Visuomotor association orthogonalizes visual cortical population codes

Samuel W. Failor[a*], Matteo Carandini[b], Kenneth D. Harris[a]
[a]UCL Queen Square Institute of Neurology, University College London, London, United Kingdom
[b]UCL Institute of Ophthalmology, University College London, London, United Kingdom
*Correspondence: s.failor@ucl.ac.uk

In principle, the brain should be best able to associate distinct behavioral responses to two sensory stimuli when these stimuli evoke sensory population response vectors that are close to orthogonal. To investigate whether task training orthogonalizes the population code in primary visual cortex (V1), we measured the orientation tuning of 4,000-neuron populations in mouse V1 before and after training on a visuomotor association task. In the task, two orientations were associated with opposite behavioral responses, while a third was a distractor. The effect of task training on population activity could be captured by a simple mathematical transformation of firing rates, which suppressed responses to the motor-associated stimuli specifically in cells responding to them at intermediate levels. This orthogonalized the representations of the task orientations by sparsening the population responses to these stimuli. The degree of response transformation varied from trial to trial, suggesting a dynamic circuit mechanism rather than static synaptic plasticity. These results indicate a simple process by which visuomotor associations orthogonalize population codes as early as in primary visual cortex.

When an animal sees a stimulus, the stimulus triggers a pattern of activity across a multitude of neurons in the 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[1–3]. Substantial evidence suggests that task training can affect these visual cortical representations, and that these changes persist even when the stimuli are presented outside of the task context[4–17]. Nevertheless, these previous results together paint a somewhat confusing picture, with some studies suggesting increases and others decreases in the numbers of neurons representing task stimuli, and some studies suggesting broadening and others sharpening of tuning curves. Representational plasticity has been analyzed primarily at the level of single cells rather than populations, making it potentially sensitive to the exact methods to select cells for analysis and to quantify their selectivity. If it were possible to mathematically summarize the effects of task training on full population responses, this could help summarize these diverse effects, and thus help reveal their biological function.

One hypothesis for the function of cortical representational plasticity is to facilitate learning of appropriate sensory-motor associations by downstream motor systems. All learning systems exhibit "inductive biases", meaning that they learn some types of stimulus-response associations more readily than others[18–21]. An animal's inductive biases presumably reflect its neural representations: the animal is likely to generalize responses between sensory stimuli evoking similar cortical activity patterns, and to differentiate stimuli evoking different patterns[18,19], and experimental evidence suggests this is indeed the case[22,23]. Thus, plasticity of sensory cortical representations may serve to change inductive bias: for an animal to make different associations to two stimuli, the cortical representations of the stimuli must become differentiated, such as if the firing vectors they evoke become more orthogonal[24].

A second, non-exclusive, hypothesis that is often tested in learning experiments is that task training increases the fidelity of cortical stimulus coding. Cortical responses vary between repeated presentations of an identical stimulus, and this variability could limit the ability of even an ideal observer to decode the stimulus from neuronal activity. Such failures of decoding are most noticeable when trying to decode stimulus identity from the activity a single neuron, but an ideal observer would be unable to accurately decode the stimulus from even a large population if trial-to-trial variability is correlated between neurons in an "information-limiting" manner[25–27]. It has been suggested that task training changes the size and correlation structure of trial-to-trial variability, thereby improving the fidelity of the population code found in naïve cortex[4,11,12,14,15,28]. This hypothesis, of course, presupposes that the population code in naïve cortex does suffer from low fidelity, which has been questioned by recent recordings of large cortical populations[29].

We used two-photon calcium imaging to study how the tuning of V1 populations changes after mice learn to associate opposing actions to two oriented gratings. Training did not improve the fidelity of stimulus coding, which was already perfect in naïve animals

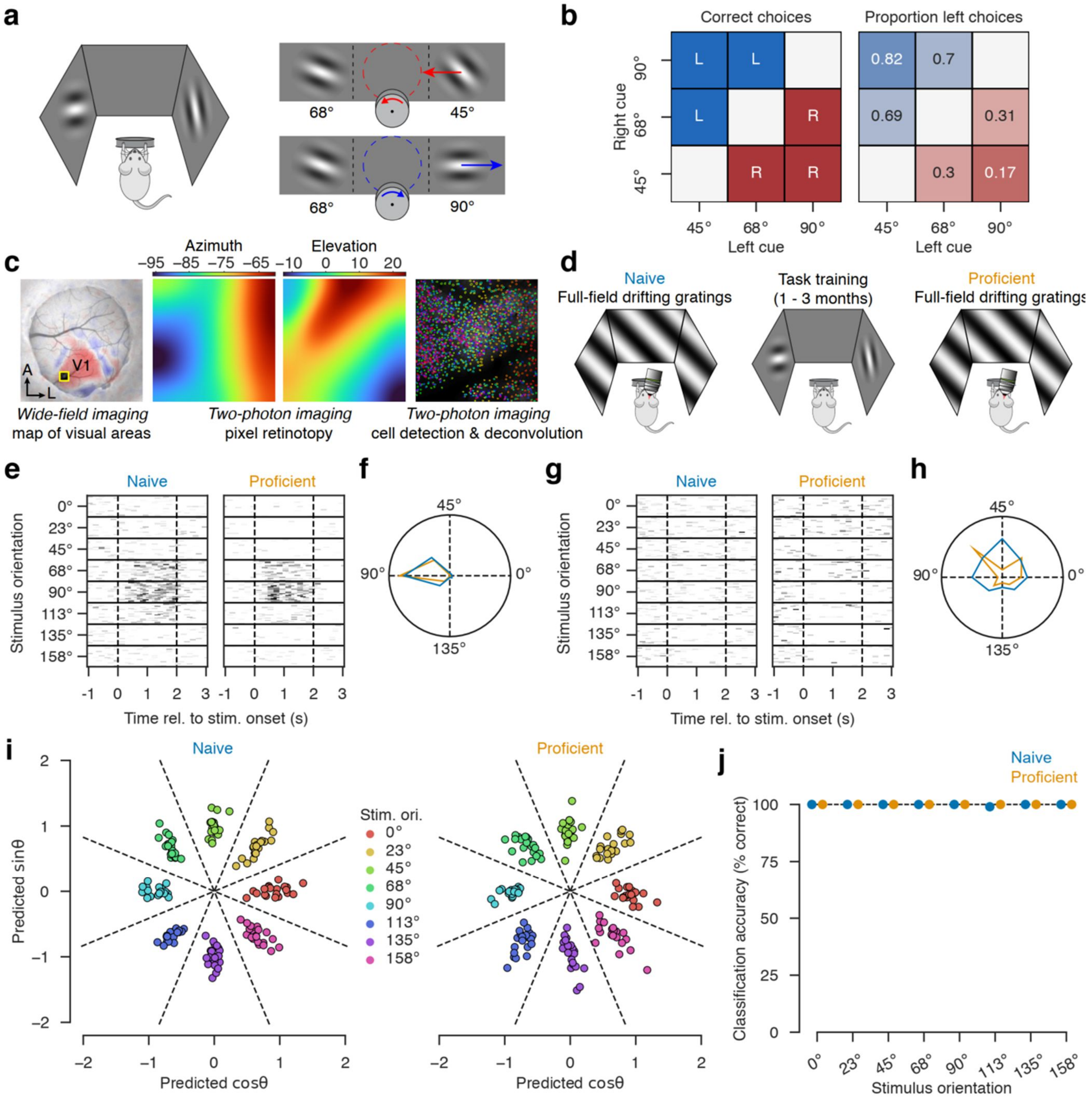


Figure 1 | Stimuli are accurately encoded by V1 populations before and after training on a visuomotor association task. a, On each trial mice are presented with two stimuli and then turn a wheel to move them on the screens. Turning towards the 45º stimulus or turning away from the 90º stimulus yields a reward, but 68º stimuli are distractors. **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. Responses to drifting grating stimuli were recorded in naïve mice, and in the same mice after they had become proficient at the task. **e,** Raster representation of responses to repeated grating stimuli for an example cell in a Naïve mouse, and a second cell from the same mouse when proficient at the task. **f**, Orientation tuning curves of the same two cells superimposed in polar coordinates (radius represents mean response of the cell to each orientation). **g,h,** Same as e,f for two more cells of weaker orientation selectivity. **i,** 2d projection of population response vectors for each orientation from one mouse before (left) and after training (right). **j,** Cross-validated classification accuracy for decoding stimulus orientation from naïve and proficient mice. Dashed line indicates perfect performance (n = 5 mice).

thanks to a subpopulation of neurons encoding the stimuli with high accuracy. Instead, training caused the mean responses to different visual stimuli to differentiate, becoming more orthogonal, an effect that was strongest for the stimuli with opposite behavioral associations. The effect of training on population activity could be fit by a simple mathematical function: a nonlinear transformation of firing rates, whose convexity is largest for motor-associated stimuli. This transformation sparsens the representations of these stimuli and makes them more orthogonal. The strength of transformation varies consistently across the population on a trial-by-trial basis, suggesting it emerges from circuit dynamics, rather than static synaptic plasticity.

Results

We trained mice in a visuomotor association task (Fig. 1a-b; Supplementary Fig. 1). Mice were shown pairs of stimuli and were trained to form motor associations with gratings of two orientations (45° and 90°) representing opposite behavioral contingencies (turn towards vs. turn away), while a third orientation was a distractor (68°) that was presented as often as the motor-associated stimuli. No other orientations were presented during task performance.

To study how task training affected cortical representations of visual stimuli, we assessed the orientation tuning of excitatory cells in V1 using two-photon calcium imaging (Fig. 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, drifting gratings were presented to passive mice in the same apparatus as the task, but the wheel was not coupled to visual stimuli and no rewards were given. Presentation of gratings in this passive condition caused pupil constriction, which was more prominent following training but not specific to any orientation (Supplementary Fig. 2a-b). Stimulus presentation evoked minimal whisking that was not significantly affected by training or orientation (Supplementary Fig. 2c-d). Thus, even though body movements modulate visual cortical activity[30–33], analyzing passive stimulus responses avoided this potential confound.

Visuomotor association does not improve decodability of task stimuli

The population code for grating orientation had extremely high fidelity, in both naïve and trained mice. Individual cells showed a range of tuning characteristics. Some neurons in both naïve and proficient mice showed sharp orientation tuning and reliable responses (Fig. 1e-f). Other neurons showed broader tuning or less reliability, with multi-peaked tuning curves particularly noticeable in proficient mice (Fig. 1g-h). Applying dimensionality reduction to the population activity (Methods), we observed that population responses to different grating stimuli showed essentially no overlap (Figure 1i). As a first test of the fidelity with V1 encoded grating orientation, we decoded the stimulus orientation from population activity using linear regression. This yielded essentially 100% cross-validated accuracy for all orientations, in both naïve and proficient mice (Fig. 1j). This result does not support the hypothesis that correlated neural noise presents a fundamental limit to the fidelity of stimulus coding in naïve animals, at least for the stimuli used here. Because this hypothesis has been influential, we examined our contradictory evidence in substantial further detail, to be sure it is valid. These analyses revealed that training-related changes in stimulus representations have no effect on the fidelity of stimulus encoding, due to the existence of a sparse subset of neurons which encode the stimuli with extremely high accuracy, in both naïve and proficient conditions (Appendix 1). We therefore next investigated how the mean response to each stimulus changed, to see if this structure matched the predictions of the inductive bias hypothesis.

