Expectation reshapes V4 neuronal activity and improves perceptual performance

Patricia L. Stan[1,2,3,4] and Matthew A. Smith[2,3,4]

[1]Center for Neuroscience, University of Pittsburgh
[2]Department of Biomedical Engineering, Carnegie Mellon University
[3]Neuroscience Institute, Carnegie Mellon University
[4]Center for the Neural Basis of Cognition, Carnegie Mellon University and University of Pittsburgh

Abstract

Recent visual experience creates expectations that heavily influence our visual perception. How does expectation shape the activity of cortical neurons to allow for improved perceptual discrimination of visual inputs? We recorded from populations of neurons in visual cortical area V4 while monkeys performed a natural image change detection task under different expectation conditions. We found that higher expectation led to an improvement in the ability to detect a change in an image. This improvement was associated with decreased neural responses to the image, providing evidence that a reduction in activity can improve stimulus encoding. The decrease in response could not be fully explained by short-timescale adaptation, suggesting partially separate mechanisms of adaptation and expectation. Additionally, higher expectation was associated with decreased trial-to-trial shared variability, indicating that a reduction in variability is a key means by which expectation influences perception. Taken together, the results of our study contribute to an understanding of how visual experience and expectation can shape our perception and behavior through modulating activity patterns across the visual cortex.

Introduction

Prior knowledge biases our perception of incoming sensory input. This can occur through a mixture of top-down cognitive factors and an accumulation of sensory experience that governs how our sensory systems process and interpret incoming information. Prior knowledge stemming from recent sensory experience creates expectations of upcoming sensory events. An increase in expectation, or the probability of a sensory event occurring, both improves our perceptual abilities and shapes our neural activity [1,2]. In the visual system, expectation has been shown to bias our visual perception [1] and improve our ability to more rapidly or accurately detect changes in incoming visual input [3–5]. Given that expectation has a profound impact on perception and behavior, understanding the neural mechanisms by which our brains incorporate sensory contexts and experiences is critical.

Expectations can be created in multiple ways, either through explicit cues that indicate the probability of an event occurring [6–14] or through repeated exposure leading to accumulation of evidence over time [15–21]. Studies using both approaches have primarily reported that expectation decreases neural activity, known as “expectation suppression”. The majority of this work interrogating the neural effects of expectation has used fMRI or EEG to look at macroscopic neural signatures of expectation across the brain [6,8,11,22–24] with an emphasis on higher visual areas such as fusiform face area [7,9,21]. Far fewer studies have evaluated the effects of expectation at the neuronal level, but evidence from neurons in macaque IT likewise support that expectation decreases responses [18–20,25,26]. Additionally, theoretical work [27–32] and experimental studies of cognitive factors such as attention [33–39] indicate that changes in the trial-to-trial correlated activity of a population of neurons can improve discriminability of stimuli, but no study has assessed the impact of expectation on these population level interactions. The lack of neuronal level interrogation, particularly in lower level visual areas, leaves many unanswered questions about how expectation shapes the responses of and interactions between visual neurons.

The impact of recent visual experience has also been studied in the context of adaptation, where a sensory stimulus is repeated on short timescales (i.e. milliseconds to seconds). The most common finding under the umbrella of adaptation effects is that of repetition suppression - a reduction in response following the repeated presentation of a stimulus [40–43]. Studies investigating if expectation modulates repetition suppression have been met with mixed results [15,16,44–49]. Therefore, it remains unclear whether short-timescale adaptation and the buildup of expectation through repeated exposure share a common mechanism. Additionally, the vast majority of expectation studies use task paradigms where expectation is irrelevant to the subject's behavioral performance [9,14–20,23,24,26] (but see [6,12]). Such designs do not offer a measurable behavioral effect of expectation, prohibiting a clear connection between neural findings and the perceptual effects of expectation.

Our goal was to investigate the neural impact of expectation on visual cortical area V4 during a task where expectation was decision-relevant and could improve visual perception. We designed a natural scene change detection task that used stimulus probability to create high and low expectation conditions and assessed the effects of expectation on V4 neuron firing

rates and population activity. We found that higher expectation led to a suppression of V4 neural activity that was correlated with an improvement in behavioral performance and which could not be fully explained by short-timescale adaptation. Additionally, we found that higher expectation was associated with reduced trial-to-trial shared variability among neurons, and this reduction was related to correct trial outcomes. Our work thus establishes neural correlates of expectation, at the level of individual neurons and populations, that appear to underlie the behavioral improvements associated with expectation.

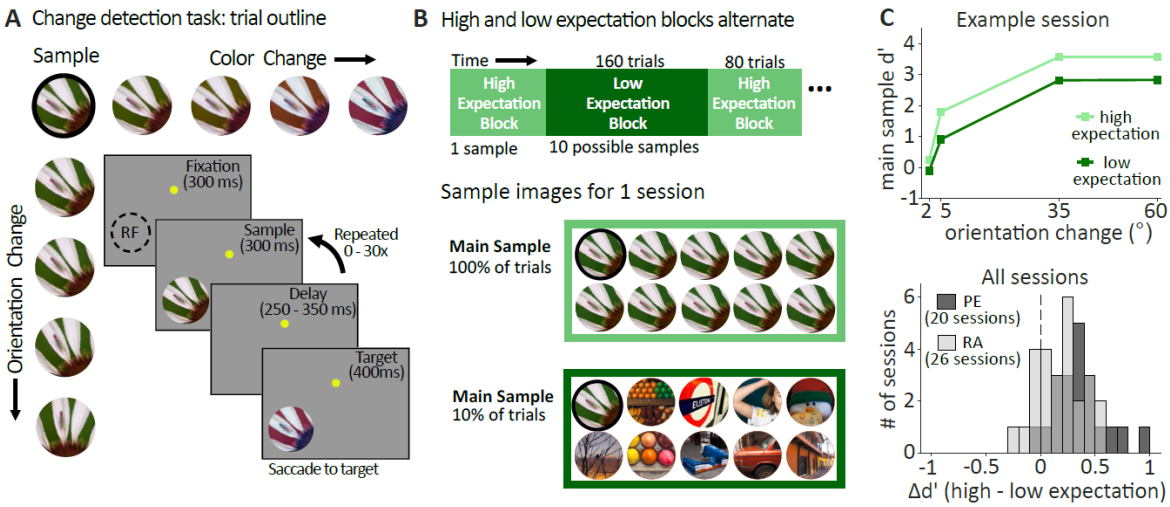

Figure 1. Modulating stimulus probability to create a higher expectation led to improved perceptual performance on a natural image change detection task

A) Natural image change detection task outline. A trial began when the monkey fixated on the yellow dot. On each trial, a sample image was chosen and repeatedly flashed on the screen in the area of the V4 receptive fields. On a flash unknown to the monkey (fixed probability), a feature of the image changed, and the monkey was rewarded for a saccade to the changed image (the target). Targets could be one of 4 difficulties. Sessions of color and orientation changes were alternated. B) Stimulus probability was used to create high and low expectation conditions. Top; High and low expectation blocks of trials alternated throughout the session. Bottom; Example images are shown for each condition for one session. To create high and low expectation conditions, we modulated the stimulus probability of a particular image (the "main sample"). In high expectation blocks, the main sample was the sample image on 100% of trials. In low expectation blocks, the main sample was the sample image on only 10% of trials, and the other 90% of trials used one of 9 different images as the sample image (each shown on 10% of trials). The main sample image for an example session is circled in black. C) Top; The behavioral performance for the main sample image in one session showed an improvement in discriminability (*d'*) for detecting each change difficulty in the high expectation condition compared to the low expectation condition. Bottom; Changes in *d'* across sessions for two monkeys indicated a robust improvement in discriminability in the high expectation condition (paired t-test, $p<0.001$ for RA, $p<0.0001$ for PE). For each session, the difference in *d'* (collapsed across target difficulties) for the main sample was determined in each condition. Values to the left of 0 (dashed line) indicate performance was better in the low expectation condition, values to the right of 0 indicate performance was better in the high expectation condition.

Results

Effects of expectation on task performance

Two monkeys performed a natural scene change detection task under high and low expectation conditions (Fig 1). On every trial, a sample image flashed on the screen, and the monkey had to

detect when a feature of the sample image had changed. There was a fixed probability of a change happening on every flash (34% or 36%, see *Methods*), meaning that 35% of trials consisted of one flash of the sample image and then the target, 23% of trials had two flashes of the sample image then the target, and so on. Because there was a fixed probability of a change, the animal could not anticipate when the change would occur. In the high expectation condition, a single image (the main sample) was used on every trial. In the low expectation condition, the main sample was used on only 10% of trials (with 9 other images comprising the other 90% of trials). By modulating the probability of the main sample being the sample image on a given trial, we modulated the animal's expectation of an upcoming image. Notably, in the 10 image condition, the main sample is not "surprising" (i.e. less likely than other images) it is simply less expected than in the 1 image condition, allowing us to interpret our effects in the context of expectation as opposed to surprise. Each day, a new main sample and accompanying 9 images were chosen (see *Methods* for a description of this process), and a change type was selected (either color or orientation). Therefore, our results generalize across different images and different change features. There were 4 change difficulties, and high and low expectation blocks of trials alternated throughout the day. Additionally, we alternated which block type (i.e., high or low expectation) was the starting block for the day.

To determine if higher expectation improves the ability to detect a change in an image, we analyzed behavior in terms of discriminability (d-prime, or *d'*), criterion, and reaction time. Importantly, we compared the performance to the main sample image in high and low expectation conditions, not to all images in the two conditions. Both monkeys showed a robust increase in *d'* for the main sample image during high expectation, with 77% of sessions (20/26) for monkey RA and 95% of sessions (19/20) for monkey PE having better performance in the high expectation condition (Fig 1C). The increase in *d'* was due to an increase in hit rate in monkey RA, and a decrease in false alarm rate in monkey PE (Supplemental Figure 1A-B). We did not see a consistent expectation effect across animals for criterion (Supplemental Figure 1C; paired t-test; monkey RA showed no difference, $p=0.4086$; monkey PE showed an increase in criterion, $p<0.0001$) nor reaction time (Supplemental Figure 1D; paired t-test; monkey RA showed a small decrease in the high expectation condition, $p=0.0340$; monkey PE showed no difference, $p=0.3393$). Therefore, the remainder of our study focuses on the improvement in discriminability (*d'*). Together, these results show that in situations when an image is expected to occur, subjects demonstrate improved abilities to detect changes in that image.

Alternative explanations for behavioral effects of expectation

We refer to our manipulation as a modulation of expectation because stimulus probability is the primary difference regarding the main sample in each condition. We found that overall behavioral performance was better in blocks in which only one main sample was shown, which we proposed is a direct behavioral correlate of expectation. However, there are two potential alternative interpretations of this behavioral result. First, behavioral performance in animals is known to be influenced directly by task difficulty. In our low expectation blocks in which ten images were used, the behavioral thresholds on the images other than the main sample were not controlled explicitly. That is, they were selected as new images each day without measuring behavioral thresholds in advance. Because of this, the overall difficulty of the low expectation

block varied from day to day based on the difficulty of the change detection task with each of the ten natural images. Second, in the low expectation condition there are in effect two things that differ from the high expectation condition. The primary difference we were concerned with was stimulus probability - the difference in the likelihood that the main sample would appear on any trial. However, the presence of the additional nine images may also have impacted the animal in a different way. Perhaps the difference in performance on the main sample was not because the animal had a perceptual improvement in the high expectation block, but rather that there was a perceptual deficit induced by the presence of the additional images in the low expectation block. We addressed these two alternative explanations in additional analyses and experiments.

