nature neuroscience

# Perceptual learning as improved probabilistic inference in early sensory areas

Vikranth R Bejjanki[1,4], Jeffrey M Beck[1,2,4], Zhong-Lin Lu[3] & Alexandre Pouget[1]

**Extensive training on simple tasks such as fine orientation discrimination results in large improvements in performance, a form of learning known as perceptual learning. Previous models have argued that perceptual learning is due to either sharpening and amplification of tuning curves in early visual areas or to improved probabilistic inference in later visual areas (at the decision stage). However, early theories are inconsistent with the conclusions of psychophysical experiments manipulating external noise, whereas late theories cannot explain the changes in neural responses that have been reported in cortical areas V1 and V4. Here we show that we can capture both the neurophysiological and behavioral aspects of perceptual learning by altering only the feedforward connectivity in a recurrent network of spiking neurons so as to improve probabilistic inference in early visual areas. The resulting network shows modest changes in tuning curves, in line with neurophysiological reports, along with a marked reduction in the amplitude of pairwise noise correlations.**



Extensive training on simple behavioral tasks such as vernier acuity or orientation discrimination leads to a gradual improvement in performance over several sessions, a form of learning known as perceptual learning[1–3]. This form of learning has been observed in a wide range of tasks and across many modalities, thus suggesting that it represents a general mechanism by which humans and animals improve their performance in response to task demands[4–9]. Although the behavioral consequences of such learning are well understood, there is debate in the literature as to the nature of the neural changes that underlie the observed behavioral changes.

Perceptual learning is typically very specific both in terms of the task itself and in terms of the location of the training[1,10–12] (but see refs. 13,14). For instance, in the visual domain, training on vernier acuity does not transfer to other tasks such as orientation discrimination or to the same vernier shown at a different retinal location[2]. This specificity, particularly the specificity to retinal location, suggests that this form of learning engages early visual areas where retinotopy is reasonably well preserved[15]. This is indeed consistent with single-cell recordings, which have revealed modifications in the response properties of cells in early and midlevel visual areas such as V1 and V4 after extensive training on orientation discrimination[8,9,16]. Two main types of changes have been reported: a sharpening and an amplification of tuning curves to orientation (**Fig. 1a**). Theoretical studies have argued that such changes could account for the observed behavioral improvement, as both types of changes increase the slope of the tuning curves, thereby increasing the ability of single neurons to discriminate orientation[17,18]. Moreover, at least one study has also proposed that perceptual learning leads to reduced internal noise through a lowering of the Fano factor (the ratio of the variance over the mean of spike counts) of visual cells[16], a change that could also potentially explain the behavioral improvement.

Although these studies offer a potential neural mechanism for perceptual learning in a fine orientation discrimination task, there are several open issues, and serious problems, with this perspective. First, the claim that tuning curve sharpening and amplification, or noise reduction, accounts for the observed behavioral improvement rests on the assumption that neural variability is independent before and after learning[17,18]. This assumption is problematic because neural variability is correlated *in vivo*[19–21] and these noise correlations are likely to change before and after learning if learning is due to changes in connectivity, as is assumed in these models[17,18]. This is all the more serious because, when noise correlations are taken into account, previous studies have shown that amplification and sharpening by means of lateral connections do not necessarily result in more informative neural representations, and therefore do not necessarily lead to behavioral improvements[22,23].

Second, it is unclear whether current neural theories can account for the effect of perceptual learning on what is known as the 'threshold versus external noise contrast', or TVC, curve, a comprehensive measure of human perceptual sensitivity that has been widely used to reveal observer characteristics in a wide range of auditory and visual tasks and changes of the perceptual limitations associated with cognitive, developmental and disease processes[24]. A TVC curve shows the perceptual threshold (detection or discrimination threshold, depending on the task) of a subject, as a function of the amount of external noise present in the stimulus (that is, the noise injected in the image on every trial; see top of **Fig. 1b**). When plotted on a log-log scale, this curve takes on a characteristic shape (**Fig. 1b**). For low levels of external noise, the perceptual threshold stays relatively constant as external noise increases because internal noise dominates. For large values of external noise, by contrast, the perceptual threshold increases linearly with the logarithm of the amplitude of the external noise because external noise is now the dominant source of variability.

[1]Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York, USA. [2]Gatsby Computational Neuroscience Unit, London, UK. [3]Department of Psychology, University of Southern California, Los Angeles, California, USA. [4]These authors contributed equally to this work. Correspondence should be addressed to A.P. (alex@bcs.rochester.edu).

Received 13 December 2010; accepted 4 March 2011; published online 3 April 2011; doi:10.1038/nn.2796

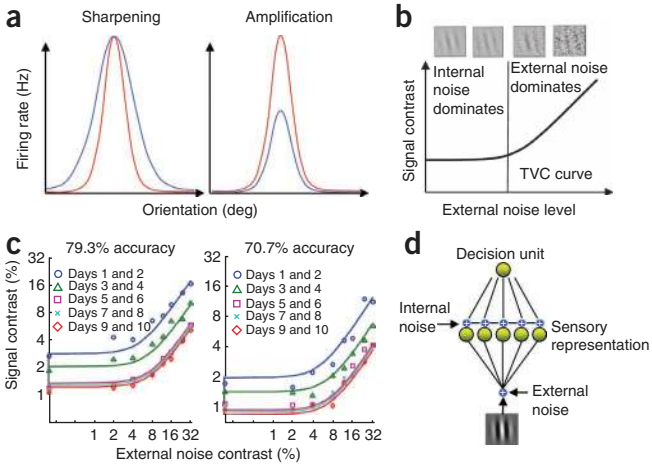


**Figure 1** Neural and behavioral correlates of perceptual learning. (**a**) An illustration of the two types of changes observed in the tuning curves of trained orientation-selective neurons. Each panel shows an illustration of a tuning curve before (blue) and after (red) learning. The *x* axis represents orientation (in degrees) and the *y* axis represents firing rate (in Hz). (**b**) The characteristic shape of a threshold-versus-contrast, or TVC curve, plotted on a log-log scale, showing the two main regimes. The *x* axis represents the external noise added to the stimulus (example stimuli are shown in the top panel). The *y* axis represents the signal contrast needed to elicit the specific level of performance. (**c**) Observed changes in TVC curves as a result of perceptual learning, at two levels of performance[27]. Signal contrasts needed to elicit a specific level of performance, estimated at each of eight levels of external noise, are shown averaged over pairs of days. Smooth TVC curves are fits of the Perceptual Template Model (PTM)[27]. The axes are the same as in **b**. (**d**) Standard model of perceptual learning. The image is preprocessed by a set of filters corrupted by noise. The noisy output of these filters is then fed into a single decision unit trained to optimize discrimination between two orientations.

Several experiments have shown that TVC curves change in a very specific manner during perceptual learning: the entire curve shifts down by a constant amount from one training session to the next (**Fig. 1c**)[4,25–27]. As argued by some authors, such a uniform shift in TVC curves is consistent with a 'late' theory of perceptual learning[4,24–28]. This conclusion relies on an engineering-inspired model containing two stages of processing: a sensory processing stage typically composed of a set of oriented spatial filters corrupted by additive and multiplicative noise, followed by a decision stage (**Fig. 1d**). In this class of models, the uniform shift in TVC curves is best explained by changes to the connections between the sensory representation and the decision stage, as opposed to a change in the early sensory representation[29]. Such results are consistent with a 'late' theory of perceptual learning, as cortical areas involved in decision making, such as the lateral intraparietal area (LIP) and the prefrontal cortex (PFC), are late in the hierarchy of cortical processing. This is also consistent with a recent report documenting a change in the response properties of LIP neurons as an animal learns a motion discrimination task[30,31].

The problem with this conclusion, however, is that training on orientation discrimination has been shown to cause changes in the response of neurons in areas V1 and V4, two areas that are not thought to be implicated in decision making. Moreover, neurally inspired models that have been developed to capture the learning-induced changes in TVC curves assume noise sources, such as multiplicative and additive noise, that cannot easily be mapped onto neural variability[4]. Indeed, although neuronal responses are known to be variable, this variability is neither additive nor multiplicative, but rather seems to show properties that are close to the Poisson distribution[32].

In this paper, we show that all these perspectives can be reconciled when we consider a neural model of perceptual learning with realistic response statistics. In such a model, learning can be implemented in early visual areas in a manner that captures all the main features of the observed TVC curve changes, while also explaining the retinal specificity of learning. In addition, the model reveals that the key to learning is not to sharpen or amplify tuning curves, as these changes are neither sufficient nor necessary for learning. Instead, the key is to improve the efficiency of probabilistic inference in early cortical circuits by adjusting the feedforward weights in a manner that brings them closer to a matched filter. We show that at the neural level, this weight adjustment has only a minor effect on the shape of the tuning curves (as has been found *in vivo*) but leads to a large decrease in the magnitude of pairwise noise correlations.

## RESULTS

### Template matching in primary visual cortex

Several aspects of the neural architecture used in our simulations (**Fig. 2a**; see **Supplementary Note**) are based on previous models of orientation discrimination, particularly the models in refs. 33 and 22. The model consists of three layers: retina, lateral geniculate nucleus (LGN) and V1. The retinal layer corresponds to grids of ON- and OFF-center ganglion cells modeled by difference-of-Gaussian filters. The output of each filter is passed through a smooth nonlinearity and used to drive the LGN cells, which generate Poisson spikes. The output spikes from the LGN cells are pooled using oriented Gabor-function (a product of cosine and Gaussian functions) receptive fields, the orientations of which are uniformly distributed around a circle. The pooled output from the LGN cells is then used as input to V1. The V1 stage represents an orientation hypercolumn—a set of neurons with receptive fields centered at the same spatial location but with different preferred orientations—of linear-nonlinear-Poisson (LNP) neurons, coupled through lateral connections (see Online Methods).

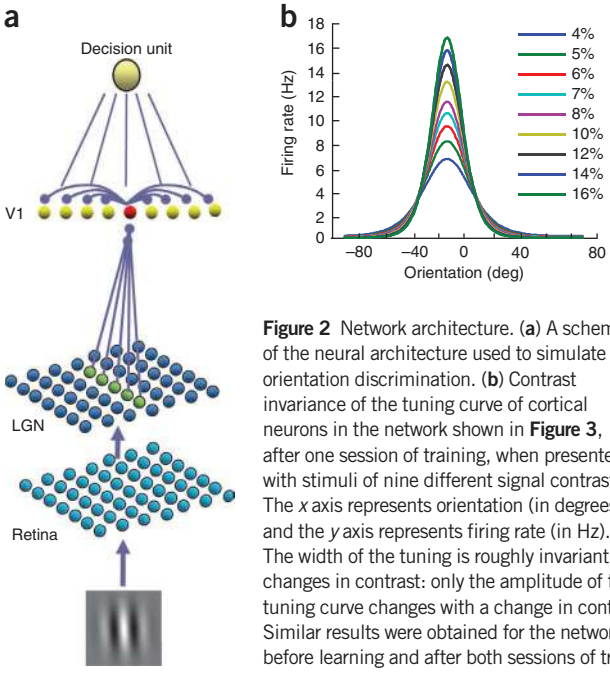


**Figure 2** Network architecture. (**a**) A schematic of the neural architecture used to simulate orientation discrimination. (**b**) Contrast invariance of the tuning curve of cortical neurons in the network shown in **Figure 3**, after one session of training, when presented with stimuli of nine different signal contrasts. The *x* axis represents orientation (in degrees) and the *y* axis represents firing rate (in Hz). The width of the tuning is roughly invariant to changes in contrast: only the amplitude of the tuning curve changes with a change in contrast. Similar results were obtained for the network before learning and after both sessions of training.

