Feedback Determines the Structure of Correlated Variability in Visual Cortex

Adrian G. Bondy[1,2] & Bruce G. Cumming[1]

[1] Laboratory of Sensorimotor Research, National Eye Institute, NIH
49 Convent Drive, Rm. 2A50
Bethesda, MD 20892

[2] Brown-NIH Neuroscience Graduate Partnership Program
185 Meeting Street, Box GL-N
Providence, Rhode Island 02912

Correspondence:
Adrian Bondy
49 Convent Drive, Room 2A50
Bethesda, MD 20892
adrian.bondy@gmail.com
301-451-4926

The variable spiking discharge of sensory neurons in response to a fixed stimulus tends to be weakly correlated (spike-count correlation, r_{sc}). This is widely thought to reflect stochastic noise in shared sensory afferents, in which case it places strict limits on the fidelity of sensory coding. However, it may also be generated by changes over time in feedback from higher-order brain regions. We tested this alternative directly by measuring spiking activity in populations of primary visual cortical (V1) neurons in rhesus monkeys performing different visual discrimination tasks on the same set of visual inputs. We found that the structure of r_{sc} (the way r_{sc} varied with neuronal stimulus preference) changed dramatically with task instruction, despite identical retinal input. This demonstrates that r_{sc} structure primarily reflects feedback dynamics engaged by the task, not noise in sensory afferents. As a consequence, previous analyses of how r_{sc} constrains sensory processing need not apply. Furthermore, these results imply that decision-related activity in sensory neurons is a consequence of task-dependent changes in feedback.

The firing rate of neurons in sensory cortex depends on sensory input, but is also variable given a fixed stimulus. This response variable is weakly correlated between neurons[1]. The origin of these spike-count correlations (r_{sc}) is not well understood. A predominant assumption is that stochastic processes, such as random fluctuations in shared sensory afferent pathways, are the primary source of r_{sc}. Consistent with this proposal, r_{sc} correlates with physical proximity and similarity in stimulus preference[2–6], both of which are predictive of greater shared afferent input.

Because perceptual decisions are thought to be generated by pooling responses of many sensory neurons, correlated "noise" in the sensory pathway could be highly detrimental: while independent variability can be averaged away by pooling enough neurons, correlated variability cannot. Following this logic, r_{sc} is widely thought to constrain the fidelity of sensory information in the brain[2,7–14] and, relatedly, to influence the choices subjects make on individual trials, yielding choice-related activity in sensory neurons[15–17].

These two important consequences of r_{sc} only follow given the widespread view that it arises from the sensory afferents. However, sensory cortical areas receive only a minority of their inputs from the upstream brain regions conveying sensory information from the periphery[18,19]. Consequently, variation over time in shared inputs from downstream areas (i.e. "top-down"; "feedback"), may make a significant contribution to r_{sc}. These signals may reflect endogenous processes like attention or arousal. This source of correlated variability need not confound downstream sensory decoding, since downstream areas may have knowledge of the state of the feedback inputs responsible.

In the present study, we directly investigated the relative contributions of feedforward and feedback pathways to r_{sc} in sensory neurons. We recorded spiking activity in populations of primary visual cortical (V1) neurons in macaque monkeys performing different orientation discrimination tasks using the same set of stimuli. The only difference between the tasks was the pair of orientations being discriminated. If r_{sc} primarily reflects noise in sensory afferents, it should be invariant to changes in the task given fixed retinal input. Alternatively, the pattern of r_{sc} may change dynamically, reflecting top-down input that changes with the task. Population-level recordings allowed us to estimate the full r_{sc} structure of V1 (in our case, how r_{sc} varies as a function of all possible combinations of pairwise orientation preference) under different task contexts. This allowed us to quantitatively compare the contribution of feedforward and feedback inputs. Strikingly, we observed profound and systematic changes with task context, and little fixed r_{sc} structure. Thus, r_{sc} structure in V1 predominantly reflects the dynamics of feedback signaling.

This result has two important implications. First, we show that the observed r_{sc} structure would degrade the task performance of a standard decoder applied to V1. However, our discovery of the feedback origin of these correlations points to the possibility that the brain can, in principle, outperform such a decoder by including knowledge of the changing state of downstream brain areas when decoding V1 activity. Second, we show quantitatively that these feedback dynamics are the primary source of the choice-related activity we observed in V1, clarifying an ongoing debate[20] about the interpretation of choice-related signals in sensory neurons. Taken together, our results suggest that r_{sc} in sensory neurons reveals less than previously thought about the fidelity of sensory information in the brain, but potentially much more about the interareal computations underlying perceptual processing.

Results

We trained two rhesus monkeys (*Macaca mulatta*) to perform a two-alternative forced choice (2AFC) coarse orientation discrimination task (Fig. 1), used previously[21]. On a given trial, the subject was shown a dynamic, 2D filtered noise stimulus for 2 seconds, after which it reported the stimulus orientation by making a saccade to one of two choice targets (oriented Gabor patches). Different task contexts were defined by the orientations of the discriminanda. The stimuli were bandpass filtered in the Fourier domain to include only orientations within a predetermined range. The stimulus filter was centered on one of the two discriminandum orientations and its orientation bandwidth was used to control task difficulty. We included 0%-signal trials, for which the stimuli were unfiltered for orientation (and thus the same regardless of context), to examine the effect of task context on r_{sc} in the presence of a fixed retinal input.

In order to detect any effect of task context on r_{sc} structure, it is critical that subjects based their choices on the presence of the correct orientation signals. To ensure this, we used psychophysical reverse correlation[21–23] to directly measure the influence of different stimulus orientations on the subject's choices (the "psychophysical kernel"). We found that subjects required multiple days of retraining after a change in the task context to fully update their psychophysical kernel. For this reason, we kept the task context fixed for the duration of each recording session, and only undertook recordings in a new task context after subjects had updated their kernel (Supplementary Fig. 1). This is a significant advance over past studies of the effect of task context on neuronal responses, which typically have not quantified the extent to which behavioral strategy truly matches task instruction.

We recorded spiking activity in populations of single V1 neurons using multi-electrode arrays while the subjects performed the task. We determined the preferred orientation of each neuron by measuring its response to oriented stimuli (see Methods) in separate blocks of trials during which subjects passively fixated. Neurons were excluded from analysis if they were not well orientation tuned. The final dataset includes 811 simultaneously recorded pairs from 200 unique cells across 41 recording sessions. For each pair, we calculated its r_{sc} value as the Pearson correlation between the set of trial-duration spike-counts across trials of the same stimulus condition. While measuring r_{sc} only across 0%-signal trials isolated any changes due to the task context, we found similar results within each signal level (Fig. 7). Therefore, to increase statistical power, we report r_{sc} values measured across all trials, after normalizing spike counts to remove the effect of stimulus drive on firing rates (see Methods).

Predicting the form of r_{sc} structure

If task-dependent feedback contributes to r_{sc}, the structure observed will depend on how that feedback affects individual neurons. Consider the simple case in which feedback acts via a single modulatory factor, increasing the activity of neurons associated with one choice, while suppressing those favoring the other choice. We illustrate this selective coupling as a function of neuronal orientation preference for two different task context (cardinal and oblique discrimination) in Fig. 2a,b. Crucially, this feedback would introduce a source of stimulus-independent variability, with correlations determined by the product of the coupling weights for a given pair. By taking the outer product of the coupling weight function, we directly obtain a matrix yielding the r_{sc} predicted for pairs of all possible pairwise orientation preferences. The resulting matrices for the two task contexts (Fig. 2c,d) exhibit a lattice-like pattern that changes its location with the task, yielding high correlation for pairs tuned to the same discriminandum orientation and low correlations for pairs tuned to opposing discriminanda.

We note that observing such a distribution of r_{sc} would, on its own, be consistent with a number of functional interpretations. In particular, qualitatively identical predictions have been generated by considering the effect of fluctuations across trials in the allocation of feature attention[24] or Bayesian priors[25] in a discrimination task, rather than feedback directly related to choice. This is because these other effects can also be expressed in terms of a modulatory input to which neurons are selectively coupled in a task-dependent manner. We return to the issue of functional interpretation later.

Importantly, the predominant view that the structure of r_{sc} in a sensory area is primarily determined by noise in sensory afferents makes a very different prediction for the r_{sc} matrix. This view requires that the pattern of r_{sc} stay fixed across tasks. To be consistent with the inverse relationship observed between r_{sc} and similarity in stimulus preference[2–6], the only possible fixed matrix has a diagonal, banded pattern (Fig. 2e), such that r_{sc} depends simply on the difference in prefered orientation. These "limited-range" correlations are widely assumed to be the way r_{sc} is distributed in sensory populations[7–9,26]. However, note that the prediction based on task-dependent feedback is also consistent with the empirical data: on average, pairs closer to the diagonal have higher r_{sc} values under that prediction, as well (Fig. 2f). Therefore, only by measuring the full matrix across multiple task contexts, as the present study is the first to do, can these two divergent predictions be properly tested.

R_{sc} structure changes systematically with task context

To assess the presence of task-dependent r_{sc} structure in the data, we first divided the recording sessions into two groups based on the task context used (Fig. 3a). Within each subset of sessions, the task context was closely similar. To estimate the r_{sc} matrix for a given subset of sessions, we used the subset of recorded r_{sc} values, along with estimates of their preferred orientations, to populate the matrix. We applied a smoothing kernel to obtain a continuous, smooth estimate. We observed clearly distinct patterns in the matrices derived from the two subsets of sessions. The highest values of r_{sc} tended to occur amongst pairs that both preferred the same discriminandum orientation and the lowest values of r_{sc} tended to occur amongst pairs preferring opposite discriminanda. Because the task context differed between the two subsets, this yielded matrices with a similar lattice-like pattern, in each case offset along the diagonal by an amount reflecting the task context (Fig. 3d,e). In other words, r_{sc} structure changed dramatically with task context, consistent with the presence of task-dependent feedback (Fig. 3b,c) and inconsistent with a fixed r_{sc} structure primarily driven by sensory afferent noise.

