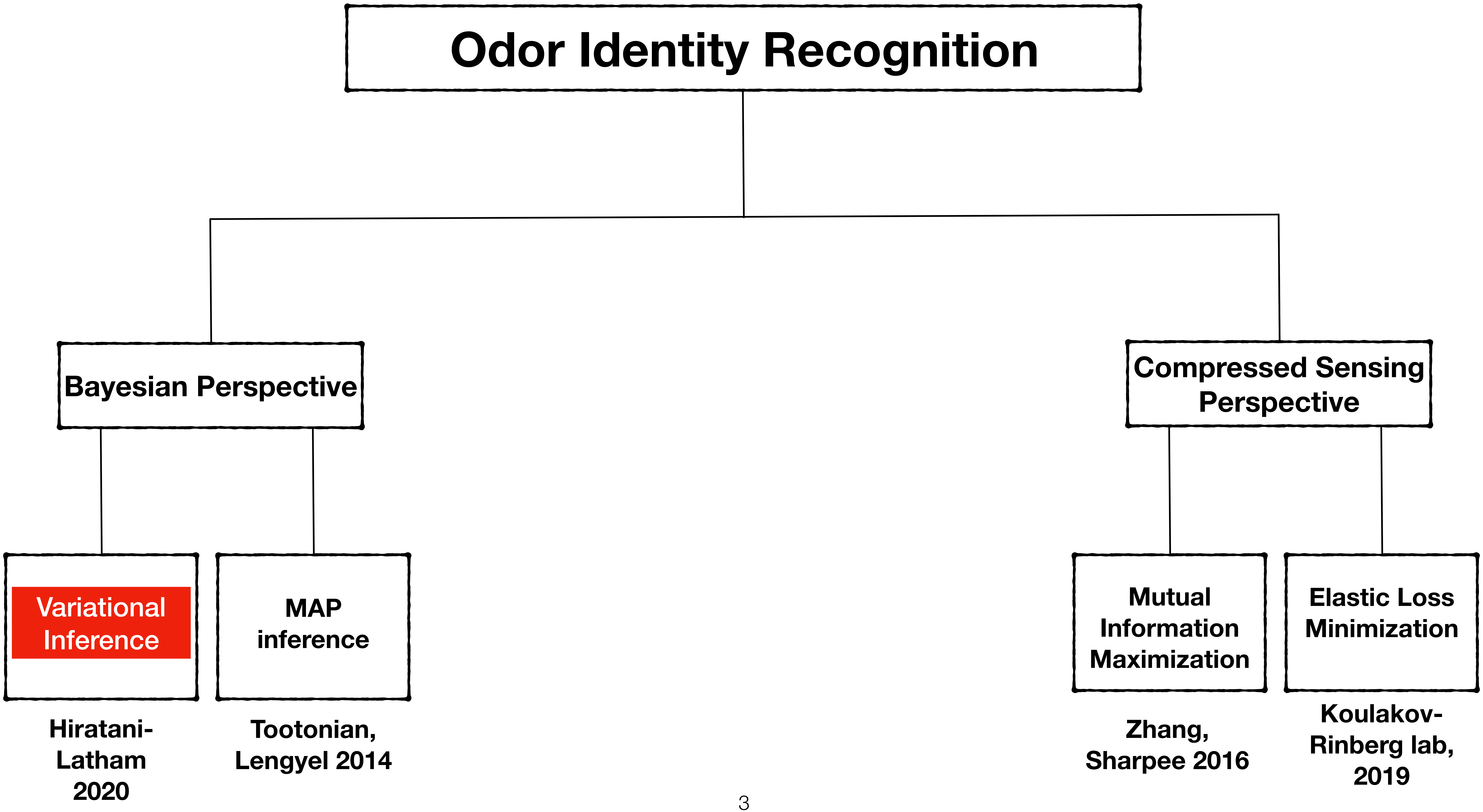
Rapid Bayesian Learning in Mammalian Olfactory System