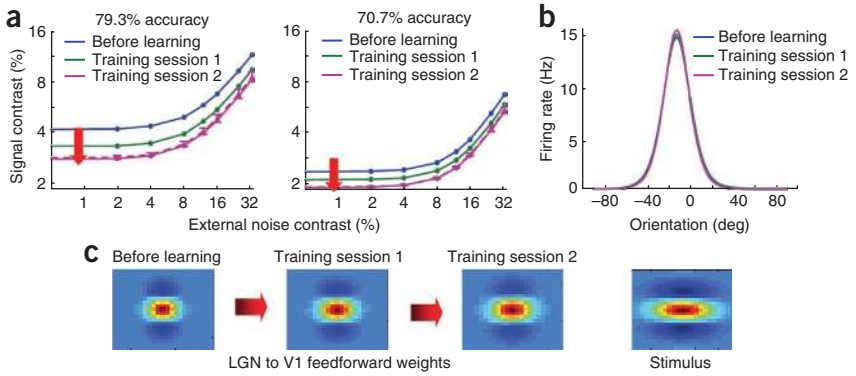


**Figure 3** Modeling perceptual learning using a realistic neural model of orientation discrimination. (**a**) Replicating the uniform shifts in TVC curves (**Fig. 1b**,**c**) using the neural model of orientation discrimination. Feedforward connections between the LGN and V1 were adjusted in a manner that moved them toward a matched filter for the stimulus. After training, we reran ten new simulations with ±10% independent noise added to the final (training session 2) feedforward thalamo-cortical weights. Pink dashed line, average TVC curve across the ten runs; error bars, 1 s.d. (**b**) Tuning curves of cortical neurons from the network, demonstrating modest amplification and sharpening as a result of learning. (**c**) Moving the thalamo-cortical feedforward weights toward a matched filter. The rightmost panel shows the two-dimensional spatial profile of a stimulus. The leftmost panel shows the two-dimensional spatial profile of the feedforward weights before learning, the second panel from left shows the spatial profile of the feedforward weights after one training session and the panel second from the right shows the spatial profile of the feedforward weights after two training sessions. Together these figures show the spatial profile of the thalamo-cortical feedforward connections moving toward the spatial profile of the stimulus, a manipulation that led to the changes in TVC curves shown in **a**.



The lateral connections are tuned to ensure that the resulting orientation tuning curves are contrast invariant. In other words, changing the contrast of the image only affects the gain of the cortical response (**Fig. 2b**), while keeping the width of the tuning curves constant, as has been reported in the primary visual cortex[34]. We constrained our parameter search to preserve contrast invariance in all the networks discussed in this paper.

This model was used to simulate the orientation discrimination task used in refs. 4,27. In those experiments, subjects were asked to report the orientation (clockwise or counterclockwise) of Gabor

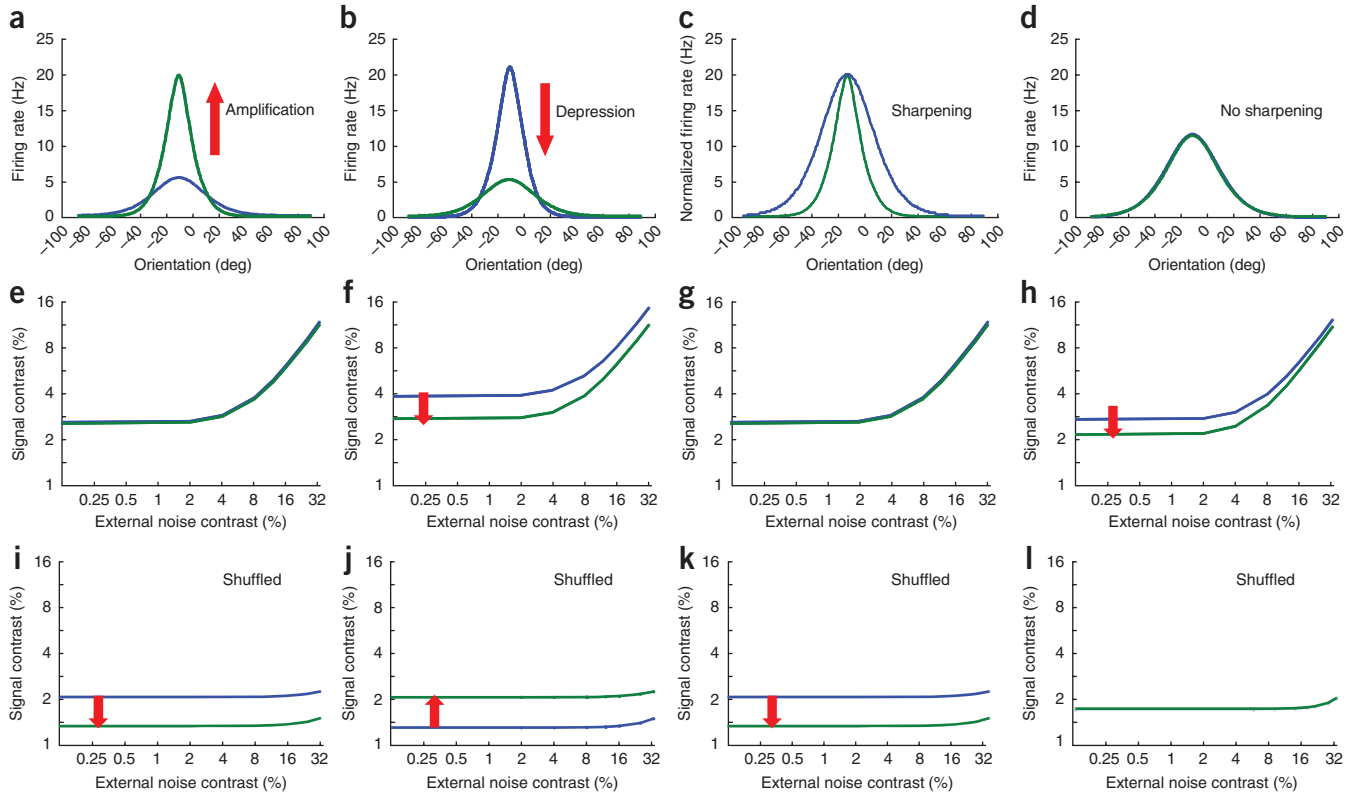


**Figure 4** Perceptual learning and tuning curves: the role of amplification and sharpening. (**a**–**l**) Taking noise correlations into account demonstrates that amplification and sharpening are neither necessary nor sufficient for learning. (**a**,**e**,**i**) Amplification is not sufficient for learning. Parameters in this network were changed in a manner that led to amplification of the orientation tuning curves (**a**). At the level of TVC curves, the same change led to no improvement in performance—the TVC curve did not shift (**e**)—thereby showing that amplification is not sufficient for learning. Single-cell recordings in such a network would incorrectly conclude that performance improved during training, as illustrated by the drop of the TVC curve when computed with $I_{shuffled}$ (**i**). (**b**,**f**,**j**) Amplification is not necessary for learning. Performance can improve in a network (**f**) in which the gain of the tuning curve decreases during learning (**b**). (**c**,**g**,**k**) Sharpening is not sufficient for learning. This network shows no change in performance (**g**) despite substantial sharpening of the tuning curves (**c**). (**d**,**h**,**l**) Sharpening is not necessary for learning. Performance can improve (**h**) even in the absence of any sharpening (**d**). All TVC curves were obtained for the 89% correct performance criterion.

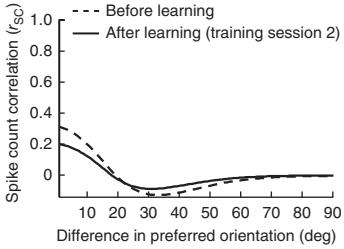


**Figure 5** The effect of perceptual learning on noise correlations. The dashed curve shows the correlation coefficients as a function of the difference in preferred orientation before learning. The correlation coefficients were computed using a similar procedure to that used in ref. 38, in response to a stimulus with 8% signal contrast. As has been reported *in vivo*[20,37,38], correlations are between 0.4 and –0.1, and their magnitude decreases as a function of orientation difference. The solid curve shows the correlations after learning took place. The main effect of perceptual learning is to reduce the magnitude of the correlations while preserving the overall pattern of correlations.



patches corrupted by pixel noise and oriented at either −12° or 12° from vertical (**Fig. 1b**). To model this task, we added a decision stage to our network in the form of a linear classifier. The linear classifier is equivalent to a decision unit whose activity is determined by the dot product of a weight vector with the population activity in the cortical layer (**Fig. 2a**). The weights were tuned to optimize classification performance in the pre-training condition and were left untouched thereafter (see Online Methods). Network TVC curves were obtained using an analytical approach combined with numerical simulations. Specifically, we derived a lower bound on Fisher information (which is proportional to the inverse of the discrimination threshold of an ideal observer) in the cortical layer, as a function of specific network parameters such as the feedforward and recurrent connectivity, using our recent work on computing Fisher information in recurrent networks of spiking neurons[35]. We then used this expression to derive Fisher information in the decision unit. Finally, we numerically estimated TVC curves using the expression for Fisher information in the decision unit, taking advantage of the inverse relation between Fisher information and the discrimination threshold[19] (see Online Methods).

As several neural models of early perceptual learning modify lateral connections[17,18], we first explored whether such changes could capture the main features of perceptual learning, such as the uniform shift of the TVC curves. Despite an extensive parameter search (see **Supplementary Note**), this approach failed: sharpening tuning curves by adjusting lateral connections often resulted in worse performance, and in an upward shift of the TVC curve, because sharpening tuning curves often modified the noise correlations between neurons in a way that decreased Fisher information. In the few cases in which the TVC curves shifted downward, we could never find a configuration in which the curves shifted uniformly. Of course, given the size of the parameter space, we cannot rule out the possibility that there exists a solution based only on changes to the lateral connections that accounts for perceptual learning. However, we can conclude that changes to lateral connections that result in sharpening or amplification do not necessarily result in behavioral improvement, in contrast to what previous models have suggested.

Next, we considered changes to feedforward connections (**Fig. 3**) and found that this approach could indeed shift the TVC curves nearly uniformly (**Fig. 3a**; see **Supplementary Table 1** for parameter values). Notably, the ratio of any two TVC curves between training sessions was approximately constant across external noise levels (with a maximum value of 1.26 and a minimum value of 1.23), as has been reported experimentally[27]. Moreover, the ratios of any two TVC curves between criterion levels were approximately constant across noise levels (1.81 ± 0.05 in the pre-training session, 1.63 ± 0.07 in training session 1 and 1.53 ± 0.08 in training session 2) and quantitatively similar to values observed experimentally[27] (1.82 ± 0.51 for the TVC curve fit to the average thresholds over days 1 and 2, 1.67 ± 0.22 for the TVC curve fit to the average thresholds over days 3 and 4 and 1.27 ± 0.15 for the TVC curve fit to the average thresholds over days 5 and 6). Crucially, to get these results, the spatial profile of the feedforward weights, between LGN and V1, had to be changed so as to match more closely the spatial profile of the stimulus (**Fig. 3c**). We repeated the simulations for three new sets of initial weights and found that the TVC curves always shifted uniformly as long as the weights were moved toward a matched filter (**Supplementary Fig. 1**). Our results were also robust to variability in the value of the final weights. Adding independent noise of ±10% to the final weights resulted in variability in the TVC curves (error bars in **Fig. 3a**), but the shift remains close to uniform.

These results show that it is possible to capture the main features of perceptual learning–induced TVC curve changes by adjusting the feedforward connectivity in early visual areas. This solution has the added advantage of capturing the specificity of perceptual learning to retinal location because it involves changing weights at a specific location on the retinotopic map.

## The role of amplification and sharpening

The amplitude of the tuning curve of a trained cortical neuron grows slightly with training, and the width is slightly reduced, thereby revealing modest amounts of amplification and sharpening (**Fig. 3b**). These changes are consistent with neurophysiological recordings in animals trained on orientation discrimination, which have reported that the tuning curves in both V1 and V4 are sharpened and/or amplified during perceptual learning[8,9,16]. However, these effects have been reported to be small, particularly in V1. Indeed, some studies did not find any appreciable sharpening or amplification[36], whereas others found only a small amount of sharpening or amplification[8,9], as is the case in our network.

Nonetheless, the fact that the tuning curves in our model show amplification and sharpening would seem to be consistent with conclusions from previous models that have invoked these mechanisms as the neural basis of perceptual learning[17,18]. Yet this would be misleading because, in our model, amplification and sharpening are neither

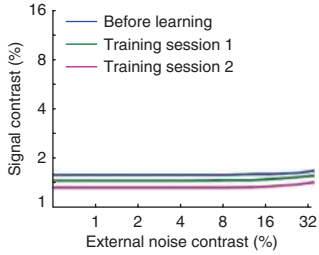


**Figure 6** TVC curves computed from responses in which noise correlations have been removed through shuffling. Curves for the network shown in **Figure 3**, with correlations removed using $I_{shuffled}$ (see main text). The TVC curves still shift downward by a uniform amount, as was the case in **Figure 3a**. However, the magnitude of shift is significantly reduced and the TVC curves no longer show the characteristic linear increase in signal contrast as external noise increases, at high levels of external noise. These curves were computed for 79.3% accuracy.