Training specifically suppresses responses to task stimuli in weakly-tuned cells

To analyze how visuomotor association changes the V1 population code, we examined the mean responses of individual neurons to gratings of all orientations, as summarized by their orientation tuning curves (Fig. 2a-b). In naïve animals, tuning curves typically had a standard single-peaked profile (Fig. 2a). In proficient animals, however, tuning curves often showed an irregular, multipeaked form (Fig. 2b). Closer examination suggested that these multipeaked orientation tuning curves had dips at the visuomotor associated orientations 45° and 90°, suggesting that mean responses to these stimuli are suppressed after training, in at least some cells.

The suppression of responses to task orientations was strongest in weakly-tuned cells (Fig. 2c-f). We first computed each cell's modal orientation preference, i.e. the orientation that drove it most strongly. We found that task training decreased the fraction of cells modally preferring the motor-associated orientations (45° and 90°), but not the distractor orientation (68°) (Fig. 2c; 45°: p = 0.012; 68°: p = 0.228; 90°: p = 0.006, paired-sample *t*-test, n = 5 mice), consistent with suppression of responses specifically to motor-associated stimuli. The decrease in cells modally preferring the motor-associated orientations came specifically from cells of low orientation selectivity (assessed by the length of the circular mean response vector; arrows in Figs. 2a,b), with no decrease in the number of cells strongly tuned for motor-associated orientations (Fig. 2d; 45°: p = 0.005 and 0.037 for orientation selectivity 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).

Tuning curves also changed shape after training, in a manner dependent on a cell's preferred orientation and selectivity (Fig. 2e-f). We grouped the recorded cells by

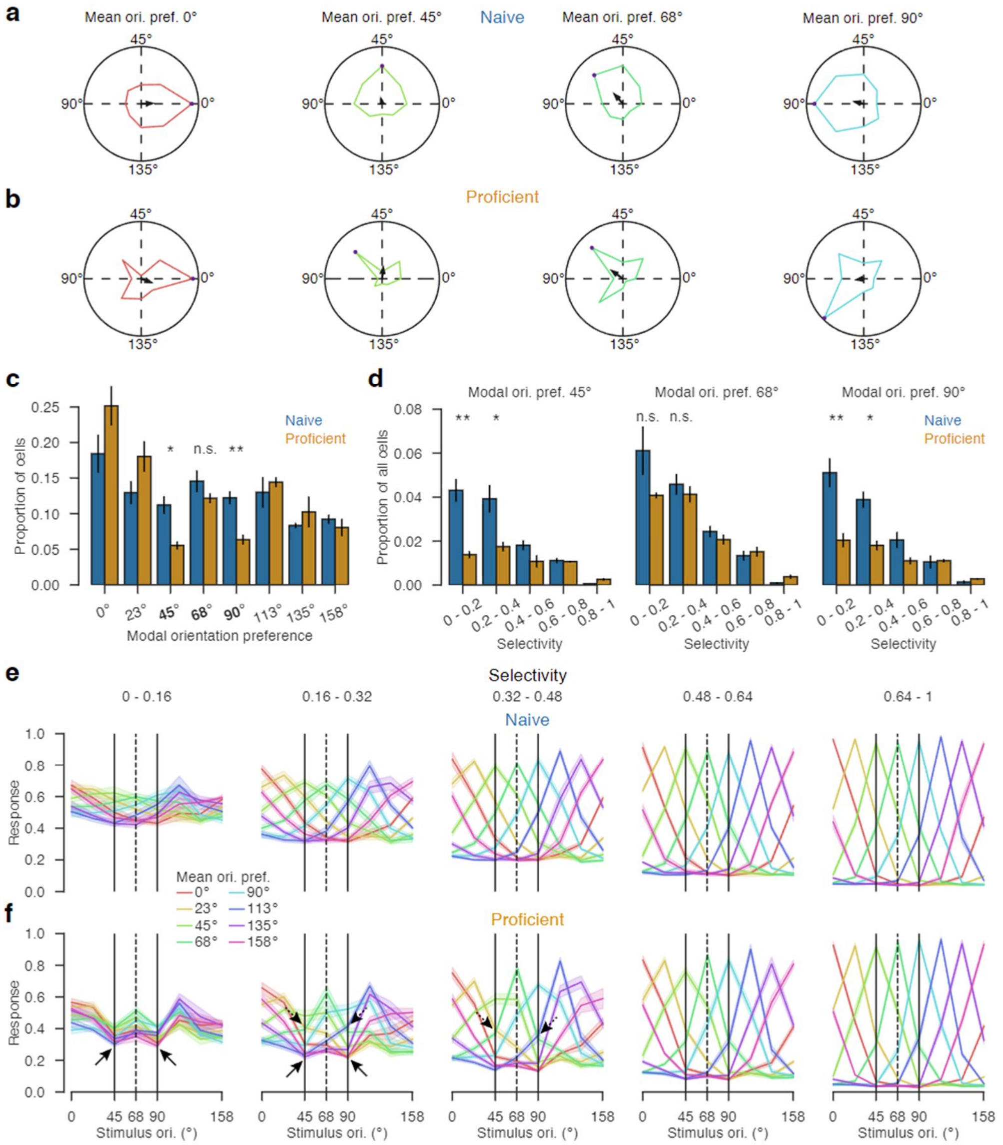


Figure 2 | a, Single-cell orientation tuning curves from naïve mice, for four cells with mean orientation preference 0°, 45°, 68°, and 90°. Colored polar curves: neural response to each orientation; dots: response to modal orientation; arrows: circular mean vectors representing mean orientation preference (angle) and orientation selectivity (magnitude). **b,** Similar plots for four other cells from mice proficient at the task. **c,** Proportion of cells with each modal orientation preference, in naïve and proficient mice. Error bars: SEM (n = 5 mice). **d,** 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. **e,** Average orientation tuning curves for cell groups defined by mean orientation preference (color) and selectivity (column) in naïve mice. Solid vertical lines indicate motor-associated orientations, dashed the distractor (68°). **f,** Same plot for proficient mice. Solid arrows highlight suppression of cell responses to the motor-associated orientations 45° and 90°. Shading: SEM (n = 5 mice)

their orientation 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 naïve mice, tuning curves had a uniform structure (Fig. 2e). 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 (Fig. 2f). Weakly tuned neurons were suppressed by the motor-associated orientations regardless of their preference. Cells whose mean orientation preference was at or close to a motor-associated orientation exhibited multimodal tuning curves after training, for which the mean and modal orientation preference differed (examples in Fig. 2b). For more strongly tuned cells, suppression by motor-associated orientations were still visible, primarily in neurons with a mean orientation preference adjacent to them. This suppression led to an asymmetry in tuning curve slopes (Supplementary Fig. 3), as previously reported in primate [16].

A mathematical model for how training changes tuning curves

Although the training-related changes to tuning curves appeared complex when analyzed in terms of single-cell statistics, they could be accurately summarized by a simple mathematical model (Fig. 3). We will first describe and verify this model, before considering its computational implications for V1 population coding.

In the model, visuomotor learning transforms V1 population responses by applying a convex nonlinear transformation to the response of each cell (Fig. 3a): if cell c's response to orientation θ was $f_{c,\theta}$ before training, then after training it is $f'_{c,\theta} = g_\theta(f_{c,\theta})$, where the function g_θ depends on the stimulus θ, but not on the cell c. If the function g_θ is convex, then cells that responded modestly to orientation θ will have their responses to θ further suppressed after training, but cells that responded to θ either strongly or not at all will be unaffected. The model accurately summarized the effects of task training: responses in naïve and proficient mice could be accurately related by piecewise linear functions whose shape varied between orientations but not between cells (Fig. 3b). The convexity of the function g_θ relating naïve to proficient responses was larger for motor-associated orientations

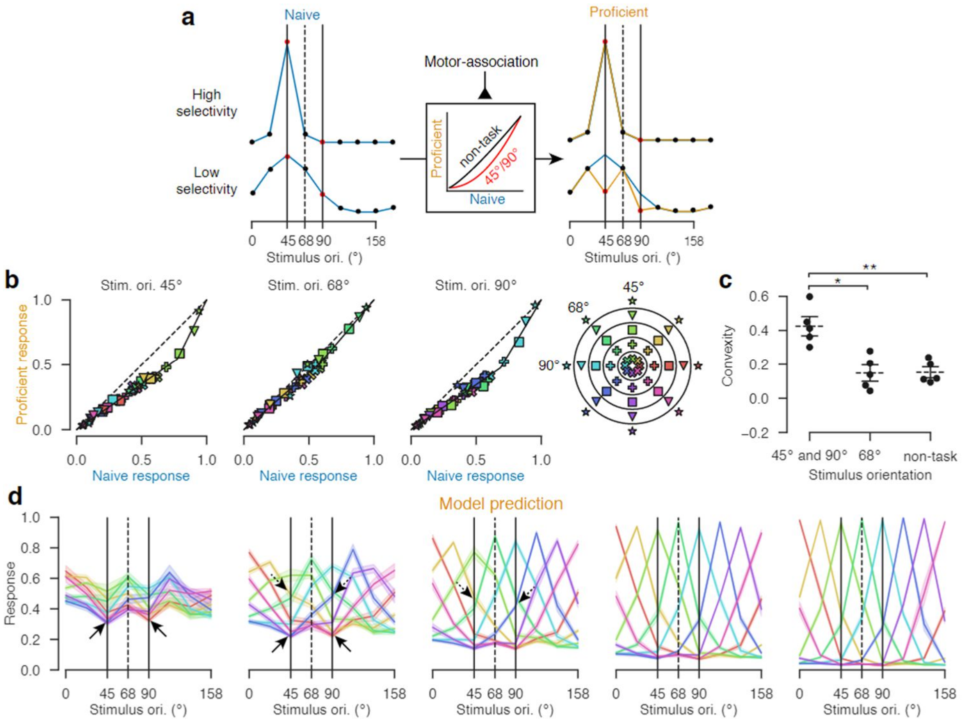


Figure 3 | Mathematical model for transformation of population activity by task training. a, Model schematic. Following task training, the naive response $f_{c,\theta}$ of cell c to stimulus θ is transformed by nonlinear function g_θ, which depends on the stimulus θ but not the cell c. Blue curves on the left illustrate tuning curves $f_{c,\theta}$ of two hypothetical cells in naïve condition. Middle box illustrates the function g_θ, which is more convex for motor-associated stimuli (red curve) than for non-task stimuli (black curve). Orange curves to the right show the proficient responses $g_\theta(f_{c,\theta})$, superimposed on original naïve curves (blue). This transformation specifically suppresses moderate responses to the motor-associated stimuli, but does not affect strong or zero responses to motor-associated stimuli, or any responses to non-task stimuli. Thus, a cell that was highly selective to 45° is unaffected (top right), while a cell that was weakly selective to 45° develops a multi-peaked tuning curve. **b,** Empirical fits of the function g_θ for θ =45°, 68°, and 90°. Each symbol shows the mean response of the same cell groups analyzed in Fig. 3e,f) to the orientation θ in Naïve vs Proficient conditions. Symbol color indicates orientation preference and glyph indicates selectivity following the code illustrated in polar coordinates on the right. Each point shows the average response of cells from all experiments. Black lines are stimulus-specific fits of piecewise linear functions g_θ relating naïve responses to proficient responses. **c,** Convexity of g_θ, for motor associated orientations 45° and 90°, distractor orientation 68°, and all other orientations. Points indicate individual mice. Error bars: mean and SEM (n = 5 mice). **d,** Proficient 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).

than for the distractor orientation or for nontask orientations (Fig. 3c**;** motor-associated vs 68°: p = 0.003; motor-associated vs non-task: p = 0.001; Independent samples *t*-test, n = 5 mice). Applying this transformation to the naïve tuning curves, we were able to predict neuronal responses in proficient subjects with remarkable accuracy (Fig. 3d**;** compare to Fig. 2f).

