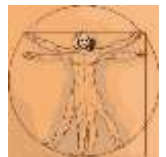


Modeling how the brain learns to represent the world: *~Abstraction and Probability~*

Taro Toyoizumi
RIKEN Center for Brain Science

The brain facilitates rapid adaptation to its environment.
How to learn the representation of the environment?



The brain's constraints

A bounded rationality perspective

(Simon 1955; Gigerenzer and Gaissmaier 2011)

- The brain opts for satisfactory rather than optimal solutions, constrained by its cognitive and biological limits.
- Despite the challenge, it effectively learns to select appropriate heuristics from its adaptive toolbox.
- These constraints likely shape how the brain represents abstract and probabilistic information.

Abstraction via dimensionality reduction

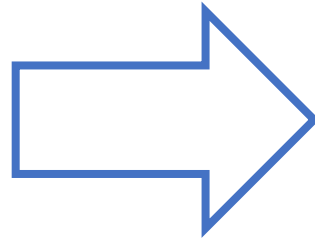
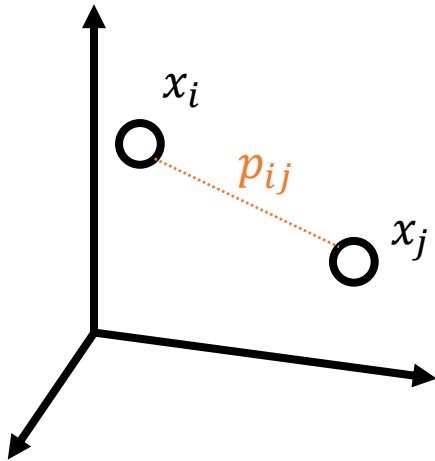
Dimensionality reduction is useful for humans

Unlike large-scale machine learning models, the brain processes only a limited amount of information—a constraint that may have driven the development of abstract representation.

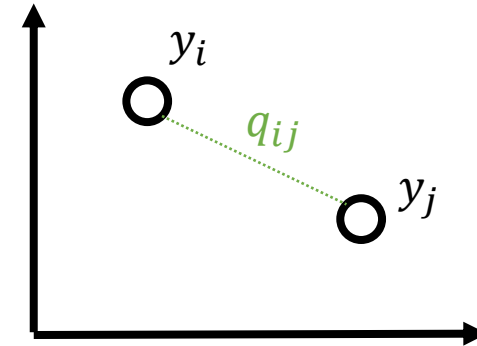
- Linear dimensionality reduction (e.g. Principal component analysis; PCA)
Biologically-plausible implementations have been proposed.
- Nonlinear dimensionality reduction (e.g. t-distributed stochastic neighbor embedding; t-SNE)
A biological implementation remains unknown.

t-SNE (Maaten and Hinton, 2008)

High-dimensional inputs



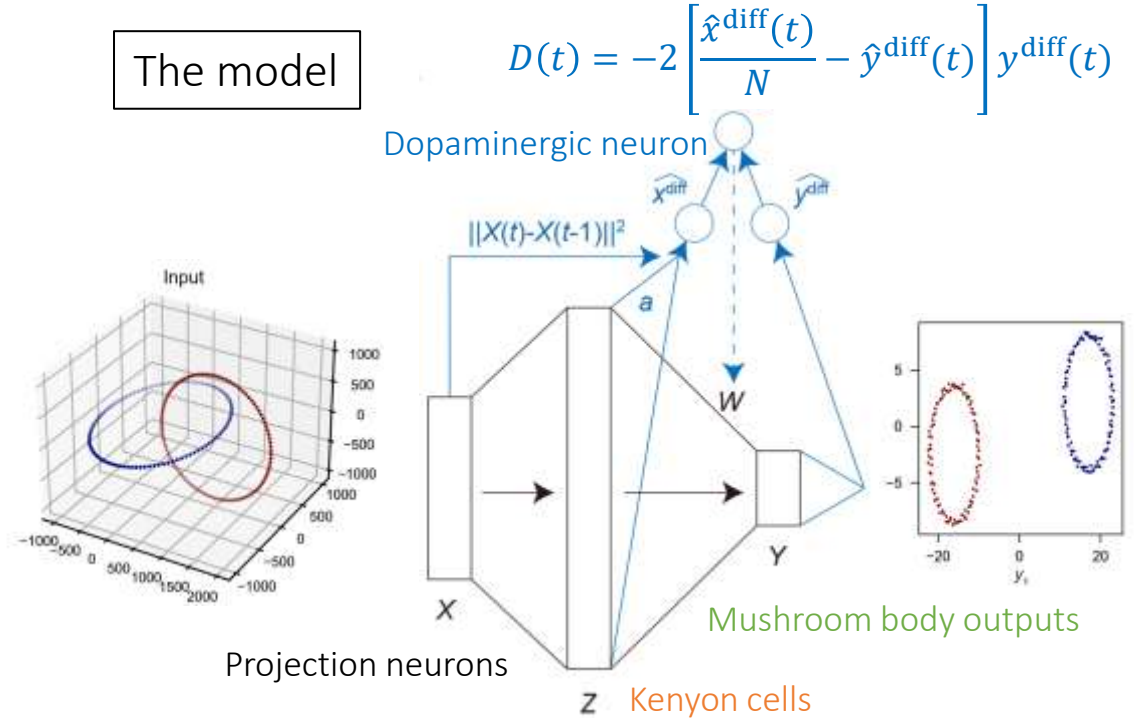
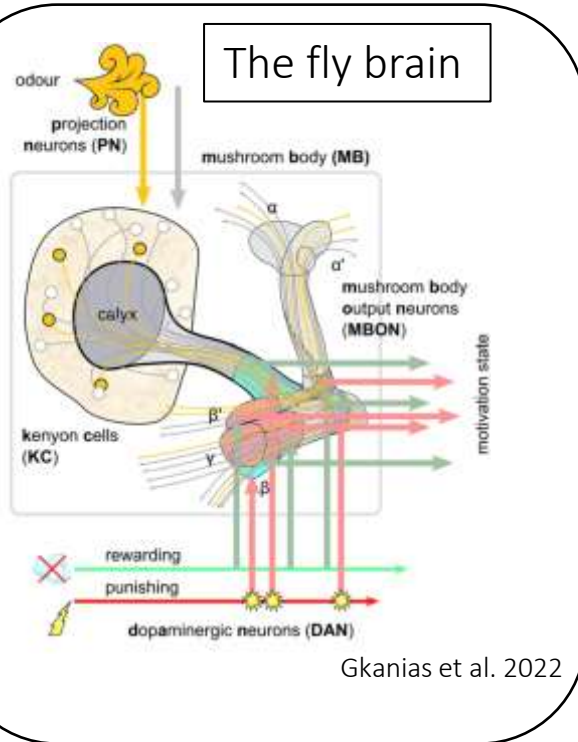
Low-dimensional outputs



Adjust the output points so that the pairwise similarity q between outputs closely approximate the corresponding pairwise similarity p between inputs.