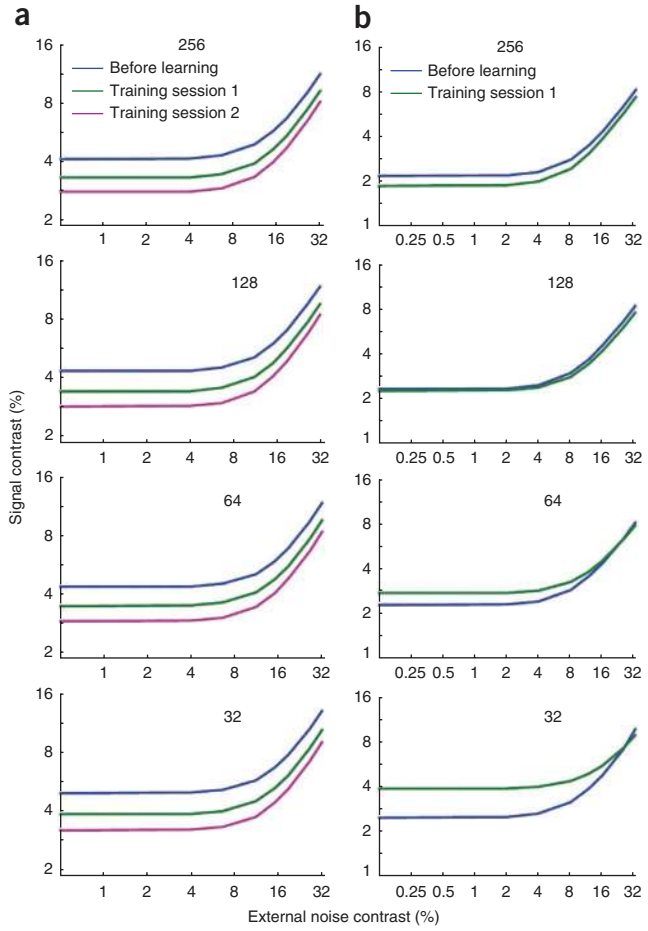


**Figure 7** The effect of subsampling. (**a**,**b**) TVC curves obtained from subsets of neurons in the networks shown in **Figure 3** (**a**) and **Figure 4b**,**f**,**j** (**b**). The plots in each panel were obtained by simulating the full network, but computing the TVC curve based on the response of a randomly picked sample of neurons. Number of neurons in each subset is indicated on each panel and varies from 256 to 32. (**a**) TVC curves obtained by subsampling the network shown in **Figure 3**. The results are qualitatively the same, even when we sample as few as 32 neurons. (**b**) TVC curves obtained by subsampling the network shown in **Figure 4b**,**f**,**j**. With 256 neurons, the results are similar to the results obtained with the full network (**Fig. 4f**), whereas with only 32 neurons, the result starts mimicking the results obtained with $I_{shuffled}$ (**Fig. 4j**). The upward shift of the TVC curve with learning, observed with 32 cells, is therefore an artifact of subsampling and does not reflect the behavior of the whole neuronal population. All TVC curves were obtained for the 79% correct performance criterion.

sufficient nor necessary to account for perceptual learning (**Fig. 4**). For instance, with appropriate choice of parameters, it is possible to find a network in which the tuning curves are amplified (**Fig. 4a**) but in which the TVC curve shows no shift (**Fig. 4e**), corresponding to no change in performance and a lack of perceptual learning. This indicates that amplification is not sufficient for perceptual learning. The opposite scenario is also possible (**Fig. 4b**): in this case, the tuning curves show depression, as opposed to amplification, but the TVC curve shifts downward (**Fig. 4f**) in line with changes observed during perceptual learning. This implies that amplification is not necessary for learning. Equivalently, sharpening, like amplification, is neither sufficient (**Fig. 4c**,**g**) nor necessary (**Fig. 4d**,**h**) for learning. (See **Supplementary Table 2** for parameter values.)

These results demonstrate that single cell responses alone are insufficient to predict behavioral performance. To get a comprehensive neural theory of perceptual learning, one must also consider the correlations between cells.

## The role of noise correlations

Noise correlations, before learning, were within the range −0.1 to 0.4 (**Fig. 5**), which is consistent with the values reported *in vivo*[20,37,38]. Moreover, correlations tended to decrease with the difference in preferred orientations between pairs of cells, as is typically found in models of orientation selectivity and *in vivo*[20,38]. After learning, correlation coefficients showed the same dependence on the difference in preferred orientation but the overall amplitude of the coefficients was reduced (**Fig. 5**).

It would be tempting to conclude that the increase in information is due to this general decrease in correlations, but one must be cautious with such conclusions. Just as sharpening the tuning curve does not guarantee an increase in information, neither does a decrease in correlation coefficients. The increase in information is determined by a combination of a change in the pattern of correlations and a change in the shape of the tuning curves. To determine more quantitatively the contribution of correlations to the increase in Fisher information in our network, we used a recently proposed metric[20]. We compared the increase in information across training sessions for the network shown in **Figure 3** against the change in information in a virtual population of neurons with the same change in correlations but with identical tuning curves across training sessions. This analysis revealed that correlations contribute to 70% of the increase in Fisher information, indicating that correlations are responsible for much of the performance improvement.

In addition, to illustrate the kind of errors that one would encounter by ignoring correlations, we plotted the TVC curves for all the networks we have described so far, but using an information theoretic quantity called $I_{shuffled}$: the Fisher information in a set of neurons with the same single-cell response statistics as in the original data but without correlations[19]. This is effectively the measure used in some previous models[17,18] and in neurophysiological studies using single-cell recordings.

This analysis revealed that the changes in TVC curves derived from $I_{shuffled}$ do not necessarily reflect changes in the TVC curves derived from the true Fisher information (**Fig. 4i–l**). For instance, consider a network in which the tuning curve sharpens as a result of training (**Fig. 4c**,**g**,**k**). As one would expect from previous work on population codes with independent noise, a sharpening of the tuning curves increases $I_{shuffled}$, which results in a downward shift of the TVC curve (**Fig. 4k**). Yet the true TVC curve for this particular network does not shift during training, indicating no change in performance (**Fig. 4g**). In addition, the TVC curve derived from $I_{shuffled}$ does not have the right profile: it remains basically flat over the range of external noise values tested (**Fig. 4k**). In a different network, one in which the tuning curves are depressed as a result of training (**Fig. 4b**,**f**,**j**), the shift in the TVC curve derived from $I_{shuffled}$ differs from the shift in the true TVC curve: the TVC curve derived from $I_{shuffled}$ shifts upward (**Fig. 4j**), whereas the true TVC curve shifts downward (**Fig. 4f**).

Finally, we also computed the TVC curves from $I_{shuffled}$ for the network shown in **Figure 3**, in which we modify the feedforward connections. In this network, the TVC curves from $I_{shuffled}$ move in the right direction, but the curves are much too flat compared to the real ones, and the amount by which they move is considerably less than the shift obtained with the true information (**Fig. 6**). Therefore, although the TVC curves derived from $I_{shuffled}$ move in

the right direction in this case, they still do not accurately reflect the TVC curves obtained from Fisher information.

### Subsampling neurons: how many does it take?

Our results suggest that the neural basis of perceptual learning can only be revealed by recording the tuning curves and noise correlations of all the neurons involved in the task within a cortical area. However, such recordings are not available. Although multi-electrode arrays make it possible to record from many neurons simultaneously, this technique can only record the responses from a small fraction of all the neurons present in a given cortical patch[38,39].

To determine whether the same conclusions would be reached in our simulated networks when recording from fewer neurons, we derived the TVC curves from subsets of neurons. More specifically, we simulated each of the networks considered so far with 256 neurons in the cortical layer, but we computed information from the responses of only a subset of randomly sampled cortical neurons. We then generated TVC curves based on this measure of information. The TVC curves derived for the network shown in **Figure 3**, based on the responses of 256, 128, 64 and 32 neurons, are qualitatively the same, even when we sample as little as 32 neurons (**Fig. 7a**).

However, this does not hold for all networks, as can be seen with the TVC curves derived for the network shown in **Figure 4b**,**f**,**j** using subsets of varying size (**Fig. 7b**). In this network, training led to a depression of the tuning curve amplitude and the TVC curves derived from the true Fisher information with all 256 neurons shifted downward, indicating a performance improvement. In contrast, the TVC curves estimated from 32 neurons show the reverse trend: performance is degraded as a result of training (**Fig. 7b**).

These results suggest that, in some networks, when we only consider the responses of a subset of a population of cortical neurons, we may not retain all the qualitative results obtained by considering the entire population, even when we have enough data to accurately characterize the correlations (see **Supplementary Fig. 2** for more examples). This in turn implies that a discrepancy between behavioral results and the information content of a subpopulation of neurons should not necessarily be taken as evidence that the recorded area does not play a role in learning.

## DISCUSSION

We explored the neural basis for the improvement in behavioral performance observed during perceptual learning in a fine orientation discrimination task. Classical neural theories have argued that perceptual learning is mediated through a steepening—by means of amplification or sharpening—of the tuning curves of neurons in early sensory areas. Further, they have proposed that tuning curves are steepened through changes in the lateral connections between neurons in early cortical sites[17,18]. Our analysis suggests a different picture: given the observed changes in TVC curves, learning can be explained by changes in the feedforward connections between the thalamus and the primary visual cortex. More specifically, the feedforward connections must be modified in a manner that moves them toward a matched filter for the stimulus. These modifications also result in modest changes in tuning curves, in line with what has been reported *in vivo*, and in a decrease in the magnitude of pairwise noise correlations. To our knowledge, this is the first neural theory of perceptual learning that is consistent with the mean change in neural response in early visual areas, with the response statistics of sensory neurons, and with the uniform downward shift in TVC curves documented in human observers.

One of the limitations of some previous theories of perceptual learning has been the assumption that neural variability remained the same before and after learning[28,31]. Under this assumption, steepening tuning curves can indeed improve behavioral performance. In contrast, we have shown here that when learning-induced changes in correlations are taken into account, steepening of tuning curves is neither necessary nor sufficient for learning. Therefore, the key to the neural basis of perceptual learning may have less to do with how tuning curves change and more to do with how the connectivity is adjusted to improve the inference performed by neural circuits. Such changes to neural connectivity can still affect tuning curves, but it is the combination of the change in tuning curves and the change in correlations that ultimately determines the presence or absence of learning.

Although this paper focuses on changes in early sensory areas, our work is conceptually related to late theories of perceptual learning. In late theories of perceptual learning, neural processing is decomposed into two stages: a sensory processing stage followed by a readout or decision stage. The readout stage is typically formalized as a probabilistic inference stage, whose goal is to compute the probability distribution over possible choices given the sensory evidence[28,31,40]. During learning, the weights between the early and late areas are adjusted so as to bring the probability distribution over choices closer to the optimal posterior distribution that would be obtained by applying Bayes' rule to the output of the early sensory stage, which in turn leads to a uniform shift in TVC curves. This perspective is in line with several neurophysiological and modeling results that strongly suggest that neurons in areas such as LIP encode probability distributions, or likelihood functions, over choices and actions[41,42].

Here we argue that perceptual learning might be due to improved probabilistic inference induced by changes at the sensory processing stage rather than at the decision stage (at least in the case of orientation discrimination). This possibility has been ignored in the past, in part because it is not common to think of early visual areas as performing probabilistic inference. Instead, the responses of neurons in early sensory areas are typically modeled as noisy nonlinear filters encoding scalar estimates of variables such as orientation. Recent theories[43,44] have challenged this perspective and suggested that, even in early sensory areas, neural patterns of activity might in fact represent probability distributions or likelihood functions. This raises the possibility that learning in early visual areas involves an improvement in probabilistic inference that could in turn lead to a uniform shift in TVC curves. This is very much the logic that we have pursued here. Therefore, our solution is computationally very similar to the one proposed by late theories. Indeed, our model does not supersede previous models of late perceptual learning[28,31,40] but instead complements them by showing how similar results can be obtained by modifying population codes in early sensory areas. Our approach, however, has one significant advantage: it accounts for the fact that neural responses change in early and midlevel visual areas as a result of training on an orientation discrimination task[8,9,16], an experimental observation that is difficult to reconcile with late theories.

Ultimately, whether perceptual learning involves early or late stages (or both) is likely to depend on the task, the sensory modalities involved and the nature of the feedback and training received by subjects. For example, in the somatosensory and auditory systems, large changes have been reported in primary sensory cortices as a result of perceptual learning[6,7]. In the visual system, by contrast, whereas training on orientation discrimination seems to engage early and midlevel areas such as V1 and V4 (refs. 8,9,16,45), training on other tasks, such as motion discrimination, triggers changes in late areas such as LIP[30]. Moreover, double training on contrast and orientation discrimination seems to trigger learning in both early and late areas, with nearly complete transfer of learning across trained locations[13,14]. Our approach could be extended to all of these forms of learning, as none of what we have presented is specific to the LGN-V1 architecture used here.