The model provides a simple, quantitative explanation for the qualitative features of tuning curve changes seen earlier. It explains why training affects mostly the cells that are broadly tuned and gives them multipeaked tuning curves: these cells exhibit intermediate levels of response that are affected most by the convex nonlinearity, and thus suppressed specifically to the task orientations. In contrast, strongly tuned cells fire close to either the minimum or maximum possible for all stimuli, so are not affected by the nonlinearity.

Tuning curve transformation varies dynamically from trial to trial

Plasticity of cortical representations is often assumed to arise from long-term plasticity of local excitatory synapses that changes the sensory drive received by cortical neurons[34,35]. Our observations, however, suggest an alternative hypothesis. Under this hypothesis, cortical neurons receive a sensory drive that is unaffected by training, but motor-associated stimuli engage a circuit process that suppresses the firing of cells receiving weak sensory drive while sparing strongly driven cells. Multiple physiological mechanisms could underlie this process, for example if motor-associated stimuli caused increased activation of a particular inhibitory cell class or neuromodulatory pathway.

The hypothesis makes an experimental prediction: if training-related changes in sensory tuning arise from a dynamic process, then engagement of this process should vary between repeats of an identical stimulus. Thus, the degree of transformation in sensory responses should fluctuate between stimulus repeats, and since the circuit process would affect all neurons similarly, trial-to-trial variations in response transformation should be consistent across the population. Finally, it is possible that the degree of transformation on each trial correlates with current behavioral state. Trial-to-trial variability in neuronal responses is well-documented and has been reported to take additive and multiplicative forms [36–38]. The current hypothesis 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 (Fig. 4a).

To test this prediction, we examined population responses in proficient mice on single trials (Fig. 4b-d). We divided cells randomly into two groups, balanced for orientation preference and selectivity, and within each cell group examined the transformation from trial-averaged population responses in the naïve condition, to single-trial population activity in the proficient condition. The convexity of the transformation varied substantially between trials, even within repeats of a single stimulus orientation, but was consistent across cell groups (Fig. 4c-d; correlation coefficient significantly exceeds 0 at $p < 0.05$ for each stimulus orientation, one sample *t*-test, n = 5 mice). The strength of tuning curve transformation on a given trial did not reflect stimulus-evoked movements, which did not vary between different grating orientations (Supplementary Fig. 2). However, the strength of tuning curve transformation to motor-associated stimuli on a given trial did vary with ongoing behavioral state, being strongest on trials where the mouse was whisking prior to stimulus onset, and this effect was only seen for motor-associated stimuli (Fig. 4e, Supplementary Fig. 4; linear mixed effects model: $p = 6.4 \times 10^{-5}$ for effect of whisking on convexity for motor-associated stimuli; $p = 1.4 \times 10^{-2}$; for difference between effect of distractor vs. motor-associated stimuli on convexity; $p = 1.2 \times 10^{-3}$ for difference between effect of non-task vs. motor-associated stimuli). Furthermore, transformation of activity on trials of high convexity was largest in areas of V1 topographically representing the task stimulus location, and affected neuropil as well as cellular activity as would be expected if it were driven by local inhibitory neurons (Fig. 4f-g; Supplementary Fig. 5).

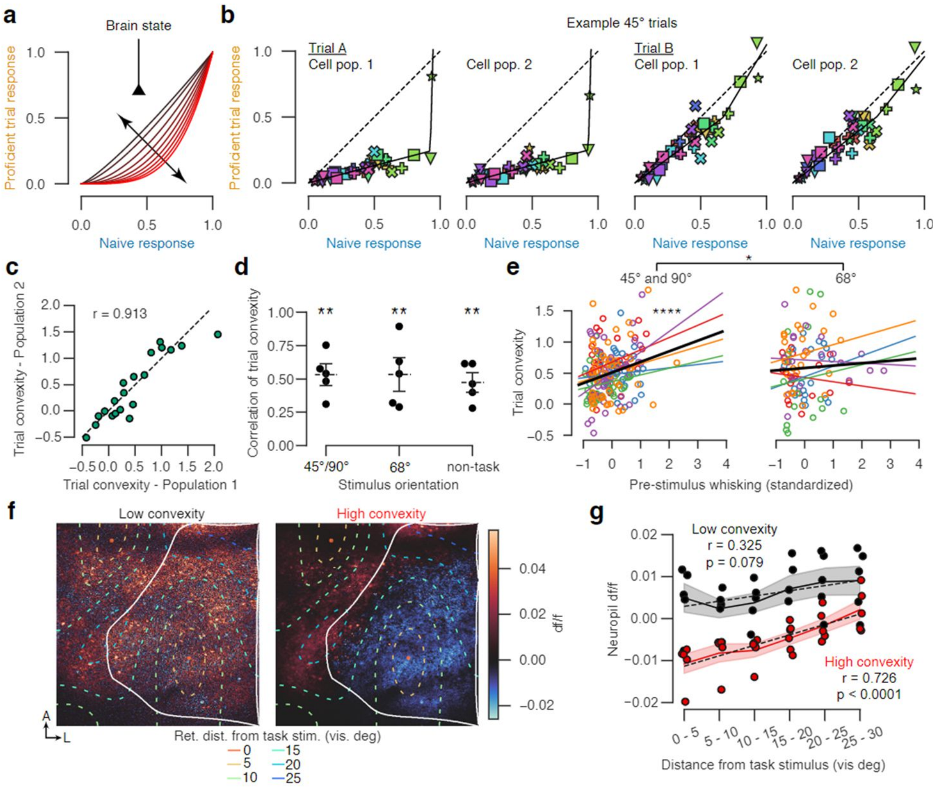


Figure 4 | Trial-to-trial variability of response transformation and dependence on behavioral state. a, Dynamic sparsening model: activity undergo varying levels on different trials, depending on instantaneous brain state. **b,** Single-trial transformation functions for two example presentations of 45° gratings in the same recording session, plotted as Fig. 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 average over motor-associated, distractor, or non-task stimuli in one experiment. Error bars: mean and SEM (n = 5 mice). **e,** Correlation of trial convexity with pre-stimulus whisking. Each point represents a stimulus presentation, color coded by mouse identity. Colored lines are linear regression fits for individual mice, black line the mean over mice. Left: motor-associated stimuli; right: distractor stimuli. **f,** Trial-to-trial variability of neuropil responses. Left and right plots show mean df/f of two-photon imaging frames to motor-associated orientations for low convexity (< 0) and high convexity (> 0.3) trials. Colored contours correspond to retinotopic distances from task stimulus location (see legend). **g,** 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). *, $p < 0.05$, **, $p < 0.01$.

Training sparsens and orthogonalizes responses to motor-associated orientations

The training-related changes in sensory tuning we observed differentiated population responses to the motor-associated stimuli by sparsening them and making them more orthogonal (Fig. 5; Appendix 2). To visualize changes in the population code, we developed a "bullseye plot" (Fig. 5a), which enables one to visually compare the responses of all recorded neurons to two stimuli. The response of each recorded cell is plotted as a circle in a location given in polar coordinates by the cell's preferred orientation and orientation selectivity. The color of this circle represents the cell's response to the two task stimuli using a two-dimensional colormap, with cells responding exclusively to 45° showing in green, cells responding exclusively to 90° in magenta, and cells responding to both in black. This visualization suggested that the population code to the stimuli grew sparser after task training, and that the number of cells responding to both stimuli decreased, reflecting an orthogonalization of the codes for the two stimuli.

To quantify changes in population sparseness we used the Treves-Rolls measure[39,40], which increased for the motor-associated orientations 45° and 90°, to a

