Fast efficient coding and sensory adaptation in gain-adaptive recurrent networks

Arthur Prat-Carrabin[1,⋆], Maximilian V. Harl[2,3], and Samuel J. Gershman[1]

[1]Department of Psychology and Center for Brain Science, Harvard University, Cambridge, MA, USA
[2]Neuroscience Center Zurich, University of Zurich and the ETH Zurich, Zurich, Switzerland
[3]TUM School of Computation, Information and Technology, Technical University of Munich, Garching, Germany

⋆*email: arthurpc@fas.harvard.edu*

July 11, 2025

Abstract

As the statistics of sensory environments often change, neural sensory systems must adapt to maintain useful representations. Efficient coding prescribes that neuronal tuning curves should be optimized to the prior, but whether they can adapt rapidly is unclear. Empirically, tuning curves after repeated stimulus presentations exhibit "adapter repulsion", whose underlying mechanism remains uncertain, and which contrasts with the "prior attraction" expected under many efficient-coding models. We propose a gain-adaptive, recurrent sensory network model in which gains optimize a novel efficient-coding objective balancing accuracy and spiking cost. From the propagation of modulated gains throughout the network emerge quickly adaptive tuning curves. The model accounts for subtle adapter-repulsion effects under peaked priors and predicts fast prior attraction under broader distributions, for which we provide supporting behavioral evidence. Our framework reconciles seemingly contradictory adaptive phenomena, under a unified theoretical and mechanistic model of efficient coding mediated by gain modulation in recurrent circuits.

Since Barlow's seminal proposal [1], a prominent idea in neuroscience is that sensory systems are efficiently adapted to the statistics of the stimuli that they represent [2–11]. In many models of neural sensory encoding, such 'efficient coding' is achieved by optimally matching neuronal sensitivities to the prior distribution of stimuli, with more encoding resources allocated to the more frequently encountered stimuli. Typical models of efficient coding posit a distribution of stimuli that is unchanging (perhaps learned over long timescales), producing static encoding schemes. Yet in natural environments the statistics of sensory inputs are not fixed: instead, they shift over time and across contexts. For an organism to maintain accurate perception, its sensory encoding must adapt accordingly, on a relatively short timescale. If the brain is able to quickly implement efficient coding — what we refer to as 'fast efficient coding' — then the tuning curves of sensory neurons should shift towards the most frequent stimuli, when the relative frequencies of stimuli change; a behavior we call 'prior attraction'. A large body of experimental work, however, has reported shifts in tuning curves in the seemingly opposite direction: after repeated exposure to a stimulus (the 'adapter'), neuronal tuning curves often shift *away* from it, i.e., away from the stimulus that has become more frequent [12–17]. This behavior, which we call 'adapter repulsion', is moreover accompanied by several additional effects; for example, tuning curves close to the adapter typically become wider, while those further away become narrower [12–14]. Adapter repulsion occurs over durations on the order of tens of milliseconds, suggesting that neural encoding can be reconfigured fairly quickly.

These observations raise several questions. First, what mechanism underlies the fast modulation of neuronal tuning curves? Second, could the brain leverage this flexibility to implement fast efficient coding, through the prior attraction of tuning curves? Third, why then should we observe adapter repulsion (and the related effects), and how can it be reconciled with prior attraction?

Studies addressing adapter repulsion have generally attributed shifts in neuronal sensitivity to changes in synaptic weights, typically through short-term synaptic depression [12–14, 16–19]. But the implied timescales and exponentially-decaying dynamics seem ill-suited for the demands of efficient coding in dynamic environments, which needs to be both fast and sustained as long as necessary.

An alternative proposal is that adaptation is an emergent property of recurrent neural networks, even without synaptic plasticity [20, 21]. Duong *et al.* (2023) [21] show that tuning shifts can arise from the modulation of neuronal gains in a recurrent network. Inspired by this approach, we introduce a recurrent network model whose neuronal gains optimize an efficient-coding objective, formulated as the sum of a lower bound on the Bayesian mean squared decoding error and a cost proportional to the neuronal spiking activity. We derive analytical approximations for the resulting tuning curves, and for the optimal gains when the prior is wide (in comparison to the width of the tuning curves). When the priors are relatively broad, the neurons' locations (preferred stimuli) under optimal gains shift towards frequently encountered stimuli — thus exhibiting prior attraction. The same optimization framework, when applied to sharply peaked priors (corresponding to the case of a repeated adapter), yields gain profiles that deviate from the prior's peak, with maximal gains located away from it. This distinct gain structure induces repulsion from the adapter, producing tuning shifts consistent with empirical reports of adapter repulsion, but also capturing the subtle related effects, such as the widening of tuning curves near the adapter and their narrowing

further away. Our results unify prior attraction and adapter repulsion as outcomes of a single efficient coding principle, achieved solely through rapid gain modulation, without changes in synaptic connectivity.

The paper is structured as follows. First, we present behavioral evidence, obtained in a pre-registered experiment, showing that human observers rapidly adapt their sensory encoding in response to changing priors, supporting the hypothesis of fast efficient coding. We then provide schematic illustrations of prior attraction and adapter repulsion. The core of the paper focuses on the gain-adaptive recurrent-network model that we propose as a candidate neural mechanism underlying these adaptive phenomena. We show how, through the modulated gains and the recurrent interactions, a neuron's effective location emerges as a weighted average of the locations of all the network's neurons. We then examine Bayesian decoding within this framework, and derive a lower bound on the expected squared error of the Bayesian mean. Next, we formulate an efficient-coding optimization problem and derive its analytical solution in the case of wide priors. This solution yields prior attraction. Turning to narrowly-peaked distributions, we show that the optimal gain configuration in this scenario places maximal gains at some distance from the peak, yielding adapter repulsion. Together, our results provide a mechanistic understanding of how sensory systems may dynamically optimize neural encoding in changing environments, and reconcile two seemingly contradictory adaptive patterns (prior attraction and adapter repulsion) as complementary outcomes of a single efficient-coding principle.

Results

Behavioral evidence of fast efficient coding

To test whether sensory encoding can rapidly adapt to changes in the prior, we conducted a pre-registered, online numerosity-estimation experiment. On each trial, subjects were briefly presented (for 500 ms) with a cloud of dots (Fig. 1A) and then asked to report, using a slider, their best estimate of the number of dots in the cloud. This number was chosen randomly on each trial from one of three uniform prior distributions: the Narrow prior ranged from 50 to 70 dots; the Medium prior, from 40 to 80 dots; and the Wide prior, from 30 to 90 dots. Importantly, the prior was randomly selected at every trial (with equal probabilities). Subjects were informed of the chosen prior at the onset of each trial (this information was displayed for at least one second, before subjects could make a response).

The variability of subjects' estimates increases with the width of the prior (Fig. 1B). With the Narrow prior, the across-subjects average standard deviation of estimates is about 4, while with the Wide prior (which is three times larger), it is around 7 (we estimate these quantities through Bayesian probabilistic methods; see Methods). A Levene's test rejects the equality of the variances across conditions ($p < 0.0001$). More quantitatively, the variance of estimates increases linearly with the prior width, consistent with previous results [10]. This scaling is predicted by an efficient-coding model formalizing a trade-off between the precision of estimates and the cost of this precision [10]. Our data thus suggest that in the context of numerosity representations, subjects implement fast efficient coding: they reorganize their representational resources on the order of one second, to adapt to changing stimuli statistics.

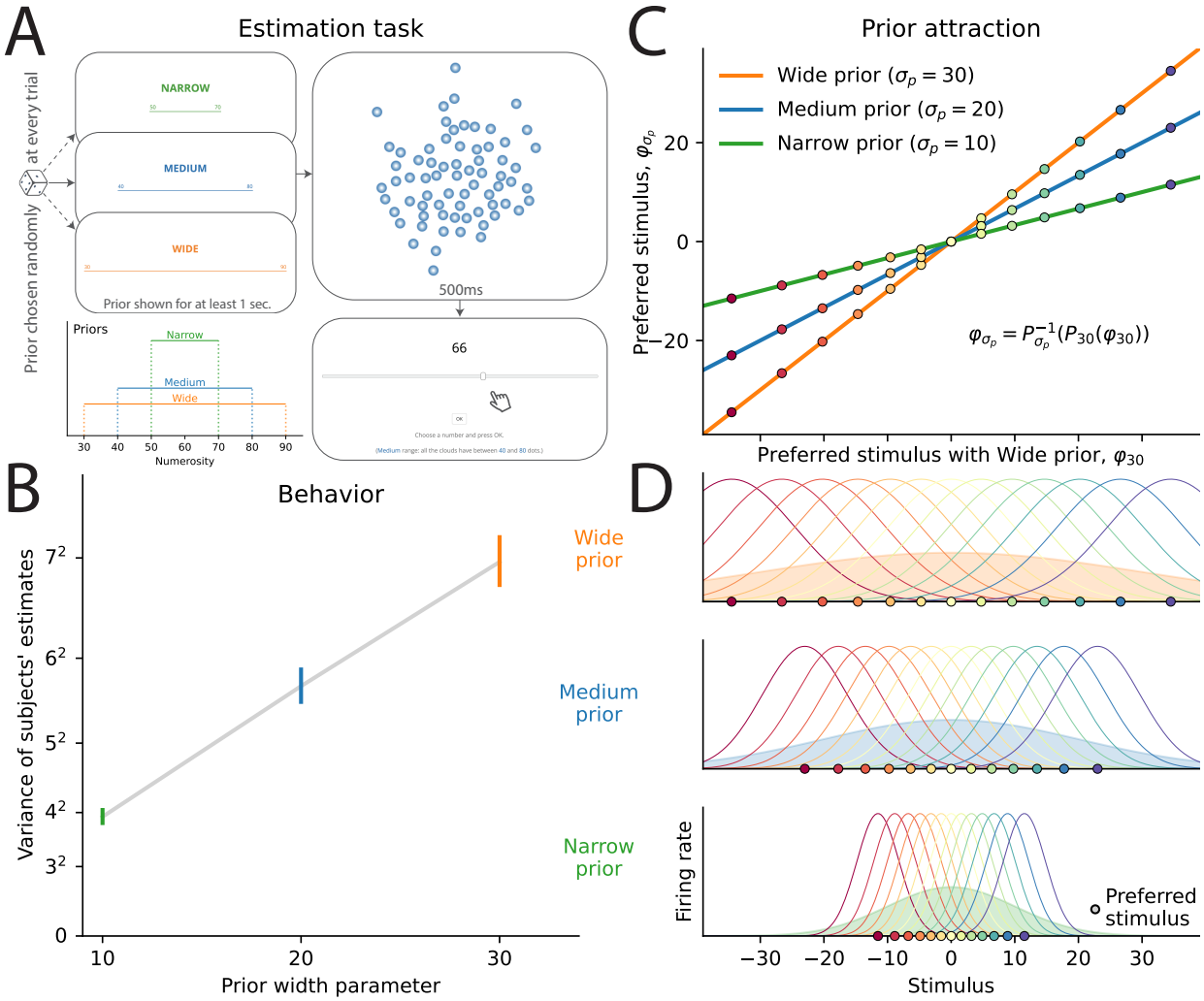

Fig. 1: Fast efficient coding: behavioral evidence and illustration of prior attraction. A. Numerosity estimation task. At each trial, a uniform prior (Narrow, Medium, or Wide) is randomly chosen, and explicitly shown for at least one second and until the subject presses the spacebar. A cloud of dots is then presented for 500ms, after which the subject is asked to estimate its numerosity. **B.** Variance of subjects' estimates as a function of the width of the prior. Values shown are Bayesian posterior mean estimates (see Methods). Error bars show the 5%-95% credible intervals. The estimate variance appears to be an affine function of the prior width. **C.** Neurons' locations φ_{σ_p} under efficient coding with each of the three priors, as a function of the locations with the Wide prior, φ_{30}. The Wide line is thus the identity line; the contraction of tuning curves with narrower priors results in the gentler slopes of the other lines. Specifically, efficient coding predicts that a neuron encodes for the same quantile, under all priors, i.e., $\forall \sigma_p, P_{\sigma_p}(\varphi_{\sigma_p}) = P_{30}(\varphi_{30})$, where P_{σ_p} and P_{30} denote the priors CDFs. **D.** Prior attraction: illustration of the tuning curves predicted by an efficient-coding model, for three Gaussian priors of mean 0 and standard deviations σ_p equal to 10 (Narrow; bottom), 20 (Medium; middle), and 30 (Wide; top). Each colored line shows the tuning curve for a different neuron. With the Narrow prior, the tuning curves are narrower and more densely distributed around the prior mean.

