details remained identical with the *Rotating* condition. We ran 13 sessions in 2 mice for all three conditions to compare the effect of the control. For all conditions, there was a balanced number of presentations of all the orientations.

Data analysis

To determine the effect of prediction, we averaged the calcium response from 250 to 1000 ms after stimulus presentation to derive tuning curves for each condition. Orientation selectivity index was defined as (1):

$$OSI = \frac{Rpref - Rorth}{Rpref + Rorth} \quad (1)$$

Where *Rpref* was the response to the preferred orientation (defined by the largest calcium response) and *Rorth* was the orientation orthogonal to the preferred orientation. To ensure the OSI fell between 0 and 1, we normalized the lowest calcium response to equal 0 in the lowest condition. To quantify how expectation affected the gain and selectivity of orientation-selective neurons we fit circular Gaussian distributions with a constant offset (2) using non-linear least square regression.

$$G(x) = Aexp - \frac{(x - \phi - j * 180)^2}{2\sigma^2} + C \quad (2)$$

Where A is the amplitude of the Gaussian, ϕ is the preferred orientation of the neuron (in degrees), $\boldsymbol{\sigma}$ is the width (in degrees) and C a constant offset to allow for baseline shifts in the activity of the neuron. We searched for best fitting solutions with parameter j, with a search space from -4 to +4 in integer steps.

Multivariate encoding analysis

We used a multivariate encoding approach (forward encoding modelling) to determine how the population activity carried information about the orientation of the presented grating on a trial-by-trial basis. This approach is based on human neuroimaging approaches examine orientation/feature selectivity from multivariate non-invasively recorded neural activity[14,15,52–55], but is similar to encoding approaches used to describe neuronal response to sensory stimuli[56,57]. Compared to the encoding-only, forward encoding takes the individual neuron activity to reconstruct the stimulus representation from the population activity. The technique goes beyond more commonly used multivariate pattern analysis procedures by producing tuning curves showing the full representation (in both amplitude, width, and bias) relative to the accuracy-only score.

The data were pooled across all experimental sessions with both orientation and non-orientation selective neurons used. In the first instance, we examined how the number of neurons affected decoding on a fixed time interval (300 to 600 ms) and in the second we found the time-resolved selectivity by applying the decoding procedure at each time point around the presentation of the stimulus (-500 to 2000 ms). A 20-fold cross-validation procedure was used in both instances for test and training data. The procedure evenly splits each test block to have the most even distribution of stimulus in each fold.

We used the presented orientations to construct a regression matrix with 8 regression coefficients. This regression matrix was convolved with a tuned set of nine basis functions (half cosine functions raised to the eighth power, Equation 5) centered from 0° to 160° in 20° steps. This helps pool similar orientations and reduces overfitting[57]. This tuned regression matrix was

used to measure orientation information across trials. This was done by solving the linear equation (3):

$$B_1 = WC_1 \tag{3}$$

Where B_1 (Neurons x N training trials) is the electrode data for the training set, C_1 (8 channels x N training trials) is the tuned channel response across the training trials, and W is the weight matrix for the sensors to be estimated (Neurons x 8 channels). We separately estimated the weights associated with each channel individually. W was estimated using least square regression to solve equation (4):

$$W = (C_1\, C_1^T)^{-1}\, C_1^T\, B_1 \tag{4}$$

We removed the correlations between neurons, as these add noise to the linear equation. To do this, we first estimated the noise correlation between neurons (which stops finding the true solution to the equation) and removed this component through regularization by dividing the weights by the shrinkage matrix[55,58]. The channel response in the test set C_2 (8 channels x N test trials) was estimated using the weights in (7) and applied to activity in B_2 (Neurons x N test trials), as per Equation 5:

$$C_2 = (W\, W^T)\, W^T\, B_2 \tag{5}$$

To avoid overfitting, we used 10-fold cross validation, where X-1 epochs were used to train the model, and this was then tested on the remaining (X) epoch. This process was repeated until all epochs had served as both test and training trials. We also repeated this procedure for each point in the epoch to determine time-resolved feature-selectivity. To re-align the trials with the

exact presented orientation, we reconstructed the item representation by multiplying the channel weights (9 channels x time x trial) against the basis set (180 orientations x 9 channels). This resulted in an Orientations (-89° to 90°) × Trial × Time reconstruction.