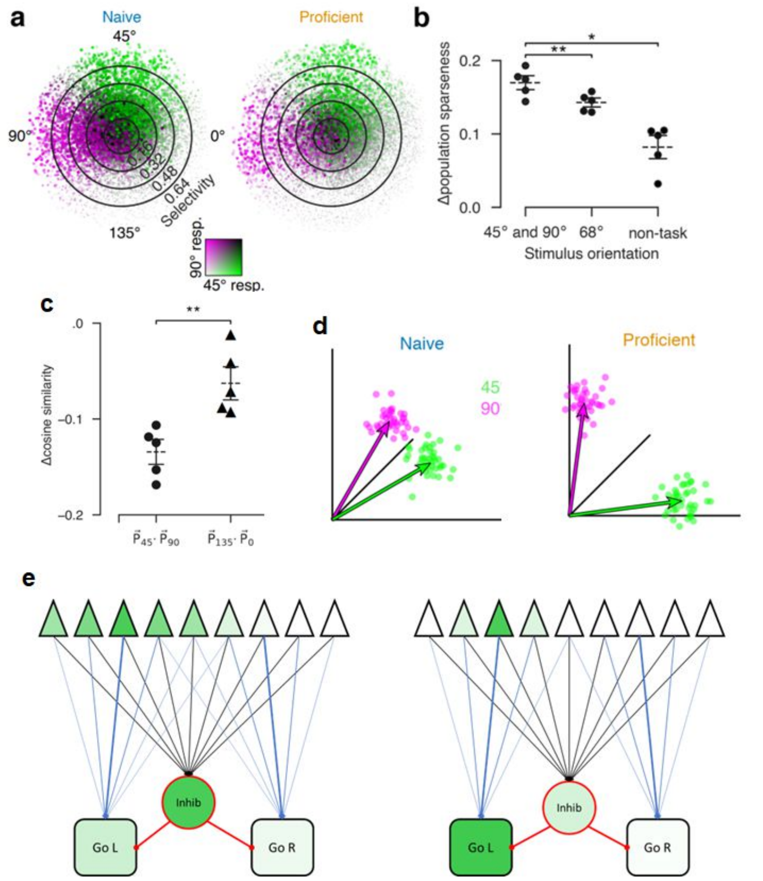


Figure 5 | Task training sparsens and orthogonalizes cortical population codes. a, "Bullseye plots" showing mean population responses to the motor-associated orientations 45° and 90° for naïve and proficient conditions. Each point represents a cell, at a polar location determined by the cell's circular mean orientation preference (angle) and selectivity (radius). The point's color represents the cell's response to the 45° (green) and 90° (magenta) stimulus orientations on an additive scale so points responding to both stimuli appear grey; the point's size and brightness (light to dark) represents the cell's maximal response to these two stimuli. **b,** Change in population sparseness between naïve and proficient conditions, as a function of stimulus orientation. Each point represents one mouse. **c,** Change following training in cosine similarity between population responses to the two motor-associated stimuli, and two non-task stimuli (see Supplementary Fig. 6 for all stimulus pairs). **d,** Cartoon illustration of geometrical effect of task training on population response vectors. Green and magenta dots represent single-trial population vectors evoked by the two motor-associated stimuli, arrows represent trial averages. Convex transformation of firing rates suppresses cells firing weakly in the naïve condition, thereby sparsening population activity and moving the population vectors closer to the coordinate axes. This orthogonalizes responses, increasing the angle between the corresponding vectors. **e,** Hypothesized consequence of sparsening and orthogonalization. Cortical cells (triangles) project to a downstream motor structure containing decision neurons that promote two separate actions (rectangles), via fixed weights (blue arrows; thickness represents synaptic strength) and nonspecific feedforward inhibition (red circle; black and red arrows). Green shading level represents activity of each cell. In the naïve condition, a dense firing pattern overlaps with the input weight vectors of both downstream neurons and also drives strong feedforward inhibition. In proficient mice, a sparser cortical code still strongly drives the correct decision neuron but drives the incorrect decision neuron and inhibitory neuron weakly, leading to stronger firing of the correct decision neuron and weaker firing of the incorrect one.

significantly greater degree than for the distractor stimulus 68°, and for non-task stimuli (Fig. 5b; 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). To quantify the orthogonality of the population codes to the two stimuli we computed the cosine similarity between their mean response vectors $(\mathbf{f}_{45} \cdot \mathbf{f}_{90} / |\mathbf{f}_{45}||\mathbf{f}_{90}|)$, which decreased after training, by a significantly larger amount than response vectors to control stimuli (Fig. 5c; p = 0.006. Independent samples *t*-test, n = 5 mice; see Supplementary Fig. 6 for all stimulus pairs). 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 (Fig. 5d).

We hypothesize that sparsening and orthogonalization of population codes could help produce correct motor outputs to stimuli, without requiring downstream synaptic plasticity (Fig. 6e). Consider a downstream motor structure, in which decision neurons receive excitatory input from V1 neurons tuned to the corresponding task stimuli, as well as feedforward inhibition reflecting the summed activity of the V1 population. When V1 population activity is dense, both decision cells receive excitatory input and strong feedforward inhibition, so both decision cells show weak activity, albeit slightly stronger for the one representing the correct choice (Fig. 5e, left). If V1 activity is sparsened, two benefits occur. First, while the excitatory drive to both decision cells reduces, it reduces more strongly to the cell producing the wrong choice. Second, the strength of feedforward inhibition goes down, resulting in a net increase of activity of the correct decision neuron (Fig. 5e, right). Thus, even though sparsening decreases total V1 activity, it could still increase the activation of downstream decision circuits.

Discussion

Training in a visuomotor task transformed population responses to oriented grating stimuli in a manner that sparsened and orthogonalized the population codes for motor-associated orientations. These changes could be explained to high quantitative accuracy by a simple mathematical 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 for motor-associated orientations. This convex transformation sparsens population responses to motor-associated orientations by suppressing neurons responding at intermediate levels, and makes the resulting population vectors more orthogonal to each other. This orthogonalization may help downstream circuits produce different behavioral responses to the two motor-associated orientations.

The way task training transformed stimulus coding was simple when described at the population level, but appeared complex if analyzing each cell's tuning individually. Training transformed the activity of all cells by a single, stimulus-dependent nonlinear function: $f_{c,\theta} \mapsto g_\theta(f_{c,\theta})$. Despite its simplicity, this transformation resulted in apparently complex changes to single-cell tuning curves, such as the emergence of multipeaked tuning, and a dependence of tuning curve plasticity on a cell's original tuning profile. Thus, even though population-level plasticity can be described by a simple formula, questions like "do tuning curves sharpen following training" need not have simple answers. Our model can nevertheless explain several of the apparently complex effects of visuomotor task training observed in previous studies of V1: it predicts a reduction in the number of cells responding modally to the trained orientations[7], an asymmetrical increase in tuning curve slope specifically at these orientations[16], and the suppression neuronal activity following learning in particular for cells with preferred orientation close to task stimuli[6,13].

Despite this concordance with previous results in visual cortex, our findings do not appear fully congruent with results from auditory and somatosensory cortex. Indeed, training on 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[41–43]. 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 experience, 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 associative learning in somatosensory cortex did observe sparsening[44], 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.[44]) 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[45]; furthermore, induction of map expansion by means other than task training can actually worsen task performance[46], in particular by increasing the rate of false responses to non-target stimuli[47]. Our task required a long training period, potentially allowing time for map expansion to reverse; it also requires differentially responding to the two stimuli while not responding to the similar distractor stimulus, for which map expansion might actually impair performance.

We did not observe an increase in the fidelity of orientation coding following training, as we found that stimuli could be decoded from population activity with 100% accuracy even in naïve mice. This result contrasts with some previous studies[4,11,12,14,15], for which we offer three non-exclusive possible explanations. The stimuli we were decoding – high-contrast full-screen drifting gratings, with orientations separated by 45° and no superimposed noise – were very distinct. The idea that cortical representations of such distinct stimuli would be of low enough fidelity that decoding them is difficult, is controversial. Indeed, a recent study found that gratings separated by just 1° could be decoded accurately[29]. The fact that some previous studies have failed to accurately decode such distinct stimuli does not prove it cannot be done, as there are several technical factors that can compromise stimulus decoding. First, two-photon microscopy is subject to artifacts such as brain movement and neuropil contamination, which unless corrected with appropriate software will introduce noise with correlations of exactly the form that compromise decoding[48–50]. Second, activity in V1 encodes not only visual stimuli, but also non-visual features such as ongoing movements[32,33]. This non-visual information may compromise decoder performance, particularly for recordings performed during performance of the behavioral task. Finally, the performance obtained by any one decoder represents a lower bound on the performance of an ideal observer, as decoder performance is sensitive to parameters such as regularization methods, particularly when decoding from with large numbers of cells.

It is often assumed that plasticity of cortical representations arises from plasticity of excitatory inputs onto the cells being recorded. This form of plasticity does not seem most likely to explain our results, given that long-term changes to synaptic strengths are presumably static, while the amount of population code transformation we observed varies from trial to trial. Clearly, synaptic or cellular plasticity must occur somewhere to explain the change in mean tuning, but we suggest that this plasticity occurs in a circuit carrying feedback to V1 pyramidal cells rather than in their feedforward sensory drive. Several possibilities for this feedback circuit are consistent with our results. Local feedback inhibition contributes to V1 orientation tuning[51]. Furthermore, stimulation of parvalbumin-positive interneurons narrows tuning curves in a manner consistent with convex transformation of firing rates, and improves behavioral orientation discrimination[52]. Our results could thus arise from strengthening of inputs onto these interneurons from local pyramidal cells tuned to motor-associated stimuli[53,54]. Alternatively, the feedback could arise from more distal cortical regions or neuromodulators, which also modulate local inhibitory classes[55–57]. The fact that cortical representations were most strongly transformed at times of higher alertness (as indicated by pre-stimulus whisking), suggests that the operation of this hypothesized feedback circuit may also be modulated by cognitive state.

Regardless of the underlying mechanism, the fact that training-related sparsening leads to orthogonalization of the representations of the motor-associated stimuli suggests a function for this process. Orthogonalizing the representations of these stimuli may allow the brain to reduce behavioral generalization between them, allowing the mouse to respond to them differently[24]. Gratings are not natural stimuli, and if a mouse ever did encounter one in the wild, it is unlikely that the grating's orientation would be of any behavioral significance. Thus, one might expect mice by default to generalize their behavioral responses from one orientation of grating to another; only after extensive training should behavioral responses to gratings of different orientations diverge. Orthogonalization of cortical representations may override this default generalization, and so allow different orientations to evoke different behavioral responses. Applications of similar techniques to artificial learning systems might provide a new mechanism to boost their learning capacity.

Acknowledgments

We thank Charu Reddy for assistance with mouse surgeries, Miles Wells, Hamish Forest, and Laura Funnel for assistance with task training, and 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

	SWF	MC	KDH
Conceptualization	•	•	•
Methodology	•	•	•
Investigation	•		
Data curation	•		
Formal analysis	•		•
Funding acquisition		•	•
Project administration	•		
Supervision		•	•
Visualization	•		•
Writing	•	•	•

Competing interests

The authors have no competing interests to declare.

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.

Visuomotor association task

The task is a modification of a two-alternative forced choice contrast discrimination task previously developed by our lab [58]. 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 1 - 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 [59]. 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 [60]. 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. [61], 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 [50]. 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 [39].

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 visuomotor association-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.

References

1. Goodfellow, I., Bengio, Y., and Courville, A. (2017). Deep Learning (The MIT Press).

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

3. DiCarlo, J.J., Zoccolan, D., and Rust, N.C. (2012). How does the brain solve visual object recognition? Neuron *73*, 415–434. 10.1016/j.neuron.2012.01.010.

4. Adab, H.Z., and Vogels, R. (2011). Practicing Coarse Orientation Discrimination Improves Orientation Signals in Macaque Cortical Area V4. Curr. Biol. *21*, 1661–1666. 10.1016/j.cub.2011.08.037.

5. Chowdhury, S.A., and DeAngelis, G.C. (2008). Fine Discrimination Training Alters the Causal Contribution of Macaque Area MT to Depth Perception. Neuron *60*, 367–377. 10.1016/j.neuron.2008.08.023.

6. Corbo, J., McClure, J.P., Erkat, O.B., and Polack, P.-O. (2022). Dynamic Distortion of Orientation Representation after Learning in the Mouse Primary Visual Cortex. J. Neurosci. *42*, 4311–4325. 10.1523/JNEUROSCI.2272-21.2022.

7. Ghose, G.M., Yang, T., and Maunsell, J.H.R. (2002). Physiological Correlates of Perceptual Learning in Monkey V1 and V2. J. Neurophysiol. *87*, 1867–1888. 10.1152/jn.00690.2001.

8. Goltstein, P.M., Meijer, G.T., and Pennartz, C.M. (2018). Conditioning sharpens the spatial representation of rewarded stimuli in mouse primary visual cortex. eLife *7*, e37683. 10.7554/eLife.37683.