$$\min_{\{y_i\}} \text{KL}[p_{ij} || q_{ij}]$$

Hebbian t-SNE: A biologically plausible implementation

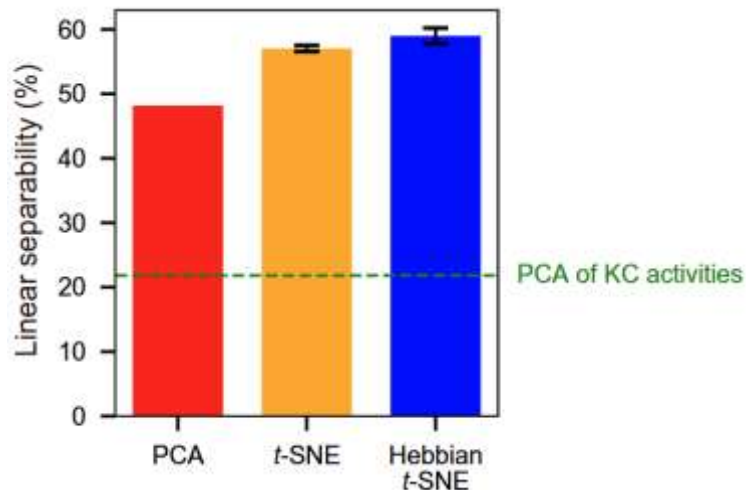


A three-factor Hebbian synaptic plasticity rule

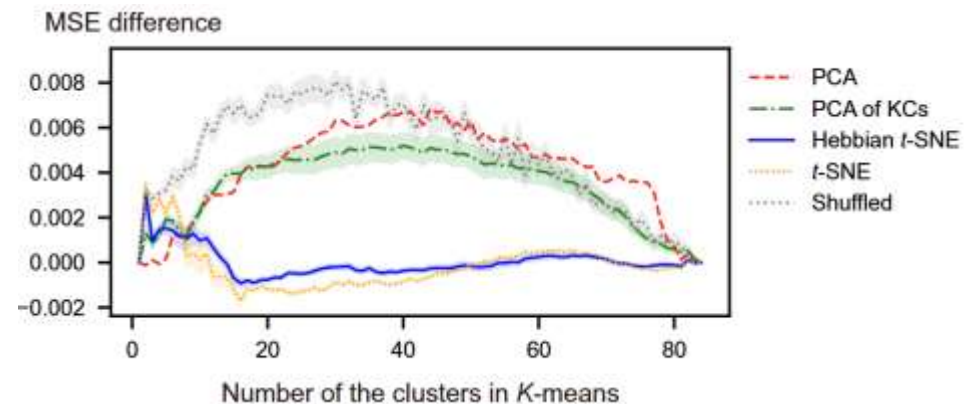
$$\Delta W \propto \underbrace{D(t)}_{\text{Global factor}} \underbrace{[y(t) - y(t-1)]}_{\text{Postsynaptic}} \underbrace{[z(t) - z(t-1)]^T}_{\text{Presynaptic}}$$

Hebbian t-SNE based on fly olfactory coding data

It produced odor representations that were well separated by chemical structure in an unsupervised manner using the activity of olfactory receptor neurons.



It effectively generalized odorant valence to similar odors via semi-supervised learning based on the activity of olfactory projection neurons.



Summary

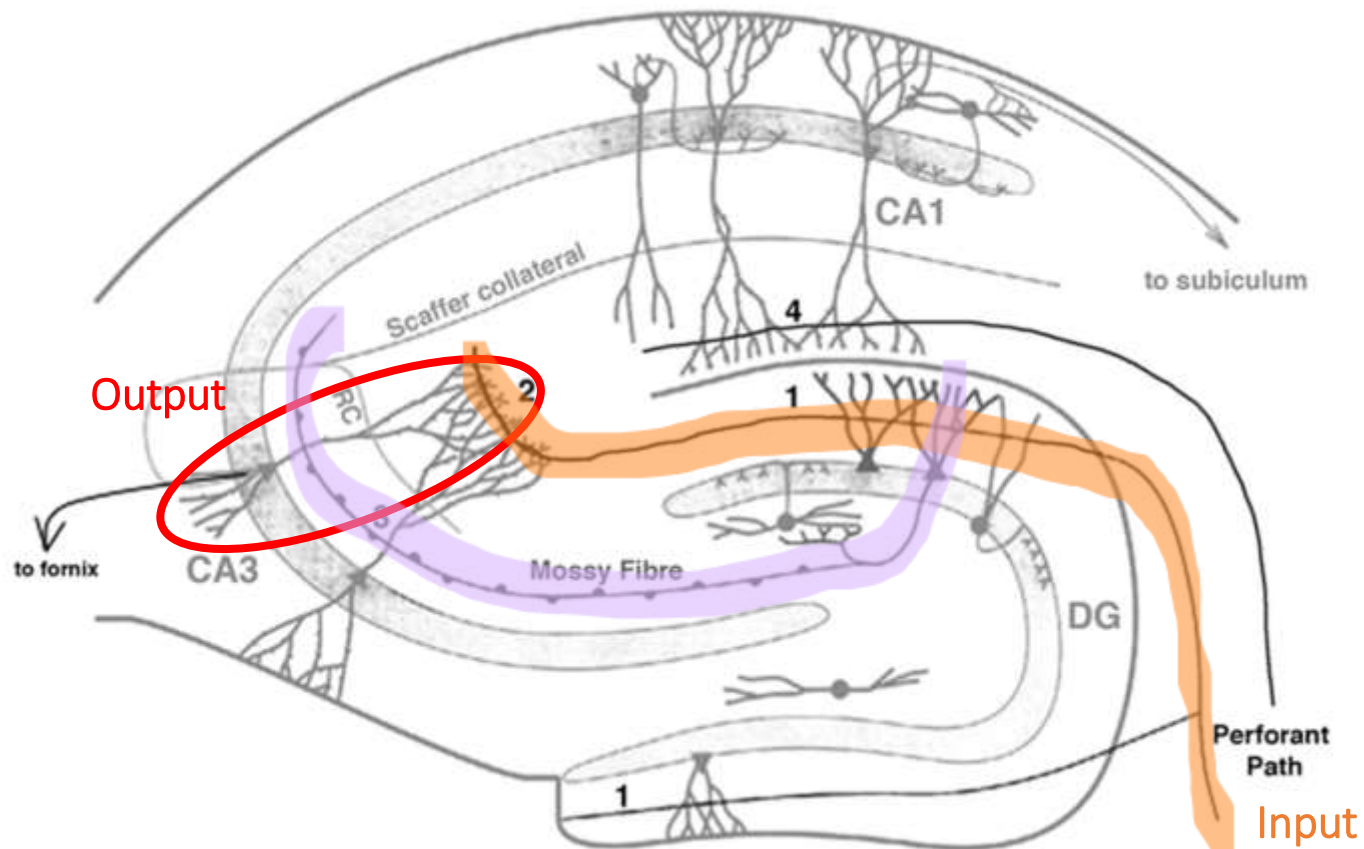
- We present Hebbian t-SNE, a biologically plausible dimensionality reduction algorithm.
- Hebbian t-SNE operates on a simple three-layer feedforward network that emulates the *Drosophila* olfactory circuit.
- The corresponding three-factor Hebbian synaptic plasticity rule effectively disentangled benchmark datasets with performance comparable to that of the original t-SNE.
- Hebbian t-SNE learned representations from recorded neural data more efficiently than alternative methods in both unsupervised and semi-supervised modes.