Finally, although we have focused on perceptual learning, our approach can be generalized to other domains such as adaptation or attention. In the case of attention, our approach (as well as our previous work; see ref. 46) would suggest that changes in noise correlations might be critical in the behavioral improvement associated with enhanced attention, perhaps more so than changes in tuning curves, which are typically the focus of single-cell studies[47,48]. Of note, it was recently reported[20] that changes in correlations account for 86% of the behavioral improvement triggered by attention. Likewise, another recent study[37] found that correlations are critical for adaptation. It will be important to investigate whether these neural changes can also account for the changes in TVC curves induced by attention or adaptation.

## METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

*Note: Supplementary information is available on the Nature Neuroscience website.*


### ACKNOWLEDGMENTS

J.M.B. was supported by the Gatsby Charitable Foundation. Z.-L.L. is supported by US National Eye Institute grant 9 R01 EY017491-05 and A.P. by Multidisciplinary University Research Initiative grant N00014-07-1-0937, US National Institute on Drug Abuse grant BCS0346785 and a research grant from the James S. McDonnell Foundation. This work was also partially supported by award P30 EY001319 from the US National Eye Institute.


### AUTHOR CONTRIBUTIONS

V.R.B. conceived the project, built the network model, ran all the simulations and analyses and wrote the paper. J.M.B. developed the analytic derivations, helped with building the network model and wrote the paper. Z.-L.L. worked on the link between the neural model and TVC curves and helped with parameter tuning. A.P. conceived the project, supervised the simulations and analyses and wrote the paper.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/natureneuroscience/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.




1. Ramachandran, V.S. & Braddick, O. Orientation-specific learning in stereopsis. *Perception* **2**, 371–376 (1973).
2. Fahle, M., Edelman, S. & Poggio, T. Fast perceptual learning in hyperacuity. *Vision Res.* **35**, 3003–3013 (1995).
3. Fiorentini, A. & Berardi, N. Perceptual learning specific for orientation and spatial frequency. *Nature* **287**, 43–44 (1980).
4. Dosher, B.A. & Lu, Z.L. Mechanisms of perceptual learning. *Vision Res.* **39**, 3197–3221 (1999).
5. Fine, I. & Jacobs, R.A. Comparing perceptual learning tasks: A review. *J. Vis.* **2**, 190–203 (2002).
6. Recanzone, G.H., Merzenich, M.M., Jenkins, W.M., Grajski, K.A. & Dinse, H.R. Topographic reorganization of the hand representation in cortical area 3b owl monkeys trained in a frequency-discrimination task. *J. Neurophysiol.* **67**, 1031–1056 (1992).
7. Recanzone, G.H., Schreiner, C.E. & Merzenich, M.M. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.* **13**, 87–103 (1993).
8. Schoups, A., Vogels, R., Qian, N. & Orban, G. Practising orientation identification improves orientation coding in V1 neurons. *Nature* **412**, 549–553 (2001).
9. Yang, T. & Maunsell, J.H.R. The effect of perceptual learning on neuronal responses in monkey visual area V4. *J. Neurosci.* **24**, 1617–1626 (2004).
10. Ahissar, M. & Hochstein, S. Task difficulty and the specificity of perceptual learning. *Nature* **387**, 401–406 (1997).
11. Fahle, M. & Morgan, M. No transfer of perceptual learning between similar stimuli in the same retinal position. *Curr. Biol.* **6**, 292–297 (1996).
12. Karni, A. & Sagi, D. Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proc. Natl. Acad. Sci. USA* **88**, 4966–4970 (1991).
13. Xiao, L.Q. *et al.* Complete transfer of perceptual learning across retinal locations enabled by double training. *Curr. Biol.* **18**, 1922–1926 (2008).
14. Zhang, J.-Y. *et al.* Rule-based learning explains visual perceptual learning and its specificity and transfer. *J. Neurosci.* **30**, 12323–12328 (2010).
15. Watanabe, T. *et al.* Greater plasticity in lower-level than higher-level visual motion processing in a passive perceptual learning task. *Nat. Neurosci.* **5**, 1003–1009 (2002).
16. Raiguel, S., Vogels, R., Mysore, S.G. & Orban, G.A. Learning to see the difference specifically alters the most informative V4 neurons. *J. Neurosci.* **26**, 6589–6602 (2006).
17. Schwabe, L. & Obermayer, K. Adaptivity of tuning functions in a generic recurrent network model of a cortical hypercolumn. *J. Neurosci.* **25**, 3323–3332 (2005).
18. Teich, A.F. & Qian, N. Learning and adaptation in a recurrent model of V1 orientation selectivity. *J. Neurophysiol.* **89**, 2086–2100 (2003).
19. Averbeck, B.B., Latham, P.E. & Pouget, A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7**, 358–366 (2006).
20. Cohen, M.R. & Maunsell, J.H.R. Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* **12**, 1594–1600 (2009).
21. Zohary, E., Shadlen, M.N. & Newsome, W.T. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* **370**, 140–143 (1994).
22. Seriès, P., Latham, P.E. & Pouget, A. Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. *Nat. Neurosci.* **7**, 1129–1135 (2004).
23. Spiridon, M. & Gerstner, W. Effect of lateral connections on the accuracy of the population code for a network of spiking neurons. *Network* **12**, 409–421 (2001).
24. Levi, D.M. & Klein, S.A. Noise provides some new signals about the spatial vision of amblyopes. *J. Neurosci.* **23**, 2522–2526 (2003).
25. Gold, J., Bennett, P.J. & Sekuler, A.B. Signal but not noise changes with perceptual learning. *Nature* **402**, 176–178 (1999).
26. Li, R.W., Levi, D.M. & Klein, S.A. Perceptual learning improves efficiency by re-tuning the decision 'template' for position discrimination. *Nat. Neurosci.* **7**, 178–183 (2004).
27. Dosher, B.A. & Lu, Z.L. Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proc. Natl. Acad. Sci. USA* **95**, 13988–13993 (1998).
28. Petrov, A.A., Dosher, B.A. & Lu, Z.L. The dynamics of perceptual learning: an incremental reweighting model. *Psychol. Rev.* **112**, 715–743 (2005).
29. Lu, Z.-L., Liu, J. & Dosher, B.A. Modeling mechanisms of perceptual learning with augmented Hebbian re-weighting. *Vision Res.* **50**, 375–390 (2010).
30. Law, C.-T. & Gold, J.I. Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat. Neurosci.* **11**, 505–513 (2008).
31. Law, C.-T. & Gold, J.I. Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nat. Neurosci.* **12**, 655–663 (2009).
32. Tolhurst, D.J., Movshon, J.A. & Dean, A.F. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* **23**, 775–785 (1983).
33. Somers, D.C., Nelson, S.B. & Sur, M. An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.* **15**, 5448–5465 (1995).
34. Ferster, D. & Miller, K.D. Neural mechanisms of orientation selectivity in the visual cortex. *Annu. Rev. Neurosci.* **23**, 441–471 (2000).
35. Beck, J.M., Bejjanki, V.R. & Pouget, A. Insights from a simple expression for linear Fisher information in a recurrently connected population of spiking neurons. *Neural Comput.* (in the press).
36. Ghose, G.M., Yang, T. & Maunsell, J.H.R. Physiological correlates of perceptual learning in monkey V1 and V2. *J. Neurophysiol.* **87**, 1867–1888 (2002).
37. Gutnisky, D.A. & Dragoi, V. Adaptive coding of visual information in neural populations. *Nature* **452**, 220–224 (2008).
38. Smith, M.A. & Kohn, A. Spatial and temporal scales of neuronal correlation in primary visual cortex. *J. Neurosci.* **28**, 12591–12603 (2008).
39. Montani, F., Kohn, A., Smith, M.A. & Schultz, S.R. The role of correlations in direction and contrast coding in the primary visual cortex. *J. Neurosci.* **27**, 2338–2348 (2007).
40. Jacobs, R.A. Adaptive precision pooling of model neuron activities predicts the efficiency of human visual learning. *J. Vis.* **9**, 22 (2009).
41. Beck, J.M. *et al.* Probabilistic population codes for Bayesian decision making. *Neuron* **60**, 1142–1152 (2008).
42. Gold, J.I. & Shadlen, M.N. Neural computations that underlie decisions about sensory stimuli. *Trends Cogn. Sci.* **5**, 10–16 (2001).
43. Ma, W.J., Beck, J.M., Latham, P.E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nat. Neurosci.* **9**, 1432–1438 (2006).
44. Simoncelli, E.P., Adelson, E.H. & Heeger, D.J. Probability distributions of optical flow. in *IEEE Conference on Computer Vision and Pattern Recognition* 310–315 (1991).
45. Hua, T. *et al.* Perceptual learning improves contrast sensitivity of V1 neurons in cats. *Curr. Biol.* **20**, 887–894 (2010).
46. Pouget, A., Deneve, S. & Latham, P.E. The relevance of Fisher information for theories of cortical computation and attention. in *Visual Attention and Cortical Circuits* (eds. Braun, J., Koch, C. & Davis, J.L.) (2001).
47. Spitzer, H., Desimone, R. & Moran, J. Increased attention enhances both behavioral and neuronal performance. *Science* **240**, 338–340 (1988).
48. Treue, S. & Maunsell, J.H.R. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* **382**, 539–541 (1996).

## ONLINE METHODS

**Stimulus design.** We generated stimulus displays that mimicked those used in earlier perceptual learning experiments[4,27]. The signals were Gabor patterns tilted $\theta_s$° to the right or left of vertical. Each stimulus image was created by assigning grayscale values to image pixels according to the following function:

$$Z(x,y,\theta) = Z_0\left(1.0 + \left(c\,e^{-\left[\frac{C_x^2}{2\sigma_x^2}+\frac{C_y^2}{2\sigma_y^2}\right]}\cos(2\pi K C_x)\right)\right) \quad (1)$$

where

$C_x = x\cos\theta + y\sin\theta$
$C_y = y\cos\theta - x\sin\theta$
$\theta = \mathrm{rad}(90 \pm \theta_s)$

and $x$ and $y$ are the horizontal and vertical coordinates respectively, $K$ is the spatial frequency of the Gabor pattern, $\sigma_x$ and $\sigma_y$ are the s.d. (extent) of the Gabor pattern in the $x$ and $y$ directions respectively, $Z_0$ is the mean, or background, grayscale value for the image and $c$ is the maximum contrast of the Gabor pattern as a proportion of the maximum achievable contrast. In all our experimental conditions, $\theta_s$ was set to 12°, the spatial frequency $K$ was set to 0.75 cycles per degree, $\sigma_x$ was set to 0.4, $\sigma_y$ was set to 0.4 and $Z_0$ was set to 126.22, which was the equivalent mean grayscale value used in the earlier perceptual learning studies[4,27]. The maximum contrast of the Gabor, labeled the signal contrast, varied depending on the experimental condition.

Pixel gray levels for the external noise were drawn from a Gaussian distribution with mean 0 and s.d. depending on the experimental condition. As in the earlier studies[4,27], we used eight external noise levels in which the s.d. of the external noise distribution was varied as a proportion of the maximum achievable contrast. The effective noise levels we used were 0.005%, 2%, 4%, 8%, 12%, 16%, 25% and 33%. Each noise element included a single pixel and the noise gray level values were added to the stimulus gray level values on a pixel-by-pixel basis to generate the noise-injected image.

**Modeling orientation selectivity.** We simulated the circuits involved in one orientation hypercolumn of primary visual cortex using a network model of spiking neurons subject to realistic variability. Several aspects of the model are based on previous models of orientation discrimination, particularly the models in refs. 33 and 22. The model consists of three layers: retina, LGN and V1. The retinal and LGN stages closely follow those described in ref. 22, with one difference being that we only model the spatial receptive fields of retinal and LGN cells, owing to the temporally stationary nature of our stimuli. The retinal stage corresponds to grids of uncoupled, ON and OFF ganglion cells modeled by difference-of-Gaussian filters. The output of each retinal filter is passed through a nonlinearity to produce an analog firing rate that accounts for stimulus contrast sensitivity. These firing rates are then used by the cells at the LGN stage to generate spikes according to a Poisson process.

The V1 stage represents a hypercolumn of orientation-selective layer 4 simple cells. It comprises 256 LNP units that are coupled to each other through lateral connections. LNP units represent a mathematical description that provides a good model of integrate-and-fire neurons in the physiologically realistic high-noise limit[49], while still being analytically tractable. The cortical simple-cell receptive field structure is established through a segregation of ON and OFF LGN inputs into ON and OFF subfields and is modeled using a Gabor function. Each cortical cell receives connections from all the LGN cells within a subfield boundary, with ON subfields yielding connections from all ON-center LGN cells and OFF subfields yielding connections from all OFF-center LGN cells. We implement full lateral connectivity: every cell is coupled with every other cell in the cortical layer. Unlike previous models of orientation selectivity, the cortical layer in our model includes only excitatory neurons. We are nevertheless able to implement the full range of excitatory and inhibitory lateral connectivity to a given cell by allowing the strength of a recurrent connection between two cells to be either positive or negative. We model all the lateral connections as being inhibitory in polarity by making the baseline connection strength substantially negative. However, the pattern of connection strengths between neurons versus the difference in their preferred orientations is chosen so that the connection strengths form a 'Mexican hat' function[33], relative to baseline.