Dynamic tuning curves: Prior attraction and adapter repulsion

Traditional efficient-coding models posit a static prior and predict that the allocation of encoding resources should match this prior. Specifically, models of neural populations with unimodal tuning curves predict that the density of their locations should be proportional to the prior [3, 6–8] (or to the prior raised to some exponent, depending on the specifics of the model; here, for simplicity we consider the exponent 1). We are interested in the implications of 'fast efficient coding', i.e., the hypothesis that sensory networks quickly implement such optimal configuration, when the prior changes. To illustrate, we consider throughout this article the problem of encoding a scalar stimulus, s, drawn from a Gaussian prior with mean $\mu = 0$, and a variable standard deviation, σ_p. For the sake of illustration we consider three values, $\sigma_p = 10$, 20, or 30. (These widths match the parameters used in our estimation task, but the scalar stimulus considered here could represent any arbitrary one-dimensional variable, such as luminosity or loudness, and the specific scale chosen has no significant implications.) If the prior narrows and efficient coding can be implemented quickly, then tuning curves should move closer to the prior center (Fig. 1D). More precisely, a neuron's location should correspond to the same quantile across different priors (Fig. 1C). This 'prior attraction' of encoding resources would explain the changes in subjects' precision in the numerosity estimation task.

Prior attraction is a prediction regarding the behavior of tuning curves when the relative frequencies of stimuli change. A number of experimental studies have documented the changes in sensory tuning curves following manipulations of the stimuli frequencies. In these experiments, typically, an adapter stimulus is repeatedly presented (Fig. 2A), and the common finding is that tuning curves shift *away* from this adapter (instead of toward the adapter as would be predicted by prior attraction) [12–17]. This 'adapter repulsion' is not a single phenomenon but rather encompasses a rich set of related effects observed in different systems. Specifically, we note three consistent experimental observations that have been made in different studies, in the primary visual cortices of macaque monkeys [13] and of cats [12, 14]. First, the tuning curves of sensory neurons whose locations are close to the adapter shift away from it. Second, this effect decreases with distance, and neurons with further locations shift less strongly. Third, neurons close to the adapter see their tuning curves widen, while those further away see their tuning curves become narrower. These stylized facts are illustrated in Figure 2B (solid and dashed lines). In addition, Refs. [12, 13] find some evidence for a slight attractive shift in neurons whose locations are even further away (Fig. 2B, dotted lines). In spite of this apparent adapter attraction, we label these observations taken together as 'adapter repulsion', since repulsion appears to be the dominant effect. Below, we examine a gain-adaptive recurrent-network model, and show how it accounts for both prior attraction and adapter repulsion under a single efficient-coding objective.

Gain-adaptive recurrent neural network

We consider a network of N neurons encoding a stimulus $s \in \mathbb{R}$. Following Ref. [21], we posit that their firing rates $\boldsymbol{r} = (r_1, \ldots, r_N) \in \mathbb{R}_+^N$ are driven, first, by the "feedforward tuning curves" response functions $\boldsymbol{f}(s) = (f_1(s), \ldots, f_N(s)) \in \mathbb{R}_+^N$, each multiplied by a gain, $g_i \in \mathbb{R}_+$; and second, by the recurrent activity of the network. Specifically, we assume

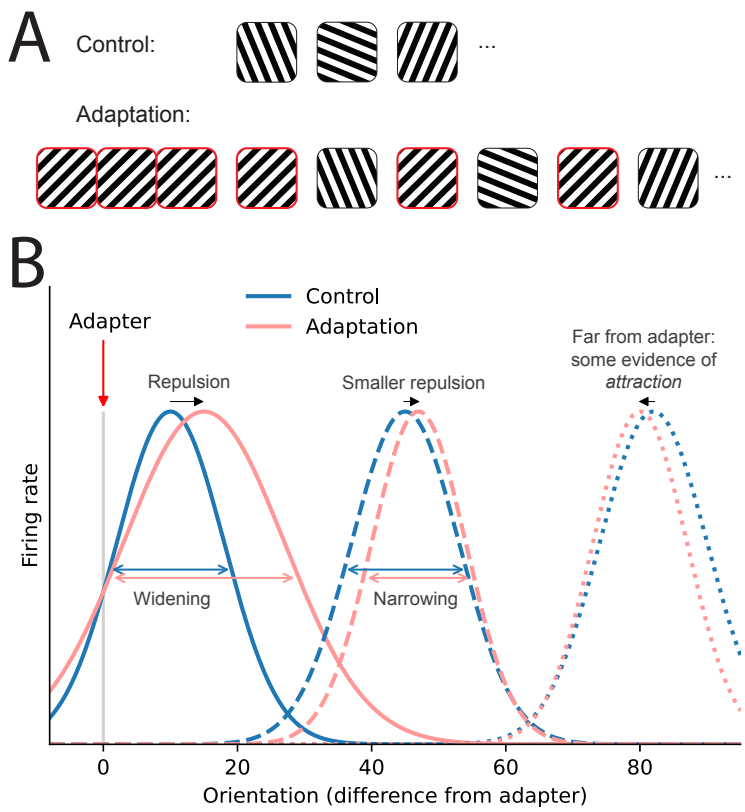

Fig. 2: Adapter repulsion phenomena. A. Typical adaptation experiments (panel inspired by Dragoi *et al.*, 2000 [12]). In the Control condition, a sequence of gratings with random orientations is shown. In the Adaptation condition, a single grating, the adapter, is presented for a long time or repeatedly. Then a sequence of gratings with random orientations, regularly interspersed with 'top-up' presentations of the adapter, is shown. **B.** Stereotypical results of adaptation experiments [12–14]. Tuning curves, i.e., firing rate as a function of the presented orientation, for three sensory neurons, in the Control condition (blue lines) and in the Adaptation condition (pink lines). For a neuron whose preferred orientation is close to the adapter (solid lines), the tuning curve widens and shifts away from the adapter. For a neuron whose preferred orientation is less close to the adapter (dashed lines), the tuning curve narrows and shifts away from the adapter, by a smaller extent. For a neuron whose preferred orientation is further from the adapter (dotted lines), some studies find a modest adapter attraction, i.e., the tuning curve slightly shifts towards the adapter [12, 13].

linear dynamics, as

$$\tau \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} = -\boldsymbol{r} + \boldsymbol{g} \circ \boldsymbol{f}(s) + W\boldsymbol{r}, \tag{1}$$

where $\boldsymbol{g} = (g_1, \ldots, g_N)$ is the vector of the gains, $\boldsymbol{g} \circ \boldsymbol{f}(s)$ is the Hadamard (element-wise) product of the vectors $\boldsymbol{g}$ and $\boldsymbol{f}(s)$, $W \in \mathbb{R}^{N \times N}$ represents the network's net synaptic connectivity matrix (we do not model differentiated excitatory and inhibitory neurons), and τ is a time constant. We are interested in the steady-state activity of the network. Setting

$\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} = 0$ yields the steady-state responses as functions of the stimulus s:

$$\boldsymbol{r}(s) = (I - W)^{-1} (\boldsymbol{g} \circ \boldsymbol{f}(s)). \tag{2}$$

We call "effective tuning curves" the functions $r_i(s)$; they result from the feedforward tuning curves $\boldsymbol{f}(s)$, the gains $\boldsymbol{g}$, and the network's recurrence W. The effective tuning curves describe the neurons' firing rates; further below, we examine how the resulting spikes can be decoded into an estimate of the stimulus.

We assume that the feedforward tuning curves $f_i(s)$ are Gaussian functions centered on the 'feedforward locations', s_i, which are ordered and equally spaced by a distance $\ell > 0$ such that $s_i = s_1 + (i-1)\ell$. The curves have amplitude 1 and width parameter σ_f, i.e.,

$$f_i(s) = \exp\left(-\frac{(s - s_i)^2}{2\sigma_f^2}\right). \tag{3}$$

The connectivity matrix, W, is assumed to be structured such that the neurons whose locations are closer to each other are more strongly connected (this is consistent with studies on the connectivity of sensory networks, for instance in the mouse primary visual cortex [22–24]). Specifically, we assume that the connectivity matrix is built with a Gaussian kernel with width parameter σ_{rec}, as

$$W_{ij} = \lambda_0 \frac{\ell}{\sigma_{rec}\sqrt{2\pi}} \exp\left(-\frac{(s_i - s_j)^2}{2\sigma_{rec}^2}\right), \tag{4}$$

where $0 < \lambda_0 < 1$ is a parameter representing the overall strength of the connectivity (it is also an approximation of the largest eigenvalue of W; see Methods).

Effective tuning curves

In the absence of recurrence (e.g., $\lambda_0 = 0$), the effective tuning curves are simply the feedforward tuning curves, multiplied by the gains (i.e., $\boldsymbol{r}(s) = \boldsymbol{g} \circ \boldsymbol{f}(s)$; Fig. 3A). With recurrence ($\lambda_0 > 0$), the activity of each neuron is influenced by all the other neurons, to varying degrees that depend on $\boldsymbol{g}$ and W. With uniform gains (i.e., if all the gains g_i are equal), these lateral inputs impact the overall activity of the neurons, but not their locations (Fig. 3B). If instead the gains are non-uniform, then each neuron's effective location will depend on the gain profile and on the recurrence matrix. Informally, a given neuron's location is pulled towards the feedforward locations of the neurons with high gains, and this effect is stronger with nearby locations (Fig. 3C). Formally, Equation 2 provides the effective tuning curves, but not in a way that makes readily explicit their effective locations and widths. Here we provide analytical approximations to these quantities, that make them more transparent. Our derivations are detailed in Methods.

