Learning orthogonalizes visual cortical population codes

Samuel W. Failor[1*], Matteo Carandini[2†], Kenneth D. Harris[1†]
[1]UCL Queen Square Institute of Neurology, University College London, London, United Kingdom
[2]UCL Institute of Ophthalmology, University College London, London, United Kingdom
*Correspondence: s.failor@ucl.ac.uk
†Co-senior authors

The response of a neuronal population to a stimulus can be summarized by a vector in a high-dimensional space. Learning theory suggests that the brain should be most able to produce distinct behavioral responses to two stimuli when the rate vectors they evoke are close to orthogonal. To investigate how learning modifies population codes, we measured the orientation tuning of 4,000-neuron populations in visual cortex before and after training on a visual discrimination task. Learning suppressed responses to the task-informative stimuli, most strongly amongst weakly-tuned neurons. This suppression reflected a simple change at the population level: sparsening of population responses to relevant stimuli, resulting in orthogonalization of their rate vectors. A model of F-I curve modulation, requiring no synaptic plasticity, quantitatively predicted the learning effect.

When an animal sees a stimulus, this triggers a pattern of activity across a multitude of neurons in its visual cortex. These neurons' firing rates together define a representation of the stimulus in a high-dimensional vector space, similar to the high-dimensional representations constructed by machine learning algorithms (*1*, *2*). The similarity of the representations of two stimuli can be quantified by the angle or dot product between the corresponding vectors. In machine learning this similarity measure, known as the "kernel function" (*2*), determines the generalizability of stimuli: stimuli evoking identical representations will generalize perfectly, while stimuli evoking orthogonal representations will not generalize at all. Allowing the representations to themselves change with learning is the heart of flexible learning algorithms such as neural networks (*1*).

In the brain, stimuli that evoke similar neural representations are likely to evoke similar behavioral responses (*3*, *4*). Furthermore, stimulus representations evolve as animals learn, even in primary sensory cortices. One might expect that after learning, the number of neurons selective for behaviorally-important stimuli increases, as has been observed in auditory (*5*, *6*), somatosensory (*7*, *8*), and visual cortex (*9*, *10*). Other studies, however, have found a paradoxical decrease in the number of cortical neurons responding optimally to learned stimuli (*11*, *12*), and in primary visual cortex (V1), neurons increase their slope at the task stimulus in a manner dependent on orientation preference (*13*). It is not yet clear whether these complex and apparently contradictory findings result from a single principle governing plasticity of visual cortical representations at a population level.

Here, we use two-photon calcium imaging to show how the tuning of populations of thousands of V1 neurons changes after mice learn an orientation discrimination task. At a single-cell level, the results appear complex: neuronal tuning curves evolve according to a lawful but complicated dependence on their prior orientation preference and tuning strength. At the population level, a simple principle emerges: learning transforms response vectors by a nonlinear function, whose convexity is largest for task-informative stimuli. This transformation sparsens the population representations and makes them more orthogonal. The degree of sparsening varied consistently across the population on a trial-by-trial basis, suggesting it emerges from rapid circuit dynamics, rather than slower plasticity mechanisms.

Results

An orientation discrimination task for mice

To study how cortical representations change with learning, we trained mice in an orientation discrimination task (**Figure 1A,B**). This task required turning a steering wheel to select one of two oriented cues, each of which could take on three different orientations. Two of these orientations were informative (45° and 90°) but

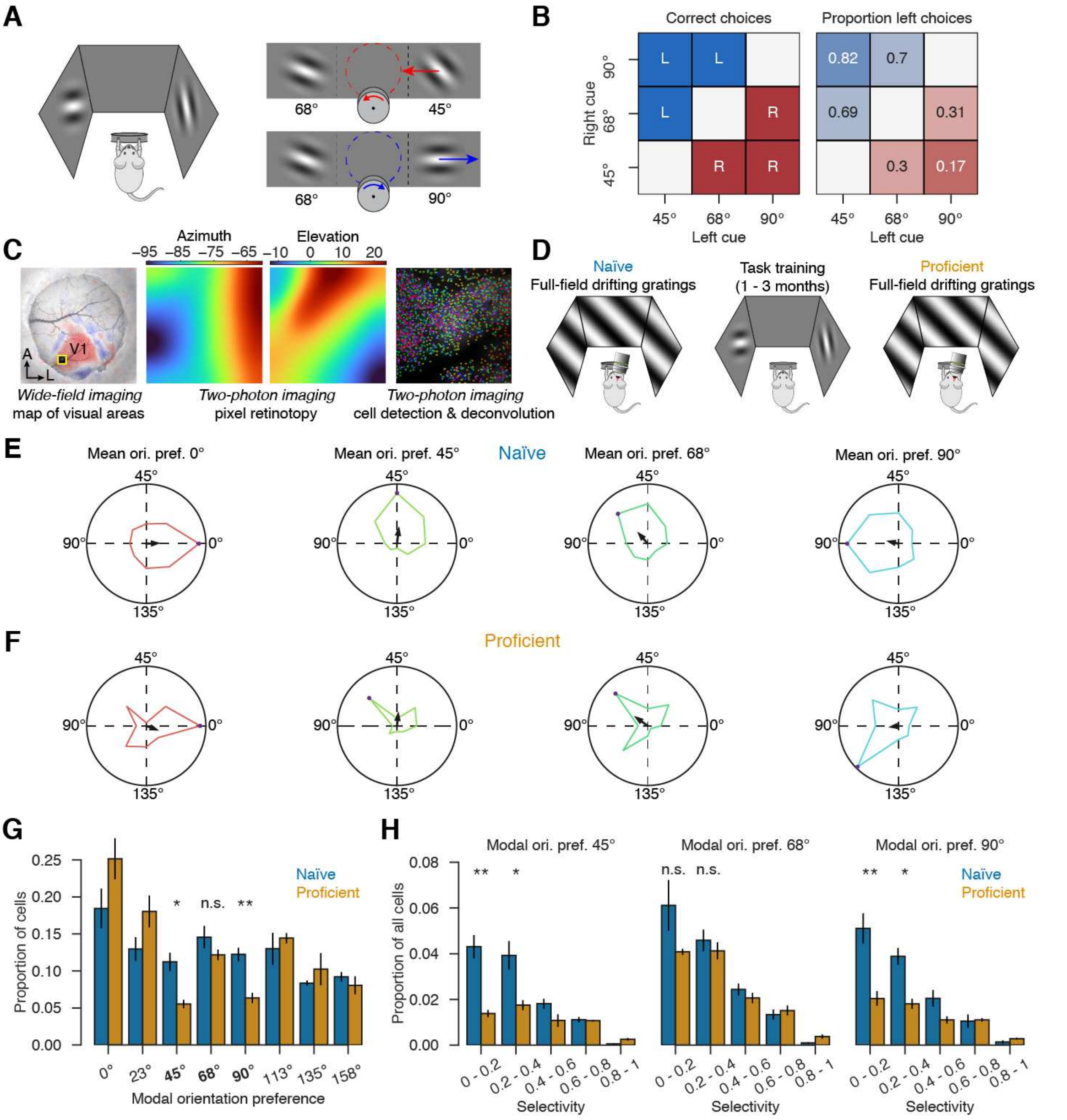

Figure 1. Learning an orientation discrimination task reduces the proportion of neurons responding maximally to task-informative orientations. (**A**) The orientation discrimination task. On each trial mice are presented with two stimuli and then turn a wheel to move them on the screen. Bringing a 45º stimulus to the center, or a 90º stimulus away from the center yields a reward, but 68º stimuli are uninformative. (**B**) Correct choices for all stimulus pairings (left) and the average proportion of left choices across mice taken from their ten highest performing sessions (right). (**C**) Pipeline for imaging neural activity. Left: V1 was located using widefield imaging with sparse noise stimuli (red/blue: sign map; yellow outlined square: region selected for two-photon imaging). Middle: retinotopy map for the two-photon field of view. Right: colored outlines of detected cells. (**D**) Timeline of experiments. **(E)** Single-cell orientation tuning curves from naïve mice, for four cells with mean orientation preference 0°, 45°, 68°, and 90°. Colored polar curves: neural response to each orientation; dots: response to modal orientation; arrows: circular mean vectors representing mean orientation preference (angle) and orientation selectivity (magnitude). (**F**) Similar plots for mice proficient at the task (**G**) Proportion of cells with each modal orientation preference, in naïve and proficient mice. Error bars: SEM (n = 5 mice). (**H**) Proportion of cell population that had modal orientation preference 45° (left), 68° (center), and 90° (right) and specified orientation selectivity. *, p < 0.05, **, p < 0.01.

had opposite behavioral contingencies (select vs. avoid) and a third was an uninformative distractor (68°). All mice included in this study successfully learned the task (**Figure S1**).

To study how training in the task affected the neural representations of visual stimuli, we assessed the orientation tuning of excitatory cells in V1 using two-photon calcium imaging (**Figure**