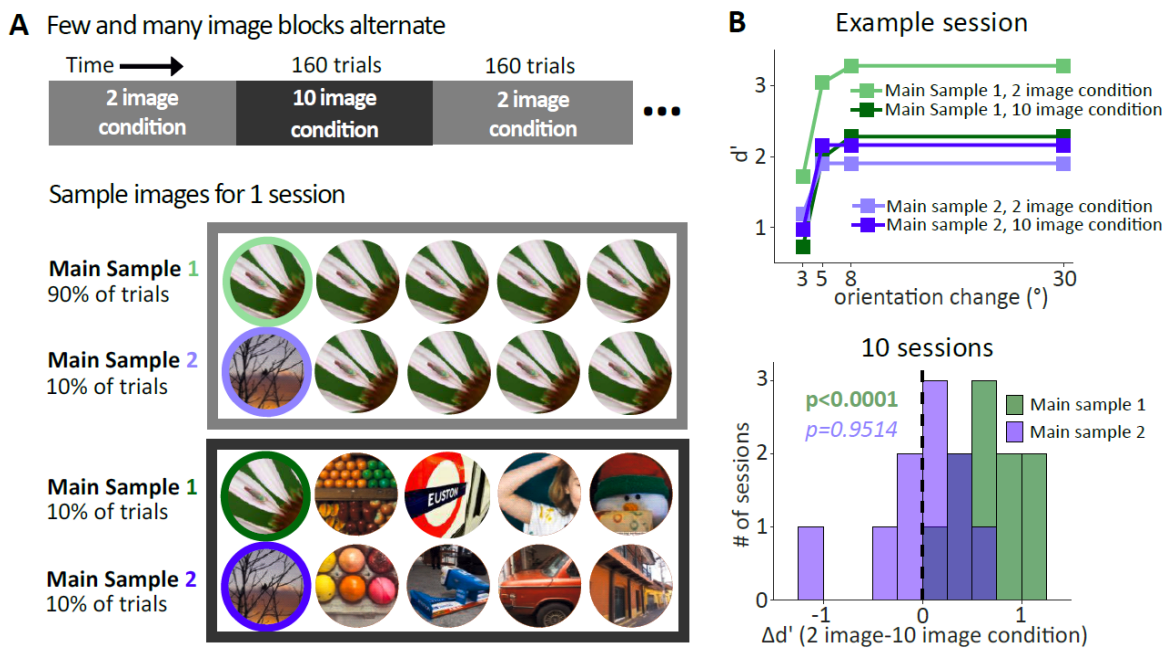

Figure 2. Improvement in behavioral performance for the main sample is due to stimulus probability, not the number of images in each condition

A) Modified task design to test the impact of stimulus probability vs. perceptual load. Blocks of trials with 2 possible sample images and 10 possible sample images alternated throughout the session. The 2 image condition contained two potential sample images, one of which was shown on 90% of trials (main sample 1, green), the other was shown on 10% of trials (main sample 2, purple). The 10 image condition contained 10 potential sample images each shown on 10% of trials (equivalent to original task, Fig 1). Therefore, there were two main samples. B) Behavioral results for main samples 1 and 2. Main sample 1 showed a strong effect of expectation, but main sample 2 elicited similar behavior between blocks (top, *d'* for each main sample in the two conditions in one example session). The change in *d'* between conditions for 10 sessions from monkey RA (bottom). Paired t-test p-values are shown.

An increase in task difficulty has been shown to improve an animal's performance in a change detection task - an effect that has been attributed to an increase in effort or motivation [50–53]. To consider this possibility, we assessed the difficulty of the main sample relative to the other images (Fig S2A), and how that might have impacted the performance for the main sample in the two conditions. We found that in some sessions the overall difficulty of the other images selected for the 10 image condition were harder than the main sample (Fig S2B, negative values on the Y axis) and in other sessions they were easier (Fig S2B, positive values on the Y

axis). There was a small positive trend (significant in one animal) that indicated the overall difficulty of the 10 image condition may have impacted the day-to-day effects of expectation. Irrespective of the value of this relative difficulty, the performance for the main sample was still better when it was alone (Fig S2B, almost all points lie to the right of 0 on the X axis). Taken together, these results suggest that a difference in difficulty was not the main driver of the improvement in discriminability for the main sample in the high expectation condition.

An additional possible explanation of our behavioral results is that the condition with 10 images could be more demanding in terms of perceptual load or working memory. We designed a control experiment that would allow us to distinguish between the effects of varying stimulus probability (which creates differences in expectation) and varying the number of images in each condition (which could create differences in working memory, perceptual load, etc.) by making a small modification to our original task design (Fig 2A). The 10 image condition remained as is, with 10 different images each being shown on 10% of trials. Instead of a 1 image condition, we used a 2 image condition, where one image (main sample 1) was used on 90% of trials, and a second image (main sample 2) was used on 10% of trials. Both main samples were also found in the 10 image condition.

In this control experiment, the stimulus probability of main sample 1 is akin to our original experiment, with only a decrease from 100% to 90% in the 2 image condition, and a 10% probability in both experiments in the 10 image condition. Main sample 2, however, has equal stimulus probability in both conditions (10%). For main sample 1, we recapitulated our initial finding: there was a significant improvement in behavioral performance for main sample 1 in the 2 image (higher stimulus probability) condition compared to the 10 image condition (Fig 2B, paired t-test, $p<0.0001$). This indicates that the reduction from 100% to 90% stimulus probability in this control experiment did not remove the effect of expectation. The key additional comparison this control allowed us to perform was in the behavior of the animal for main sample 2 in the two conditions. The stimulus probability of this image was equal in both blocks, but the perceptual load from working memory demands might be considered higher in the 10 image condition. However, there was no significant difference in behavioral performance for main sample 2 across these conditions (Fig 2B, paired t-test, $p=0.9514$). These results support our initial interpretation: the improvement in discriminability of the main sample in the 1 image high expectation condition is not due to an impairment in performance when a large number of images are used across trials, but rather due to an enhanced expectation imparted by the increased stimulus probability.

Effects of expectation on firing rates

Single neuron firing rates in the visual system reflect both the preferences of each neuron for visual input as well as modulation by numerous cognitive and contextual factors. We sought to understand whether expectation of a particular visual input modulates firing rates of individual neurons. Because there are well known effects of recent visual experience on neuronal responses, typically labeled as short-timescale adaptation or repetition suppression, we focused on the responses to the first visual stimulus presented in each trial. When we compared the firing rates for the main sample (45-300ms after image onset) in high and low expectation

conditions we found, in line with past studies, that average firing rates were decreased with higher expectation (Fig 3). The majority of neurons exhibited expectation suppression of varying magnitudes, although some neurons had enhanced firing rates in the high expectation condition (Fig 3C). The decrease in firing rate was robust throughout the session as we alternated blocks of each expectation condition (Fig 3D), and this effect was evident in every session in both monkeys (Fig 3E). Thus, the predominant effect of expectation on the firing rates of individual neurons was one of suppression.

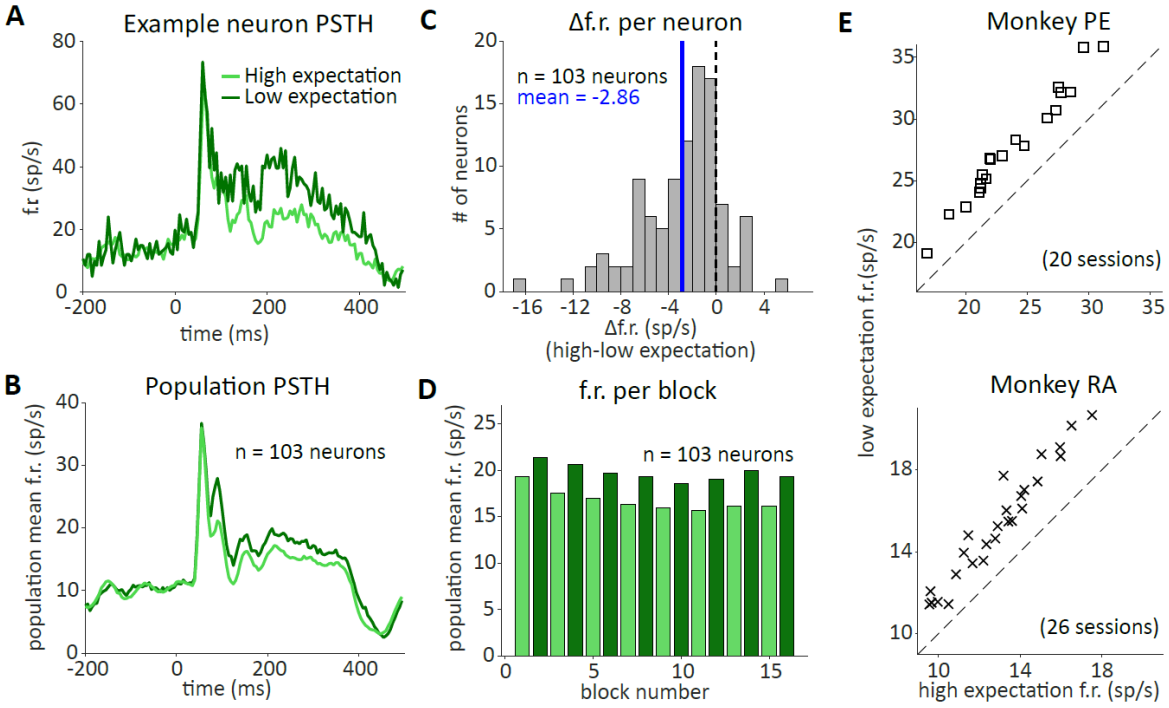

Figure 3. Higher expectation is associated with decreased firing rates

A-D are from one example session from Monkey RA. Only responses to the first flash on trials of the main sample image were included. Unless otherwise noted, responses were averaged across blocks of the same condition. A) Example neuron peri-stimulus time histogram (PSTH) for the main sample image during high and low expectation conditions, aligned to stimulus onset (time 0). Stimulus offset was at 300ms. The vertical axis is the firing rate (f.r.) in spikes per second (sp/s). B) Population average PSTHs across all neurons in 1 session. C) Difference in firing rate (average over an epoch 45-300ms following stimulus onset) for each neuron in an example session. Values to the left of 0 indicate that the firing rate was lower in the high expectation condition. The blue line indicates the mean across neurons. D) The average firing rate across neurons for each block in an example session. E) The average firing rates for the main sample image across neurons were lower in the high than the low expectation conditions for all sessions from two monkeys. Each point represents one session. The dashed line indicates unity slope.

We next sought to link these firing rate changes to the behavioral effects of expectation we had observed in each animal. We analyzed this at a session-by-session level, measuring the firing rate difference along with the behavioral *d'*. We found that these two measures were significantly correlated across sessions in both monkeys (Fig 4A; right-tailed Pearson's correlation; Monkey PE r=0.391, p=0.0440; Monkey RA r=0.340, p=0.0447), indicating that in sessions with a larger suppression of firing rates in the high expectation condition, there was also a greater improvement in discriminability in the high expectation condition. Thus, suppression of firing in individual neurons may represent a neural correlate of expectation.