Bridging abstract and concrete representations of memory

The hippocampal CA3 receives two input pathways that exhibit distinct coding properties.

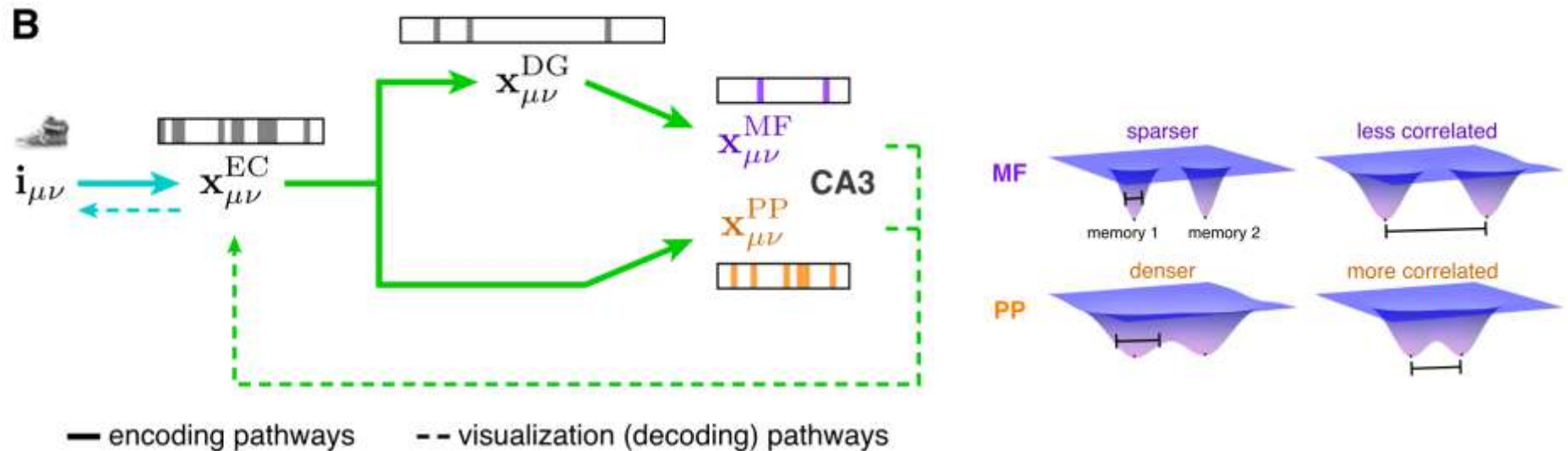
The direct pathway with dense neural activity (The perforant pathway)

The indirect pathway with sparse neural activity (The mossy fiber pathway)



Association between abstract and concrete information

- Modeling **direct** and **indirect** input paths.



A dual-pathway model of associative memory

Abstract categories

Concrete examples



FashionMNIST

Memory is embedded as the superposition of input patterns from the two pathways.

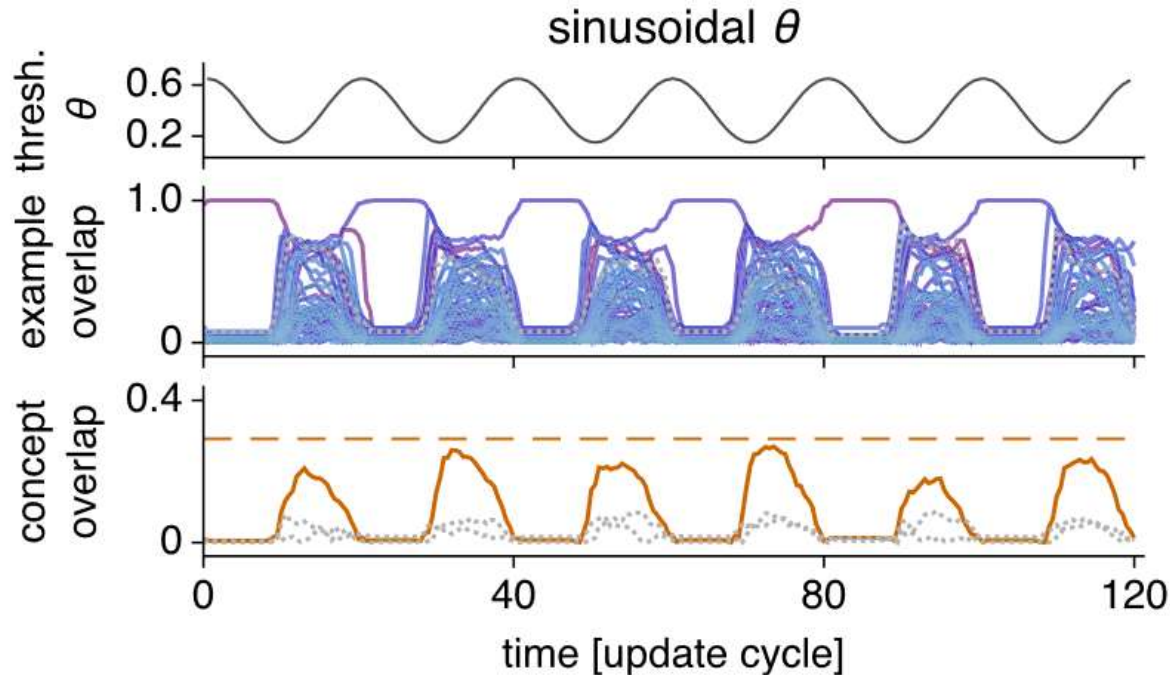
Sparse
(to distinguish)

Dense
(to relate)

0.9 ×  + 0.1 × 

Hetero association with theta wave modulation

Switching the recalled input pattern is accomplished by adjusting the neural threshold.



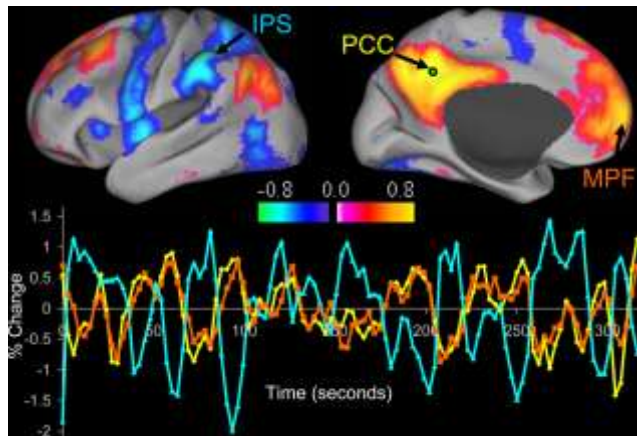
Summary

- We model CA3 as a Hopfield-like network that stores both dense and sparse representations.
- As additional memories are stored, dense representations merge based on shared features, while sparse representations remain distinct.
- Depending on the theta phase, the model retrieves either concrete examples via sparse activity or abstract categories via dense activity.
- These complementary encoding mechanisms may offer computational advantages for solving complex tasks.

