

# Adaptivity of Tuning Functions in a Generic Recurrent Network Model of a Cortical Hypercolumn

Lars Schwabe and Klaus Obermayer

Department of Computer Science and Electrical Engineering, Berlin University of Technology, 10587 Berlin, Germany

The representation of orientation information in the adult visual cortex is plastic as exemplified by phenomena such as perceptual learning or attention. Although these phenomena operate on different time scales and give rise to different changes in the response properties of neurons, both lead to an improvement in visual discrimination or detection tasks. If, however, optimal performance is indeed the goal, the question arises as to why the changes in neuronal response properties are so different. Here, we hypothesize that these differences arise naturally if optimal performance is achieved by means of different mechanisms. To evaluate this hypothesis, we set up a recurrent network model of a visual cortical hypercolumn and asked how each of four different parameter sets (strength of afferent and recurrent synapses, neuronal gains, and additive background inputs) must be changed to optimally improve the encoding accuracy of a particular set of visual stimuli. We find that the predicted changes in the population responses and the tuning functions were different for each set of parameters, hence were strongly dependent on the plasticity mechanism that was operative. An optimal change in the strength of the recurrent connections, for example, led to changes in the response properties that are similar to the changes observed in perceptual learning experiments. An optimal change in the neuronal gains led to changes mimicking neural effects of attention. Assuming the validity of the optimal encoding hypothesis, these model predictions can be used to disentangle the mechanisms of perceptual learning, attention, and other adaptation phenomena.

**Key words:** visual cortex; orientation tuning; model; plasticity; perceptual learning; attention

## Introduction

Orientation tuning is the paradigmatic example of stimulus selectivity in the visual cortex. It first arises in the primary visual cortex (V1) and is preserved in higher visual areas such as V2 and V4. The typically bell-shaped orientation tuning functions, however, have been shown experimentally to be highly adaptive. They depend on the temporal context of the stimulation (Muller et al., 1999; Dragoi et al., 2002) and the current behavioral demands (Moran and Desimone, 1985; McAdams and Maunsell, 1999, 2000; Treue and Martinez Trujillo, 1999), and they also change during the course of training a perceptual task (Schoups et al., 2001; Ghose et al., 2002; Yang and Maunsell, 2004). This raises the question of why the representations of physically unchanged stimuli in the adult visual cortex are so adaptive. Would it not be sufficient for an animal to act successfully if one proper representation of the current “state” of the environment were computed in the visual cortex and then forwarded to neuronal structures responsible for planning and initiating actions?

Many previous theoretical works are based on “optimal coding hypotheses” to explain the observed changes. Paradiso (1988), Seung and Sompolinsky (1993), Clifford et al. (2000),

Nakahara et al. (2001), and Bethge et al. (2003), for example, used descriptive models of tuning functions and assessed the quality of the sensory representation as a function of the parameters of tuning functions. In reality, however, neuronal response properties are computed in a recurrent cortical network in which architecture and plasticity mechanisms constrain the set of available tuning functions and their possible changes. Hence, the changes predicted by descriptive models may not be realizable. Teich and Qian (2003) set up a physiologically plausible model to explain the changes in orientation tuning functions in V1 during adaptation and perceptual learning. In this study, however, the synaptic changes were not derived from a functional principle; rather, they were determined *ad hoc* to fit experimental data.

This motivated us to combine both approaches and to evaluate an optimal coding principle for a physiologically realistic model of a visual cortical hypercolumn. A recurrent neuronal network encodes a stimulus (e.g., the orientation  $\theta$  of an oriented grating) by the activity of its output neurons. The quality of this representation can then be assessed using a hypothetical ideal observer (“decoder” or “read-out”). Within such a setting (see Fig. 1), we address the following two questions: (1) How do the tuning functions and population responses change if the quality of representation is optimally improved for a “relevant” set of stimuli? (2) How are these changes affected, if plasticity is restricted to one of the four “sites of plasticity”: maximum conductance of the afferent or recurrent synapses, gain of the excitatory neurons, and strength of an additive (feedback) input current?

We find that optimal changes in response properties are dif-

Received Nov. 2, 2004; revised Feb. 16, 2005; accepted Feb. 16, 2005.

This work was supported by the Deutsche Forschungsgemeinschaft (SFB 618).

Correspondence should be addressed to Lars Schwabe, Department of Computer Science and Electrical Engineering, Berlin University of Technology, FR2-1, Franklinstrasse 28/29, 10587 Berlin, Germany. E-mail: schwabe@cs.tu-berlin.de and oby@cs.tu-berlin.de.

DOI:10.1523/JNEUROSCI.4493-04.2005

Copyright © 2005 Society for Neuroscience 0270-6474/05/253323-10\$15.00/0

ferent for different sites of plasticity and that specific changes can even go into opposite directions, and still improve coding quality. This finding stresses how important it is to consider physiological constraints when interpreting data in light of a functional principle. We also find that published experimental data on perceptual learning and on attentional modulations of tuning functions in the visual cortex are broadly consistent with the predictions of the model if the recurrency (for perceptual learning) or the values of the neuronal gain (for attentional modulation) are changed. This motivates the hypothesis that seemingly unrelated phenomena may be explained by one functional principle and that diversity emerges because different routes are taken to calibrate cortical representations according to the same goal.

### Materials and Methods

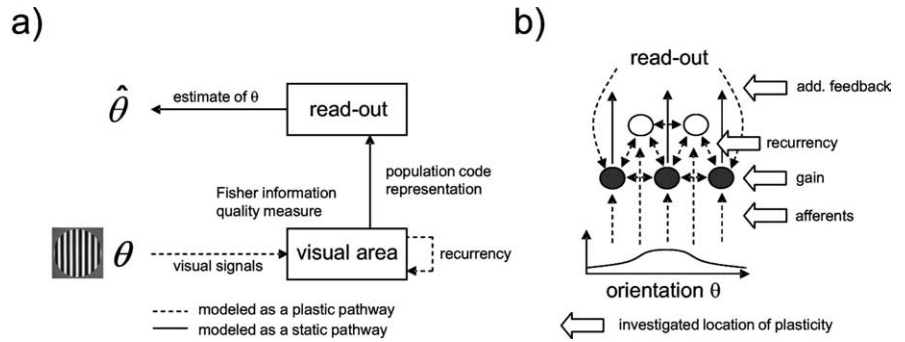
In the following section, we describe the hypercolumn model (Fig. 1*b*), the single-cell model, the basic quality measure, and the two objective functions used to measure coding quality. All parameter values refer to the “unadapted” system and were adjusted such that the responses of the model are consistent with neuronal responses from area V4 of the macaque monkey (Yang and Maunsell, 2004). An encoder/decoder framework (Fig. 1*a*) is used to construct a principled quality measure for the neuronal representation, which is independent of a concrete neuronal read-out. We do not claim that the encoder/decoder framework applies one-to-one to the feedforward/feedback interactions between two connected visual areas; rather, this framework is used to construct a principled measure of the quality of a neuronal code.

**Mean-field network model of the cortical hypercolumn.** The architecture of the recurrent network model is shown in Figure 1*b*. Excitatory and inhibitory neurons receive already tuned afferent inputs from a lower visual area as well as additive and background (feedback) inputs. The latter are not described explicitly, but their overall effects are summarized by a fluctuating background conductance and an additive input current (“feedback”). Parameterizations were chosen such that the model reproduced the “average response” of a V4 cell in the control trial of a perceptual learning experiment (Yang and Maunsell, 2004).

We used a simplified rate model with a biophysical interpretation following Shriki et al. (2003). Consider a presynaptic neuron *j* making a synapse to a postsynaptic neuron. Whenever neuron *j* fires a spike, the postsynaptic conductance makes an instantaneous jump of magnitude  $1/\tau_j$  and then decays with the time constant  $\tau_j$  as described by  $\tau_j dg_j/dt = -g_j + \sum_k (t - t_k^j)$ , where  $t_k^j$  denotes the time of the *k*th spike fired by neuron *j*. Depending on the type of the synapse, we set  $\tau_j = \tau_E = 5$  ms for excitatory synapses and  $\tau_j = \tau_I = 10$  ms for inhibitory synapses. Let  $f_i = F_i(I_i)$  denote the firing-rate response of neuron *i* to the input current  $I_i$ , where  $f_i$  is a state variable denoting the firing rate of neuron *i* and  $F_i$  is the current–frequency function of neuron *i*. Then one obtains the following steady-state condition (Shriki et al., 2003):

$$f_i = F_i \left( I_i^{\text{aff}} + I_i^{\text{add}} + \sum_j W_{ij} f_j - \Delta I_i(g_i^L) \right),$$

where  $W_{ij} = G_{ij}(E_j - E_i^L - V_i^c)$  is the “synaptic weight” for the connection between the neurons *i* and *j*.  $I_i^{\text{aff}}$  and  $I_i^{\text{add}}$  are the afferent and additive input currents, and  $E_j$  is the reversal potential for the synaptic connections of (presynaptic) neuron *j* ( $E_j = E_E = 0$  mV for excitatory synapses and  $E_j = E_I = -80$  mV for inhibitory synapses). The current  $I_i^c$  and the voltage  $V_i^c$  determine the shift  $\Delta I_i(g_i^L) = I_i^c + V_i^c g_i^L$  of the current–frequency function toward higher or lower input currents (values for  $I_i^c$  and  $V_i^c$  are given below), and  $g_i^L$  and  $E_i^L$  are the leak conductance and



**Figure 1.** The hypercolumn model and the encoder/decoder framework for assessing the quality of sensory representations. *a*, The encoder/decoder framework. An ideal observer computes a point estimate  $\hat{\theta}$  of the stimulus  $\theta$  based on the neuronal responses of the cortical hypercolumn. The variance of this estimate should be minimal; therefore, the Fisher information should be maximized. *b*, Recurrent network model of a cortical hypercolumn of excitatory (filled circles) and inhibitory (open circles) neurons. The thick arrows point to the sites of plasticity (dashed lines) considered in this report. add., Additive.

reversal potential of neuron *i*. The steady-state responses  $f_i$  are computed by integrating  $\tau_i df_i/dt = -f_i + F_i(I_i)$ .