To quantify the orientation selective response, we found the vector sum of the orientation for each trial (Figure 3D) to determine the decoded orientation. The difference between the decoded and presented orientation was the orientation error. For each condition (and time point where applicable) we found the distribution of orientation errors and calculated the histogram of responses.

In the temporal classification analysis, groups of 500 neurons were used on each instance for both training and test data with the cross-validation procedure applied to each time point around stimulus presentation. We permuted new groups of 500 neurons 24 times. For the second analysis, we averaged the evoked activity from (250 to 1000 ms after stimulus presentation. The same classification was then used as in the previous analysis but with different numbers of neurons (2 to 1944 neurons in steps of 2 neurons) to determine how decoding was affected by population size. Again, we selected different groups of neurons 24 times so as not to skewer the results by the neurons that were selected.

Computational model

The analytic model is based on previous work accounting for feature (i.e. orientation, spatial) adaptation based on neuronal response and human psychophysical data[36,38,39,59,60]. The model consists of a bank of six orientation-selective information channels with preferred orientations evenly spaced between 0 and 150° (in 30° steps). Each channel's sensitivity profile was given by a Gaussian function (6).

$$G(x) = Aexp - \frac{(x-\phi)^2}{2\sigma^2} \quad (6)$$

Where A is the amplitude (set to 1 a.u.), $\phi$is the channel's preferred orientation, $\sigma$is the width of the channel (set to be 40° consistent with the neuronal data). The number of channels, along with the width means the model is equally sensitive to all orientations. The population response to any presented orientated stimulus is given by the sensitivity profiles of the channels (See Supplementary Figure 4). In an unadapted state (Supplementary Figure 3A), the model will show a maximal response around the presented orientation with the vector average of the population response will be the presented orientation.

To account for adaptation, the gain of the information channels is reduced by an inverse proportion to their response by the previous stimulus (Supplementary Figure 4B). For instance, if a 90° stimulus is the adapting stimulus the sensitivity of the channels around 90° will be maximally reduced while orthogonal channels will be unaffected. The magnitude of this reduction (*adaptation ratio*) can be varied to allow for greater or less adaptation and was included as a free parameter in the analysis. The adaptation aspect of the model is consistent with previous models used to psychophysical data[36,38,39,59,60]. The new model accurately predicts serial dependency effects (where the current orientation is biased away from the previous orientation) seen in human behavioral and neuronal data (Supplementary Figure 2)[42,43,61].

Prediction gain modulation works in a similar manner as adaptation except that the stimulus sensitivity, rather than channel sensitivity, is modulated. Furthermore, the gain modulation occurs before the stimulus and is for the orientation that is expected rather than actually presented. The modulation of stimulus sensitivity is consistent with a previous study

624 which found that uncommon stimuli result in stimulus-specific adaptation in the auditory
625 cortex[40]. Stimulus-specific adaptation has been used in modelling neuronal adaptation[41]. To
626 model stimulus-selective gain modulation, the tuned Gaussian function was found using
627 Equation 1 and inverted before being applied to the channels. The amount of gain modulation
628 by expectation was a free parameter (*expectation gain*).

629 To account for long-lasting effects of gain modulation, the channel's sensitivity was
630 normalized by towards 1 (maximum sensitivity) on each trial. This causes the model to have
631 adaptation for a number of stimuli back, with the number determined by the *modulation factor*
632 and will lead to serial dependency-like effects for adaptation and prediction[42,43]. We regressed
633 the adaptation-only model against the neuronal data and found a factor of 3.0 best fit the data
634 which was set for other modelling experiments. This expectation factor means ~4 trials back
635 lead to detectable adaptation effects (Supplementary Figure 2B).