Naoki Hiratani and Peter Latham, 2020

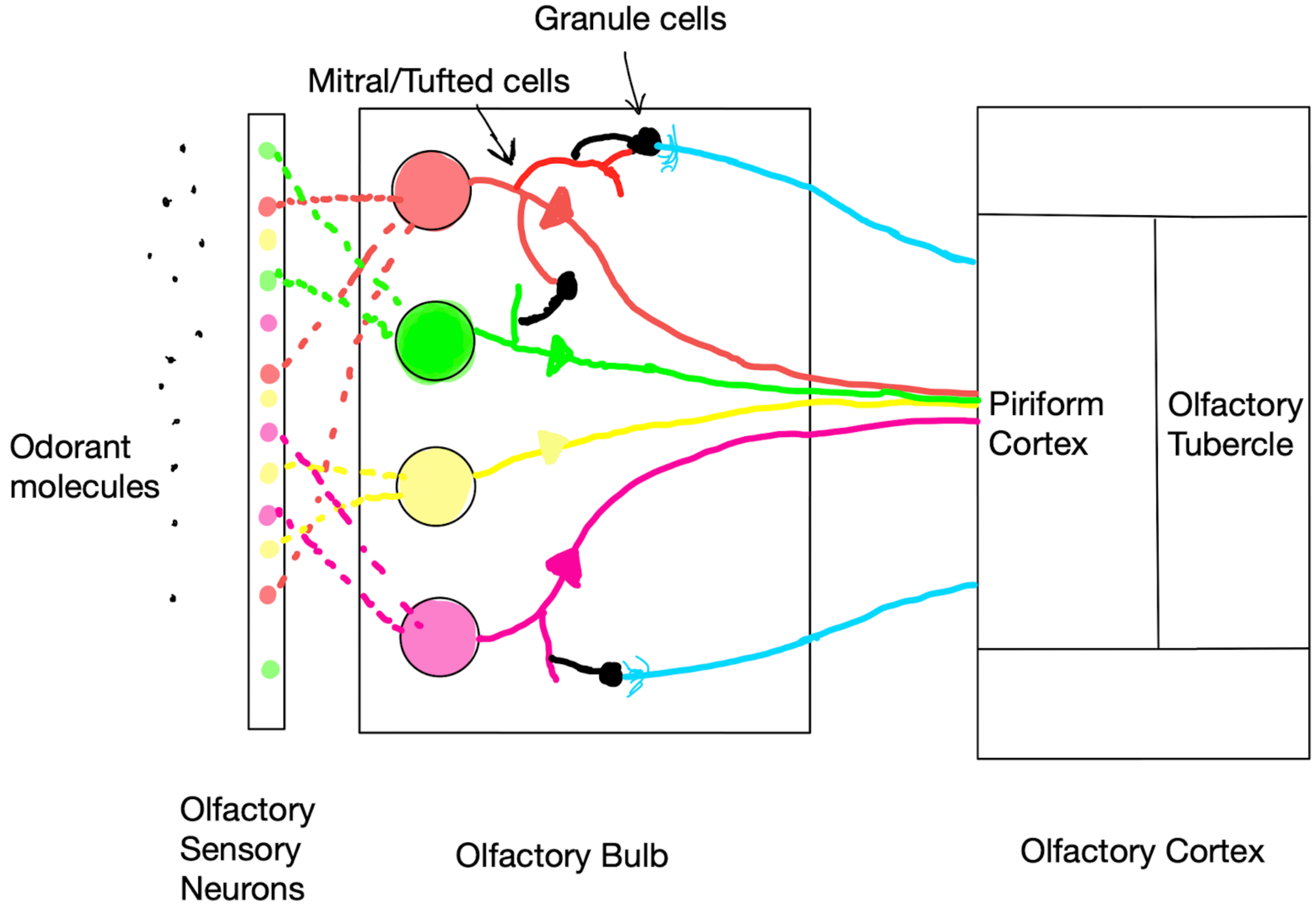
Presented By:
Achint Kumar

Central Question

- How olfactory system learns odor identity?
- How olfactory system performs odor-reward association?


Odor Identity Recognition
Bayesian Perspective
Compressed Sensing Perspective
Variational Inference
MAP inference
Mutual Information Maximization
Elastic Loss Minimization
Hiratani-Latham 2020
Tootonian, Lengyel 2014
Zhang, Sharpee 2016
Koulakov-Rinberg lab, 2019