9. Goltstein, P.M., Coffey, E.B.J., Roelfsema, P.R., and Pennartz, C.M.A. (2013). In vivo two-photon Ca2+ imaging reveals selective reward effects on stimulus-specific assemblies in mouse visual cortex. J. Neurosci. Off. J. Soc. Neurosci. *33*, 11540–11555. 10.1523/JNEUROSCI.1341-12.2013.

10. Gu, Y., Liu, S., Fetsch, C.R., Yang, Y., Fok, S., Sunkara, A., DeAngelis, G.C., and Angelaki, D.E. (2011). Perceptual Learning Reduces Interneuronal Correlations in Macaque Visual Cortex. Neuron *71*, 750–761. 10.1016/j.neuron.2011.06.015.

11. Henschke, J.U., Dylda, E., Katsanevaki, D., Dupuy, N., Currie, S.P., Amvrosiadis, T., Pakan, J.M.P., and Rochefort, N.L. (2020). Reward Association Enhances Stimulus-Specific Representations in Primary Visual Cortex. Curr. Biol. *30*, 1866-1880.e5. 10.1016/j.cub.2020.03.018.

12. Jurjut, O., Georgieva, P., Busse, L., and Katzner, S. (2017). Learning Enhances Sensory Processing in Mouse V1 before Improving Behavior. J. Neurosci. *37*, 6460–6474. 10.1523/JNEUROSCI.3485-16.2017.

13. Poort, J., Wilmes, K.A., Blot, A., Chadwick, A., Sahani, M., Clopath, C., Mrsic-Flogel, T.D., Hofer, S.B., and Khan, A.G. (2021). Learning and attention increase visual response selectivity through distinct mechanisms. Neuron *110*, 1–12. 10.1016/j.neuron.2021.11.016.

14. Poort, J., Khan, A.G., Pachitariu, M., Nemri, A., Orsolic, I., Krupic, J., Bauza, M., Sahani, M., Keller, G.B., Mrsic-Flogel, T.D., et al. (2015). Learning Enhances Sensory and Multiple Non-sensory Representations in Primary Visual Cortex. Neuron *86*, 1478–1490. 10.1016/j.neuron.2015.05.037.

15. Raiguel, S., Vogels, R., Mysore, S.G., and Orban, G.A. (2006). Learning to See the Difference Specifically Alters the Most Informative V4 Neurons. J. Neurosci. *26*, 6589–6602. 10.1523/JNEUROSCI.0457-06.2006.

16. Schoups, A., Vogels, R., Qian, N., and Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. Nature *412*, 549–553. 10.1038/35087601.

17. Yan, Y., Rasch, M.J., Chen, M., Xiang, X., Huang, M., Wu, S., and Li, W. (2014). Perceptual training continuously refines neuronal population codes in primary visual cortex. Nat. Neurosci. *17*, 1380–1387. 10.1038/nn.3805.

18. Bordelon, B., and Pehlevan, C. (2022). Population codes enable learning from few examples by shaping inductive bias. 2021.03.30.437743. 10.1101/2021.03.30.437743.

19. Sinz, F.H., Pitkow, X., Reimer, J., Bethge, M., and Tolias, A.S. (2019). Engineering a Less Artificial Intelligence. Neuron *103*, 967–979. 10.1016/j.neuron.2019.08.034.

20. Wolpert, D.H. (1996). The Lack of A Priori Distinctions Between Learning Algorithms. Neural Comput. *8*, 1341–1390. 10.1162/neco.1996.8.7.1341.

21. Wolpert, D.H. (1996). The Existence of A Priori Distinctions Between Learning Algorithms. Neural Comput. *8*, 1391–1420. 10.1162/neco.1996.8.7.1391.

22. Hong, H., Yamins, D.L.K., Majaj, N.J., and DiCarlo, J.J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. Nat. Neurosci. *19*, 613–622. 10.1038/nn.4247.

23. Majaj, N.J., Hong, H., Solomon, E.A., and DiCarlo, J.J. (2015). Simple Learned Weighted Sums of Inferior Temporal Neuronal Firing Rates Accurately Predict Human Core Object Recognition Performance. J. Neurosci. *35*, 13402–13418. 10.1523/JNEUROSCI.5181-14.2015.

24. Barak, O., Rigotti, M., and Fusi, S. (2013). The Sparseness of Mixed Selectivity Neurons Controls the Generalization–Discrimination Trade-Off. J. Neurosci. *33*, 3844–3856. 10.1523/JNEUROSCI.2753-12.2013.

25. Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. Nat. Neurosci. *17*, 1410–1417. 10.1038/nn.3807.

26. Rumyantsev, O.I., Lecoq, J.A., Hernandez, O., Zhang, Y., Savall, J., Chrapkiewicz, R., Li, J., Zeng, H., Ganguli, S., and Schnitzer, M.J. (2020). Fundamental bounds on the fidelity of sensory cortical coding. Nature *580*, 100–105. 10.1038/s41586-020-2130-2.

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

28. Jeanne, J.M., Sharpee, T.O., and Gentner, T.Q. (2013). Associative learning enhances population coding by inverting inter-neuronal correlation patterns. Neuron *78*, 352–363. 10.1016/j.neuron.2013.02.023.

29. Stringer, C., Michaelos, M., Tsyboulski, D., Lindo, S.E., and Pachitariu, M. (2021). High-precision coding in visual cortex. Cell *184*, 2767-2778.e15. 10.1016/j.cell.2021.03.042.

30. Bimbard, C., Sit, T.P., Lebedeva, A., Harris, K.D., and Carandini, M. (2021). Behavioral origin of sound-evoked activity in visual cortex. 2021.07.01.450721. 10.1101/2021.07.01.450721.

31. Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., and Churchland, A.K. (2019). Single-trial neural dynamics are dominated by richly varied movements. Nat. Neurosci. *22*, 1677–1686. 10.1038/s41593-019-0502-4.

32. Niell, C.M., and Stryker, M.P. (2010). Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. Neuron *65*, 472–479. 10.1016/j.neuron.2010.01.033.

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

34. Cooke, S.F., and Bear, M.F. (2014). How the mechanisms of long-term synaptic potentiation and depression serve experience-dependent plasticity in primary visual cortex. Philos. Trans. R. Soc. B Biol. Sci. *369*, 20130284. 10.1098/rstb.2013.0284.

18. Cooper, L. N., Intrator, N., Blais, B. S. & Shouval, H. Z. *Theory of Cortical Plasticity*. (World Scientific, 2004).

36. Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of Ongoing Activity: Explanation of the Large Variability in Evoked Cortical Responses. Science *273*, 1868–1871. 10.1126/science.273.5283.1868.

37. Goris, R.L.T., Movshon, J.A., and Simoncelli, E.P. (2014). Partitioning neuronal variability. Nat. Neurosci. *17*, 858–865. 10.1038/nn.3711.

38. Lin, I.-C., Okun, M., Carandini, M., and Harris, K.D. (2015). The Nature of Shared Cortical Variability. Neuron *87*, 644–656. 10.1016/j.neuron.2015.06.035.

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

40. Treves, A., and Rolls, E.T. (1991). What determines the capacity of autoassociative memories in the brain? Netw. Comput. Neural Syst. *2*, 371–397. 10.1088/0954-898X/2/4/004.

41. Buonomano, D.V., and Merzenich, M.M. (1998). CORTICAL PLASTICITY: From Synapses to Maps. Annu. Rev. Neurosci. *21*, 149–186. 10.1146/annurev.neuro.21.1.149.

42. Weinberger, N.M. (2004). Specific long-term memory traces in primary auditory cortex. Nat. Rev. Neurosci. *5*, 279–290. 10.1038/nrn1366.

43. Feldman, D.E., and Brecht, M. (2005). Map Plasticity in Somatosensory Cortex. Science *310*, 810–815. 10.1126/science.1115807.

44. Gdalyahu, A., Tring, E., Polack, P.-O., Gruver, R., Golshani, P., Fanselow, M.S., Silva, A.J., and Trachtenberg, J.T. (2012). Associative Fear Learning Enhances Sparse Network Coding in Primary Sensory Cortex. Neuron *75*, 121–132. 10.1016/j.neuron.2012.04.035.

45. Reed, A., Riley, J., Carraway, R., Carrasco, A., Perez, C., Jakkamsetti, V., and Kilgard, M.P. (2011). Cortical Map Plasticity Improves Learning but Is Not Necessary for Improved Performance. Neuron *70*, 121–131. 10.1016/j.neuron.2011.02.038.

46. Han, Y.K., Köver, H., Insanally, M.N., Semerdjian, J.H., and Bao, S. (2007). Early experience impairs perceptual discrimination. Nat. Neurosci. *10*, 1191–1197. 10.1038/nn1941.

47. Thomas, M.E., P, L.C., Chaudron, Y.J.M., Cisneros-Franco, J.M., and Villers-Sidani, É. de (2020). Modifying the adult rat tonotopic map with sound exposure produces frequency discrimination deficits that are recovered with training. J. Neurosci.

48. Harris, K.D., Quiroga, R.Q., Freeman, J., and Smith, S.L. (2016). Improving data quality in neuronal population recordings. Nat. Neurosci. *19*, 1165–1174. 10.1038/nn.4365.

49. Pachitariu, M., Stringer, C., and Harris, K.D. (2018). Robustness of Spike Deconvolution for Neuronal Calcium Imaging. J. Neurosci. *38*, 7976–7985. 10.1523/JNEUROSCI.3339-17.2018.

43. Pachitariu, M. *et al.* Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv* 061507 (2017).

51. Ringach, D.L., Hawken, M.J., and Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. Nature *387*, 281–284. 10.1038/387281a0.

52. Lee, S.H., Kwan, A.C., Zhang, S., Phoumthipphavong, V., Flannery, J.G., Masmanidis, S.C., Taniguchi, H., Huang, Z.J., Zhang, F., Boyden, E.S., et al. (2012). Activation of specific interneurons improves V1 feature selectivity and visual perception. Nature *488*, 379–383. 10.1038/nature11312.

53. Bannon, N.M., Chistiakova, M., and Volgushev, M. (2020). Synaptic Plasticity in Cortical Inhibitory Neurons: What Mechanisms May Help to Balance Synaptic Weight Changes? Front. Cell. Neurosci. *14*.

54. Khan, A.G., Poort, J., Chadwick, A., Blot, A., Sahani, M., Mrsic-Flogel, T.D., and Hofer, S.B. (2018). Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. Nat. Neurosci. *21*, 851–859. 10.1038/s41593-018-0143-z.

55. Fu, Y., Tucciarone, J.M., Espinosa, J.S., Sheng, N., Darcy, D.P., Nicoll, R.A., Huang, Z.J., and Stryker, M.P. (2014). A cortical circuit for gain control by behavioral state. Cell *156*, 1139–1152. 10.1016/j.cell.2014.01.050.

56. Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J.P., Chang, W.C., Jenvay, S., Miyamichi, K., Luo, L., and Dan, Y. (2014). Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. Science *345*, 660–665. 10.1126/science.1254126.

57. Kuchibhotla, K.V., Gill, J.V., Lindsay, G.W., Papadoyannis, E.S., Field, R.E., Sten, T.A.H., Miller, K.D., and Froemke, R.C. (2017). Parallel processing by cortical inhibition enables context-dependent behavior. Nat. Neurosci. *20*, 62–71. 10.1038/nn.4436.

