
Learning efficient task-dependent representations with synaptic plasticity

Colin Bredenberg

Center for Neural Science
New York University
cjb617@nyu.edu

Eero P. Simoncelli*

Center for Neural Science,
Howard Hughes Medical Institute
New York University
eero.simoncelli@nyu.edu

Cristina Savin*

Center for Neural Science,
Center for Data Science
New York University
csavin@nyu.edu

Abstract

Neural populations do not perfectly encode the sensory world: their capacity is limited by the number of neurons, metabolic and other biophysical resources, and intrinsic noise. The brain is presumably shaped by these limitations, improving efficiency by discarding some aspects of incoming sensory streams, while preferentially preserving commonly occurring, behaviorally-relevant information. Here we construct a stochastic recurrent neural circuit model that can learn efficient, task-specific sensory codes using a novel form of reward-modulated Hebbian synaptic plasticity. We illustrate the flexibility of the model by training an initially unstructured neural network to solve two different tasks: stimulus estimation, and stimulus discrimination. The network achieves high performance in both tasks by appropriately allocating resources and using its recurrent circuitry to best compensate for different levels of noise. We also show how the interaction between stimulus priors and task structure dictates the emergent network representations.

1 Introduction

A variety of forces shape neural representations in the brain. On one side, sensory circuits need to faithfully represent their inputs, in support of the broad range of tasks an animal may need to perform. On the other side, the neural ‘wetware’ is intrinsically noisy [1] and computing resources are highly limited, in terms of number of neurons and metabolic energy. It remains a mystery how local synaptic learning rules can overcome these limitation to yield robust representations at the circuit level.

Past work has focused on individual aspects of this problem. Studies of efficient coding have successfully explained features of early sensory representations [2, 3] in terms of the interaction between stimulus statistics and resource limitations, and several models have proposed how such representations could emerge through local unsupervised learning [4, 5, 6]. However, the bulk of this theoretical work has ignored task constraints. This oversight might seem justified, considering that we generally think of sensory cortices as performing unsupervised learning, however a growing body of experimental evidence suggests that behavioral goals shape neural receptive fields in adult sensory cortices (A1 [7, 8, 9], S1 [10, 11], V1 [12, 13]), usually in the presence of neuromodulation [14, 15, 16]. This kind of plasticity has been modelled using tri-factor learning rules, which provide a mechanism for learning *task-specific* representations using only local information [17, 18, 19]. However, the interaction between the task, input statistics, and biological constraints remains largely unexplored (but see [20]).

Here we use a stochastic recurrent neural network model to derive a class of tri-factor Hebbian plasticity rules capable of solving a variety of tasks, subject to metabolic constraints. The framework leverages internal noise for estimating gradients of a task-specific objective; thus, noise provides an essential ingredient for learning, and is not simply an impediment for encoding. We systematically

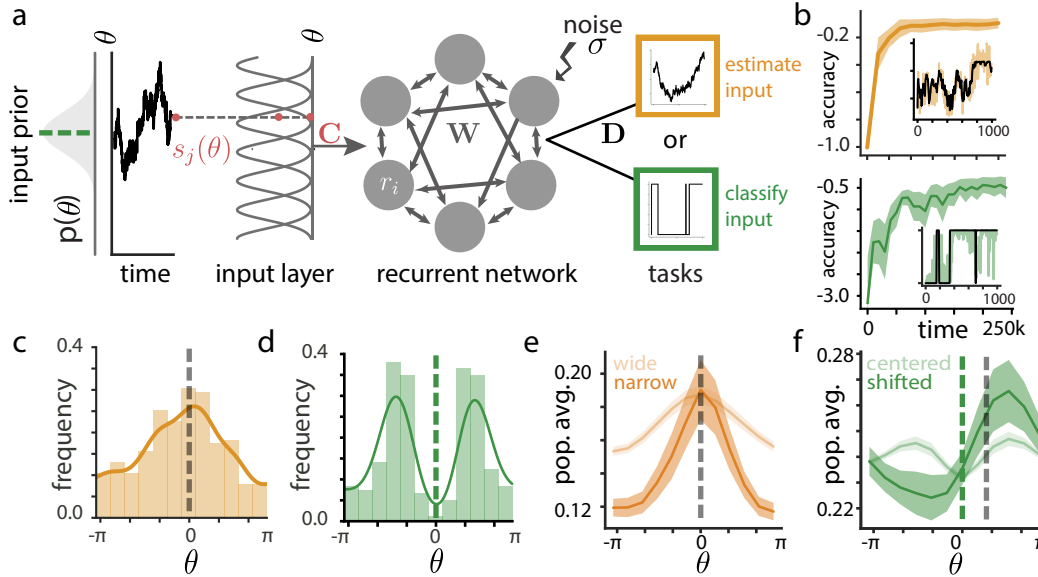


Figure 1: Recurrent neural network architecture and task learning. **a.** Model schematic. Stimuli drawn from a prior distribution (gray) are encoded in the responses of a static input layer, which feeds into the recurrent network; a linear decoder in turn produces the network output for either an estimation or a classification task. **b.** Network performance as a function of training time for the estimation (top) and classification (bottom) tasks. Shaded intervals indicate ± 1 s.e.m. across 45 test trials. Inset shows an example trained network output, with the correct output in black. **c.** Histogram of preferred stimuli and corresponding kernel density estimates (line). **d.** Same as **c**, for the classification task; decision boundary at $\theta = 0$ (dashed line). **e.** Population-averaged firing rates for two input priors. Shaded intervals are ± 1 s.e.m. averaging across all neurons in the network (dark gold) or all neurons across 20 simulations (light gold). **f.** Effects of shifting the stimulus prior relative to the classification boundary; dashed lines show common decision boundary (green) and shifted prior (gray). Shaded intervals as in **e**.

examine the interactions between input statistics, task constraints and resource limitations, and show that the emerging representations select task-relevant features, while preferentially encoding commonly occurring stimuli. The network also learns to reshape its intrinsic noise in a way that reflects prior statistics and task constraints.