The parameters of the model are adjusted so that the response properties of individual cells in the cortical layer of our network closely match the response properties of V1 neurons *in vivo*. First, because we model cortical cells as LNP units, the variability of generated spikes is guaranteed to be of the Poisson form, which is a good description of neural variability *in vivo*. Second, the response of cortical cells in our network shows the characteristic contrast-invariance that has been reliably demonstrated in orientation-selective cells in the primary visual cortex[34]. Finally, units in the network are subject to realistic inter-neuronal noise correlations that arise from an interaction between the variability in the stimulus, the variability in neuronal firing and the pattern of network connections, and not from any artificial injection of variability.

In the perceptual learning experiments we modeled[4,27], subjects were asked to report the orientation of Gabor patches corrupted by pixel noise and oriented at either −12° or 12° from vertical. To model this task, we added a decision stage to our network in the form of a linear classifier. The linear classifier is equivalent to a decision unit whose activity is determined by the dot product of a weight vector with the population activity in the cortical layer. The weights were tuned to optimize classification performance in the pretraining condition and were left untouched thereafter.

**Computing discrimination performance.** We compute the orientation discrimination performance of our network, when presented with the noisy oriented Gabor stimuli (described above), by estimating Fisher information. Discrimination thresholds can be computed by means of Fisher information because Fisher information is inversely proportional to the discrimination threshold of an ideal observer; that is, it directly predicts performance in discrimination tasks. Recently, we have derived an analytic expression for the linear Fisher information in a population of LNP neurons driven to a noise-perturbed steady state[35]. Linear Fisher information corresponds to the fraction of Fisher information that can be recovered by a locally optimal linear estimator. In practice, linear Fisher information has been found to provide a tight bound on total Fisher information, both in simulations[22] and *in vivo*[19]. This expression can be written as follows:

$$I = (M\mathbf{h}')^T\left[M\Gamma_{\mathrm{hh}}M^T + D^{-1}GD^{-1}\right]^{-1}(M\mathbf{h}') \quad (2)$$

where $M$ represents the matrix of thalamo-cortical feedforward connections, $\mathbf{h}$ represents the mean input firing rates from the thalamus, $\mathbf{h}'$ represents the derivative of $\mathbf{h}$ with respect to orientation, $\Gamma_{\mathrm{hh}}$ represents the covariance matrix of the input firing rates from the thalamus, $G$ is a diagonal matrix whose entries give the mean response of the LNP neurons and $D$ is a diagonal matrix that gives the derivative, or slope, of the activation function $G$ of the cortical neurons in steady state.

The expression in equation (2) is based on the assumption that the decoder used to read out the activity of the cortical layer is the optimal decoder for the specific network. However, in this study, we are interested in highlighting the changes in early response properties that could lead to the observed behavioral improvements, while keeping the decoder constant. Thus, on the basis of previous work[50], we derived another form of the expression in equation (2), for the Fisher information in our network using a fixed decoder, which takes the following form:

$$I(\mathbf{W}_{\mathrm{dec}}) = \frac{\left(\mathbf{W}_{\mathrm{dec}}^T\mu'\right)^2}{\mathbf{W}_{\mathrm{dec}}^T\Gamma\mathbf{W}_{\mathrm{dec}}} \quad (3)$$

where

$$\mu' = \left(D^{-1} - W\right)^{-1}M\mathbf{h}'$$
$$\Gamma = \left(D^{-1} - W\right)^{-1T}\left[M\Gamma_{\mathrm{hh}}M^T + D^{-1}GD^{-1}\right]\left(D^{-1} - W\right)^{-1}$$

$\boldsymbol{W}_{\mathrm{dec}}$ represents the pattern of connection weights from the cortical layer to the decision stage—that is, the fixed decoder described earlier—and $W$ represents the matrix of cortical recurrent connections.

**Deriving TVC curves.** Using the expression in equation (3) allows us to compute the Fisher information, and hence the discrimination threshold, at the decision stage in response to a specific stimulus. In the perceptual learning studies we




doi:10.1038/nn.2796

modeled[4,27], a staircase over signal contrast was used to generate a TVC curve that represents the signal contrast needed to elicit a specific level of performance (represented by percentage correct performance), given a particular level of external noise. In our simulations of those experiments, we numerically obtain TVC curves using the analytic expression in equation (3). Specifically, we first compute the Fisher information at the decision stage using stimuli with a wide range of signal contrasts and a specific level of external noise. We used 15 signal contrast levels: 1.25%, 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 5%, 6%, 7%, 8%, 10%, 12%, 14% and 16%. We then repeat this process with the eight levels of external noise used by the earlier studies[4,27]. Finally, we compute an iso-information contour, for a value of information that is equivalent to the percentage correct criterion used by the earlier studies[4,27] (computed by means of signal detection theory), through the matrix of information generated by the previous two steps, to obtain a TVC curve. As a result of this process, we are able to generate from our network TVC curves, or orientation discrimination performance curves, equivalent to those generated for the human subjects in the perceptual studies[4,27].

**$I_{\text{shuffled}}$.** We denote as $I_{\text{shuffled}}$ the information available in an artificial data set in which the activity of the cortical units was shuffled across trials to remove all noise correlations across cells. This shuffling operation is analogous to making single-cell recordings and generating artificial population patterns of activity by grouping the activity of different cells collected under the same stimulus conditions. To compute $I_{\text{shuffled}}$ for a given stimulus, we consider only the diagonal elements of the cortical covariance matrix (that is, we consider only the variance terms and assume no covariance) when computing the information at the decision stage using the analytic expression in equation (3). To then compute TVC curves based on $I_{\text{shuffled}}$, we use the same numerical approach as that used in computing TVC curves based on the true information (described above).

**Subsampling neuronal populations.** To quantify the results we would have obtained from our simulations had we been able to record only from a subset of the simulated cortical neurons, we also computed information on the basis of only the responses of a subset of randomly sampled cortical neurons. To do this, we again simulated all of the networks considered in the main paper with the full 256 neurons in the cortical layer. However, when we compute information using the expression in equation (3), we only use the activities and the correlations derived from a randomly sampled subset of neurons, and the weights from these neurons to the decision stage, in computing the Fisher information at the decision stage. We then use the same numerical approach as that used in computing TVC curves on the basis of the true information, to compute TVC curves based on the subsampled information. To quantify the effect of varying sample size, we repeated this process for random samples of 128, 64 and 32 neurons.


49. Gerstner, W. & Kistler, W. *Spiking Neuron Models: An Introduction* (Cambridge Univ. Press, New York, 2002).
50. Wu, S., Nakahara, H. & Amari, S.I. Population coding with correlation and an unfaithful model. *Neural Comput.* **13**, 775–797 (2001).

**Supplementary Information for:**

**Title: Perceptual Learning as Improved Probabilistic Inference in Early Sensory Areas**

**Authors:** Vikranth R. Bejjanki, Jeffrey M. Beck, Zhong-Lin Lu and Alexandre Pouget

## Table of Contents





Nature Neuroscience: doi:10.1038/nn.2796

## Supplementary Figures

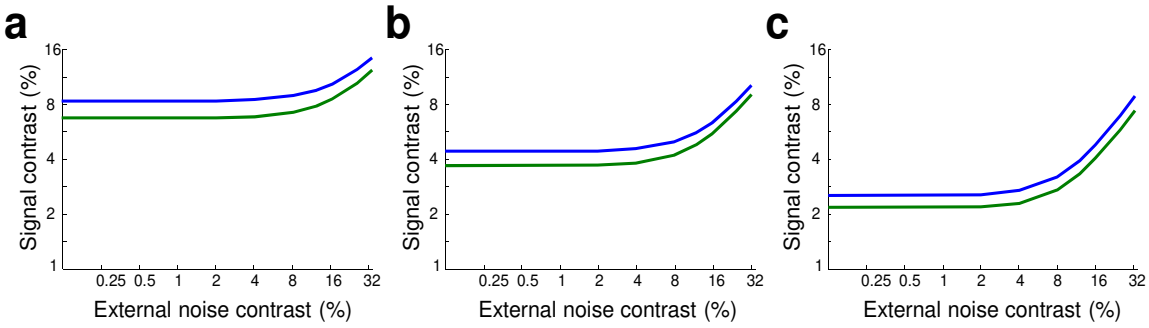


**Figure S1:** Sensitivity to initial weights. **a−c.** TVC curves before (blue) and after (green) learning for three different networks in each of which we used a different initial pattern of feed-forward thalamo-cortical weights but in all of which we moved the weights towards a matched filter. Moving towards a matched filter leads to a uniform shift in the TVC curve, independent of the precise pattern of initial weights. All TVC curves were obtained for the 79.3% correct performance criterion.



Nature Neuroscience: doi:10.1038/nn.2796

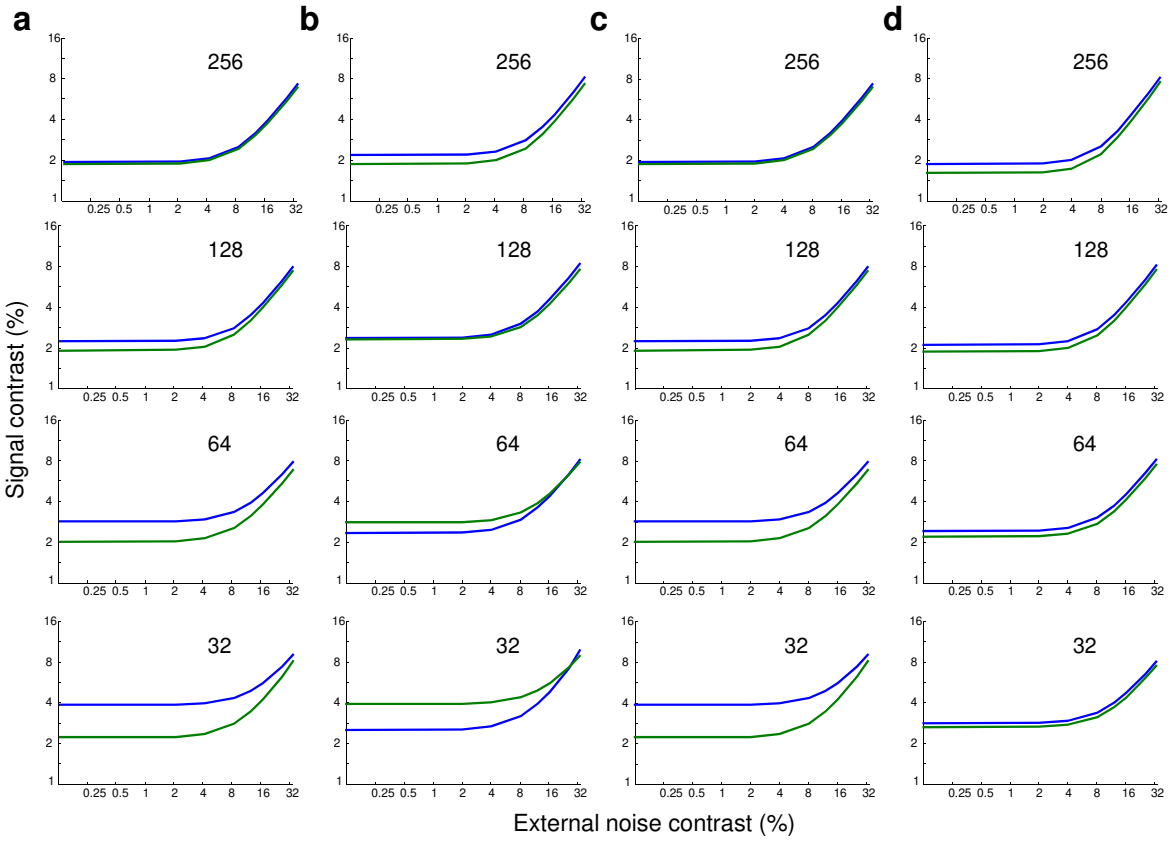


**Figure S2:** The effect of subsampling. **a–d.** TVC curves obtained from subsets of neurons, using the same procedure as in Fig. 7 in the main text, for the networks shown in Fig. 4 in the main text. Each column corresponds to the same column as in Fig. 4. The blue curves correspond to performance before learning, while the green curves correspond to performance after training session 1. With 256 neurons, the results are similar to the results obtained with the full networks (Fig. 4e–h). With only 32 neurons, the results start mimicking the results obtained with $I_{shuffled}$ (Fig. 4i–l). This reflects the fact that neurons tend to be independent in small subsets. All TVC curves were obtained for the 79% correct performance criterion.