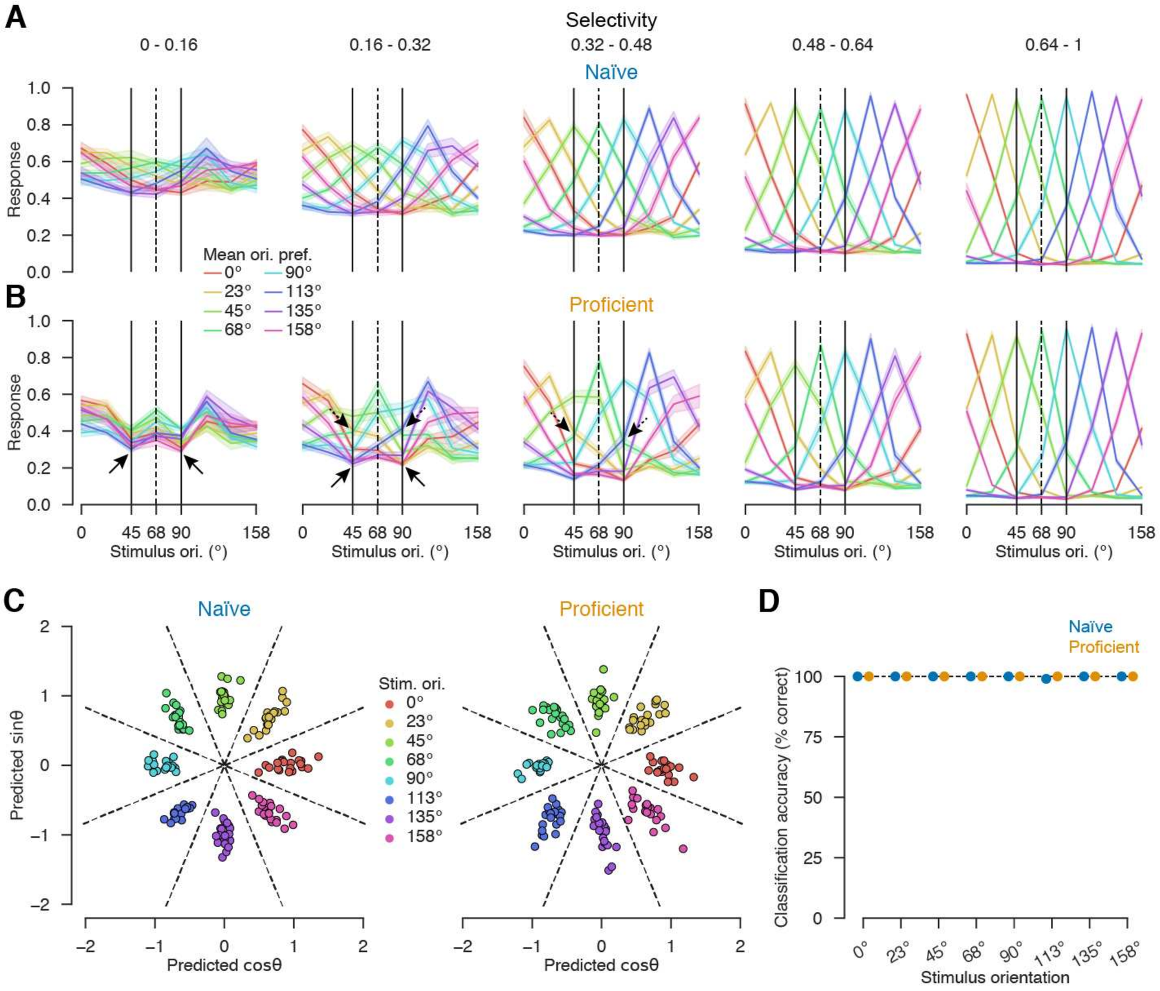

Figure 2. Learning has multifarious effects on orientation tuning curves. (**A**) Average orientation tuning curves for cell groups defined by mean orientation preference (color) and selectivity level (column) for naïve mice. Solid vertical lines indicate task-informative orientations, dashed uninformative (68°). **(B)** Same plot for proficient mice. Solid arrows highlight suppression of cell responses to the informative task orientations 45° and 90°. Dashed arrows highlight tuning curve asymmetry for cells with preferences near the informative orientations. Shading: SEM (n = 5 mice) (**C**) 2d projection of population response vectors for each orientation from one mouse before (left) and after learning (right). (**D**) Cross-validated classification accuracy for decoding stimulus orientation from naïve and proficient mice. Dashed line indicates perfect performance. (n = 5 mice)

1C, D). We obtained two recordings in passive conditions: one before task training began (naïve condition), and one after training was complete (proficient condition). In both cases, mice were placed in the same apparatus where they performed the task, and drifting gratings were presented; no rewards were delivered, and the wheel was not coupled to visual stimuli. In this passive condition, the presentation of visual stimuli triggered minimal whisking, and a pupillary light reflex, neither of which differed significantly between stimuli or training conditions (**Figure S2**). Thus, even though body movements and changes in arousal strongly modulate visual cortical activity (*14–16*), analyzing passive stimulus responses avoids this potential confound.

Training in the task decreased the fraction of neurons preferentially tuned to the task-informative orientations, and this decrease was specific to weakly tuned cells (**Figure 1E-H**). We defined a cell's modal orientation preference to be the stimulus orientation driving it to fire maximally (dots in **Figure 1E-F**). Task training significantly decreased the fraction of neurons whose modal orientation preference was one of the two task-informative orientations (45° and 90°), but not the fraction of neurons preferring the distractor orientation (68°) (**Figure 1G;** 45°: p = 0.012; 68°: p = 0.228; 90°: p = 0.006, paired-sample *t*-test, n = 5 mice). To further characterize tuning curves, we defined a cell's circular mean response as a vector in a complex plane (arrows in **Figure 1E-F**); the length and angle of this vector defined the cell's selectivity index and mean orientation preference. This analysis showed that the decrease in cells modally preferring the task-informative orientations came only from weakly-tuned cells: there was no decrease in the number of cells strongly tuned to the informative stimuli (**Figure 1H**; 45°: p = 0.005 and 0.037 for selectivity indices 0 - 0.2 and 0.2 - 0.4; 68°: p = 0.130 and 0.390; 90°: p = 0.001 and 0.013, paired samples *t*-test, n = 5 mice).

Task-informative orientations suppress weakly-tuned cells

Tuning curves also changed shape after training, in a manner dependent on a cell's preferred orientation and selectivity (**Figure 2A,B**). We grouped the recorded cells by their selectivity and mean orientation preference and plotted the mean tuning curves of cells in each group before and after training, using held-out repeats. In mice that had not learned the task, tuning curves had a uniform structure (**Figure 2A**). By construction, these curves peaked at the cells' mean orientation preference, and the depth of modulation increased with the cells' selectivity index. For trained mice, however, a different structure appeared (**Figure 2B**). Weakly tuned neurons were suppressed by the task-informative orientations regardless of their preference. Cells whose mean orientation preference was at or close to a task-informative orientation exhibited bimodal tuning curves after training, for which the mean and modal orientation preference differed (examples in **Figure 1F**). For more strongly tuned cells, suppression by task-informative orientations was still visible, primarily in neurons with a mean orientation preference adjacent to them. This suppression led to an asymmetry in tuning curve slopes (**Figure S3A**), as previously reported in primate (*13*). At a single-cell level, we observed a training-dependent increase in the magnitude of the d′ statistic that measures distinguishability of task-informative orientations (**Figure S3C-D**) as previously observed (*9*), primarily attributable to a decrease in the standard deviation of responses to these stimuli (**Figure S3E**).

These changes in cellular tuning did not improve the ability to decode stimulus orientation from population activity, because the stimulus could be decoded exceptionally well even prior to task training (**Figure 2C,D**). Failures of stimulus decoding can occur even in large populations of well-tuned cells, if the structure of trial-to-trial variability matches differences between stimuli (*17–19*). In the current case, however, a simple two-dimensional projection showed that trial-to-trial variability did not interfere with stimulus coding before or after training (**Figure 2C**), and linear discriminant analysis gave near perfect accuracy in both training conditions and for all orientations (**Figure 2D;** naïve: 799/800 trials correct; proficient: 808/808 trials correct**)**. Indeed, populations of V1 neurons can reliably encode much finer stimulus orientation than demanded by our task (*20*).

Training sparsens and orthogonalizes responses to task-informative orientations

Although task training did not improve the decodability of the population activity, it did change its character, sparsening the population responses to task-informative orientations (**Figure 3A,B**). We quantified the sparseness of population activity using kurtosis and found that proficient mice exhibited significantly higher population sparseness for the task-informative orientations than for the distractor (**Figure 3B**; 45° vs 68°: p = 0.008; 68° vs 90°: p = 0.023; 45° vs 90°: p = 0.340. Welch's *t*-test, n = 5 mice).

This sparsening took a specific form: it made the population responses to the task-informative orientations more orthogonal to each other (**Figure 3A,C,D**). Training reduced the cosine similarity between population response vectors to the task-informative orientations compared to control orientations (**Figure 3C-D**; p = 0.006. Independent samples *t*-test, n = 5 mice). Thus, by increasing the number of zero components in the population response vectors (i.e. sparsening), training moved them closer to the coordinate axes of N-dimensional space, and thereby orthogonalized them (**Figure 3E**).

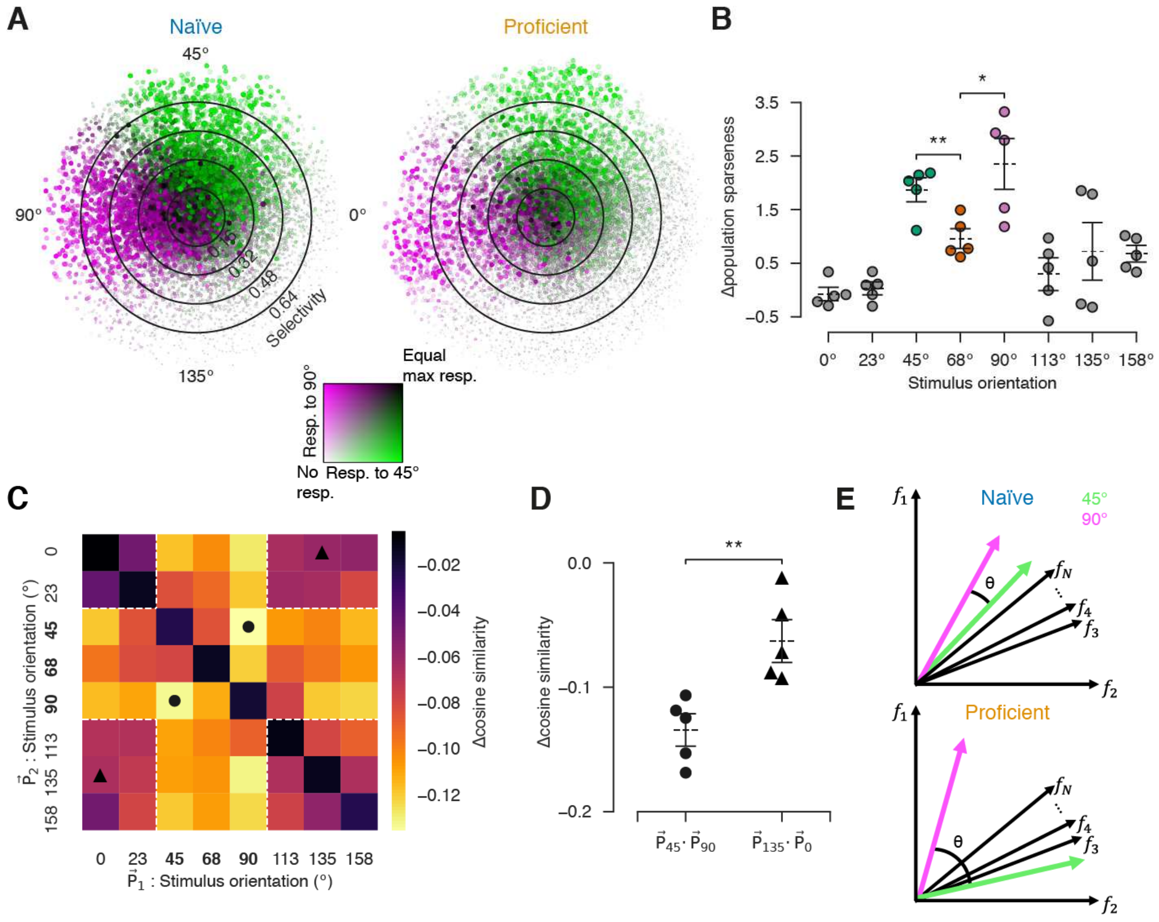

Figure 3. Learning sparsens and orthogonalizes responses to informative task orientations. (**A**) "Bullseye plots" showing structure of population responses to the informative task orientations 45° and 90°. Each point represents a cell. The point's location in polar coordinates indicates the cell's circular mean orientation preference (angle) and selectivity (distance from the origin). The point's hue represents the cell's relative response to the 45° and 90° stimulus orientations (green to magenta); the point's size and brightness (light to dark) represents the cell's maximal response to these two stimuli. (**B**) Change in population sparseness following task learning, as a function of stimulus orientation. Error bars: mean and SEM (n = 5 mice). (**C**) Change in cosine similarity between mean population responses to each pair of orientations after task learning. White dashed lines demarcate task stimuli. Black circles and triangles indicate the orientation pairs shown in (D). (**D**) Change in cosine similarity of population responses to 45° and 90°, and between 135° and 0°, after task learning. Error bars: mean and SEM (n = 5 mice). (**E**) Illustration of learning's effect on population response vectors to task-informative stimuli. Sparsening of population responses moves the vectors closer to coordinate axes and increases the angle between them. *, $p < 0.05$, **, $p < 0.01$.