2 Task-dependent local learning

Stochastic circuit model. We consider a simple sensory coding paradigm, in which a stimulus orientation θ is drawn from a von Mises distribution and encoded in the responses of an input population with fixed tuning, $s_j(\theta)$ (Fig. 1a). The input activity provides input to the recurrent network via synapses with weight specified by matrix \mathbf{C} . The recurrent network is nonlinear and stochastic, with neuron activities \mathbf{r} and recurrent synaptic strengths \mathbf{W} . A linear decoder with synaptic weights parameterized by matrix \mathbf{D} transforms the network activity into a task-specific output.

The stochastic dynamics governing the activity of recurrent neurons take the form:

$$dr_i = \left[-f^{-1}(r_i) + \sum_{j=1}^{N_r} w_{ij}r_j + \sum_{k=1}^{N_s} c_{ik}s_k + b_i \right] dt + \sigma dB_i, \quad (1)$$

where $f(\cdot)$ is the nonlinear response function (for simplicity, same for all neurons), N_s and N_r are the number of neurons in the input and recurrent populations, respectively, and b_i is a bias reflecting neural excitability. The parameter σ controls the standard deviation of the Brownian noise, B_i , added independently to each neuron. This intrinsic noise is one of the main constraints on the network.

The $f^{-1}(r_i)$ term is included to allow analytic tractability of the nonlinear dynamics. Furthermore, this formulation allows us to add Brownian noise to the current, such that fluctuations outside of those allowed by the F-I nonlinearity are sharply attenuated. In the small-noise limit, $\sigma \rightarrow 0$, the network has the same steady-state dynamics as a traditional nonlinear recurrent neural network. At equilibrium, $\bar{r}_i = f\left(\sum_j w_{ij}r_j + \sum_k c_k s_k + b_i\right)$, with nonlinearity f determining the steady-state F-I curve for each neuron; intrinsic noise induces fluctuations about this fixed point.

When the recurrent connectivity matrix \mathbf{W} is symmetric, the network dynamics are a stochastic, continuous analog of Hopfield's network formulation [21, 22], and of the Boltzmann machine [23]. Its corresponding energy function is:

$$E(\mathbf{r}, \mathbf{s}; \mathbf{W}) = -\frac{1}{2} \sum_{i,j} w_{ij} r_i r_j + \sum_i \int_0^{r_i} f^{-1}(x_i) dx_i - \sum_{ij} c_{ij} r_i s_j - \sum_i b_i r_i. \quad (2)$$

The network dynamics implement stochastic gradient descent on this energy, which corresponds to Langevin sampling [24] from the stimulus-dependent steady-state probability distribution:

$$p(\mathbf{r}|\mathbf{s}; \mathbf{W}) = \frac{\exp\left[-\frac{E(\mathbf{r}, \mathbf{s}; \mathbf{W})}{\sigma^2}\right]}{\int_{\mathbf{r}} \exp\left[-\frac{E(\mathbf{r}, \mathbf{s}; \mathbf{W})}{\sigma^2}\right]}. \quad (3)$$

The steady-state distribution is in the exponential family, and offers a variety of useful mathematical properties. Most importantly, the probabilistic description of the network, $p(\mathbf{r}|\mathbf{s}; \mathbf{W})$, can be used to calculate the gradient of an objective function with respect to the network weights via a procedure similar to node perturbation [25]. In practice, we use approximate solutions to Eq. (1) (or equivalently, Langevin sampling from Eq.3) using Euler-Maruyama integration.

Task-dependent objectives. We consider a task-specific objective function of the general form:

$$\mathcal{O}(\mathbf{W}, \mathbf{D}) = \iint \alpha(\mathbf{D}\mathbf{r}, \mathbf{s}) p(\mathbf{r}|\mathbf{s}; \mathbf{W}) p(\mathbf{s}) d\mathbf{r} d\mathbf{s} - \lambda_W \|\mathbf{W}\|_2^2, \quad (4)$$

where $\alpha(\cdot, \cdot)$ is a task-specific loss function, computed as a function of the linear readout $\mathbf{D}\mathbf{r}$, in a downstream circuit. The second term ensures that synaptic weights do not grow indefinitely [4]; it is a mathematically convenient way of introducing metabolic constraints, although regularizing the neural activity itself is also possible. For brevity, we have only included the constraint on \mathbf{W} in Eq. (4). In practice, we also include similar constraints on \mathbf{C} , \mathbf{D} , and \mathbf{b} , with corresponding Lagrange multipliers λ_C , λ_D and λ_b .

