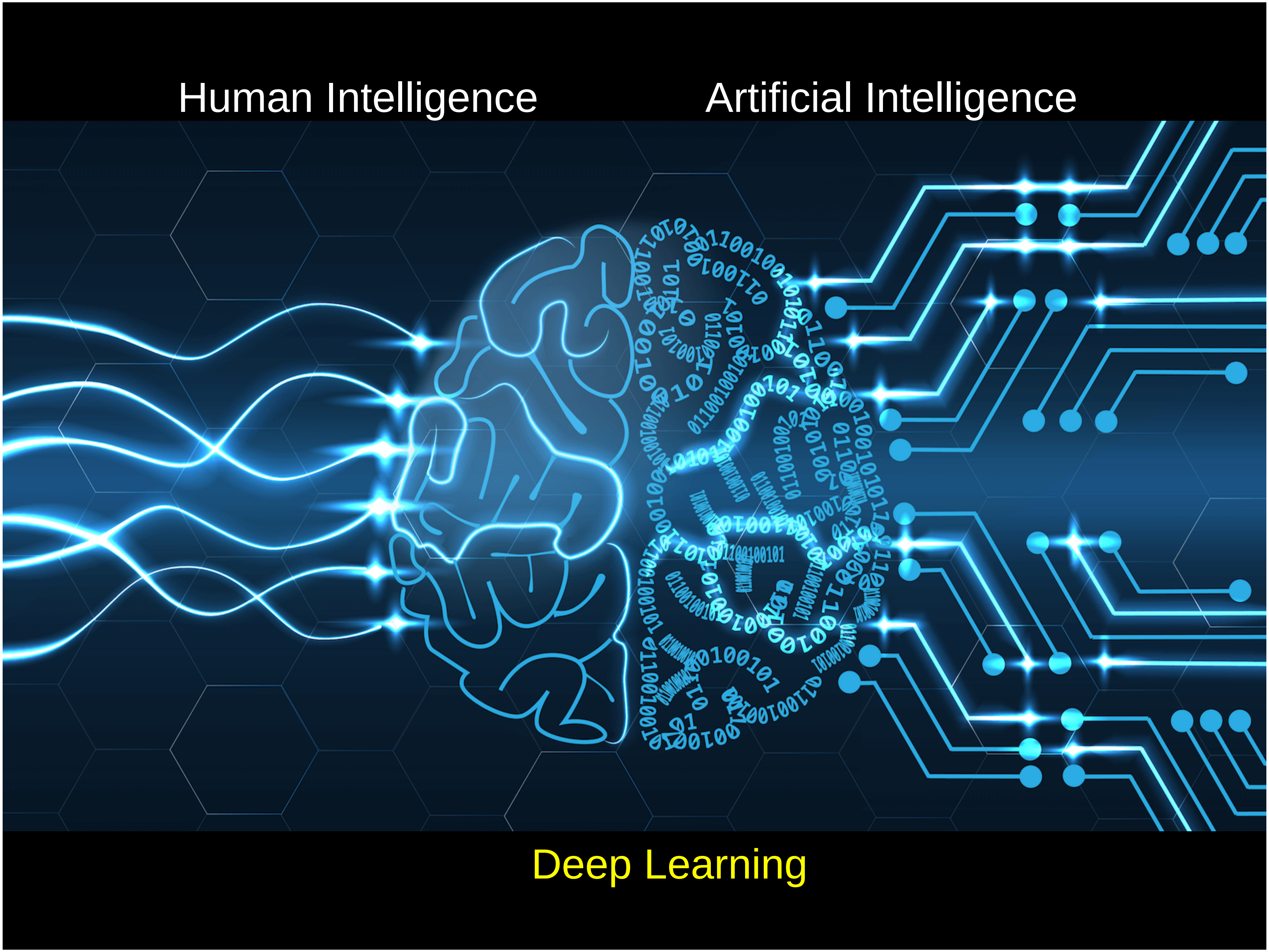