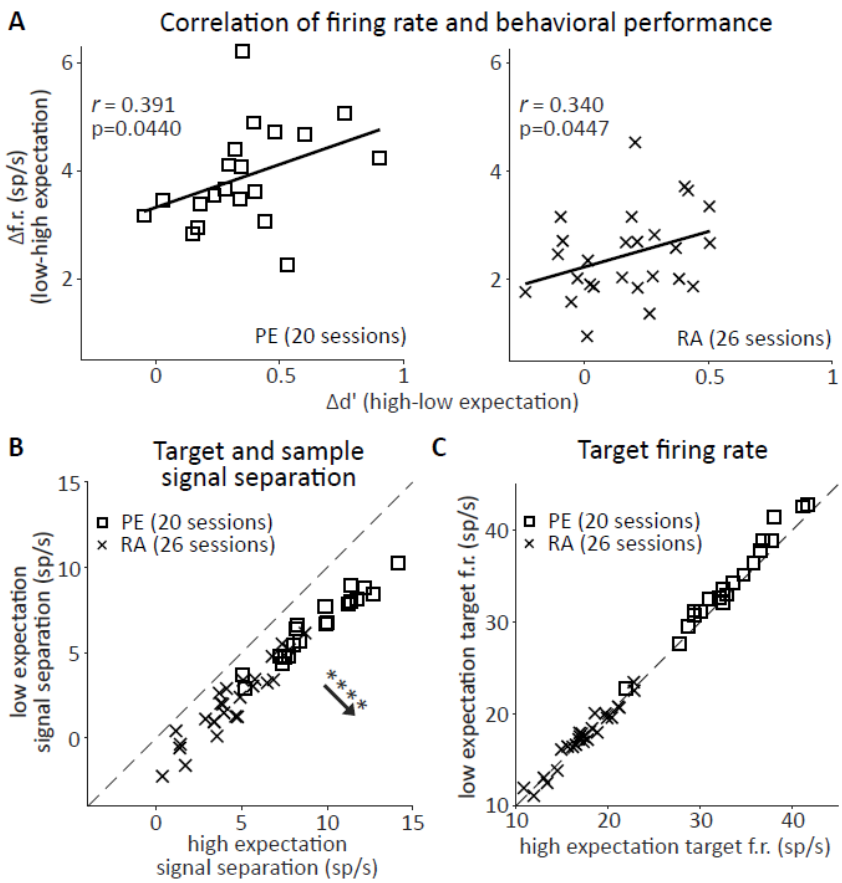

Figure 4. Reduction in activity is correlated with behavioral improvement and leads to larger signal separation

A) The magnitude of the firing rate difference between high and low expectation was correlated with the behavioral effect in *d'* (right-tailed Pearson's correlation). A best-fit line to the data is shown in black. Each point represents one session for monkey PE (left) and RA (right). B) Signal separation was calculated as the average response to the target (collapsed across the target difficulties) minus the average response to the first flash of the main sample image. Unity lines are dotted for reference. The high expectation condition had a significantly higher signal separation (paired t-test p<0.0001). C) The average firing rates across neurons in each session for the target image on trials of the main sample. Each point represents one session. Target firing rates were similar in the low and high expectation conditions. While the small difference was significant (paired t-test, p<0.0001, data combined across animals), this effect was present in only one animal (PE).

Next, we looked to identify a possible mechanistic explanation for how suppression of firing rates could lead to improved stimulus discriminability. Theoretical frameworks such as predictive coding have suggested that suppression of activity to an image could make a new incoming image more salient, as the difference in the signal would be greater and therefore more easily decoded by a higher cortical area [54]. While we found a clear impact of expectation on the neuronal response to the main sample image, the ability of the visual system to discriminate between two images in principle should depend not only on the response to the main sample, but on the response to the target (the main sample with shifted color or orientation). We therefore considered whether the changes in response to the main sample led to a greater separation between responses to sample and target, which might support improved detection performance. For high and low expectation conditions, we subtracted the population average

response to flash 1 of the main sample from the response to the target. We found a robust increase in the separation between sample and target in the high expectation condition (paired t-test, $p<0.0001$, Fig 4B). This effect appeared to be largely driven by the modulation in response to the main sample, because firing rates to the target stimulus were overall very similar in the two expectation conditions (Fig 4C), although there was a small effect on target response in one animal (paired t-test; Monkey PE, $p<0.0001$; Monkey RA, $p=0.3965$). These results indicate that there was a robust increase in signal separation between the main sample image and the target image - an important potential neural substrate for the behavioral effects of expectation - which was largely driven by changes in the neuronal response landscape that were forged before the target image was shown.

The role of stimulus adaptation

One possible mechanism underlying the decrease in firing rates we observe with heightened expectation is that of short-timescale adaptation. Adaptation refers to the changes in neural activity that occur when a sensory input is repeated [40,41]. Suppression of neural responses (i.e. repetition suppression [55,56]) is most common and can occur at various timescales ranging from several hundreds of milliseconds to many seconds following prolonged exposure. Although our task design involved eye movements between the trials that provided variety to the visual inputs, and there were variable lengths of pauses between trials depending on how quickly the animal engaged in the task, it is still that case that in the high expectation block the recent visual input was dominated by repeated presentations of the main sample image. Thus, the well-studied effects of short-timescale adaptation on neurons could be the means by which expectation modulated firing rates and in turn behavior. If the firing rate changes seen between high and low expectation conditions were caused by the same mechanisms as within-trial adaptation, we would make the following predictions: 1) the amount of firing rate change due to adaptation and expectation should be correlated across neurons, and 2) for each neuron, the sign of firing rate change due to adaptation and expectation should match (e.g. a neuron whose firing rate decreases with adaptation should also have a firing rate decrease with expectation).

To assess the effect of expectation, we focused our analysis on the difference in firing rates between flash 1 of the main sample in the high and low expectation conditions (as was done in Fig 3 & 4). We refer to this difference as "expectation" (Fig 5A, left). Constraining to just the first flash allowed us to better isolate the effect of expectation prior to any short-timescale (within-trial) adaptation effects. To assess the effect of short-timescale adaptation, we focused our analysis on the difference in firing rates between flash 1 and flash 2 of the main sample within a trial in each expectation condition (Fig 5A, right). We refer to this difference as "adaptation" which can take the form of repetition suppression or repetition facilitation. When comparing the effects of expectation and short-timescale adaptation (Fig 5B & 5D), we used each neuron’s adaptation value from the low expectation condition, as the adaptation found in that condition was expected to be solely confined to the time scale of individual trials since the sample images varied from trial to trial.

First, we looked to see if the magnitude of firing rate changes due to adaptation and expectation were correlated. For each session, we measured the correlation between each neuron's

adaptation effect (low expectation flash 1 firing rate - low expectation flash 2 firing rate) and expectation effect (low expectation flash 1 firing rate - high expectation flash 1 firing rate). For the majority of sessions, adaptation and expectation were correlated across our populations of neurons (Fig 5B; 96% of sessions Monkey RA, 85% of sessions Monkey PE, Spearman's correlation p<0.05), indicating that neurons that have larger changes in firing rates due to adaptation also tend to have larger changes in firing rates due to expectation.

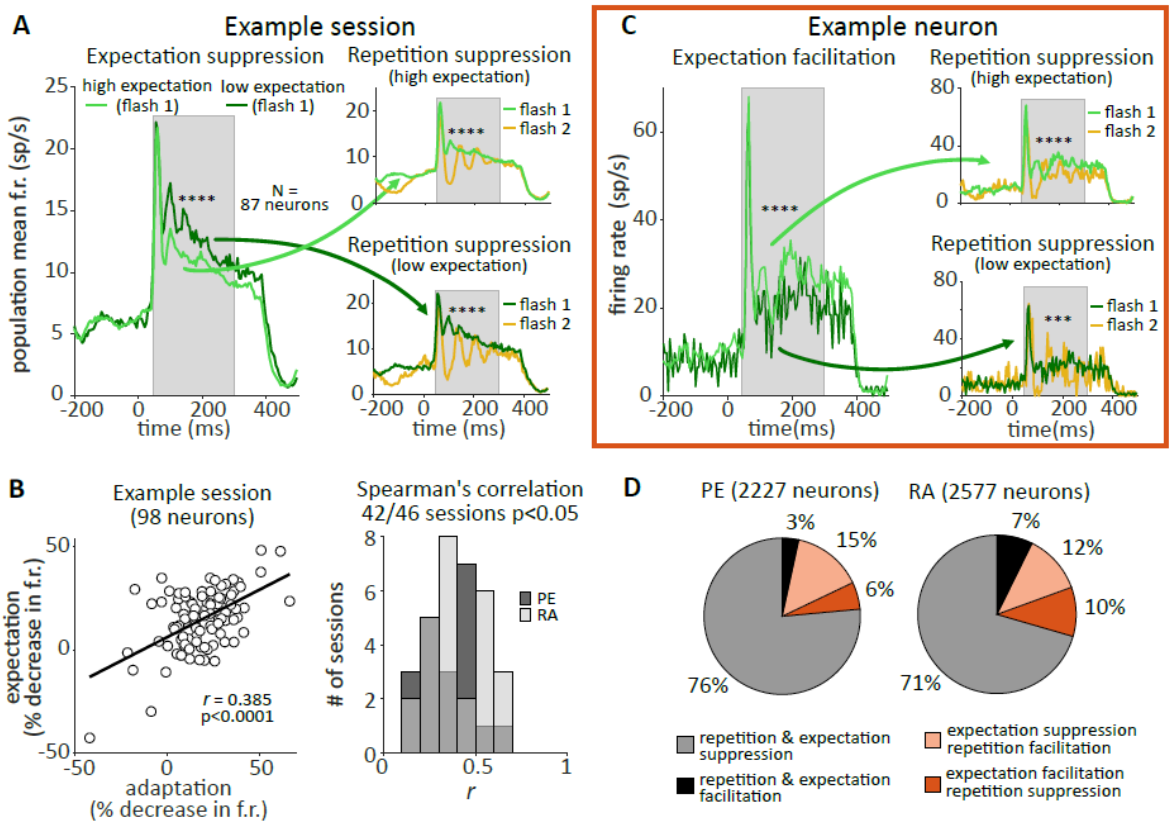

Figure 5. Short-timescale adaptation cannot fully account for expectation suppression

A) Population averaged responses on trials of the main sample image for one example session. PSTHs are aligned to stimulus onset. Left; population PSTH for the main sample in high and low expectation conditions (akin to Fig 3B, but a different session is shown here) showing expectation suppression. Responses were calculated using the first flash of the main sample in each condition. Right; population PSTH for the first and second flash in the high (top) and low (bottom) expectation condition showing repetition suppression, or adaptation. The green PSTHs are identical to those on the left. The gray shaded region indicates the time period over which responses per trial were calculated. B) The change in firing rates due to adaptation and expectation are correlated. Left; example session showing the Spearman's correlation between the percentage change in firing rate due to adaptation in the low expectation condition vs. the percentage change in firing rate due to expectation for each neuron (best-fit line in black). Right; The correlation values for all sessions for each monkey show a strong positive trend. C) The impact of adaptation and expectation can differ for individual neurons. PSTHs for an example neuron show it is possible to exhibit both expectation facilitation (left) and repetition suppression (right). D) The percentage of neurons falling under each combination of facilitation and suppression due to expectation and repetition. Opposing effects of expectation and repetition are in orange, and represented more than one fifth of the neurons in each animal.

Second, if adaptation and expectation shared the same mechanism, we would predict that all neurons exhibiting within-trial repetition suppression would exhibit expectation suppression, and all neurons exhibiting repetition facilitation would exhibit expectation facilitation. For each

neuron, we assessed whether firing rates were increased or decreased with adaptation and expectation. Interestingly, some neurons that had typical decreased responses within a trial (repetition suppression, Fig 5B, right) had overall higher firing rates in the high expectation condition where the main sample is repeated on every trial (Fig 5B, left). In fact, ~20% of neurons did not show the same sign in firing rate change with adaptation as with expectation (Monkey PE 21%, Monkey RA 22%; Fig 17B). This result indicates that although decreases in firing rate are most prevalent with both adaptation and expectation, mechanisms underlying adaptation cannot on their own fully account for changes in firing rates due to expectation.

Effects of expectation on neuronal populations

Our analysis so far has focused on changes in firing rates in individual neurons and in the population average. However, in addition to changes in firing rates, expectation could be acting on population activity to change the structure of variability shared among groups of neurons. Studies of other cognitive factors such as learning [37,57,58], attention [33,34,38,59], and task context [60,61] report a reduction in pairwise noise correlations (the trial-by-trial correlations between the responses of pairs of neurons) with improved perceptual abilities. Therefore, to investigate if changes in expectation affect noise correlations (also known as spike count correlations, or r_{sc}), we first compared the mean and standard deviation of the distributions of r_{sc} for the main sample during high and low expectation. On average, noise correlations were slightly positive in both high and low expectation conditions (average of 0.048 and 0.040, respectively). However, we found no difference when considering all pairs of neurons in the standard deviation of the r_{sc} distribution between low and high expectation (Fig 6A, middle), and an inconsistent effect on mean r_{sc} (Fig 6A, right; paired t-test; Monkey RA no difference in r_{sc} mean, p=0.4427; Monkey PE a small increase in r_{sc} mean, p=0.0007). Therefore, in contrast to studies of other cognitive factors, we do not find evidence that expectation decreases mean noise correlations.