Nature Neuroscience: doi:10.1038/nn.2796

## Supplementary Note

### I. Replicating orientation discrimination with external noise

*Stimulus design:* As described in the main text, we replicated the experiment of Dosher & Lu (1999)[1]. In Dosher & Lu (1999), subjects carried out a central fixation task while simultaneously being exposed to orientation discrimination stimuli in the periphery. It was the peripheral task that subjects became better at as a result of perceptual learning. In our study, we only considered the peripheral part of the task – in all our conditions, the stimulus consisted of an oriented Gabor pattern that was tilted $\theta_s^o$ to the right or left of vertical. Formally, each stimulus image was created by assigning grayscale values to image pixels according to the following function:

$$Z(x,y,\theta) = Z_0\left(1.0 + \left(c\ e^{-\left[\frac{Cx^2}{2\sigma_x^2}+\frac{Cy^2}{2\sigma_y^2}\right]}\cos(2\pi KCx)\right)\right) \qquad \text{(S1)}$$

where

$$Cx = x\cos\theta + y\sin\theta;$$
$$Cy = y\cos\theta - x\sin\theta;$$
$$\theta = rad\left(90 \pm \theta_s\right)$$

In the above formalism, x and y are the horizontal and vertical co-ordinates respectively, $K$ is the spatial frequency of the Gabor pattern, $\sigma_x$ and $\sigma_y$ are the standard deviation (extent) of the Gabor in the x and y directions respectively, $Z_0$ is the mean, or background, grayscale value for the image and $c$ is the maximum contrast of the Gabor



Nature Neuroscience: doi:10.1038/nn.2796

pattern as a proportion of the maximum achievable contrast. In all our experimental conditions, $\theta_s^o$ was set to $12^o$, $K$ was set to 0.75 cycles/deg, $\sigma_x$ was set to 0.4, $\sigma_y$ was set to 0.4 and $Z_0$ was set to 126.22 which was the equivalent mean grayscale value used in Dosher & Lu's studies[1, 2]. The maximum contrast of the Gabor, labeled the signal contrast, varied depending on the experimental condition (see **Section III** below). Each Gabor pattern extended over $2.3^o$ X $2.3^o$ of visual angle and was rendered on a 23 X 23 pixel grid. However, it should be noted that the actual stimulus image extended over a span of 45 X 45 pixels. In addition to the Gabor pattern, each stimulus image was padded with 22 extra pixels in the horizontal and vertical dimensions (11 at each end of the image), which were set to the background grayscale value. The reason for this padding was that the retinal cells in our model of orientation discrimination, in line with experimental data, had circular center-surround receptive fields (described in **Section II** below) and unless the image was padded, the receptive fields of the retinal cells at the corners of the square retinal array would extend beyond the extent of the image, leading to possible edge effects in the response of these cells. We added padding for an extent equivalent to two retinal surround field standard deviations, thereby accounting for the vast majority of the corner cells' receptive field extents.

*Adding external noise:* External noise was added to the Gabor stimulus in a manner similar to that used by Dosher & Lu. Pixel gray levels for the external noise were drawn from a Gaussian distribution with mean zero and standard deviation depending on the experimental condition (see **Section III** below). As in Dosher & Lu's studies, we used eight external noise levels in which the standard deviation of the external noise



Nature Neuroscience: doi:10.1038/nn.2796

distribution was varied as a proportion of the maximum achievable contrast. The effective noise levels we used were: 0.005%, 2%, 4%, 8%, 12%, 16%, 25% and 33%. Each noise element included a single pixel and the noise gray level values were added to the stimulus gray level values on a pixel by pixel basis to generate the noise-injected image.

**II. Network model of orientation discrimination**

*Network architecture*: We developed a network model of orientation discrimination containing four stages: retina, LGN, V1 and a decoder. Several aspects of the model are based on previous models of orientation discrimination, particularly the models in Somers et al. (1995)[3] and Series et al. (2004)[4]. The first three stages, which take as input an image of a noisy, oriented Gabor pattern and which give as output a pattern of V1 orientation-tuned activity, model the major processing steps involved in early sensory processing in the visual system. The retina consists of uncoupled analog units that are driven by the image and output an analog firing rate. The retina feeds into the Lateral Geniculate Nucleus (LGN), which consists of a layer of uncoupled, spiking neurons. The LGN in turn feeds into V1, which also consists of spiking neurons, but which are coupled through lateral connections. The fourth stage of the network models the processing that occurs in later decision stages. This stage includes a single unit, with connections to each of the cortical cells, that takes as input the activities of the units in the V1 layer and which gives as output an estimate of the orientation of the stimulus. We describe each layer of the network in the following sections.



Nature Neuroscience: doi:10.1038/nn.2796

*Retina*: The retina is modeled after Series et al. (2004) (in turn modeled after Somers et al. 1995) and contains units in two layers. One difference between the retina modeled in this network and that modeled in previous networks is that we only model the spatial receptive field properties of retinal cells, and not the temporal receptive field properties. We feel justified in doing so since the stimuli we use are temporally stationary. One retinal layer consists of ON center-surround cells and the other OFF center-surround cells. Each layer contains 529 cells arranged in a 23 by 23 array, and the center-to-center spacing between cells, expressed in degrees of visual angle, is 0.1º.

The firing rate of a cell at location (x,y) is determined by the firing rates of the associated center and surround subfields. Specifically,

$$\begin{aligned} r_{ON}(x,y) &= G\left[r_{baseline} + r_{center}(x,y) - r_{surround}(x,y)\right] \\ r_{OFF}(x,y) &= G\left[r_{baseline} - r_{center}(x,y) + r_{surround}(x,y)\right] \end{aligned} \tag{S2}$$

The expressions in (S2) highlight another difference between the retina modeled in this network and that modeled in previous networks. In Series et al. (2004), the firing rate of a retinal cell was rectified using a threshold-linear function in a manner that set the firing rate of the cell to zero, when the expressions within the square brackets on the right hand side of (S2) were less than zero. In our model, we implemented a smooth rectification function G which had the following form:

$$G(x) = \frac{\mu}{\lambda}\ln\left(1 + e^{\lambda(x-\theta)}\right) \tag{S3}$$

We used $\mu = 1.0,\ \lambda = 0.2$ and $\theta = 0.0$



Nature Neuroscience: doi:10.1038/nn.2796

The center and surround subfield responses were generated by convolving the image with a spatial receptive field. Letting α ={center, surround}, the subfield responses are given by:

$$r_\alpha(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F_\alpha(x-x', y-y')I(x',y')dx'dy' \quad \text{(S4)}$$

The center and surround receptive fields $F_\alpha(x,y)$ are modeled as circularly symmetric Gaussians,

$$F_\alpha(x,y) = \frac{K_\alpha}{2\pi\sigma_\alpha^2} e^{\frac{x^2+y^2}{2\sigma_\alpha^2}} \quad \text{(S5)}$$

We used $\sigma_{center}$ =0.176º, $\sigma_{surround}$=0.53º, $K_{center}$=16, $K_{surround}$=16.64, and $r_{baseline}$=15 spk/s.

`

As described in **Section I**, the stimulus is a stationary, Gabor pattern of width w, length l and percent contrast c. Without loss of generality, we center the Gabor pattern at the origin (0,0) and compute $r_\alpha(x,y)$ via equations (S4) and (S5) above as:

$$r_\alpha(x,y) = q(c)K_\alpha\left(\int_{-\frac{w}{2}}^{\frac{w}{2}} \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} e^{-\frac{(x-x')^2}{2\sigma_\alpha^2}} dx'\right)\left(\int_{-\frac{l}{2}}^{\frac{l}{2}} \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} e^{-\frac{(y-y')^2}{2\sigma_\alpha^2}} dy'\right) \quad \text{(S6)}$$

where q(c) is the effective intensity of the stimulus at contrast c.



Nature Neuroscience: doi:10.1038/nn.2796

Note that in the above expression $r_\alpha(x, y)$ is implicitly dependent on the orientation of the Gabor pattern because the image co-ordinates $I(x', y')$ are a function of the Gabor orientation. In all of our experimental conditions, as described in **Section I**, the Gabor pattern's dimensions in degrees of visual angle were $w = 2.3^\circ$ and $l = 2.3^\circ$. The effective intensity q(c) was defined to be

$$q(c) = \left( \frac{\beta\left[\log_{10}(c)\right]}{c} \right) \tag{S7}$$

with $\beta = 0.275$.

The expression in (S7) was chosen to account for the contrast dependence of LGN responses (see below).

<u>*LGN*</u>: Following Series et al. (2004), we assume a one-to-one correspondence between retinal ganglion cells and LGN cells, so that the response of each ganglion cell is passed on to one LGN cell of the same center polarity. The firing rate of an LGN cell at location (*x*,*y*) is either $r_{\mathrm{ON}}(x,y)$ or $r_{\mathrm{OFF}}(x,y)$, depending on whether the LGN cell has ON or OFF polarity. Again, as with the retina, we only model the spatial response properties of LGN cells and not the temporal properties, owing to the temporally stationary nature of our stimuli. Given the parameters used in modeling the retina, the peak response of the LGN cells (that is, the LGN cell with the largest firing rate) versus percent contrast (c), denoted $R_{\mathrm{LGN}}$, is well fit by:



Nature Neuroscience: doi:10.1038/nn.2796

$$R_{LGN}(c) = r_{baseline} + 25\left[\log_{10}(c)\right] \quad \text{(S8)}$$

where, as in the previous section, $r_{baseline}$=15 spk/s is the spontaneous firing rate. This relation is consistent with Somers *et al.* (1995) and with experimental data.

Note that this LGN model is simplified in several ways. It does not account for the mild orientation bias that has been reported in LGN responses or for the precise firing statistics and bursting in LGN. These properties are likely to influence the information available in the input to the cortex. However, it is unlikely that the fraction of this information that is transmitted to the cortical stage will depend critically on these assumptions, at least for the network regimes that we explored in this study.

*V1 – Feed forward thalamo-cortical connections*: The cortical simple cell receptive field structure is established through a segregation of ON and OFF LGN inputs into 3 main subfields (OFF-ON-OFF). We model the receptive field of each cortical cell, with respect to the LGN, using a Gabor function $gab(x, y, \theta)$ defined by:

$$gab(x, y, \theta) = e^{-\left[\frac{C_x^2}{2\sigma_x^2} + \frac{C_y^2}{2\sigma_y^2}\right]} \cos(2\pi k C_x) \quad \text{(S9)}$$

where

$$C_x = x\cos\theta + y\sin\theta;$$
$$C_y = y\cos\theta - x\sin\theta;$$



Nature Neuroscience: doi:10.1038/nn.2796

The parameters $\sigma_x$ and $\sigma_y$ determine the size (or extent) of the receptive field for each cortical cell in the horizontal and vertical dimensions respectively. The anisotropy of the receptive fields is controlled by the parameter $\gamma = \frac{\sigma_y^2}{\sigma_x^2}$. The parameter $k$ determines the preferred spatial frequency of the receptive field for each cortical cell. The receptive fields of all cortical cells are centered at the same position in space; they differ only by their orientation $\theta$. Positive regions of the Gabor function correspond to ON subfields; negative regions correspond to OFF subfields. Each cortical cell receives connections from all the LGN cells within a subfield boundary, with ON-subfields yielding connections from all ON-center LGN cells and OFF subfields yielding connections from all OFF-center LGN cells. The strength of each connection was set to $\alpha * |gab(x, y, \theta)|^2$ where $\alpha$ represents a uniform gain or amplification parameter. The baseline parameters ($\sigma_x$, $\sigma_y$, $k$ and $\alpha$) for the thalamo-cortical connectivity were set as $\sigma_x = 0.36$, $\sigma_y$=0.2, $k = 0.7$, and $\alpha = 0.7$. These parameters were changed in the conditions that involved changes to thalamo-cortical feed forward connectivity as a model of perceptual learning (see **Section IV**).

*V1 – Lateral Connections*: Units in the cortical layer are modeled as spiking units that are coupled to each other through lateral connections. In previous models of orientation selectivity[3, 4], the cortical layer consisted of both excitatory and inhibitory neurons, reflecting cortical layers *in vivo*. In such models, a cortical cell received lateral connections from both excitatory and inhibitory neurons and the net polarity of recurrent connectivity, to a particular cell, was determined by the proportion and strength of lateral



Nature Neuroscience: doi:10.1038/nn.2796

connections from excitatory neurons, relative to lateral connections from inhibitory neurons. In contrast to these previous models, we only include excitatory cortical cells in the cortical layer of our model. However, we are nevertheless able to implement the full range of excitatory and inhibitory lateral connectivity to a particular cell by allowing the strength of a recurrent connection between two cells to be either positive or negative (in previous models, and *in vivo*, the strength of a connection between two cells could only be positive). This allows us to model both excitatory and inhibitory lateral connections without including inhibitory neurons in the cortical layer.