Fluctuating brain activity and chaos

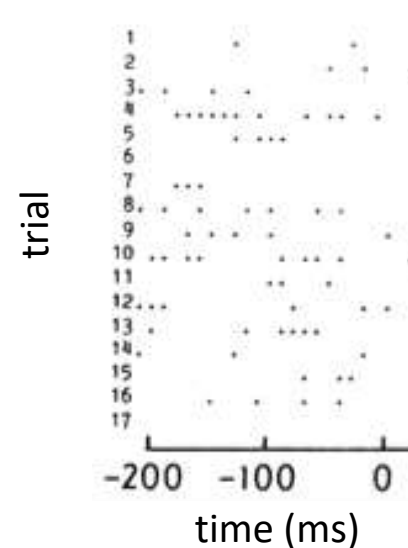
The brain exhibits variability in neural activity

BOLD signal varies even at rest.



Fox and Greicius 2010

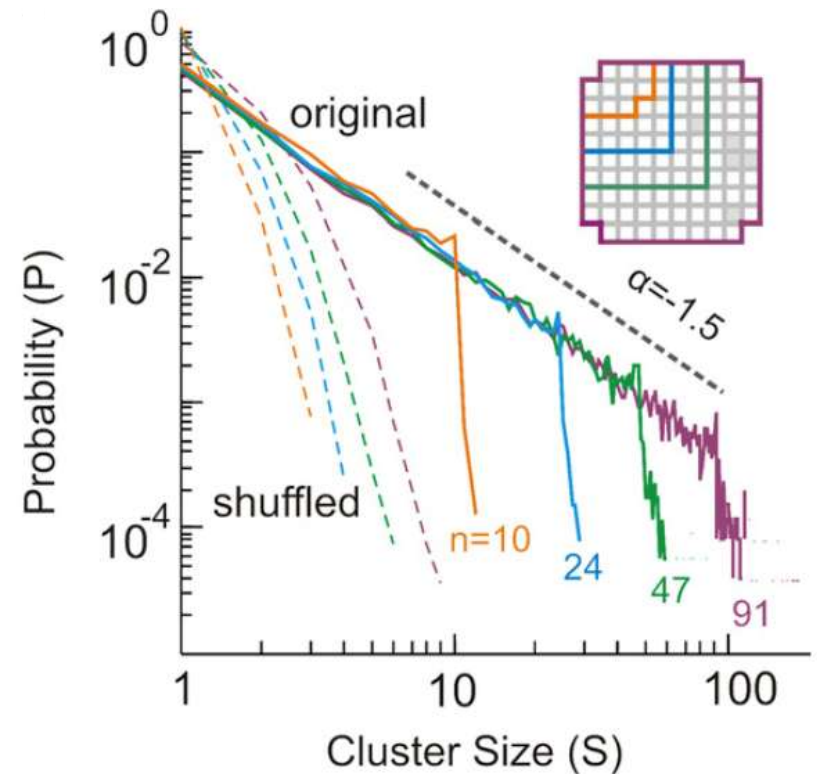
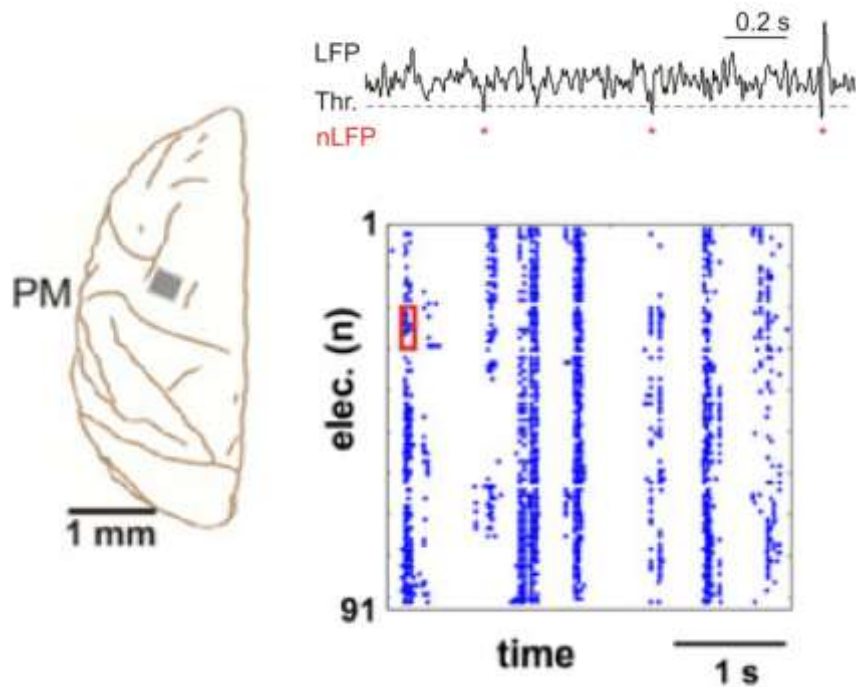
Apparently random spiking patterns



Perrett et al. 1982

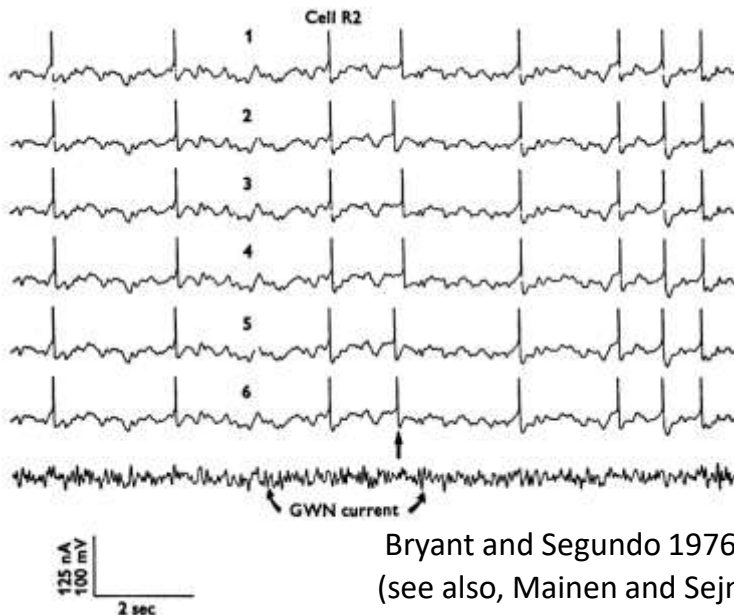
Variability is an intrinsic property of neural dynamics.