We find that the effective location of neuron i, $\varphi(s_i)$, and its effective tuning-curve

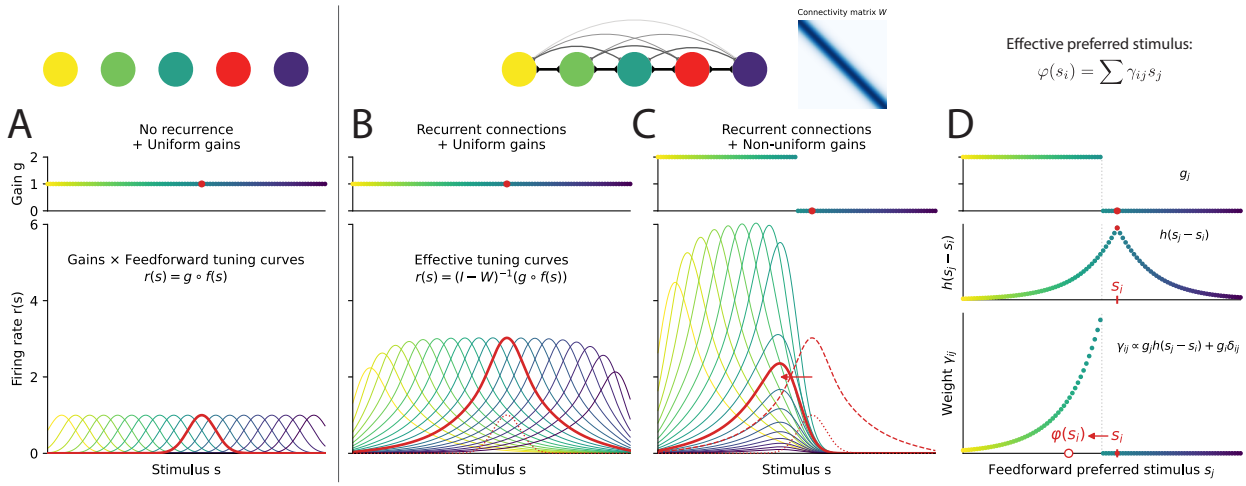

Fig. 3: Tuning curve shifts via gain modulation in a recurrent network. A. Without recurrence and with gains uniformly equal to 1 (top), the effective tuning curves are not different from the feedforward tuning curves. Thus all the neurons have identical tuning curves, except for their locations (bottom). **B.** With recurrence, and maintaining the gains uniformly equal to 1, the activity of each neuron propagates in the network, resulting in a general increase of the amplitudes of the tuning curves, but no changes in their locations. **C.** With non-uniform gains (here, half the neurons have gain 2 and half have gain 0), the tuning curves locations of the neurons with lower gain shift towards those of the neurons with higher gains. In red, tuning curve of a neuron (solid line), compared with its tuning curve with uniform gains (dashed line, solid line in **B**) and with no recurrence (dotted line, solid line in **A**). The inset shows the connectivity matrix (darker blue corresponds to stronger connections). **D.** The effective location of a neuron i is a weighted sum of the feedforward stimuli of all the neurons. The weights γ_{ij} (bottom panel) are determined by the gains g_j (top panel) and by a decreasing function $h(s_j - s_i)$ of the distance between the two neurons' locations (middle panel).

squared-width, $\sigma_r^2(s_i)$, can be expressed approximately with weighted averages, as

$$\varphi(s_i) \simeq \sum_{j=1}^{N} \gamma_{ij} s_j \tag{5}$$

$$\text{and} \quad \sigma_r^2(s_i) \simeq \sigma_f^2 + \sum_{j=1}^{N} \gamma_{ij} (s_j - \varphi(s_i))^2, \tag{6}$$

where

$$\gamma_{ij} = \frac{g_i \delta_{ij} + g_j h(s_i - s_j)}{G(s_i)}, \quad \text{and} \quad G(s_i) = g_i + \sum_{j=1}^{N} g_j h(s_i - s_j), \tag{7}$$

where δ_{ij} is the Kronecker delta, and $h(z)$ is a decreasing function of $|z|$ (see Fig. 3D, middle panel). In other words, the effective location of neuron i, $\varphi(s_i)$, is a weighted average of the feedforward locations of all the neurons, s_j, each weighted, through γ_{ij}, by its gain, g_j, and

by a decreasing function of its distance from s_i, $|s_i - s_j|$. Figure 3D illustrates, for a neuron i (in red) and given a gain profile (top panel), the unimodal function $s_j \mapsto h(s_j - s_i)$ (middle panel), the weights γ_{ij} of the neurons j (bottom panel), and the resulting effective location of neuron i, $\varphi(s_i)$, shown to be shifted away from its feedforward location, s_i.

Finally, we use these results to approximate the neuron's tuning curve as a Gaussian function. We obtain:

$$r_i(s) \simeq G(s_i) \frac{\sigma_f}{\sigma_r(s_i)} \exp\left(-\frac{(s - \varphi(s_i))^2}{2\sigma_r(s_i)^2}\right). \tag{8}$$

Equation 8 provides an analytical expression describing the probabilistic population code carried out by this network given a gain profile $\boldsymbol{g}$. The quality of our approximations of W, $\varphi(s_i)$, $\sigma_r^2(s_i)$, and $r_i(s)$ depends on the values of the various parameters, including σ_f (the width of the feedforward tuning curves), σ_{rec} (the width parameter of the connectivity matrix), λ_0 (the strength of the connectivity), and $\boldsymbol{g}$ (the gains). In Supplementary Information we present a comparison of these quantities with their approximations.

Bayesian decoding

What does the activity of this network tell us about the stimulus it is encoding? We now look at the question of decoding, and examine the Bayesian posterior implied by the spikes emitted by the network during one unit of time. We denote by $\boldsymbol{k} = (k_1, \ldots, k_N)$ the vector of the counts of spikes emitted by the N neurons. As an approximation to the joint statistics of this network, we consider the scenario of independent, Poisson-distributed spikes with rate parameters $r_i(s)$. This specifies the likelihood. As for the prior, we posit a Gaussian distribution, which we denote as $p(s) = N(s; \mu, \sigma_p^2)$. The resulting posterior is approximately Gaussian (see Methods):

$$p(s|\boldsymbol{k}) \simeq N\left(s; \hat{s}(\boldsymbol{k}), \tau(\boldsymbol{k})^{-1}\right), \tag{9}$$

where

$$\hat{s}(\boldsymbol{k}) = \frac{\frac{\mu}{\sigma_p^2} + \sum_{i=1}^N k_i \frac{\varphi(s_i)}{\sigma_r(s_i)^2}}{\frac{1}{\sigma_p^2} + \sum_{i=1}^N \frac{k_i}{\sigma_r(s_i)^2}} \tag{10}$$

and the precision (inverse variance) is

$$\tau(\boldsymbol{k}) = \frac{1}{\sigma_p^2} + \sum_{i=1}^N \frac{k_i}{\sigma_r(s_i)^2}. \tag{11}$$

The optimal estimate, if one seeks to minimize the expected squared error, is the posterior mean, $\hat{s}(\boldsymbol{k})$. Note that this estimator is a weighted average of the prior mean weighted by the prior precision, $1/\sigma_p^2$, and of each neuron's effective location, $\varphi(s_i)$, weighted by the number of spikes emitted by the neuron, k_i, multiplied by its effective 'precision', $1/\sigma_r(s_i)^2$. In other words, each spike is akin to a "vote" for the location of the corresponding neuron, weighted by a measure of the specificity of the neuron. This type of decoder has been known as population vector [25, 26], population average [27, 28], or center-of-gravity decoder [29]; see also Refs. [30, 31].

Furthermore, noting that the expected squared error of the Bayesian-mean estimator is the expected posterior variance, and using Jensen's inequality, we find (see Methods) a useful lower bound on the mean squared error (MSE), which we denote by L, as

$$\mathbb{E}_{s,\boldsymbol{k}}\left[\left(\hat{s}(\boldsymbol{k})-s\right)^2\right] \geq L \equiv \mathbb{E}_s\left[\left[\mathbb{E}_{\boldsymbol{k}|s}\left[\tau(\boldsymbol{k})|s\right]\right]^{-1}\right] = \int \frac{p(s)}{\frac{1}{\sigma_p^2}+\sum_{i=1}^{N}\frac{r_i(s)}{\sigma_r(s_i)^2}}\mathrm{d}s. \tag{12}$$

Below we use this lower bound L as our loss function. In our analyzes we have compared the squared error to the inverse of the expected precision, and we have found that the difference is relatively small, with our parameter regime (see Supplementary Information).

Efficient coding under spiking cost

We now explicitly address the question of optimal coding. We consider an observer equipped with the described recurrent sensory network, and asked to estimate, under a quadratic loss, a presented stimulus s that is randomly drawn from the prior $p(s)$. Prior to the presentation of s, we inform the observer of the prior $p(s)$, as in our experiment above, and we allow for the adjustment of the gains, $\boldsymbol{g}=(g_1,\ldots,g_N)$, of the sensory network. The synaptic connectivity matrix, W, and the parameters of the feedforward tuning curves, however, remain fixed. Crucially, we posit that each spike is costly. In particular, we posit a cost proportional to the expected number of emitted spikes [32, 33], as

$$C = \int p(s)\sum_{i=1}^{N} r_i(s)\mathrm{d}s. \tag{13}$$

The gains are chosen so as to minimize the lower bound on the MSE derived above, L, under this cost, i.e., to solve the optimization problem

$$\min_{\boldsymbol{g}} L + \alpha C, \tag{14}$$

where α is a coefficient specifying the relative weight of the cost.

Optimal gains under a prior larger than the tuning curves

Given a regime of parameters (i.e., the specifications of σ_f, σ_{rec}, λ_0, ℓ, N, σ_p, and α), the optimization problem just presented can be numerically solved, but it requires minimizing a function of the N parameters $g_1,\ldots,g_N$, which can be computationally intensive. Here we provide instead an approximate optimal solution, which relies on the assumption that the prior is wide in comparison to the widths of the tuning curves (see Methods). We take the network of neurons to the continuous limit ($N\to\infty$), and consider the functions $g(s)$, $\varphi(s)$, etc., instead of the discrete sets $(g_1,\ldots,g_N)$, $(\varphi(s_1),\ldots,\varphi(s_N))$, etc., allowing us to approximate discrete sums with continuous integrals. An insightful intermediate result is

$$\varphi(s) \simeq s + \tilde{\sigma}_h^2\frac{g'(s)}{g(s)}, \tag{15}$$

where $\tilde{\sigma}_h$ is proportional to σ_{rec} and depends on the other fixed parameters. Equation 15 shows that the shift of the effective location, $\varphi(s)$, as compared to the feedforward location, s, is driven by the variations of the gains, g. Intuitively, if the gains increase as a function of s (i.e., $g'(s) > 0$), then the location of the tuning curve is pushed towards greater stimuli (and thus $\varphi(s) > s$); and conversely if $g'(s) < 0$.

We solve the efficient-coding problem (Eq. 14) for the (continuous) gains $g(s)$ by deriving the corresponding Euler-Lagrange equation under a small-fluctuations assumption on the gains, yielding an ordinary differential equation for $g(s)$ wherever it is non-vanishing ($g(s) > 0$; see Methods). The approximate solution we obtain equals the positive part of an affine function of the log prior, as

$$g(s) = \left[\gamma_0 + \gamma_1 \log p(s)\right]_+, \tag{16}$$

where γ_0, γ_1 are constants and $[x]_+ = \max(0, x)$. Intuitively, the neurons whose locations are more probable under the prior should see their gains increased. This in turn determines the shift of the tuning curves, as the effective locations directly depend on the derivative of the gains (Eq. 15). Specifically, the gain profile in Eq. 16 implies that for the neurons whose feedforward locations are below the central value of the prior ($s < \mu$), the effective locations will be greater than the feedforward one ($\varphi(s) > s$), while feedforward locations above the central value ($s > \mu$) will result in lower effective locations ($\varphi(s) < s$). Put simply, the effective locations are attracted towards the central value of the prior, and thus the tuning curves contract towards the center.

Gain adaptation yields prior attraction

We now optimize the gains of the recurrent network to three Gaussian priors, labeled 'Narrow', 'Medium', and 'Wide', with mean $\mu = 0$ and widths $\sigma_p = 10$, 20, and 30, respectively (Fig. 4A). We look both at our approximation (Eq. 16) and at the numerically-optimized gains. For these analyzes we fix the various parameters as follows: $\sigma_f = 5$, $\sigma_{rec} = 10$, $\lambda_0 = 0.85$, $\ell = 0.5$, $N = 801$, and $\alpha = 5$. These values are comparable to those chosen in similar studies, they yield tuning-curve widths of the same order of magnitude as those measured in numerosity-sensitive neurons [34–36], and result in estimate variability comparable to that obtained in the behavioral task. Finally, in our numerical optimization we add a regularization term on the second derivative of the gains, $\|g''\|_2^2$, in order to avoid unstable solutions.