636 To determine the effects of the different stimulus conditions (Random, Expected and
637 Unexpected) on the model's channels, we presented sequences of orientations to the model
638 and split the responses into conditions. To allow for easier comparison, we aligned the six
639 orientation channels to their preferred orientation and collapsed the results across conditions.
640 The same effects were evident before collapsing.

641 Lastly, we examined how the actual neuronal responses could be predicted by the
642 model's predictions with different values of the free parameters. To do this, we used to model
643 to predict responses to the orientations presented to the mice during the session for all
644 stimulus conditions. For each neuron, we used the model's responses to the stimuli as
645 regressors to predict the neuron's response (averaged from 250to 1000 ms) for each stimulus

condition. We iterated this procedure with different values for adaptation and expectation gain to determine what values best predicted the data.

Psychophysics experiment

Three observers (one naive to the experimental aims) participated in the experiment. The observers viewed sequences of rotating orientations with the same features as the rodent two-photon experiment, except that the full-field gratings were replaced with Gabors (3c/°) presented ~10° from fixation. The final item in the sequence was a low-contrast Gabor with an orientation presented around vertical. The observers were required to indicate whether the final item was rotated clockwise or anti-clockwise from vertical using the keyboard. The true orientation of the Gabor was varied using the Method of Constant Stimuli (MOCS) from -6 (anti-clockwise) to +6° in 3° steps. A new trial began ~1 s after the response. Separate method of constant stimuli were used for each expected orientation (-90° to +60° from vertical, in 30° steps) with 30 repetitions of each point (a total of 300 trials) in each block. Each observer completed 4 blocks of the procedure.

For each block, the proportion of total responses for each MOCS level and expected orientation was found. Cumulative Gaussian functions were fitted to these data for each expected orientation to determine the point of subjective equality (Supplementary Figure 3). This shows what orientation the observers perceived as being vertical. The results are shown in Figure 6D. Following the tilt aftereffect literature, we fit first derivative of Gaussian (D1) functions to the points of subjective equality to quantify the orientation-selective bias results. This was done for each block, and a paired-samples t-test was applied to each observer's results to determine whether they showed a significant expectation aftereffect.

Data availability

The data are available at: https://osf.io/t2vb3. The code is available at: https://github.com/MatthewFTang/PredictionOrientationSelectivityMouseV1

Acknowledgements

This work was supported by the Australian Research Council (ARC) Centre of Excellence for Integrative Brain Function (CE140100007) to JBM and EA. The NHMRC supported MFT, JBM and EA with a Project grant (APP1165337) and an Ideas grant (APP1181643) to EK, CL and EA. MFT was supported by the NVIDIA corporation who donated a TITAN V GPU. EA was supported by an ARC Discovery Project (DP170100908). JBM was supported by an ARC Australian Laureate Fellowship (FL110100103) and by the Canadian Institute for Advanced Research (CIFAR).

Author Contributions

MFT and EA convinced the experiments, EK and CL performed the experiments, MFT analyzed the data, JED and MFT developed the model, MFT, EK, CYL, JED, JBM and EA wrote the paper

References

1. Lundstrom, B. N. & Fairhall, A. L. Decoding stimulus variance from a distributional neural code of interspike intervals. *J. Neurosci.* **26**, 9030–9037 (2006).
2. Gross, C. G., Rocha-Miranda, C. E. & Bender, D. B. Visual properties of neurons in inferotemporal cortex of the Macaque. *J. Neurophysiol.* **35**, 96–111 (1972).
3. Treue, S. & Martínez Trujillo, J. C. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* **399**, 575–579 (1999).
4. Luck, S. J., Chelazzi, L., Hillyard, S. A. & Desimone, R. Neural mechanisms of spatial selective