Next we generated a single r_{sc} matrix summarizing the task-dependent structure across the entire dataset. To do this, we expressed each neuron's preferred orientation relative to the discriminandum orientations on its respective recording session, such that 0º and 90º always indexed the discriminandum orientations. This combined matrix even more closely resembled the lattice-like pattern predicted by task-dependent feedback (Fig. 4a,b). When we analyzed its eigenspectrum (Fig. 4e), we found that its rank-1 eigenvalue was much greater than chance, demonstrating quantitatively that the task-dependent correlations can be largely explained by a single source of covariability. Because the chance distribution was obtained by randomly translating each individual r_{sc} measurement along the diagonal, this also rules out the possibility that we observed task-dependent structure simply due to noisy sampling of a fixed diagonal, banded pattern ($p<0.005$, permutation test). Furthermore, the rank-1 eigenvector (Fig. 4d) closely resembled a sinusoid with peak and trough at 0º and 90º. This can be interpreted as the pattern of coupling across neuronal preferred orientations to this source of task-dependent covariability. Its shape implies a feedback input to the V1 population that selectively targets the two task-relevant groups of neurons, as described in the initial prediction. These features were also present in the task-aligned r_{sc} matrix when computed separately for each subject (Supplementary Fig. 2).

We observed a different result during separate blocks of trials in the same recording sessions, during which the subject fixated passively for reward but the same set of stimuli was shown. During these blocks, the task-aligned r_{sc} matrix could not be distinguished from a diagonal, banded pattern (Supplementary Fig. 3). This demonstrates that the task-dependent pattern observed during task performance depends on active task engagement, and cannot be explained, for instance, simply as an effect of recent task experience. We performed a number of additional analyses to rule out any possibility that our findings could be explained as an effect of changing retinal input across task contexts (see Supplementary Discussion §1 and Supplementary Figs. 4-7). Taken together, these controls strengthen our interpretation that centrally-generated signals reflecting task engagement underlie the observed correlations, rather than, for instance, slow time scale changes in local V1 circuitry with learning or changes in retinal input with task context.

Segregating fixed and task-dependent components of r_{sc} structure

To quantify the degree to which r_{sc} structure changed with task context, and to determine if there was also a component that remained fixed, we turned to a quantitative model. The model described the r_{sc} structure across sessions using two components: a fixed component (an r_{sc} matrix that did not change with task context), and a task-dependent component (an r_{sc} matrix whose alignment changed systematically with task context). The shape of the two components was fit to the data (i.e. the set of 811 r_{sc} measurements). By construction, if r_{sc} depended only on the raw orientation preferences of neuronal pairs, with no effect of task context, then the model would assign large coefficients to the fixed component and coefficients of zero to the task-dependent component. If r_{sc} was entirely task-dependent, the reverse would be true.

When fitted to the observed r_{sc} measurements (see Methods), the task-dependent component of the model explained most of the explainable variance in the data (82%, Fig. 5a). Not surprisingly, its shape recapitulated the lattice pattern in the task-aligned r_{sc} matrix (Fig. 4b). The fixed component had a markedly smaller amplitude, with a less organized structure (Fig. 5b). Removing the fixed component from the model altogether had little effect, while removing the task-dependent component dramatically impaired model performance (Fig. 5c). Thus, we failed to reliably identify a fixed source of r_{sc} structure, such as the limited-range correlations postulated previously[7–9,26], during task performance. Instead, the predominant feature was task-dependent changes in r_{sc} structure. (We were also able to reproduce these model results individually for one subject).

Effect of task-dependent r_{sc} structure on neural coding

R_{sc} in sensory neurons is typically studied as a source of noise that impacts the ability of a downstream brain area to decode a sensory input[2,7–14]. Our results show that the predominant source of r_{sc} structure in V1 is top-down in origin. It is difficult to say whether this reflects an additional source of uncertainty in the sensory representation or not. One possibility is that the spikes in V1 introduced by feedback can be adaptively discounted by the decoder, removing any such uncertainty. Current approaches for understanding the impact of r_{sc} on neural coding have not considered this possibility. Therefore, to gain quantitative insight into the potential impact of the observed r_{sc} structure on sensory information coding, we made the assumption of a standard, linear decoder (which is blind to the presence of feedback) applied to V1. In this case, it has been shown[27] that a particular r_{sc} structure places a strict upper bound on decoding accuracy (although assessing the quantitative impact of spike-count correlations on information coding is an area of active investigation[28]). These so-called "differential" correlations are those that mimic the correlations produced by changes in the stimulus along the axis defining the task. For the task we used, the differential correlation for a neuronal pair is given by the product of the slopes of their mean responses as a function of signal strength (Fig. 6a) – a metric of similarity in tuning for the task. When we plotted these values as a smooth, task-aligned matrix (Fig. 6b), we observed a lattice-like pattern strikingly similar to the observed r_{sc} matrix (Fig. 4b). Confirming this similarity, the task-dependent component of r_{sc} structure identified by the regression model was highly correlated on a pair-by-pair basis with the differential correlations (r=0.62, Fig. 6c). In other words, the structure of stimulus-independent covariability in the V1 population (introduced by feedback) was closely similar to the structure of covariability introduced by stimulus variation. Perhaps this is not surprising, as the r_{sc} structure was consistent with feedback that alternatingly targeted the task-relevant neuronal pools, similar to the effect of varying the stimulus along the axis defining the task. However, the implication is profound: task-dependent feedback appears to degrade, rather than improve, the sensory representation in V1.

Of course, this is only true if the feedback cannot be taken into account by the decoder – a decoder that had access to the activity of neurons providing the feeback would not be limited in this way. As an illustration, we consider a simple algorithm for decoding V1 activity in the presence of task-dependent feedback that fully eliminates any deleterious effects of the resulting correlations. Because feedback introduced "differential" correlations, this mean it moves the V1 population along the dimension orthogonal to the optimal decision boundary for the task. This implies that, in principle, subjects could simply adjust their decision criterion trial-to-trial to eliminate any influence on choice. Doing so would only require computing the appropriate criterion offset given knowledge of the state of the feedback input to the sensory population. Whether or not this sort of adaptive decoding is used, this example illustrates the difficulty of assessing the impact of r_{sc} in a sensory population on perceptual performance in light of our results.

Relationship between r_{sc} structure and perceptual choice

A key motivation for our investigation was the frequent observation throughout sensory cortex of choice-related activity (correlations between trial-to-trial variability of single neurons and choice)[29,30] as this has been proposed to reflect the effect of trial-to-trial variability in feedback related to the upcoming choice[25,31,32]. We also observed significant choice-related activity in our recorded neurons. For each neuron, we calculated its Choice Probability (CP), a metric which quantifies the probability with which an ideal observer could correctly predict the subject's choices from that neuron's spike count on each trial[29,30]. Across the population, we found an average CP of 0.54 for task-relevant neurons, significantly above chance level (Fig. 8a) and similar in magnitude to another study using the same task[21].

Theoretical studies have emphasized that widespread CP across a large population of sensory neurons depends on the presence of spike-count correlation[15–17,30,33]. After all, if many sensory neurons have variability that is correlated with choice, this implies the variability of individual neurons is also correlated. However, this could be compatible with either or both of two causal interpretations: 1) correlated fluctuations directly affect the choices a subject makes trial to trial (a feedforward source of CP); or 2) the correlated fluctuations reflect variation across trials in a feedback signal related to the upcoming choice (a feedback source).

A feedback source of CP makes the unique prediction of r_{sc} structure in V1 that is weaker across trials in which there is less variability in choice. Consistent with this prediction, we found that the amplitude of the r_{sc} structure was attenuated on high-signal trials relative to 0% signal trials, in a manner which depended systematically on signal strength (Fig. 7). However, this attenuation was modest, even at the highest signal level we analyzed (11% reduction) and despite the highly uneven distribution of choices. These data suggest that the feedback generating the r_{sc} structure is correlated with choice, but also rule out a post-decisional mechanism in which the state of feedback is completely determined by the final report. We also found that the r_{sc} structure, when calculated using only spikes from different 200-ms windows during the trial, showed a stable timecourse and did not grow in amplitude with decision formation (Supplementary Fig. 8). Taken together, these observations support the conclusion that the r_{sc} structure reflects variation in feedback signals only partially correlated with the subject's final choices. These could include bias, attention to orientation, reward history, intertrial dependencies, and/or a decision variable.

Fig. 7 demonstrates at least the presence of a feedback source of CP. Next, we considered the possibility that some of the observed CP is due to a feedforward effect on choice of correlated fluctuations in V1. Since we were unable to reliably identify any fixed source of r_{sc} structure, this would require that CP be generated by the feedforward effect on choice of r_{sc} structure introduced by feedback. Such a self-reinforcing loop is plausible given that feedback introduced "differential" correlations, mimicking the effect of varying the stimulus along the axis of the task and therefore necessarily influencing choices[34] unless it can somehow be discounted. To probe this possibility quantitatively, we made use of the known analytical relationship between spike-count correlations, readout weights, and CPs, under the assumption of a linear decoder applied to a population of sensory neurons[15] (where CP solely reflects the feedforward influence of neuronal variability). We found that the r_{sc} structure we observed would be sufficient to produce a pattern of CP across the population consistent with the data, across a wide range of possible readout schemes. This result could not be replicated when we considered only the weak "fixed" component of r_{sc} structure identified in the model (Fig. 8b,c). Thus, our data rule out the view that CP reflects the feedforward effect of stochastic noise in the afferent sensory pathway on perceptual choice. Instead, any feedforward source of CP appears to depend on task-dependent changes in the structure of interneuronal correlations that subsequently influence perceptual judgments.