Olfactory System in a Nutshell



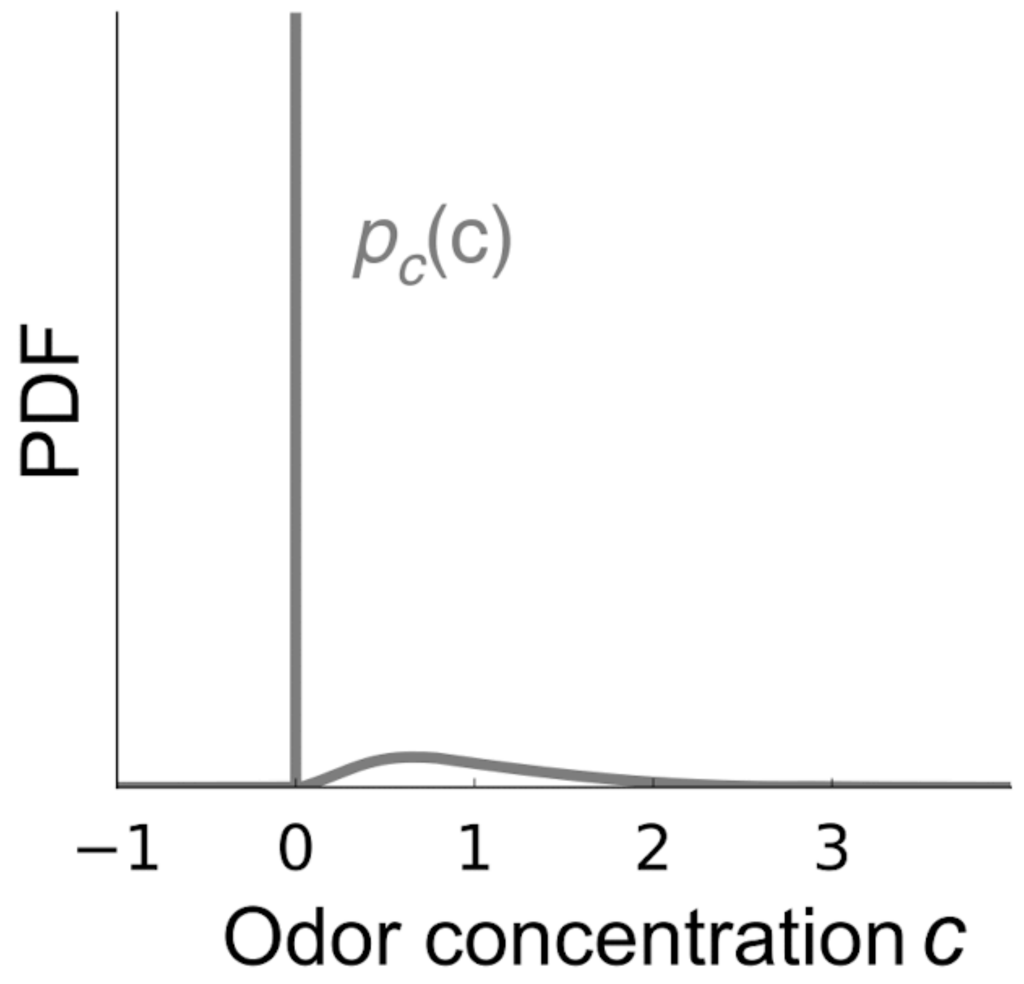
Problem Setting

$$p(c_j) = (1 - c_0)\delta(c_j) + \text{Gamma}(\alpha, \alpha)$$

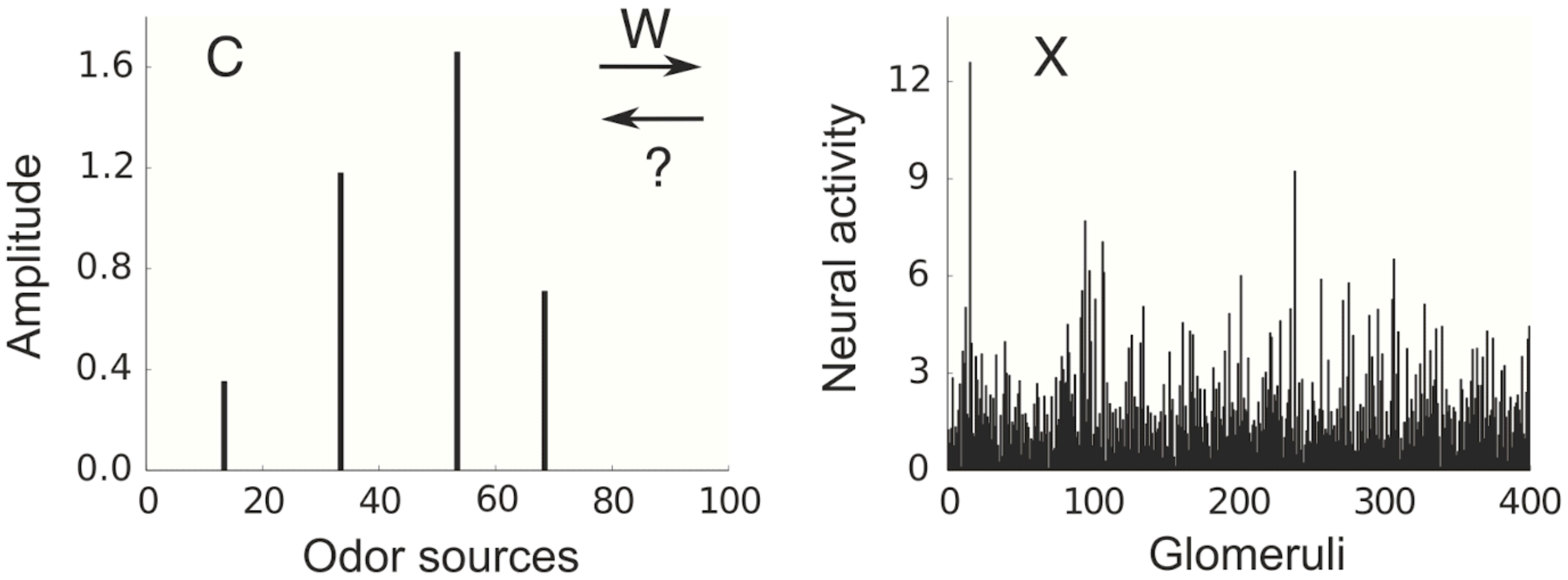
- Let **c** represent a Mx1 vector of odor concentration
- Let **x** represent a Nx1 vector of glomerular activity

$$x_i = \sum_{j=1}^{M} w_{ij} c_j + \xi_i$$

- Goal: Given **x**, find **c**



Goal: Find p(c|x)



Olfactory Learning as Bayesian Inference

Prelude: Why we need Variational Approximation?

Step 1: Use Variational Approximation to get variational odor and weight distributions

Step 2: Minimize KL divergence between variational distributions and the true distribution to get iterative relations for the variational distributions

Step 3: Solve the iterative relations to get expression for the variational distributions

Step 4: Show that the parameters of the variational distributions can be learnt by neural network dynamics and synaptic plasticity rules

Prelude: Why variational approximation?

$$
\begin{aligned}
p(\mathbf{c}_t|\mathbf{x}_{1:t}) &= \int d\mathbf{w}\, p(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t}) \\
&= \int d\mathbf{w}\, p(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t-1}, \mathbf{x}_t) \\
&= \int d\mathbf{w}\, \frac{p(\mathbf{x}_t|\mathbf{c}_t, \mathbf{w}, \mathbf{x}_{1:t-1})p(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t-1})}{p(\mathbf{x}_t)} \\
&\propto \int d\mathbf{w}\, p(\mathbf{x}_t|\mathbf{c}_t, \mathbf{w})p(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t-1}) \\
&= \int d\mathbf{w}\, p(\mathbf{x}_t|\mathbf{c}_t, \mathbf{w})p(\mathbf{c}_t)p(\mathbf{w}|\mathbf{x}_{1:t-1})
\end{aligned}
$$

$$p(A|B) = \frac{p(A,B)}{p(B)}$$

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

The goal is to evaluate the above integral. The first term in the integrand is Gaussian distribution, i.e. $p(\mathbf{x}_t|\mathbf{c}_t, \mathbf{w}) \sim \mathcal{N}(Wc, 1)$ and we also know the expression for $p(\mathbf{c}_t)$. It is the third term which makes it intractable to evaluate the integral.

Olfactory Learning as Bayesian Inference

Prelude: Why we need Variational Approximation?

Step 1: Use Variational Approximation to get variational odor and weight distributions

Step 2: Minimize KL divergence between variational distributions and the true distribution to get iterative relations for the variational distributions

Step 3: Solve the iterative relations to get expression for the variational distributions

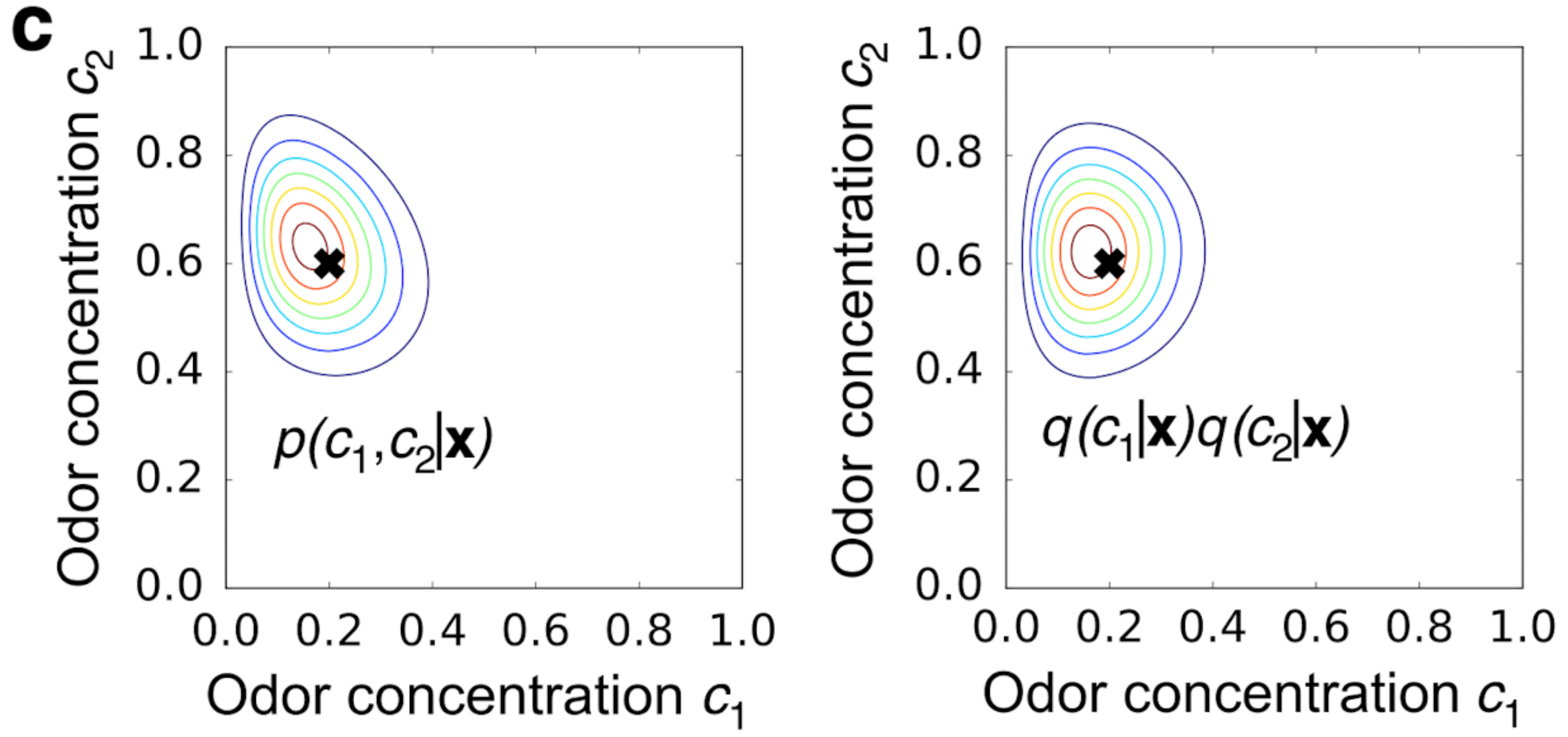
Step 4: Show that the parameters of the variational distributions can be learnt by neural network dynamics and synaptic plasticity rules