Our analytical approximation of the optimal gains is relatively close to the optimum we find numerically. The optimal gains have an inverted-U shape centered on μ (Fig. 4B); they are non-zero near the prior mean, and zero far from the prior mean: as a result, the effective locations of the network's neurons are attracted towards the prior mean, and this contraction is stronger for the narrower priors (Fig. 4C,D). As mentioned above, fast efficient coding predicts that from one prior to the other, a neuron should encode for the same quantile. More precisely, denoting by φ_{σ_p} and $\varphi_{\sigma'_p}$ the effective locations of a given neuron under the priors with widths σ_p and σ'_p, respectively, and by P_{σ_p} and $P_{\sigma'_p}$ the CDFs of these priors, then the prediction is $P_{\sigma_p}(\varphi_{\sigma_p}) = P_{\sigma'_p}(\varphi_{\sigma'_p})$. We thus look at the neurons' effective locations under the Narrow and Medium priors, in comparison with those predicted on the basis of the Wide case, i.e., we compare φ_{σ_p} and $P_{\sigma_p}^{-1}(P_{30}(\varphi_{30}))$ as a function of φ_{30},

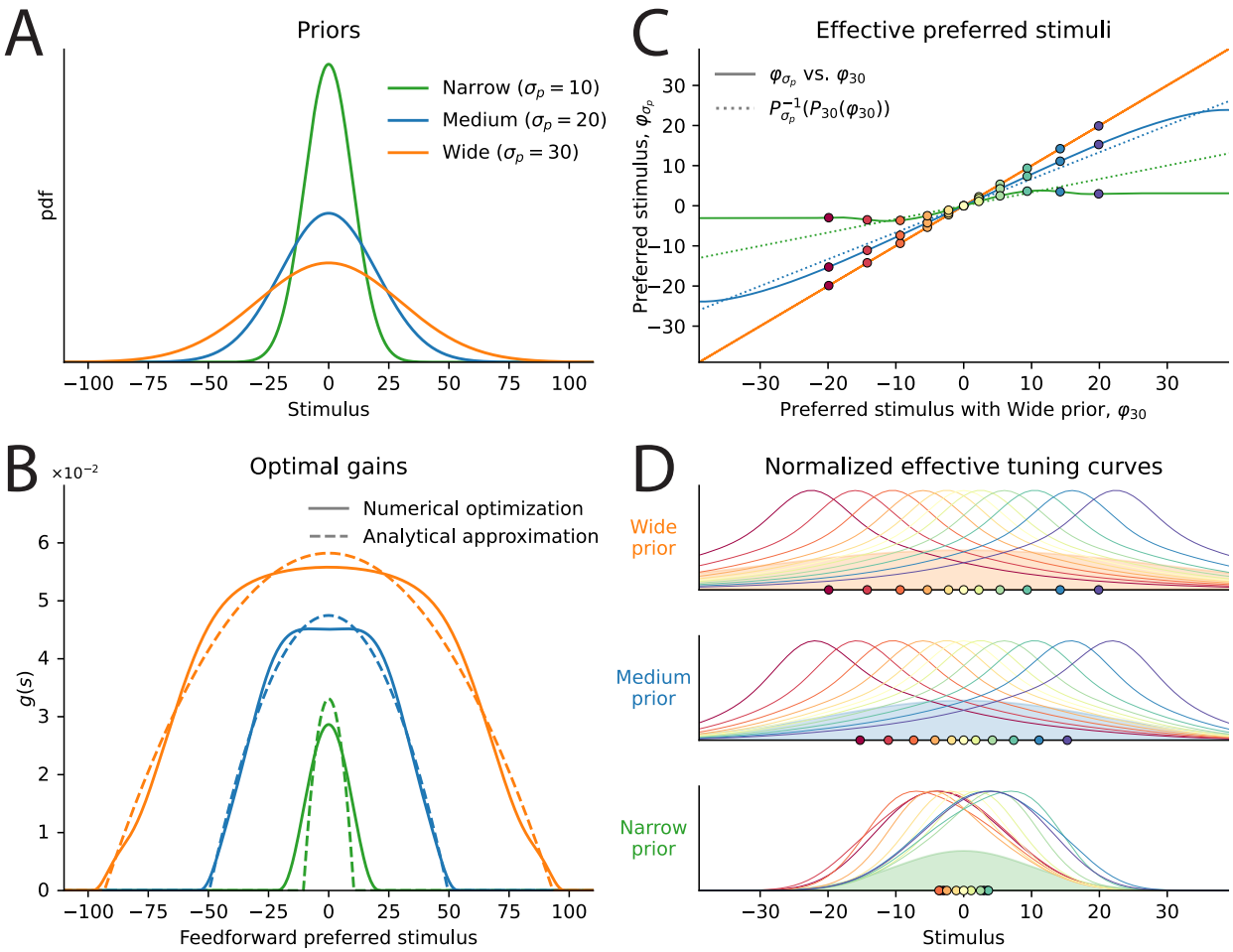

Fig. 4: Optimal gains yield efficient coding through prior attraction. A. Narrow, Medium, and Wide Gaussian priors. **B.** Optimal gains $g(s)$ as a function of the feedforward location, for the three priors, obtained through numerical optimization (solid lines) and approximate analytical derivation (Eq. 16; dashed lines). **C.** Neurons' effective locations, φ_{σ_p}, with the gains optimized to the different priors, as a function of their effective locations under the Wide prior, φ_{30} (compare to Fig. 1C). **D.** Normalized effective tuning curves, $r_i(s)$, exhibiting prior attraction resulting from the optimization of the gains (compare to Fig. 1D).

with $\sigma_p \in \{10, 20\}$. We find that the effective locations under the Medium prior are closer to the prior mean than under the Wide prior, and that they are well predicted by the relation just mentioned (Fig. 4C, blue lines). As for the locations under the Narrow prior, they are also well predicted by this relation, although only over an interval of roughly two standard deviations around the prior mean (Fig. 4C, green lines). As a result, the effective tuning curves exhibit 'prior attraction', i.e., they contract towards the center of the prior when the prior narrows (Fig. 4D).

We emphasize that the mentioned prediction, $P_{\sigma_p}(\varphi_{\sigma_p}) = P_{\sigma'_p}(\varphi_{\sigma'_p})$, is obtained in efficient-coding models in which the locations are free parameters, optimized to an objective but oth-

erwise unconstrained. By contrast, the efficient-coding problem we have considered is that of optimizing the gains in a strongly structured network of neurons. These gains together determine the effective locations, and we find them to behave similarly to the predictions of the "unconstrained" models, suggesting that recurrent networks can implement fast efficient coding through gain adaptation.

Anatomy of the network's Fisher information

The locations and shapes of the neurons' tuning curves determine the precision of the encoding carried out by the network. We now examine a statistical measure of this precision, the Fisher information $I(s)$. Generally, for a continuum of Poisson neurons with locations s, gain $g(s)$, Gaussian tuning curves with width $\sigma(s)$, and density $d(s)$, if these functions vary slowly on the scale of $\sigma(s)$, then the Fisher information is approximately (see Methods):

$$I(s) \simeq \sqrt{2\pi}\frac{g(s)d(s)}{\sigma(s)}. \tag{17}$$

We examine, with our recurrent network, how these three quantities, $g(s)$, $\sigma(s)$, and $d(s)$, vary with the prior width when the gains are adapted to the prior, and we look at how together they result in the Fisher information of the population. (Note that in our network the density is directly related to the locations, as $d(s) = (\varphi^{-1})'(s)\frac{1}{\ell}$.) For simplicity, we only consider these quantities at the central value, μ.

The gains increase as a function of the prior width (Fig. 5A). This results in larger spiking activity for larger widths (i.e., the expected number of spikes increases; Fig. 5D, dashed line). As the prior widens, the tuning curves spread out, but they also widen, i.e., their density decreases and their widths increase (Fig. 5B,C). Informally, if tuning curves are half as dense, then they should double in width to cover the same ground. Here we find indeed that the effective widths of tuning curves are an affine function of the inverse of the density (Fig. 5F).

The population's Fisher information results from the combination of these three factors. We find that it decreases as a function of the prior width: informally, the neural population is less precise when there are more stimuli to be represented (Fig. 5E). The Fisher information as a function of the prior width, σ_p, is well fit by a function proportional to $1/\sigma_p$. This is consistent with our behavioral data, in which the variance of estimates is an affine function of the width of the prior (and not of its square; Fig. 1B). This dependence of the Fisher information in $1/\sigma_p$ (rather than $1/\sigma_p^2$) reflects the endogenous choice in precision that results from the efficient-coding problem that features a cost (rather than a bound) on the spiking activity. The terms of the trade-off, L and C (Eqs. 12-13), are shown in Figure 5D. The loss, L, is the expected mean squared error (MSE). If the network dedicated the same amount of encoding resources in the context of the different priors, the loss would increase as the square of the width, σ_p^2. This would result in larges losses. But the MSE in fact increases at a lower rate (Fig. 5D, solid line), because the gains increase with larger priors (Figs. 4B and 5A), resulting in an intensification of the spiking activity (Fig. 5D, dashed line). This increased activity is costly, but it optimally mitigates the worsening of the loss [10].

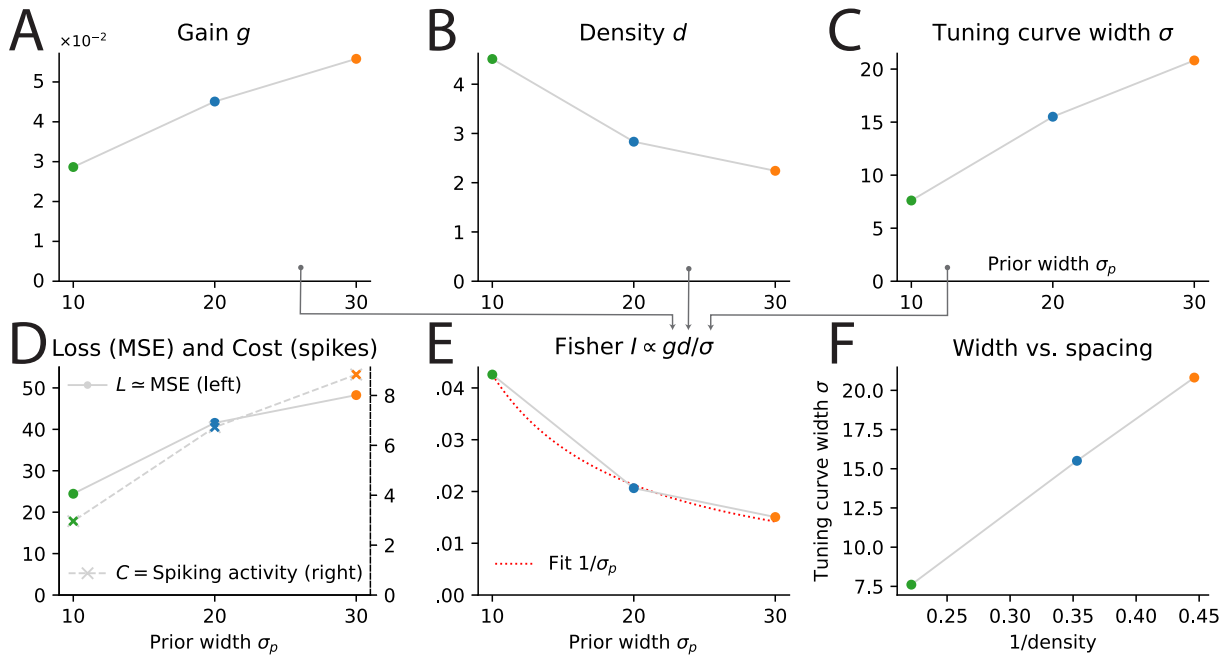

Fig. 5: Properties of the encoding network across priors of different widths. Gain (**A**), density (**B**), tuning curve width (**C**), loss function L and cost C (**D**), and Fisher information (**E**), as a function of the width parameter of the Gaussian priors, σ_p. **F.** Tuning curve width as a function of the inverse of the density (which indicates the spacing between tuning curves' locations), exhibiting a linear relationship: as tuning curves become more widely spaced, their width increases.