Discussion

The predominant paradigm describes r_{sc} in populations of sensory neurons as a feature of the noise corrupting feedforward sensory pathways[2,7–14]. The effect of top-down input has been studied with the view that it may modify r_{sc} in a way that adaptively reduces the noise[11,13,35]. The present study requires a different view. We found that the pattern of correlated variability in V1 is almost entirely determined by the task context, independent of any changes in retinal input. This demonstrates how r_{sc} can be, in the first place, driven by variability in top-down inputs whose dynamics change depending on the context, quite different from purely stochastic noise. Furthermore, the changes we saw in the sensory population do not appear to be beneficial to task performance (Fig. 6), at least not in the manner this has typically been examined (i.e. assuming a linear decoder of the sensory population alone). These results reveal the need for a new functional interpretation of stimulus-independent covariability in sensory populations. They also highlight the importance of considering the interconnectedness of cortical areas as a crucial aspect of sensory information processing.

These data complement recent reports about the influence of spatial attention on r_{sc} in visual cortex[11,13,35]. In those studies, changes in r_{sc} with attention appeared to adaptively improve the accuracy with which a model observer of visual cortex could perform the task, reinforcing the current thinking about r_{sc} as a problematic source of noise. However, a recent reanalysis[36] of neuronal data from a classic spatial attention study[11] shows how these data may require a different interpretation. The reanalysis revealed that the reduction in r_{sc} under spatial attention could be accounted for as an effect of attenuation of the ongoing variability of a small number of shared gain-modulating inputs, presumably feedback in origin. This demonstrates that the r_{sc} attenuated under spatial attention is likely caused, in the first place, by variation in feedback signals, which are not necessarily noise.

An important consequence of this emerging body of data is that the functional implications of correlated variability for information transmission depend greatly on how downstream brain regions treat spikes in an upstream population introduced by feedback, a complex question that has received little consideration. Earlier, we showed how knowledge of feedback inputs to a sensory population could, in principle, be exploited by a decoder to improve its performance. This example illustrates how past studies that have simply applied a decoder to a population of neurons without regard to activity elsewhere in the brain may have led to misleading conclusions about the functional implications of correlated variability.

On a more constructive note, our results also show how r_{sc} can reveal important aspects of the interareal computations underlying perceptual processing. In particular, our results significantly further our understanding of choice-related signals in the brain. Initial findings of choice-related activity were viewed as evidence for a feedforward effect of neuronal variability on choice[37]. Subsequent work has emphasized the relationship between choice-related activity and correlated fluctuations amongst sensory neurons[15–17,30,33], as well as the possibility of choice-related feedback[25,31,32]. Our work shows quantitatively that choice-related activity is a consequence of task-dependent changes in rsc structure introduced by feedback. We found that correlated fluctuations in V1 are more pronounced on trials where the subject's choices were more variable (Fig. 7), indicating they reflect information related to (although not completely determined by) the upcoming choice. Importantly, correlated fluctuations may subsequently act as an input to the decision through feedforward pathways. If they do, this would imply choice-related activity comes about through self-reinforcing loops of reciprocal connectivity between cortical areas, an intriguing possibility that has also been suggested by other studies[25,32,38].

We currently lack an established, normative account for our central finding: the introduction of r_{sc} structure in a sensory area by feedback. However, one emerging body of theoretical work[25,39,40], influenced by the longstanding idea of perception as probabilistic inference[41,42], does offer an explanation. This work starts with the premise that the goal of a perceptual system is to generate valid inferences about the structure of the outside world. Given the highly impoverished nature of sensory input, the task is impossible without bringing substantial prior knowledge to bear. The novel proposal is that this computation takes place directly in sensory cortical neurons, such that their activity reflects a posterior belief about the presence of a particular sensory feature, requiring both sensory input and prior beliefs conveyed by feedback. Fluctuations in prior beliefs therefore introduce correlated variability. As a consequence, the theory provides a powerful, normative framework for predicting the form this covariability will take in the laboratory, where experimeters can directly manipulate prior beliefs with a psychophysical task. With few free parameters, this framework succeeds in predicting the novel results reported here in the context of an orientation discrimination task[25,39]. While more empirical tests are needed, this novel theory does provide a principled account of data like ours, which are otherwise puzzling.

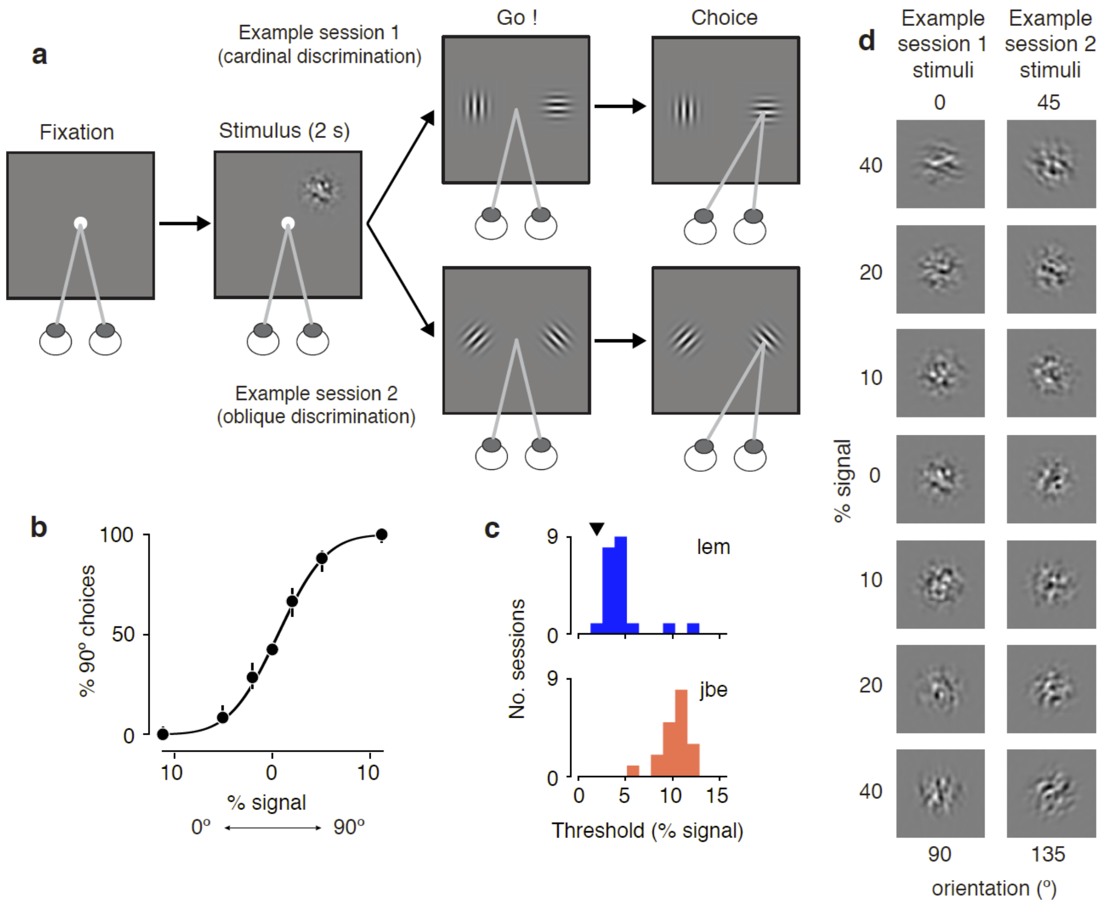

Figure 1. Orthogonal orientation discrimination task. **a**. Schematic illustration of the task. After the subject acquired fixation, a dynamic, filtered noise stimulus appeared for a fixed duration of 2 s. Then the subject had to saccade to the one of two orthogonal choice targets (Gabor patches) whose orientation matched the stimulus. Two example task contexts shown (cardinal and oblique discriminations). The task context was fixed in a given recording session, but varied across sessions. **b**. Psychometric function for monkey *'lem'*, example session. Black curve is a probit fit, and error bars are 95% confidence intervals. **c**. Histograms showing the distribution of psychometric thresholds across sessions for the two subjects. Threshold is defined as the signal level eliciting 75% correct performance. Black triangle indicates the threshold corresponding to the example session in (b). **d**. Example single stimulus frames corresponding to the two example task contexts in (a). The stimuli consisted of dynamic, white noise filtered in the Fourier domain for orientation (see Methods). The filter was centered on one of the two discriminandum orientations and its bandwidth determined signal strength. A given trial consisted of many frames of

independent noise with a fixed filter. 0% signal stimuli were unfiltered for orientation and were statistically identical across task contexts.