This mean-field model describes steady states of large networks that do not possess a high degree of synchrony. One of its key assumptions is that changes in the input conductance of the cell lead to subtractive changes in its current–frequency curve. This assumption is fulfilled by the model neurons we use (see below, Single-cell model). Our model differs slightly from the one used by Shriki et al. (2003). First, we used a different model neuron. Second, we considered current–frequency functions in the presence of fluctuating balanced background inputs. The latter leads to smaller effective membrane time constants so that the network dynamics is likely to be dominated by the synaptic time constant.

**Input currents.** We separate the overall input current  $I_i = I_i^{\text{aff}} + I_i^{\text{ec}} + I_i^{\text{add}}$  into an afferent, a recurrent, and an additive background component. A large fraction of the latter is assumed to be a result of direct feedback received from a downstream area. For excitatory and inhibitory neurons, we initially set  $I_i^{\text{add}}$  to 0.6 and 0.64 nA, respectively. This gives rise to a background activity of 3.6 spikes/s (sp/s) for the excitatory and inhibitory neurons. When a stimulus  $\theta$  is presented to the network, it leads to an afferent input to neuron *i*, which is calculated using a bell-shaped input tuning function with a peak at  $\theta_i^{\text{pr}}$ :

$$I_i^{\text{aff}}(\theta) = W_i^{\text{aff}} \text{Mexp}(|\cos(2\pi d_i \theta) - 1|),$$

$$W_i^{\text{aff}} = G_i^{\text{aff}}(E_E - E_i^L - V_i^c).$$

$d_i(\theta) = \min(|\theta - \theta_i^{\text{pr}}|, 1 - |\theta - \theta_i^{\text{pr}}|)$  is the circular distance between stimulus  $\theta$  and the “afferent” preferred stimulus (PS)  $\theta_i^{\text{pr}} = i/N$  of neuron *i*, *N* is the number of excitatory or inhibitory neurons, depending on the type of neuron *i*, and  $M = 2000$  sp/s. For convenience, we consider dimensionless one-dimensional “circular” stimuli with  $0 \leq \theta \leq 1$ . To obtain numerical values for stimulus domains such as “orientation” or “direction,”  $\theta$  needs to be multiplied by 180 and 360°, respectively.

The recurrent input is a weighted sum of the output activities of the neuron and is given by the following equations:

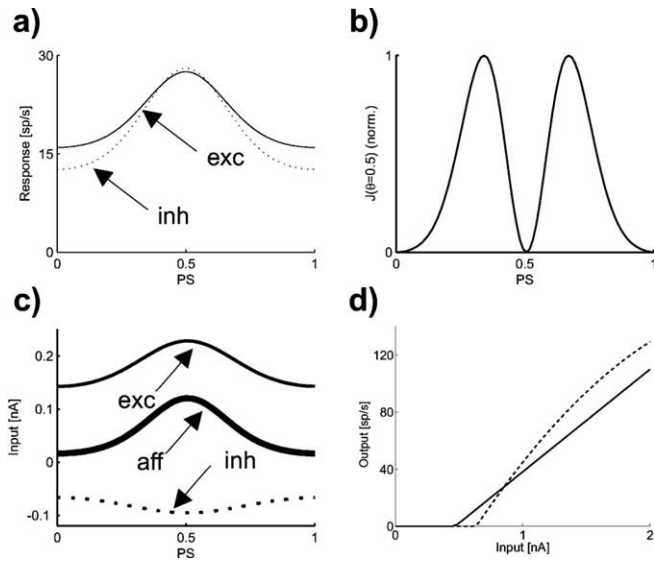
$$I_i^{\text{rec}} = \sum_j W_{ij}^{\text{rec}} f_j,$$

$$W_{ij}^{\text{rec}} = G_{ij}^{\text{rec}}(E_j - E_i^L - V_i^c),$$

$$G_{ij}^{\text{rec}} = Z_i \exp(\kappa_i \cos(2\pi d_i(\theta_i^{\text{pr}}))),$$

where  $\kappa_i$  ( $\kappa_i = \kappa_I = 1$  for inhibitory neurons and  $\kappa_i = \kappa_E = 4$  for excitatory neurons) and  $Z_i$  determines the specificity and the strength of the recurrent connections. The strengths  $Z_i$  are set so that  $\sum_j G_{ij}^{\text{rec}} = 0.135$  nS for the excitatory neurons and 0.2813 nS for the inhibitory neurons *j*. The responses of the network with its initial parameterization and the corresponding input currents are shown in Figure 2, *a* and *c*.

**Single-cell model.** The single-cell model is the Hodgkin–Huxley-type



**Figure 2.** Response properties and quality of representation of the model hypercolumn for the choice of parameters given in Materials and Methods. *a*, Responses of excitatory (exc; solid line) and inhibitory (inh; dotted line) neurons to a stimulus with  $\theta = 0.5$ . *b*, Contribution of the excitatory neurons to the Fisher information of the network for  $\theta = 0.5$  normalized (norm.) to the maximal value. *c*, Input currents at the steady state for an excitatory neuron with PS = 0.5 as a function of the PS of the presynaptic neuron. aff, Afferent. *d*, Fits to the current–frequency functions of the conductance-based model neurons. exc neuron, Solid line; inh neuron, dashed line.

neuron described by Destexhe et al. (2001). The dynamics of the membrane potential  $V_i$  of neuron  $i$  is described by the following equation:

$$C_i^m \frac{dV_i}{dt} = -g_i^L(V_i - E_i^L) - \sum_{\text{int}} I_{\text{int}} - I_i,$$

where  $I_{\text{int}}$  denotes intrinsic voltage-dependent currents,  $g_i^L$  and  $E_i^L$  are the leak conductances ( $g_i^L = g_E^L = 22.74$  nS for excitatory neurons and  $g_i^L = g_I^L = 2g_E^L$  for inhibitory neurons) and reversal potentials ( $E_i^L = E_E^L = E_I^L = -80$  mV),  $C_i^m$  is the membrane capacitance ( $C_i^m = C_E^m = C_I^m = 0.5$  nF), and  $t$  is the time. Each current  $I_{\text{int}}$  is described by a Hodgkin–Huxley equation:

$$I_{\text{int}}(t) = \bar{g} m^M(t) h^N(t) (V(t) - E),$$

where  $\bar{g}$  is the peak conductance,  $E$  is the reversal potential, and  $m(t)$  and  $h(t)$  are the activation and inactivation variables. Three voltage-dependent currents are included: a fast  $\text{Na}^+$  current and a delayed-rectifier  $\text{K}^+$  current for the generation of action potentials and a slow noninactivating  $\text{K}^+$  current responsible for spike-frequency adaptation.

For the  $\text{Na}^+$  current, we used the following equations:

$$I_{\text{Na}} = \bar{g}_{\text{Na}} m^3 h (V - E_{\text{Na}}),$$

$$\frac{dm}{dt} = \alpha_m(V)(1 - m) - \beta_m(V)m,$$

$$\frac{dh}{dt} = \alpha_h(V)(1 - h) - \beta_h(V)h,$$

$$\alpha_m = \frac{-0.32(V - V_T - 13)}{\exp(-(V - V_T - 13)/4) - 1},$$

$$\beta_m = \frac{0.28(V - V_T - 40)}{\exp((V - V_T - 40)/5) - 1},$$

$$\alpha_h = 0.128 \exp(-(V - V_T - V_S - 17)/18),$$

$$\beta_h = \frac{4}{1 + \exp(-(V - V_T - V_S - 40)/5)},$$

with parameter values  $V_T = -58$  mV,  $V_S = -10$  mV,  $\bar{g}_{\text{Na}} = 17.87$   $\mu\text{S}$ , and  $E_{\text{Na}} = 50$  mV.

For the “delayed-rectifier”  $\text{K}^+$  current, we used the following equations:

$$I_{\text{Kd}} = \bar{g}_{\text{Kd}} n^4 (V - E_K), \quad \frac{dn}{dt} = \alpha_n(V)(1 - n) - \beta_n(V)n,$$

$$\alpha_n = \frac{-0.032(V - V_T - 15)}{\exp(-(V - V_T - 15)/5) - 1},$$

$$\beta_n = 0.5 \exp(-(V - V_T - 10)/40),$$

with parameter values  $E_K = -90$  mV and  $\bar{g}_{\text{Kd}} = 3.46$   $\mu\text{S}$ .