The specific choices of the loss, α , determines the nature of the task. Here, we chose two example objective functions – input encoding and binary classification. For reproducing the input, we use a mean squared error (MSE) objective:

$$\alpha_{\text{MSE}}(\mathbf{D}\mathbf{r}, \mathbf{s}) = -\|\mathbf{s} - \mathbf{D}\mathbf{r}\|_2^2, \quad (5)$$

with a negative sign reflecting the fact that we are maximizing, rather than minimizing, the objective. For classification, we use the log-likelihood (LL) objective:

$$\alpha_{\text{LL}}(\mathbf{D}\mathbf{r}, \mathbf{s}) = \phi(\mathbf{s}) \log(\psi(\mathbf{D}\mathbf{r})) + (1 - \phi(\mathbf{s})) \log(\psi(\mathbf{D}\mathbf{r})), \quad (6)$$

where $\psi(\cdot)$ is a sigmoid nonlinearity and $\phi(\mathbf{s})$ gives the mapping from stimulus \mathbf{s} to corresponding binary class.

Local task-dependent learning. We derive synaptic plasticity rules by maximizing $\mathcal{O}(\mathbf{r}, \mathbf{s}; \mathbf{W})$ using gradient descent, averaging across the stimulus distribution $p(\mathbf{s})$, and taking advantage of the closed-form expression for the steady-state stimulus-dependent response distribution $p(\mathbf{r}|\mathbf{s}; \mathbf{W})$.

Taking the derivative of the objective function (Eq. 4) with respect to w_{ij} yields:

$$\frac{\partial \mathcal{O}}{\partial w_{ij}} = \iint p(\mathbf{s}) \alpha(\mathbf{D}\mathbf{r}, \mathbf{s}) \frac{\partial p(\mathbf{r}|\mathbf{s})}{\partial w_{ij}} d\mathbf{r} d\mathbf{s} - 2\lambda_W w_{ij}. \quad (7)$$

As in [26], differentiating Eq. 3, we note that:

$$\frac{\partial}{\partial w_{ij}} p(\mathbf{r}|\mathbf{s}; \mathbf{W}) = \frac{1}{\sigma^2} \left[\frac{\partial}{\partial w_{ij}} E(\mathbf{r}, \mathbf{s}; \mathbf{W}) - \left\langle \frac{\partial}{\partial w_{ij}} E(\mathbf{r}, \mathbf{s}; \mathbf{W}) \right\rangle_{p(\mathbf{r}|\mathbf{s})} \right] p(\mathbf{r}|\mathbf{s}; \mathbf{W}), \quad (8)$$

where the brackets denote the conditional expectation with respect to $p(\mathbf{r}|\mathbf{s}; \mathbf{W})$. Rearranging, and substituting Eq. (8) into Eq. (7) yields:

$$\begin{aligned} \frac{\partial \mathcal{O}}{\partial w_{ij}} &= \frac{1}{\sigma^2} \iint \alpha(\mathbf{D}\mathbf{r}, \mathbf{s}) \left(\frac{\partial E(\mathbf{r}, \mathbf{s}; \mathbf{W})}{\partial w_{ij}} - \left\langle \frac{\partial E(\mathbf{r}, \mathbf{s}; \mathbf{W})}{\partial w_{ij}} \right\rangle_{p(\mathbf{r}|\mathbf{s})} \right) p(\mathbf{r}, \mathbf{s}; \mathbf{W}) \, d\mathbf{r} d\mathbf{s} - 2\lambda_W w_{ij} \\ &= \frac{1}{\sigma^2} \iint \alpha(\mathbf{D}\mathbf{r}, \mathbf{s}) \left(r_i r_j - \langle r_i r_j \rangle_{p(\mathbf{r}|\mathbf{s})} \right) p(\mathbf{r}, \mathbf{s}; \mathbf{W}) \, d\mathbf{r} d\mathbf{s} - 2\lambda_W w_{ij}. \end{aligned}$$

To update weights via gradient ascent, the learning rule takes the form:

$$\Delta w_{ij} \propto \frac{\partial \mathcal{O}}{\partial w_{ij}} = \mathbb{E} \left[\alpha(\mathbf{D}\mathbf{r}, \mathbf{s}) \left(r_i r_j - \langle r_i r_j \rangle_{p(\mathbf{r}|\mathbf{s})} \right) \right] - 2\lambda_W w_{ij},$$

where the expectation is approximated by sampling \mathbf{r} as part of the network dynamics:

$$\Delta w_{ij} \propto \alpha(\mathbf{D}\mathbf{r}, \mathbf{s}) \left(r_i r_j - \langle r_i r_j \rangle_{p(\mathbf{r}|\mathbf{s})} \right) - 2\lambda_W w_{ij}. \quad (9)$$

This expression for the weight update is similar to a standard reward-modulated Hebbian plasticity rule. It is driven by correlations between pre- and post-synaptic activity, with reward, α , having a multiplicative effect on weight changes. The subtractive term ensures that updates only occur in the presence of deviations from the average correlation level. It is symmetric in the indices $\{i, j\}$, and thus preserves the symmetry of the weight matrix \mathbf{W} . Finally, the weight regularization adds a contribution similar in form to Oja's rule [4]. In practice, we approximate the conditional expectation by a running average computed using a low-pass filter. We can derive similar updates for the input weights, and biases:

$$\begin{aligned} \Delta c_{ik} &\propto \alpha(\mathbf{D}\mathbf{r}, \mathbf{s}) \left(r_i s_k - \langle r_i s_k \rangle_{p(\mathbf{r}|\mathbf{s})} \right) - 2\lambda_C c_{ik} \\ \Delta b_i &\propto \alpha(\mathbf{D}\mathbf{r}, \mathbf{s}) \left(r_i - \langle r_i \rangle_{p(\mathbf{r}|\mathbf{s})} \right) - 2\lambda_b b_i, \end{aligned}$$

Hence, our framework allows us to optimize parameters of the network using local operations.