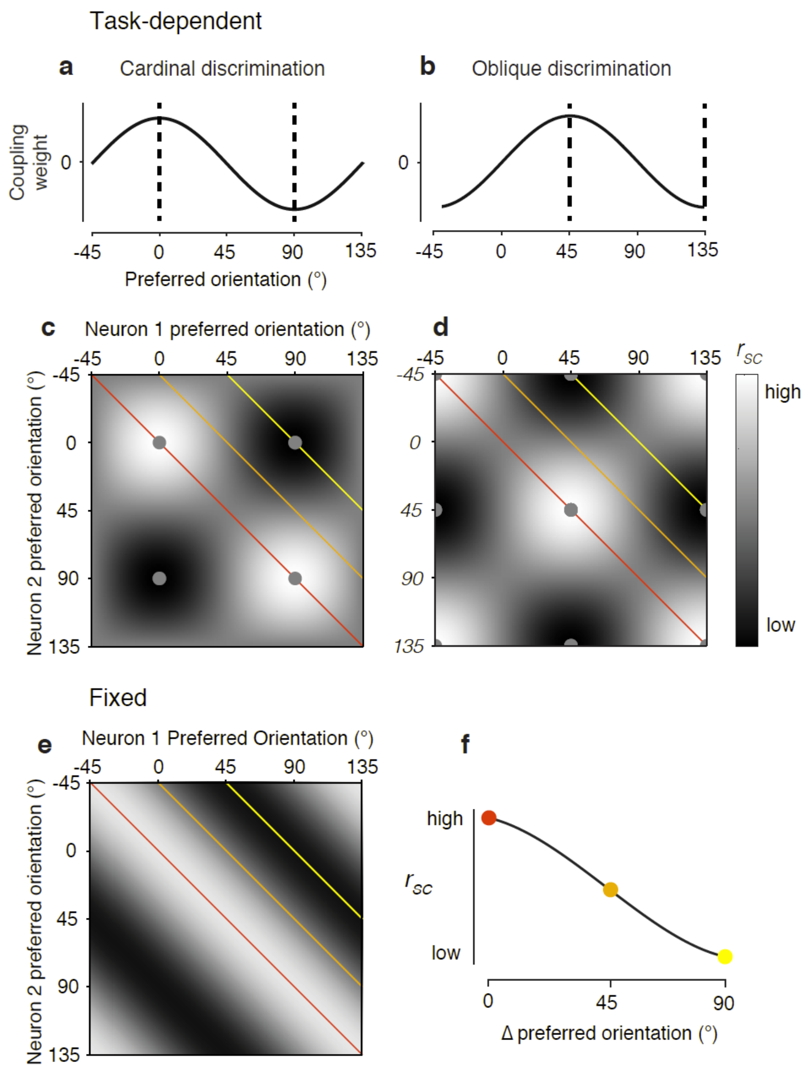

Figure 2. Predictions for task-dependent and fixed sources of r_{sc} structure. Schematic illustration of potential sources of r_{sc} structure during performance of the orientation discrimination task. **a-b.** The effect on r_{sc} structure of feedback selectively (and alternatingly) targeting the two subpopulations of task-relevant neurons. This was assumed to act via a single modulatory factor, to which neurons are coupled with a strength and sign that varies with orientation preference. Coupling weight functions for two tasks context (cardinal and oblique discrimination) shown. The effect on V1 firing rates on a given trial is simply a scalar multiple of these functions. **c-d**. Task-dependent feedback introduces variability in neuronal activity whose correlations depend on pairwise orientation preference and which change systematically with task context. We illustrate this using r_{sc} matrices indexed by neuronal prefered orientation. Ignoring any differences in spike-count variance across orientations, the pattern in the matrices is simply the outer product of the coupling weight functions in (a) and (b). (In other words, (a) and (b) are their sole eigenvectors). In both cases, the result is a lattice-like pattern in the r_{sc} matrix, offset along the diagonal by an amount reflecting the task context. Colored lines indicate constant differences in neuronal preferred orientation (0º, red; 45º, orange; 90º, yellow). Gray dots indicate regions of high and low correlation corresponding to pairs preferring the same or opposite discriminandum orientations. Note that exact values of r_{sc} will also depend on other sources of

covariability so are not well constrained by this prediction. **e.** Alternatively, r_{sc} structure could reflect a fixed source of stochastic afferent noise. To be consistent with the observed inverse relationship between r_{sc} and similarity in stimulus preference[2–6], this would imply a diagonal, banded pattern such that r_{sc} depends only on Δ preferred orientation. Colored lines are as in (c-d). **f**. All predicted r_{sc} matrices (c-e) contain an identical relationship between r_{sc} and Δ preferred orientation. Thus, they cannot be readily distinguished using existing experimental observations, but require measuring the full r_{sc} matrix under different task contexts.

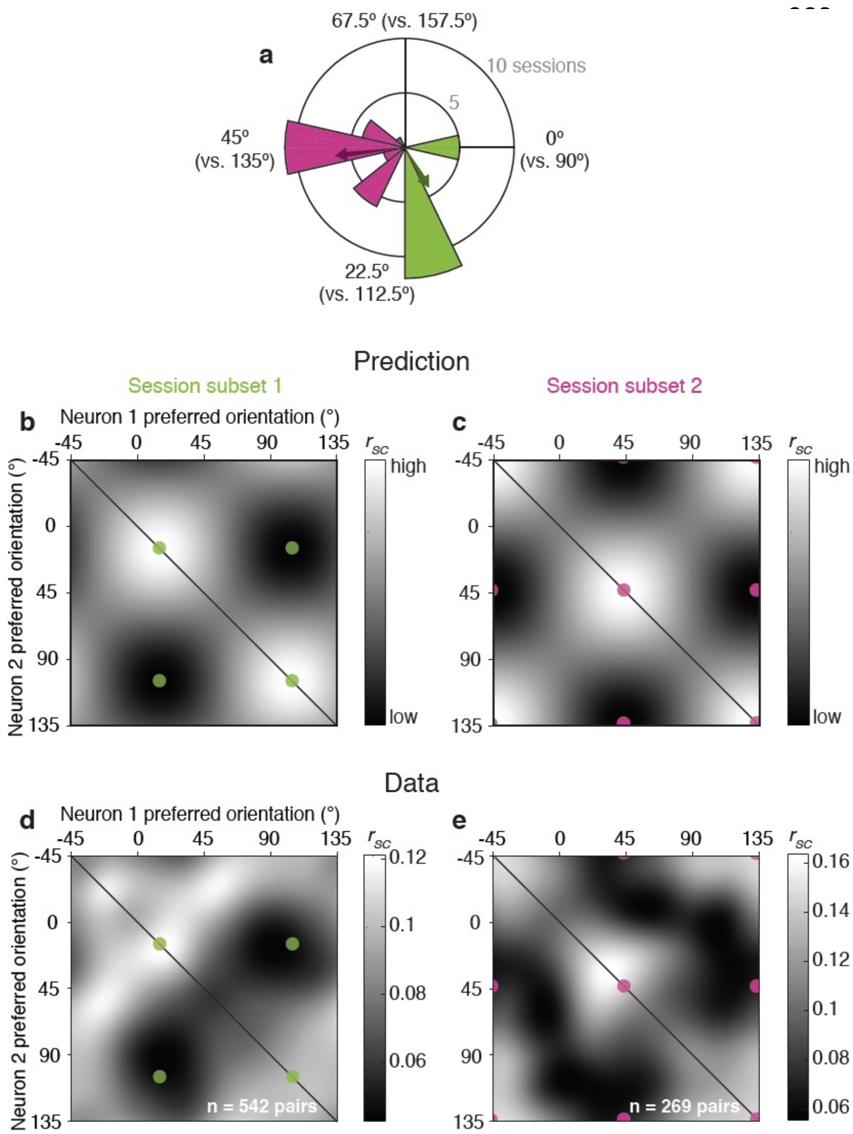

Figure 3. R_{sc} structure in V1 depends systematically on task context. a. We divided the set of recording sessions into two groups based on the task context used. Polar histogram shows the distribution of task contexts used across sessions, with color indicating the division into two subsets. Note that the period is 90° because of the orthogonality of the discriminanda. Colored arrows indicate the mean task context associated with each subset. **b-c**. The presence of task-dependent feedback would predict distinct r_{sc} matrices associated with the two subsets of sessions. The locations in the matrix where peaks and troughs in r_{sc} are predicted are highlighted with colored circles. These correspond to the mean discriminandum orientations indicated with arrows in (a). **d-e**. Observed r_{sc} matrices for the two subsets of sessions. These are obtained by populating the matrices with the set of r_{sc} measurements made within each subset, at locations determined by the orientation preferences of the pairs. We applied a von Mises smoothing kernel (approximating a 2D wrapped Gaussian with 15° s.d.). The observed pattern is distinct across the two matrices, closely matching the predictions in (b-c). Note that the smoothed matrices contain only positive values, even amongst pairs that would be decorrelated by task-dependent feedback, suggesting unknown sources of global correlation.