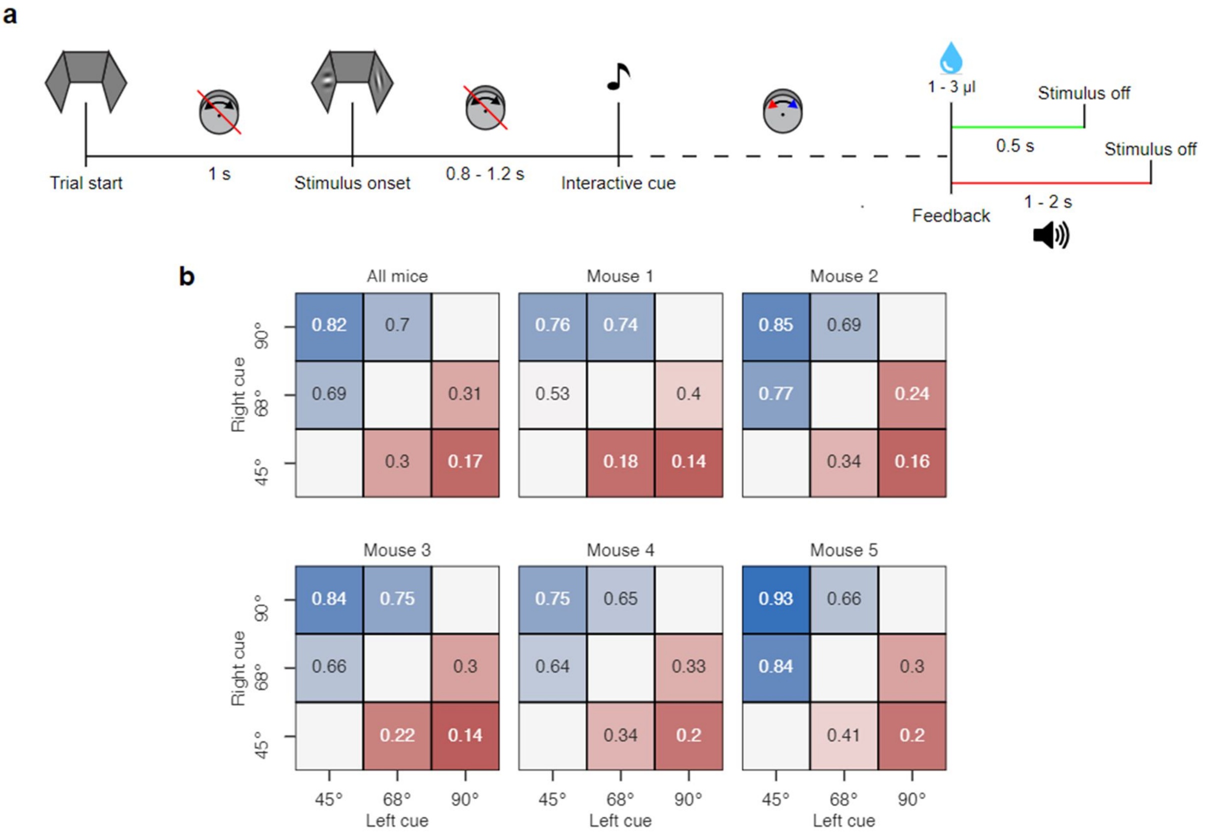
58. Burgess, C.P., Lak, A., Steinmetz, N.A., Zatka-Haas, P., Bai Reddy, C., Jacobs, E.A.K., Linden, J.F., Paton, J.J., Ranson, A., Schröder, S., et al. (2017). High-Yield Methods for Accurate Two-Alternative Visual Psychophysics in Head-Fixed Mice. Cell Rep. *20*, 2513–2524. 10.1016/j.celrep.2017.08.047.

59. Peters, A.J., Fabre, J.M.J., Steinmetz, N.A., Harris, K.D., and Carandini, M. (2021). Striatal activity topographically reflects cortical activity. Nature *591*, 420–425. 10.1038/s41586-020-03166-8.

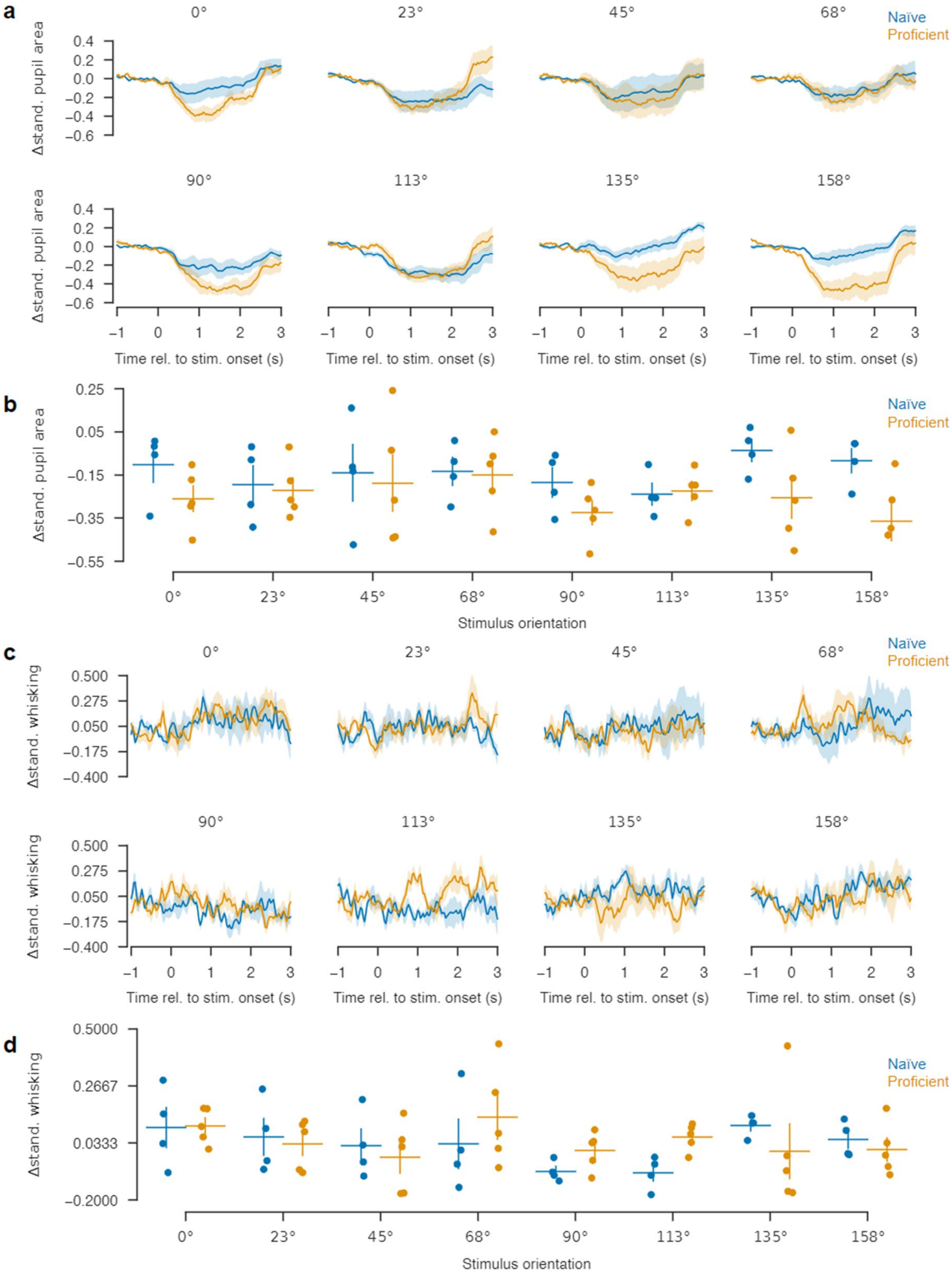
60. Pologruto, T.A., Sabatini, B.L., and Svoboda, K. (2003). ScanImage: Flexible software for operating laser scanning microscopes. Biomed. Eng. OnLine *2*, 13. 10.1186/1475-925X-2-13.

61. Sereno, M.I., McDonald, C.T., and Allman, J.M. (1994). Analysis of Retinotopic Maps in Extrastriate Cortex. Cereb. Cortex *4*, 601–620. 10.1093/cercor/4.6.601.

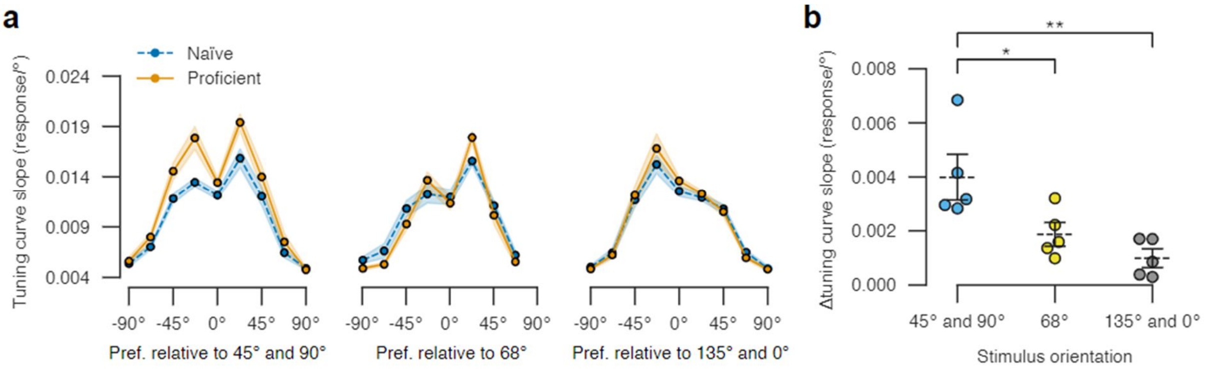
62. Hastie, T., Tibshirani, R., and Friedman, J.H. (2001). The elements of statistical learning data mining, inference, and prediction : with 200 full-color illustrations (Springer).