A model for learning-evoked sparsening

These apparently complex learning-induced tuning changes could be accurately predicted by a simple computational model (**Figure 4A**). Plasticity of cortical representations is often assumed to arise from coordinated plasticity of local excitatory synapses (*21*, *22*), but our observations suggested an alternative possibility. Because training further attenuated the responses to task-informative orientations in cells that already responded weakly to them, we hypothesized that the effects of training could be explained by a stimulus-dependent modulation in the relationship between excitatory input and firing rate (the *f-I* curve; **Figure 4A**). Under this hypothesis, cortical neurons receive a stimulus-dependent bottom-up excitatory input that is unaffected by learning, but also receive a feedback signal that, after learning, is activated by salient stimuli. This signal changes the way that local cells respond to excitatory inputs, specifically reducing responses to weak excitation. This could be instantiated by multiple possible network mechanisms, for example feedback inhibition from a local inhibitory cell class, or activation of a long-range neuromodulatory system by task-informative stimuli.

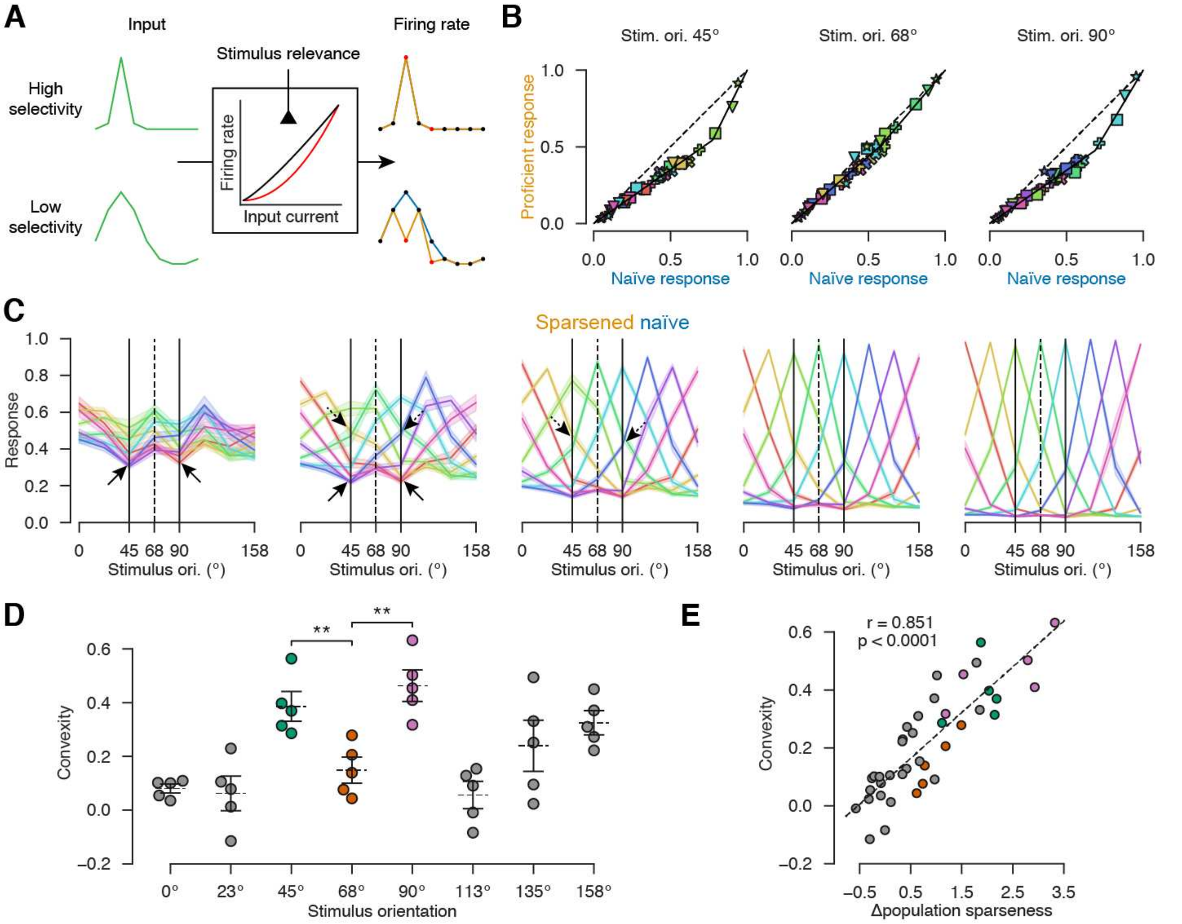

Fig. 4. Model of learning-evoked sparsening by *f-I* curve modulation. (**A**) Model schematic. Following task learning the *f-I* curve is stimulus-dependent, becoming more convex when informative task stimuli are presented. This spares responses in cells with high selectivity (top) but suppresses responses to informative stimuli in cells with low selectivity (bottom). (**B**) Effect of learning on responses to task stimulus orientations as a function of cell orientation preference (color) and selectivity (symbol). Symbols correspond to the selectivity bins of the columns in Fig. 2A-B ordered **X**, ✚, ■, ▼, ★. Each point shows the average response of cells from all experiments. Black lines are stimulus-specific fits of piecewise linear functions relating naïve responses to proficient responses. (**C**) Orientation tuning curves predicted by the model, obtained by applying the functions fit in (B) to naïve tuning curves. Solid and dashed arrows highlight the same features seen in the actual proficient responses, as shown in Fig. 2B. Shading: SEM (n = 5 mice). (**D**) Convexity of naïve-to-proficient transformation functions, for each stimulus orientation. Points indicate individual mice. Error bars: mean and SEM (n = 5 mice). (**E**) Relationship between learning-evoked changes in population sparseness and convexity of naïve-to-proficient transformation. Each point represents a stimulus orientation in a single experiment. Points for 45°, 68°, and 90° are colored as in (D). *, p < 0.05, **, p < 0.01.

This hypothesis makes a strong and testable prediction: that population responses before and after training can be related by a single function that depends on the stimulus but not the cell. Denote the bottom-up input received by cell c following stimulus θ by $I_{c,\theta}$, and the pre-training f-I curve by $h(I)$; thus, before training, the response of cell c to stimulus θ is $f_{c,\theta} = h(I_{c,\theta})$. Our model holds that after training the f-I curve depends on the stimulus, but not the cell; denoting it by $h'_\theta(I)$, the post-training response of cell c to stimulus θ is $f'_{c,\theta} = h'_\theta(I_{c,\theta})$. The firing rates before and after learning are therefore predicted to be related as $f'_{c,\theta} = g_\theta(f_{c,\theta})$, where the function $g_\theta(f) = h'_\theta(h^{-1}(f))$ predicts proficient from naïve responses, in a manner that depends on the stimulus θ, but not on the cell. Furthermore, it can be proven that if the function g_θ is convex, then population sparseness will increase after learning (Appendix).

To test this prediction, we attempted to relate pre- and post-learning responses through a function which varies between stimulus orientations but not between cells, with good success (Figure 4B-E). Responses before and after training could be accurately related by piecewise linear functions (**Figure 4B; Figure S4**). The shape of the function

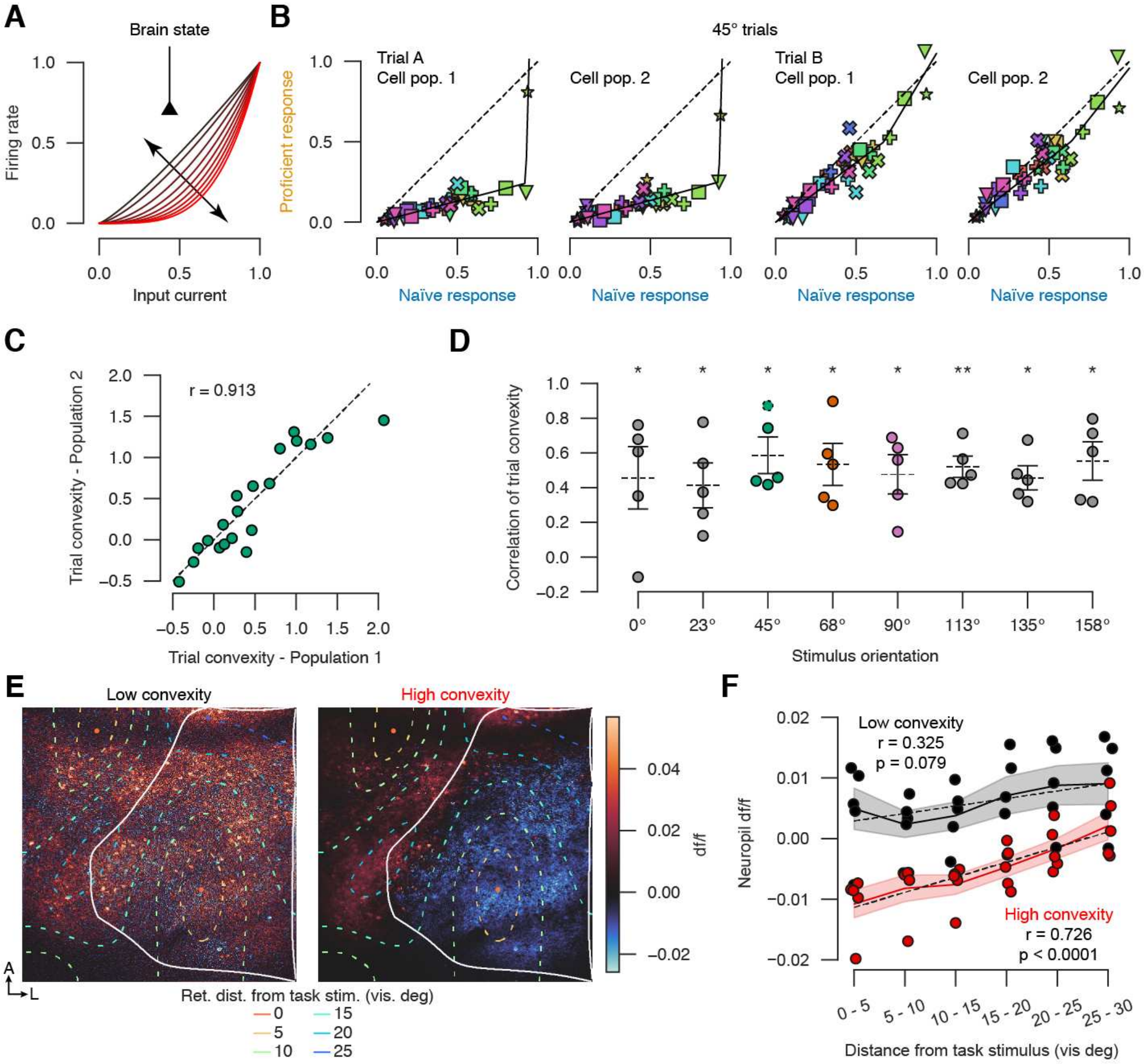

Figure 5. Trial-to-trial variability in response sparsening. (**A**) Dynamic sparsening model: cells undergo varying levels of *f-I* curve modulation dependent on brain state. (**B**) Single-trial sparsening functions for two example 45° trials from the same recording session, plotted as Figure 4B. For each trial, responses of separate halves of the cell population are shown. (**C**) Similarity of single-trial convexities between two different halves of the cell population, for the recording in (B). Each point represents a single presentation of the 45° stimulus. (**D**) Correlation of single-trial convexities between two halves of cells, with each point representing one stimulus orientation in one experiment. Point with dashed outline is the session shown in (C). Error bars: mean and SEM (n = 5 mice). (**E**) Trial-to-trial variability of neuropil responses. Left and right plots show mean df/f of two-photon imaging frames to task-informative orientations for low convexity (< 0) and high convexity (> 0.3) trials. Colored contours correspond to retinotopic distances from task stimulus location (see legend). (**F**) V1 neuropil responses to task-informative orientations, as a function of distance from retinotopic position of the task stimulus, for trials with low and high convexity. Dashed lines are least-squares fits. Shading: SEM (n = 5 mice). *, p < 0.05, **, p < 0.01.

varied between orientations, but for each orientation, a single function fit the responses of cells in all tuning categories, as predicted by the model. Applying these functions to the naïve tuning curves, we were able to predict neuronal responses in proficient subjects with remarkable accuracy (**Figure 4C;** compare **Figure 2B**). Specifically, the model explains why learning affects mostly the cells that are broadly tuned: these cells exhibit intermediate levels of response that are affected most by the change in nonlinearity. The convexity of the nonlinear function g_θ relating naïve to proficient responses was larger for task-informative orientations than for the distractor orientation (**Figure 4D;** 45° vs 68°: p = 0.006; 68° vs 90°: p = 0.002; 45° vs 90°: p = 0.311. Independent samples *t*-test, n = 5 mice), and accurately predicted the increased sparseness of population responses to each stimulus (**Figure 4E**). These changes were local to the region of V1 representing the task stimulus location, where they affected neuropil as well as cellular activity, as might be expected from activation of local inhibitory cells (**Figure S5**).