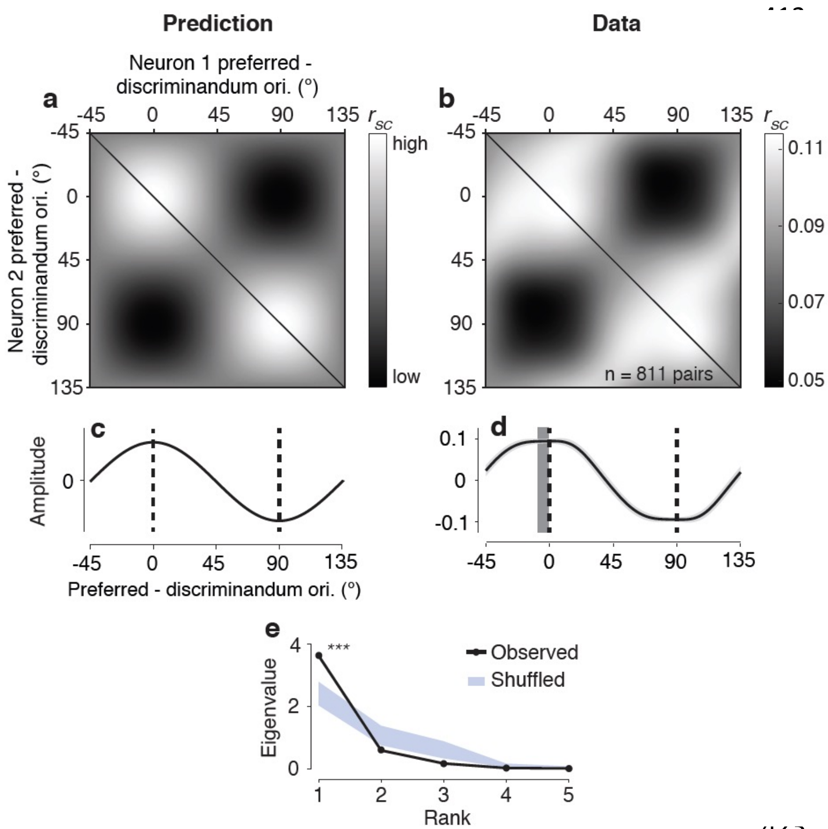

Figure 4. Task-aligned r_{sc} matrix. a. Predicted r_{sc} matrix based on the hypothesis of task-dependent feedback, shown in a task-aligned coordinate frame. Each pair's preferred orientations are expressed relative to the discriminandum orientations, so this prediction is the same across all tasks. 0º and 90º therefore index the discriminandum orientations. **b.** Observed r_{sc} matrix in the task-aligned coordinate frame, combining data recorded across all sessions and smoothed identically to the data in Fig. 3c,d. Note the striking similarity to the prediction in (a). **c,d.** The rank-1 eigenvectors of the predicted and observed r_{sc} matrices in (a) and (b) are closely similar. (We first removed the mean r_{sc} value from the matrices to ignore any flat eigenvectors). The eigenvector corresponding to the observed data is shown with an error bar of 1 s.e. The dark gray vertical bar indicates the peak in the eigenvector +/- 1 bootstrap s.e. This was not significantly different from 0º, indicating close alignment between the dynamic pattern of r_{sc} in V1 and the subject's task. **e**. Eigenspectrum for the observed matrix in (b). Most of the variance in the matrix was explained by its rank-1 eigenvector, significantly more than would be predicted by chance ($p<0.001$, permutation test). Chance level was determined by adding a random offset to the preferred orientations of each pair (i.e. permuting each included r_{sc} value along the diagonal).

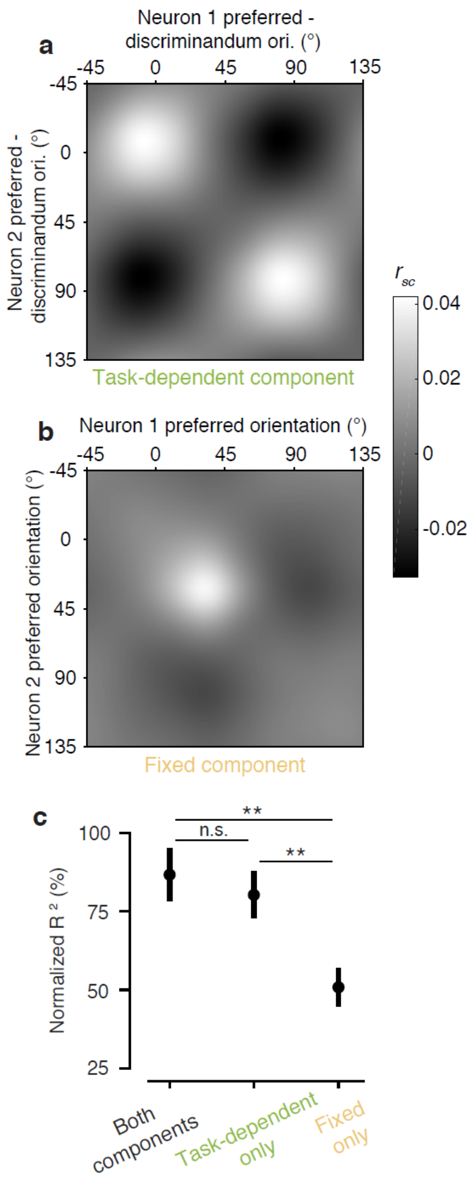

Figure 5. Segregating fixed and task-dependent components of r_{sc} structure. a-b. We used a quantitative model that described the r_{sc} structure across sessions using two components: a fixed component (an r_{sc} matrix for orientation that did not change with task context), and a task-dependent component (one whose alignment changed systematically with the task). The shape of the two components was fit to the data (i.e. the set of 811 r_{sc} measurements; see Methods). We found the amplitude of the task-dependent component (a) was considerably larger than the fixed component (b), showing the majority of r_{sc} structure changes with the subject's task. Not surprisingly, the form of the task-dependent component closely resembled the form of the task-aligned r_{sc} matrix (Fig. 4b). The fixed component had a less organized structure. Note that preferences for the task-dependent component, but not the fixed component, are expressed relative to the discriminandum orientations. Mean r_{sc} values are close to 0 due to the inclusion of a model constant. **c**. Goodness-of-fit for the joint model and two reduced models that included only one of the two components. Values are expressed relative to an estimate of the explainable variance in the data (see Methods). Removing the task-dependent component (but not the fixed component) significantly reduced goodness-of-fit. Error bars are +/- 1 s.e. obtained from repeated 10-fold cross-validation, and ** indicates goodness-of-fit values that are significantly different at the $p<0.01$ level.

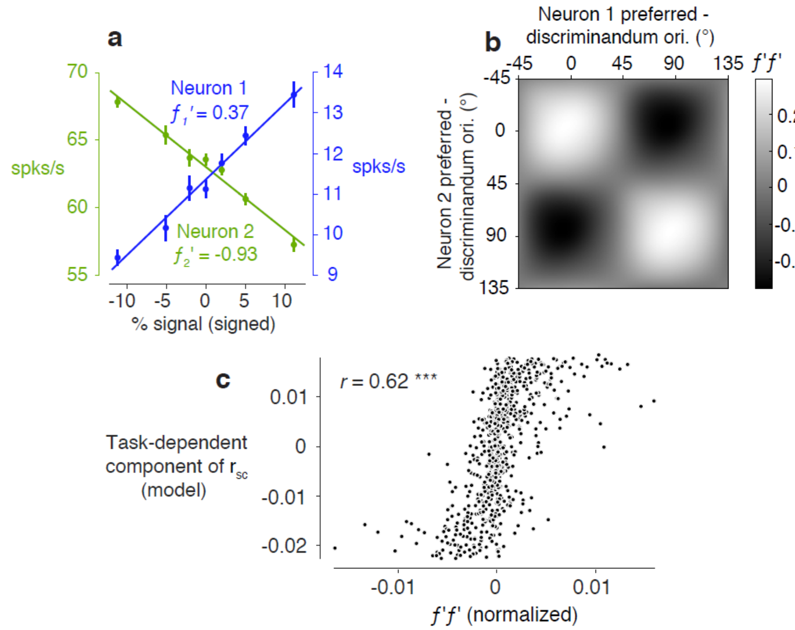

Figure 6. Task-dependent feedback introduces differential correlations. a. Responses (mean +/- 1 s.e.) to the stimuli used in the task at various signal strengths for two example neurons. For the purposes of illustration, the two discriminanda orientations are simply labeled positive and negative. Calling these response functions f_1 and f_2, the differential correlation for this pair is proportional to the product of the derivatives $f_1'f_2'$ [27]. This product can be viewed as a metric of similarity in tuning for the task. Therefore, differential correlations are those that resemble the effect of changes in the stimulus along the axis defining the task. **b**. The matrix of $f'f'$ values, as a function of task-aligned pairwise orientation preference, obtained using kernel smoothing as in Fig. 4b. We observed a lattice-like pattern that was extremely similar to the structure of task-dependent r_{sc} we observed during task performance (Fig. 4b), suggesting task-dependent feedback introduces a source of differential correlation to the V1 population. **c**. Scatter plot of the task-dependent (putatively top-down) component of r_{sc} (Fig. 5a) against $f'f'$ values (normalized; see Methods) for each recorded neuronal pair. The two were highly correlated across the population ($r=0.62$, $p<10^{-5}$).

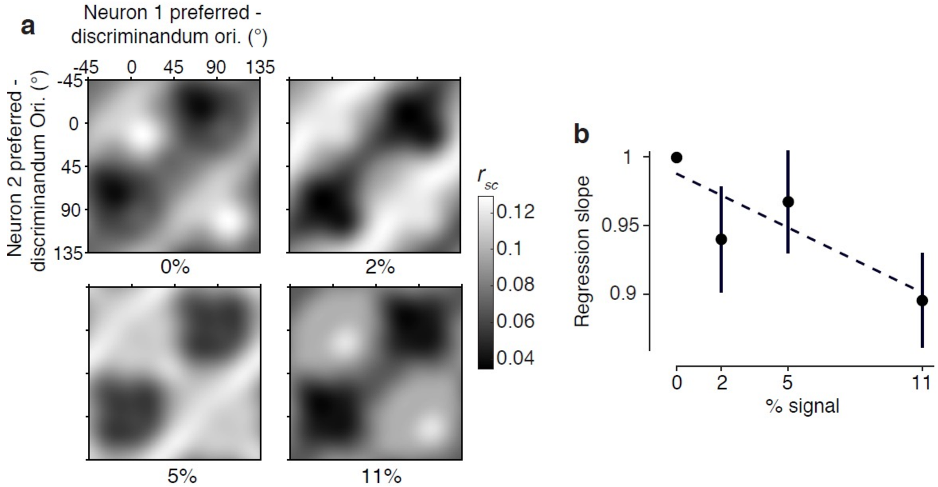

Figure 7. R_{sc} structure depends on variability in choice. a. The average, task-aligned r_{sc} matrix, shown separately for each stimulus strength. The matrix associated with the 0% signal trials reflects dynamic changes in r_{sc} structure across task contexts despite identical retinal input. A qualitatively similar structure was apparent at non-zero signal levels, even though the retinal input was not strictly identical across sessions (although spike counts were z-scored to eliminate the effect of stimulus drive). **b.** We quantified the slope of a regression line comparing the r_{sc} values measured at each signal level against the r_{sc} values measured at the 0% signal level. The magnitude of the slope indicates the degree of attenuation of the r_{sc} structure at a given signal level. We observed a weak but significant negative correlation ($p<0.05$, bootstrap test) between this slope value and signal strength (error bars are +/- 1 bootstrap s.e.), suggesting the feedback variability generating the r_{sc} structure is attenuated on high-signal trials, when there was also less variability in choice.