691 attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* **77**, 24–42 (1997).
692 5. Friston, K. A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 815–836
693 (2005).
694 6. Rao, R. P. N. & Ballard, D. H. Predictive coding in the visual cortex: A functional interpretation of
695 some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).
696 7. Maravall, M., Petersen, R. S., Fairhall, A. L., Arabzadeh, E. & Diamond, M. E. Shifts in coding
697 properties and maintenance of information transmission during adaptation in barrel cortex. *PLoS*
698 *Biol.* **5**, 0323–0334 (2007).
699 8. Adibi, M., Clifford, C. W. G. & Arabzadeh, E. Informational basis of sensory adaptation: Entropy
700 and single-spike efficiency in rat barrel cortex. *J. Neurosci.* **33**, 14921–14926 (2013).
701 9. Bichot, N. P., Rossi, A. F. & Desimone, R. Parallel and serial neural mechanisms for visual
702 search in macaque area V4. *Science (80-.).* **308**, 529–534 (2005).
703 10. Näätänen, R., Paavilainen, P., Rinne, T. & Alho, K. The mismatch negativity (MMN) in basic
704 research of central auditory processing: A review. *Clin. Neurophysiol.* **118**, 2544–2590 (2007).
705 11. Garrido, M. I., Kilner, J. M., Stephan, K. E. & Friston, K. J. The mismatch negativity: A review of
706 underlying mechanisms. *Clin. Neurophysiol.* **120**, 453–463 (2009).
707 12. Näätänen, R., Gaillard, A. W. K. & Mäntysalo, S. Early selective-attention effect on evoked
708 potential reinterpreted. *Acta Psychol. (Amst).* **42**, 313–329 (1978).
709 13. Näätänen, R. & Michie, P. T. Early selective-attention effects on the evoked potential: A critical
710 review and reinterpretation. *Biol. Psychol.* **8**, 81–136 (1979).
711 14. Tang, M. F., Smout, C. A., Arabzadeh, E. & Mattingley, J. B. Prediction error and repetition
712 suppression have distinct effects on neural representations of visual information. *Elife* **7**, (2018).
713 15. Smout, C. A., Tang, M. F., Garrido, M. I. & Mattingley, J. B. Attention promotes the neural
714 encoding of prediction errors. *PLoS Biol.* **17**, 1–22 (2019).
715 16. Kok, P., Mostert, P. & De Lange, F. P. Prior expectations induce prestimulus sensory templates.

Proc. Natl. Acad. Sci. U. S. A. **114**, 10473–10478 (2017).

17. King, J. R. *et al.* Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *Neuroimage* **83**, 726–738 (2013).

18. King, J. R., Gramfort, A., Schurger, A., Naccache, L. & Dehaene, S. Two distinct dynamic modes subtend the detection of unexpected sounds. *PLoS One* **9**, e85791 (2014).

19. Leinweber, M., Ward, D. R., Sobczak, J. M., Attinger, A. & Keller, G. B. A Sensorimotor Circuit in Mouse Cortex for Visual Flow Predictions. *Neuron* **95**, 1420-1432.e5 (2017).

20. Meyer, T. & Olson, C. R. Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 19401–19406 (2011).

21. Schneider, D. M., Sundararajan, J. & Mooney, R. A cortical filter that learns to suppress the acoustic consequences of movement. *Nature* **561**, 391–395 (2018).

22. Zmarz, P. & Keller, G. B. Mismatch Receptive Fields in Mouse Visual Cortex. *Neuron* **92**, 766–772 (2016).

23. Keller, G. B., Bonhoeffer, T. & Hübener, M. Sensorimotor Mismatch Signals in Primary Visual Cortex of the Behaving Mouse. *Neuron* **74**, 809–815 (2012).

24. Heindorf, M., Arber, S. & Keller, G. B. Mouse Motor Cortex Coordinates the Behavioral Response to Unpredicted Sensory Feedback. *Neuron* **99**, 1040-1054.e5 (2018).

25. de Vries, S. E. J. *et al.* A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nat. Neurosci.* **23**, 138–151 (2020).