Step 1: Using Variational Approximation

We make the following assumption:

$$p(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t}) \approx q^t(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t}) \equiv \prod_{ij} q_{ij}^{w,t}(w_{ij}|\mathbf{x}_{1:t}) \times \prod_{j} q_j^c(c_j|\mathbf{x}_{1:t})$$

We are basically assuming each element of w_{ij} and c_j are independent. While this assumption for c_j is always true but w_{ij} independence is only true at the beginning of trials and this approximation become progressively worse as the number of trials increases.



Olfactory Learning as Bayesian Inference

Prelude: Why we need Variational Approximation?

Step 1: Use Variational Approximation to get variational odor and weight distributions

Step 2: Minimize KL divergence between variational distributions and the true distribution to get iterative relations for the variational distributions

Step 3: Solve the iterative relations to get expression for the variational distributions

Step 4: Show that the parameters of the variational distributions can be learnt by neural network dynamics and synaptic plasticity rules

Step 2: Minimizing KL divergence

$$p(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t}) \approx q^t(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t}) \equiv \prod_{ij} q_{ij}^{w,t}(w_{ij}|\mathbf{x}_{1:t}) \times \prod_j q_j^c(c_j|\mathbf{x}_{1:t})$$

We minimize the KL divergence between: $p(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t})$ and $q^t(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t})$

$$D_{KL}[q^t(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t})||p(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t})] = \int d\mathbf{c}d\mathbf{w}\, q^t(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t})\log\frac{q^t(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t})}{p(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t})}$$

Taking functional derivative with respect to $q_{ij}^{w,t}(w_{ij}|\mathbf{x}_{1:t})$ and equating to zero gives,

$$\begin{aligned} 0 &= \int d\mathbf{c}\frac{\partial}{\partial q_{ij}^{w,t}}\int d\mathbf{w}\, q^t(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t})(\log[q^t(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t})] - \log[p(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t})]) \\ &= -\int d\mathbf{c}d\mathbf{w}_{\backslash w_{ij}} \log[p(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t})] + \int d\mathbf{c}d\mathbf{w}_{\backslash w_{ij}} \log[q^t(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t})] \\ &+ \int d\mathbf{c}d\mathbf{w}_{\backslash w_{ij}} [q^t(\mathbf{c}_t, \mathbf{w}|\mathbf{x}_{1:t})] \end{aligned}$$

Next, we take functional derivative with respect to$q_j^c(c_j|\mathbf{x}_{1:t})$. We are left with these two relations:

$$\begin{aligned} \log q_{ij}^{w,t} &\sim \langle \log p(\mathbf{x}|\mathbf{c}, \mathbf{w})\rangle_{\backslash w_{ij}} + \langle \log p(\mathbf{w}|\mathbf{x}_{1:t-1})\rangle_{\backslash w_{ij}} \\ \log q_j^c &\sim \langle \log p(\mathbf{x}|\mathbf{c}, \mathbf{w})\rangle_{\backslash c_j} + \log p(c_j) \end{aligned}$$

Olfactory Learning as Bayesian Inference

Prelude: Why we need Variational Approximation?

Step 1: Use Variational Approximation to get variational odor and weight distributions

Step 2: Minimize KL divergence between variational distributions and the true distribution to get iterative relations for the variational distributions

Step 3: Solve the iterative relations to get expression for the variational distributions

Step 4: Show that the parameters of the variational distributions can be learnt by neural network dynamics and synaptic plasticity rules

Step 3: Solving for variational odor distribution

$$q_j^c(c_j) \propto p_c(c_j) \exp\left[\langle \log p(\mathbf{x}|\mathbf{c}, \mathbf{w}) \rangle_{\setminus c_j} \right]$$

$$q_j^c(c_j) \propto p_c(c_j) \ \exp\left[-\frac{\lambda_j^t}{2} \left(c_j - \mu_j^t \right)^2 \right]$$

$$\mu_j^t \equiv \frac{1}{\sum_i \langle {w_{ij}^{t-1}}^2 \rangle} \sum_i \langle w_{ij}^{t-1} \rangle \left[m_i^t + \langle w_{ij}^{t-1} \rangle \langle c_j \rangle \right]$$

$$\lambda_j^t \equiv \sum_i \langle {w_{ij}^{t-1}}^2 \rangle$$

$$m_i^t \equiv x_i^t - \sum_{j=1}^{M} \langle w_{ij}^{t-1} \rangle \langle c_j \rangle$$

Olfactory Learning as Bayesian Inference

Prelude: Why we need Variational Approximation?