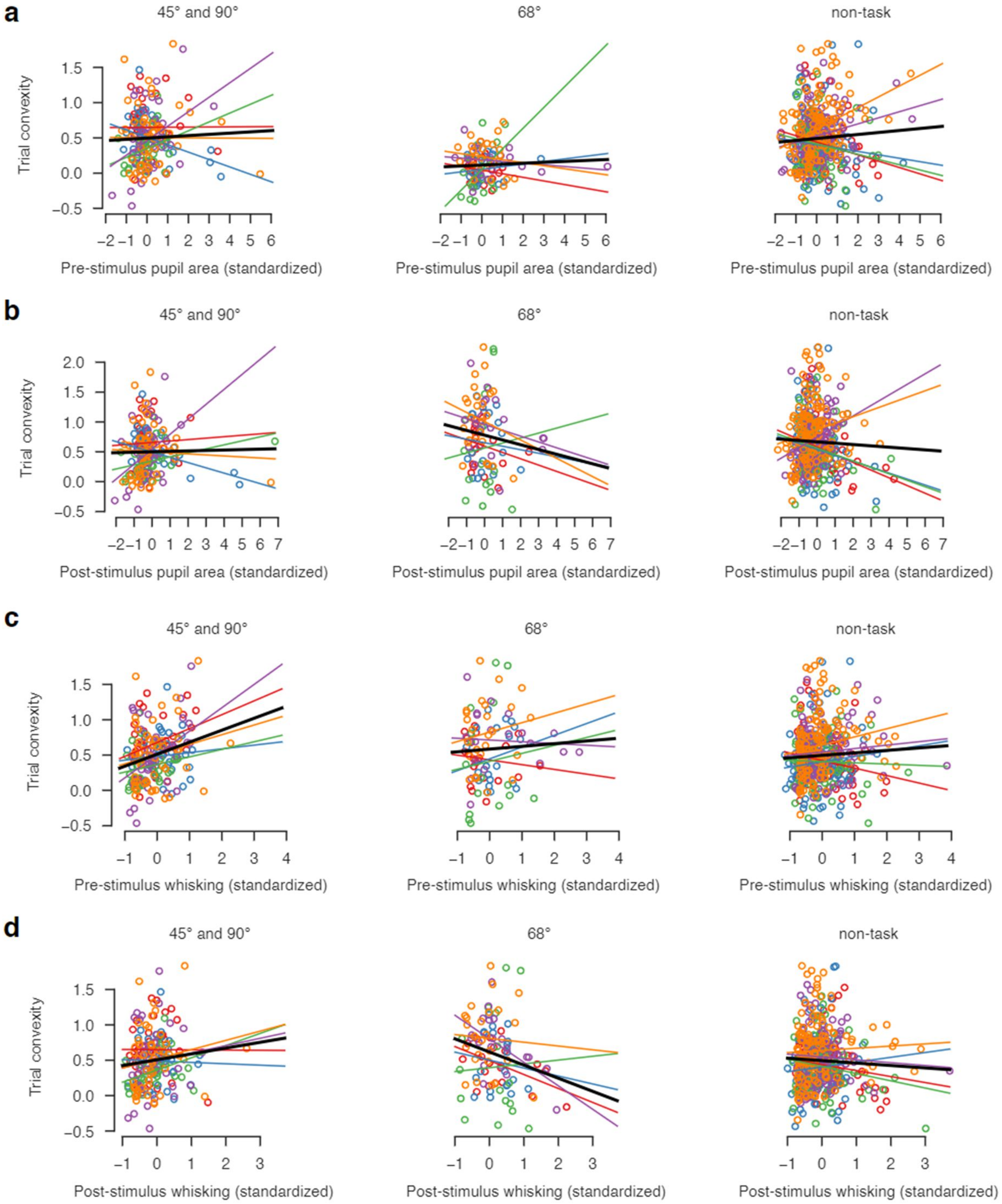
Supplementary Figure 1 | Task details. a. Temporal structure of the task. **b.** Behavioral performance for all mice. Matrices show the proportion of left choices for all cue pairings averaged over ten highest performing sessions. Cue pairings that were not presented are shown in white.



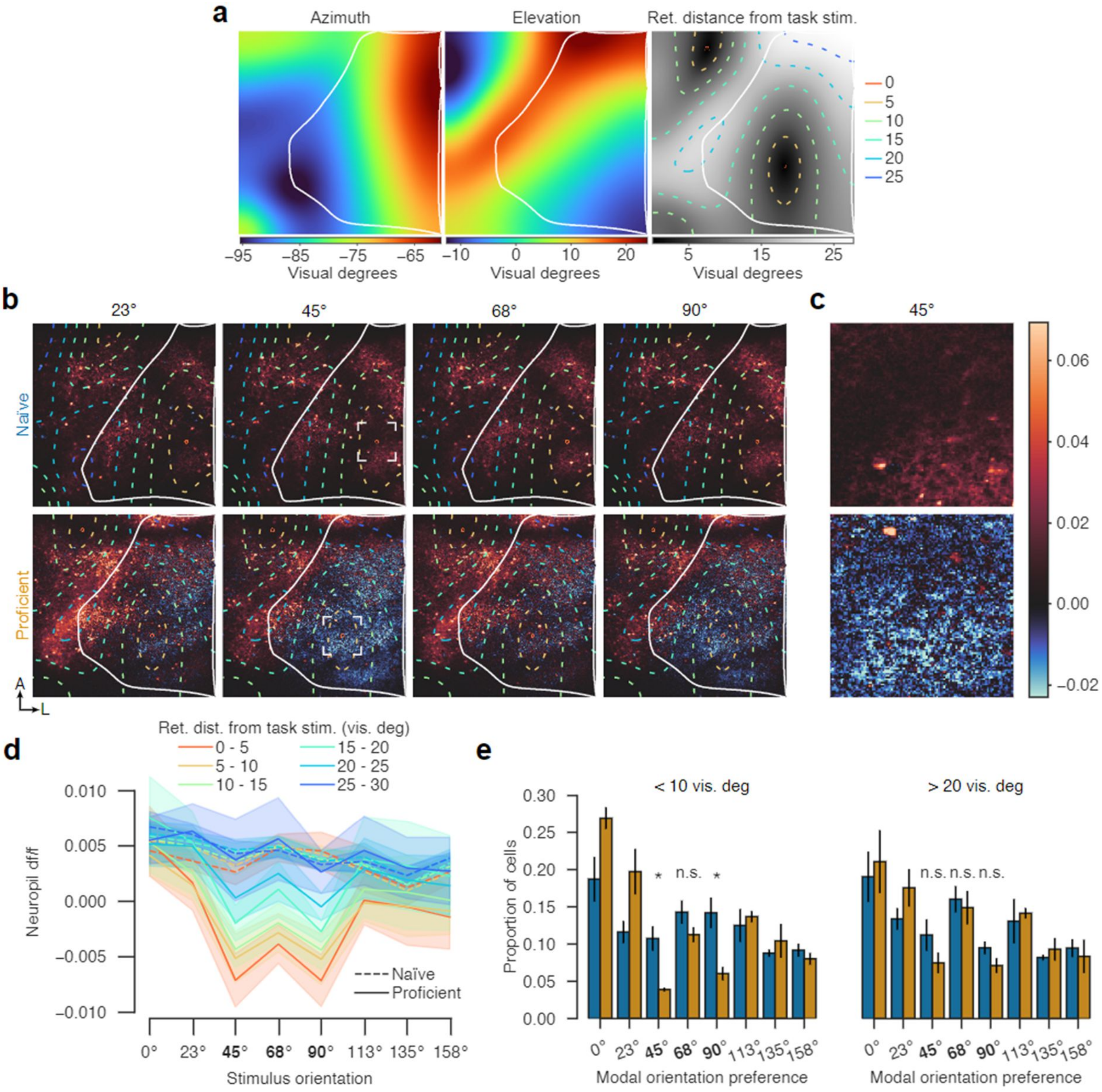
Supplementary Figure 2 | Measures of behavioral responses during passive viewing of grating stimuli. a. Stimulus-triggered pupil area time course, averaged over all trials of each stimulus orientation and training condition. Stimulus presentation causes pupil constriction, but pupil responses to motor-associated orientations 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 a marginal effect of training (p = 0.053), and no effect of 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).



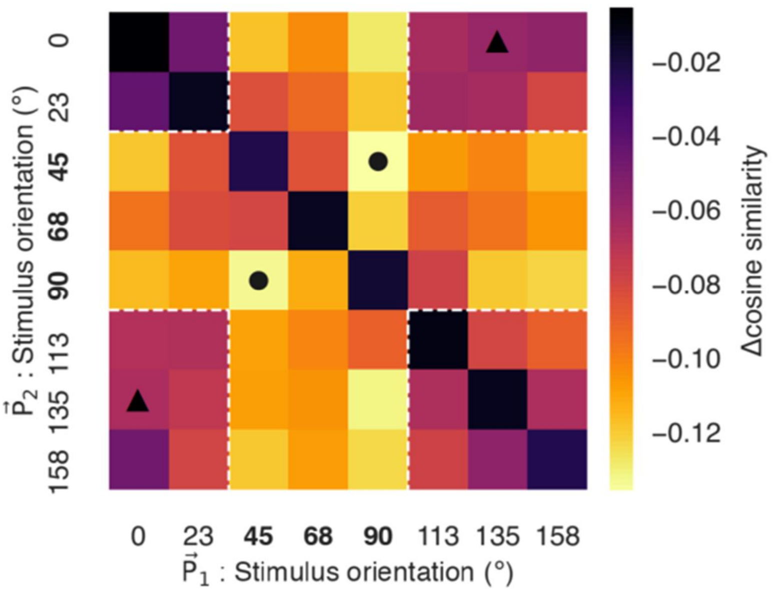
Supplementary Figure 3 | Additional metrics of single-cell tuning. a. Tuning curve slope as a function of mean orientation preference relative to the informative task orientations (45° and 90°; left), uninformative distractor orientation (68°, center), and non-task orientation controls (135° and 0°; 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).



Supplementary Figure 4 | Correlation of trial convexity with multiple measures of behavioral state. a, Correlation of single trial convexity pre-stimulus pupil area, plotted as in Fig. 5e, for motor-associated stimuli (left), distractor stimuli (center), and non-task stimuli (right). **b-d,** similar plots for post-stimulus pupil area, pre- and post-stimulus whisking. Convexity correlated positively with pre-stimulus whisking for the motor-associated stimuli (linear mixed effects model; p = 6.4 x 10^{-5}) and the effect was significantly larger than for distractor (p = 0.014) and non-task stimuli (p = 0.001). Convexity was not significantly correlated with post-stimulus whisking for the motor-associated stimuli (p = 0.065), but the distractor (p = 0.006) and non-task stimuli (p = 0.028) showed significant lower levels of correlation. The correlation of the distractor stimulus with post-stimulus whisking was significantly negative (linear mixed effects model; p = 0.009).



Supplementary Figure 5 | 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). *, $p < 0.05$.



Supplementary Figure 6 | Orthogonalization of responses to all orientation pairs. Pseudocolor matrix showing change in cosine similarity between mean population responses to each pair of orientations following task training. White dashed lines demarcate task stimuli. Black circles and triangles indicate the orientation pairs shown in Fig. 5c.