For the noninactivating  $\text{K}^+$  current, we used the following equations:

$$I_M = \bar{g}_M n (V - E_K), \quad \frac{dn}{dt} = \alpha_n(V)(1 - n) - \beta_n(V)n,$$

$$\alpha_n = \frac{0.0001(V + 30)}{1 - \exp(-(V + 30)/9)}, \quad \beta_n = \frac{-0.0001(V + 30)}{1 - \exp((V + 30)/9)},$$

with  $\bar{g}_M = 0.28$   $\mu\text{S}$  for excitatory neurons and  $\bar{g}_M = 0.028$   $\mu\text{S}$  for inhibitory neurons.

The model neurons additionally receive balanced excitatory and inhibitory synaptic background inputs. The corresponding conductances are described by a stochastic process similar to an Ornstein–Uhlenbeck process with the following update rule (Gillespie, 1996):

$$g_{\text{bg}}(t + \Delta t) = g_{\text{bg}}^0 + [g_{\text{bg}}(t) - g_{\text{bg}}^0] \exp(-\Delta t/\tau) + A \times N(0, 1),$$

where  $g_{\text{bg}}^0$  is the average conductance,  $\tau$  is a synaptic time constant,  $A$  is the amplitude coefficient, and  $N(0, 1)$  is a normally distributed random number with zero mean and unit SD. The amplitude coefficient has the following analytic expression:

$$A = \sqrt{\frac{D \times \tau}{2} \left[ 1 - \exp\left(-\frac{\Delta t}{\tau}\right) \right]},$$

where  $D = 2\sigma^2\tau^{-1}$  is the diffusion coefficient. Numerical values for the background conductances are  $\sigma = 3$  nS and  $\sigma = 6.6$  nS for the variances of the excitatory and inhibitory conductance,  $\tau = 2.7$  ms for the excitatory time constant and  $\tau = 10.5$  ms for the inhibitory time constant, and  $g_{\text{bg}}^0 = 12.1$  nS for the mean excitatory conductance and  $g_{\text{bg}}^0 = 57.3$  nS for the mean inhibitory conductance. The reversal potentials were 0 and  $-75$  mV for the excitatory and inhibitory conductances.

We simulated the spike responses of this model neuron to current injections for different values of the leak conductance in the presence of the fluctuating background conductance. We considered only the already adapted responses, which were then best fitted with thresholded polynomials. These fits were done by minimizing the mean-squared error between the simulated firing rate and the one predicted by the thresholded polynomials. We obtained the following equations:

$$F_i^E(I) = \max(0, a_i^E(I - \Delta I(g_L)))$$

$$F_i^I(I) = \max(0, a^I(I - \Delta I(g_L)) + b^I(I - \Delta I(g_L))^2),$$

with  $a_i^E = 71.9$  sp/s (nA) $^{-1}$ ,  $I_E^c = 0.13$  nA, and  $V_E^c = 15.2$  mV for excitatory neurons and  $a_i^I = 133$  sp/s (nA) $^{-1}$ ,  $b_i^I = -28$  sp/s (nA) $^{-2}$ ,  $I_i^I = 0.02$  nA, and  $V_i^I = 14.6$  mV for inhibitory neurons (Fig. 2*d*).

The average output activity of the mean-field (rate) model is then converted into a noisy spike output activity with Poisson statistics with the spikes being conditionally independent given the stimulus. The probability to count  $n$  spikes fired by neuron  $i$  in a time interval of duration  $\tau$  is given by the following equation:

$$P_i(n; \theta) = \frac{[\tau \cdot f_i(\theta)]^n}{n!} \exp(-\tau \times f_i(\theta)),$$

where  $\theta$  is the presented stimulus and  $f_i(\theta)$  is the steady-state response  $f_i$  of neuron  $i$  to the stimulus  $\theta$ . We always used  $\tau = 1$ .

**Fisher information.** To quantify the quality of the representation of the stimulus  $\theta$ , we consider a hypothetical ideal observer whose task is to provide the best possible estimate  $\hat{\theta}$  of the stimulus given a set of spike counts and knowledge of the probability distribution  $P_i(n; \theta)$ . In an estimation, task the Fisher information is a useful quantity to measure the quality of a representation, because (for a one-dimensional continuous stimulus  $\theta$ ) the Fisher information  $J(\theta)$  provides, via  $1/J(\theta)$ , a lower bound for the variance of any unbiased estimator of  $\theta$  (Kay, 1993). If  $P_i(n; \theta)$  is the probabilistic description of how the spike count  $n$  of neuron  $i$  relates to the stimulus value  $\theta$ , then no unbiased estimate of  $\theta$  based on the spike count  $n$  can have a lower variance than  $1/J(\theta)$ .

For a population of  $N$  neurons with their “noise” being statistically independent given the stimulus, the population Fisher information is  $J(\theta) = \sum_i J_i(\theta)$ , where

$$J_i(\theta) = E \left[ \left( \frac{d}{d\theta} \log P_i(n; \theta) \right)^2 \right] P_i(n; \theta).$$

The Fisher information is also monotonically related to the mutual information between the stimulus  $\theta$  and the whole vector  $\mathbf{n} = (n_1, n_2, \dots, n_N)$  of the spike counts (Brunel and Nadal, 1998) as well as to the measure  $d'$  often used in the psychophysics literature (Seung and Sompolinsky, 1993). For Poisson statistics of the spike response, one obtains the following equation:

$$J_i(\theta) = \tau \left( \frac{\partial f_i(\theta)}{\partial \theta} \right)^2 / f_i(\theta).$$

Figure 2*b* shows  $J_i(\theta = 0.5)$  as a function of neuron  $i$  (normalized to the maximum) for the initial network parameterization.

Note that here the Fisher information of a single neuron  $i$ ,  $J_i(\theta)$ , is proportional to  $[f_i'(\theta)]^2 / f_i(\theta)$ , where  $f_i(\theta)$  and  $f_i'(\theta)$  are the tuning function of that neuron and its derivative. To determine how much changes in the absolute values of  $f_i(\theta)$  and the changes in the slopes  $f_i'(\theta)$  contributed to the total improvement of the Fisher information, we calculated the quantities  $J_{\text{add}}(\theta)$  and  $J_{\text{slp}}(\theta)$ . For  $J_{\text{add}}(\theta)$  we used the amplitudes  $f_i(\theta)$  after we changed the model parameters and the derivatives  $f_i'(\theta)$  before the reparameterization. For  $J_{\text{slp}}(\theta)$ , we used the derivatives  $f_i'(\theta)$  after the reparameterization but the amplitudes  $f_i(\theta)$  before the reparameterization. Thus, an increase/decrease in the encoding accuracy only attributable to changes in the slope is reflected by a large change in  $J_{\text{slp}}(\theta)$ , whereas an increase/decrease only attributable to changes in the response magnitude is reflected by a large change in  $J_{\text{add}}(\theta)$ .

**Objective functions.** A full optimization of the network is only a reasonable approach if additional constraints on the plausible range of values of the network parameters are imposed. Because we want our results to depend as little as possible on other constraints than the chosen network architecture and the chosen site of plasticity, we consider how the objective functions change as a function of an optimal but small change in the values of the model parameters. Therefore, we slightly vary the relevant model parameters along the gradient of the objective function, similar to Nakahara et al. (2001), but without further constraints.

In this report, we consider the two objective functions,  $J(\theta = 0.5)$  and  $\int J(\theta) d\theta$ . The first objective function quantifies how well the stimulus  $\theta = 0.5$  is encoded. It is an example for the task to improve coding accuracy for a small set of relevant orientations, which may happen, for example, during a perceptual learning experiment. The second objective function quantifies how well all stimuli are encoded. It is an example for the task to improve coding accuracy overall, which may happen in a spatial attention experiment. The derivatives of  $J_i(\theta)$  w. r. t., the conductance  $G_{rs}^{\text{rec}}$  at the recurrent synapses, the conductance  $G_r^{\text{aff}}$  of the afferent synapses, the gain  $a_r^E$  of an excitatory neuron  $r$ , and the additive input  $I_r^{\text{add}}$  are given in the Appendix (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material) (Table 1).