Peaked prior and adapter repulsion

The analytical results above rely on approximations that are less accurate when the prior width is close to, or lower than, the widths of the tuning curves. What are the optimal tuning curves under very narrow, that is, "peaked" priors? We consider this question in relation with the experimental finding that the tuning curves of some sensory neurons, after repeated exposure to an 'adapter' stimulus, shift away from this stimulus, as detailed above (Fig. 2). Hence we examine the tuning curves of the recurrent network in a scenario similar to these studies. It is not clear what the prior should be after repeated presentations of a stimulus; thus to simplify we consider a 'control' prior (pre-adaptation) that is uniform over a large interval, in comparison with a peaked, 'adaptation' prior that is a mixture between the same uniform distribution and a very narrow Gaussian prior of width $\sigma_p = 1$ and centered on $\mu = 0$ (Fig. 6A; see [4] for a similar approach). We numerically optimize the gains to these priors. Importantly, we keep all the other parameters of the network fixed and identical to the values that we have used above. Hence it is exactly the same efficient-coding optimization problem (Eq. 14) that is solved: only the prior changes.

With the peaked prior, the optimal gains exhibit two local maxima symmetrically located away from the peak, while at the peak ($\mu = 0$) the gains reach a local minimum, resulting in a bimodal M shape of the optimal gains as a function of the location (Fig. 6B). The shape of the gains in turn determines the shapes of the neurons' tuning curves. For a neuron whose

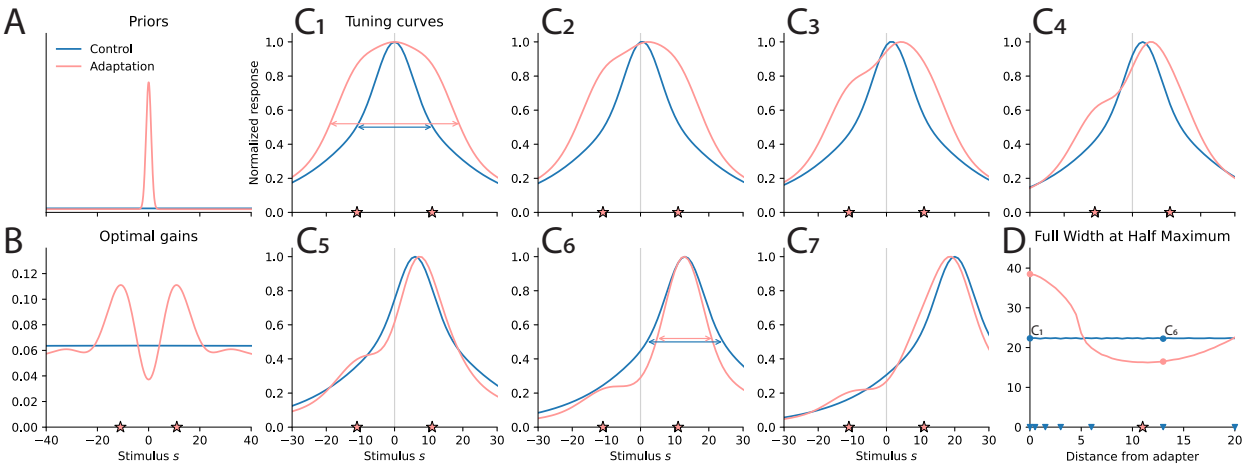

Fig. 6: Efficient coding via gain adaptation captures the rich adapter-repulsion phenomena. A. Uniform, Control-condition prior (blue line), and Adaptation-condition prior (pink line). **B.** Gains optimized for the two priors. With the Adaptation prior, the optimal gains exhibit two local maxima whose locations, marked with pink stars, are away from the peak of the prior. **C1-7.** Tuning curves for seven different neurons, in the two conditions. The pink stars indicate the locations of the gains' maxima. In the Adaptation condition, the tuning curves are 'attracted' towards these maxima, resulting in changes in their locations and widths. In C1 and C6, the widths of the tuning curves are shown, to emphasize the widening (C1) or the narrowing (C6). **D.** Full Width at Half Maximum of the neuronal tuning curves as a function of the distance from the adapter of their locations, in the two conditions. On the abscissa, the triangles show the locations of the neurons depicted in C1-7, while the pink star shows the location of the gains' maximum. The dots on the curves corresponds to the neurons whose widths have been emphasized in C1 and C6.

location is exactly at the adaptation peak, the tuning curve is "pulled" by the two gain maxima in both directions, resulting in a tuning curve almost twice as large as that obtained under the control prior (Fig. 6C1).

The neurons not located at the peak are unequally impacted by the two gain maxima, resulting in asymmetric tuning curves. The maximum of each curve is located, roughly, between the neuron's location under the control prior (which is essentially the feedforward location), and the location of the closest gains' maximum (Fig. 6C2-7). Thus for locations between the adaptation peak and the gains' maximum, there is a repulsive shift (away from the peak; Fig. 6C2-5). In other words, the adapter repulsion, in this model, is understood as an attraction towards the gains' local maximum. Feedforward locations further away from the peak and closer to the gains' maximum result in shorter shifts, and the shift vanishes for a location close to the gains' maximum (Fig. 6C6). Finally, for locations beyond the gains' maximum, we obtain a shift towards the peak, i.e., 'adapter attraction', but it is better understood here as an attraction towards the gain maximum (Fig. 6C7).

The widths of the tuning curves, similarly, are determined by the bimodal gain profile. The 'Full Width at Half Maximum' (FWHM) is maximal for neurons whose locations are at the adaptation peak (Fig. 6C1). Further, the impact of one maximum decreases while

that of the other increases, resulting in narrowing tuning curves (even narrower than under the control prior; Fig. 6C2-7). The minimum FWHM is reached near the location where the gains reach their local maximum (Fig. 6D). In sum, our model captures all the experimental findings that we have found to be consistently reported in the literature (and illustrated in Fig. 2B): namely, repulsive shifts close to the adapter, weaker repulsive shifts further from the adapter, and attractive shifts even further; also the widening of the tuning curves close to the adapter, and their narrowing away from it [12–14].

Discussion

We have studied a computational model of sensory representations that predicts a diverse range of adaptive changes in neuronal tuning curves following changes in stimulus frequencies, under a single principle — efficient coding — and using a single mechanism — neuronal gain adaptation. We presented behavioral evidence obtained in a pre-registered estimation task suggesting that human observers dynamically adjust their perceptual encoding in response to rapidly changing priors. Motivated by these findings, we introduced a gain-adaptive recurrent neural network in which neuronal gains minimize an objective combining decoding error and a metabolic cost proportional to neuronal spiking. Using analytical approximations and numerical simulations, we showed that this optimization yields prior attraction under wide stimulus priors and adapter repulsion under sharply peaked stimulus distributions, thus unifying apparently contradictory adaptive patterns. Our model suggests that fast efficient coding and the rich phenomena of adapter repulsion can be obtained through gain modulation alone, without requiring changes in synaptic connectivity.

Our behavioral experiment builds on previous work that had demonstrated subjects' adaptation to the prior when it is manipulated during the experiment but unchanging in blocks of successive trials [10], a finding substantiated by results obtained with similar experimental settings [11, 37–39]. Ni *et al.* (2023) [38] in addition re-analyze data from visual averaging tasks and argue that efficient adaptation occurs at each trial. Our experiment explicitly tests this hypothesis by examining behavior when the prior distribution changes from one trial to the next. Our results replicate those obtained previously with blocked designs [10], supporting the hypothesis that sensory representations efficiently adapt to changing priors on rapid timescales (on the order of one second).

Several neurophysiological studies support the idea that neural responses dynamically adapt to changing stimulus statistics in a manner consistent with fast efficient coding [40, 41]. In the blowfly visual system, H1 neurons (selectively responsive to horizontal motion) adjust their responses when the variance of the stimulus distribution changes [42, 43]. Specifically, the (monotonic) tuning curves are invariant when plotted as a function of the velocity normalized by the standard deviation of the prior. Similarly, in cat lateral geniculate nucleus [44], in songbird auditory neurons [45], and in rat barrel cortex [46], gains have been showed to vary as a function of the variance of the stimulus amplitude.

Related observations have been made in the domain of value-based decision-making: dopamine neurons adapt their coding to the range of reward magnitudes [47], and neurons in the orbitofrontal cortex dynamically adjust their response functions to represent economic value across different ranges [48]. Most of these examples involve monotonic tuning functions.

Two studies in which tuning curves seem to be unimodal functions of the stimulus report results consistent with fast efficient coding. These studies investigate the encoding of time in the rat hippocampus [49] and striatum [50]. Both works report a scalable representation of time, with patterns of neural activity that stretch or compress in response to changes in the duration being timed. This suggests a scalable and dynamic encoding — although not on a trial-to-trial basis, in these studies, but on the timescale of experimental sessions.

Fast efficient coding, in our model, is mediated by gain modulation. A substantial body of research has explored the role of gain modulation in shaping neuronal responses, and in particular how it may benefit coding efficiency. Gain modulation has been proposed as a multiplicative mechanism by which an 'attention field' is implemented, in order to improve performance in a given task [51]. Gain control also enables the decorrelation of multi-dimensional inputs, thereby carrying out a form of efficient coding that eliminates statistical dependencies in natural signals [52, 53]. The adaptations of gains to the variance of stimuli, mentioned just above [44–46], are examples of fast efficient coding mediated by gains, in single neurons. They are however specific to monotonic tuning curves, as in these cases gain adaptation essentially allows for mapping the range of stimuli onto a limited range of firing rates. The role of gains in the mechanism of fast efficient coding that we have investigated here is of a different nature. Indeed here the unequal gains across the recurrent population induce shifts in the neurons' locations, which, when optimized to a prior, allow for an efficient 'coverage' of the stimulus space. Hence this mechanism inherently operates at the population level (it does not apply to single neurons), and it is intrinsically tied to the recurrent nature of the considered population.

Some models of adapter repulsion also leverage the recurrence of sensory networks, but through synaptic plasticity. Westrick *et al.* (2016) proposes a divisive-normalization model of adaptation in which the normalization weights between each pair of neurons are updated through a Hebbian rule [19]. Such synaptic adaptation implies updating a number of parameters that scales as the square of the size of the neural population, which may yield instability and overfitting [21]. Perhaps more importantly, these synaptic updates typically occur on timescales longer than those of the adaptation phenomena we have considered here—on the order of a second, maybe less, for fast efficient coding, and less than 100ms for adapter repulsion.

A faster synaptic mechanism, often mentioned in studies reporting adapter repulsion [12–14], is short-term synaptic depression, whereby the amplitude of postsynaptic responses is scaled down whenever the presynaptic neuron fires an action potential [18]. This synaptic-level gain control adapts the sensitivity of postsynaptic neurons to the fluctuations in the firing rates of afferent neurons. Such short-term plasticity occurs on a short timescale, but its effects spontaneously relax within hundreds of milliseconds. This limits its suitability for efficient coding, where the optimal gain configuration should be maintained as long as need be. Similarly, Quiroga *et al.* (2016) [20] show that repulsive shifts can emerge from the transient dynamics of a recurrent network, an account that does not resort to synaptic plasticity, but that predicts only short-lived adaptive effects.