Step 1: Use Variational Approximation to get variational odor and weight distributions

Step 2: Minimize KL divergence between variational distributions and the true distribution to get iterative relations for the variational distributions

Step 3: Solve the iterative relations to get expression for the variational distributions

Step 4: Show that the parameters of the variational distributions can be learnt by neural network dynamics and synaptic plasticity rules

Step 4: Inferring the neural dynamics and learning rule

- From the variational odor distribution:

$$\tau_{\mathrm{r}} \frac{\mathrm{d}m_i}{\mathrm{d}\tau} = -m_i - \sum_{j=1}^{M} w_{ij}^{\mathrm{L}} \bar{c}_j + x_i$$

$$\tau_{\mathrm{r}} \frac{\mathrm{d}\bar{c}_j}{\mathrm{d}\tau} = -\bar{c}_j + F_j\left(\sum_{i=1}^{N} w_{ji}^{\mathrm{F}} m_i\right)$$

- From the variational weight distribution:

$$\Delta w_{ji}^{F,t} = \frac{m_i c_j}{t \rho_j}$$

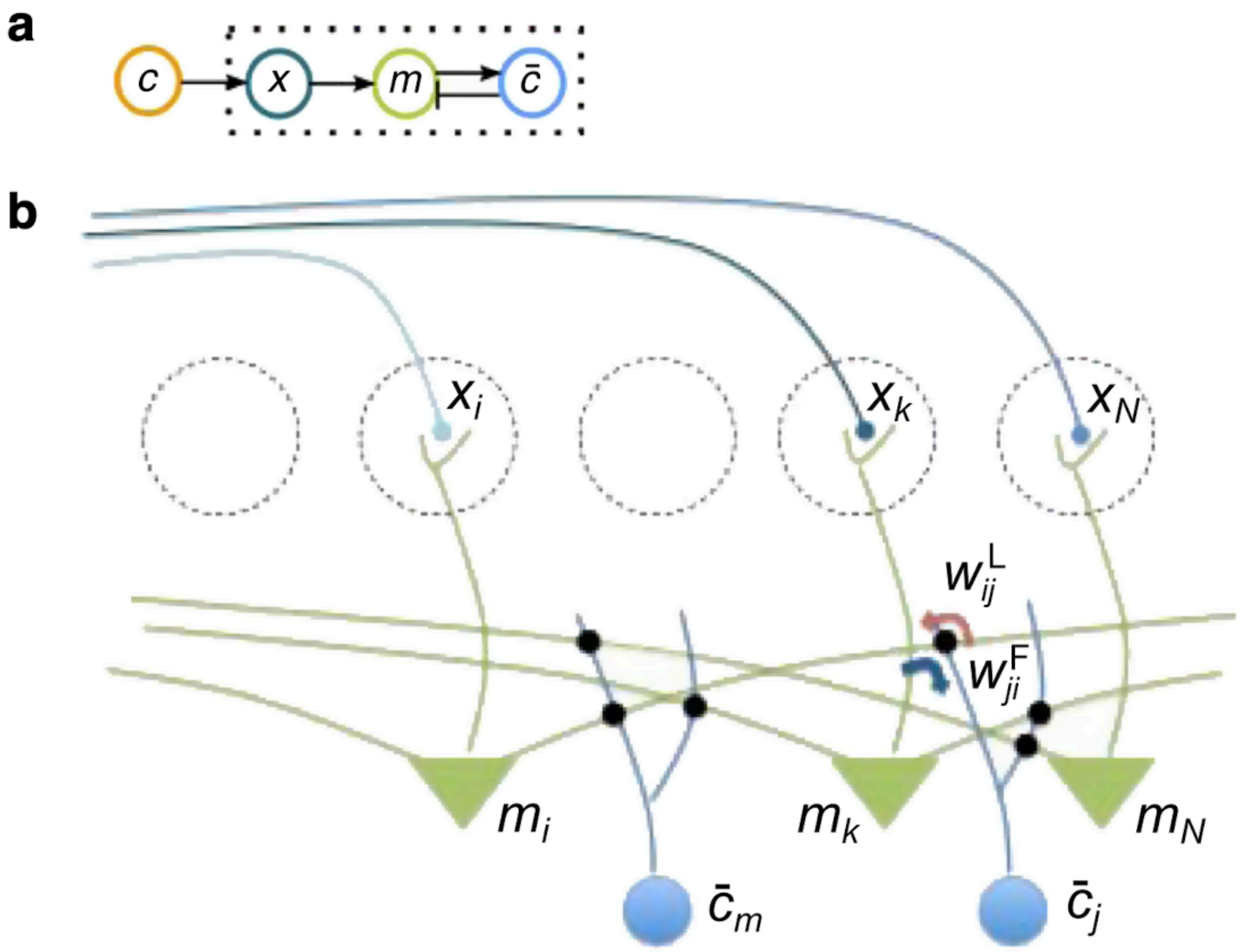
$$\Delta w_{ji}^{L,t} = \frac{m_i c_j}{t \rho_j}$$

Neural Implementation of Bayesian Learning

$$\tau_{\mathrm{r}} \frac{\mathrm{d}m_i}{\mathrm{d}\tau} = -m_i - \sum_{j=1}^{M} w_{ij}^{\mathrm{L}} \bar{c}_j + x_i$$

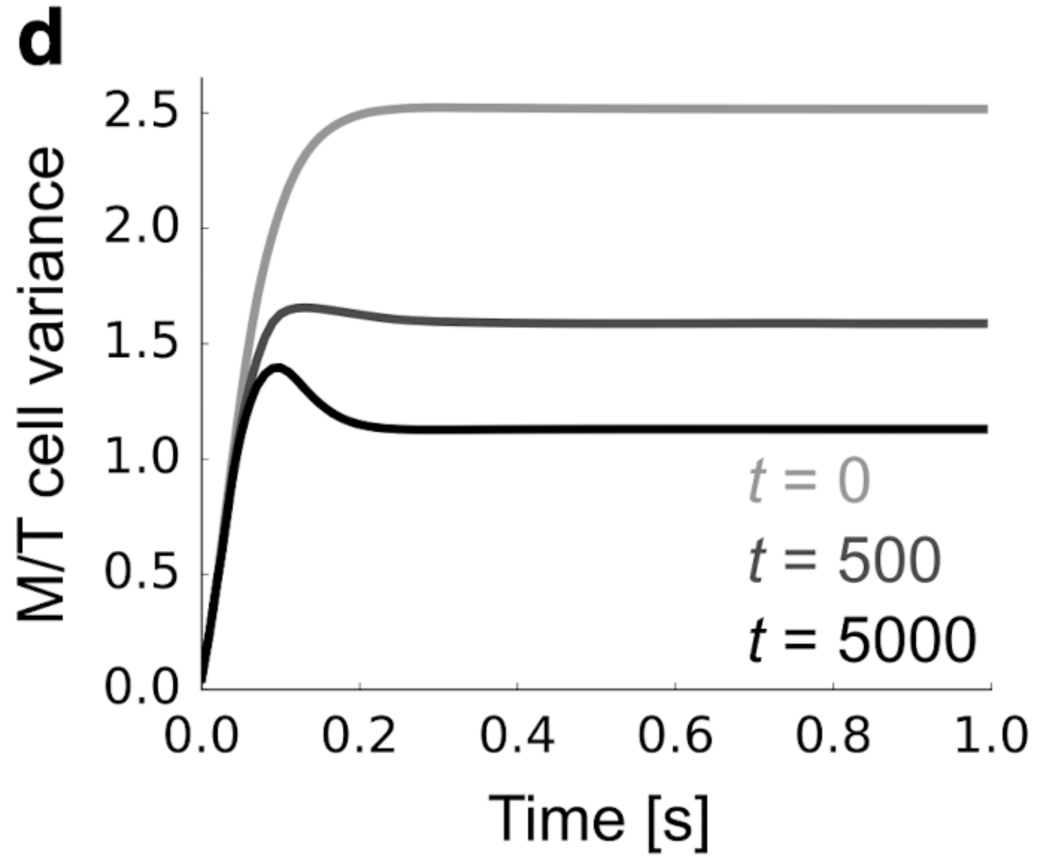
$$\tau_{\mathrm{r}} \frac{\mathrm{d}\bar{c}_j}{\mathrm{d}\tau} = -\bar{c}_j + F_j\left(\sum_{i=1}^{N} w_{ji}^{\mathrm{F}} m_i\right)$$