Even though the distribution of noise correlations were not greatly affected by expectation, it is possible that correlations between pairs were still changing between conditions, but in such a way as to maintain the overall distribution of r_{sc}. This would represent changes in the neuronal population variability that might remain hidden from summary statistics of r_{sc} distributions [62]. We sought to determine if r_{sc} for each pair of neurons was changing between conditions. For each pair, we compared the difference in r_{sc} between conditions (high-low expectation r_{sc}) to a shuffled control where we shuffled the trial order in both conditions, recomputed r_{sc}, and took the difference (high-low expectation). We did not see consistent effects on r_{sc} mean (paired t-test; Monkey RA p=0.4411; Monkey PE p=0.0009), but in every session we found an increase in the standard deviation of the difference in r_{sc} distribution compared to the control distribution (Fig 6B). This indicates that more pairs of neurons exhibited changes in r_{sc} between conditions than would be expected by chance. Together, these results indicate that co-fluctuations between neurons are indeed changing between high and low expectation, but the changes are not well captured by a difference in the mean and standard deviation of pairwise noise correlations.

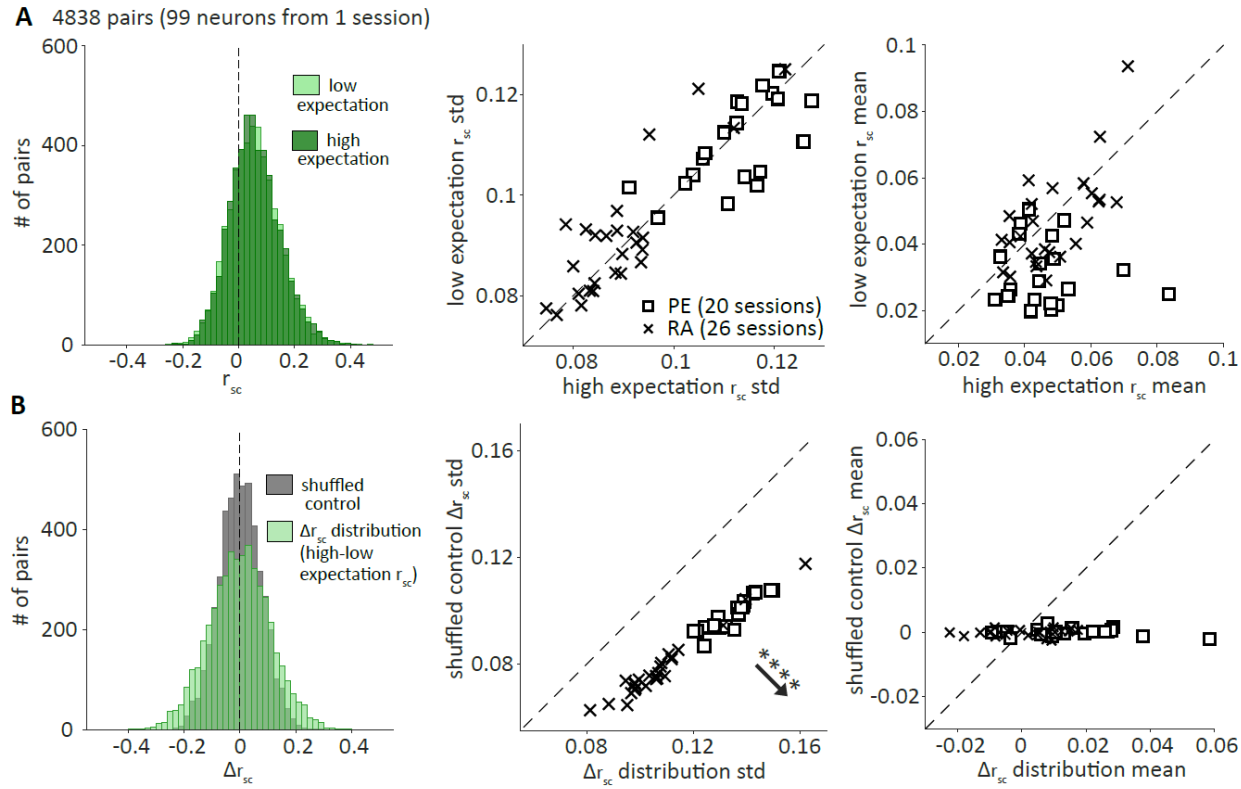

Figure 6. Noise correlations are modulated by expectation

A) Left; The distribution of noise correlations (r_{sc}) for an example session in high and low expectation conditions. r_{sc} was calculated for the first flash of trials of the main sample. The dashed line indicates a value of 0. Middle; The standard deviation of the r_{sc} distributions were similar in both expectation conditions across sessions for two animals . Right; The mean of the r_{sc} distributions for each session were similar in both expectation conditions across sessions. While there was a small difference that was significant (paired t-test, p=0.0017, data combined across animals), this effect was present in only one animal (PE). B) Left; The same example session as in (A) showing the distribution of the change in r_{sc} (Δr_{sc}) between high and low expectation conditions for each pair (green) and the Δr_{sc} for each pair when trials were shuffled (gray). The change in r_{sc} between expectation conditions was more variable in the real data (i.e., the distribution was wider) than in the shuffled control. Middle; The standard deviation of the Δr_{sc} and shuffled control distributions for each session for two monkeys showing a significant difference (paired t-test, p<0.0001), verifying the observation of a single session in the left panel. Right; The mean of the Δr_{sc} and shuffled control distributions for each session for two monkeys. While there was a small difference that was significant (paired t-test, p=0.0018, data combined across animals), this effect was present in only one animal (PE).

Because population-level changes in variability can be sometimes difficult to detect in distributions of r_{sc} [62], we considered an alternative approach – factor analysis (FA) – that would allow us to characterize the structure of population covariability [63]. FA is a dimensionality reduction method that has revealed changes in population covariability during decision making [64–66], learning [37,67,68], and attention [36,62,69–71]. We analyzed the shared variability of neuronal populations separately in the two conditions, high and low expectation, using factor analysis applied to the spikes from every flash of the main sample (see *Methods*). To characterize the structure of population activity, we looked at 3 commonly reported metrics from FA: percent shared variance, shared dimensionality, and loading similarity. Percent shared variance (%sv) is a measure of the strength of shared variability and was calculated as the average %sv across all neurons in each session. Shared dimensionality (d_{shared}) is a measure of the number of

activity patterns present in the population activity and was calculated as the number of dimensions needed to explain 95% of the shared variance. Loading similarity (ls) is a measure of the degree to which neurons co-fluctuate together and was calculated for the first dimension.

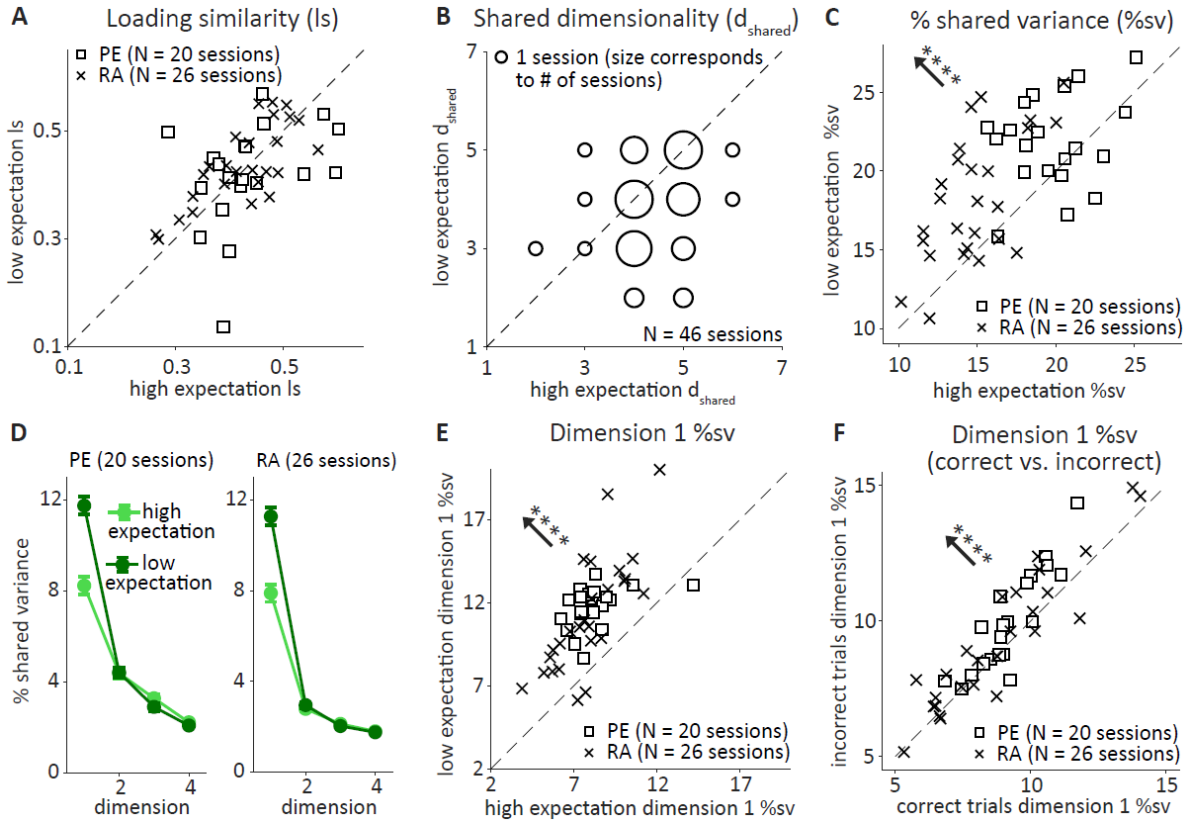

Figure 7. Higher expectation is associated with a decrease in shared variability.

A) Loading similarity for the first dimension in each condition for two monkeys. Each point represents one session. B) Shared dimensionality (d_{shared}) in each condition for two monkeys. The size of the circle is proportional to the # of sessions that fell at the coordinate of each plotted point (collapsed across monkeys). C) Average percent shared variance (%sv) across neurons in each condition for two monkeys. Each point represents one session. There was a significant decrease in %sv in the high expectation condition (paired t-test, $p<0.0001$). D) %sv for each dimension following singular value decomposition. Dimensions after 4 are not plotted as there were very few sessions with a dimensionality greater than 4. The points indicate the mean and standard error across sessions. E) The average %sv across neurons for the first dimension was significantly lower in the high expectation condition (paired t-test, $p<0.0001$). Each point represents one session. F) The %sv for the first dimension on correct vs. incorrect trials showed that correct trials had a significantly lower %sv (paired t-test, $p=0.0001$).

There was no significant difference in loading similarity (Fig 7A, paired t-test; $p=0.9144$), and only a small difference in shared dimensionality which was driven by subject PE (Fig 7B, paired t-test; together $p=0.0022$, RA $p=0.1613$, PE $p=0.0031$). The most consistent result we observed was a decrease in %sv across both animals in the high expectation condition (Fig 7C, paired t-test, together $p<0.0001$, PE $p=0.0153$, RA $p<0.0001$). Recent studies have indicated that variability in cortex is primarily low-rank (i.e., a variability structure in which most of the variability is confined to one dimension). A reduction in low rank variability has been associated with a variety of sensory and cognitive factors that are linked to improved perception [37,62,71,72]. To determine if expectation likewise modulates low-rank variability, we used singular value

decomposition to assess the %sv along each dimension (Fig 7D). In both monkeys, we found a robust and significant decrease in %sv along the first dimension in the high expectation condition (Fig 7E; paired t-test, $p<0.0001$), indicating that higher expectation is associated with a decrease in low-rank variability. Additionally, we wondered if a decrease in low-rank %sv was more generally related to improved behavioral performance. We ran FA separately on correct trials and incorrect (misses and false alarm) trials (see *Methods*) and found that low-rank %sv was lower on correct trials than incorrect trials (Fig 7E, $p=0.0001$), indicating that a decrease in shared variability is associated with correct performance on our perceptual task. Together with existing literature on changes in variability due to other cognitive factors such as attention, this finding supports the notion of shifts in low-rank variability as a canonical mechanism by which reshaping cortical activity can impact behavior.