## Results

In this section, we report the changes in the four sets of network parameters and the resulting changes in the neuronal responses

**Table 1. Summary of the vector notation used in the Appendix to derive expressions for the gradients of the Fisher information objective function**

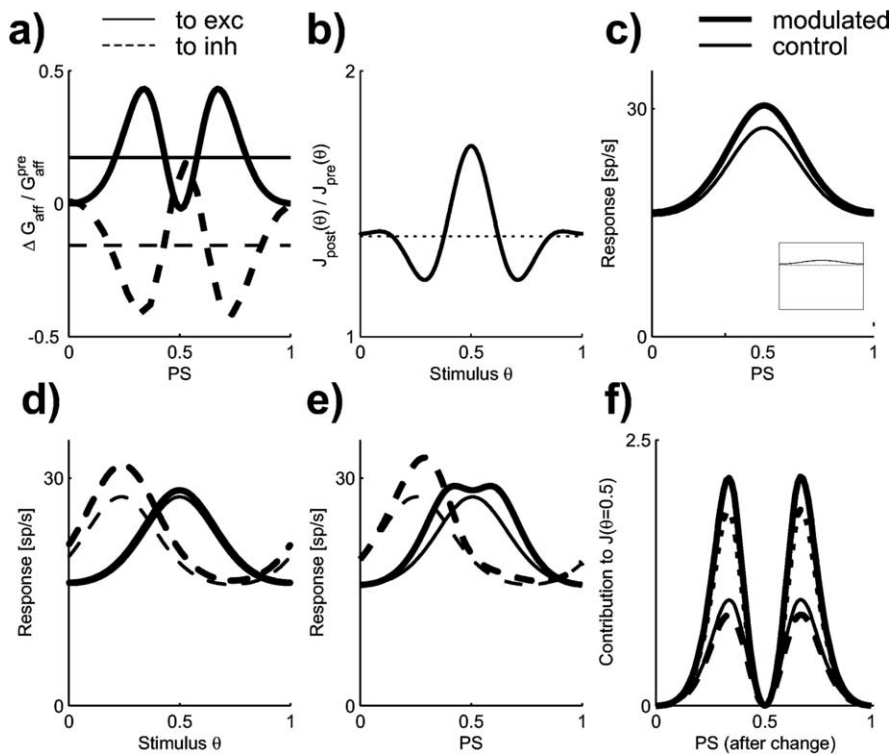
Symbol	Definition	Symbol	Definition
$\tilde{f}$	$\tilde{f}_i = \partial f_i / \partial \theta$	$\mathbf{F}$	$f_i = \partial f_i / \partial I_i$
$\tilde{f}^{\text{aff}, r}$	$\tilde{f}_i^{\text{aff}, r} = \partial f_i / \partial G_r^{\text{aff}}$	$\mathbf{F}^{\text{aff}, r}$	$F_i^{\text{aff}, r} = \partial^2 F_i / \partial G_r^{\text{aff}} \partial I_i$
$\tilde{f}^{\text{rec}, rs}$	$\tilde{f}_i^{\text{rec}, rs} = \partial^2 f_i / \partial G_{rs}^{\text{rec}} \partial \theta$	$\mathbf{F}^{\text{rec}, rs}$	$F_i^{\text{rec}, rs} = \partial^2 F_i / \partial G_{rs}^{\text{rec}} \partial I_i$
$\tilde{f}^{\text{rec}, rs}$	$\tilde{f}_i^{\text{rec}, rs} = \partial f_i / \partial G_{rs}^{\text{rec}}$	$\mathbf{F}^{\text{mul}, r}$	$F_i^{\text{mul}, r} = \partial^2 F_i / \partial a_r^E \partial I_i$
$\tilde{f}^{\text{rec}, rs}$	$\tilde{f}_i^{\text{rec}, rs} = \partial^2 f_i / \partial G_{rs}^{\text{rec}} \partial \theta$	$\mathbf{F}^{\text{add}, r}$	$F_i^{\text{add}, r} = \partial^2 F_i / \partial I_r^{\text{add}} \partial I_i$
$\tilde{f}^{\text{mul}, r}$	$\tilde{f}_i^{\text{mul}, r} = \partial f_i / \partial a_r^E$	$\tilde{I}$	$\tilde{I}_i = \partial I_i^{\text{aff}} / \partial \theta$
$\tilde{f}^{\text{mul}, r}$	$\tilde{f}_i^{\text{mul}, r} = \partial^2 f_i / \partial a_r^E \partial \theta$	$\tilde{I}^r$	$\tilde{I}_i^r = \partial I_i^{\text{aff}} / \partial G_r^{\text{aff}}$
$\tilde{f}^{\text{add}, r}$	$\tilde{f}_i^{\text{add}, r} = \partial f_i / \partial I_r^{\text{add}}$	$\tilde{I}^r$	$\tilde{I}_i^r = \partial^2 I_i^{\text{aff}} / \partial G_r^{\text{aff}} \partial \theta$
$\tilde{f}^{\text{add}, r}$	$\tilde{f}_i^{\text{add}, r} = \partial^2 f_i / \partial I_r^{\text{add}} \partial \theta$	$\mathbf{h}^r$	$h_i^r = \delta_{ri}$
		$\mathbf{h}^{\text{rs}}$	$h_i^{\text{rs}} = \delta_{ri} V_{rs} g_{rs}^{\infty}$
		$\tilde{\mathbf{h}}^{\text{rs}}$	$\tilde{h}_i^{\text{rs}} = \delta_{ri} V_{rs} (\partial F_i / \partial I_i) (\partial f_i / \partial \theta)$

for an optimal increase in the quality of the representation. As a measure of quality, we use the Fisher information for a particular stimulus value as well as the integral of the Fisher information terms over all stimulus values. The first case corresponds, for example, to the task of improving the ability to judge orientations close to a reference orientation (e.g., vertical). The second case corresponds, for example, to the task of improving this ability for all oriented stimuli such as at a particular visual field location. We choose the initial parameterization of our network such that it reproduced the “control” responses in area V4 of the macaque as reported by Yang and Maunsell (2004). We checked whether the reported response changes were different when the initial parameters were chosen differently (“Mexican-hat”-like recurrent connections with weaker and stronger values for the maximum conductances), but we found no qualitative differences.

### Plasticity at afferent synapses

We first asked how to change the maximum conductances  $G_i^{\text{aff}}$  of the afferent synapses to excitatory and inhibitory neurons to increase the Fisher information of the network specifically for  $\theta = 0.5$ . We computed the gradient of the corresponding objective function w. r. t., the maximum conductances of the afferent synapses, and changed their values by a small amount,  $\Delta G_i^{\text{aff}}$ , proportional to this gradient (see Materials and Methods). Figure 3*a* shows the  $\Delta G_i^{\text{aff}}$  normalized by the unadapted initial values  $\Delta G_{i, \text{pre}}^{\text{aff}}$  as a function of the PS of the postsynaptic neuron  $i$ . The changes were strongest for neurons with the PS differing approximately  $\pm 0.16$  from  $\theta = 0.5$ . For excitatory neurons with these preferred stimuli, the afferent synapses became stronger (thick solid line), whereas the synapses to inhibitory neurons with these preferred stimuli became weaker (thick dashed line). The afferent synapses to excitatory neurons with the PS very close to  $\theta = 0.5$  do not change, but inhibitory neurons with these preferred stimuli receive slightly more excitation via their afferents after the adjustment. We also asked how to change the afferent synapses to increase the Fisher information for all stimuli. The resulting changes were uniform. All synapses to excitatory neurons became stronger (thin solid line), and all synapses to inhibitory neurons became weaker (thin dashed line). Figure 3*b* compares the population Fisher information before and after the adjustments and demonstrates that for the stimulus-specific changes, the strongest increase was for the stimulus  $\theta = 0.5$  (solid line). However, the performance also increased for stimuli close to  $\theta = 0.5$ , because the neurons that increase their contribution to the encoding of this stimulus also contribute to the encoding of the nearby stim-





**Figure 3.** Adjusting the afferent synapses. *a*, Predicted changes in the afferent synapses to excitatory (exc; solid line) and inhibitory (inh; dashed line) neurons to optimally increase the objective functions  $J(\theta = 0.5)$  (thick lines) and  $\int J(\theta)d\theta$  (thin lines) (see Materials and Methods). These changes were computed by following the gradient with a step size  $\eta = 10^{-8}$ . *b*, Relative change in the Fisher information after having adjusted the afferent synapses for the  $J(\theta = 0.5)$  and  $\int J(\theta)d\theta$  objective functions (solid and dotted lines). *c*, Population response before and after (thin and thick lines) having made the adjustment to increase  $\int J(\theta)d\theta$ . In this case, the shape of the population response after the adjustment is the same as the shape of the tuning functions of the individual neurons. The inset shows the ratio of the responses after the adjustment to the responses with the initial parameterization (solid line; dotted line marks a ratio of 1). *d*, Tuning functions for two neurons with preferred stimuli of 0.25 and 0.5 (dashed and solid lines) before and after (thin and thick lines) optimizing  $J(\theta = 0.5)$ . *e*, Population response to the same two stimuli,  $\theta = 0.25$  and  $\theta = 0.5$  (thin and thick lines), for the same adjustment. *f*, Fisher information of the individual neurons for  $\theta = 0.5$  before and after (thin solid and thick dotted lines) optimizing  $J(\theta = 0.5)$ . The solid lines show the Fisher information of the reparameterized network if it were only attributable to the changes in the response magnitude [ $J_{\text{add}}(\theta = 0.5)$ ; thick dashed line] and tuning function slopes [ $J_{\text{slp}}(\theta = 0.5)$ ; thick solid line] at  $\theta = 0.5$  (see Materials and Methods for details). The normalization is w. r. t. the maximal Fisher information before the reparameterization.

uli. For the uniform change, the Fisher information was also increased in a uniform manner (dotted line).