Learning the decoder. Because the readout weights enter $\alpha(\mathbf{D}\mathbf{r}, \mathbf{s})$, the optimization of D requires a slightly different treatment. Since $p(\mathbf{r}|\mathbf{s}; \mathbf{W})$ does not depend on \mathbf{D} , taking the derivative of Eq. (4) yields:

$$\frac{\partial \mathcal{O}}{\partial D_{ij}} = \int p(\mathbf{s}, \mathbf{r}; \mathbf{W}) \frac{\partial}{\partial D_{ij}} \alpha(\mathbf{D}\mathbf{r}, \mathbf{s}) \, d\mathbf{r} d\mathbf{s} - 2\lambda_D D_{ij}.$$

Using the same stochastic update scheme as in Eq. (9) yields:

$$\Delta D_{ij} \propto \frac{\partial}{\partial D_{ij}} \alpha(\mathbf{D}\mathbf{r}, \mathbf{s}) - 2\lambda_D D_{ij}.$$

This partial derivative will be different for each choice of α . For α_{MSE} , we get:

$$\Delta D_{ij}^{(\text{MSE})} \propto 2 \left(\sum_k D_{ik} r_k - s_i \right) r_j - 2\lambda_D D_{ij}, \quad (10)$$

and for α_{LL} (noting that \mathbf{D} here has only one row):

$$\Delta D_j^{(\text{LL})} \propto (\phi(\mathbf{s}) - \psi(\mathbf{D}\mathbf{r})) r_j - 2\lambda_D D_j. \quad (11)$$

3 Numerical results

To simulate the realistic scenario of slowly changing input stimuli constantly driving the circuit, we sample inputs from a von Mises prior distribution using Langevin dynamics with a significantly slower time constant than that of the recurrent network dynamics ($\tau_s = 375$). We set the noise level to an intermediate regime, so that its effects on circuit computation are non-negligible, but not pathological ($\sigma_B = 0.2$), and calibrate the hyperparameters that control the metabolic constraints (strength of regularization) to a level where they start to interfere with network performance.¹ As

¹For exact parameter values, see source code: [anonymizedlink.git](https://github.com/anonymizedlink).

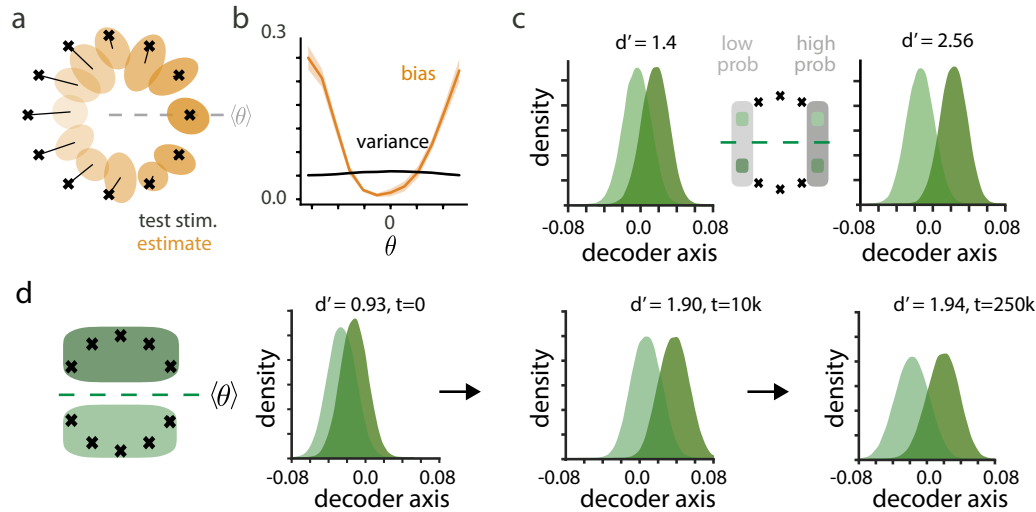


Figure 2: Task-specific stimulus encoding. **a.** Outputs for the estimation task: crosses mark the locations of test stimuli; ellipses show the 95% probability region for the associated network response distributions. **b.** Squared bias and variance for the estimator task as a function of stimulus angle. Shaded intervals are ± 1 SEM across 20 simulations. **c.** Decoder response distribution for stimuli near the boundary in the low-probability region of the space ($\theta = \pi$, left), and in the high-probability region ($\theta = 0$, right), as indicated on the center schematic. **d.** Discriminability of stimulus classes in the network output, shown before learning, after 10 trials, and after 250 trials.

learning progresses, our derived local plasticity rules quickly converge to a good solution, for both estimation and categorization (Fig. 1 b).

The emerging representations are noticeably different for the two tasks (compare Fig. 1c and d): for estimation, the distribution of preferred orientations is concentrated in the highly probable stimulus region, whereas the preferred distribution is bimodal in the case of classification. The average population activity for any stimulus provides additional quantification of the way in which learning allocates network resources (here, neural activity) to different stimuli (light colors in Fig. 1e and f). For the representation task, this metric confirms that neural resources are preferentially allocated to commonly occurring stimuli (Suppl. Fig. S1a, Fig. 1c; the most likely stimulus is 0). Hence learning has converged to an efficient representation of stimulus statistics. Moreover, the average population activity encodes the prior probability of input stimuli (Fig. 1e).