Appendix 1

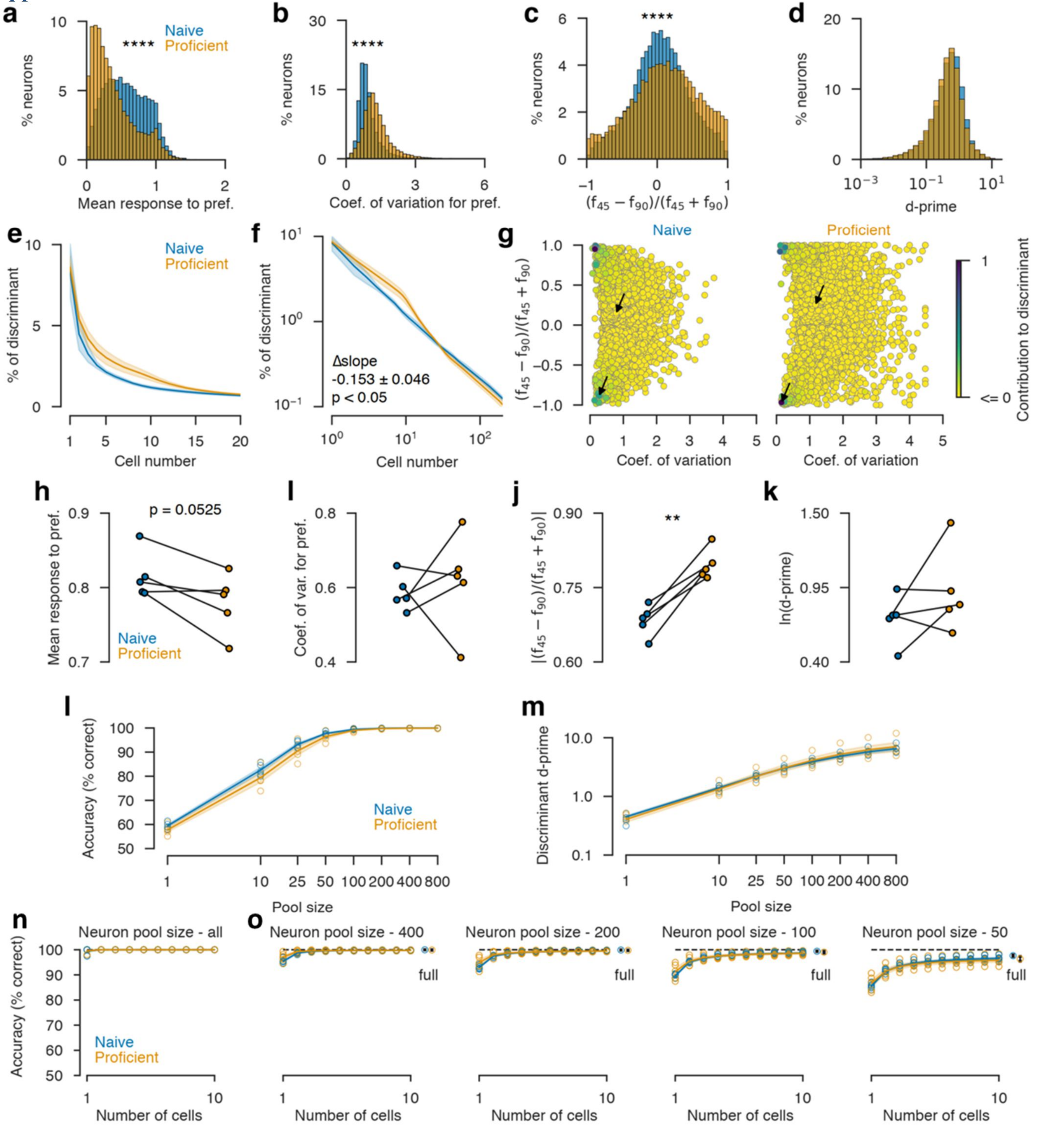


Figure A1 | Deeper analysis of coding fidelity for motor-associated stimuli in naïve and trained mice. a, Histogram of mean response of each neuron to whichever of the two motor associated orientations (45° and 90°) drove it most strongly. **b,** Histogram of coefficients of variation (standard deviation divided by mean) of each neuron's responses to its preferred stimulus. **c,** Histogram of response index comparing activity evoked by the two stimuli, for all cells. **d,** Histogram of d' discriminability for all cells (difference between means, divided by RMS standard deviation). For **a-d**, significance was assessed by a linear mixed effects model incorporating a random effect and slope for each mouse. **e,** Percentage of discriminant function accounted for by successive neurons, for an L2-regularized discriminant analysis classifier. Shading shows mean and SE over mice. **f,** same plot on a log-log scale. **g,** Analysis of cells contributing to discriminant function. Each circle represents a cell, in a position determined by its response index and coefficient of variation. Color represents percentage contribution to discriminant function. Arrows show locations of example neurons from Figs. 1e-h. **h-k,** Average over neurons contributing to the decoder of the same statistics shown in (a-d), weighted by the neurons' contributions to the discriminant function. **l,** Performance of a decoder trained on a randomly-subselected pool of neurons, as a function of decoder size. No significant difference between naïve and proficient conditions was seen for any pool size. **m,** Similar plot measuring d' of the discriminant function. Again no difference was seen for any pool size. **n,** Accuracy of decoding from an optimal cell subset of neurons, selected from the whole population by a greedy method, as a function of number of subset size. No significant difference between naïve and proficient conditions was found. **o,** Same analysis for optimal subsets greedily selected from random pools of the indicated size. In no case was a significant difference between naïve and proficient conditions found.

To more deeply investigate our result that task training did not improve representational fidelity, we focused on coding of the motor-associated 45° and 90° stimuli, which require opposite behavioral contingencies in the task. We started by analyzing the coding properties of all recorded neurons individually. The mean response of a typical cell was lower after training, even when considering each neuron's preferred motor-associated stimulus (Fig. A1a; linear mixed effects model with random intercept and slope; $p = 1.3 \times 10^{-16}$). Neuronal variability, assessed by the coefficient of variation of the response to each cell's preferred stimulus, typically increased after training (Fig. A1b; linear mixed effects model with random intercept and slope, $p = 2 \times 10^{-8}$) indicating that the decrease in mean response was not compensated by an equivalent decrease in standard deviation. Selectivity of neurons between the two motor-associated stimuli, assessed by a response index $\frac{f_{90}-f_{45}}{f_{90}+f_{45}}$, however typically grew stronger, reflecting an increase in the percentage of neurons responding almost exclusively to one stimulus (Fig. A1c; linear mixed effects model with random intercept and slope on absolute value of the response index, $p = 0.01$). Finally, the d′ statistic, which measures how well a single neuron can distinguish between the two stimuli in the face of trial-to-trial variability, did not differ significantly between naïve and trained mice (Fig. A1d, note the log x-axis; linear mixed effects model with random intercept and slope, $p = 0.1$), with small fraction of cells of very high d′ values (~10) present in both cases. Thus, the effect of training on the average neuron was mixed: an increase in the difference between the task stimuli but also an increase in coefficient of variation, leading no systematic change in d′.

These changes had no effect on decodability of the stimuli, which was perfect for both naïve and trained populations. To understand why, we analyzed the solution found by L2-regularized discriminant analysis, which computes a weighted sum of population activity (the "discriminant function") with weights that maximize the reliable difference between the 45° and 90° stimuli. The decoder had 100% accuracy in all naïve and trained experiments when given access to the full ~4000-cell population. To understand why changes in individual neuronal tuning did not affect performance, we investigated which neurons the decoder selected to base its decision on.

The decoder based its output on a sparse subset of neurons (Fig. A1e-g), in both naïve and trained conditions. To show this, we measured the percentage of the discriminant function accounted for by each neuron's activity. The contribution of the recorded neurons to the discriminant function followed a power-law over the first ~100 neurons (Fig. 2e, f): the proportion of the discriminant function accounted by the n^{th} neuron was approximately proportional to $n^{-\alpha}$, where the scaling exponent α was -0.760 ± 0.040 in naïve subjects and -0.913 ± 0.071 in proficient subjects, reflecting a small but significant increase in slope with training ($p = 0.04$, paired t-test). The single best neuron accounted for 8.4 ± 1.6% (naïve) or 8.6 ± 0.65% (proficient) of the discriminant function, and the top 20 neurons (~0.5% of the recorded population) together accounted for 36.9 ± 3.1% and 46.7 ± 3.9% of the discriminant function (naïve and trained; $p < 0.05$, paired t-test). The decoder thus based its decision on a highly sparse set of neurons, which became slightly but significantly sparser after training. Importantly, the L2-regularization approach that we used (unlike L1-based methods[62]) does not preferentially seek sparse weights; the fact that it nevertheless found them indicates that a sparse subset of neurons encoded the stimulus in a particularly advantageous manner.

The neurons selected by the decoder were strongly selective between the two task stimuli and had low variability (Fig. A1g), and in both naïve and proficient subjects there were enough such neurons to produce perfect decoding. The cells picked by the decoders again responded less in proficient than in naïve mice ($p=0.05$, paired t-test), and showed higher selectivity ($p=0.002$, paired t-test), but with no significant change in variability or d′ ($p>0.05$; Fig. A1h-k). The increased sparsity of the ensembles selected by the decoder in proficient mice likely results from an increase in the fraction of extremely selective cells, allowing the decoder to focus on a smaller subset of highly selective cells than in the naïve case. To further demonstrate how accurately this sparse set of neurons encoded the stimulus, we sequentially added neurons to our model based on their cross-validated performance (i.e., sequential feature selection), limiting the number of total neurons in our model to 10. Remarkably, decoding from just one optimally-selected neuron yielded cross-validated performance of 99.5 ± 0.5% in naïve mice, 99.6 ± 0.4% in proficient (Fig. A1n, left; $p > 0.05$, paired t-test).

The 100% accuracy of stimulus decoding in naïve and trained conditions therefore arises because in both conditions there exists a sparse subpopulation of cells that encoded the stimulus extremely accurately. It remains possible however that a decoder denied access to these rare but exceptionally accurate neurons might work better in the trained condition. If so, this could constrain decoding both for downstream neurons in the brain, which might only have access to a subset of V1 axons, as well as to previous experiments which recorded from smaller populations.

We therefore asked if a difference between naïve and trained decodability might appear for randomly-selected cell pools, which will usually exclude the very best cells (Figure A1l). When decoding from one randomly chosen neuron performance was 59.4 ± 0.6% in naïve mice, 57.8 ± 0.9% in proficient (p = 0.076, paired t-test), and increased in both cases to reach an asymptote of 100% at around 400 random neurons. For no pool size did we see a significant difference between naïve and proficient conditions. We also assesses decoder performance by using the d' of the discriminant function, but again found no significant difference (Fig. A1m). We conclude that even for a decoder without access to the best neurons in the recorded population, decoding fidelity does not increase following task training.

In a final attempt to find a decoder whose performance is better for proficient than naïve mice, we again picked an optimal sparse subset of each random cell pool in a sequential manner (Fig. A1o). In each case, decoding reach asymptotic performance using just a few neurons, and once again no significant difference was found between the naïve and trained conditions (p>.05 in all cases).

We conclude that while the structure of the V1 population code for orientation changes following task training, coding fidelity does not significantly improve in proficient mice, even after considering multiple methods aimed at revealing such a difference.

Appendix 2

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[39,40] 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. For any scalar $\alpha, S_k[x_i] = S_k[\alpha x_i]$. 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 $\sum_i x_i = \sum_i y_i$. Thus, $S_k[y_i] \geq S_k[x_i]$ and the theorem is proved.