Discussion

In this study, we sought to identify neural signatures of expectation in populations of visual cortical neurons, and link those neural signatures to the behavioral improvements associated with recent visual experience. We trained monkeys to perform a natural scene change detection task where we modulated stimulus probability to create different expectation conditions that were decision-relevant. At the neuronal level, we found that higher expectation was associated with a suppression of neuronal activity that increased signal separation and was correlated with task performance. The suppression of activity due to expectation was distinct from that due to short-timescale stimulus repetition. At the population level, higher expectation was associated with modulation of noise correlations and a decrease in shared variability. Together, these neural effects were linked with the behavioral improvements associated with expectation.

Our task relied on recent visual experience to create an expectation of upcoming sensory input, leading to a measurable improvement in perceptual abilities. Critically, this allowed us to relate neural findings to a concrete behavioral outcome. Previous studies of expectation and adaptation have often used tasks that are unrelated to the expectation or only require passive viewing of images [9,14–20,23,24,26] (although there are notable exceptions [6,12]), thus necessitating an assumption that an expectation has been created. The absence of a behavioral readout may be one reason why some studies of expectation have variable neural outcomes [73]. By ensuring that expectation is task-relevant as indicated by an improvement in behavioral performance, our experiment maximized the ability to relate neural and behavioral changes of expectation.

At the same time, the introduction of a task also introduces cognitive factors that relate to performance. There are multiple cognitive factors at play in our change detection task that are known to impact behavioral performance, including expectation [1,2,73], feature attention [74,75], task difficulty [50–52], and working memory [76]. Therefore, it is important to consider other cognitive factors that differ in our two task conditions. For instance, the multiple images in the low expectation condition could increase task difficulty, perceptual load, or working memory demands. To address this possibility, we ran an additional experiment that allowed us to separate the effect of many images from stimulus probability (i.e. expectation). Our results indicated that it is in fact higher stimulus probability, and not the lower number of images, that led to a behavioral improvement in the high expectation condition, thus making an increase in

expectation (due to modulated stimulus probability) the most likely explanation for the improvement in change detection.

Our results provide evidence for "expectation suppression" at the single-neuron level. Given that short-timescale adaptation due to stimulus repetition also decreases neural activity [40–43], we sought to determine if mechanisms of short-timescale adaptation could account for the reduction in response in the high expectation condition. We found at the population level a strong association between adaptation and expectation, indicating some level of shared neural circuitry. However, in comparing the effects of adaptation and expectation, we found a subset of neurons were differentially modulated, with decreases in activity due to adaptation, but increases due to expectation (or vice versa). This indicates that short-timescale adaptation effects are insufficient to account for the entirety of the expectation effects on neuronal firing rates. Our results are thus in line with other studies that suggest expectation and repetition suppression are separable neural effects [49,77,78]. Furthermore, there is limited data showing that adaptation leads to an improvement in behavior and perception [42], unlike the robust changes we observed due to expectation. Overall, regardless of the precise mechanism that results in decreased responses, our study is one of few to directly show that a reduction of neural activity in a sensory area is associated with improved perceptual abilities.

How might the single neuron and population results we see benefit sensory processing? Changes in both the signal (the mean sensory evoked activity across neurons) and noise (the trial-to-trial fluctuations in neuronal responses) can impact the information content of a population of neurons. Theoretical frameworks such as predictive coding posit that a decrease in firing rate to an expected stimulus would allow for a larger difference in activity to a new sensory input, thereby making the input more easily detectable [54,56]. Such a process could be implemented by a hierarchical framework analogous to the structure of the primate visual system [79]. Our results provide support for this idea. We found that the difference in response between the sample image and the target image increased in the high expectation condition, thereby increasing the signal-to-noise ratio of the target and making it more easily detectable. Recent work with temporal prediction sequences on a faster time scale did not find neural prediction signals in V1/V4, but did identify them in EEG when the subject was instructed to look for violations of the standard sequence [80]. This suggests that our task design in which the expectation signals are particularly relevant may have been critical to our pattern of results.

Finally, we found that higher expectation was associated with a decrease in shared variability, particularly along the first dimension. A reduction in % shared variance could result in a decrease in the overlap between the representation of two stimuli, which would allow more accurate decoding of responses by a subsequent brain area [62]. This is supported by findings that low-rank decreases in shared variability are associated with multiple cognitive processes that impact behavior such as learning, attention, and task switching [37,72]. Our results provide further evidence that different cognitive factors, perhaps through different mechanisms, may operate through common effects on cortical circuits. Overall, the results of this study represent one of the first reports of the effects of expectation on individual neurons and populations along

with the consequences for behavioral performance. This sheds light on the ways in which expectation affects mid-level visual processing to in turn shape behavior.

Methods

Experimental Subjects

All experimental procedures were conducted in accordance with the United States National Research Council's Guide for the Care and Use of Laboratory Animals and approved by the Institutional Animal Care and Use Committee of Carnegie Mellon University. Data from two male rhesus macaque monkeys were included in this study. Prior to the start of behavioral training, monkeys were surgically implanted with a titanium headpost using aseptic techniques under isoflurane anesthesia. Once monkeys showed proficiency in the task, we surgically implanted a 100-electrode Utah array (Blackrock Microsystems) in left hemisphere V4.

Electrophysiological recordings

Signals from the microelectrode array were bandpass filtered (0.3Hz to 7.5 kHz), digitized at 30 kHz, and recorded by a Grapevine system (Ripple Neuro). Waveform segments exceeding a threshold (-3 times the root mean square noise on each array channel) were saved for offline processing. Following data collection, we performed spike sorting in two stages. First, waveforms were automatically sorted as either spikes or noise using a neural network classifier developed in our laboratory [81]. Second, we manually spike-sorted all of the waveforms that passed the first stage using custom software developed in our laboratory [82] using the time-voltage waveforms, interspike interval distribution, and waveform shape characterized by principal component analysis. The output of this spike sorting process included both well-isolated single units and small multi-unit groups with similar waveforms. We refer to each of these as a “neuron”.

Receptive field mapping

All visual stimuli were shown on a luminance-corrected CRT monitor positioned 57 cm away from the monkey's eyes. Monocular eye position and pupil diameter were constantly monitored using an infrared camera (EyeLink 1000). We mapped receptive fields (RFs) of the V4 neurons to determine appropriate size and location for the images used in the visual change detection task. Sinusoidal gratings of 4 possible orientations were displayed one at a time in a grid of positions that covered the likely RF area based on the anatomical location of the implant. For monkey PE, 1.9 degree gratings covered an area 1.9 degrees above fixation to 11.0 degrees below fixation and 1.9 degrees to the left of fixation to 14.7 degrees to the right of fixation. For monkey RA, 0.9 degree gratings covered an area 2.8 degrees above fixation to 7.4 degrees below fixation and 1.9 degrees to the left of fixation to 7.4 degrees to the right of fixation. We identified the center and extent of each neuron's receptive field by analyzing the mean response to gratings presented at each location. The size and position of images for the change detection task were chosen to cover the aggregate of the V4 receptive fields recorded by the array. For Monkey PE, the image was 10.6 degrees in diameter and centered at 4.4 degrees below and 6.2 degrees to the right of fixation. For Monkey RA, the image was 8.0 degrees in diameter and centered at 3.8 degrees below and 3.8 degrees to the right of fixation.

Visual stimulus selection

We chose to use natural images (as opposed to oriented gratings or other parameterized images) in order to capture more diverse and naturalistic visual responses from our populations of neurons. To this end, the precise symbolic content of the image (i.e. how likely our subjects are to encounter the scene in their natural environment) was less important than ensuring that the images selected drove strong and diverse responses in our neuronal population. Natural images were acquired from Yahoo Flickr Creative Commons 100 Million Dataset and consisted of 112x112 pixel images of a variety of natural scenes and objects. To select the images used for daily experiments, we used a stimulus selection paradigm [83] to identify images that drove strong and diverse responses in the particular V4 population being recorded. We ran multiple recording sessions (6 for monkey PE, 2 for monkey RA) where in each session we presented 1200-2000 new natural images. Monkeys performed an active fixation task while passively viewing sequences of 6 (Monkey PE) or 8 (Monkey RA) images. On each trial, 6 or 8 images were pseudo randomly selected from the larger pool for that day. Each image was displayed for 100 ms with a 100 ms inter-stimulus interval. Each image was repeated 10-24 times, depending on the number of images and duration of that session.

After gathering image responses, the 600 images eliciting the strongest population firing rate (average across neurons) were selected. We then manually curated the group of images to eliminate those that were either lacking in color (e.g. a grayscale image) or orientation information (e.g. a blue sky) which would have made it nearly impossible to detect a color or orientation change, respectively. For the remaining 300-400 images, we used principal component analysis (PCA) on the average response to each image to identify the principal components (PCs) that explained the greatest response variance. We projected the neural responses into this PCA space. To select groups of 10 images for each session of the main change detection task, we used an objective function that favored large responses that were far away from each other in PCA space.

(1) $$f(\mathbf{r}, \mathbf{r}_1, \ldots, \mathbf{r}_M) = \|\mathbf{r}\|_2 + \frac{1}{M}\sum_{j=1}^{M} \|\mathbf{r} - \mathbf{r}_j\|_2$$

Here **r** is a vector of responses for each neuron. *M* is the number of images. The term $\|\mathbf{r}\|_2$ maximizes the responses of the group of neurons, while the term $\frac{1}{M}\sum_{j=1}^{M} \|\mathbf{r} - \mathbf{r}_j\|_2$ maximizes the Euclidean distance between **r** and $\mathbf{r}_1,\ldots,\mathbf{r}_M$. Using this method, we were able to select new images for each session that would result in strong and diverse responses in our recorded V4 neurons.

Natural image change detection task paradigm

Subjects performed a natural image change detection task where the subject's goal was to determine and report when a flashing natural image had changed (Fig 1). Following a fixation period (300 ms), a natural image (the “sample”) was displayed on the screen for 300 ms, followed by a gray screen inter-stimulus interval (lasting 250-350 ms, chosen randomly from a uniform distribution within that range). The image was repeatedly presented ("flashed") a

variable number of times, with a fixed probability (34% for Monkey PE and 36% for Monkey RA) that the target would appear on each flash after the original sample. This created a flat hazard function to discourage guessing. The sample image was fixed for each trial and remained the same across flashes until the target (a changed version of the sample image) appeared. Once the target was presented, the subject had 400 ms to saccade to the target and receive a liquid reward. Subjects had to detect a change in either orientation or color of the sample image on that trial. The change type was fixed for the entirety of a session, and we alternated orientation and color sessions. There were 4 difficulties of targets which were determined by a rotation in orientation or L*a*b color space. Difficulties were titrated to obtain comparable average performance for both change types. To additionally encourage constant behavior within a trial, the reward was adjusted based on the number of flashes, such that longer trials resulted in a higher reward. The ramping of reward was dependent on the animal and set at a value that encouraged a flat false alarm rate over the duration of each trial. This was done to promote relatively constant attention and motivation throughout the trial.

To create differences in expectation of an incoming sensory input, we manipulated the probability of a particular sample image (the "main sample") appearing on a given trial. In high expectation blocks of trials, the main sample was used on 100% of trials, meaning the subject always had to detect a change in the main sample. In low expectation blocks of trials, the main sample was used on only 10% of trials, and 9 other images were used for the remaining 90% (each image appearing on 10% of trials). 10 new images were chosen each day (see Stimulus selection), one of which was selected as the main sample for that day. In the low expectation block, images were pseudo randomly chosen for each trial. High expectation blocks contained 80 trials, and low expectation blocks contained 160 trials (16 trials of each of the 10 images). Block transitions were not cued, so subjects had to rely on visual experience for knowledge of the block identity. Importantly, block transitions were triggered by 80 or 160 completed trials (i.e. either a correct detection of the change, or a failure to notice the image had changed). Broken fixations, false alarms, or failure to engage with the task did not count towards this trial count. Therefore, the time it took to complete a block was quite variable, meaning the animals could not use specific timing cues to determine block switches. High and low expectation blocks alternated throughout the day (a total of 6-16 blocks were completed each session), and which condition was the starting block was alternated for each subsequent session.