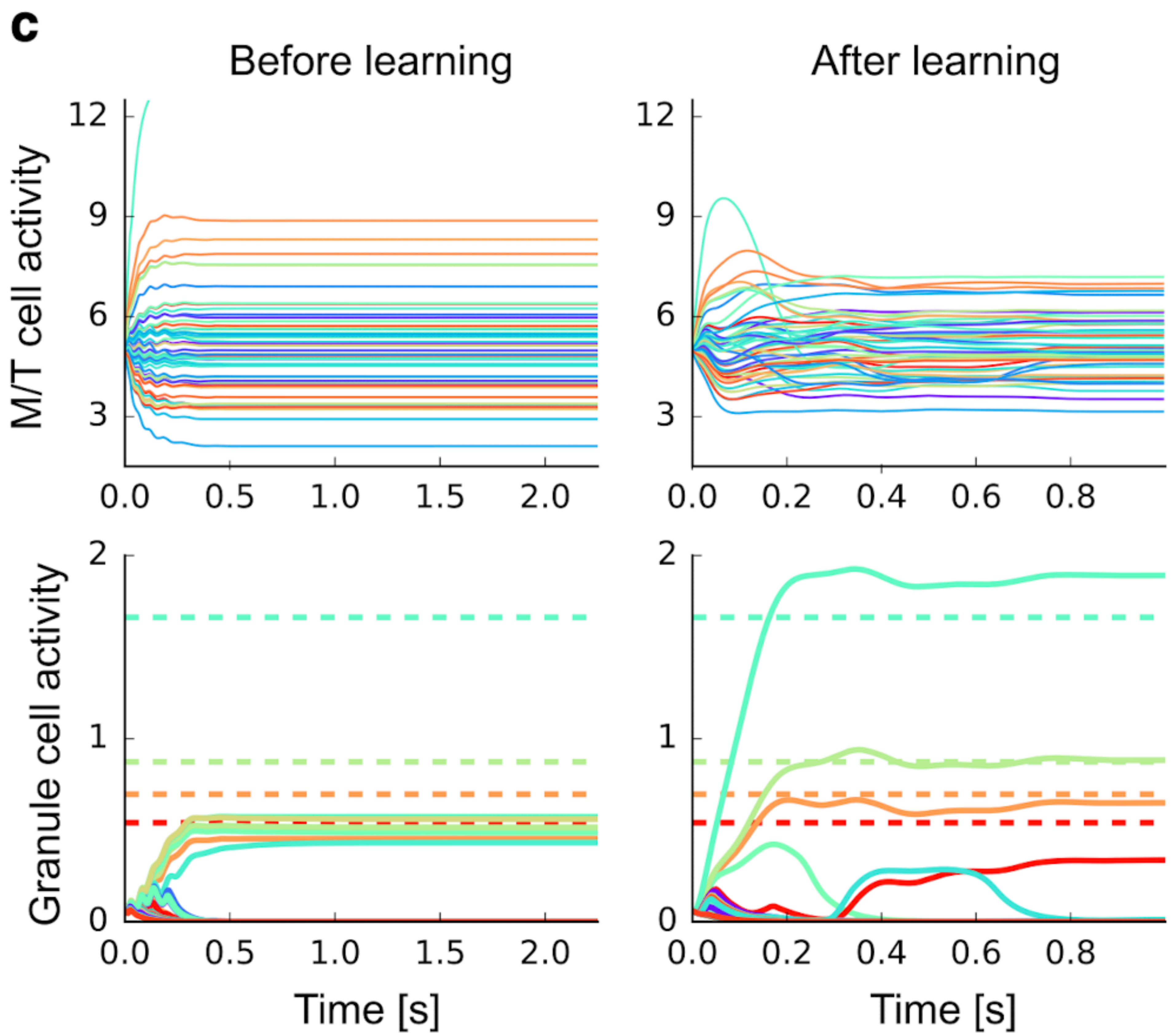
Bayesian Learning model maps perfectly onto the circuitry of the Olfactory Bulb



Firing Rate Dynamics before and after learning

- M/T cells show both positive and negative responses relative to baseline
- Granule cells show very selective responses with activity levels roughly matching the concentration of the corresponding odors
- M/T cells response range decreases with learning





Transfer Function

$$\tau_{\mathrm{r}} \frac{\mathrm{d}m_i}{\mathrm{d}\tau} = -m_i - \sum_{j=1}^{M} w_{ij}^{\mathrm{L}} \bar{c}_j + x_i$$

$$\tau_{\mathrm{r}} \frac{\mathrm{d}\bar{c}_j}{\mathrm{d}\tau} = -\bar{c}_j + F_j\left(\sum_{i=1}^{N} w_{ji}^{\mathrm{F}} m_i\right)$$