The hypothesis that the sparseness of a population response to a stimulus depends on local or distal feedback makes a second prediction: if the strength of this feedback varies between trials, then the amount of sparsening should also vary between trials. Furthermore, since this signal modulates all neurons similarly, the degree of sparsening should be consistent across the population. Trial-to-trial variability in neuronal responses is well-documented, and has been reported to take additive and multiplicative forms (*23*–*25*). The current model predicts a different type of trial-to-trial variability: it predicts that responses follow a nonlinear transformation whose convexity varies from one trial to the next.

To test this second prediction, we examined population responses on single trials (**Figure 5**). We divided cells randomly into two groups, balanced for orientation preference and selectivity, and within each group examined how single-trial population activity was related to naïve trial-averaged responses. The convexity of the population response varied substantially between trials, even within repeats of a single stimulus orientation, but was consistent across cell groups (**Figure 5A-D;** correlation coefficient significantly exceeds 0 at $p < 0.05$ for each stimulus orientation, one sample *t*-test with Holm-Sidak correction, n = 5 mice.). Furthermore, suppression of activity on trials of high convexity was largest in areas of V1 topographically representing the task stimulus location, as would be expected if it were driven by local inhibitory neurons (**Figure 5E-F**).

Discussion

Although learning-related changes to orientation tuning curves were apparently complex, they could be explained to high quantitative accuracy by a simple principle: neuronal outputs on each trial reflect a nonlinear transformation of the mean naïve responses, whose convexity varies from trial to trial but is largest on average for task-informative orientations after learning. This convex transformation sparsens population responses to task-informative orientations and makes them more orthogonal to each other. This orthogonalization may help downstream circuits produce different behavioral responses to the two.

This model can explain many of the apparently complex effects of learning observed in previous studies of V1. It predicts a reduction in the number of cells responding modally to the task-informative orientations (*12*) and an asymmetrical increase in tuning curve slope specifically at these orientations (*13*). Additionally, nonlinear suppression of the task-informative orientations predicts an increase in the fraction of cells that are significantly selective between these stimuli (*9*, *10*), as confirmed by an increase in the d′ statistic. Thus, one simple principle can explain several apparently diverse results observed in visual cortex of multiple species.

Despite this concordance with previous results in visual cortex, our findings do not appear fully congruent with results from auditory and somatosensory cortex. Indeed, learning of multiple tasks, as well as stimulation of neuromodulatory systems under anesthesia, causes an increase in the number of electrophysiological recording sites responding modally to the task stimuli (*5*–*7*). We suggest three, non-exclusive, reasons for this apparent discrepancy. First, it would be surprising if there were only one mechanism by which cortical representations evolve with learning, and it is reasonable to expect that different mechanisms are employed to a different extent in different cortical regions and different tasks. In fact, one study of learning in somatosensory cortex did observe sparsening (*11*), suggesting that this mechanism is at least sometimes also employed in non-visual cortices. Second, methodological differences may explain at least some of the difference. Our study (like Ref. (*11*)) used two-photon imaging to record excitatory cells in superficial layers. Auditory and somatosensory studies have typically used electrophysiological multi-unit recordings, which are biased toward fast-spiking interneurons, and increased activity of these cells is one possible mechanism by which sparsening of pyramidal cell activity could occur. Finally, expansion of sites responding to task stimuli is a transient phenomenon. After continued training or stimulus exposure, expanded maps can "renormalize" to their original state without compromising behavioral performance (*26*); furthermore, induction of map expansion by means other than task training can actually worsen task performance (*27*), in particular by increasing the rate of false responses to non-target stimuli (*28*). Our task required long training, potentially allowing time for map expansion to reverse; it also requires differentially responding to the two informative stimuli while not responding to the similar distractor stimulus, for which map expansion might actually impair performance.

It is often assumed that plasticity of cortical representations arises from plasticity of excitatory inputs onto the cells being recorded. Our model suggests that this form of plasticity is not required to explain our results. Clearly, synaptic or cellular plasticity must occur somewhere to change the tuning curves; our model suggests that it occurs upstream of the circuit carrying the feedback signal. Several identities of this circuit are consistent with our data. Sparsening could be mediated by a class of local interneurons, whose inputs from local pyramidal cells tuned to task-informative stimuli are strengthened after learning (*29*, *30*). Alternatively, it could be mediated by feedback from more distal cortical regions or neuromodulators, which target local inhibitory circuits to cause retinotopically-aligned suppression. Although we did not observe any videographic correlates of cortical sparsening (such as increased pupil diameter or whisking), our data are not inconsistent with a covert cognitive state change such as an increase in attention caused by the task-informative orientations.

Regardless of the underlying mechanism, the fact that learning-related sparsening leads to orthogonalization of the representations of the task-informative stimuli suggests a function for this process. We suggest that orthogonalizing the representations of these stimuli allows the brain to produce different behavioral responses to them. Gratings are not natural stimuli, and if a mouse ever did encounter one in the wild, it seems unlikely that the grating's orientation would be of any behavioral significance. Thus, one might expect mice by default to generalize from one orientation of grating to another; only after extensive training should behavioral responses to them diverge. Orthogonalization of cortical representations of these stimuli may override this default generalization and encourage differing behavioral responses. Applications of similar techniques to artificial learning systems might provide a new mechanism to boost their learning capacity.

References

1. I. Goodfellow, Y. Bengio, A. Courville, F. Bach, *Deep Learning* (MIT Press, Cambridge, Massachusetts, 2017).

2. B. Schölkopf, A. J. Smola, F. Bach, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, Cambridge, Massachusetts, 2018).

3. H. Hong, D. L. K. Yamins, N. J. Majaj, J. J. DiCarlo, Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat. Neurosci.* **19**, 613–622 (2016).

4. N. J. Majaj, H. Hong, E. A. Solomon, J. J. DiCarlo, Simple Learned Weighted Sums of Inferior Temporal Neuronal Firing Rates Accurately Predict Human Core Object Recognition Performance. *J. Neurosci.* **35**, 13402–13418 (2015).

5. D. V. Buonomano, M. M. Merzenich, Cortical plasticity: from synapses to maps. *Annu.Rev.Neurosci.* **21**, 149–186 (1998).

6. N. M. Weinberger, Specific long-term memory traces in primary auditory cortex. *Nat.Rev.Neurosci.* **5**, 279–290 (2004).

7. D. E. Feldman, M. Brecht, Map plasticity in somatosensory cortex. *Science*. **310**, 810–5 (2005).

8. G. H. Recanzone, M. M. Merzenich, W. M. Jenkins, K. A. Grajski, H. R. Dinse, Topographic reorganization of the hand representation in cortical area 3b owl monkeys trained in a frequency-discrimination task. *J. Neurophysiol.* **67**, 1031–1056 (1992).

9. J. Poort, A. G. Khan, M. Pachitariu, A. Nemri, I. Orsolic, J. Krupic, M. Bauza, M. Sahani, G. B. Keller, T. D. Mrsic-Flogel, S. B. Hofer,

Learning Enhances Sensory and Multiple Non-sensory Representations in Primary Visual Cortex. *Neuron*. **86**, 1478–1490 (2015).

10. J. Poort, K. A. Wilmes, A. Blot, A. Chadwick, M. Sahani, C. Clopath, T. D. Mrsic-Flogel, S. B. Hofer, A. G. Khan, Learning and attention increase visual response selectivity through distinct mechanisms. *bioRxiv* (2021), doi:10.1101/2021.01.31.429053.

11. A. Gdalyahu, E. Tring, P.-O. Polack, R. Gruver, P. Golshani, M. S. Fanselow, A. J. Silva, J. T. Trachtenberg, Associative Fear Learning Enhances Sparse Network Coding in Primary Sensory Cortex. *Neuron*. **75**, 121–132 (2012).

12. G. M. Ghose, T. Yang, J. H. R. Maunsell, Physiological Correlates of Perceptual Learning in Monkey V1 and V2. *J. Neurophysiol.* **87**, 1867–1888 (2002).

13. A. Schoups, R. Vogels, N. Qian, G. Orban, Practising orientation identification improves orientation coding in V1 neurons. *Nature*. **412**, 549–553 (2001).

14. C. Stringer, M. Pachitariu, N. Steinmetz, C. B. Reddy, M. Carandini, K. D. Harris, Spontaneous behaviors drive multidimensional, brainwide activity. *Science*. **364**, 255 (2019).

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

16. C. M. Niell, M. P. Stryker, Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*. **65**, 472–9 (2010).

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

18. R. Moreno-Bote, J. Beck, I. Kanitscheider, X. Pitkow, P. Latham, A. Pouget, Information-limiting correlations. *Nat. Neurosci.* **17**, 1410–1417 (2014).

19. O. I. Rumyantsev, J. A. Lecoq, O. Hernandez, Y. Zhang, J. Savall, R. Chrapkiewicz, J. Li, H. Zeng, S. Ganguli, M. J. Schnitzer, Fundamental bounds on the fidelity of sensory cortical coding. *Nature*. **580**, 100–105 (2020).

20. C. Stringer, M. Michaelos, D. Tsyboulski, S. E. Lindo, M. Pachitariu, High-precision coding in visual cortex. *Cell.* **184,** 2767-2778.e15 (2021).

21. S. F. Cooke, M. F. Bear, How the mechanisms of long-term synaptic potentiation and depression serve experience-dependent plasticity in primary visual cortex. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **369**, 20130284 (2014).

22. L. N. Cooper, N. Intrator, B. S. Blais, H. Z. Shouval, *Theory of cortical plasticity* (World Scientific, New Jersey, 2004).

23. A. Arieli, A. Sterkin, A. Grinvald, A. Aertsen, Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science*. **273**, 1868–71 (1996).