For both the stimulus-specific and the uniform changes in the afferent synapses, the Fisher information increased, because the tuning functions were changed. If the adjustment was uniform, then the initially “symmetric” network parameterization (all neurons and associated connections have equal values of their model parameters) remained symmetric. For symmetric parameterizations, the population response to a stimulus has the same shape as the tuning functions of the neurons. However, for stimulus-specific adjustments, this is no longer the case. Therefore, here as well as for the other three mechanisms, we report only the population responses for uniform adjustments, but for stimulus-specific adjustments, we show both the population response and the tuning functions.

Figure 3c shows how the population response of the excitatory neurons to the stimulus  $\theta = 0.5$  (and hence the shape of all tuning functions) changed after the uniform adjustment of the afferent synapses. All excitatory neurons received more afferent excitation, and all inhibitory neurons received less afferent excitation. Without recurrent connections, this would have caused a multiplicative effect on all stimulus-driven activations, but because of

the recurrent interactions shaping the activation profile, the activity increase is stronger for neurons with already high activity. The inset of Figure 3c (thick line) shows the ratio of the population response after the adjustment to the initial response. The changes are not strictly multiplicative, because a multiplicative change would correspond to a horizontal line.

Figure 3, *d* and *e*, shows examples of tuning functions and population responses after the stimulus-specific adjustments. Figure 3*d* shows the tuning functions for two neurons with PS 0.5 and 0.25 (solid and dashed lines) before and after the reparameterization (thin and thick lines). The peak activation of the neuron with PS = 0.5 increased only very little but showed a much stronger increase for the neuron with PS = 0.25. The tuning functions gave rise to population activations shown in Figure 3*e*. Here, two stimuli with  $\theta = 0.5$  and  $\theta = 0.25$  were used (solid and dashed lines). With the initial parameterization (thin lines), the population activations for every stimulus had the same shape; only the location of the peak activity was dependent on the stimulus. However, after the reparameterization, the shape of the population activation depends on the stimulus. For example, the activation profile for  $\theta = 0.5$  became bimodal, because the tuning functions for neurons with a PS close to 0.5 were not changed, whereas the peak amplitudes of neurons with a PS differing by  $\Delta\theta = \pm 0.16$  from  $\theta = 0.5$  increased. For  $\theta = 0.25$ , the profile was again unimodal. The peak activity increased and was slightly shifted toward  $\theta = 0.5$ .

Let us now analyze how the reparameterized network achieved its increase in the quality of the representation of the stimuli around  $\theta = 0.5$ . Figure 3*f* (thin line) shows the contribution of every neuron to the population Fisher information for the initial parameterization (normalized to the maximal contribution) (compare Fig. 2*b*) as well as the contribution after the reparameterization (dotted line; relative to the normalization). Neurons with an initially high contribution increased their Fisher information even more, whereas the contribution of neurons with initially low Fisher information for  $\theta$  remained low.

For conditionally independent Poisson spike trains, the Fisher information of a single neuron  $i$  for a stimulus  $\theta$ ,  $f_i(\theta)$ , is proportional to  $[f_i'(\theta)]^2/f_i(\theta)$ , where  $f_i(\theta)$  and  $f_i'(\theta)$  are the tuning function of that neuron and its derivative (see Materials and Methods). Figure 3*f* shows how much of the overall improvement is attributable to changes in the response magnitudes and the slopes. For neurons with a high contribution to the population Fisher information, the value  $J_{\text{add}}(\theta = 0.5)$  after the change is below the value  $J(\theta = 0.5)$  before the change (thick dashed vs thin solid line), whereas the value  $J_{\text{slp}}(\theta = 0.5)$  after the change is above the value  $J(\theta = 0.5)$  before the change (thick solid vs dotted lines). The increased response magnitudes would have caused a decrease in the encoding accuracy, but this effect was compen-

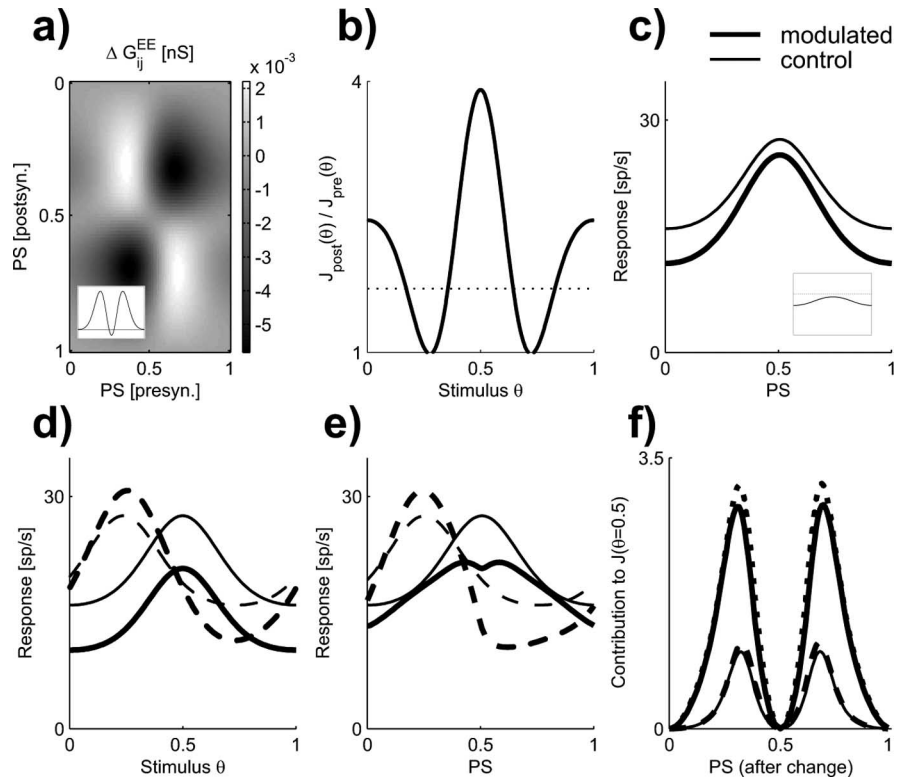
sated by the increase in the slopes at  $\theta = 0.5$ , and overall the encoding accuracy for  $\theta = 0.5$  was increased.

### Plasticity at recurrent synapses

Let us now consider the consequences of only adjusting the conductances at the recurrent synapses. We first adjusted the synapses to increase the population Fisher information specifically for the stimulus  $\theta = 0.5$  (see Materials and Methods). Figure 4*a* shows the values  $\Delta G_{ij}^{EE}$  by which we changed the synapses at the recurrent connections between excitatory neurons as a function of the presynaptic and postsynaptic initial PS. For neurons with a PS different from  $\theta = 0.5$ , synaptic changes depend on the PS of the postsynaptic neuron. The excitation from presynaptic neurons with a PS similar to the PS of the postsynaptic neuron increased, and the increase was strongest for postsynaptic neurons with the PS differing approximately  $\pm 0.16$  from  $\theta = 0.5$  (Fig. 4*a*, inset). The excitation from presynaptic neurons with a PS different from the PS of the postsynaptic neuron decreased, and the decrease was also strongest for postsynaptic neurons with the PS differing approximately  $\pm 0.16$  from  $\theta = 0.5$ . The changes for the other three types of recurrent connections (EI, IE, and II) are complementary to the changes shown in Figure 4*a*. Where the excitation of excitatory neurons increased, the inhibition of excitatory neurons decreased, the excitation of inhibitory neurons decreased, and the inhibition of inhibitory neurons increased.

These adjustments caused changes in the tuning functions, which in turn gave rise to the population Fisher information shown in Figure 4*b*. For the objective function  $J(\theta = 0.5)$ , the strongest increase was for stimuli around  $\theta = 0.5$  (solid line), whereas the increase was “uniform” for the uniform objective function  $\int J(\theta)d\theta$  (dotted line). For the latter, the population response to a stimulus with  $\theta = 0.5$  is shown in Figure 4*c*. The responses of all neurons in the reparameterized network were lower compared with the responses with the initial parameterization (thick vs thin line). The shape of the population activation (and hence of the individual tuning functions) changed as well, and overall the change was not strictly multiplicative (Fig. 4*c*, inset). The reduced activity and the shape change resulted in an increased encoding accuracy for all stimuli.