26. Kok, P., Jehee, J. F. M. & de Lange, F. P. Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron* **75**, 265–270 (2012).

27. Loewy, D. H., Campbell, K. B. & Bastien, C. The mismatch negativity to frequency deviant stimuli during natural sleep. *Electroencephalogr. Clin. Neurophysiol.* **98**, 493–501 (1996).

28. Atienza, M., L. Cantero, J. & Gómez, C. M. The mismatch negativity component reveals the sensory memory during REM sleep in humans. *Neurosci. Lett.* **237**, 21–24 (1997).

29. Näätänen, R., Paavilainen, P., Titinen, H., Jiang, D. & Alho, K. Attention and mismatch negativity. *Psychophysiology* vol. 30 436–450 (1993).

30. Bekinschtein, T. A. *et al.* Neural signature of the conscious processing of auditory regularities. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1672–1677 (2009).

31. Perrin, F. *et al.* Brain response to one's own name in vegetative state, minimally conscious state, and locked-in syndrome. *Arch. Neurol.* **63**, 562–569 (2006).

32. Wijnen, V. J. M., van Boxtel, G. J. M., Eilander, H. J. & de Gelder, B. Mismatch negativity predicts recovery from the vegetative state. *Clinical Neurophysiology* vol. 118 597–605 (2007).

33. Melanie, B. *et al.* Preserved Feedforward But Impaired Top-Down Processes in the Vegetative State. *Science (80-.).* **332**, 858–862 (2011).

34. Heinke, W. *et al.* Sequential Effects of Increasing Propofol Sedation on Frontal and Temporal Cortices as Indexed by Auditory Event-related Potentials. *Anesthesiology* **100**, 617–625 (2004).

35. Goltstein, P. M., Montijn, J. S. & Pennartz, C. M. A. Effects of isoflurane anesthesia on ensemble patterns of Ca2+ activity in mouse V1: Reduced direction selectivity independent of increased correlations in cellular activity. *PLoS One* **10**, 1–31 (2015).

36. Tang, M. F., Dickinson, J. E., Visser, T. A. W. & Badcock, D. R. The broad orientation dependence of the motion streak aftereffect reveals interactions between form and motion neurons. *J. Vis.* **15**, (2015).

37. Dickinson, J. E., Harman, C., Tan, O., Almeida, R. A. & Badcock, D. R. Local contextual interactions can result in global shape misperception. *J. Vis.* **12**, 3 (2012).

38. Clifford, C. W. G. Perceptual adaptation: Motion parallels orientation. *Trends Cogn. Sci.* **6**, 136–143 (2002).

39. Westrick, Z. M., Heeger, D. J. & Landy, M. S. Pattern adaptation and normalization reweighting. *J. Neurosci.* **36**, 9805–9816 (2016).

40. Ulanovsky, N., Las, L. & Nelken, I. Processing of low-probability sounds by cortical neurons. *Nat.*

Neurosci. **6**, 391–398 (2003).

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

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

43. Zavitz, E., Yu, H. H., Rowe, E. G., Rosa, M. G. P. & Price, N. S. C. Rapid adaptation induces persistent biases in population codes for visual motion. *J. Neurosci.* **36**, 4579–4590 (2016).

44. Gibson, B. Y. J. J. & Radner, M. Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. *J. Exp. Psychol.* **20**, 453–467 (1937).

45. von Helmholtz, H. *Treatise on Physiological Optics Vol. III*. vol. 3 (Dover Publications, 1867).

46. Garrido, M. I. *et al.* The functional anatomy of the MMN: A DCM study of the roving paradigm. *Neuroimage* **42**, 936–944 (2008).

47. Ayaz, A., Saleem, A. B., Schölvinck, M. L. & Carandini, M. Locomotion controls spatial integration in mouse visual cortex. *Curr. Biol.* **23**, 890–894 (2013).