24. R. L. Goris, J. A. Movshon, E. P. Simoncelli, Partitioning neuronal variability. *Nat Neurosci*. **17**, 858–65 (2014).

25. I. C. Lin, M. Okun, M. Carandini, K. D. Harris, The Nature of Shared Cortical Variability. *Neuron*. **87**, 644–56 (2015).

26. A. Reed, J. Riley, R. Carraway, A. Carrasco, C. Perez, V. Jakkamsetti, M. P. Kilgard, Cortical Map Plasticity Improves Learning but Is Not Necessary for Improved Performance. *Neuron*. **70**, 121–131 (2011).

27. Y. K. Han, H. Köver, M. N. Insanally, J. H. Semerdjian, S. Bao, Early experience impairs perceptual discrimination. *Nat. Neurosci.* **10**, 1191–1197 (2007).

28. M. E. Thomas, C. P. Lane, Y. M. J. Chaudron, J. M. Cisneros-Franco, É. de Villers-Sidani, Modifying the Adult Rat Tonotopic Map with Sound Exposure Produces Frequency Discrimination Deficits That Are Recovered with Training. *J. Neurosci.* **40**, 2259–2268 (2020).

29. A. G. Khan, J. Poort, A. Chadwick, A. Blot, M. Sahani, T. D. Mrsic-Flogel, S. B. Hofer, Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. *Nat. Neurosci.* **21**, 851–859 (2018).

30. N. M. Bannon, M. Chistiakova, M. Volgushev, Synaptic Plasticity in Cortical Inhibitory Neurons: What Mechanisms May Help

to Balance Synaptic Weight Changes? *Front. Cell. Neurosci.* **14**, 204 (2020).

Acknowledgments

We thank Charu Reddy for assistance with mouse surgeries, Miles Wells, Hamish Forest, and Laura Funnel for assistance with task training, Michael Krumin for technical assistance with two-photon calcium imaging.

Funding

This work was supported by the Wellcome Trust (grants 205093, 108726 to MC and KDH).

Contributions

	Samuel W. Failor	Matteo Carandini	Kenneth D. Harris
Conceptualization	X	X	X
Methodology	X	X	X
Investigation	X		
Data curation	X		
Formal analysis	X		X
Funding acquisition		X	X
Project administration	X		
Supervision		X	X
Visualization	X		X
Writing	X	X	X

Competing interests

The authors have no competing interests to declare.

Supplemental Materials

Materials and Methods

Experimental procedures

All experimental procedures were conducted according to the UK Animals Scientific Procedures Act (1986). Experiments were performed at University College London under personal and project licenses released by the Home Office following appropriate ethics review.

Surgical procedure

Five transgenic adult mice (60 days or older) expressing GCaMP6s in excitatory neurons (CaMK2a-tTA;tetO-GCaMP6s) underwent a procedure to implant cortical windows over right primary visual cortex (V1). Mice were anesthetized with isoflurane, an ophthalmic ointment was applied to the eyes, and injections of carprofen and dexamethasone were administered. The hair on the head at the planned incision site was shaved away, and the mouse was transferred to a stereotaxic apparatus where its skull was secured with ear bars. The scalp was cleaned with 70% ethanol to remove loose hairs and other detritus, after which a lidocaine ointment was applied. Following a final application of iodine and ethanol, the scalp over visual cortex was excised, and the edges of the incision were sealed to the skull with a cyanoacrylate adhesive. A sterilized metal head plate with a circular well was cemented onto the skull using dental acrylic resin. A 4 mm circular craniotomy was made over right V1 using a biopsy punch, and a glass window was sealed in place with a cyanoacrylate adhesive and dental acrylic resin. At the end of the procedure, mice were removed from anesthesia and placed on a heating pad to recover. Carprofen was added to the mice's drinking water for three days following surgery to mitigate post-operative pain, and mice were checked daily for any adverse outcomes.

Following recovery, mice were habituated for handling and head-fixation before carrying out recordings.

An orientation discrimination task

The task is a modification of a two-alternative forced choice contrast discrimination task previously developed by our lab (*31*). Mice were head-fixed with their body and hindlimbs resting on a stage, leaving their front forepaws free to turn a small wheel left or right. Three computer screens surrounded the mouse, spanning -135 to +135 visual degrees (v°) along the azimuth axis and -35 to +35 v° along the elevation axis. Trials began after 2 s of continuous quiescence (no wheel movement), after which two full contrast Gabors with sigmas of 18 v° and spatial frequencies of 0.04 cycles/v° were presented simultaneously and centered at -80 and +80 v° azimuth. These Gabors were randomly oriented at either 45°, 68°, or 90°, though the pair were never identical. After an additional quiescence period of approximately 1 s, an auditory cue (12 kHz, 100 ms) would sound, signaling to the mouse that the horizontal position of the Gabors could be manipulated via wheel movement. If the mouse moved the wheel before the auditory cue, the Gabors remained stationary while the quiescence requirement remained in force. When a Gabor was moved to the center screen, a choice was recorded for that trial, and a feedback period was initiated. Correct choices (driving a 45° stimulus to the center, or a 90° stimulus away) were rewarded with 1 - 5 µl of water and a short 0.25 s delay, while incorrect choices (driving a 90° stimulus to the center, or a 45° stimulus away) resulted in a 1 - 2 s burst of white noise.

The Gabor was locked at the center position during the feedback, following which it would disappear, and the next pre-trial period of enforced quiescence would begin. During task training, mice were water restricted in line with the approved project license. Mice were considered proficient at the task when they consistently made the correct choice on over 70% of trials.

Recording visual responses in V1

Two sessions of two-photon calcium imaging were performed: one before task training (naïve) and one after mice had achieved high performance in the task (proficient). Imaging in the proficient condition was performed immediately after a behavioral session and in the same apparatus.

Location of visual areas

Prior to the first two-photon imaging session, we determined the location of V1 in each mouse's cortical window by recording cortical responses to sparse noise under mesoscopic wide-field calcium imaging and then generating a visual sign map, as previously described (*32*). Mice were placed on a stage of the same type used in the task, and white squares of width 7.5° visual angle were shown on a black background at a frame rate of 6 Hz for 10 minutes. Squares appeared randomly at fixed positions in a 12 by 36 grid, spanning the retinotopic range of the computer screens. 12% of the squares shown at any one time.

Two-photon calcium imaging

Layer 2/3 in V1 was imaged using a commercial two-photon microscope (Bergamo II, Thorlabs Inc) controlled by ScanImage (*33*). A ti:sapphire laser (Chameleon Vision, Coherent) was set to a wavelength between 940 and 980 nm, and the beam was focused with a 16X water-immersion objective (0.8 NA, Nikon). Images were acquired at a frequency of 30 Hz across six planes (5 Hz per plane), a resolution of 512 x 512 pixels, with a frame width between 730 and 810 µm. The fly-back plane was excluded from further analysis. During recordings, mice were head-fixed and placed on the same type of stage used for the task. Three computer screens surrounded the mouse, spanning -135 to +135 v° along the azimuth axis and -35 to +35 v° along the elevation axis.

Sparse noise

To map the retinotopy of V1 under two-photon imaging (Fig. 1C, middle), sparse noise stimuli were again presented. Black or white squares of width 4.5° visual angle were shown on a gray background at a frame rate of 5 Hz for 8 – 30 minutes. Squares appeared randomly at fixed positions in a 16 by 60 grid, spanning the retinotopic range of the computer screens. 1.5% of the squares were shown at any one time.

Drifting gratings

At least 16 blocks of drifting grating stimuli were presented in each recording. In each block, gratings spanning 16 directions (22.5° intervals) and a blank stimulus were each presented once in a randomized sequence. Each grating lasted 2 s, with an inter-trial interval sampled randomly from a uniform distribution with a range of 2 – 3 s. Drifting gratings were full contrast and sinusoidal, with a spatial frequency of 0.04 cycles/v° and a temporal frequency of 4 cycles/s, that either encompassed all three screens (full-field, three mice) or the entire left screen (two mice), contralateral to the recorded hemisphere. Data from the two directions for each of the eight orientations covering 180° were analyzed together.

Face recording

An infrared LED illuminated the mouse's face, and a camera with an infrared filter was used to capture any changes in pupil area or whisking behavior.

Data analysis

Pixel map of retinotopy

To obtain a retinotopic map of the two-photon imaging frame (Fig. 1C middle, Fig. S4A), we analyzed the two-photon recordings during sparse noise stimuli on a pixel-by-pixel basis, without cell detection. To accelerate the computation and denoise the data, analyses were performed after singular value decomposition (SVD), which produces valid results as these computations are linear. First, we z-scored each pixel's time course independently. Next, we applied single-value decomposition (SVD) on the z-scored image frames, $F = USV^T$, where F was the full movie encoded as a matrix of size $N_{pixels} \times T$, U was size $N_{pixels} \times N_{SVDs}$, S was a diagonal matrix of singular values, and V was size $T \times N_{SVDs}$ with T being the number of two-photon imaging frames. A matrix Y was computed summarizing the mean response of each of the first 100 columns of V to each noise frame, as the time-averaged activity in a window 0.2 to 0.6 s after stimulus onset minus the time-averaged activity in a 1 s pre-stimulus window. This matrix was of size $F \times 100$, where F is the number of noise stimulus frames. The dependence of these responses on individual noise pixels was estimated using ridge regression: $\beta = (X^T X + \lambda I)^{-1} X^T Y$, where X was a $F \times N_{noise\_squares}$ matrix containing 1 if a particular square was white or black on a particular frame (0 if it was grey), λ was a ridge parameter ($\lambda = 100$), and I was the identity matrix. The stimulus dependence of each pixel was then obtained by matrix multiplication $R = US\beta$, resulting in a matrix R of size $N_{pixels} \times N_{noise\_squares}$, encoding the receptive field map of each 2p imaging pixel. To generate retinotopic maps of the imaging frame, each pixel's receptive field map was smoothed with a Gaussian (sigma 12 v°) and a peak found, giving retinotopic positions along the elevation and azimuth axes for each pixel.

Pixel retinotopy maps were used to ensure that the two-photon imaging frames were retinotopically aligned with the position of the left task stimulus (0 v° elevation, -80 v° azimuth) during drifting grating recordings. When the optimal imaging location in V1 was identified in naïve mice, an image of the cortical vasculature was saved for positioning subsequent imaging experiments.

Visual sign maps

Due to the retinotopic eccentricity of the imaging location in V1 and the large field of view used, it was occasionally the case that areas outside V1 were also recorded. To differentiate V1 from adjacent visual areas, visual sign maps were obtained using the above pixel retinotopy maps averaged across planes (Fig. S4). First, elevation and azimuth maps were smoothed with a median (width 10 pixels) and a Gaussian (sigma 60 pixels) filter. Similar to the process described in Ref. (*34*), the sine of the difference in angle between the gradients of the elevation and azimuth maps was calculated. This sign map was then thresholded to values above 0.31, and pixels that were members of the largest patch were considered to be in V1. This process was consistent in isolating V1, as verified by visual inspection of the elevation and azimuth retinotopic maps.

Pixel map of orientation responses

To obtain a pixel map of orientation preference (Fig. S4), the average df/f of each pixel was calculated in response to each stimulus orientation. For each trial, df was defined as the average fluorescence in a post-stimulus window spanning 0 – 2 s, minus the baseline defined as the average fluorescence in a pre-stimulus window spanning -1 to 0 s relative to stimulus onset. This value was divided by f_0, the baseline measurement. To isolate neuropil responses (Fig. S4D), only pixels that did not belong to a cell, as determined by Suite2P and subsequent manual curation, were included in the analysis.