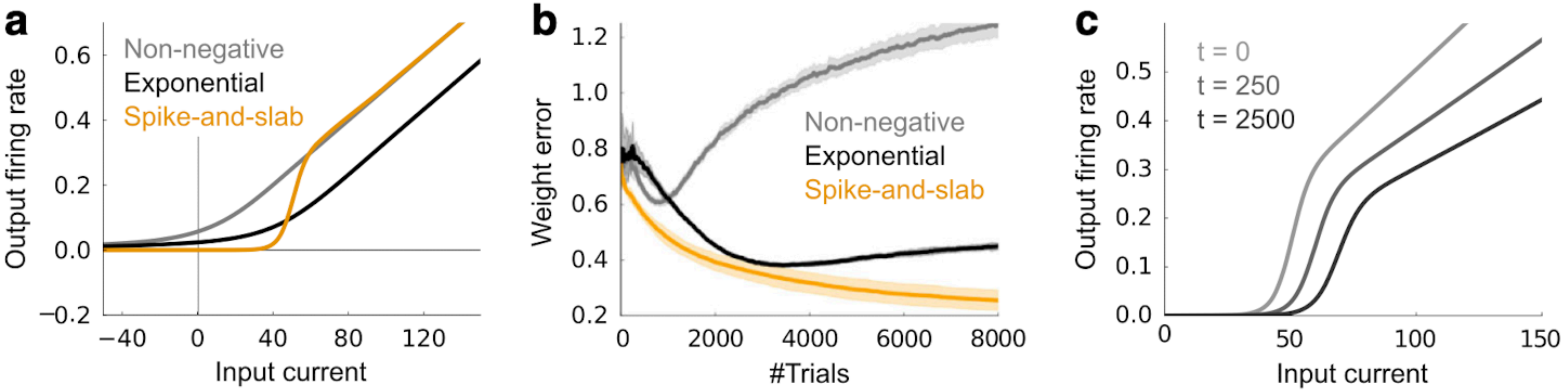
$$\Delta w_{ji}^{F,t} = \frac{m_i c_j}{t \rho_j}$$