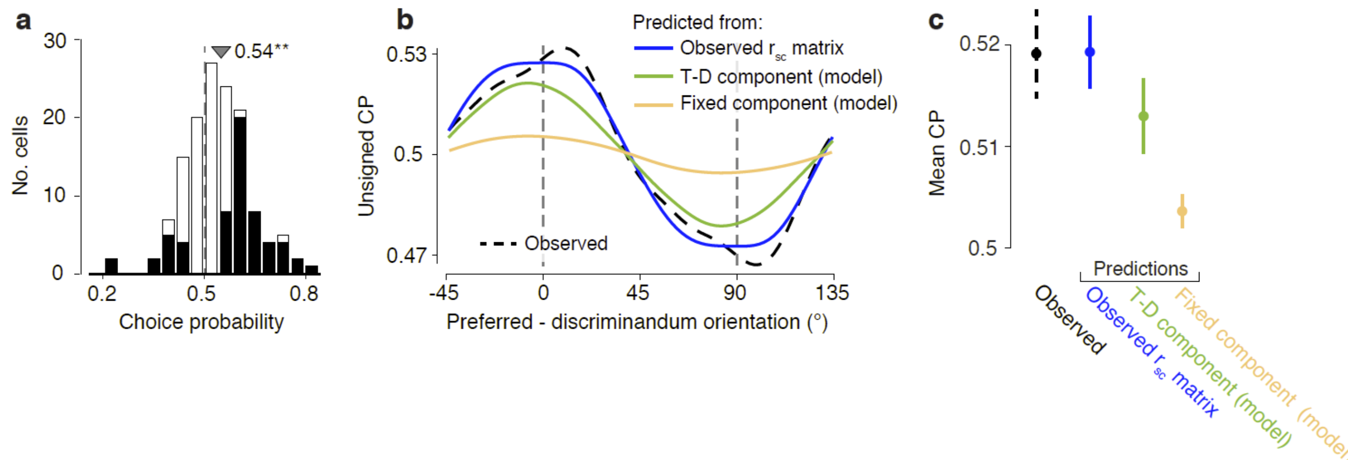

Figure 8. The task-dependent component of r_{sc} structure accounts for choice-related activity. a. Histogram of observed CPs, from the subset of neurons (n=138) significantly preferring one of the two discriminandum orientations (d'>0.9 at highest signal level). The mean of 0.54 was significantly above chance (bootstrap test, cell resampling, $p<0.01$). CPs that were individually significant ($p<0.05$; bootstrap test, trial resampling) are shown in black. **b**. To quantitatively assess the possibility of a feedforward source of CP, we made use of the known analytical relationship between spike-count correlations, readout weights, and CPs, under the assumption of a linear decoder applied to a population of sensory neurons[15] (see Methods). This allowed us to determine whether this framework could account for the observed CP simply as a consequence of reading out a population of neurons with the observed r_{sc} structure, without any reference to feedback. To perform this analysis, we defined CP as a continuous function of task-aligned preferred orientation, analogous to our description of the r_{sc} matrix in Fig. 4b. The dashed black line shows the profile of CP observed across preferred orientations, after smoothing with a wrapped Gaussian with 10° s.d. We applied a fixed sign convention to the CP values across all neurons, equivalent to arbitrarily calling the 0°-choice the preferred one. The predicted CP profiles (solid lines) show the CP elicited by reading out a sensory population with different r_{sc} structures. The readout weights across orientations were unobserved and had to be assumed. The profiles shown are averages of a large set generated from different assumed readout weight profiles. Strikingly, the results were highly insensitive to the readout weights (see Supplementary Fig. 10). The prediction using the observed r_{sc} matrix (Fig. 4b) closely match the observed

CP profile. This could be nearly replicated using only the task-dependent component (Fig. 5a) of r_{sc} structure we identified. However, using only the fixed component (Fig. 5b) produced a much smaller magnitude of CP than observed. **c**. Mean CP (using the traditional sign convention) associated with the profiles in (b), +/- 1 bootstrap s.e. obtained by resampling from the data. Note that the mean observed CP is lower here than in (a) because all neurons are included, regardless of their orientation preference.

Methods

Electrophysiology

We recorded extracellular spiking activity from populations of V1 neurons in two awake, head-fixed rhesus monkeys (*Macaca mulatta*). Both monkeys were implanted with a head post and scleral search coils under general anaesthesia[43]. In monkey '*lem*', a recording chamber was implanted over a craniotomy above the right occipital operculum, as described previously[44], by which we introduced linear microelectrode arrays (U- and V-probes, Plexon; 24-contacts, 50 or 60 µm spacing) at an angle approximately perpendicular to the cortical surface with a custom micro-drive. We positioned the linear arrays so that we roughly spanned the cortical sheet, as confirmed with current-source density analysis, and removed them after each recording session. In monkey '*jbe*', a planar "Utah" array (Blackrock Microsystems; 96 electrodes 1mm in length inserted to target supragranular layers, 400 um spacing) was chronically implanted, also over the right occipital operculum. All procedures were performed in accordance with the U.S. Public Health Service Policy on the humane care and use of laboratory animals and all protocols were approved by the National Eye Institute Animal Care and Use Committee.

Broadband signals were digitized at 30 or 40 kHz and stored to disk. Spike sorting was performed offline using custom software in MATLAB®. First, spikes were detected using a voltage threshold applied to high-pass filtered signals. Next, triggered waveforms were projected into spaces defined either by principal components or similarity to a template. Clusters boundaries were finally estimated with a Gaussian mixture model, and then rigorously verified and adjusted by hand when needed. In the linear array recordings, spike sorting yield and quality was substantially improved by treating sets of three or four neighboring contacts as "n-trodes". As this was not possible with the Utah array due to the greater interelectrode spacing, we excluded pairs of neurons recorded on the same electrode to avoid contamination by misclassification. Neurons from separate recording sessions were treated as independent. To reduce the possibility that a single neuron from the Utah array contributed to two datasets, we included only sessions that were separated by at least 48 hours (with a median separation of

5 days). We excluded from analysis those neurons whose mean evoked firing rate did not exceed 7 spikes/second.

Visual stimuli

All stimuli were presented binocularly on two gamma-corrected cathode ray tube (CRT) monitors viewed through a mirror haploscope, at 85 or 100Hz. The monitors subtended 24.1° x 19.3° of visual angle (1280 x 1024 pixels). The stimuli presented during performance of the discrimination task consisted of bandpass filtered dynamic white noise, as described previously[21]. Briefly, stimuli were filtered in the Fourier domain with a polar-separable Gaussian. The peak spatial frequency was optimized for the recorded neuronal population (1 and 4 cpd medians for *'lem'* and '*jbe*', respectively) while the peak orientation could take one of two orthogonal values the animal had to discriminate in a given session. The angular s.d. of the filter modulated the orientation bandwidth and was varied trial to trial. A 2D Gaussian contrast envelope was applied to the stimulus so that its spatial extent was as small as possible while still covering the minimum response fields of the neuronal populations being recorded. The median envelope s.d. was 0.6 degrees for both animals. The median stimulus eccentricity was 5.4 degrees for *'lem'* and 0.5 degrees for '*jbe*'. In Fig. 1, we quantify orientation bandwidth as % signal strength. This was calculated as $100 * R$, where R is the length of the resultant vector associated with the angular component of the stimulus filter.

We estimated neuronal orientation preferences in separate blocks of trials, using 420-ms presentations of the following types of stimuli, presented at a range of orientations: 1) an orientation narrowband version of the stimulus described above (10° angular s.d.); 2) sinusoidal gratings; and 3) circular patches of dynamic 1D noise patterns (random lines). The preferred orientation of a neuron was calculated as the circular mean of its orientation tuning curve. For each neuron, from among the set of tuning curves elicited by the different stimulus types described above, we chose as the final estimate of preferred orientation the one with the smallest standard error, obtained by resampling trials. We excluded from further analysis all neurons where this exceeded 5°. On a subset of sessions, we also used these orientation-tuning blocks to present examples of the 0%-signal orientation-filtered noise stimuli. These were presented at the same location and size as during task performance, allowing us to calculate r_{sc} structure in the absence of task engagement but with identical retinal input.

Orthogonal orientation discrimination task

The animals performed a coarse orientation discrimination task using the orientation-filtered noise stimuli, as described previously[21]. To initiate a trial, the subject had to acquire a central fixation square. After a delay of 50 ms, the stimulus appeared for a fixed duration of 2 seconds. The trial was aborted if the subject broke fixation at any point during the stimulus presentation, defined as either 1) making a microsaccade covering a distance greater than a predefined threshold (typically 0.5°) or 2) a deviation in mean eye position from the center of the fixation point of more than a predefined threshold, typically 0.7°. At the end of the stimulus presentation, two choice targets appeared. These were Gabor patches of 2-3° in spatial extent, oriented at each of the two discriminandum orientations. The locations of the choice targets depended on the task. For discriminandum pairs near horizontal and vertical (-22.5° – +22.5° and 67.5° – 112.5°), the choice targets were positioned along the vertical meridian, at an eccentricity of about 3°, with the more vertically-oriented target appearing always in the upper hemifield. For orientation pairs near the obliques (22.5° – 67.5° and 112.5° – 157.5°), the choice targets were positioned along the horizontal meridian, at the same range of eccentricities, with the smaller of the two orientations always appearing in the left hemifield. (We use the convention that horizontal is 0° and that orientation increases with clockwise rotation.) To penalize random guessing, the volume of liquid reward delivered after correct choices was doubled with each consecutive correct choice, up to a maximum of four times the initial amount. Since we were primarily interested in the effect of task engagement on neuronal activity, we applied a behavioral criterion to our data, excluding sessions where the subject's psychophysical threshold (defined as the signal level eliciting 75% correct performance) exceeded 14% signal. A two-pass presentation procedure was used. Each instance of a stimulus (generated with a given noise seed) was shown twice per experimental block. This allowed us to account for any effect of fluctuations in the stimulus on r_{sc} (see Supplementary Discussion §1.1 and Supplementary Fig. 5).