Variability is observed at multiple scales: neuronal avalanches

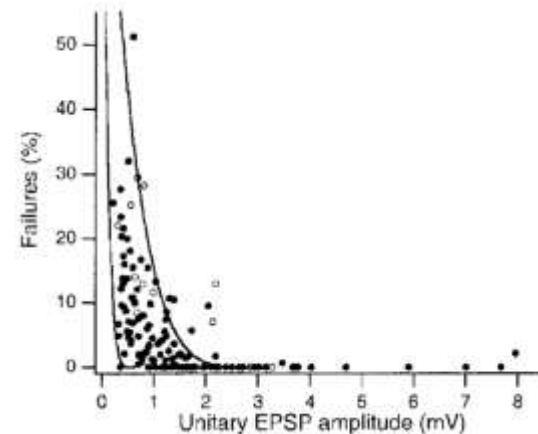


What is the source of neural variability?

- Dynamical behavior of a neuron can be reliable in isolation.



Bryant and Segundo 1976
(see also, Mainen and Sejnowski 1995)



Feldmeyer et al. 1999

- One hypothesis is that the *chaotic* activity of a network of neurons produces the apparently stochastic nature of observed background activity (van Vreeswijk and Sompolinsky 1996).

(See also, Amari 1974, Freeman 1975, Babloyantz et al. 1985, etc.)

Variability due to chaotic neural dynamics

Sompolinsky et al. 1988; Molgedey et al. 1992

A deterministic update equation:

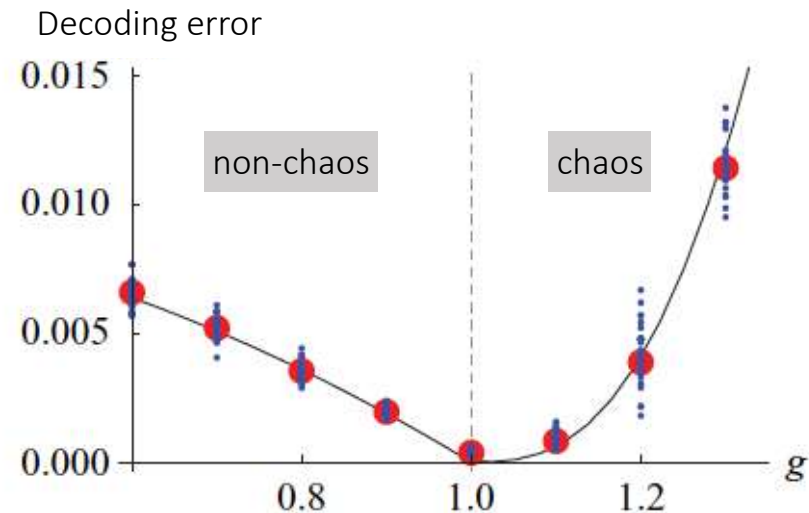
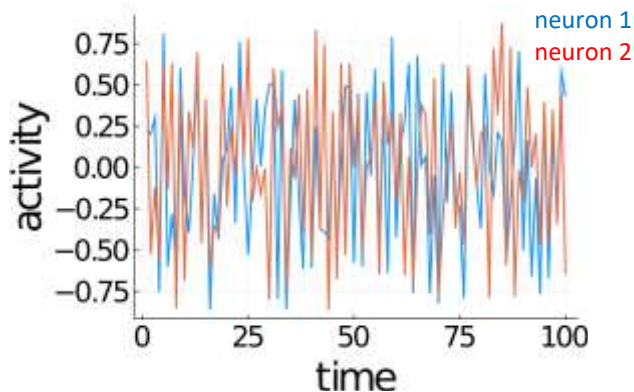
$$\overset{\text{neural activity}}{x_i(t+1)} = \underset{\text{tanh nonlinearity}}{\phi} \left(\sum_{j=1}^N W_{ij} x_j(t) \right)$$

The intrinsic dynamics affect information coding.
The sensitivity to input is maximized at the edge of chaos.

Random Gaussian synaptic weights:

$$W_{ij} \sim \mathcal{N} \left(0, \frac{g^2}{N} \right)$$

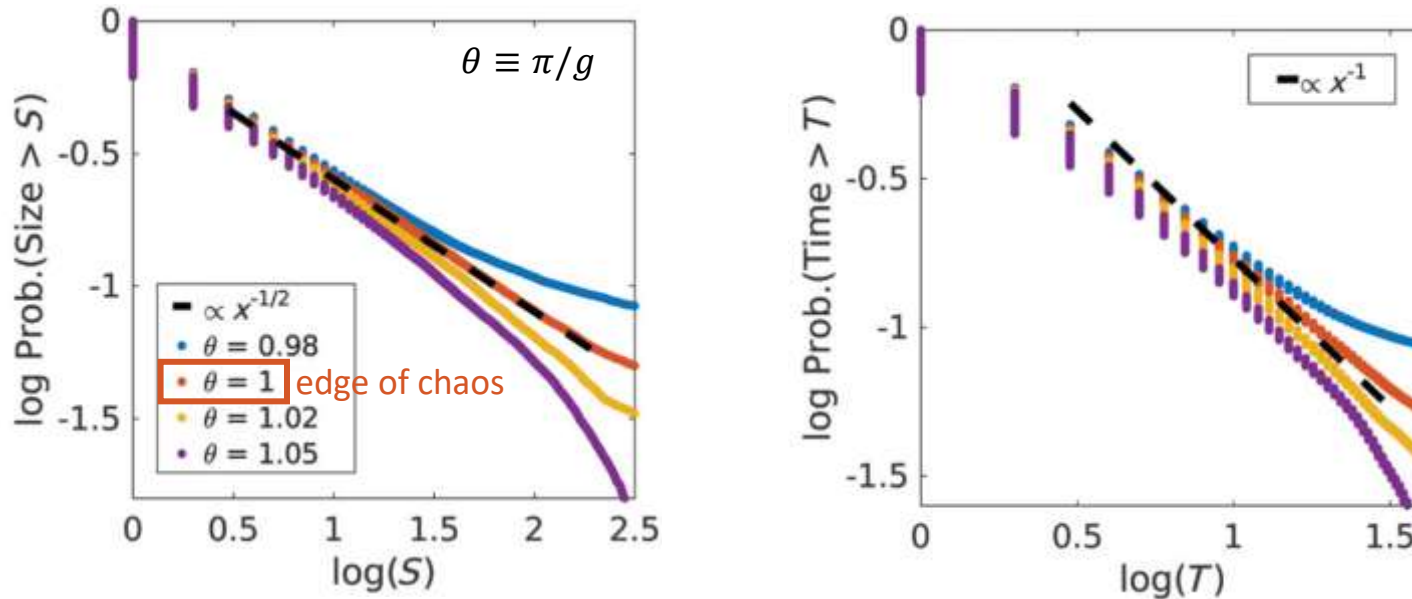
Variability due to chaotic dynamics:



Toyoizumi and Abbott 2011

Neuronal avalanches at the edge of chaos

Power-law distributed avalanche size and duration exhibited by random networks with heavy-tailed synaptic weight distribution:
Neural dynamics $\Rightarrow$ the critical branching process



At the edge of chaos, each neuron has on average one strong enough synaptic weight that can drive the downstream neuron by itself.