The alternative mechanism investigated here is that variations in the neuronal gains of a recurrent circuit induce shifts in the location of tuning curves. Dragoi *et al.* (2000,2002) allude to this possibility [12, 13], and Duong *et al.* (2023) show that, under an efficient-coding objective, it predicts tuning shifts away from an adapter [21]. We build on this

study and significantly extend it in several key ways. We provide analytical approximations to the main parameters of the effective tuning curves, as well as the optimal gains under wide priors, thus providing clear insights into the otherwise opaque behavior of the network; we introduce (and solve) an efficient-coding problem that includes only two terms, both readily justified: the loss function is derived from optimal, Bayesian inference, and the cost is proportional to the total expected number of action potentials, each of which consumes some amount of energy [32, 33]; we show in addition how our model captures the subtle features of adapter repulsion phenomena, including the widening and narrowing of tuning curves and the attraction far from the adapter; simultaneously within the same framework we predict prior attraction (and provide supporting behavioral data), thereby tying under a single paradigm the literature on efficient coding—traditionally understood as an optimal adaptation to a static prior—and that on adapter repulsion—the dynamic adaptation to a repeated stimulus.

The idea that adapter repulsion results from a form of efficient coding has been proposed elsewhere [4, 14, 15, 21], although a 2014 review on sensory adaptation concludes that efficient coding is "unlikely to be the sole goal of adaptation" [54]. Our results support the hypothesis that efficient coding is indeed the cause (if not the goal) of some adaptive phenomena, but our findings also suggest that this might be obscured by the non-trivial, and sometimes counter-intuitive, implications of efficient coding. Indeed, we have seen that with some priors the tuning curves were attracted towards the prior, while with other priors they are repelled from it. Further, we obtain adapter repulsion and adapter attraction simultaneously, depending on the distance between the location and the adapter (as experimentally observed). Our model explains this seemingly contradictory behavior: in both cases tuning curves are attracted to a local peak of the gains; and the gains exhibit such peaks because they are optimally adapted to the statistics of the stimuli.

We have focused on the steady-state of the network, as in many studies on adaptation, although there are rich transient dynamics during adaptation [13, 17, 20]. We conjecture that gain changes first have an impact "locally" (i.e., on the neurons whose gains have changed), before propagating throughout the network. Thus there should be some latency in the tuning shifts, and the tuning curves of the neurons whose locations are further from the adapter should adapt later than the closer ones. This roughly corresponds to the findings reported by Dragoi *et al.* (2002) [13]; but we leave to future studies the investigation of the recurrent network's transient dynamics.

Methods

Effective tuning curves

In order to obtain an analytical approximation of $\boldsymbol{r}(s)$, we derive first an approximation of $M = (I - W)^{-1}$ through the diagonalization of W. The eigenfunctions of a continuous Gaussian kernel can be expressed as a combination of cosine functions; in the discrete case, the 'discrete cosine transform' provides a way to obtain approximate eigenvalues and eigenvectors of the matrix W. From these we can derive their counterparts for M. This enables us to find an approximate expression for the elements M_{ij}, as $M_{ij} \simeq \delta_{ij} + h(s_i - s_j)$,

where δ_{ij} is the Kronecker delta, and $h(z)$ is a decreasing function of $|z|$ (see Fig. 3D, middle panel). Depending on the strength of connectivity, λ_0, the function h resembles either a Gaussian function (small λ_0) or a Laplace function (large λ_0; i.e., a function proportional to $\exp(-|z|/\nu)$, for some width $\nu > 0$). The latter has heavier tails, implying a greater range of each neuron's influence. The details of the mathematical derivations can be found in Supplementary Information.

Substituting M in Eq. 2 with its approximation yields an approximation of the effective tuning curve of a neuron i, $r_i(s)$, as a weighted sum of the feedforward tuning curves of all the neurons:

$$r_i(s) \simeq g_i f_i(s) + \sum_{j=1}^{N} g_j h(s_i - s_j) f_j(s). \tag{18}$$

Thus the contribution of neuron j to the effective tuning curve of neuron $i \neq j$ is equal to its feedforward tuning curve $f_j(s)$ multiplied by its gain g_j and by a decreasing function of the distance between the feedforward locations of the two neurons, $h(s_i - s_j)$.

We then define the effective location of neuron i, $\varphi(s_i)$, and its effective tuning curve squared-width, $\sigma_r^2(s_i)$, as

$$\varphi(s_i) \equiv \frac{\int s r_i(s) \mathrm{d}s}{\int r_i(s) \mathrm{d}s} \tag{19}$$

$$\text{and} \quad \sigma_r^2(s_i) \equiv \frac{\int (s - \varphi(s_i))^2 r_i(s) \mathrm{d}s}{\int r_i(s) \mathrm{d}s}. \tag{20}$$

With the approximation of $r_i(s)$ in Eq. 18, we obtain expressions for $\varphi(s_i)$ and $\sigma_r^2(s_i)$ (Eqs. 5,6).

Bayesian decoding

Posterior

The posterior on the stimulus s is given by Bayes' rule, as

$$p(s|\boldsymbol{k}) = \frac{p(s)p(\boldsymbol{k}|s)}{p(\boldsymbol{k})}, \tag{21}$$

where $p(s)$ is the prior belief about s and $p(\boldsymbol{k}|s)$ is the likelihood of the spike counts $\boldsymbol{k}$. As mentioned, we approximate the spike counts as independent and Poisson-distributed, with rate parameters $r_i(s)$. In this case the log-likelihood is

$$\log p(\boldsymbol{k}|s) = \sum_{i=1}^{N} \log p(k_i|s) \tag{22}$$

where

$$p(k_i|s) = \frac{e^{-r_i(s)} r_i(s)^{k_i}}{k_i!}, \tag{23}$$

thus the log-posterior is

$$\log p(s|\boldsymbol{k}) = \log p(s) - \sum_{i=1}^{N} r_i(s) + \sum_{i=1}^{N} k_i \log r_i(s) + c_{\boldsymbol{k}}, \tag{24}$$

where $c_{\boldsymbol{k}}$ is a constant that depends on $\boldsymbol{k}$ but not on s. We now posit a Gaussian prior, which we denote as $p(s) = N(s; \mu, \sigma_p^2)$. As for the sum of the rates, $\sum_{i=1}^{N} r_i(s)$, we assume that it varies slowly on the scale of the width of the posterior and thus that it can be neglected here—we note that in many similar models, this term is simply assumed constant [31, 55, 56] (we also further justify our assumption in Supplementary Information). With our Gaussian approximation of the effective rates (Eq. 8), we obtain the approximate Gaussian posterior presented in Eq. 9.

A lower bound on the Bayesian MSE

The unconditional, expected squared error of the Bayesian-mean estimator is the expected posterior variance, as

$$\mathbb{E}_{s,\boldsymbol{k}}\left[(\hat{s}(\boldsymbol{k}) - s)^2\right] = \mathbb{E}_{\boldsymbol{k}}\left[\mathbb{E}_{s|\boldsymbol{k}}\left[(\hat{s}(\boldsymbol{k}) - s)^2 \,|\boldsymbol{k}\right]\right] = \mathbb{E}_{\boldsymbol{k}}\left[\tau(\boldsymbol{k})^{-1}\right]. \tag{25}$$

In turn, the unconditional expected posterior variance is naturally expressed (by the law of total expectation) as

$$\mathbb{E}_{\boldsymbol{k}}\left[\tau(\boldsymbol{k})^{-1}\right] = \mathbb{E}_{s,\boldsymbol{k}}\left[\tau(\boldsymbol{k})^{-1}\right] = \mathbb{E}_{s}\left[\mathbb{E}_{\boldsymbol{k}|s}\left[\tau(\boldsymbol{k})^{-1}|s\right]\right], \tag{26}$$

where $\mathbb{E}_{\boldsymbol{k}|s}\left[\tau(\boldsymbol{k})^{-1}|s\right]$ is the expected value, conditional on a stimulus s, of the posterior variance. By Jensen's inequality, a lower bound on this quantity is the inverse of the expected posterior precision, conditional on s, i.e.,

$$\mathbb{E}_{\boldsymbol{k}|s}\left[\tau(\boldsymbol{k})^{-1}|s\right] \geq \left[\mathbb{E}_{\boldsymbol{k}|s}\left[\tau(\boldsymbol{k})|s\right]\right]^{-1}. \tag{27}$$

This conditional expected precision is readily derived from Eq. 11 as

$$\mathbb{E}_{\boldsymbol{k}|s}\left[\tau(\boldsymbol{k})|s\right] = \frac{1}{\sigma_p^2} + \sum_{i=1}^{N} \frac{r_i(s)}{\sigma_r(s_i)^2}. \tag{28}$$

Thus a lower bound on the mean squared error, unconditional on s, is the expected value, taken over s, of the inverse of the conditional expected precision, which we denote by L, i.e.,

$$\mathbb{E}_{s,\boldsymbol{k}}\left[(\hat{s}(\boldsymbol{k}) - s)^2\right] \geq L \equiv \int \frac{p(s)\mathrm{d}s}{\frac{1}{\sigma_p^2} + \sum_{i=1}^{N} \frac{r_i(s)}{\sigma_r(s_i)^2}}. \tag{29}$$

Continuous approximations

Because we consider the case of relatively wide priors, we assume that the various functions characterizing the network ($g(s)$, $\varphi(s)$, etc.) vary slowly on the scale of the neurons'

widths. In addition to Eq. 15, we obtain the following useful continuous approximations (see Supplementary Information). The effective squared-width is approximately

$$\sigma_r^2(s) \simeq \bar{\sigma}_r^2 - \left(\frac{g'(s)}{g(s)}\right)^2 \tilde{\sigma}_h^4 + \frac{1}{2}\frac{g''(s)}{g(s)}\left(\tilde{\mu}_4 - \tilde{\sigma}_h^4\right), \tag{30}$$

where $\tilde{\sigma}_h^2$ is a constant that represents the squared-width of the function h, and that is proportional to σ_{rec}^2, $\tilde{\mu}_4$ is a 'fourth-moment' quantity characterizing h (it is proportional to σ_{rec}^4); and where we denote by $\bar{\sigma}_r^2$ the typical width of the tuning curves:

$$\bar{\sigma}_r^2 = \sigma_f^2 + \tilde{\sigma}_h^2. \tag{31}$$

This is in particular the width of the tuning curves when the gains are uniform ($g' = 0$).

The total expected number of spikes, given a stimulus s, is approximately

$$\sum_{i=1}^{N} r_i(s) = cg(s), \tag{32}$$

where

$$c = \frac{\sigma_f\sqrt{2\pi}}{\ell(1-\lambda_0)}. \tag{33}$$

In other words, the expected number of spikes emitted by the whole network, conditional on a stimulus s, is essentially determined by the gain of the neuron whose location is s. Finally, the expected precision of the decoder (Eq. 28) is driven by the quantity

$$\sum_{i=1}^{N} \frac{r_i(s)}{\sigma_r(s_i)^2} \simeq c\frac{g(s)}{\sigma_r(s)^2}. \tag{34}$$

Solving the efficient-coding problem

Using the continuous approximations just derived, we have

$$L \simeq \int \frac{p(s)\mathrm{d}s}{\frac{1}{\sigma_p^2} + c\frac{g(s)}{\bar{\sigma}_r^2 - \left(\frac{g'(s)}{g(s)}\right)^2 \tilde{\sigma}_h^4 + \frac{1}{2}\frac{g''(s)}{g(s)}\left(\tilde{\mu}_4 - \tilde{\sigma}_h^4\right)}}, \tag{35}$$

and

$$C \simeq c\int p(s)g(s)\mathrm{d}s. \tag{36}$$

The Euler-Lagrange equation implied by Eq. 14 and L and C is not amenable to an analytic solution. Thus we adopt a crude approximation strategy, whereby wherever $g(s)$ is nonzero, we consider that in the denominator of the integrand of L the precision term that comes from the prior, $1/\sigma_p^2$, is small in comparison to the other term (which is the precision term that comes from the likelihood). Under this assumption the problem simplifies (see Supplementary Information) to minimizing the integral of the Lagrangian

$$\mathcal{L}(s, g, g') = p(s)\left(\frac{\bar{\sigma}_r^2}{cg(s)} + \alpha cg(s)\right) + \frac{1}{cg(s)}\left(\kappa_0 p(s)\left(\frac{g'(s)}{g(s)}\right)^2 - \kappa_1\frac{p'(s)g'(s)}{g(s)}\right), \tag{37}$$

where κ_0 and κ_1 are constants. The two terms in the first parenthesis are readily interpreted. First, $\frac{\bar{\sigma}_r^2}{cg(s)}$ is the mean squared-error when the gains are uniform ($g' = 0$): it scales as the squared-width of the tuning curves, and it is inversely related to the gains. Second, $\alpha c g(s)$ is the spiking cost, which increases proportionally to the gains. The second parenthesis captures the impact of the variations of the gains on the loss function. The first term increases with $(g')^2$ and is thus minimized by 'flat' gain profiles that spread out evenly the encoding resource. The second term, crucially, decreases with $p'g'$, and thus it favors gain profiles that allocate stronger gains towards the more frequent stimuli.