48. Niell, C. M. & Stryker, M. P. Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. *Neuron* **65**, 472–479 (2010).

49. Summerfeld, C., Wyart, V., Johnen, V. M. & de Gardelle, V. Human scalp electroencephalography reveals that repetiti suppression varies with expectation. *Front. Hum. Neurosci.* **5**, 1–13 (2011).

50. Todorovic, A., van Ede, F., Maris, E. & de Lange, F. P. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An MEG study. *J. Neurosci.* **31**, 9118–9123 (2011).

51. Bastos, A. M. *et al.* Canonical Microcircuits for Predictive Coding. *Neuron* **76**, 695–711 (2012).

52. Brouwer, G. J. & Heeger, D. J. Decoding and reconstructing color from responses in human visual cortex. *J. Neurosci.* **29**, 13992–14003 (2009).

53. Ester, E. F., Sprague, T. C. & Serences, J. T. Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron* **87**, 893–905 (2015).

54. Sprague, T. C. & Serences, J. T. Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nat. Neurosci.* **16**, 1879–1887 (2013).

55. Tang, M. F. *et al.* Neural dynamics of the attentional blink revealed by encoding orientation selectivity during rapid visual presentation. *Nat. Commun.* **11**, 1–14 (2020).

56. Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. & Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* **22**, 1677–1686 (2019).

57. Park, I. M., Meister, M. L. R., Huk, A. C. & Pillow, J. W. Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nat. Neurosci.* **17**, 1395–1403 (2014).

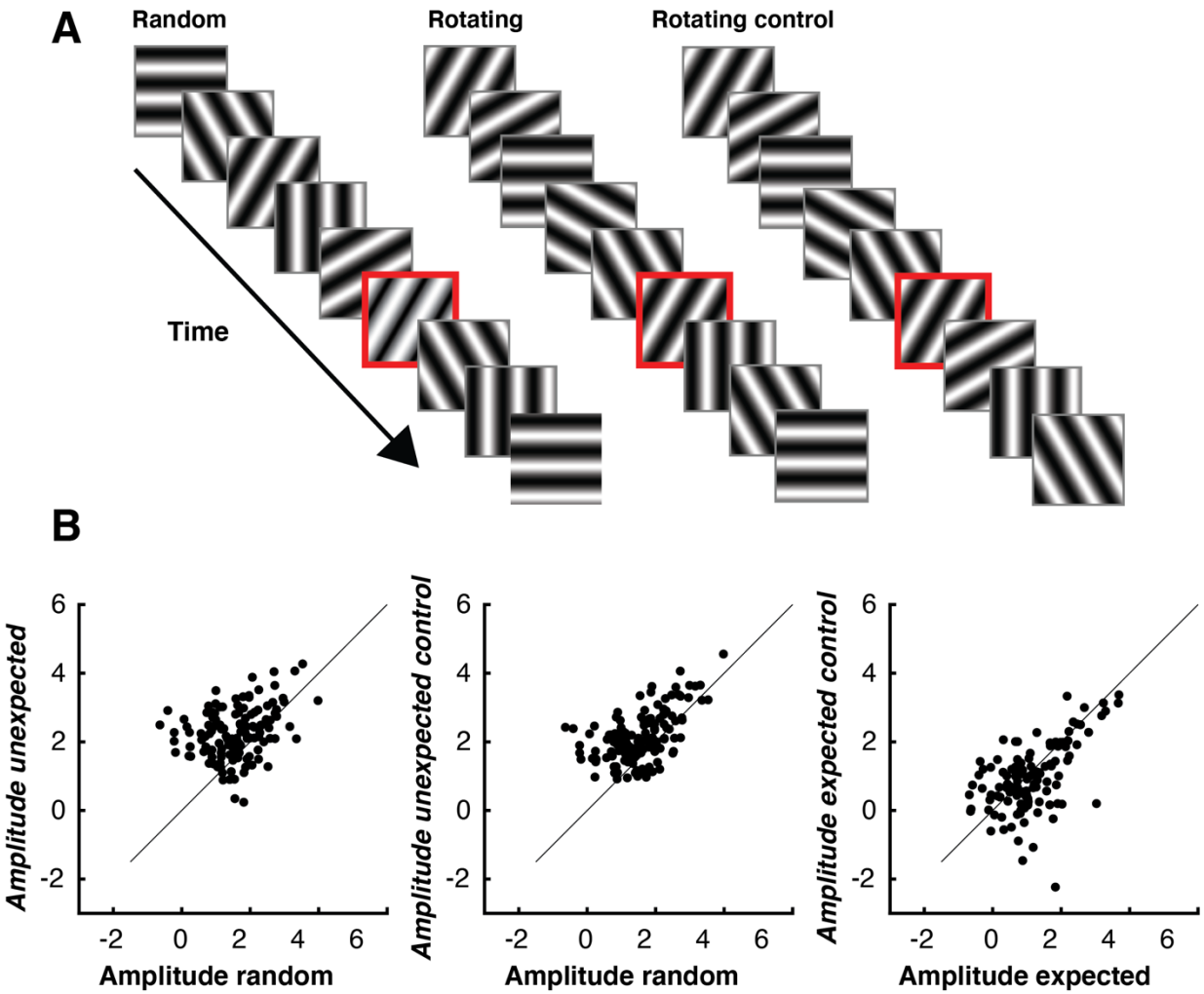
58. Wolff, M. J., Jochim, J., Akyürek, E. G. & Stokes, M. G. Dynamic hidden states underlying working-memory-guided behavior. *Nat. Neurosci.* **20**, 864–871 (2017).