What is the role of chaos in brain computation?

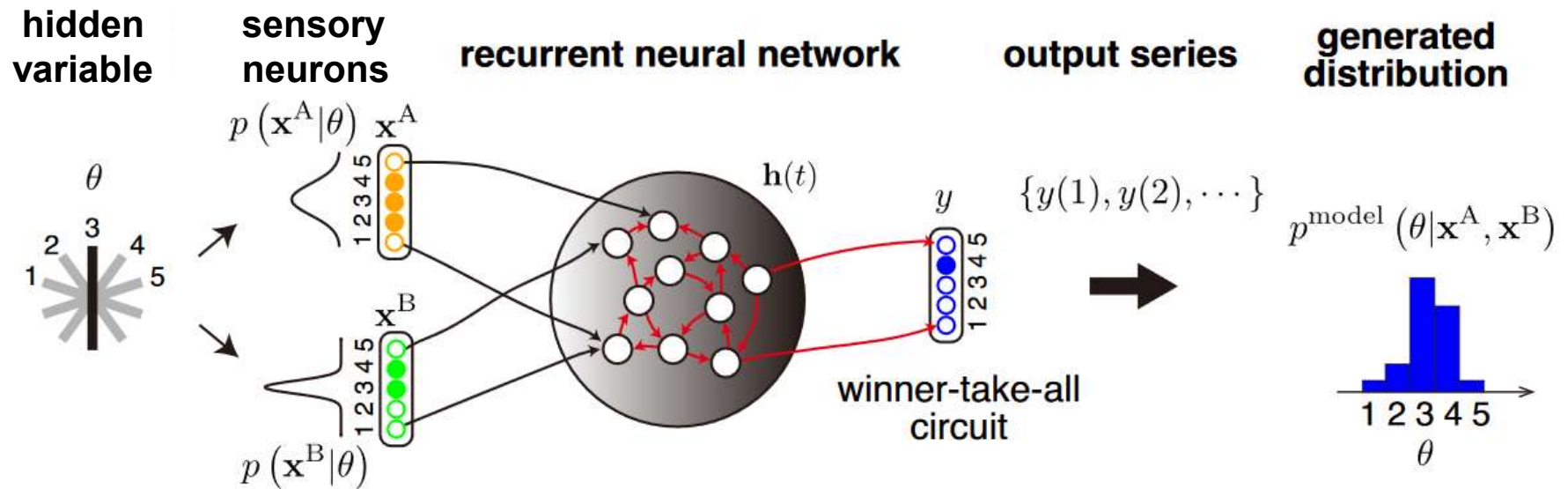
Bayesian inference, which relies on uncertain information, necessitates considering a range of possibilities that underlie observed data.

When exact evaluation of the Bayesian posterior is challenging, random sampling—such as Monte Carlo methods—proves highly effective.

[Hypothesis]

Neural networks in the brain actively generate fluctuations in neural dynamics through chaos, thereby enabling Monte Carlo sampling.

Multimodal sensory integration using a recurrent neural network



Train the weights of the **red** synapses so that the output histogram approximates the posterior distribution of the hidden variable.

Node-perturbation learning

Barto and Anandan 1985

[Model] The network dynamics are deterministic, except during the learning period, when random perturbations are added:

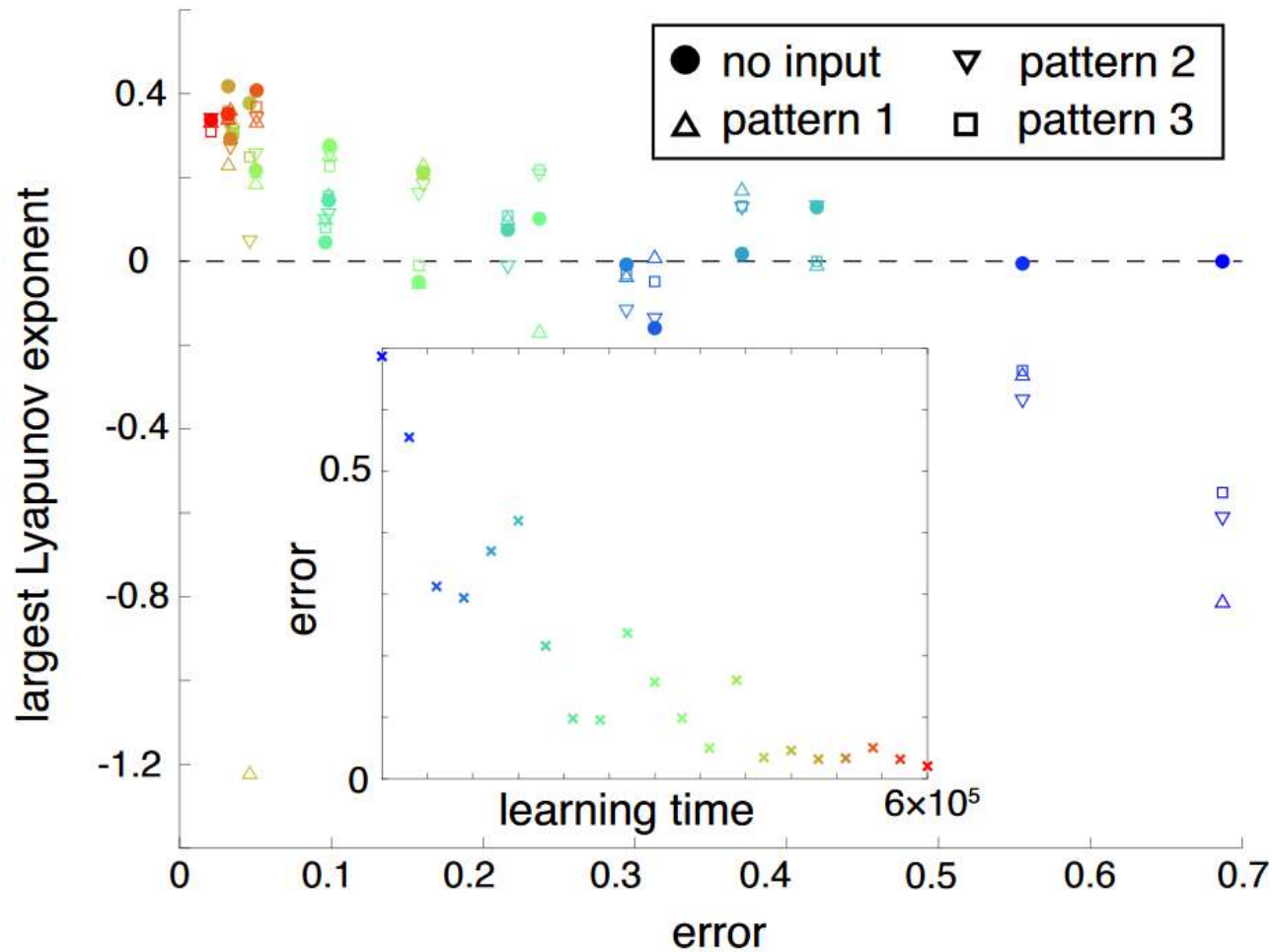
$$\mathbf{x}(t + 1) = \tanh[J\mathbf{x}(t) + \boldsymbol{\xi}(t)]$$

 exploratory perturbation

We compute the gradient by correlating neural perturbations with the resulting error, i.e., the Hellinger distance between the estimated and true posterior distributions.