Figure 4, *d* and *e*, shows examples of tuning functions and population responses after improving  $J(\theta = 0.5)$ . Figure 4*d* shows the tuning functions for two neurons with PS 0.5 and 0.25 (solid and dashed lines) before and after the reparameterization (thin and thick lines). The peak activity of the first neuron decreased, but the shape of its tuning function remained unchanged, whereas the tuning function of the second neuron became sharper and its peak activity increased. These tuning functions explain the population responses of the reparameterized network to the two stimuli  $\theta = 0.5$  and  $\theta = 0.25$  shown in Figure 4*e*. The profile of the response to  $\theta = 0.5$  was bimodal and below the initial responses, whereas the response profile for  $\theta = 0.25$  became sharper and its peak activity increased. Figure 4*f* shows the



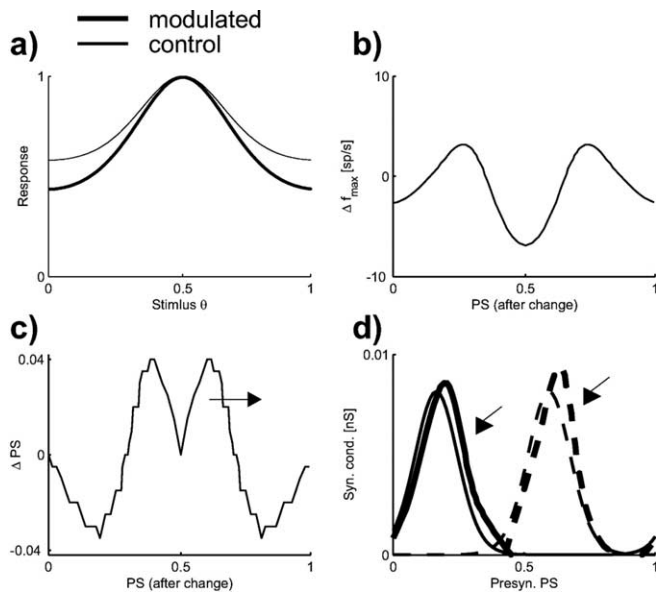
**Figure 4.** Adjusting the recurrent synapses. *a*, Predicted changes in the recurrent excitatory synapses to excitatory neurons for optimizing  $J(\theta = 0.5)$ . The inset shows the change in the self-excitation (the horizontal line indicates no change). These changes were computed by following the gradient of the objective function with a step size  $\eta = 3 \times 10^{-6}$ . postsyn., Postsynaptic; presyn., presynaptic. *b–f*, Same as in Figure 3*b–f*.

two “hypothetical” Fisher information terms  $J_{\text{add}}(\theta = 0.5)$  and  $J_{\text{slp}}(\theta = 0.5)$  (see Materials and Methods) before the reparameterization as well as the Fisher information  $J(\theta = 0.5)$  after the reparameterization. In contrast to the simulations in which only the afferent synapses were adjusted, the reparameterization of the recurrency changed the tuning functions so that now the  $J_{\text{add}}(\theta = 0.5)$  are above the initial  $J(\theta = 0.5)$  (thick dashed vs thin solid line). The  $J(\theta = 0.5)$  for the reparameterized network are above the  $J_{\text{slp}}(\theta = 0.5)$  (thick dotted vs. thick solid line), because now both the decreased response magnitudes and the changes in the slopes contributed to increasing the encoding accuracy for  $\theta = 0.5$ .

Figure 5*a* compares the shape of the tuning functions averaged over all excitatory neurons before (thin line) and after (thick line) the recurrency was adjusted. In addition to this sharpening, the peak responses were modulated depending on the PS of the neurons (Fig. 5*b*), and the preferred stimuli themselves were also changed (Fig. 5*c*). The underlying synaptic mechanisms for the shifts of the PS are shown in Figure 5*d* for the two neurons with maximal shifts of their PS toward and away from  $\theta = 0.5$  (solid and dotted lines). After the adjustment, the neuron that shifted its PS toward  $\theta = 0.5$  received more excitation from neurons with a PS closer to  $\theta = 0.5$  (Fig. 5*d*, left arrow). The neuron that shifted its PS away from  $\theta = 0.5$  received more excitation from neurons with the PS differing even more from  $\theta = 0.5$  (Fig. 5*d*, right arrow) after the reparameterization.

### Changing the gain of excitatory neurons

Another mechanism we considered is the adjustment of the gain for the excitatory neurons. We first asked how to change the gains  $a_i^E$  (see Materials and Methods) to increase the Fisher informa-



**Figure 5.** Differential changes in tuning functions after having adjusted the recurrent synapses to increase the Fisher information specifically for  $\theta = 0.5$ . **a**, All tuning functions, normalized to their peak value and aligned so that their PSs coincide, were pooled and are shown before (thin line) and after (thick line) the adjustment of the recurrent synapses. **b**, Change of the peak response after the adjustment. Positive values correspond to an increase in the response. **c**, Changes in the PS of the excitatory neurons. Positive values correspond to a shift away from 0.5. **d**, Strength of the recurrent excitation of two excitatory neurons before (thin lines) and after (thick lines) the adjustment. The first neuron (solid lines) had the maximal shift of its PS toward 0.5, and the second neuron (dashed lines) had the maximal shift of its PS away from 0.5. Note that after the adjustments to the recurrency, the PS also changed. PS denotes the optimal stimulus and not the afferent PS (see Materials and Methods) used to determine the afferent input. Syn. cond., Synaptic conductance.

tion of the network specifically for the stimulus  $\theta = 0.5$  (Fig. 6a, solid line). One possible realization of this gain modulation is to change the variance of the balanced background inputs (Chance et al., 2002), which could be realized rapidly by, for example, adjusting top-down feedback inputs. Similar to the case of changing the strength of afferent inputs to the excitatory neurons, the gains were increased mainly for neurons with the PS differing approximately  $\pm 0.16$  from  $\theta = 0.5$ . The encoding accuracy is enhanced around  $\theta = 0.5$ . The changes  $\Delta a_i^E$  necessary to increase the Fisher information of the network for all stimuli were again uniform (Fig. 6a, dotted line, b).

In contrast to the previous two mechanisms, the uniform increase in the gains for the objective function  $\int J(\theta)d\theta$  resulted in a strictly multiplicative modulation of the population response to a stimulus ( $\theta = 0.5$ ) (Fig. 6c and inset). If  $J(\theta)$  is optimized, the individual tuning functions and population responses are similar, but not identical, to the changes induced when adjusting the strength of the afferent synapses. The responses for the neuron with PS = 0.25 (Fig. 6d, dashed lines) increased for all stimuli, but the tuning function for the neuron with PS = 0.5 was unaffected (solid lines). The population responses to a stimulus with  $\theta = 0.5$  were also bimodal, and the peak activation of the responses to a stimulus with  $\theta = 0.25$  was increased and slightly shifted toward  $\theta = 0.5$  (Fig. 6e). Figure 6f also parallels the twofold effects of the changed tuning functions on the encoding accuracy (compare Fig. 3f). For neurons with an already high contribution to the population Fisher information, the values of  $J_{\text{add}}(\theta = 0.5)$  are below the values of  $J(\theta = 0.5)$  for the initial parameterization, whereas the values of  $J_{\text{slp}}(\theta = 0.5)$  are higher than the initial values of  $J(\theta = 0.5)$ . Thus, the increased slopes at  $\theta = 0.5$  com-

pensated the reduced Fisher information because of the increase in activity, and encoding accuracy was enhanced for  $\theta = 0.5$ .

### Changing the additive feedback inputs

The last mechanism we consider is the adjustment of the additive input currents  $I_i^{\text{add}}$  for both the excitatory and inhibitory neurons, as given by the gradient of the objective function w. r. t.  $I_i^{\text{add}}$ . Figure 7a shows how the input to the excitatory (solid lines) and inhibitory (dashed lines) neurons were changed to increase the population Fisher information specifically for  $\theta = 0.5$  (thick lines) as well as for all stimuli (thin lines), respectively. The additive input currents to all excitatory neurons were decreased, whereas they were increased for inhibitory neurons. For the objective function  $J(\theta = 0.5)$ , the strongest reductions of the inputs to excitatory neurons were for neurons with the PS differing approximately  $\pm 0.16$  from  $\theta = 0.5$ . Inhibitory neurons with those preferred stimuli had the strongest increase in their additive inputs. To increase the Fisher information for all stimuli, the additive inputs had to be changed in a uniform manner (thin lines). These changes resulted in an increase in the encoding accuracy around  $\theta$  and for all stimuli (Fig. 7b).

Because of the reduced excitation and increased inhibition, the population response (and the tuning functions) shown in Figure 7c was reduced. This reduction for the case of the objective function  $\int J(\theta)d\theta$  was strictly subtractive. For the case of the objective function  $J(\theta = 0.5)$ , the tuning functions were also shifted toward lower activation levels. These stimulus-specific changes gave rise to the population responses to the two stimuli  $\theta = 0.5$  and  $\theta = 0.25$  as shown in Figure 7e. Because the reduction of excitation and the increased inhibition were strongest for neurons with the PS differing approximately  $\pm 0.16$  from  $\theta = 0.5$ , the strongest reduction of the responses was observed for those stimuli. Changing the values of  $I_i^{\text{add}}$  led to a change in the value of  $J_{\text{add}}(\theta = 0.5)$  but not of  $J_{\text{slp}}(\theta = 0.5)$ . Figure 7f shows that the  $J_{\text{slp}}(\theta = 0.5)$  (after the reparameterization) and the  $J(\theta = 0.5)$  (before the reparameterization) are identical (the thin and thick solid lines are superimposed). In other words, when adjusting only the additive inputs, the improvement of the encoding accuracy around  $\theta = 0.5$  is only attributable to the subtractive shifts of the tuning functions.