Spike-count correlation measurements

Spike-count correlations were calculated as the Pearson correlation between spike counts, counted over the entire duration of the stimulus, with a 50-ms delay to account for the typical V1 response latency. Spike counts were first z-scored separately within each experimental block (typically a set of 100-200 trials lasting about 10 minutes) and each stimulus condition. This removed correlations related to long-term firing rate nonstationarities and allowed us to combine trials at different signal levels without introducing correlations related to similarity in stimulus preference. We used a balanced z-scoring method proposed recently to prevent bias related to differences in choice distributions across signal levels[45]. We excluded pairs that were not simultaneously isolated for at least 25 trials total. The median number of trials per pair during task performance was 752.

A main goal of the study was to measure how spike-count correlation varies with pairwise orientation. We describe this relationship as a smoothed function estimated from measures of r_{sc} combined across multiple recording sessions, which we then sampled discretely with 1° resolution. The smoothed estimates were obtained using a bivariate von Mises smoothing kernel. A point in the correlation matrix **C** was given as:

$$\mathbf{C}(x,y) = \tanh \frac{\sum_{i=1}^{n} z_i K(x,y,\theta_i,\phi_i)}{\sum_{i=1}^{n} K(x,y,\theta_i,\phi_i)}, \text{ where } K(x,y,\theta_i,\phi_i) = e^{\kappa\,(\cos(\theta_i - x) + \cos(\phi_i - y))}, \quad (1)$$

z_i is the i^{th} (Fisher z-transformed) r_{sc} measurement, θ_i and ϕ_i are the preferred orientations of the i^{th} pair, and κ is the von Mises dispersion parameter. We set $\kappa = 1.3\pi$, yielding a smoothing kernel closely approximating a bivariate wrapped Gaussian with 15° s.d. In some cases, we expressed the r_{sc} matrix in a task-aligned coordinate frame (e.g. Fig. 4b), for which the preferred orientations of the i^{th} pair relative to the discriminandum orientations were used for θ_i and ϕ_i. Since there were always two orthogonal task orientations, we averaged across both possible alignments, such that $\mathbf{C}(x, y) = \mathbf{C}(x + 90°, y + 90°)$. All angular quantities were doubled for the calculations, as orientation has a period of 180°. To generate the kernel-smoothed profile of CP (Fig. 8), we used a one-dimensional equivalent of the procedure above, in which preferred orientations were parameterized only by a single parameter.

Regression model

We used a multilinear regression model to identify fixed and task-dependent components of the structured correlations we observed. Our approach was to describe the set of observations (811 individual pairwise r_{sc} measurements, Fisher z-transformed to produce normal error) in terms of a set of two underlying correlation structures: one defining r_{sc} as a function of pairwise preferred orientation alone ("fixed") and the other defining r_{sc} as a function of pairwise preferred orientation relative to the discriminandum orientations ("task-dependent"). In order to provide a continuous and smooth description of the data, each component was parameterized as the sum of an array of $n\ x\ n$ evenly spaced basis functions. Each observation, y_i, was expressed as:

$$y_i = x_i^{fixed} \cdot \beta^{fixed} + x_i^{task} \cdot \beta^{task} + \beta_0 + \varepsilon_i \tag{2}$$

x_i^{fixed} and x_i^{task} are length-n^2 vectors of loadings onto the basis functions, which were given by evaluating the basis functions at the location corresponding to the pairwise orientation preference of the i^{th} pair. β^{fixed} and β^{task} are the length-n^2 vectors of amplitudes of the basis functions (coefficients to be fit), β_0 is a model constant, and $\cdot$ is the element-wise product. For the basis functions, we used bivariate von Mises functions, with no correlation and equal dispersion in both dimensions. Thus the k^{th} loading ($x_i^{fixed}(k)$ or $x_i^{task}(k)$) was given by:

$$x_i(k) = \frac{e^{\kappa\left(\cos\left(\theta_i - \mu_k^1\right) + \cos\left(\phi_i - \mu_k^2\right)\right)}}{Z} \tag{3}$$

where θ_i and ϕ_i are the preferred orientations of the i^{th} pair (relative to the discriminandum orientations in the case of the task-dependent loadings), μ_k is a pair of orientations defining the location of the k^{th} basis function, Z is a normalization constant such that the sum of all loadings for observation i ($x_i^{fixed} + x_i^{task}$) is 1, and κ is the von Mises dispersion parameter. Again, angular quantities were doubled and κ was set to 1.3π. We found that arrays of 8x8 were sufficient to describe the structure of the two components. Because the observations were pairwise correlations, it was sufficient only to fit the upper triangular portion of the array of basis functions. Thus, the two-component model contained 73 parameters (36 for each component, plus the model constant).

We fit the model by finding the parameters (β^{fixed}, β^{task} & β_0) that minimized the L1 error (to encourage sparseness) plus two additional terms that encouraged smoothness and symmetric positive semi-definiteness, as the two components were meant to represent correlation matrices. The solution was obtained as:

$$\hat{\beta}^{fixed}, \hat{\beta}^{task}, \hat{\beta}_0 = \underset{\beta^{fixed}, \beta^{task}, \beta_0}{\operatorname{argmin}} \sum_i |\varepsilon_i| + \alpha_1 \Gamma\left(\beta^{fixed} + \beta^{task}\right) + \alpha_2 D_{SPD}\left(\beta^{fixed} + \beta^{task}\right) \tag{4}$$

where Γ is the discrete 2D Laplace operator, corresponding to circular convolution with the kernel:

$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ and $D_{SPD}(X)$ is the 2-norm between X and the nearest symmetric positive semidefinite matrix $\hat{X}$, which is given by $(B + H)/2$ where H is the symmetric polar factor of $B = \frac{(X+X')}{2}$ [46]. The α's controlled the strength of regularization and were chosen to produce the best fit (as measured with R^2 under 10-fold cross-validation). The solution was obtained by gradient descent using the MATLAB function *fminsearch*.

While this model did not explain more than a small percentage of the variance of the raw observed r_{sc} values, this is not surprising as the raw correlation data do not vary smoothly with preferred orientation (reflecting both noise, and the fact that r_{sc} is known to depend on parameters other than orientation.[1,5,6]). For this reason, we measured goodness-of-fit relative to an estimate of the explainable variance, which we took as the variance explained simply by a smoothed version of the raw data (sum of values in fixed and task-aligned matrices was 3.6%).

Choice probability predictions

Choice Probability was calculated in the standard way[29]. We only used 0%-signal trials, as the uneven choice distributions elicited by signal trials yield noisier CP measurements. Assuming feedforward pooling with linear readout weights, the relationship between the covariance matrix for a population of neurons, the readout weight of each neuron, and the Choice Probability (*CP*) of each neuron is:

$$CP_k = \frac{1}{2} + \frac{2}{\pi}\operatorname{sgn}(\xi_k)\arctan\sqrt{2{\xi_k}^{-2} - 1}^{\,-1} \quad \text{with} \quad \xi_k = \frac{(\mathbf{C}\beta)_k}{\sqrt{C_{kk}\beta^{\mathrm{T}}\mathbf{C}\beta}} \tag{5}$$

where CP_k is the CP of neuron k with respect to choice 1, β is the vector of readout weights and **C** is the covariance matrix[15]. We used this known relationship to quantify the CPs that would be associated with the r_{sc} structure we observed and the fixed and task-dependent components we identified, assuming only a feedforward source of CP (Fig. 8). CPs, r_{sc} structure, and readout weights were described as task-aligned functions of preferred orientation (i.e. with orientation expressed relative to the discriminandum orientations). This is equivalent to assuming a population of infinite size that is homogeneous at a given orientation. For the fixed component of r_{sc}, which was indexed relative to raw orientation preferences, we generated a task-aligned version by substituting the observed r_{sc} values with model fits (using only a fixed component of the model) and then generating a smoothed task-aligned matrix as in Fig. 4b, using these substituted values. To guarantee real-valued CPs on [0,1], we performed the calculations using a symmetric positive definite approximation[46] of the r_{sc} matrices, which introduced negligible error.

Since the readout weights were unknown, we generated a random distribution of plausible readout weights that could support task performance. To do this, we started with a vector of random weights (drawn from a normal distribution) and applied the 90° symmetry inherent in the task, such that $\beta_\theta = -\beta_{\theta+90}$, where β_θ is the weight assigned to neurons with task-relative preferred orientation θ. Then, we smoothed the readout weight profiles with a wrapped Gaussian kernel with 15° s.d. and excluded profiles which did not have a circular mean within 22.5° of choice 1 (0°). In practice, we found the CP predictions to be remarkably insensitive to the readout weights (Supplementary Fig. 10). This may be due to the use of a smoothed estimate of the r_{sc} matrix, which necessarily limits the effect of varying the readout weights. If we had enough data to reliably detect sharp discontinuities in the r_{sc} matrix, then the exact assumptions of readout could have a larger influence on predicted CP[15].

Estimating mean covariance for a population of neurons is necessarily more error-prone than estimating mean correlation, as the former is sensitive to sampling error in measurements of average spike-count variance (and therefore firing rate), so for this reason we chose to perform the calculations using correlations (see Supplemental Discussion §2). We can use correlations interchangeably with covariances in equation 1, under the simplifying assumption that the variance is uniform as a function of preferred orientation. If $\mathbf{\Sigma}$ is the correlation matrix for a population with uniform variance α, then it follows that:

$$\xi_k = \frac{a(\mathbf{\Sigma}\beta)_k}{\sqrt{a\Sigma_{kk}\beta^{\mathrm{T}}(a\mathbf{\Sigma})\beta}} = \frac{(\mathbf{\Sigma}\beta)_k}{\sqrt{\Sigma_{kk}\beta^{\mathrm{T}}\mathbf{\Sigma}\beta}} \tag{6}$$

where $\Sigma_{kk} \equiv 1$ for all k. We felt that spike-count variance that depended systematically on preferred orientation was unlikely to be a feature of the V1 representation, and thus that the advantages of using correlations outweighed the cost.