Cell detection

Registration, cell detection, neuropil correction, and deconvolution of the two-photon imaging data were carried out using Suite2P (*35*). Imaged planes were aligned with non-rigid registration (four blocks, 128 x 128), and spiking activity was deconvolved from calcium fluorescence using a kernel with a timescale of 2 s.

Characterizing single-cell orientation tuning

All cells identified by Suite2P were analyzed for orientation responses. First, each cell's trial responses were computed by time-averaging its deconvolved activity on each trial over a window of width 0 - 2 s from drifting grating onset. Next, the mean response of each cell to each orientation and to the blank stimulus was computed by averaging over the respective stimulus trials. Each cell's trial responses were then normalized by dividing by its mean response to its preferred stimulus condition.

A cell's orientation preference was defined in two ways: the orientation it responded maximally to (preferred modal orientation; Fig. 1E-F) or its preferred mean orientation, the argument of the complex number $z = \frac{\sum_\theta r_\theta e^{2i\theta}}{\sum_\theta r_\theta}$, where r_θ is the cell's mean response to orientation θ. The orientation selectivity of a cell was defined as the modulus of z. To determine the tuning curve of each cell as a function of its orientation preference and selectivity (Fig. 2A-B), a cross-validated approach was used to avoid erroneously detecting tuning due to random fluctuations in responses. The preferred mean orientation and selectivity of each cell were calculated using odd-numbered trials, while the tuning curves were generated using the mean response to each orientation on even-numbered trials.

Tuning curve slope (Fig. S2A) was quantified as the absolute difference between the cell's response at a stimulus orientation, and the orientation 22.5° closer to the cell's preferred mean orientation, divided by 22.5. The cell's tuning curve slope at its preferred mean orientation was defined as the absolute difference between orientations -22.5° or +22.5° from preferred, divided by 45. Thus, in cases where these responses were equal, the tuning curve slope at the preferred orientation was zero.

Discriminability index

The discriminability index (d') of a cell, its ability to discriminate between two orientations (θ_a and θ_b), was defined as $\frac{\mu_{\theta_a} - \mu_{\theta_b}}{\sqrt{\frac{\sigma^2_{\theta_a} + \sigma^2_{\theta_b}}{2}}}$ where μ and σ^2 are the mean and variance of the respective orientation responses. The mean and variance for each stimulus orientation was the average of the mean and variance of the two corresponding stimulus directions.

Population sparseness

Population sparseness was summarized as the kurtosis of the mean population response to each orientation, i.e., $k = \frac{\mu_4}{\sigma^4}$, where μ_4 is the fourth central moment and σ is the standard deviation of mean orientation cell responses (*36*).

Orthogonalization of population responses

To calculate the orthogonalization of population responses between different stimulus orientations (Fig. 3), we split the trials into odd and even halves, and computed the N_{cells}-dimensional population response vectors $\boldsymbol{P}_i(\theta)$ to orientation θ for the trial set i (i = 1: odd trials; i = 2: even trials). We computed the cosine similarity between orientations θ_1 and θ_2 as $\frac{\boldsymbol{P}_1(\theta_1) \cdot \boldsymbol{P}_2(\theta_2)}{\|\boldsymbol{P}_1(\theta_1)\| \|\boldsymbol{P}_2(\theta_2)\|}$. This process resulted in an eight-by-eight matrix of similarity values for each mouse and training condition. Computing this similarity between two separate halves ensured that the diagonal was not 1 by definition.

Dimensionality reduction

To display population responses in a 2-dimensional plot (Fig. 2C), we trained a linear regression model to predict a 2-dimensional vector $(\cos\theta, \sin\theta)$ for each trial, where θ is the stimulus orientation, from the N_{cells}-dimensional population response vector on that trial. The model was trained on odd trials, and then applied to population responses on even trials to obtain a two-dimensional projection of population activity that separates points by stimulus orientation.

Stimulus prediction

Orientation was also decoded from population activity using linear discriminant analysis (LDA; Fig. 2D). An LDA model was fit using the population responses in odd trials, and its performance was assessed on even trials. To build the model, we used the class *LinearDiscriminantAnalysis* from the Python library scikit-learn, with solver set to "eigen" and the shrinkage coefficient automatically calculated.

Modeling learning-evoked changes to orientation responses

For each mouse, cells in the naïve and proficient recordings were divided into classes by binning mean orientation preference (eight bins, *0°*: 168.75 – 11.25°, *23°*: 11.25 – 33.75°, *45°*: 33.75 – 56.25°, *68°*: 56.25 – 78.75°, *90°*: 78.75 – 101.25°, *113°*: 101.25 – 123.75°, *135°*: 123.75 – 146.25°, *158°*: 146.25 – 168.75°) and selectivity (five bins, 0 – 0.16, 0.16 – 0.32, 0.32 – 0.48, 0.48 – 0.64, 0.64 – 1). The mean response of each cell class to each stimulus was determined by cross-validation, using odd trials to determine the cell's tuning class, and using even trials to compute its tuning, as described above. Responses in the proficient mice were fit by piecewise linear functions of responses in naïve mice, $r_p = f_{a,b}(r_n)$, where

$$f_{a,b}(x) = \begin{cases} xb/a, & r_n \leq a \\ (x-1)\dfrac{b-1}{a-1} + 1, & r_n > a \end{cases}$$

The function $f_{a,b}$ is the piecewise linear function constrained to pass through (0,0), (a, b), and (1,1). The parameters a and b were fit for each mouse and stimulus by nonlinear least squares (Python library SciPy, *optimize.curve_fit*), constrained to values between 0 and 1.

The convexity of the transformation from naïve to proficient population responses to a stimulus was quantified as $C = \frac{m_{pref}}{m_{non-perf}} - 1$, where m_{pref} was the slope of a line from the origin to the point representing the cell class with the strongest selectivity to this stimulus, and $m_{non-pref}$ was the slope of a linear regression on the points corresponding to cell classes whose mean orientation preference was not the stimulus shown. This approach was used to measure convexity on mean responses, relating the trial-averaged population response in the same mouse prior and after training (Fig. 4D), and on single trials (Fig. 5), where the population responses in single trial in a proficient mouse was compared to the trial-averaged population response in that mouse prior to training (Fig. 5).

To assess the consistency of trial-to-trial fluctuations in sparsening across the population (Fig. 5C-D), we randomly divided the proficient cells into two populations balanced for orientation preference and selectivity. Trial-by-trial convexity was measured, as described above, for each cell population, and the correlation coefficient of these convexities was computed. This process was repeated 2000 times, and the average correlation in convexity over orientations was found for each mouse.

Pupil area and whisking

Facial recordings were processed with the toolkit FaceMap (www.github.com/MouseLand/FaceMap) to obtain traces of pupil area and whisking intensity. The pupil area was defined as the area of a Gaussian fit on thresholded pupil frames, where pixels outside the pupil were set to zero. Whisking intensity was defined as the average change in individual pixels between frames for a region of interest limited to the whisker pad. From these resulting traces, trial-evoked changes in pupil area and whisking were calculated. First, for each trial pupil area and whisking were averaged in a post-stimulus time windows spanning 0.5 to 3 s for pupil and 0 to 3 s for whisking. Next, to compare across sessions, pupil and whisking trials were normalized by the blank stimulus trial average. Lastly, stimulus-evoked changes in pupil area and whisking were calculated by subtracting from the normalized trials a pre-stimulus baseline, defined as the average normalized pupil area and whisking in a -1 to 0 s window.

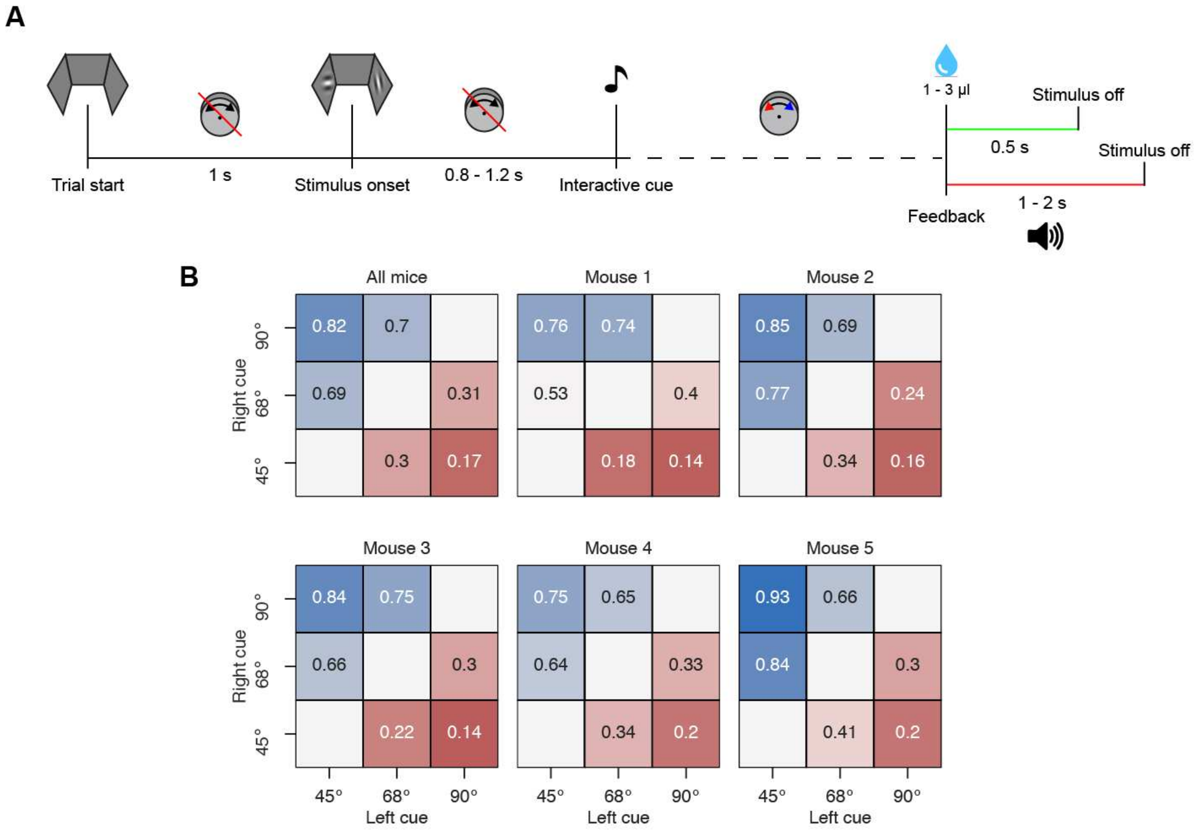

Figure S1. Orientation discrimination task. (**A**) Temporal structure of the task. **(B)** Behavioral performance for all mice. Matrices show the proportion of left choices for all cue pairings averaged over ten highest performing sessions. Cue pairings that were not presented are shown in white.