The picture looks very different in the case of categorization: improbable stimuli still have a small contribution to the emerging neural representation, but so do the most probable stimuli (which in our example are centered on the categorization boundary). This is reflected in the distribution of preferred stimuli: the neurons distribute their tuning functions on either side of the most probable stimulus so that the most sensitive part of their tuning function lies on the decision boundary (Suppl. Fig. S1b), and their peak responses are tuned for class identity (Fig. 1d). Overall, the emerging representations reflect both input statistics and task constraints.

So far all results used a particular choice of prior, but the same intuitions hold under prior variations. In the estimation task, tightening the prior leads to a corresponding tightening of the population tuning (Fig. 1e, Suppl. Fig. S1c). Firing rates decrease for peripheral stimuli; under synaptic weight regularization, the network will reduce firing rates for less probable stimuli, as they have little impact on the average error. This firing rate reduction is coupled with a concomitant increase in error for less probable stimuli (Suppl. Fig. S1d). For the classification task, we shifted the input distribution (Suppl. Fig. S1e) as a way to segregate the effects of the decision boundaries from those of the prior distribution Fig. 1f). This break in symmetry leads to asymmetric errors, with larger errors for the less probable class (Suppl. Fig. S1f), though the effect is more subtle than for tightening the input distribution. The corresponding network representation also becomes asymmetric, with higher firing rates concentrating on the side of the more probable stimulus. Thus, even under transformations

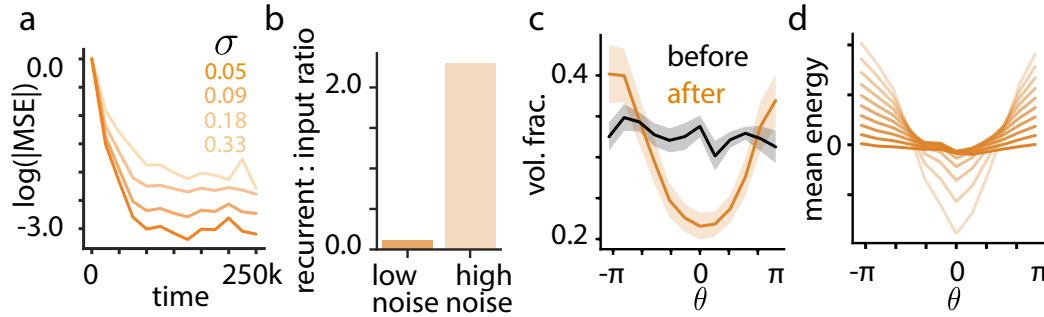


Figure 3: The effects of internal noise. **a.** Estimation performance during learning for different magnitudes of noise σ (for reference, the average external drive to the neuron is roughly twice the size of the largest noise level). **b.** Ratio of recurrent to input current for neurons in a low vs. high noise regime. **c.** Volume fraction of noise within/outside the readout manifold, as a function of stimulus angle ($\sigma = 0.33$), before and after learning. Shaded intervals are ± 1 s.e.m. across 20 simulations. **d.** Mean energy of $p(\mathbf{r}|\mathbf{s})$ as a function of input angle and different magnitudes of noise σ .

of our original prior, trained neurons show higher firing rates for both more probable, and more task-relevant features of the input, such that performance is consistently better for frequent stimuli.

How do these task-specific changes in representation manifest themselves in the decoded outputs? First, in the case of estimation, we can probe the distribution of outputs of the linear readout (Fig. 2a, gold ellipses) for a set of test stimuli (Fig. 2a, black crosses).² We found that responses are systematically biased for less probable stimuli, whereas the bias is negligible for frequently occurring stimuli. The effect on variance is much weaker (Fig. 2b). Second, for classification, we need to measure decoder output variability as a function of the prior stimulus probability. We use the fact that the classification boundary intersects the circle on which the input stimuli lie for the most probable and the least probable stimuli under the prior (Fig. 2c, inset). We compare the degree of output overlap for two test stimuli equally spaced on the two sides of the decision boundary at these two extremes. We find substantially higher discriminability for the high probability stimuli, relative to the low probability ones. As for estimation, this difference is due primarily to a better separability of the two means rather than a difference in variance. In summary, the network exhibits better performance for probable stimuli across tasks. Given limited resources, input statistics dictate not only the precision of the representation, but also task performance.

Next, we investigate the dynamics of learning and associated network representations, focusing on the classification task. We measure output discriminability at different stages of learning, with d' values estimated averaging over the stimulus distribution. At the outset, the output distributions for the two stimulus categories are indistinguishable (Fig. 2d). As learning progresses, the means of the two distributions segregate, while their variance remains approximately the same; this result prompts a more detailed examination of the degree to which the network is able to compensate for its own intrinsic noise.

Given that output variance changes little across stimuli and over learning, is noise strictly deleterious for the network, and how can the network learn to counteract its effects? We found that increasing intrinsic noise leads to both slower learning and worse asymptotic performance (Fig. 3a). Interestingly, the more intrinsic noise we add in the network, the more the network relies on recurrent connectivity for its representation. Increases in noise cause an increased engagement of recurrent connectivity after learning (Fig. 3b). The fact that recurrent connectivity is not strictly needed in absence of noise is intuitive, given that our chosen tasks involve a simple, instantaneous map of inputs to outputs (although note that in our framework some noise is needed to enable synaptic plasticity to occur). Though it is clear that recurrent connections alleviate some of the negative consequences of noise, their exact mechanism of action requires more investigation.