To understand how manipulating stimulus probability in this way could produce a difference in expectation, it is helpful to consider the following example. In the high expectation block, following 5 trials where the main sample was used on every trial, at the start of fixation on the 6th trial the observer would likely expect to see the main sample again given their recent visual experience. In a low expectation block, following 5 trials where a different image was the sample on each trial, at the start of fixation on the 6th trial the observer would not know which image to expect (it could be any one of the 10). In this case, if the main sample appeared on the 6th trial, the expectation would be much lower than in the high expectation scenario. Our analyses thus focused on an identical visual input (the main sample) when it was experienced in two different task contexts (high and low expectation). Importantly, the low expectation block contained 10

images of equal probability, therefore the main sample was not unexpected or surprising, allowing us to assess the effects of expectation without the confound of surprise.

Behavioral analysis

There are 3 different trial outcomes that we used for measuring behavioral performance: corrects, misses, and false alarms. A "correct" trial was one in which the monkey succeeded in detecting a change in the sample image, made a saccade to the target within the appropriate time window, and received a reward. A "miss" trial was one in which the sample changed, but the monkey failed to detect the change and continued fixating until the target display ended. A "false alarm" trial was any time the monkey made a saccade toward a flashed sample image to report a change, but the sample image had not yet changed. Because the change could never occur on the first flash, a false alarm could only occur on flash 2 or later. The fourth true outcome in a two-alternative forced choice task, a “correct withhold”, also occurred in our task. In each trial if the animal maintained fixation through flashes of the sample image, those flashes were considered “correct withholds”. Because every trial ended with a change if the animal maintained fixation long enough, the trial itself could not terminate in a correct withhold.

We quantified several measures of behavioral performance for the main sample separately for high and low expectation conditions. Blocks of the same condition were combined. Depending on the analysis, we only included trials for a particular stimulus in the low expectation condition (e.g. when determining the performance for the main sample, we exclude trials of the other 9 stimuli). Hit rate was calculated as HR = (# correct trials)/(# of correct + miss trials). False alarm rate was calculated as FAR = (# of false alarm trials/total number of opportunities to false alarm). The opportunities to false alarm were the sum of total flashes where the subject could have looked at the image (leading to a false alarm trial) but did not (i.e., correct withholds). Using hit rate and false alarm rate, we calculated two measures from signal detection theory: d-prime (d') and criterion. d' is a measure of sensitivity that allows us to measure a subject's ability to detect a signal. d' = z(HR)-z(FAR) where z(HR) and z(FAR) are the z transforms of hit rate and false alarm rate. Criterion (c) is a measure of the bias in reporting a signal and is calculated as c= -0.5(z(HR) + z(FAR)). Additionally, we determined reaction times for correct trials, calculated as the time between the target onset and the moment the subject's eyes reached the target.

For Supplemental Figure 2, relative difficulty was calculated as the average performance (d') for the other 9 images in the low expectation condition minus the d' of the main sample in the low expectation condition. Positive values indicated that the performance for the main sample was worse than the other images, therefore the main sample was on average a more difficult image in which to detect a change. Negative values indicated that the performance for the main sample was better than the other images, therefore the main sample was on average a less difficult image in which to detect a change.

Stimulus probability control task

We collected 10 sessions of data from 1 monkey on a modified task design to determine if the number of images in each block had an effect on behavioral performance. The overall trial structure of the task was the same as described above, the only difference was in the number and frequency of images shown in the different conditions (Fig 2). The low expectation blocks described above remained unchanged, we simply renamed it as the "10 images" condition to remove any assumptions of expectation effects. The second condition (the "2 images" condition) contained two images, one of which appeared on 10% of trials, the other which appeared on 90% of trials. Therefore, we had two "main samples". Main sample 1 was the image that most closely matched our initial task design: it had a high probability of appearing in the 2 images condition (90%) and a lower probability of appearing in the many images condition (10%). Main sample 2 was the image that had equal probability of appearing in the two conditions. Behavioral performance was calculated separately for each image in each condition, collapsed across blocks of the same condition. We report the change in *d'* as the performance in the 2 image block minus the performance in the many images block.

Analysis of neuronal responses

We excluded from further analysis all neurons with a firing rate less than 1 spike per second (sp/s) to the main sample. Peri-stimulus time histograms (PSTHs) were created by aligning spike trains to stimulus onset and averaging the spike rate in 5ms time bins across trials. Because we detected effects during both the transient and sustained portion of the response (Fig 3), all analyses of neuronal firing rates used the 45-300ms window following stimulus onset. Only complete presentations of the first flash were included in the analyses for Figures 3-6. Figure 7 included all complete flashes of the main sample because the greater trial count provided by using all flashes is important given trial limitations of factor analysis [84]. Unless otherwise noted, firing rates were averaged across blocks of the same condition. Firing rates to the target were calculated during the 45-100ms bin following target onset on correct trials. This was to avoid any possible artifacts due to eye movements (e.g. the saccade to the target to receive a reward) which can occur after 100ms.

Adaptation analyses

We sought to compare changes in firing rates due to expectation with known short-timescale adaptation that occurs within each trial due to the repeated flashes of an identical visual stimulus. We defined "adaptation" as the change in firing rate between flash 1 and flash 2 in each condition, and "expectation" as the change in firing rate between flash 1 in the high expectation condition and flash 1 in the low expectation condition. For analyses involving adaptation, only trials with 2 or more flashes were included, even when determining firing rates on the first flash. For each neuron, we calculated its adaptation and expectation effect. To determine if there was a correlation between the decrease due to expectation and the decrease due to adaptation, for each session we measured the Spearman's correlation between the % decrease in firing due to expectation and the % decrease in firing rate due to adaptation in the low expectation condition (Fig 5).

Analysis of noise correlation

The pairwise noise correlation (also referred to as spike-count correlation, or r_{sc}) for the main sample was calculated by taking the Pearson's correlation between two neuron's spike counts on each trial of the main sample. This was then repeated for all pairs of neurons, excluding pairs that were on the same array channel. Given that the number of trials can impact the estimate of noise correlation, we equalized the number of trials by randomly sub-selecting a number of trials from the high expectation condition (which always had more trials) to equal those in the low expectation condition. To ensure that the sub-selection was representative of the noise correlation values we would have obtained if we included all trials, we repeated this sub-selection 2000 times, calculated the average r_{sc} for each sub-selection, and took the average across all 2000 sub-selections. Then, we chose the sub-selection of trials that was closest to the average r_{sc} across all sub-selections, and used those trials for further analysis. r_{sc} was calculated separately for each condition for flash 1. We report the mean and standard deviation of the distributions of noise correlations for the main sample during high and low expectation in Figure 6A.

To assess if each pair's r_{sc} value was changing between conditions, we took the difference in r_{sc} (high-low expectation) and compared to a shuffled control. To obtain the shuffled control, for each neuron and each condition, we shuffled the order of trials and re-calculated r_{sc}. We then took the difference in r_{sc} (high-low expectation) with shuffled trials. We report the mean and standard deviation of the change in r_{sc} distribution and shuffled control in Figure 6B.

Factor analysis

We used the dimensionality reduction method factor analysis (FA) [62–64,71,84–87] to characterize the covariability of our neuronal population under different task conditions. FA allowed us to separate the variance shared among a population of neurons from each neuron's independent variance, making it particularly well suited for analyzing population activity.

We performed FA on the spike of every flash of the main sample separately in high and low expectation conditions. We used the spike counts during the 45-300ms time window following stimulus onset for each complete stimulus presentation (i.e. all flashes) in each trial. Only one neuron per channel was included. FA results are dependent on the number of trials, so we matched trial numbers across conditions. For each session, we sub-selected trials from the high expectation condition (which has more repetitions of the main sample than the low expectation condition). To ensure our results were robust to the variability that comes with randomly selecting a particular set of trials, in the high expectation condition we first ran FA using 30 groups of randomly selected trials. We then calculated the mean %sv and loading similarity across groups, and identified the group that led to the smallest difference from those means. We then used this group for further analysis. For Figure 7E, we ran FA on the spike counts of every flash of the main sample separately for correct and incorrect (misses and false alarm) trials regardless of expectation condition. All subsequent procedures are as described above.

FA is defined as:

(2) $$\mathbf{x} \sim N(\mu, \mathrm{LL}^{\mathrm{T}} + \psi)$$

Where **x** is a *n* x 1 vector of spike counts with n as the number of neurons, **μ** is a *n* x 1 vector of mean spike counts. LL^T is the shared component, where *L* is a *n* x *i* loading matrix with *i* as the number of latent dimensions, and Ψ is a *n* x *n* diagonal matrix containing each neuron's individual variance. To determine the number of latent dimensions *i*, we ran FA using different numbers of latents and chose the value of *i* that maximized the cross-validated log-likelihood.

% shared variance refers to the percentage of the total variance that is shared versus independent [84]. The % shared variance for neuron n is calculated as:

(3) $$\%\mathrm{sv}_{\mathrm{n}} = \frac{\mathrm{L_n L_n^T}}{\mathrm{L_n L_n^T} + \psi_{\mathrm{n}}}$$

In figure 7C we reported the average %sv across neurons. In figure 7C-E we used singular value decomposition to partition the shared variance along each latent. We then computed the %sv for each neuron *n* along each latent *i* as:

(4) $$\%\mathrm{sv}_{i\mathrm{n}} = \frac{\lambda_i \mathrm{v}_{i\mathrm{n}}^2}{\mathrm{L_n L_n^T} + \psi_{\mathrm{n}}}$$

Where λ is the eigenvalue of LL^T corresponding to the *i*th latent, and v_{in} is the *n*th entry in the *i*th eigenvector of LL^T. We then averaged this value across all neurons to determine the %sv along each dimension. In figure 7D-E we report the average %sv across neurons for the first dimension.

Loading similarity is the similarity of loading weights across neurons for each latent [88]. We compute loading similarity as:

(5) $$\text{loading similarity } (\mathbf{u}_{\mathrm{i}}) = 1 - \frac{\mathrm{var}(\mathbf{u}_{\mathrm{i}})}{1/\mathrm{n}}$$

Where *n* is the number of neurons, and u_i is an *n* x 1 vector of weights for a given latent *i*. A value of 1 indicates that the weights for all neurons are the same (either all positive or all negative). A value of 0 indicates that the loading weights are as variable as possible. We report the loading similarity for the first latent (the dominant dimension).

Shared dimensionality (referred to as simply "dimensionality" in the results, or d_{shared}) was computed as the number of latents needed to explain 95% of the shared covariance, LL^T [84].

Firing rate controls for shared variance

Firing rates can affect %sv, with higher firing rates leading to lower %sv values. Therefore, we sought to control for population firing rate differences between the high and low expectation conditions by sub-selecting neurons in each condition that would result in matched population

firing rates. From the analyses described above, we obtained %sv values for each individual neuron in each condition. We also calculated the firing rate for each neuron and ranked them from lowest to highest, separately for each condition. We removed one neuron at a time from each condition, recalculated the difference in population firing rate, and repeated until the difference was reduced while keeping as many neurons as possible. The identity of the neurons in each condition may have differed, but the total number of neurons were matched. Using this method, we significantly brought down the difference in firing rate between high and low expectation in each session (difference in firing rate was <1 sp/s for each session). Then, we re-calculated the average 1st dimension %sv for each condition. We found in these firing rate matched controls that the high expectation condition still had a lower %sv as in the unmatched data, and the results for Fig 7C-D followed the same overall pattern. We repeated this procedure for Figure 7E, and likewise found the results of the firing rate matched control were similar to those shown in Figure 7.