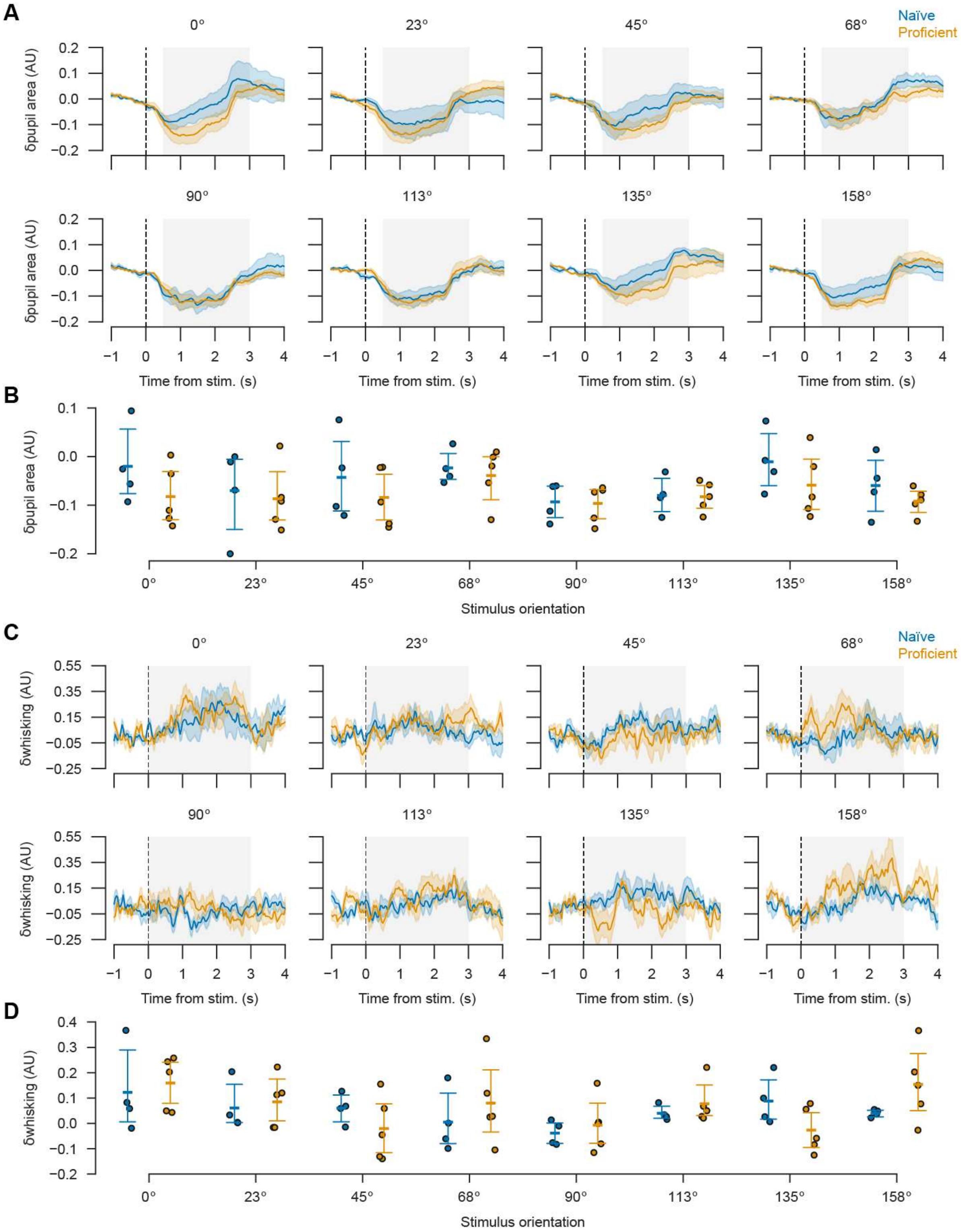

Figure S2. Measures of behavioral responses during passive viewing of grating stimuli. (**A**) Stimulus-triggered pupil area time course, averaged over all trials of each stimulus orientation and training condition. Stimulus presentation causes pupil constriction, but pupil responses to task-informative stimuli do not appear substantially different to those to other stimuli. Shaded regions: SEM (n = 5 mice). (**B**) Average change in pupil area within gray shaded time windows shown in (A). ANOVA indicated no significant effect of training (p = 0.053), stimulus orientation (p = 0.279), or their interaction (p = 0.951). Error bars: mean and SEM (n = 5 mice). (**C** and **D**) Same as in (A and B) but for whisking, assessed by video motion energy over the whisker pad. ANOVA indicated no significant effect of training (p = 0.547), stimulus orientation (p = 0.061), or their interaction (p = 0.372).

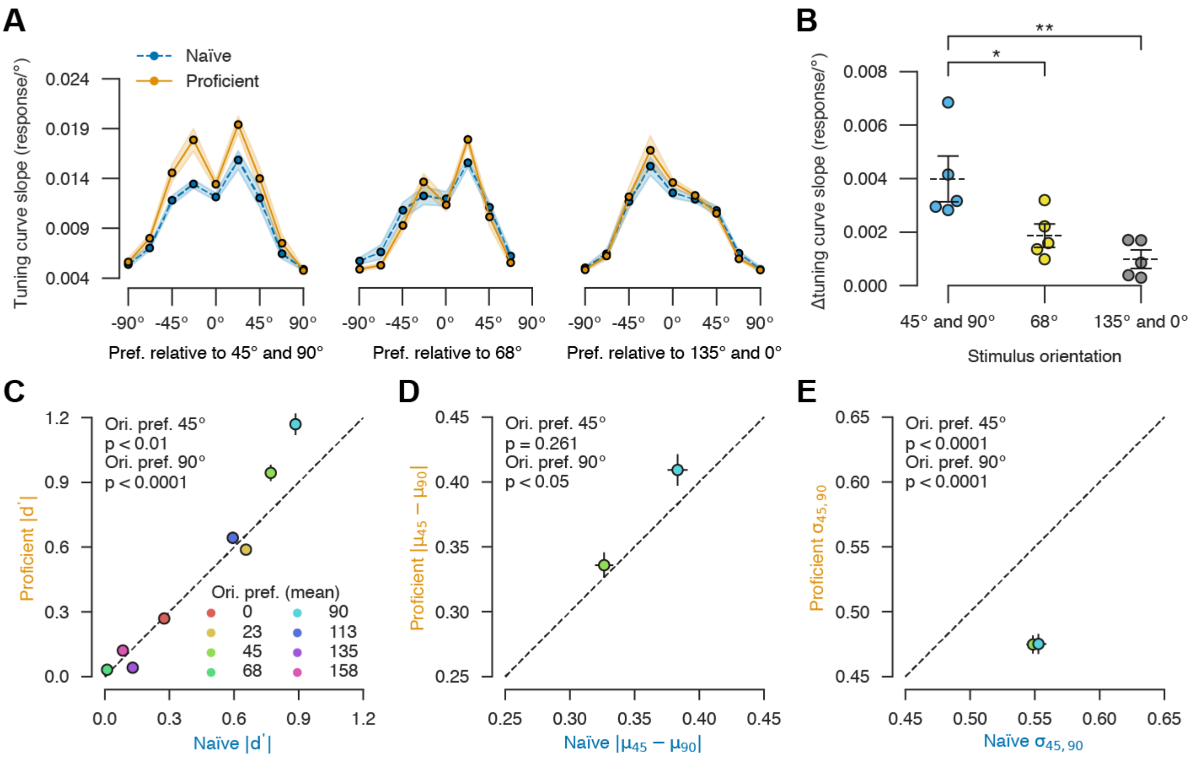

Figure S3. Additional metrics of single-cell tuning. (**A**) Tuning curve slope as a function of mean orientation preference relative to the informative task orientations (45° and 90°; left), uninformative distractor orientation (68°, center), and non-task orientation controls (135° and 0°; right). Shading: SEM (n = 5 mice). Note that the slope increases with training specifically for stimuli adjacent to task-informative stimuli (*13*). (**B**) Change in tuning curve slope at the informative, distractor, and control orientations for cells with adjacent orientation preferences. Comparisons: 45° and 90° vs 68°, p = 0.036; 45° and 90° vs 135° and 0°, p = 0.0006. Independent samples *t*-test. Error bars: mean and SEM (n = 5 mice). (**C**) Learning-evoked changes in the d' statistic, measuring the discriminability of 45° from 90° stimuli in individual cells. Each point shows the absolute value of the average d' across cells of a single mean orientation preference (color) pooled across all mice. The d' magnitude increases significantly for cells preferring the task-informative orientations 45° and 90° (p = 0.010, p = 5.6 x 10^{-8}, Linear mixed effects model with random intercept). (**D**) Absolute value of average difference of mean responses to 45° and 90°, for 45° and 90° preferring cells, before and after training (p = 0.261, p = 0.042, Linear mixed effects model with random intercept.) (**E**) Average standard deviation of responses to 45° and 90° for 45° and 90° preferring cells, before and after training. Learning reduced the standard deviation of responses to 45° and 90°, contributing to an increase in d' (p = 2.7 x 10^{-18}, p = 1.8 x 10^{-14}, Linear mixed effects model with random intercept). Error bars: SEM (n = 5 mice, naïve cells = 20,857, proficient cells = 16,876). *, p < 0.05, **, p < 0.01.