²Plasticity was disabled for test stimuli here, to isolate effects of internal noise without the confound of changes in network parameters.

It seems like the overall level of noise in the network does not change dramatically during learning, at least not as reflected in output fluctuations. Yet, performance is systematically better for probable stimuli, indicating the possibility of noise compensation. To investigate whether (and if so, how) recurrent connections shape internal noise, we asked what fraction of the internal noise lies in the decoding manifold given by \mathbf{D} and if it depends on the stimulus. Since the entropy of the conditional stimulus response distribution and its marginals are not analytically tractable, we resorted to numerical approximations: we used the network dynamics (with frozen weights) to sample from this conditional distribution, projected these responses a) in the readout manifold and b) in the 2D manifold of maximum variance, as defined by the first two PCA axes of the neural responses. Our final metric, which we refer to as the noise volume fraction, is computed as the ratio between the estimated noise in each of the two manifolds (using the determinant of the empirical covariance of the projected responses as a proxy for noise magnitude). This metric is 1 when the primary axes of internal noise lie in the decoding manifold; it is 0 when the two spaces are orthogonal, such that internal noise does not affect network outputs. After learning, the volume fraction is much smaller for probable stimuli (Fig. 3c), indicating that the network has learned to effectively ‘hide’ more of its noise for frequent inputs. This is not to say that output variance is lower for more probable stimuli: as we have already seen, if anything, output variance increases slightly for more probable stimuli. In general, the variability of the network output increases with the firing rates of its neurons, such that high network activity *necessarily* produces increased variability (one cannot amplify the signal without amplifying the noise to some degree). However, when normalizing for this increase, the network projects a smaller fraction of its total noise onto the decoder for more probable stimuli than it does for less probable stimuli.

An alternative way to think about the effects of intrinsic noise on the network activity is in terms of the energy function (Eq. 2), and corresponding steady-state stimulus response distribution (Eq. 3). Here, the noise variance acts as a temperature: the energy landscape flattens with increasing noise. Formally, one way to compensate for an increase in σ is to rescale the network’s stimulus-conditioned energy in proportion to the increase in σ^2 , thus preserving the stimulus response distribution $p(\mathbf{r}|\mathbf{s})$. The network does employ this kind of compensation (Fig.3d). As the intrinsic noise increase, the network boosts the gain of its energy function, with the mean energy increasing monotonically as a function of σ^2 (Suppl. Fig. S1g). Hence, the network learns to compensate for its intrinsic noise so as to maintain a good signal to noise ratio.

4 Discussion

Despite recent successes in the supervised training of recurrent networks [27, 28, 29, 30], it remains a mystery how biological circuits are capable of robust, flexible local learning in the face of constraints such as internal noise and limited metabolic resources. Here we have derived a class of synaptic learning rules that optimize a task-specific objective by combining local information about the firing rates of each pair of connected neurons with a single performance signal, implementable by global neuromodulation. Intrinsic noise in the recurrent dynamics is a key component of our solution, as it allows us to derive closed-form probabilistic expressions for the objective function gradients. Online learning naturally follows, since the network dynamics sample from a well-defined stimulus response distribution.³ We further show that the derived learning rules lead to emerging neural representations that allocate neural resources and reshape internal noise to reflect both input statistics and task constraints.

The use of stochasticity as a way to estimate gradients has a long tradition [26] and has been used to account for biological phenomenology, in particular the role of HVC variability in bird song learning [25]. Our model is conceptually closely related, and can be thought of as a mathematically tractable instantiation of the REINFORCE framework [26], where stochastic network responses are given by a continuous variant of the Boltzmann machine [23]. Existing literature has focused exclusively on the role of intrinsic noise on learning dynamics, ignoring its interactions with the circuit function and the emerging neural representations that support it. Here we show that there is a conflict between the positive role of stochasticity on learning and its deleterious effects on encoding. Over the course of learning, the network converges to a solution that trades off between the two, by appropriately reshaping the noise so as to increase the signal-to-noise ratio of the output.

³Note that although we are using sampling dynamics as well, our approach is different from sampling theory [31] in that here the network computations are not explicitly cast in a Bayesian framework.

Prior statistics play a key role in the emerging representations, with more neurons tuned to commonly occurring stimuli, and overall population activity weaker for infrequent stimuli. The inhomogeneous distribution of tuning functions aligns with optimal encoding of priors, as derived for abstract encoding models [32]. But unlike previous models, our network also encodes the prior as a population tuning function, a discrepancy that most likely reflects differences in the exact form of the regularizer enforcing metabolic constraints. Different choices of regularizer, in particular a sparseness-encouraging constraint on neural activity, would likely lead to representations more similar to traditional efficient coding models [2]. Nonetheless, our approach provides explicit local learning dynamics for these abstract models and, importantly, is successful in regimes where analytic methods are intractable, e.g. for multivariate stimuli, and for tasks that go beyond simple stimulus reproduction.