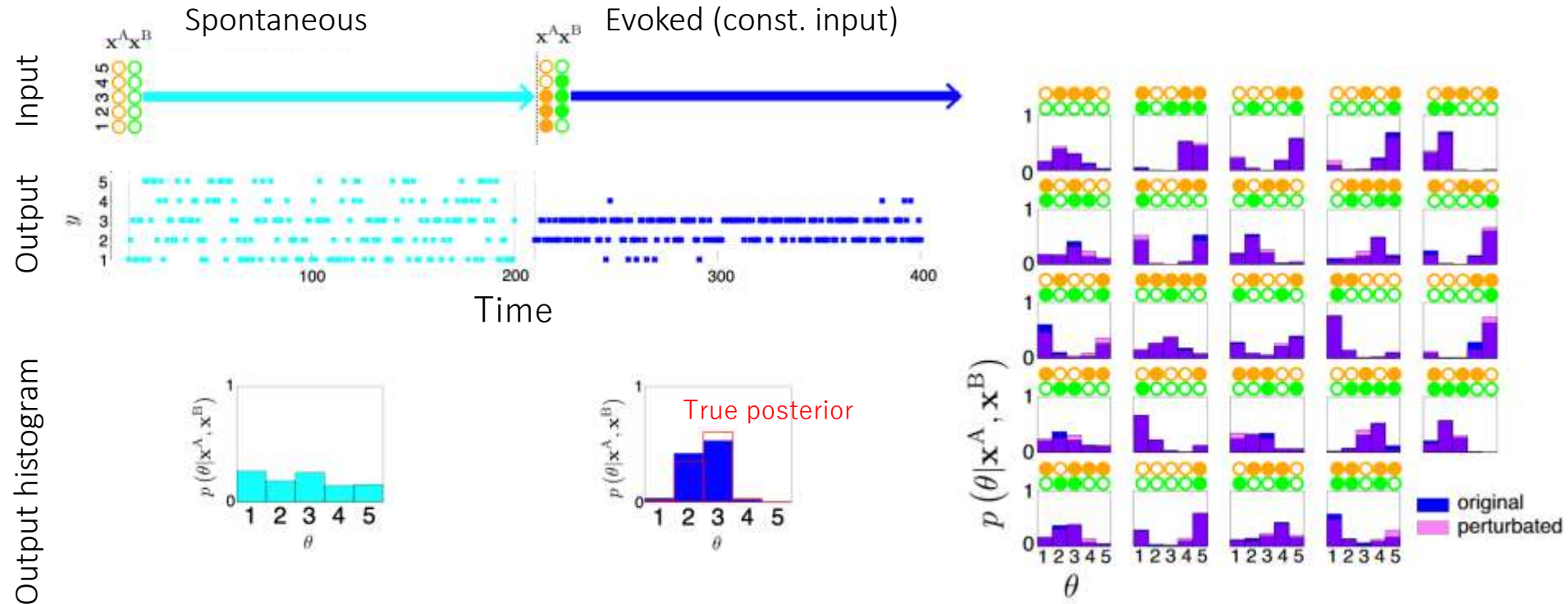
$$\Delta J \propto -\mathcal{L} \boldsymbol{\xi}(t) \mathbf{x}^\top(t)$$

The network develops chaos as learning progresses.
While neural dynamics are sensitive to perturbations, the estimated distribution is robust.

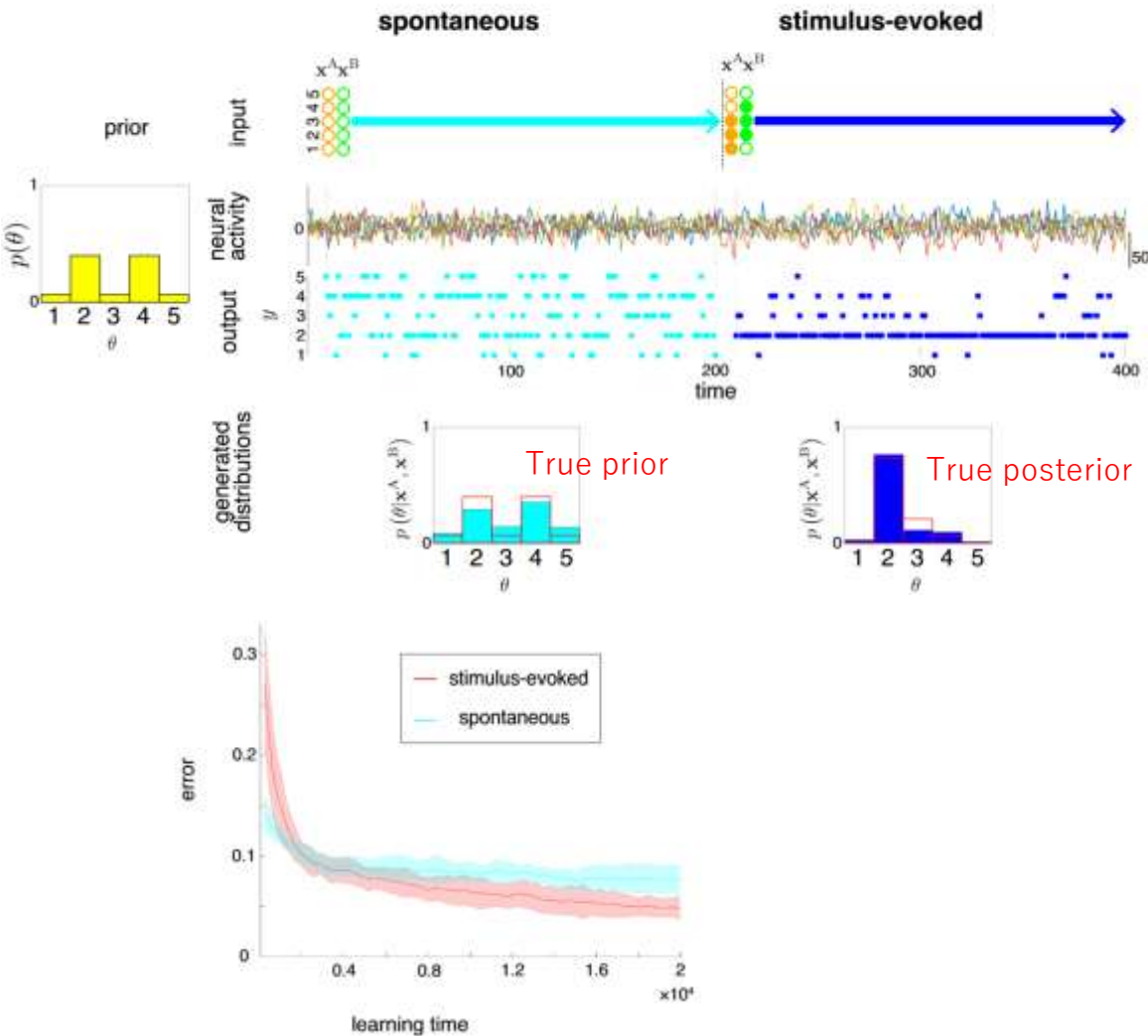


Fluctuations in the recurrent neural network's activity encoded the posterior distribution of the hidden variable based on sensory inputs.