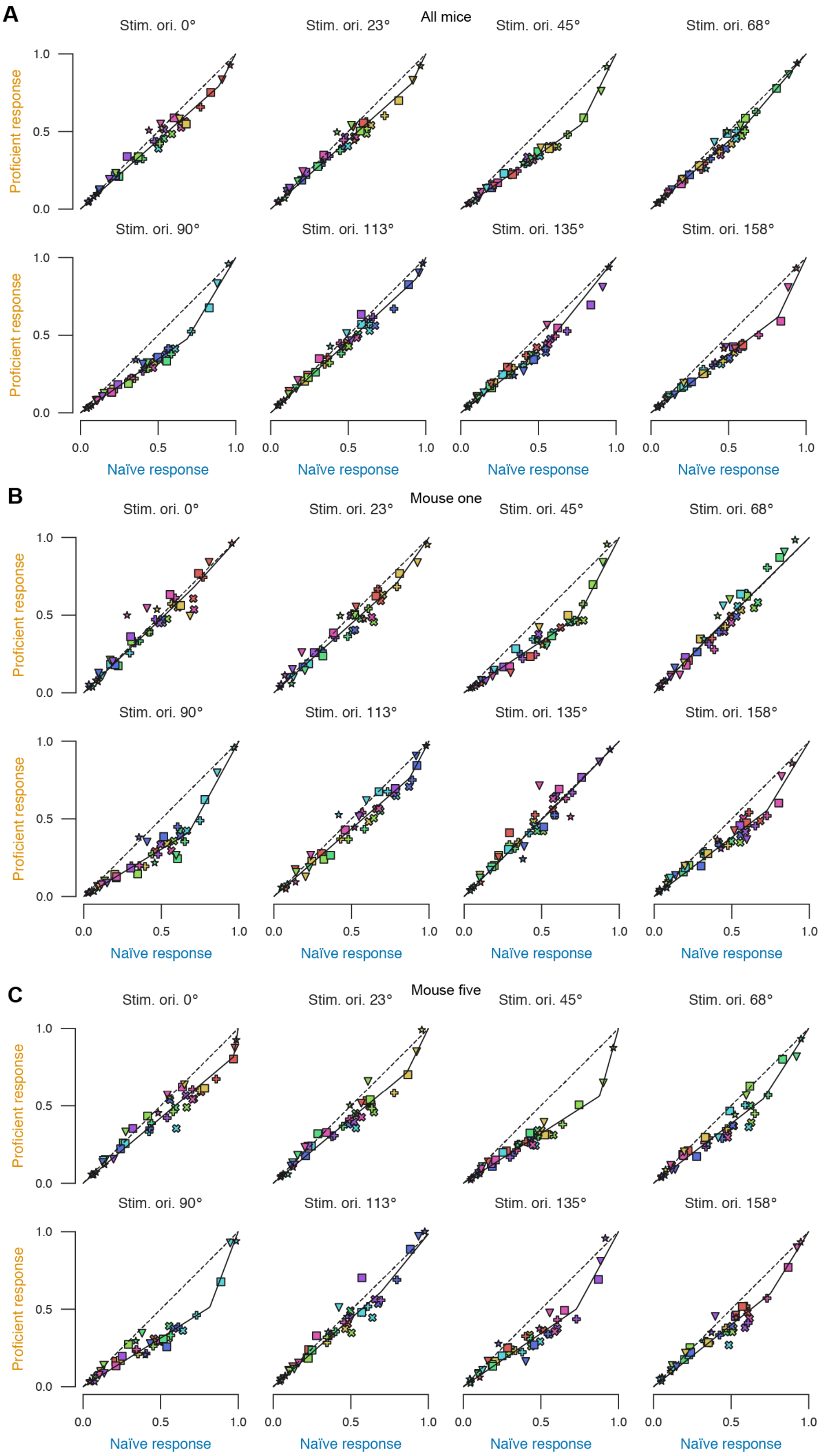

Figure S4. Naïve and proficient population responses to all stimulus orientations. (**A**) Learning-evoked changes in orientation responses averaged over all mice, like figure 4B but showing analyses for all stimulus orientations. (**B** and **C**) Same as (A) but for two representative mice individually.

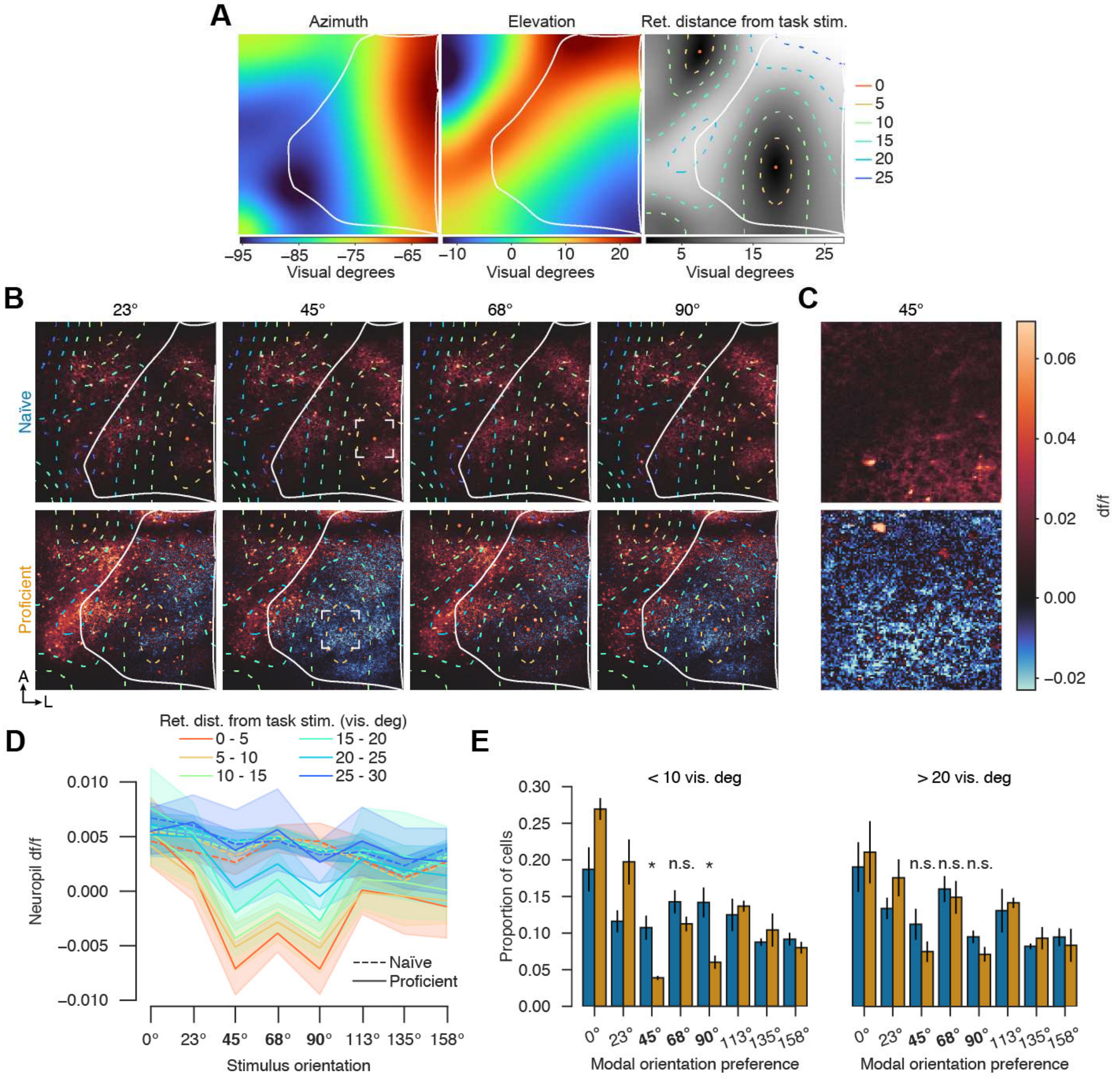

Figure S5. Response suppression is aligned with the retinotopic location of the task stimulus. (**A**) Retinotopic mapping of visual cortex, for an example mouse. Left two pseudocolor plots show preferred azimuth and elevation for each pixel in the field of view, assessed by analyzing responses to sparse noise stimuli. White line demarcates the border of V1. Right panel shows distance in degrees of visual angle from each pixel's preferred retinotopic location to the retinotopic position of the task stimulus, in pseudocolor (grayscale), and with contour representation (dashed colored lines). (**B**) Mean df/f of two-photon imaging frames during presentation of full-field gratings of the marked orientations in the same mouse prior to (top) and after training (bottom). White lines and colored contours mark V1 boundary and retinotopic distance to stimulus location, as in (A). (**C**) Zoom into boxed regions in (B). Note that after training, neuropil is suppressed in the region retinotopically matching the stimulus, although individual cells continue to respond strongly there. (**D**) V1 neuropil responses as a function of stimulus orientation and retinotopic distance from the task stimulus position (colors), for naïve and proficient mice (dashed and solid lines). Shading: SEM (n = 5 mice). Note specific suppression of responses to task orientations in pixels retinotopically close to the stimulus location. (**E**) Histogram of modal orientation preferences of V1 cells in naïve and proficient mice, for cells close to (left) and distant from (right) the retinotopic position of the task stimulus, plotted as in Figure 1G. The proportion of cells preferring 45° and 90° but not 68° changes significantly amongst cells within 10 v° of the task stimulus location (p = 0.020, p = 0.045, p = 0.121, paired samples *t*-test). For cells further than 20 v° from the task stimulus location, all three changes are insignificant (p = 0.206, p = 0.132, p = 0.762, paired samples *t*-test). Error bars: SEM (n = 5 mice). *, p < 0.05.

Appendix

Here we prove that applying a convex transformation to a neural population response vector increases its sparseness. Intuitively, the argument works as follows. Sparseness measures the degree to which a small number of neurons fire more than the mean firing rate. Applying a convex transformation causes a disproportionate boost in the firing rate of these few highly active neurons, increasing the sparseness of the population response.

Formally, we will prove that this holds for a wide family of sparseness metrics, which includes those described by Treves and Rolls and Willmore and Tolhurst (*36*, *37*) as a special case corresponding to $k(x) = x^2$.

Theorem. *Let $k(x)$ be a convex function. Let $\{x_i : i = 1 \dots N\}$ be a finite set of non-negative real numbers. We define the sparseness measure*

$$S_k[x_i] = \sum_{i=1}^{N} k\left(\frac{x_i}{\bar{x}}\right),$$

where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$. Let g be a convex non-decreasing function with $g(0) = 0$, and write $y_i = g(x_i)$. Then

$$S_k[y_i] \geq S_k[x_i].$$

Proof. It is clear that for any scalar α, $S_k[x_i] = S_k[\alpha x_i]$. So without loss of generality, we can rescale x and g so that $\bar{x} = 1$ and $\bar{y} = 1$. After this rescaling,

$$S_k[y_i] - S_k[x_i] = \sum_{i=1}^{N} k(y_i) - k(x_i)$$

Now because $\sum_i x_i = \sum_i g(x_i)$, and g is continuous, there must exist an x_0 with $g(x_0) = x_0$. Because g is convex and $g(0) = 0$, $x_i \geq x_0$ implies $y_i \geq x_i$, and $x_i \leq x_0$ implies $y_i \leq x_i$. Let d be a subgradient of k at x_0, so if either $a \geq b \geq x_0$ or $a \leq b \leq x_0$, then $k(a) - k(b) \geq d(a - b)$. If $x_i \geq x_0$ then $y_i \geq x_i \geq x_0$ and if $x_i \leq x_0$ then $y_i \leq x_i \leq x_0$. For all i one of these two conditions is true so $k(y_i) - k(x_i) \geq d(y_i - x_i)$. Thus $S_k[y_i] - S_k[x_i] = \sum_{i=1}^{N} k(y_i) - k(x_i) \geq d \sum_i y_i - x_i = 0$, as we have rescaled so that $\sum_i x_i = \sum_i y_i$. So $S_k[y_i] \geq S_k[x_i]$ and the theorem is proved.

References

31. C. P. Burgess, A. Lak, N. A. Steinmetz, P. Zatka-Haas, C. Bai Reddy, E. A. K. Jacobs, J. F. Linden, J. J. Paton, A. Ranson, S. Schröder, S. Soares, M. J. Wells, L. E. Wool, K. D. Harris, M. Carandini, High-Yield Methods for Accurate Two-Alternative Visual Psychophysics in Head-Fixed Mice. *Cell Rep.* **20**, 2513–2524 (2017).

32. A. J. Peters, J. M. J. Fabre, N. A. Steinmetz, K. D. Harris, M. Carandini, Striatal activity topographically reflects cortical activity. *Nature*. **591**, 420–425 (2021).

33. T. A. Pologruto, B. L. Sabatini, K. Svoboda, ScanImage: Flexible software for operating laser scanning microscopes. *Biomed. Eng. OnLine*. **2**, 13 (2003).

34. M. I. Sereno, C. T. McDonald, J. M. Allman, Analysis of Retinotopic Maps in Extrastriate Cortex. *Cereb. Cortex*. **4**, 601–620 (1994).

35. M. Pachitariu, C. Stringer, M. Dipoppa, S. Schröder, L. F. Rossi, H. Dalgleish, M. Carandini, K. D. Harris, Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv*, 061507 (2017).

36. B. Willmore, D. J. Tolhurst, Characterizing the sparseness of neural codes. *Netw. Bristol Engl.* **12**, 255–270 (2001).

37. A. Treves, E. T. Rolls, What determines the capacity of autoassociative memories in the brain? *Netw. Comput. Neural Syst.* **2**, 371–397 (1991).