We find an approximate solution by positing small fluctuations of $g(s)$, i.e., we assume that $g(s) = g_0 + \tilde{g}(s)$, where $|\tilde{g}(s)| \ll g_0$. The expansion of $g(s)$ in $\mathcal{L}$ using this small-fluctuation assumption yields a simpler functional-minimization problem; in turn, the corresponding Euler-Lagrange equation ($\mathcal{L}_{\tilde{g}} - \frac{\mathrm{d}}{\mathrm{d}s}\mathcal{L}_{\tilde{g}'} = 0$) yields a differential equation for $\tilde{g}(s)$ whose solution is dominated by a linear function of $\log p(s)$, resulting in our solution for the gain profile (Eq. 16). Note that with the Gaussian prior that we have assumed, the optimal gains take a quadratic form (bounded below by 0), as

$$g(s) = g_0 \left[1 - \left(\frac{s - \mu}{\Delta \sigma_p} \right)^2 \right]_+ , \tag{38}$$

where g_0 is the maximal gain in the network, while Δ determines the neurons whose gains are non-vanishing, as those whose feedforward locations s belongs in the interval $[\mu - \sigma_p \Delta, \mu + \sigma_p \Delta]$. The details of these derivations are given in Supplementary Information.

Task and subjects

The experiment was a variant of the one presented in Ref. [10]. In the original design, each prior was tested in a separate block of trials, with the distribution remaining constant throughout each block. By contrast, here the prior changed randomly from trial to trial, enabling us to examine whether relatively quick changes in the stimulus distribution resulted in changes in subjects' estimates. The experiment was pre-registered on AsPredicted (https://aspredicted.org/48yv-zs3c.pdf). The task was divided in two sections. In the first section, the three ranges were presented "separately", i.e., in this section there were three blocks of successive trials, and each block featured a prior that did not change across trials. Subjects were explicitly informed of the prior, at the beginning of each block. In the second section, at each trial the prior was chosen randomly and explicitly told to the subject. In all trials the size of the slider stayed the same, and ranged from 25 to 95. Furthermore, each section comprised three kinds of phases. In the 'Learning' phases (15 trials), the correct answer (that is, the correct number of dots) was directly shown to the subject, together with the cloud of dots. The only action required in this phase was to press the spacebar to move on to the next trial. In the 'Feedback' phases (30 trials), subjects were asked to estimate the number of dots in each cloud, after which feedback was provided, i.e., the correct number was shown. In the 'No-Feedback' phases (10 rounds in each block of the first section, 150 rounds in the second section), subjects estimated the number of dots, and no feedback was provided.

The performance of subjects in both the Feedback phases and the No-Feedback phases impacted their reward. Data analysis was conducted only with the data from the No-Feedback phase of the second section. The reward comprised a \$3 (USD) base pay and a performance bonus, calculated as follows. Subjects accumulated points throughout the experiment. The amount of points obtained in a trial was equal to $100(1 - ((\hat{s} - s)/12)^2)$, where s was the correct number of dots and $\hat{s}$ was the subject's estimate. Every 1000 points obtained in the experiment was worth 40¢ in the performance bonus. Each trial in the Feedback and in the No-Feedback phases resulted in points. The points formula just presented allowed for negative numbers. In this case, it was deducted from the cumulated number of points. If at the end of the experiment a subject had a negative total number of points, then the performance bonus was set to zero, and the responses of this subject were not included in the analysis (as described in the pre-registration). This was the case for one subject.

A total of 60 subjects participated in the experiments (36 male, 24 female, average age: 43.1, standard deviation (sd): 11.8). The average performance bonus was \$7.09 (sd: 1.89). The average duration of the experiment was 52 minutes (sd: 19.3 minutes). Subjects were recruited on Amazon Mechanical Turk through CloudResearch [57]. The study protocol was approved by the Institutional Review Board (IRB) of Harvard University (protocol IRB15-2048). The task was coded with jsPsych [58].

Data analysis

The variance estimates presented in Figure 1B correspond to the posterior-mean estimates of the fixed-effect components of a statistical model that included subject-specific random effects. In other words, they are Bayesian estimates of the average variance across subjects. More precisely, the statistical model was specified by the following equations, where c is the condition (Narrow, Medium, or Wide), $\hat{x}_{sic}$ is the response of subject s in trial i of condition c, and N_+ is the Gaussian distribution truncated to the positive numbers:

$$\begin{aligned} \hat{x}_{sic}|x &\sim N\left(m_{sc}(x), \sigma_{sc}(x)^2\right), \\ m_{sc}(x) &\sim N\left(m_{0c}(x), \tau^2\right), \\ \text{and } \ln\left(\sigma_{sc}(x)\right) &\sim N\left(\ln\left(\sigma_{0c}(x)\right), \nu^2\right), \end{aligned} \tag{39}$$

with the priors

$$\begin{aligned} m_{0c}(x) &\sim N\left(x, 20^2\right), \\ \sigma_{0c}(x) &\sim N_+\left(7, 7^2\right), \\ \tau &\sim N_+(5, 10^2), \\ \text{and } \nu &\sim N_+(5, 10^2). \end{aligned} \tag{40}$$

This statistical model was estimated using Stan with the HMC-NUTS sampler [59] (10 chains of 1000 samples each, following 1000 warmup iterations.) The error bars in Figure 1B correspond to the 5th and 95th percentile of the posterior. All other data analyses were conducted using NumPy and Scipy, and figures were made using Matplotlib [60–62].

Data and code availability statements

The data for the behavioral experiment is publicly available at https://osf.io/vhdqe/. The code used for the model and the figures is publicly available at https://osf.io/wa7hm/.

Acknowledgements

We thank Lyndon Duong for helpful discussions. This research was supported by the Kempner Institute for the Study of Natural and Artificial Intelligence and a Polymath Award from Schmidt Sciences (S.J.G.).

References

[1] H. B. Barlow. "Possible Principles Underlying the Transformations of Sensory Messages". In: *Sensory Communication*. Ed. by Walter A. Rosenblith. Cambridge, MA: The MIT Press, Sept. 1961, pp. 217–234. DOI: 10.7551/mitpress/9780262518420.003.0013.

[2] Simon Laughlin. "A simple coding procedure enhances a neuron's information capacity". In: *Zeitschrift fur Naturforschung - Section C Journal of Biosciences* 36.9-10 (1981), pp. 910–912. ISSN: 18657125. DOI: 10.1515/znc-1981-9-1040.

[3] Nicolas Brunel and J P Nadal. "Mutual information, Fisher information, and population coding." In: *Neural computation* 10.7 (1998). arXiv: 1011.1669v3 ISBN: 0899-7667, pp. 1731–57. ISSN: 0899-7667. DOI: 10.1162/089976698300017115.

[4] Martin J. Wainwright. "Visual adaptation as optimal information transmission". In: *Vision Research* 39.23 (Nov. 1999), pp. 3960–3974. ISSN: 0042-6989. DOI: 10.1016/S0042-6989(99)00101-7.

[5] Mark D. McDonnell and Nigel G. Stocks. "Maximally Informative Stimuli and Tuning Curves for Sigmoidal Rate-Coding Neurons and Populations". In: *Physical Review Letters* 101.5 (Aug. 2008). arXiv: 0802.1570v3, p. 058103. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.101.058103.

[6] Deep Ganguli and Eero P Simoncelli. "Implicit encoding of prior probabilities in optimal neural populations". In: *Advances in Neural Information Processing Systems 23*. Ed. by J. D. Lafferty et al. Curran Associates, Inc., 2010, pp. 658–666.

[7] Deep Ganguli and Eero P. Simoncelli. "Efficient Sensory Encoding and Bayesian Inference with Heterogeneous Neural Populations". In: *Neural Computation* 26.10 (Oct. 2014), pp. 2103–2134. ISSN: 0899-7667. DOI: 10.1162/NECO_a_00638.

[8] Zhuo Wang, Alan A. Stocker, and Daniel D. Lee. "Efficient Neural Codes That Minimize Lp Reconstruction Error". In: *Neural Computation* 28.12 (Dec. 2016). arXiv: 1803.01446, pp. 2656–2686. ISSN: 0899-7667. DOI: 10.1162/NECO_a_00900.

[9] Arthur Prat-Carrabin and Michael Woodford. "Bias and variance of the Bayesian-mean decoder". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 23793–23805.

[10] Arthur Prat-Carrabin and Michael Woodford. *Endogenous Precision of the Number Sense*. Sept. 2024. DOI: 10.7554/eLife.101277.1.

[11] Jiang Mao, Constantin A. Rothkopf, and Alan A. Stocker. "Adaptation optimizes sensory encoding for future stimuli". In: *PLOS Computational Biology* 21.1 (Jan. 2025). Publisher: Public Library of Science, e1012746. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1012746.

[12] Valentin Dragoi, Jitendra Sharma, and Mriganka Sur. "Adaptation-Induced Plasticity of Orientation Tuning in Adult Visual Cortex". In: *Neuron* 28.1 (Oct. 2000). Publisher: Elsevier, pp. 287–298. ISSN: 0896-6273. DOI: 10.1016/S0896-6273(00)00103-3.

[13] Valentin Dragoi et al. "Dynamics of neuronal sensitivity in visual cortex and local feature discrimination". In: *Nature Neuroscience* 5.9 (Sept. 2002). Publisher: Nature Publishing Group, pp. 883–891. ISSN: 1546-1726. DOI: 10.1038/nn900.

[14] Gidon Felsen et al. "Dynamic Modification of Cortical Orientation Tuning Mediated by Recurrent Connections". In: *Neuron* 36.5 (Dec. 2002). Publisher: Elsevier, pp. 945–954. ISSN: 0896-6273. DOI: 10.1016/S0896-6273(02)01011-5.

[15] Gidon Felsen, Touryan Jon, and Yang Dan. "Contextual modulation of orientation tuning contributes to efficient processing of natural stimuli". In: *Network: Computation in Neural Systems* 16.2-3 (Jan. 2005). Publisher: Taylor & Francis, pp. 139–149. ISSN: 0954-898X. DOI: 10.1080/09548980500463347.

[16] Andrea Benucci, Aman B. Saleem, and Matteo Carandini. "Adaptation maintains population homeostasis in primary visual cortex". In: *Nature Neuroscience* 16.6 (June 2013). Publisher: Nature Publishing Group, pp. 724–729. ISSN: 1546-1726. DOI: 10.1038/nn.3382.