## Discussion

In this section, we first discuss the main finding of this report: that an increased encoding accuracy for a continuous stimulus variable can be achieved via different mechanisms which then result in different changes in the stimulus tuning functions. Then we relate our predicted changes in neuronal responses to experimentally observed changes during attentional modulations and perceptual learning. These two phenomena happen on distinct time scales and lead to different kinds of tuning function changes. We will show, however, that the observed changes are broadly consistent with the hypotheses that visual attention and perceptual learning can be explained by the common principle of optimally encoding sensory information and that the differences observed are a result of different plasticity mechanisms being operative. Finally, we discuss the limitations of our modeling approach.

### Tuning function changes and adaptation mechanisms

The Fisher information  $J(\theta)$  can be increased by an increase in the slopes  $f_i'(\theta)$  of the tuning functions and by a decrease in the activities  $f_i(\theta)$  for these stimuli, because the contribution of the  $i$ th neuron is proportional to  $[f_i'(\theta)]^2/f_i(\theta)$ . Multiple strategies exist to adjust their



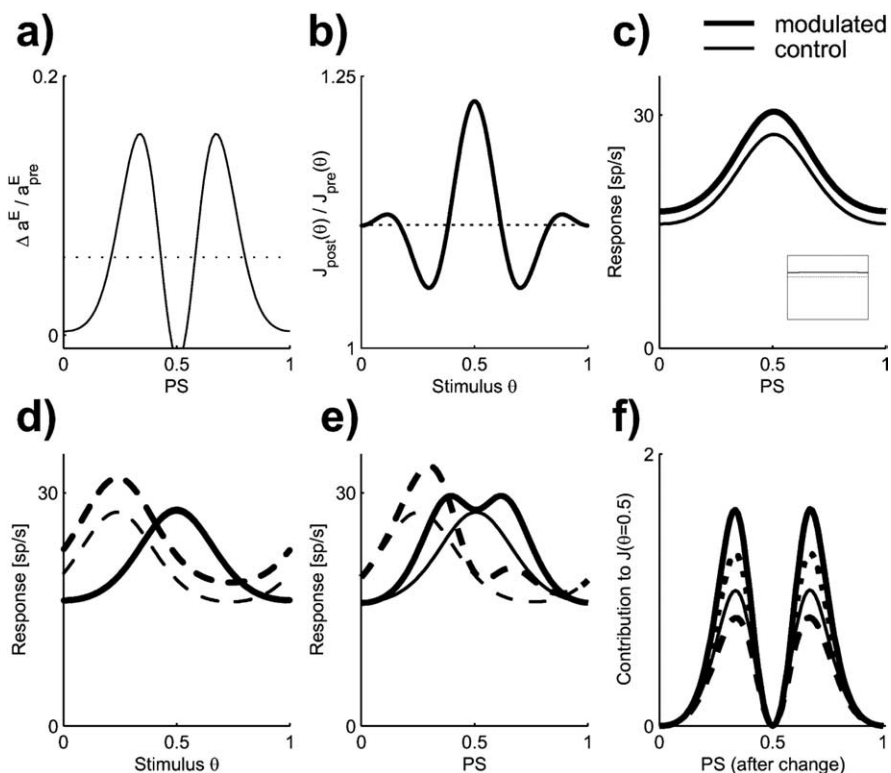
values. For example, the slope for a particular value of  $\theta$  could be increased by a multiplicative scaling of the tuning function or by shifting it toward lower or higher stimulus values.

Interestingly, our model predicted that not all mechanisms lead to optimal changes in both the slopes and the activation levels simultaneously. In two cases, the physiological constraints (the model architecture) resulted in a “compromise.” When changing the gain of the excitatory neurons or the strength of the afferent synapses, the increase in activity, which resulted in a decreased encoding accuracy, was compensated by an increase in the slopes (compare Figs. 3*f*, 6*f*). However, when adjusting the recurrent connections, both increase in the slopes and decrease in the activation levels contributed to the improvement of the Fisher information (Fig. 4*f*). In this case, the changes in the slopes were achieved via shifting the PS, sharpening of the tuning functions, or differentially adjusting the response amplitudes. When the additive feedback inputs were changed, the slopes remained unaffected, but the activations of all neurons were reduced (Fig. 7*c–f*) to increase performance.

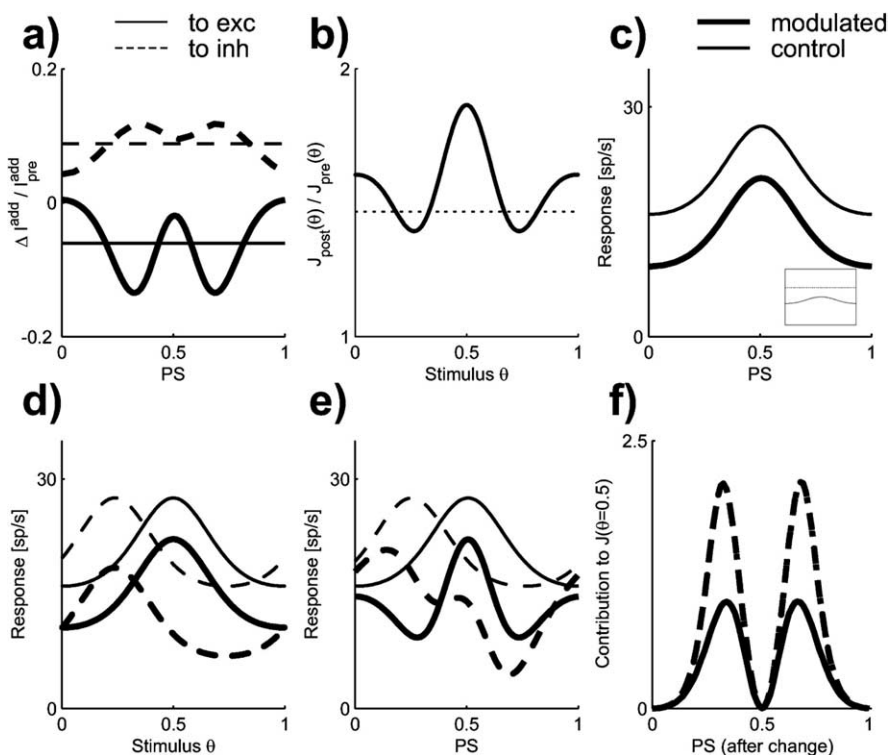
**Model predictions and perceptual learning**

Schoups et al. (2001) investigated the physiological correlate of perceptual learning in V1. They found that after monkeys were trained on an orientation discrimination task, the perceptual improvements were specific for location and orientation. The physiological correlate was an activity reduction for cells with preferred orientations (POs) around the learned orientation (Schoups et al., 1998) and an increase in the slopes of the tuning function at the learned orientation for neurons with POs differing approximately 20° from that orientation. When optimizing performance as measured by the objective function  $J(\theta = 0.5)$ , our model predicts an activity reduction for neurons with the PS close to the learned stimulus when assuming the recurrent connections or the additive (feedback) inputs as the location of plasticity but not when adjusting the afferent synapses, the recurrent synapses, or the single neuron gains. Furthermore, our model predicts an increase in the slopes of the tuning function for the learned stimulus when the afferent synapses, the recurrent synapses, or the single neuron gains were modified but not when adjusting the additive inputs.

Another group performed a similar experiment (Ghose et al., 2002) and found that the perceptual improvements were



**Figure 6.** Adjusting the gain of the excitatory neurons. *a*, Predicted changes in the gains of the excitatory neurons for optimizing  $J(\theta = 0.5)$  (solid line) and  $\int J(\theta)d\theta$  (dotted line). The changes were computed by following the gradient of the objective function with a step size  $\eta = 5$ . *b–f*, Same as in Figure 3*b–f*.



**Figure 7.** Adjusting the additive feedback input. *a*, Predicted changes in the additive feedback input to excitatory and inhibitory neurons (solid and dashed lines) to increase the objective functions  $J(\theta = 0.5)$  and  $\int J(\theta)d\theta$  (thick and thin lines). These changes were computed by following the gradient of the objective function with a step size  $\eta = 2 \times 10^{-4}$ . *b–f*, Same as in Figure 3*b–f*.



orientation specific but transferred between retinotopic locations. Similar to Schoups et al. (2001), Ghose et al. (2002) reported an activity reduction for neurons with POs close to the learned orientation. However, no increase in the tuning function slopes was found, but shifts of the POs were reported as predicted by our model when adjusting the recurrent connections. The reason for the discrepancy between the findings of the two groups is not clear.

The same group performed a similar experiment while recording neurons in area V4 (Yang and Maunsell, 2004). They reported a sharpening of orientation tuning functions and an orientation-dependent change in the response amplitude with the largest increase in the responses for neurons with POs differing from the learned orientation but almost no increase for neurons with a PO close to the learned orientation. When adjusting the recurrent connections, our model predicts a stimulus-dependent increase in the response amplitude for neurons with the PS differing from the learned stimulus and an increase in the slopes of the tuning functions, both of which are consistent with the data. However, model results are not fully consistent with the observed lack of change in activity at the learned orientation and better fit with the V1 data of Schoups et al. (1998) and Ghose et al. (2002).