$$\Delta w_{ji}^{L,t} = \frac{m_i c_j}{t \rho_j}$$

$$p_{\mathrm{c}}(c) \propto \Theta(c)$$

$$p_{\mathrm{c}}(c) = \frac{1}{c_{\mathrm{o}}} \exp(-c/c_{\mathrm{o}})$$

$$p(c_j) = (1 - c_0)\delta(c_j) + \mathrm{Gamma}(\alpha, \alpha)$$



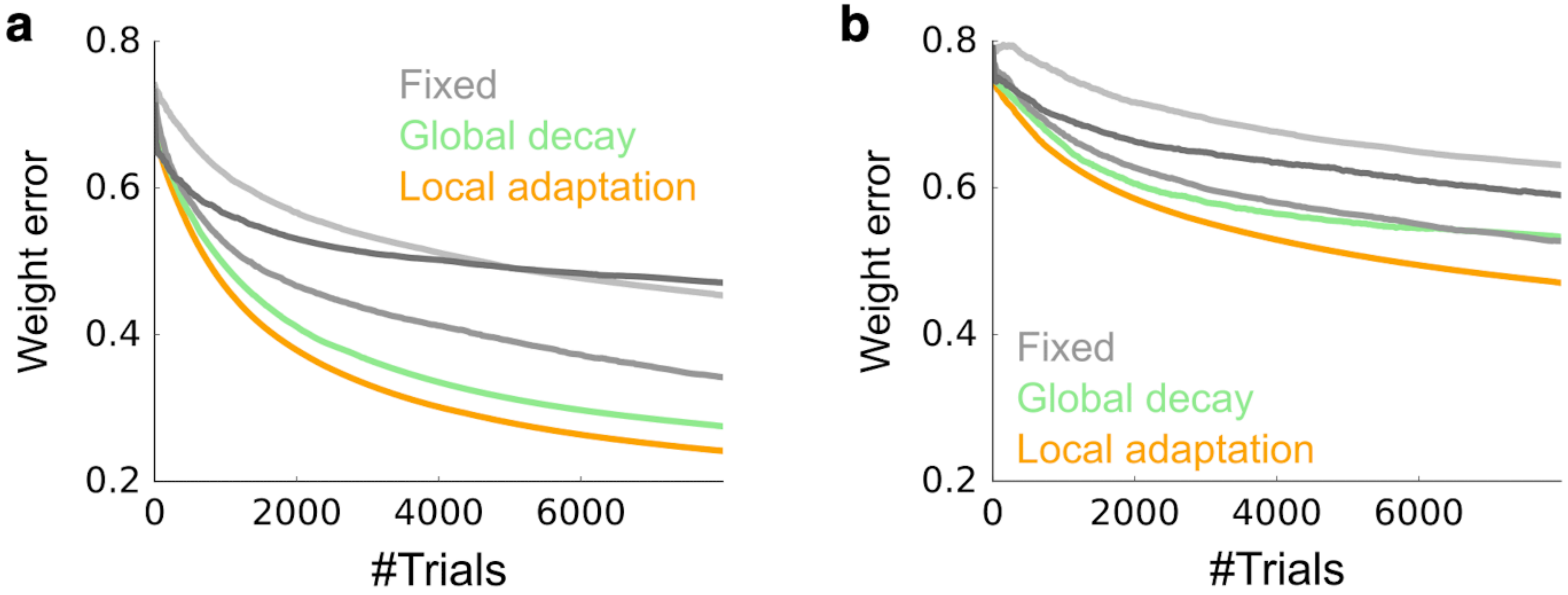
Synaptic Plasticity rule

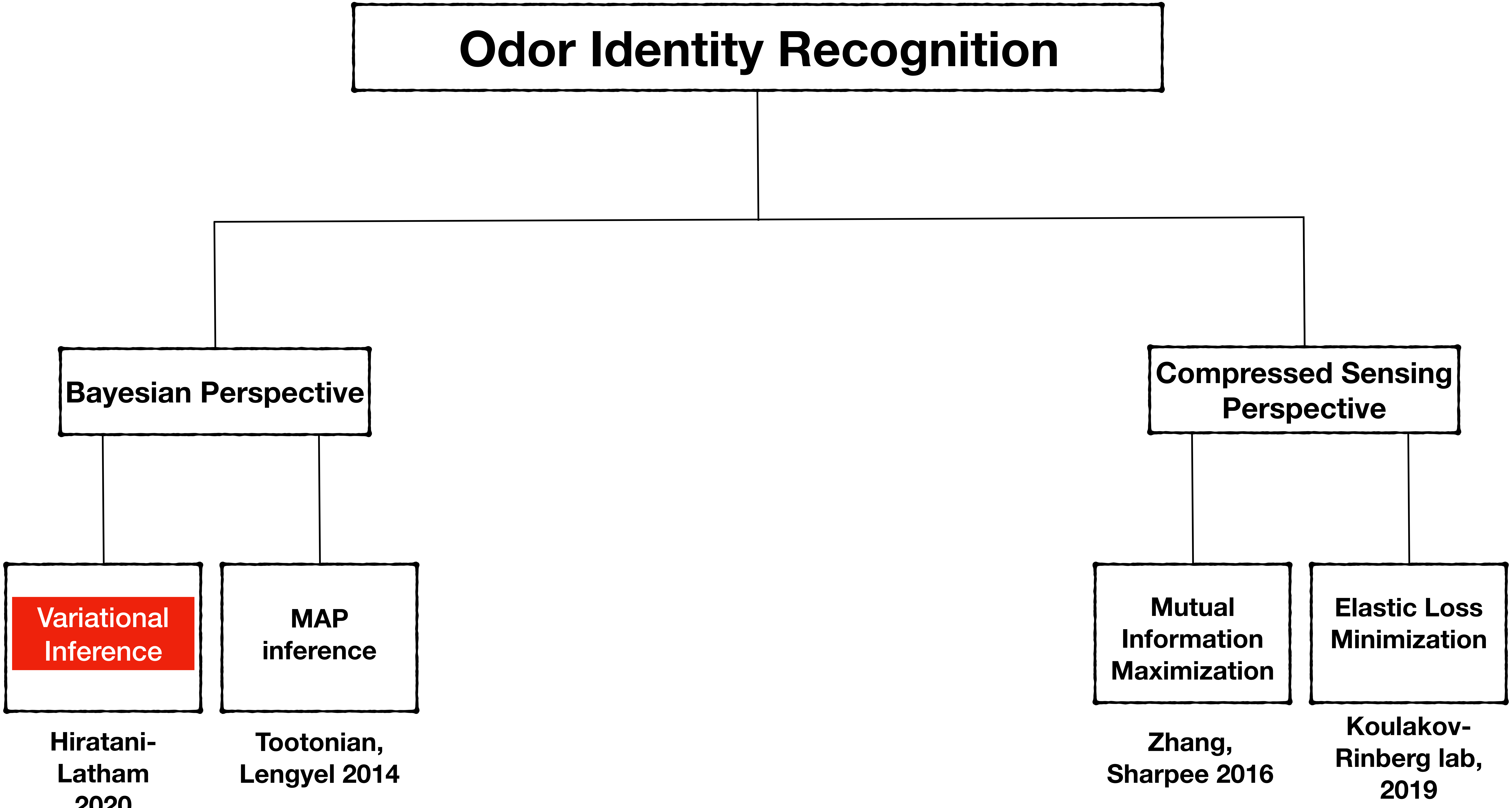
$$\Delta w_{ji}^{F,t} = \frac{m_i c_j}{t\rho_j}$$

$$\Delta w_{ji}^{L,t} = \frac{m_i c_j}{t\rho_j}$$

- Learning rate is product of two terms: $1/(t\rho)$
- $1/\rho_j \propto S_j$ (lifetime sparseness)
- Fixed: $1/(t\rho)$=constant
- Global Decay: $1/\rho$ = constant

3 odors presented/trial **7 odors presented/trial**




Odor Identity Recognition
Bayesian Perspective
Compressed Sensing Perspective
Variational Inference
Hiratani-Latham 2020
MAP inference
Tootonian, Lengyel 2014
Mutual Information Maximization
Zhang, Sharpee 2016
Elastic Loss Minimization
Koulakov-Rinberg lab, 2019

MAP inference: Quick Introduction

By Bayes rule,

$$p(\mathbf{c}_t|\mathbf{x}_t) \propto p(\mathbf{x}_t|\mathbf{c}_t)p(\mathbf{c}_t)$$

So, the Maximum a posteriori estimate is given by:

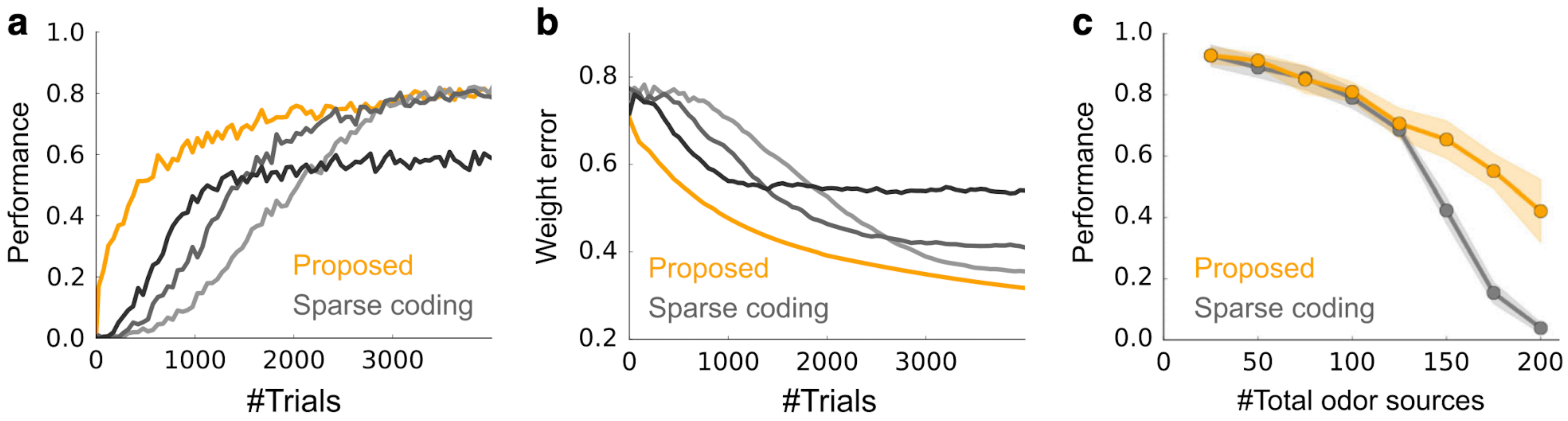
$$\widehat{\mathbf{c}}_t = \arg\max_{\mathbf{c}} p(\mathbf{x}_t|\mathbf{c}, \widehat{\mathbf{w}})p(\mathbf{c})$$

We define an objective function: $E_t \equiv \log p(\mathbf{x}_t|\widehat{\mathbf{c}}_t, \widehat{\mathbf{w}}) + \log p(\widehat{\mathbf{c}}_t)$ and maximize with respect to c_j and w_{ij}

Tootonian, Lengyel 2014

Comparison between Variational and MAP inference

- Variational approach doesn't require fine tuning and learns much faster (see a)
- Variational approach leads to lower error in weights (see b)
- Variational approach performs better when there are large number of odor sources (see c)



Conclusion

- Variational Bayesian inference leads to a biologically plausible learning rules
- Some predictions are consistent with experiments, others can tested by future experiments
- Future work could involve, including periglomerular cells which have been implicated in whitening of OB, a natural prediction of circuitry in PCx