[17] Carlyn A. Patterson, Stephanie C. Wissig, and Adam Kohn. "Distinct Effects of Brief and Prolonged Adaptation on Orientation Tuning in Primary Visual Cortex". In: *Journal of Neuroscience* 33.2 (Jan. 2013). Publisher: Society for Neuroscience Section: Articles, pp. 532–543. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.3345-12.2013.

[18] L. F. Abbott et al. "Synaptic Depression and Cortical Gain Control". In: *Science* 275.5297 (Jan. 1997). Publisher: American Association for the Advancement of Science, pp. 221–224. DOI: 10.1126/science.275.5297.221.

[19] Zachary M. Westrick, David J. Heeger, and Michael S. Landy. "Pattern Adaptation and Normalization Reweighting". In: *Journal of Neuroscience* 36.38 (Sept. 2016). Publisher: Society for Neuroscience Section: Articles, pp. 9805–9816. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.1067-16.2016.

[20] Maria Del Mar Quiroga, Adam P. Morris, and Bart Krekelberg. "Adaptation without Plasticity". In: *Cell Reports* 17.1 (Sept. 2016), pp. 58–68. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2016.08.089.

[21] Lyndon R. Duong et al. "Adaptive coding efficiency in recurrent cortical circuits via gain control". In: (2023). arXiv: 2305.19869.

[22] Ho Ko et al. "Functional specificity of local synaptic connections in neocortical networks". In: *Nature* 473.7345 (May 2011). Publisher: Nature Publishing Group, pp. 87–91. ISSN: 1476-4687. DOI: 10.1038/nature09880.

[23] Wei-Chung Allen Lee et al. "Anatomy and function of an excitatory network in the visual cortex". In: *Nature* 532.7599 (Apr. 2016). Publisher: Nature Publishing Group, pp. 370–374. ISSN: 1476-4687. DOI: 10.1038/nature17192.

[24] L. Federico Rossi, Kenneth D. Harris, and Matteo Carandini. "Spatial connectivity matches direction selectivity in visual cortex". In: *Nature* 588.7839 (Dec. 2020). Publisher: Nature Publishing Group, pp. 648–652. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2894-4.

[25] Apostolos P. Georgopoulos, Andrew B. Schwartz, and Ronald E. Kettner. "Neuronal Population Coding of Movement Direction". In: *Science* 233.4771 (1986). Publisher: American Association for the Advancement of Science, pp. 1416–1419. ISSN: 0036-8075.

[26] H. S. Seung and H. Sompolinsky. "Simple models for reading neuronal population codes". In: *Proceedings of the National Academy of Sciences of the United States of America* 90.22 (1993). ISBN: 1074910753, pp. 10749–10753. ISSN: 00278424. DOI: 10.1073/pnas.90.22.10749.

[27] David L. Sparks, Richard Holland, and Barton L. Guthrie. "Size and distribution of movement fields in the monkey superior colliculus". In: *Brain Research* 113.1 (Aug. 1976), pp. 21–34. ISSN: 00068993. DOI: 10.1016/0006-8993(76)90003-2.

[28] Choongkil Lee, William H. Rohrer, and David L. Sparks. "Population coding of saccadic eye movements by neurons in the superior colliculus". In: *Nature* 332.6162 (Mar. 1988). Publisher: Nature Publishing Group, pp. 357–360. ISSN: 1476-4687. DOI: 10.1038/332357a0.

[29] H. P. Snippe. "Parameter extraction from population codes: a critical assessment". In: *Neural Computation* 8.3 (Apr. 1996), pp. 511–529. ISSN: 0899-7667. DOI: 10.1162/neco.1996.8.3.511.

[30] Wei Ji Ma et al. "Bayesian inference with probabilistic population codes". In: *Nature Neuroscience* 9.11 (2006), pp. 1432–1438.

[31] Mehrdad Jazayeri and J. Anthony Movshon. "Optimal representation of sensory information by neural populations". In: *Nature Neuroscience* 9.5 (May 2006), pp. 690–696. ISSN: 1097-6256. DOI: 10.1038/nn1691.

[32] David Attwell and Simon B. Laughlin. "An Energy Budget for Signaling in the Grey Matter of the Brain". In: *Journal of Cerebral Blood Flow & Metabolism* 21.10 (Oct. 2001). Publisher: SAGE Publications Ltd STM, pp. 1133–1145. ISSN: 0271-678X. DOI: 10.1097/00004647-200110000-00001.

[33] Biswa Sengupta et al. "Action potential energy efficiency varies among neuron types in vertebrates and invertebrates". In: *PLoS Computational Biology* 6.7 (2010), p. 35. ISSN: 15537358. DOI: 10.1371/journal.pcbi.1000840.

[34] Esther F. Kutter et al. "Single Neurons in the Human Brain Encode Numbers". In: *Neuron* 100.3 (2018). Publisher: Elsevier Inc., 753–761.e4. ISSN: 10974199. DOI: 10.1016/j.neuron.2018.08.036.

[35] Esther F. Kutter et al. "Distinct neuronal representation of small and large numbers in the human medial temporal lobe". In: *Nature Human Behaviour* 7.11 (Nov. 2023). Publisher: Nature Publishing Group, pp. 1998–2007. ISSN: 2397-3374. DOI: 10.1038/s41562-023-01709-3.

[36] Esther F. Kutter et al. "Single-neuron representation of nonsymbolic and symbolic number zero in the human medial temporal lobe". In: *Current Biology* 34.20 (Oct. 2024), 4794–4802.e3. ISSN: 0960-9822. DOI: 10.1016/j.cub.2024.08.041.

[37] Arthur Prat-Carrabin and Michael Woodford. "Efficient coding of numbers explains decision bias and noise". In: *Nature Human Behaviour* (May 2022). Publisher: Springer US, pp. 845–848. ISSN: 2397-3374. DOI: 10.1038/s41562-022-01352-4.

[38] Long Ni and Alan A. Stocker. "Efficient sensory encoding predicts robust averaging". In: *Cognition* 232 (Mar. 2023), p. 105334. ISSN: 0010-0277. DOI: 10.1016/j.cognition.2022.105334.

[39] Arthur Prat-Carrabin and Samuel J. Gershman. *Bayesian estimation yields anti-Weber variability*. Aug. 2024. DOI: 10.1101/2024.08.08.607196.

[40] Barry Wark, Brian Nils Lundstrom, and Adrienne Fairhall. "Sensory adaptation". In: *Current Opinion in Neurobiology* 17.4 (Aug. 2007), pp. 423–429. ISSN: 0959-4388. DOI: 10.1016/j.conb.2007.07.001.

[41] Alison I. Weber, Kamesh Krishnamurthy, and Adrienne L. Fairhall. "Coding Principles in Adaptation". In: *Annual Review of Vision Science* 5.Volume 5, 2019 (Sept. 2019). Publisher: Annual Reviews, pp. 427–449. ISSN: 2374-4642, 2374-4650. DOI: 10.1146/annurev-vision-091718-014818.

[42] Naama Brenner, William Bialek, and Rob de Ruyter van Steveninck. "Adaptive Rescaling Maximizes Information Transmission". In: *Neuron* 26.3 (June 2000), pp. 695–702. ISSN: 08966273. DOI: 10.1016/S0896-6273(00)81205-2.

[43] Adrienne L. Fairhall et al. "Efficiency and ambiguity in an adaptive neural code". In: *Nature* 412.6849 (Aug. 2001). Publisher: Nature Publishing Group, pp. 787–792. ISSN: 1476-4687. DOI: 10.1038/35090500.

[44] Vincent Bonin, Valerio Mante, and Matteo Carandini. "The Statistical Computation Underlying Contrast Gain Control". In: *Journal of Neuroscience* 26.23 (June 2006). Publisher: Society for Neuroscience Section: Articles, pp. 6346–6353. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0284-06.2006.

[45] Katherine I. Nagel and Allison J. Doupe. "Temporal processing and adaptation in the songbird auditory forebrain". In: *Neuron* 51.6 (Sept. 2006), pp. 845–859. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2006.08.030.

[46] Miguel Maravall et al. "Shifts in Coding Properties and Maintenance of Information Transmission during Adaptation in Barrel Cortex". In: *PLOS Biology* 5.2 (Jan. 2007). Publisher: Public Library of Science, e19. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.0050019.

[47] Philippe N. Tobler, Christopher D. Fiorillo, and Wolfram Schultz. "Adaptive coding of reward value by dopamine neurons". In: *Science (New York, N.Y.)* 307.5715 (Mar. 2005), pp. 1642–1645. ISSN: 1095-9203. DOI: 10.1126/science.1105370.

[48] Camillo Padoa-Schioppa. "Range-adapting representation of economic value in the orbitofrontal cortex". In: *Journal of Neuroscience* 29.44 (2009), pp. 14004–14014. ISSN: 02706474. DOI: 10.1523/JNEUROSCI.3751-09.2009.

[49] Akihiro Shimbo, Ei Ichi Izawa, and Shigeyoshi Fujisawa. "Scalable representation of time in the hippocampus". In: *Science Advances* 7.6 (2021). ISSN: 23752548. DOI: 10.1126/sciadv.abd7013.

[50] Gustavo B. M. Mello, Sofia Soares, and Joseph J. Paton. "A Scalable Population Code for Time in the Striatum". In: *Current Biology* 25.9 (May 2015). Publisher: Elsevier, pp. 1113–1122. ISSN: 0960-9822. DOI: 10.1016/j.cub.2015.02.036.

[51] John H. Reynolds and David J. Heeger. "The Normalization Model of Attention". In: *Neuron* 61.2 (Jan. 2009), pp. 168–185. ISSN: 08966273. DOI: 10.1016/j.neuron.2009.01.002.

[52] Odelia Schwartz and Eero P. Simoncelli. "Natural signal statistics and sensory gain control". In: *Nature Neuroscience* 4.8 (Aug. 2001), pp. 819–825. ISSN: 1097-6256. DOI: 10.1038/90526.

[53] Lyndon R. Duong et al. "Adaptive whitening in neural populations with gain-modulating interneurons". In: *Proceedings of the 40th International Conference on Machine Learning* (2023). DOI: 10.48550/arXiv.2301.11955.

[54] Samuel G. Solomon and Adam Kohn. "Moving Sensory Adaptation beyond Suppressive Effects in Single Neurons". In: *Current Biology* 24.20 (Oct. 2014). Publisher: Elsevier, R1012–R1022. ISSN: 0960-9822. DOI: 10.1016/j.cub.2014.09.001.

[55] Richard S. Zemel, Peter Dayan, and Alexandre Pouget. "Probabilistic interpretation of population codes". In: *Neural Computation* 10.2 (Feb. 1998), pp. 403–430. ISSN: 0899-7667. DOI: 10.1162/089976698300017818.

[56] J. Beck et al. "Probabilistic population codes and the exponential family of distributions". In: *Progress in Brain Research*. Ed. by Paul Cisek, Trevor Drew, and John F. Kalaska. Vol. 165. Computational Neuroscience: Theoretical Insights into Brain Function. Elsevier, Jan. 2007, pp. 509–519. DOI: 10.1016/S0079-6123(06)65032-2.

[57] David J Hauser et al. "Evaluating CloudResearch's Approved Group as a solution for problematic data quality on MTurk". In: *Behavior Research Methods* 55.8 (2023), pp. 3953–3964.

[58] Joshua R. de Leeuw, Rebecca A. Gilbert, and Björn Luchterhandt. "jsPsych: Enabling an Open-Source Collaborative Ecosystem of Behavioral Experiments". In: *Journal of Open Source Software* 8.85 (2023), p. 5351. DOI: 10.21105/joss.05351.

[59] Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual, 2.35*. 2024.

[60] Charles R. Harris et al. "Array programming with NumPy". In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2.

[61] Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[62] J. D. Hunter. "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.