One limitation of our formulation is that reward does not depend on the history of network responses, circumventing the issue of temporal credit assignment. This differentiates our approach from traditional models of reward-modulated learning, which focus on solving temporal credit assignment, especially in spiking circuits [18, 25, 33, 34]. Nonetheless, our mathematical derivation extends previous work by making explicit the four-way interactions between intrinsic noise, metabolic constraints, input statistics, and task structure in the circuit. Preliminary results suggest that it is possible to extend the current framework to incorporate temporal dependencies in both stimuli and reward structure, better aligning it with traditional goals of reinforcement learning.

It has been argued that learning based on stochastic gradient estimates cannot match the learning capabilities of the brain [35], as they perform poorly in high dimensions [36]. This has led to renewed focus on alternatives that rely on the neural system having access to a closed-form expression for the gradient (at least approximately), in particular biologically-plausible approximations to backpropagation [37, 35, 30]. Still, these models can't be the whole story. While gradient information might be available for unsupervised [38] or intrinsic learning objectives [39], this is certainly not true for external rewards, when the loss function is specified by the environment itself. Animals rarely, if ever, have access to explicit reward functions. Moreover, neither unsupervised learning nor backpropagation can account for the critical role of neuromodulation in synaptic plasticity. Tri-factor plasticity rules such as ours are needed to explain the contribution of neuromodulation to perceptual learning [14, 15, 16]. The complementary nature of the two classes of learning rules suggests that they could *both* play an important role in perceptual learning. Bringing the two closer together is a promising direction for future research in both theory and biology.

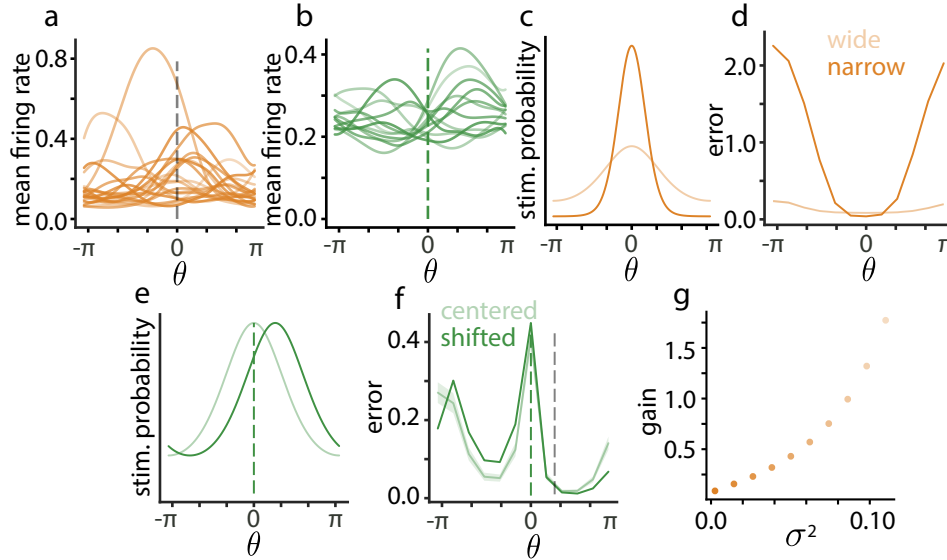


Figure S1: Manipulations of the input distribution. **a.** Sample tuning curves for the representation network. The gray dashed line indicates the prior mode. **b.** Sample tuning curves for the classification network. The green dashed line indicates the prior mode and the classification boundary. **c.** Two different input priors: the original distribution in light gold, and a sharper prior in dark gold. **d.** Error as a function of test input angle for a network trained on the original distribution (light gold), and the tightened distribution (dark gold). **e.** Schematic showing shifting the input distribution for the classification task. The original distribution is shown in light green, and the shifted distribution is shown in dark green; the green dashed line indicates the classification boundary. **f.** Error as a function of test input angle for a network trained on the original distribution (light green), and the shifted distribution (dark green). The gray dashed line indicates the shifted input distribution mode. Error bars for the original distribution indicate ± 1 s.e.m. across 20 simulations. **g.** The gain (L2 norm) of the mean energy as a function of σ^2 .

References

- [1] Aldo Faisal, Luc PJ Selen, and Daniel M Wolpert. Noise in the nervous system. *Nature reviews neuroscience*, 9(4):292, 2008.
- [2] Bruno A Olshausen et al. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.
- [3] Deep Ganguli and Eero P Simoncelli. Neural and perceptual signatures of efficient sensory coding. *arXiv preprint arXiv:1603.00058*, 2016.
- [4] Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of mathematical biology*, 15(3):267–273, 1982.
- [5] Christopher J Rozell, Don H Johnson, Richard G Baraniuk, and Bruno A Olshausen. Sparse coding via thresholding and local competition in neural circuits. *Neural computation*, 20(10):2526–2563, 2008.
- [6] Wieland Brendel, Ralph Bourdoukan, Pietro Vertech, Christian K Machens, and Sophie Denève. Learning to represent signals spike by spike. *arXiv preprint arXiv:1703.03777*, 2017.
- [7] Norman M Weinberger. Learning-induced changes of auditory receptive fields. *Current opinion in neurobiology*, 3(4):570–577, 1993.
- [8] Stephen V David, Jonathan B Fritz, and Shihab A Shamma. Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proceedings of the National Academy of Sciences*, 109(6):2144–2149, 2012.
- [9] Daniel B Polley, Elizabeth E Steinberg, and Michael M Merzenich. Perceptual learning directs auditory cortical map reorganization through top-down influences. *Journal of neuroscience*, 26(18):4970–4982, 2006.