Additionally, optimally changing the recurrent connections predicts shifts of the PS. Unfortunately, shifts of the tuning functions can be addressed experimentally only indirectly (e.g., by investigating the histograms of the POs), because the time scale of perceptual learning is too long for tracking response properties of an individual neuron. The histograms of POs shown by Yang and Maunsell (2004) are not uniform after learning, but as to whether this change is statistically significant needs to be tested.

In summary, the reported physiological correlates of perceptual learning in the visual cortex are by themselves diverse and seem to depend on the visual area. They are, however, mostly (but not completely) consistent with the model predictions, if the recurrent connections are changed to improve performance.

### Model predictions and attentional modulations

One of the most frequently reported physiological correlates of attention in the visual cortex is an increase in the stimulus-driven neuronal activity compared with control trials. As described by Treue and Martinez Trujillo (1999), the effects of spatial and feature-based attention in area MT were disentangled and were shown to contribute independently to the observed increase in activity. The effects of attention on the direction tuning curves of neurons in area MT were reported to be approximately multiplicative. Such a separation into a spatial and a feature-based component was also found in area V4 (McAdams and Maunsell, 2000), as well as an approximately multiplicative modulation of the entire orientation tuning function, presumably mainly because of spatial attention (McAdams and Maunsell, 1999). In none of these studies, a sharpening of stimulus tuning curves was reported.

Our model predicted that, to increase the encoding accuracy for all stimuli, strictly multiplicative changes in the tuning functions and the population responses are to be expected only when adjusting the neuronal gain (Fig. 6c, inset). However, an approximately multiplicative change would also be compatible with the changes predicted when adjusting the afferents (Fig. 3c, inset). Of course, adjusting the afferent synapses during attentional modulation via mechanisms such as long-term potentiation/long-term depression is out of question, but a possible mechanism could be an effective increase in the impact of feedforward inputs because

of synchronous activity in a lower area. Adjusting the recurrent connections or the additive (feedback) inputs disqualify as possible mechanisms, because they lead to a decrease in neuronal activity.

So far, no study has directly tested for how attention to particular stimulus values, which would correspond to optimizing an objective function like  $J(\theta = 0.5)$ , changes stimulus tuning functions. The study that comes closest is the study by Treue and Martinez Trujillo (1999). There, the monkey attended to a particular stimulus direction during a direction discrimination task. The authors reported an increased activity, when attention was allocated to the presented stimulus. When the afferent synapses or the gains of the excitatory neurons are assumed as the site of plasticity, our model predicts that activity increases but that this increase is small compared with the strongest changes, which are predicted for neurons with the PS differing approximately  $\pm 0.16$  from the currently relevant stimulus 0.5 (Figs. 3d, 4d, 6d, 7d, dashed lines). Those changes, however, were not investigated by Treue and Martinez Trujillo (1999). One reason for the discrepancy could be that in our model the parameters were changed for an increase in the Fisher information for only a single value. It is conceivable that such a very specific modulation is not achievable with the neuronal circuits in the visual cortex or that the range of stimuli actually selected to be represented more accurately is broader. Changes in the recurrent connections and the additive inputs lead to a decrease in neuronal activity and are therefore inconsistent with the data.

In summary, one prominent physiological correlate of attentional modulations is an increase in activity for neurons responding to the attended stimulus. Approximately multiplicative modulations of tuning functions are consistent with our model predictions derived from optimizing the objective function  $\int J(\theta)d\theta$ , if the gain of the excitatory neurons or afferent synapses are adjusted to improve performance. When optimizing the objective function  $J(\theta = 0.5)$ , for all the mechanisms we investigated, we predict the strongest changes for neurons with a PS different from 0.5, which so far has not been tested directly in experiments.

### Model limitations

In our contribution, we have considered changes in activity only for the mechanisms at the “different sites of plasticity” being operative individually. Because different mechanisms can lead to opposing changes in activity but still improve performance, it is conceivable that when considering the combined action of multiple mechanisms, some of the discrepancies between model predictions and experimental data, which were mentioned in the last sections, can be resolved.

One limitation of the encoder/decoder framework used here could be the fact, that Fisher information is not always the proper quality measure of a neuronal representation. Bethge et al. (2003), for example, demonstrated that the Fisher information as a quality measure fails if the time window for decoding is very short. For very short decoding time windows, however, the dynamics of the encoding process must be considered, and our model, which was not intended to describe the activity dynamics, is no longer applicable. Xie (2002) showed that optimal maximum-likelihood decoding cannot be achieved if only a few neurons are available for representation. However, given that within the visual cortex larger populations of neurons are believed to encode visual information, this may not be a severe limitation (Feldman, 1984). Finally, a neural system may not be able to make optimal use of the information in its activity pat-

terns, and neuronal structures may not be able to implement every conceivable type of optimal decoding. However, it has been shown that recurrent networks can implement maximum-likelihood decoding procedures (Deneve et al., 1999) for the case of continuous variables and several types of statistical models.

Our modeling framework applies to the case in which the relevant property of the environment is a continuous variable and its value has to be determined or to be discriminated from another one. Assuming the validity of the optimal encoding hypothesis, the model can then be used to disentangle the mechanisms of perceptual learning, attention, and possibly other adaptation phenomena in the visual areas. Because the model is a generic cortex model, our predictions may transfer to other continuous stimulus domains or even to the motor cortex, which is also highly adaptive in the adult (Paz and Vaadia, 2004). When discrete stimuli are considered for the perceptual tasks, however, other optimality criteria (e.g., the classification error for particular classifiers or the mutual information between the stimuli and the neuronal responses) need to be considered.

## References

- Bethge M, Rotermund D, Pawelzik K (2003) Optimal neural rate coding leads to bimodal firing rate distributions. *Network* 14:303–319.
- Brunel N, Nadal JP (1998) Mutual information, Fisher information, and population coding. *Neural Comput* 10:1731–1757.
- Chance FS, Abbott LF, Reyes AD (2002) Gain modulation from background synaptic input. *Neuron* 35:773–782.
- Clifford CW, Wenderoth P, Spehar B (2000) A functional angle on some after-effects in cortical vision. *Proc R Soc Lond B Biol Sci* 267:1705–1710.
- Deneve S, Latham PE, Pouget A (1999) Reading population codes: a neural implementation of ideal observers. *Nat Neurosci* 2:740–745.
- Destexhe A, Rudolph M, Fellous JM, Sejnowski TJ (2001) Fluctuating synaptic conductances recreate in vivo-like activity in neocortical neurons. *Neuroscience* 107:13–24.
- Dragoi V, Sharma J, Miller EK, Sur M (2002) Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. *Nat Neurosci* 5:883–891.
- Feldman ML (1984) Morphology of the neocortical pyramidal neuron. In: *Cerebral cortex, Vol 1* (Peters A, Jones EG, eds), pp 123–200. New York: Plenum.
- Ghose GM, Yang T, Maunsell JH (2002) Physiological correlates of perceptual learning in monkey V1 and V2. *J Neurophysiol* 87:1867–1888.
- Gillespie DT (1996) Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral. *Physiol Rev [E]* 54:2084–2091.
- Kay SM (1993) Fundamentals of statistical signal processing: estimation theory. Englewood Cliffs, NJ: Prentice Hall.
- McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J Neurosci* 19:431–441.
- McAdams CJ, Maunsell JH (2000) Attention to both space and feature modulates neuronal responses in macaque area V4. *J Neurophysiol* 83:1751–1755.
- Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. *Science* 229:782–784.
- Muller JR, Metha AB, Krauskopf J, Lennie P (1999) Rapid adaptation in visual cortex to the structure of images. *Science* 285:1405–1408.
- Nakahara H, Wu S, Amari S (2001) Attention modulation of neural tuning through peak and base rate. *Neural Comput* 13:2031–2047.
- Paradiso MA (1988) A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biol Cybern* 58:35–49.
- Paz R, Vaadia E (2004) Learning-induced improvement in encoding and decoding of specific movement directions by neurons in the primary motor cortex. *PLoS Biol* 2:264–274.
- Schoups A, Vogels R, Orban R (1998) Effects of perceptual learning in orientation discrimination on orientation coding in V1. *Invest Ophthalmol Vis Sci Suppl* 39:684.
- Schoups A, Vogels R, Qian N, Orban G (2001) Practising orientation identification improves orientation coding in V1 neurons. *Nature* 412:549–553.
- Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. *Proc Natl Acad Sci USA* 90:10749–10753.
- Shriki O, Hansel D, Sompolinsky H (2003) Rate models for conductance-based cortical neuronal networks. *Neural Comput* 15:1809–1841.
- Teich AF, Qian N (2003) Learning and adaptation in a recurrent model of V1 orientation selectivity. *J Neurophysiol* 89:2086–2100.
- Treue S, Martinez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399:575–579.
- Xie X (2002) Threshold behaviour of the maximum likelihood method in population decoding. *Network* 13:447–456.
- Yang T, Maunsell JH (2004) The effect of perceptual learning on neuronal responses in monkey visual area V4. *J Neurosci* 24:1617–1626.