We implement full lateral connectivity – every cell is coupled with every other cell in the cortical layer. We model all the lateral connections as being inhibitory in polarity by making the baseline connection strength significantly negative. However, the pattern of connection strengths between neurons versus the difference in their preferred orientations is chosen so that the connection strengths form a "Mexican hat" function[3, 5], relative to baseline. This implies that:

1. A cell is connected to other cells with similar preferred orientations, with connection strengths that are less inhibitory than baseline.
2. A cell is connected to other cells with significantly dissimilar preferred orientations, with connection strengths that are more inhibitory than baseline.

The polarity of the lateral connection weights, relative to the baseline weight, from a particular cell to all other cells is described mathematically as a difference between two Von-Mises functions of preferred orientations. The width of one function represents the



Nature Neuroscience: doi:10.1038/nn.2796

width of the short-range excitatory (relative to baseline) connection weights and the width of the other function represents the width of the long-range inhibitory (relative to baseline) connection weights.

Formally, the strength of the lateral connection between two cells x and y, with preferred orientations $P_x$ and $P_y$ (in radians) can be written as follows:

$$W(x,y) = \frac{G_w}{N_{out}}\left[ e^{K_e\left(\cos\left(P_y - P_x\right)-1\right)} - A_i e^{K_i\left(\cos\left(P_y - P_x\right)-1\right)} \right] + DC_w \qquad \text{(S10)}$$

The strength of a lateral connection between two neurons is described by five parameters:

1. The width of the excitatory (relative to baseline) connections is represented by the Von-Mises concentration parameter $K_e$
2. The width of the inhibitory (relative to baseline) connections is represented by the Von-Mises concentration parameter $K_i$
3. The overall baseline weight is represented by $DC_w$. As we mentioned previously, this value is set to be significantly negative (inhibitory).
4. The overall gain or amplification factor applied to the lateral connection weights is represented by $G_w$. It should be noted that $G_w$ is scaled by $N_{out}$ which represents the total number of units in the cortical layer thereby making $\frac{G_w}{N_{out}}$ the effective amplification factor.



Nature Neuroscience: doi:10.1038/nn.2796

5. The relative amplitude of the inhibitory Von-Mises distribution, with respect to the excitatory Von-Mises distribution is represented by $A_i$. For example, when $A_i$=1, both the distributions have the same amplitude.

The baseline parameters used were: $K_e$=1, $K_i$=0.5, $DC_w$=−1.0, $G_w$=100 and $A_i$=0.4. These parameters were changed in the conditions that involved changes to cortical lateral connectivity, as a model of perceptual learning (see **Section IV**).

*V1 – Neural properties*: The cortical layer contains 256 neurons which are modeled as Linear Non-Linear Poisson (LNP) neurons. LNP neurons represent a mathematical description that provides a very good model of neural data[6-8] – they represent a good model of integrate and fire neurons in the physiologically realistic high-noise limit. Furthermore, LNP neurons have the advantage of being analytically tractable thereby allowing for good analytical descriptions of neural behavior.

Each cortical cell is modeled as a single point process and the input-output relationship in the cell is composed of three distinct operations.

1. **Linear step:** LNP neurons linearly combine their input spike trains, both feed-forward and recurrent, to obtain a "membrane potential proxy" $u(t)$. The membrane potential proxy $u_i(t)$, for neuron i, is given by:

$$\tau \frac{du_i(t)}{dt} = -u_i(t) + \left[ \sum_j M_{ij} h_j(t) + \sum_j W_{ij} r_j(t) \right] \qquad \text{(S11)}$$



Nature Neuroscience: doi:10.1038/nn.2796

In this expression, $\tau$ represents the membrane time constant which was set to 20 msec, M represents the matrix of feed-forward connection weights from the LGN to the cell (described earlier), W represents the matrix of lateral connection weights from all other cortical cells to the cell (described earlier), $h_j(t)$ represents the input spikes at time *t* from the j[th] pre-synaptic LGN cell and $r_j(t)$ represents the input spikes at time *t* from the j[th] pre-synaptic cortical cell.

2. **Non-linear step:** The "membrane potential proxy" $u_i(t)$ is then passed through a static non-linearity g(u),

$$g(u_i(t)) = \frac{\mu}{\lambda}\ln\left(1+e^{\lambda(u_i(t)-\theta)}\right) \tag{S12}$$

We used $\mu = 1.0,\ \lambda = 0.07$ and $\theta = 50.0$.

This in turn gives the instantaneous probability that the neuron emits a spike.

$$\rho_i\left(t \mid u_i(t)\right) = g\left(u_i(t)\right) \tag{S13}$$

3. **Poisson step:** Finally, a Poisson process is instantiated with the instantaneous spike probability, leading to a variable number of spikes being generated by the neuron at a given time step. The variability of the generated spikes is thus guaranteed to be of the Poisson form which is a good description of biological



Nature Neuroscience: doi:10.1038/nn.2796

neural variability. The spikes emitted by the neuron in a small time interval $dt$ is given by:

$$r_i(t) = Poisson\left[dt * \rho_i(t \mid u_i(t))\right] \quad \text{(S14)}$$

These spikes are in turn transmitted to all the other cortical cells through lateral connections thereby influencing the post-synaptic cells' membrane potential proxies at the next time step.

*V1 – Single-cell Response Properties*: The response properties of individual cells in the cortical layer of our network closely match the response properties of V1 neurons *in vivo*. First, since we model cortical cells as LNP units, the variability of generated spikes is guaranteed to be of the Poisson form which is a good description of neural variability *in vivo*. Second, the response of cortical cells in our network shows the characteristic contrast-invariance that has been reliably demonstrated in orientation selective cells in the primary visual cortex[9-11]. Specifically, as shown in Figure 2b in the main text, the amplitude of the response of V1 cells in our network increases with an increase in signal contrast but the width of the response does not change – in line with responses observed *in vivo*. Furthermore, *in vivo* the amplitude of the response of orientation selective cells in the primary visual cortex has been shown to be logarithmically related to linear changes in signal contrast[10]. The cortical cells in our network reliably show this behavior in a manner that closely mimics realistic cortical neurons.



Nature Neuroscience: doi:10.1038/nn.2796

*Decoder:* The final stage of the network involves connections from all the cortical cells to a single decision unit that outputs an estimate for the orientation of the stimulus (right or left of center, for example). This stage thus represents the decoding step in which the V1 activities are translated into a task-relevant estimate. The overarching goal in this study was to show that early changes in network properties can lead to the uniform shift of TVC curves that is observed during perceptual learning. Accordingly, in simulating the effects of perceptual learning, we kept the connection weights between the cortical cells and the decision unit constant. This was to ensure that any shift in the TVC curve we observed by making early changes in our network, could in fact be linked directly to those early changes. This was particularly critical given that previous work[12] has already shown that changes to the decision weights (the weights from V1 onto the decision unit) can shift the TVC curve uniformly. Had we allowed the weights to change anywhere, and found a uniform shift, we would have been unable to determine whether the uniform shift was the result of an early change (such as a change in the LGN-V1 weights) as opposed to being the result of a change in the decision weights.

We chose to use a linear classifier as the decoder in our network. Formally, the pattern of connection weights $W_{dec}$ from the cortical cells to the decision unit can be written as:

$$W_{dec} = \Sigma^{-1}(s) f'(s) \qquad \text{(S15)}$$

where $f'(s)$ represents the derivative of the tuning curve of the cortical cells. The tuning curve is a concise representation of the response properties of cortical cells and represents



Nature Neuroscience: doi:10.1038/nn.2796

the mean activation of the cell in response to a stimulus s. $\sum^{-1}(s)$ represents the inverse of the cortical covariance matrix.

We optimized the linear classifier for the pre-learning network condition – with the parameters set to the baseline values for each of the layers (described in the sections above) – and with the signal and noise contrast of the stimulus set to an arbitrary value within the range of values used in Dosher & Lu's experiments. We then kept the classifier weights constant across all of our network and experimental manipulations. It was important to use the optimal weights to read out the activity in V1 for the pre-learning network condition, because had we used suboptimal weights, it would have been difficult to pinpoint the basis for any improvement in network performance. For instance, there might have been a way to adjust the LGN-V1 weights such that the Fisher information in V1 does not increase, but such that the performance of the network increases because the sensory representation in V1 is read out more efficiently by the decision weights. This would be a repeat of what Lu and Dosher have already published. With the method we used, on the other hand, the shift in the TVC curves observed in our network can be related solely to the increase of information in the V1 layer.

**III. Computing discrimination performance**

*Relating neural responses to discrimination performance:* In **Section II**, we described the architecture and response properties of our network model of orientation discrimination. We now describe the procedure by which we compute orientation discrimination performance in our network, when presented with the noise-injected stimuli described in



Nature Neuroscience: doi:10.1038/nn.2796

**Section I**. One standard way to relate neural response properties to discrimination performance is to consider the information-theoretic quantity of Fisher information which directly predicts performance in discriminations tasks. Fisher information is an upper bound on performance in discrimination tasks as it is proportional to the square of the discrimination threshold of an ideal observer of neural activity. Therefore, the greater the Fisher information, the better the performance.

We have recently derived an analytic expression for the Fisher information in a population of LNP neurons driven to a noise-perturbed steady state, with network properties similar to those of the network used here[13]. This expression takes advantage of the analytic tractability of LNP neurons and is able to consider the influence of network correlations, which are present in biologically realistic networks, on Fisher information. This expression can be written as follows:

$$I = \left(Mh'\right)^T \left[M\Gamma_{hh}M^T + D^{-1}GD^{-1}\right]^{-1}\left(Mh'\right) \qquad \text{(S16)}$$

The above expression contains several terms each of which map onto specific network and response properties of our network. They are:

1. **M** represents the matrix of thalamo-cortical feed-forward connections. As described earlier, this depends on four network parameters – $\sigma_x$, $\sigma_y$, $k$ and $\alpha$.
2. **h** represents the mean input firing rates from the thalamus. It is directly dependent on the percent signal contrast *c* of the stimulus (described in **Section I**).



Nature Neuroscience: doi:10.1038/nn.2796

3. $\Gamma_{hh}$ represents the covariance matrix of the input firing rates from the thalamus. The input covariance is a combination of the variance of the external noise in the stimulus and the variance resulting from the Poisson spiking of the LGN cells.
4. **G** is a diagonal matrix whose entries give the mean response of the LNP neurons. It is obtained from the non-linear function which transforms the "membrane potential proxy" $u(t)$ to the firing rate for the LNP neurons.
5. **D** is a diagonal matrix which gives the derivative or slope of the activation function **G**.

Readers are referred to[13] for an in-depth description of the techniques used to derive the above expression, including the precise assumptions that were made. It is important to note that the expression in (S16) derives the linear Fisher information in a population of LNP neurons with realistic variability. Linear Fisher information is the fraction of Fisher information that can be recovered by a locally optimal linear estimator. In practice, linear Fisher information has been found to provide a tight bound on total Fisher information, both in simulations[4] and in vivo[14]. Consequently, in the rest of this document, we consider the linear Fisher information in our network and any reference to Fisher information refers to the linear Fisher information. Another point to note is that the original derivation of linear Fisher information in[13] was based on considering a discrimination task in which the difference in stimulus orientation was small. In our case, we are concerned with a difference in orientation of ±12 degrees which is not a small difference. However, linear Fisher Information can still be computed using the analytic


Nature Neuroscience: doi:10.1038/nn.2796

expression by considering the difference between the mean activities induced by the two stimuli and the average covariance induced by the two stimuli.

Another point that should be noted is that the analytic expression as it is written in equation (S16) is based on the assumption that the decoder that is used to read out the activity of the cortical layer is the optimal decoder for the specific network. This implies that a change in network parameters, across learning for instance, might lead to a change in the optimal decoder. However, in this study, we are interested in highlighting the changes in early response properties that could lead to the observed behavioral improvements, while keeping the decoder constant. Thus, based on prior work by Wu et al (2001)[15], we derived another form of the expression in (S16) above, for the Fisher information in our network using a fixed decoder, which looks as follows:

$$I\left(W_{dec}\right)=\frac{\left(W_{dec}^{T}\mu'\right)^{2}}{W_{dec}^{T}\Gamma W_{dec}} \tag{S17}$$

where

$$\mu'=\left(D^{-1}-W\right)^{-1}Mh'$$
$$\Gamma=\left(D^{-1}-W\right)^{-1T}\left[M\Gamma_{hh}M^{T}+D^{-1}GD^{-1}\right]\left(D^{-1}-W\right)^{-1}$$

$W_{dec}$ represents the pattern of connection weights from the cortical layer to the decision stage, i.e. the fixed decoder described earlier and $W$ represents the matrix of cortical recurrent connections which, as described earlier, depends on five network parameters – $K_e, K_i, DC_w, G_w$ and $A_i$.