59. Dickinson, J. E., Mighall, H. K., Almeida, R. A., Bell, J. & Badcock, D. R. Rapidly acquired shape and face aftereffects are retinotopic and local in origin. *Vision Res.* **65**, 1–11 (2012).

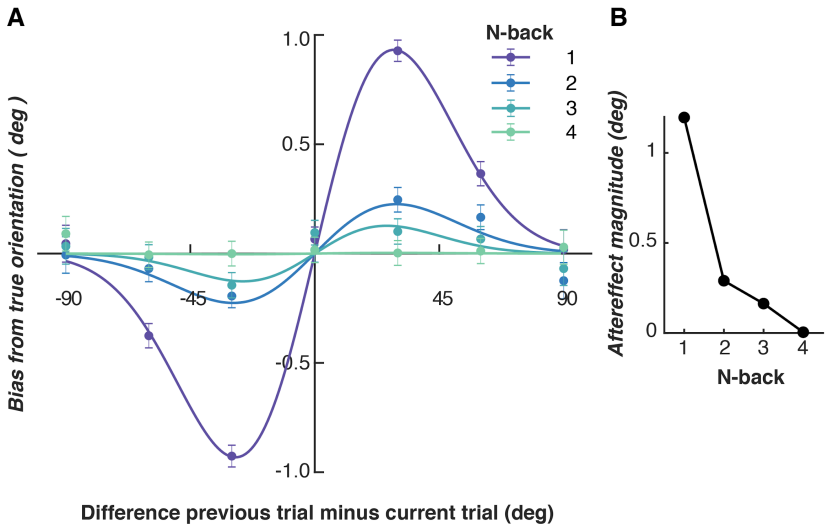
60. Dickinson, J. E., Almeida, R. A., Bell, J. & Badcock, D. R. Global shape aftereffects have a local substrate: A tilt aftereffect field. *J. Vis.* **10**, 1–12 (2010).