- [10] Gregg H Recanzone, Michael M Merzenich, William M Jenkins, Kamil A Grajski, and HUBERT R Dinse. Topographic reorganization of the hand representation in cortical area 3b owl monkeys trained in a frequency-discrimination task. *Journal of Neurophysiology*, 67(5):1031–1056, 1992.
- [11] Gregg H Recanzone, Michael M Merzenich, and William M Jenkins. Frequency discrimination training engaging a restricted skin surface results in an emergence of a cutaneous response zone in cortical area 3a. *Journal of Neurophysiology*, 67(5):1057–1070, 1992.
- [12] Wu Li, Valentin Pièch, and Charles D Gilbert. Perceptual learning and top-down influences in primary visual cortex. *Nature neuroscience*, 7(6):651–657, 2004.
- [13] Aniek Schoups, Rufin Vogels, Ning Qian, and Guy Orban. Practising orientation identification improves orientation coding in v1 neurons. *Nature*, 412(6846):549–553, 2001.
- [14] Jonathan S Bakin and Norman M Weinberger. Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. *Proceedings of the National Academy of Sciences*, 93(20):11219–11224, 1996.
- [15] Michael P Kilgard and Michael M Merzenich. Cortical map reorganization enabled by nucleus basalis activity. *Science*, 279(5357):1714–1718, 1998.
- [16] Robert C Froemke, Michael M Merzenich, and Christoph E Schreiner. A synaptic memory trace for cortical receptive field plasticity. *Nature*, 450(7168):425–429, 2007.
- [17] H Sebastian Seung. Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. *Neuron*, 40(6):1063–1073, 2003.
- [18] Nicolas Frémaux and Wulfram Gerstner. Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules. *Frontiers in neural circuits*, 9:85, 2016.
- [19] Wulfram Gerstner, Marco Lehmann, Vasiliki Liakoni, Dane Corneil, and Johanni Brea. Eligibility traces and plasticity on behavioral time scales: experimental support of neohebbian three-factor learning rules. *Frontiers in neural circuits*, 12, 2018.
- [20] Cristina Savin and Jochen Triesch. Emergence of task-dependent representations in working memory circuits. *Frontiers in Computational Neuroscience*, 8:57, 2014.
- [21] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.
- [22] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. In *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, pages 411–415. World Scientific, 1987.
- [23] David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. A learning algorithm for Boltzmann machines. In *Readings in Computer Vision*, pages 522–533. Elsevier, 1987.
- [24] Gareth O Roberts, Richard L Tweedie, et al. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.
- [25] Ila R Fiete, Michale S Fee, and H Sebastian Seung. Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. *Journal of neurophysiology*, 98(4):2038–2057, 2007.
- [26] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. In *Reinforcement Learning*, pages 5–32. Springer, 1992.
- [27] Barak A Pearlmutter. Learning state space trajectories in recurrent neural networks. *Neural Computation*, 1(2):263–269, 1989.
- [28] David Sussillo and Larry F Abbott. Generating coherent patterns of activity from chaotic neural networks. *Neuron*, 63(4):544–557, 2009.
- [29] James Martens and Ilya Sutskever. Learning recurrent neural networks with Hessian-free optimization. In *Proc of the 28th International Conference on Machine Learning (ICML-11)*, pages 1033–1040, 2011.
- [30] Owen Marschall, Kyunghyun Cho, and Cristina Savin. A unified framework of online learning algorithms for training recurrent neural networks. *arXiv preprint arXiv:1907.02649*, 2019.

- [31] József Fiser, Pietro Berkes, Gergő Orban, and Máté Lengyel. Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, 14(3):119–130, March 2010.
- [32] Deep Ganguli and Eero P Simoncelli. Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. *Neural computation*, 2014.
- [33] Thomas Miconi. Biologically plausible learning in recurrent neural networks reproduces neural dynamics observed during cognitive tasks. *Elife*, 6:e20899, 2017.
- [34] Gregor M Hoerzer, Robert Legenstein, and Wolfgang Maass. Emergence of complex computational structures from chaotic neural networks through reward-modulated hebbian learning. *Cerebral cortex*, 24(3):677–690, 2014.
- [35] Alexandre Payeur, Jordan Guerguiev, Friedemann Zenke, Blake Richards, and Richard Naud. Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits. *bioRxiv*, 2020.
- [36] Justin Werfel, Xiaohui Xie, and H Sebastian Seung. Learning curves for stochastic gradient descent in linear feedforward networks. In *Advances in neural information processing systems*, pages 1197–1204, 2004.
- [37] Timothy P Lillicrap, Daniel Cownden, Douglas B Tweed, and Colin J Akerman. Random synaptic feedback weights support error backpropagation for deep learning. *Nature communications*, 7(1):1–10, 2016.
- [38] Cengiz Pehlevan, Tao Hu, and Dmitri B Chklovskii. A hebbian/anti-hebbian neural network for linear subspace learning: A derivation from multidimensional scaling of streaming data. *Neural computation*, 27(7):1461–1495, 2015.
- [39] Ross Goroshin, Joan Bruna, Jonathan Tompson, David Eigen, and Yann LeCun. Unsupervised learning of spatiotemporally coherent metrics. In *Proceedings of the IEEE international conference on computer vision*, pages 4086–4093, 2015.