Nature Neuroscience: doi:10.1038/nn.2796

*Deriving TVC curves:* In Dosher & Lu's studies, given a level of external noise contrast, a staircase method was used to determine the level of signal contrast that was needed to elicit a criterion level of accuracy. This process allowed the experimenters to compute each subject's discrimination threshold for a particular level of external noise contrast, given a criterion level. By repeating the staircase procedure at multiple levels of external noise contrast for the same criterion level, they were then able to derive a curve of discrimination threshold as a function of external noise contrast known as a threshold-versus contrast or TVC curve. This curve depicts the change in discrimination threshold, for a given criterion level, as a function of the change in external noise contrast. In order to be able to relate the results of our network to the results of Dosher & Lu, we need to be able to generate TVC curves of our network's performance similar to the TVC curves of their subjects' performance.

The analytic expression in equation (S17) allows us to compute the Fisher information in the network at the decision stage, when presented with a stimulus generated using a particular signal and noise contrast. In order to then derive TVC curves of our network's performance, using this analytic expression, we used the following approach:

1. We computed the information at the decision stage (using eq.(S17)) when the network was presented with stimuli including a range of signal contrasts and a particular external noise contrast. We used 15 signal contrast levels. They were: 1.25%, 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 5%, 6%, 7%, 8%, 10%, 12%, 14% and 16%.



Nature Neuroscience: doi:10.1038/nn.2796

2. We repeated this process with the eight levels of external noise contrast used by Dosher & Lu.
3. As a result of steps 1 and 2, we generated a matrix which represents the information at the decision stage as a function of signal contrast and external noise contrast. As we move across columns of this matrix we have the change in information as a function of external noise contrast and as we move across rows we have the change in information as a function of signal contrast.
4. Since the criterion accuracy level used by Dosher & Lu was in terms of percent correct performance and our network's performance was in terms of Fisher information, we derived the Fisher information that was equivalent to the percent correct criterion used by Dosher & Lu using techniques from classical signal detection theory[16]. For example, the 79.3% correct criterion used by them corresponds to a Fisher information value of 0.0046 $deg^{-2}$ for the stimuli used.
5. Finally, we computed the iso-information contour through the matrix of information, for the information matching the chosen criterion (0.0046 $deg^{-2}$ in the above example). This gave us a curve of discrimination threshold as a function of external noise contrast, for the criterion accuracy level – a TVC curve.

By following the above approach we have the equivalent orientation discrimination performance curves from our network as Dosher & Lu had for their human subjects. We can now make changes to the network parameters and measure the changes to the TVC curves in an effort to model the neural basis of perceptual learning. The test will be


Nature Neuroscience: doi:10.1038/nn.2796

whether early network changes lead to similar changes in the TVC curves as the changes observed by Dosher & Lu, due to perceptual learning by their subjects.

**IV. Network models tested**

*Parametric study:* In the preceding sections, we have described the manner in which we replicated the external-noise inclusion experiments of Dosher and Lu, our network model for orientation discrimination and the analytic approach that we used to compute the orientation discrimination performance in a given network, when presented with the noise-injected stimuli of Dosher & Lu. We now discuss the specific changes to network parameters that we simulated in order to test our hypothesis that the behavioral improvements observed during perceptual learning can be obtained through changes in the neural representations in early sensory areas. We considered two possible mechanisms through which neural representations could be changed in early sensory areas – through changes to cortical lateral connections and through changes to the thalamo-cortical feed forward connections.

1. **Changes to cortical lateral connections:** As described in **Section II**, the cortical lateral connections in our network model were defined by five parameters – $K_e, K_i, DC_w, G_w$ and $A_i$. The baseline values for these parameters were set to $K_e$=1, $K_i$=0.5, $DC_w$=−1.0, $G_w$=100 and $A_i$=0.4. To test the effect of changing cortical lateral connections on the behavioral performance of the system, we modified each of these parameters across a range of values and observed the



Nature Neuroscience: doi:10.1038/nn.2796

resultant changes in TVC curves. Specifically, $K_e \in (0.05, 4)$, $K_i \in (0.05, 4)$, $DC_w \in (-2.5, 0)$, $G_w \in (0.1, 200)$ and $A_i \in (0, 1.5)$.

2. **Changes to thalamo-cortical feed forward connections:** As described in **Section II**, the thalamo-cortical feed forward connections in our network model were defined by four parameters $-\sigma_x$, $\sigma_y$, $k$ and $\alpha$. The baseline values for these parameters were set to $\sigma_x = 0.36$, $\sigma_y$=0.2, $k = 0.7$, and $\alpha = 0.7$. To test the effect of changing thalamo-cortical feed forward connections on the behavioral performance of the system, we modified each of these parameters across a range of values and observed the resultant changes in TVC curves. Specifically, $\sigma_x \in (0.2, 0.4)$, $\sigma_y \in (0.2, 0.4)$, $k \in (0.5, 1.25)$ and $\alpha \in (0.2, 2.0)$.

*Modeling Perceptual Learning:* As discussed in the main text (see Results section and Fig. 3), we were able to replicate the full range of observed behavioral changes, by only making changes to the thalamo-cortical feed forward connections, in our model. The specific parameter values that led to the reported results (Fig. 3) are as follows:

**Supplementary Table 1: Parameters for networks shown in Fig. 3 (in the main text)**

| | $\sigma_x$ | $\sigma_y$ | $k$ | $\alpha$ | $K_e$ | $K_i$ | $DC_w$ | $G_w$ | $A_i$ |
|---|---|---|---|---|---|---|---|---|---|
| Before Learning (baseline) | 0.36 | 0.2 | 0.7 | 0.7 | 1 | 0.5 | −1.0 | 100 | 0.4 |
| Training Session 1 | 0.36 | 0.23 | 0.67 | 0.6 | 1 | 0.5 | −1.0 | 100 | 0.4 |
| Training Session 2 | 0.36 | 0.27 | 0.62 | 0.5 | 1 | 0.5 | −1.0 | 100 | 0.4 |



Nature Neuroscience: doi:10.1038/nn.2796

*Exploring the role of amplification and sharpening*: In addition to a uniform shift in TVC curves, changing the thalamo-cortical feed forward connections also led to a modest amount of amplification and sharpening in cortical tuning curves (Fig. 3b in the main text). We simulated several network models which allowed us to demonstrate that amplification and sharpening were neither necessary nor sufficient for the observed shift in TVC curves. The parameters for the models used in these demonstrations (Fig. 4 in the main text) are as follows:

**Supplementary Table 2: Parameters for networks shown in Fig. 4 (in the main text)**

| | $\sigma_x$ | $\sigma_y$ | $k$ | $\alpha$ | $K_e$ | $K_i$ | $DC_w$ | $G_w$ | $A_i$ |
|---|---|---|---|---|---|---|---|---|---|
| **Amp. not sufficient** | | | | | | | | | |
| Before Learning | 0.4 | 0.3 | 0.75 | 0.5 | 2 | 1 | −1.5 | 1 | 0.75 |
| After Learning | 0.29 | 0.33 | 1.00 | 1.1 | 2 | 1 | −1.5 | 1 | 0.75 |
| **Amp. not necessary** | | | | | | | | | |
| Before Learning | 0.26 | 0.29 | 1.25 | 1.6 | 2 | 1 | −1.5 | 1 | 0.75 |
| After Learning | 0.4 | 0.3 | 0.75 | 0.5 | 2 | 1 | −1.5 | 1 | 0.75 |
| **Sharp. not sufficient** | | | | | | | | | |
| Before Learning | 0.4 | 0.3 | 0.75 | 0.5 | 2 | 1 | −1.5 | 1 | 0.75 |
| After Learning | 0.29 | 0.33 | 1.00 | 1.1 | 2 | 1 | −1.5 | 1 | 0.75 |
| **Sharp. not necessary** | | | | | | | | | |
| Before Learning | 0.32 | 0.3 | 0.75 | 1.0 | 0.1 | 0.75 | −1.2 | 40 | 1.0 |
| After Learning | 0.32 | 0.4 | 0.6 | 0.63 | 0.05 | 0.4 | −1.0 | 60 | 1.3 |

*Quantifying the contribution of correlations:* In addition to demonstrating that learning need not depend on tuning curve changes like amplification and sharpening, we also quantified the improvement in performance we would have observed in the network shown in Figure 3 in the main text, if we had not seen any tuning curve changes across



Nature Neuroscience: doi:10.1038/nn.2796

training. In other words, we quantified the contribution of only changing the inter-neuronal correlations to the overall improvement in performance that was observed in our simulations. To do this, we used the analytic relationship in (S17) to compute the contribution of a change in correlations to the training induced increase in Fisher information. Specifically, we compared the increase in information across training sessions for the network shown in Figure 3, against the change in information in a virtual population of neurons with the same change in correlations but with identical tuning curves across training sessions. We call this population 'virtual' because it is not clear that such spike statistics could be generated by an actual neural network since tuning curves and correlations cannot easily be decoupled in our model. Nonetheless, we can still compute Fisher information for such spike statistics.

This method was inspired by a metric that was recently proposed by Cohen and Maunsell[17]. However, in the case of Cohen and Maunsell, their metric was based on an approximate computation of the Fisher Information before and after learning. Here, since we have access to the analytic relationship between network parameters and the true Fisher Information, we were able to quantify the contribution of a change in correlations to the training-induced increase in the true Fisher Information.



Nature Neuroscience: doi:10.1038/nn.2796

**Supplementary References**


1. Dosher, B.A. & Lu, Z.L. Mechanisms of perceptual learning. *Vision Research* **39**, 3197-3221 (1999).
2. Dosher, B.A. & Lu, Z.L. Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 13988-13993 (1998).
3. Somers, D.C., Nelson, S.B. & Sur, M. An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.* **15**, 5448-5465 (1995).
4. Series, P., Latham, P.E. & Pouget, A. Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. *Nature Neuroscience* **7**, 1129-1135 (2004).
5. Sompolinsky, H. & Shapley, R. New perspectives on the mechanisms for orientation selectivity. *Current Opinion in Neurobiology* **7**, 514–522 (1997).
6. Gerstner, W. & Kistler, W. *Spiking Neuron Models: An Introduction* (Cambridge University Press New York, NY, USA, 2002).
7. Paninski, L. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems* **15**, 243–262 (2004).
8. Plesser, H.E. & Gerstner, W. Noise in integrate-and-fire neurons: from stochastic input to escape rates. *Neural Computation* **12**, 367–384 (2000).
9. Ferster, D. & Miller, K.D. Neural Mechanisms of Orientation Selectivity in the Visual Cortex. *Annual Reviews in Neuroscience* **23**, 441-471 (2000).




Nature Neuroscience: doi:10.1038/nn.2796

10. Sclar, G. & Freeman, R.D. Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. *Experimental Brain Research* **46**, 457-461 (1982).
11. Troyer, T.W., Krukowski, A.E., Priebe, N.J. & Miller, K.D. Contrast-Invariant Orientation Tuning in Cat Visual Cortex: Thalamocortical Input Tuning and Correlation-Based Intracortical Connectivity. *J. Neurosci.* **18**, 5908-5927 (1998).
12. Lu, Z.-L., Liu, J. & Dosher, B.A. Modeling mechanisms of perceptual learning with augmented Hebbian re-weighting. *Vision Research* **50**, 375-390 (2010).
13. Beck, J.M., Bejjanki, V.R. & Pouget, A. Insights from a simple expression for linear Fisher information in a recurrently connected population of spiking neurons. *Neural Computation* (In Press).
14. Averbeck, B.B., Latham, P.E. & Pouget, A. Neural correlations, population coding and computation. *Nat Rev Neurosci* **7**, 358-366 (2006).
15. Wu, S., Nakahara, H. & Amari, S.I. Population coding with correlation and an unfaithful model. *Neural Computation* **13**, 775-797 (2001).
16. Green, D.M. & Swets, J.A. *Signal detection theory and psychophysics* (John Wiley and Sons, Los Altos, California, USA, 1966).
17. Cohen, M.R. & Maunsell, J.H.R. Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience* **12**, 1594-1600 (2009).



Nature Neuroscience: doi:10.1038/nn.2796