Noise in the decision process after pooling (pooling noise) has the effect of uniformly scaling down CPs, such that ξ_k in Eq. 5 is substituted with: $\frac{(\mathbf{C}\beta)_k}{\sqrt{C_{kk}(\beta^{\mathrm{T}}\mathbf{C}\beta+\sigma^2_{pool})}}$, where σ^2_{pool} is the variance of the pooling noise[6]. We found that non-zero pooling noise was needed to avoid overestimating the magnitude of CP from the observed correlation structure. We used a fixed value of pooling noise in our predictions such that the average squared difference between the CP profile predicted from the observed correlation matrix and the observed CP profile was minimized. Empirically, we found that pooling noise variance of 0.6 was optimal. Since our spike counts were normalized to have unit variance, this implies pooling noise whose variance is 60% of the average spike-count variance of single neurons. This should be interpreted with care, as overestimation of CPs may also be an artefact related to the assumption of a homogeneous population[15]. Alternatively, the need to invoke pooling noise may be due to nonuniform sensory integration across the trial, which is distinct but which would also have an attenuating effect on CP when measured over the entire trial.

Calculating differential correlations

The information capacity of a sensory population, assuming a linear read out, is bounded when the spike-count covariances sufficiently match the differential correlations[27]. Since we made use of spike-count correlations, rather than covariances, in the present study, we normalized the measurements of differential correlations by the product of the standard deviations of the stimulus-independent variability of each pair.

1. Cohen, M. R. & Kohn, A. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* **14,** 811–9 (2011).

2. Zohary, E., Shadlen, M. N. & Newsome, W. T. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* **370,** 140–143 (1994).

3. Bair, W., Zohary, E. & Newsome, W. T. Correlated firing in macaque visual area MT: time scales and relationship to behavior. *J. Neurosci.* **21,** 1676–97 (2001).

4. Lee, D., Port, N. L., Kruse, W. & Georgopoulos, a P. Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex. *J. Neurosci.* **18,** 1161–70 (1998).

5. Smith, M. A. & Kohn, A. Spatial and temporal scales of neuronal correlation in primary visual cortex. *J. Neurosci.* **28,** 12591–603 (2008).

6. Kohn, A. & Smith, M. a. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J. Neurosci.* **25,** 3661–73 (2005).

7. Abbott, L. F. & Dayan, P. The effect of correlated variability on the accuracy of a population code. *Neural Comput.* **11,** 91–101 (1999).

8. Sompolinsky, H., Yoon, H., Kang, K. & Shamir, M. Population coding in neuronal systems with correlated noise. *Phys. Rev. E* **64,** 51904 (2001).

9. Snippe, H. & Koenderink, J. Information in channel-coded systems: correlated receivers. *Biol. Cybern.* **190,** 183–190 (1992).

10. Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7,** 358–66 (2006).

11. Cohen, M. R. & Maunsell, J. H. R. Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* **12,** 1594–600 (2009).

12. Graf, A. B. a, Kohn, A., Jazayeri, M. & Movshon, J. A. Decoding the activity of neuronal populations in macaque primary visual cortex. *Nat. Neurosci.* **14,** 239–45 (2011).

13. Mitchell, J. F., Sundberg, K. a & Reynolds, J. H. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* **63,** 879–88 (2009).

14. Johnson, K. O. Sensory discrimination: decision process. *J. Neurophysiol.* **43,** 1771–1792 (1980).

15. Haefner, R. M., Gerwinn, S., Macke, J. H. & Bethge, M. Inferring decoding strategies from choice probabilities in the presence of correlated variability. *Nat. Neurosci.* **16,** 235–42 (2013).

16. Nienborg, H. & Cumming, B. G. Correlations between the activity of sensory neurons and behavior: how much do they tell us about a neuron's causality? *Curr. Opin. Neurobiol.* **20,** 376–381 (2010).

17. Shadlen, M. N., Britten, K. H., Newsome, W. T. & Movshon, J. A. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci.* **16,** 1486–510 (1996).

18. Callaway, E. M. Feedforward, feedback and inhibitory connections in primate visual cortex. *Neural Netw.* **17,** 625–32 (2004).

19. Sillito, A. M., Cudeiro, J. & Jones, H. E. Always returning: feedback and sensory processing in

visual cortex and thalamus. *Trends Neurosci.* **29,** 307–16 (2006).

20. Cumming, B. G. & Nienborg, H. Feedforward and feedback sources of choice probability in neural population responses. *Curr. Opin. Neurobiol.* **37,** 126–132 (2016).

21. Nienborg, H. & Cumming, B. G. Decision-Related Activity in Sensory Neurons May Depend on the Columnar Architecture of Cerebral Cortex. *J. Neurosci.* **34,** 3579–3585 (2014).

22. Nienborg, H. & Cumming, B. G. Psychophysically measured task strategy for disparity discrimination is reflected in V2 neurons. *Nat. Neurosci.* **10,** 1608–1614 (2007).

23. Ahumada Jr, A. J. Perceptual classification images from Vernier acuity masked by noise. in *Perception ECVP abstract* **25,** 0 (Pion Ltd, 1996).

24. Ecker, A. S., Denfield, G. H., Bethge, M. & Tolias, A. S. On the Structure of Neuronal Population Activity under Fluctuations in Attentional State. *J. Neurosci.* **36,** 1775–1789 (2016).

25. Haefner, R. M., Berkes, P. & Fiser, J. Perceptual Decision-Making as Probabilistic Inference by Neural Sampling. *Neuron* **90,** 649–660 (2016).

26. Shamir, M. & Sompolinsky, H. Implications of neuronal diversity on population coding. *Neural Comput.* **18,** 1951–1986 (2006).

27. Moreno-Bote, R. *et al.* Information-limiting correlations. *Nat. Neurosci.* **17,** 1410–1417 (2014).

28. Kanitscheider, I., Coen-Cagli, R., Kohn, A. & Pouget, A. Measuring Fisher Information Accurately in Correlated Neural Populations. *PLoS Comput. Biol.* **11,** 1–27 (2015).

29. Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S. & Movshon, J. A. A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci.* **13,** 87 (1996).

30. Crapse, T. B. & Basso, M. A. Insights into Decision-Making Using Choice Probability. *J. Neurophysiol.* jn.00335.2015 (2015). doi:10.1152/jn.00335.2015

31. Nienborg, H. & Cumming, B. G. Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature* **459,** 89–92 (2009).

32. Wimmer, K. *et al.* Sensory integration dynamics in a hierarchical network explains choice probabilities in cortical area MT. *Nat. Commun.* **6,** 6177 (2015).

33. Nienborg, H., Cohen, M. R. & Cumming, B. G. Decision-Related Activity in Sensory Neurons: Correlations Among Neurons and with Behavior. *Annu. Rev. Neurosci.* **35,** 463–483 (2012).

34. Pitkow, X., Liu, S., Angelaki, D. E. E., DeAngelis, G. C. C. & Pouget, A. How Can Single Sensory Neurons Predict Behavior? *Neuron* **87,** 411–423 (2015).

35. Ruff, D. A. & Cohen, M. R. Attention can either increase or decrease spike count correlations in visual cortex. *Nat. Neurosci.* **17,** 1591–7 (2014).

36. Rabinowitz, N. C., Goris, R. L., Cohen, M. & Simoncelli, E. Attention stabilizes the shared gain of V4 populations. *Elife* **4,** e08998 (2015).

37. Newsome, W. T., Britten, K. H., Movshon, J. A. & Shadlen, M. N. in *Neural Mechanism of Visual Perception* 171–197 (1989).

38. Kwon, S. E., Yang, H., Minamisawa, G. & O'Connor, D. H. Sensory and decision-related activity propagate in a cortical feedback loop during touch perception. *Nat. Neurosci.* **19,** (2016).

39. Lange, R. D. & Haefner, R. M. Inferring the brain's internal model from sensory responses in a probabilistic inference framework. (2016). doi:10.1101/081661

40. Orbán, G., Berkes, P., Fiser, J. & Lengyel, M. Neural Variability and Sampling-Based Probabilistic Representations in the Visual Cortex. *Neuron* **92,** 530–543 (2016).

41. Von Helmholtz, H. *Handbuch der physiologischen Optik*. **9,** (Voss, 1867).

42. Knill, D. C. & Richards, W. *Perception as Bayesian Inference*. (Cambridge University Press, 1996).

43. Judge, S. J., Richmond, B. J. & Chu, F. C. Implantation of magnetic search coils for measurement of eye position: An improved method. *Vision Res.* **20,** 535–538 (1980).

44. Cumming, B. G. & Parker, A. J. Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *J. Neurosci.* **19,** 5602–18 (1999).

45. Kang, I. & Maunsell, J. H. R. Potential confounds in estimating trial-to-trial correlations between neuronal response and behavior using choice probabilities. *J. Neurophysiol.* **108,** 3403–15 (2012).

46. Higham, N. J. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra Appl.* **103,** 103–118 (1988).

Acknowledgements

We thank James McFarland for useful discussions; Ralf Haefner for advice about the analysis; Richard Krauzlis, Bevil Conway, and Ali Ghazizadeh for comments on an earlier version of the manuscript; and Beth Nagy, Irina Bunea, and Denise Parker for veterinary care.

Author Contribution

A.G.B. and B.G.C. conceived and designed the experiments. A.G.B. performed the experiments and all aspects of the analysis. A.G.B. and B.G.C. wrote the paper. B.G.C. advised at all stages.

Competing Financial Interests

The authors declare no competing financial interests.