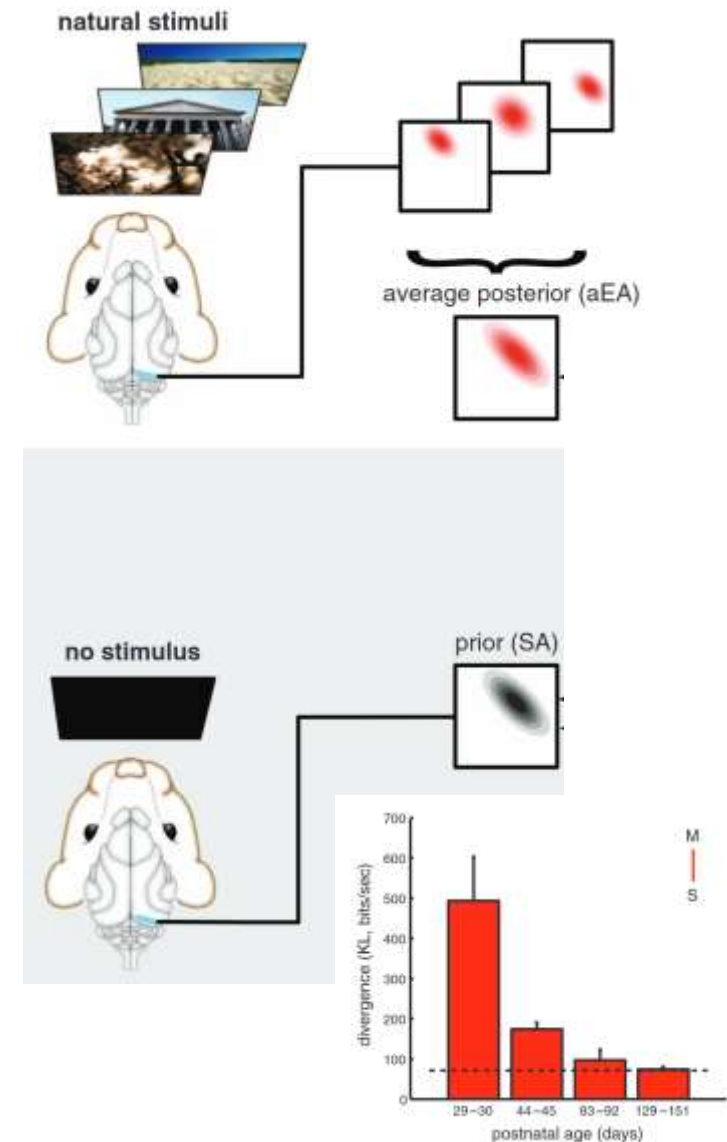
Although neural dynamics were sensitive to perturbations, the estimated distribution remained robust.



After learning, the distribution of spontaneously generated outputs approximates the prior distribution.

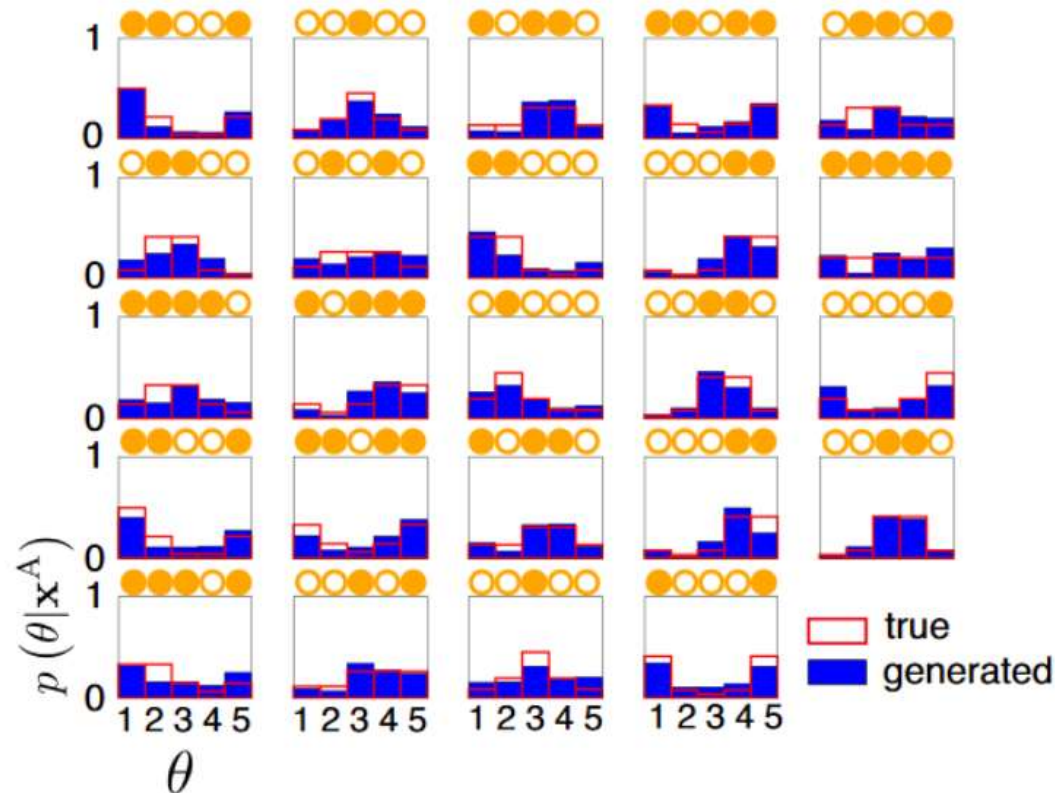


Terada and Toyoizumi 2024



Berkes et al. 2011

More generally, the network samples outputs from the averaged posterior when sensory inputs are only partially given.
(Literal averaging is not needed.)



$$p(\theta | x^A) = E_{x^B | x^A} [p(\theta | x^A, x^B)]$$

Summary

We trained RNNs to sample outputs from the Bayesian posterior distribution.

- The network developed chaotic dynamics to generate samples for probabilistic inference.
- Consistent with experimental results, spontaneous neural activity represented the prior distribution of the variable of interest.
- More generally, the network generated samples from the averaged posterior distribution given partially observed sensory inputs.

Collaborators

- Neural dimensionality reduction
Kensuke Yoshida (RIKEN)
- Dual-pathway associative memory
Louis Kang (RIKEN)
- Chaotic Bayes sampler
Yu Terada (RIKEN)



Funding



Brain/MINDS



Thank you!

toyoizumilab.brain.riken.jp