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

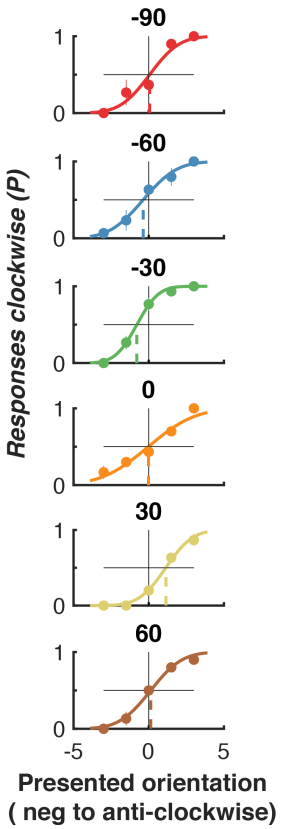
Supplementary figures



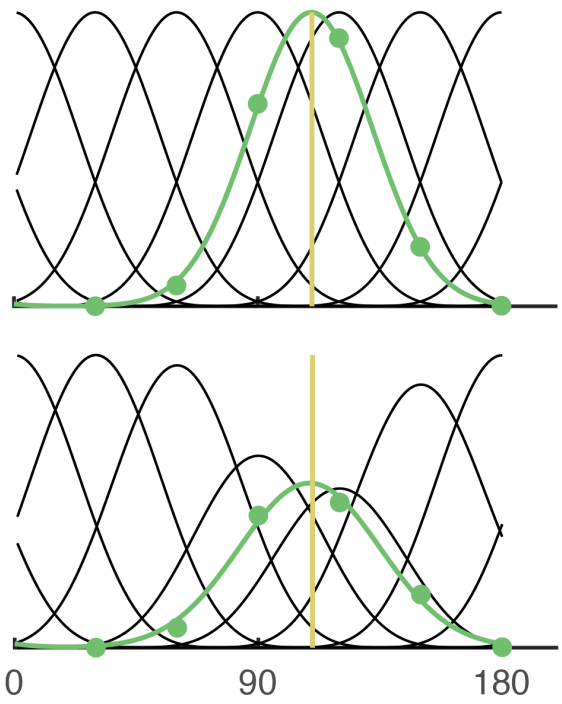
Supplementary Figure 1. Control condition to determine whether the rotation sequence caused the increased gain in the Unexpected trials. **((B).** Fitted gain values for each neuron for the three conditions. This subset of neurons showed the same effect in the original two conditions, with an increase in gain in the *Unexpected* compared to *Random* condition (t(129 = 7.74, $p < 0.001$). This effect was maintained when comparing the *Random* to the *Unexpected control* condition (t(129) = 7.81 $p < 0.001$). There was no significant difference between *Unexpected* and *Unexpected control* conditions (t(129 = 1.81, $p = 0.07$).



Supplementary Figure 2. A version of the model incorporating adaptation-only (expectation modulation set to 0) accurately predicts serial dependence effects seen in human psychophysical and neuronal data in response to oriented streams of gratings. **A** The magnitude of expected perceived bias (difference between the orientation presented and the orientation decoded from the population response by the vector mean) by the orientation of the presented and previous stimulus. **B** First derivative of Gaussian functions are fitted to the data to quantify the magnitude of the aftereffect (given by the amplitude parameter). The model predicts the immediately prior stimulus (N back = 1) elicits the largest aftereffect which decreases with larger N back steps.



Supplementary Figure 3. A single subject's results from the psychophysical task using a Method of Constant Stimuli to determine how expectation affects the perceived orientation of a target. The true orientation of the target Gabor was varied from -6 (anti-clockwise) to +6 (clockwise) degrees from vertical in 3 degree steps (x-axis). On each trial, the observer determined whether the orientation was clockwise or anti-clockwise. Cumulative Gaussian functions were fitted to the results to determine whether the point of subjective equality (i.e., the point at which the observer is equally likely to judge the orientation as clockwise or anticlockwise) is biased away from vertical. Separate functions are fit for each different expected orientation (-90 to +60 in 30 degree steps, different panels).



Supplementary Figure 4. A simple schematic example of the model. (A) The model in an unadapted state, showing its response to a 110° stimulus (yellow line). The green line shows the model's response to the stimulus in each channel. (B) Applying adaptation gain at 110° reduces the model's sensitivity to nearby, but not distant, stimuli. The model's response (green line) is reduced in magnitude relative to panel A when the same test stimuli are applied.