Acknowledgements

We are grateful to Samantha Schmitt for assistance with data collection, Ben Cowley for assistance with natural image stimulus selection, and Qichao Wu for assistance with spike sorting. PLS was supported by NIH EY031975. MAS was supported by NIH EY029250 and MH118929.

References

1. de Lange, F. P., Heilbron, M. & Kok, P. How Do Expectations Shape Perception? *Trends Cogn. Sci.* **22**, 764–779 (2018).
2. Summerfield, C. & de Lange, F. P. Expectation in perceptual decision making: neural and computational mechanisms. *Nat. Rev. Neurosci.* **15**, 745–756 (2014).
3. Wyart, V., Nobre, A. C. & Summerfield, C. Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences* **109**, 3593–3598 (2012).
4. Stein, T. & Peelen, M. V. Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *J. Exp. Psychol. Gen.* **144**, 1089–1104 (2015).
5. Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F. & Seth, A. K. Expectations accelerate entry of visual stimuli into awareness. *J. Vis.* **15**, 13 (2015).
6. 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).
7. Summerfield, C. & Koechlin, E. A Neural Representation of Prior Information during Perceptual Inference. *Neuron* **59**, 336–347 (2008).
8. John-Saaltink, E. S., Utzerath, C., Kok, P., Lau, H. C. & de Lange, F. P. Expectation Suppression in Early Visual Cortex Depends on Task Set. *PLoS One* **10**, e0131172 (2015).
9. Egner, T., Monti, J. M. & Summerfield, C. Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream. *J. Neurosci.* **30**, 16601–16608 (2010).
10. Dunovan, K. & Wheeler, M. E. Computational and neural signatures of pre and post-sensory expectation bias in inferior temporal cortex. *Sci. Rep.* **8**, (2018).
11. Davis, B. & Hasson, U. Predictability of what or where reduces brain activity, but a bottleneck occurs when both are predictable. *Neuroimage* **167**, 224–236 (2018).
12. Bell, A. H., Summerfield, C., Morin, E. L., Malecek, N. J. & Ungerleider, L. G. Encoding of Stimulus Probability in Macaque Inferior Temporal Cortex. *Curr. Biol.* **26**, 2280–2290 (2016).
13. Rungratsameetaweemana, N., Itthipuripat, S., Salazar, A. & Serences, J. T. Expectations Do Not Alter Early Sensory Processing during Perceptual Decision-Making. *J. Neurosci.* **38**, 5632–5648 (2018).
14. Amado, C. *et al.* The contribution of surprise to the prediction based modulation of fMRI responses. *Neuropsychologia* **84**, 105–112 (2016).
15. Kaliukhovich, D. A. & Vogels, R. Stimulus Repetition Probability Does Not Affect Repetition Suppression in Macaque Inferior Temporal Cortex. *Cereb. Cortex* **21**, 1547–1558 (2010).
16. Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M. & Egner, T. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* **11**, 1004–1006 (2008).
17. Kumar, S., Kaposvari, P. & Vogels, R. Encoding of Predictable and Unpredictable Stimuli by Inferior Temporal Cortical Neurons. *J. Cogn. Neurosci.* **29**, 1445–1454 (2017).
18. Meyer, T. & Olson, C. R. Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences* **108**, 19401–19406 (2011).
19. Ramachandran, S., Meyer, T. & Olson, C. R. Prediction suppression and surprise enhancement in monkey inferotemporal cortex. *J. Neurophysiol.* **118**, 374–382 (2017).
20. Ramachandran, S., Meyer, T. & Olson, C. R. Prediction suppression in monkey

inferotemporal cortex depends on the conditional probability between images. *J. Neurophysiol.* **115**, 355–362 (2016).
21. Pajani, A., Kouider, S., Roux, P. & de Gardelle, V. Unsuppressible Repetition Suppression and exemplar-specific Expectation Suppression in the Fusiform Face Area. *Sci. Rep.* **7**, 160 (2017).
22. Melloni, L., Schwiedrzik, C. M., Müller, N., Rodriguez, E. & Singer, W. Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness. *J. Neurosci.* **31**, 1386–1396 (2011).
23. Richter, D., Ekman, M. & de Lange, F. P. Suppressed Sensory Response to Predictable Object Stimuli throughout the Ventral Visual Stream. *J. Neurosci.* **38**, 7452–7461 (2018).
24. Vergnieux, V. & Vogels, R. Statistical Learning Signals for Complex Visual Images in Macaque Early Visual Cortex. *Front. Neurosci.* **14**, (2020).
25. Schwiedrzik, C. M. & Freiwald, W. A. High-Level Prediction Signals in a Low-Level Area of the Macaque Face-Processing Hierarchy. *Neuron* **96**, 89–97.e4 (2017).
26. Kaposvari, P., Kumar, S. & Vogels, R. Statistical Learning Signals in Macaque Inferior Temporal Cortex. *Cereb. Cortex* **28**, 250–266 (2018).
27. Shadlen, M. N. & Newsome, W. T. The Variable Discharge of Cortical Neurons: Implications for Connectivity, Computation, and Information Coding. *J. Neurosci.* **18**, 3870–3896 (1998).
28. Abbott, L. F. & Dayan, P. The Effect of Correlated Variability on the Accuracy of a Population Code. *Neural Comput.* **11**, 91–101 (1999).
29. Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7**, 358–366 (2006).
30. Moreno-Bote, R. *et al.* Information-limiting correlations. *Nat. Neurosci.* **17**, 1410–1417 (2014).
31. Sharpee, T. O. & Berkowitz, J. A. Linking neural responses to behavior with information-preserving population vectors. *Current Opinion in Behavioral Sciences* **29**, 37–44 (2019).
32. Bartolo, R., Saunders, R. C., Mitz, A. R. & Averbeck, B. B. Information-Limiting Correlations in Large Neural Populations. *J. Neurosci.* **40**, 1668–1678 (2020).
33. Cohen, M. R. & Maunsell, J. H. R. Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* **12**, 1594–1600 (2009).
34. Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4. *Neuron* **63**, 879–888 (2009).
35. Ruff, D. A. & Cohen, M. R. Attention can either increase or decrease spike count correlations in visual cortex. *Nat. Neurosci.* **17**, 1591–1597 (2014).
36. Snyder, A. C., Yu, B. M. & Smith, M. A. Distinct population codes for attention in the absence and presence of visual stimulation. *Nat. Commun.* **9**, (2018).
37. Ni, A. M., Ruff, D. A., Alberts, J. J., Symmonds, J. & Cohen, M. R. Learning and attention reveal a general relationship between population activity and behavior. *Science* **359**, 463–465 (2018).
38. Herrero, J. L., Gieselmann, M. A., Sanayei, M. & Thiele, A. Attention-Induced Variance and Noise Correlation Reduction in Macaque V1 Is Mediated by NMDA Receptors. *Neuron* **78**, 729–739 (2013).
39. Ruff, D. A. & Cohen, M. R. Simultaneous multi-area recordings suggest that attention

improves performance by reshaping stimulus representations. *Nat. Neurosci.* **22**, 1669–1676 (2019).
40. Solomon, S. G. & Kohn, A. Moving sensory adaptation beyond suppressive effects in single neurons. *Curr. Biol.* **24**, R1012–22 (2014).
41. Webster, M. A. Visual Adaptation. *Annual Review of Vision Science* **1**, 547–567 (2015).
42. Kohn, A. Visual Adaptation: Physiology, Mechanisms, and Functional Benefits. *J. Neurophysiol.* **97**, 3155–3164 (2007).
43. Weber, A. I., Krishnamurthy, K. & Fairhall, A. L. Coding Principles in Adaptation. *Annu Rev Vis Sci* **5**, 427–449 (2019).
44. Grotheer, M. & Kovács, G. Repetition Probability Effects Depend on Prior Experiences. *J. Neurosci.* **34**, 6640–6646 (2014).
45. Kronbichler, L., Said-Yürekli, S. & Kronbichler, M. Perceptual Expectations of Object Stimuli Modulate Repetition Suppression in a Delayed Repetition Design. *Sci. Rep.* **8**, (2018).
46. Kovacs, G., Kaiser, D., Kaliukhovich, D. A., Vidnyanszky, Z. & Vogels, R. Repetition Probability Does Not Affect fMRI Repetition Suppression for Objects. *Journal of Neuroscience* **33**, 9805–9812 (2013).
47. Kaliukhovich, D. A. & Vogels, R. Neurons in Macaque Inferior Temporal Cortex Show No Surprise Response to Deviants in Visual Oddball Sequences. *Journal of Neuroscience* **34**, 12801–12815 (2014).
48. Vinken, K., de Beeck, H. P. O. & Vogels, R. Face Repetition Probability Does Not Affect Repetition Suppression in Macaque Inferotemporal Cortex. *J. Neurosci.* **38**, 7492–7504 (2018).
49. Grotheer, M. & Kovács, G. The relationship between stimulus repetitions and fulfilled expectations. *Neuropsychologia* **67**, 175–182 (2015).
50. Spitzer, H., Desimone, R. & Moran, J. Increased Attention Enhances Both Behavioral and Neuronal Performance. *Science* **240**, 338–340 (1988).
51. Boudreau, C. E., Williford, T. H. & Maunsell, J. H. R. Effects of Task Difficulty and Target Likelihood in Area V4 of Macaque Monkeys. *J. Neurophysiol.* **96**, 2377–2387 (2006).
52. Ruff, D. A. & Cohen, M. R. Global Cognitive Factors Modulate Correlated Response Variability between V4 Neurons. *J. Neurosci.* **34**, 16408–16416 (2014).
53. Ghosh, S. & Maunsell, J. H. R. Single trial neuronal activity dynamics of attentional intensity in monkey visual area V4. *Nat. Commun.* **12**, (2021).
54. Walsh, K. S., McGovern, D. P., Clark, A. & O'Connell, R. G. Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Ann. N. Y. Acad. Sci.* **1464**, 242–268 (2020).
55. Grill-Spector, K., Henson, R. & Martin, A. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* **10**, 14–23 (2006).
56. Auksztulewicz, R. & Friston, K. Repetition suppression and its contextual determinants in predictive coding. *Cortex* **80**, 125–140 (2016).
57. Gu, Y. *et al.* Perceptual Learning Reduces Interneuronal Correlations in Macaque Visual Cortex. *Neuron* **71**, 750–761 (2011).
58. Jeanne, J. M., Sharpee, T. O. & Gentner, T. Q. Associative Learning Enhances Population Coding by Inverting Interneuronal Correlation Patterns. *Neuron* **78**, 352–363 (2013).
59. Snyder, A. C., Morais, M. J. & Smith, M. A. Dynamics of excitatory and inhibitory networks

are differentially altered by selective attention. *J. Neurophysiol.* **116**, 1807–1820 (2016).
60. Cohen, M. R. & Newsome, W. T. Context-Dependent Changes in Functional Circuitry in Visual Area MT. *Neuron* **60**, 162–173 (2008).
61. Bondy, A. G., Haefner, R. M. & Cumming, B. G. Feedback determines the structure of correlated variability in primary visual cortex. *Nat. Neurosci.* **21**, 598–606 (2018).
62. Umakantha, A. *et al.* Bridging neuronal correlations and dimensionality reduction. *Neuron* **109**, 2740–2754.e12 (2021).
63. Cunningham, J. P. & Yu, B. M. Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* **17**, 1500–1509 (2014).
64. Harvey, C. D., Coen, P. & Tank, D. W. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484**, 62–68 (2012).
65. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
66. Kaufman, M. T., Churchland, M. M., Ryu, S. I. & Shenoy, K. V. Vacillation, indecision and hesitation in moment-by-moment decoding of monkey motor cortex. *Elife* **4**, e04677 (2015).
67. Sadtler, P. T. *et al.* Neural constraints on learning. *Nature* **512**, 423–426 (2014).
68. Vyas, S. *et al.* Neural Population Dynamics Underlying Motor Learning Transfer. *Neuron* **97**, 1177–1186.e3 (2018).
69. Cohen, M. R. & Maunsell, J. H. R. A Neuronal Population Measure of Attention Predicts Behavioral Performance on Individual Trials. *J. Neurosci.* **30**, 15241–15253 (2010).
70. Rabinowitz, N. C., Goris, R. L., Cohen, M. & Simoncelli, E. P. Attention stabilizes the shared gain of V4 populations. *Elife* **4**, (2015).
71. Huang, C. *et al.* Circuit Models of Low-Dimensional Shared Variability in Cortical Networks. *Neuron* **101**, 337–348.e4 (2019).
72. Ruff, D. A., Xue, C., Kramer, L. E., Baqai, F. & Cohen, M. R. Low rank mechanisms underlying flexible visual representations. *Proceedings of the National Academy of Sciences* **117**, 29321–29329 (2020).
73. Feuerriegel, D., Vogels, R. & Kovács, G. Evaluating the evidence for expectation suppression in the visual system. *Neurosci. Biobehav. Rev.* **126**, 368–381 (2021).
74. Maunsell, J. H. R. & Treue, S. Feature-based attention in visual cortex. *Trends Neurosci.* **29**, 317–322 (2006).
75. Liu, T. Feature-based attention: effects and control. *Current Opinion in Psychology* **29**, 187–192 (2019).
76. Myers, N. E., Stokes, M. G. & Nobre, A. C. Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends Cogn. Sci.* **21**, 449–461 (2017).
77. Tang, M. F., Smout, C. A., Arabzadeh, E. & Mattingley, J. B. Prediction error and repetition suppression have distinct effects on neural representations of visual information. *Elife* **7**, 2108 (2018).
78. Feuerriegel, D., Churches, O., Coussens, S. & Keage, H. A. D. Evidence for spatiotemporally distinct effects of image repetition and perceptual expectations as measured by event-related potentials. *Neuroimage* **169**, 94–105 (2018).
79. Chao, Z. C., Takaura, K., Wang, L., Fujii, N. & Dehaene, S. Large-Scale Cortical Networks for Hierarchical Prediction and Prediction Error in the Primate Brain. *Neuron* **100**, 1252–1266.e3 (2018).

80. Solomon, S. S., Tang, H., Sussman, E. & Kohn, A. Limited Evidence for Sensory Prediction Error Responses in Visual Cortex of Macaques and Humans. *Cereb. Cortex* **31**, 3136–3152 (2021).
81. Issar, D., Williamson, R. C., Khanna, S. B. & Smith, M. A. A neural network for online spike classification that improves decoding accuracy. *J. Neurophysiol.* **123**, 1472–1485 (2020).
82. Kelly, R. C. *et al.* Comparison of recordings from microelectrode arrays and single electrodes in the visual cortex. *J. Neurosci.* **27**, 261–264 (2007).
83. Cowley, B., Williamson, R., Clemens, K., Smith, M. & Yu, B. M. Adaptive stimulus selection for optimizing neural population responses. *Adv. Neural Inf. Process. Syst.* 1396–1406 (2017).
84. Williamson, R. C. *et al.* Scaling Properties of Dimensionality Reduction for Neural Populations and Network Models. *PLoS Comput. Biol.* **12**, e1005141 (2016).
85. Churchland, M. M. *et al.* Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* **13**, 369–378 (2010).
86. Bittner, S. R. *et al.* Population activity structure of excitatory and inhibitory neurons. *PLoS One* **12**, e0181773 (2017).
87. Athalye, V. R., Santos, F. J., Carmena, J. M. & Costa, R. M. Evidence for a neural law of effect. *Science* **359**, 1024–1029 (2018).
88. Cowley, B. R. *et al.* Slow Drift of Neural Activity as a Signature of Impulsivity in Macaque Visual and Prefrontal Cortex. *Neuron* **108**, 551–567.e8 (2020).

Supplemental Information

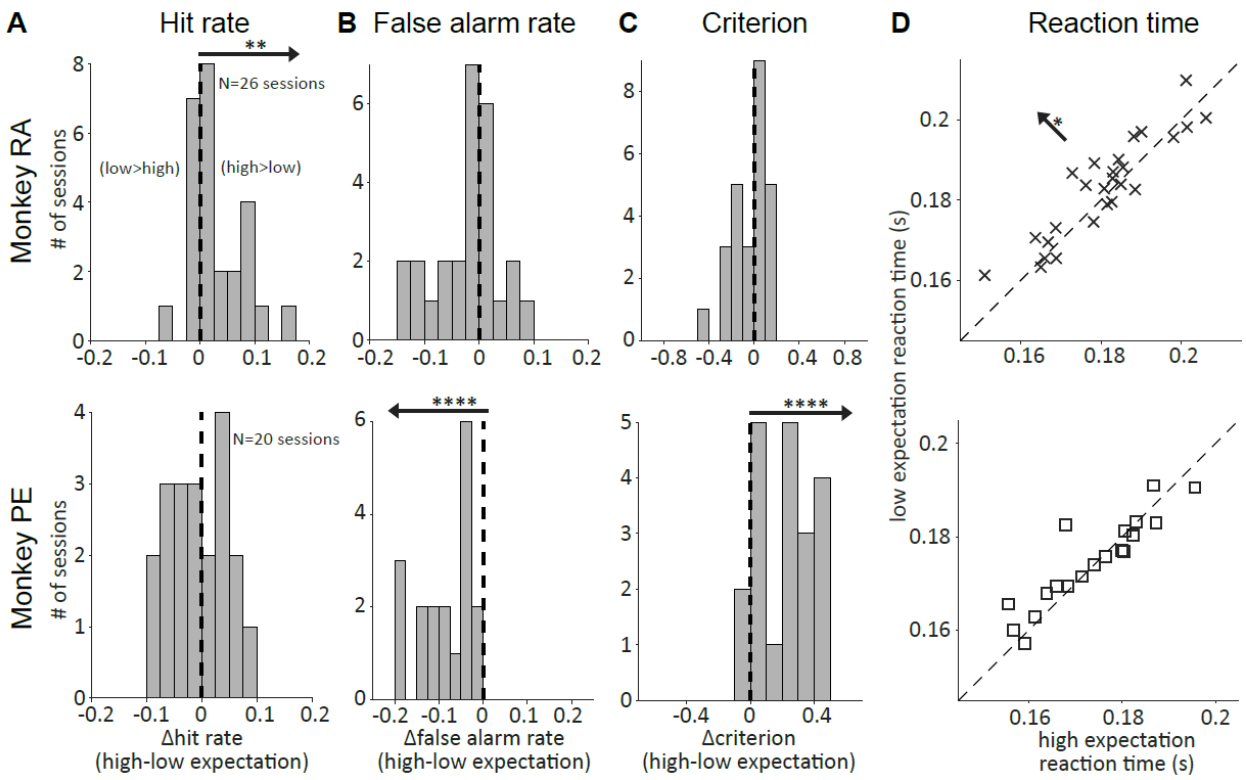

Figure S1. Additional behavioral metrics of expectation

A-B Hit rate and false alarm rate, which are used to calculate *d'*, are shown for all sessions for monkey RA (top) and monkey PE (bottom). A) The difference in hit rates between high and low expectation conditions across sessions was significant for Monkey RA (top, paired t-test $p<0.01$) but not monkey PE. Values to the left of 0 indicate hit rate was higher in the low expectation condition, values to the right of 0 indicate hit rate was higher in the high expectation condition. B) The difference in false alarm rates between high and low expectation conditions across sessions was significant for monkey PE (bottom, paired t-test $p<0.0001$) but not monkey RA. C) Criterion - a measure of bias that indicates how likely the subjects were to look at the sample - was significantly higher in the high expectation condition for monkey PE (bottom, paired t-test $p<0.0001$) but not monkey RA. D) Reaction time was calculated across all correct trials for each session. There was a significant decrease in reaction time in the high expectation condition for monkey RA (paired t-test, $p<0.05$) but not monkey PE.

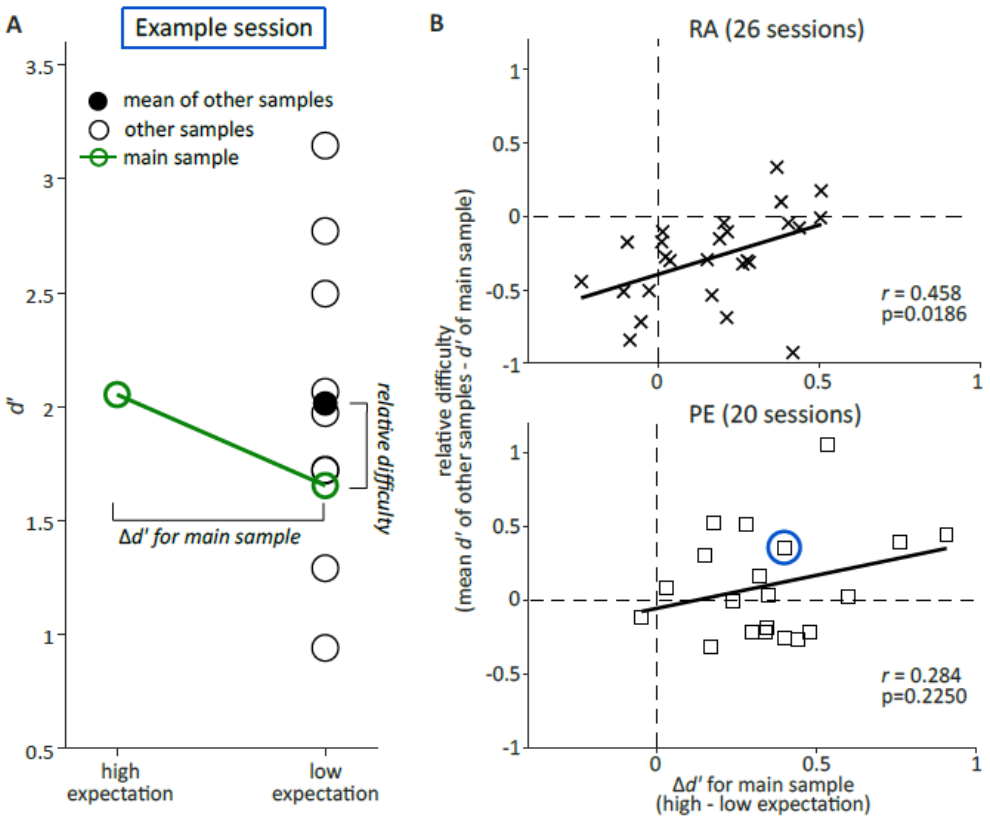

Figure S2. Behavioral effects of expectation cannot be explained by relative task difficulty

A) The behavioral performance to the main sample image in the high and low expectation conditions (green circles), each image in the low expectation condition (black open circles), and the average across images other than the main sample in the low expectation condition (filled circle) are shown for one example session. For each session, the change in d' refers to the difference in performance for the main sample image in the two conditions (the same analysis in Figure 1). The relative difficulty refers to the difference in performance between the main sample image and the average of the other images. In this session, the images other than the main sample tended to have a higher d' than the main sample in the low expectation condition. B) The relative difficulty versus the change in d' for the main sample is shown for each session for monkey RA (top) and monkey PE (bottom). The example session in (A) is highlighted in blue. Negative X axis values indicate performance for the main sample was higher in the low expectation condition, and positive X axis values indicate performance for the main sample image was higher in the high expectation condition. Most values are on the positive side of the X-axis, consistent with an overall behavioral improvement in high expectation. Positive Y axis values indicate the main sample image was more difficult to detect a change in (i.e., lower d') than the other images (such as in the example session in (A)), and negative Y axis values indicate the main sample image was easier to detect a change in than the other images. Each point represents one session. Dashed lines indicate 0 for reference. Best fit lines are in black. In both monkeys there was a positive relationship between the change in d' for the main sample and relative difficulty, indicating that on a session-by-session basis, the easier the other images in the low expectation condition were, the greater the improvement in d' in the high expectation condition. The majority of points for both animals fell on the positive side of the X axis, indicating that an effect of expectation was present regardless of the relative difficulty.