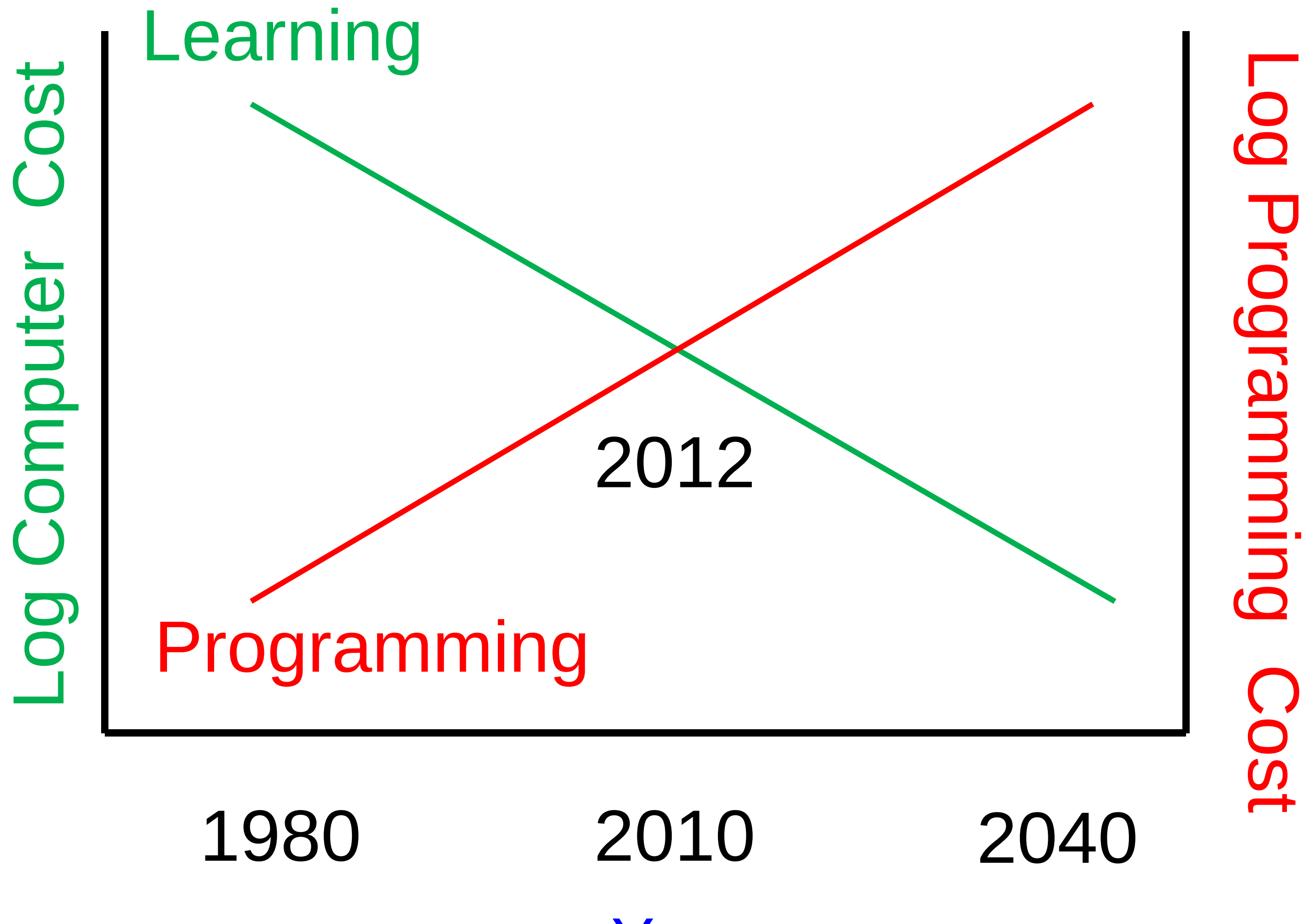
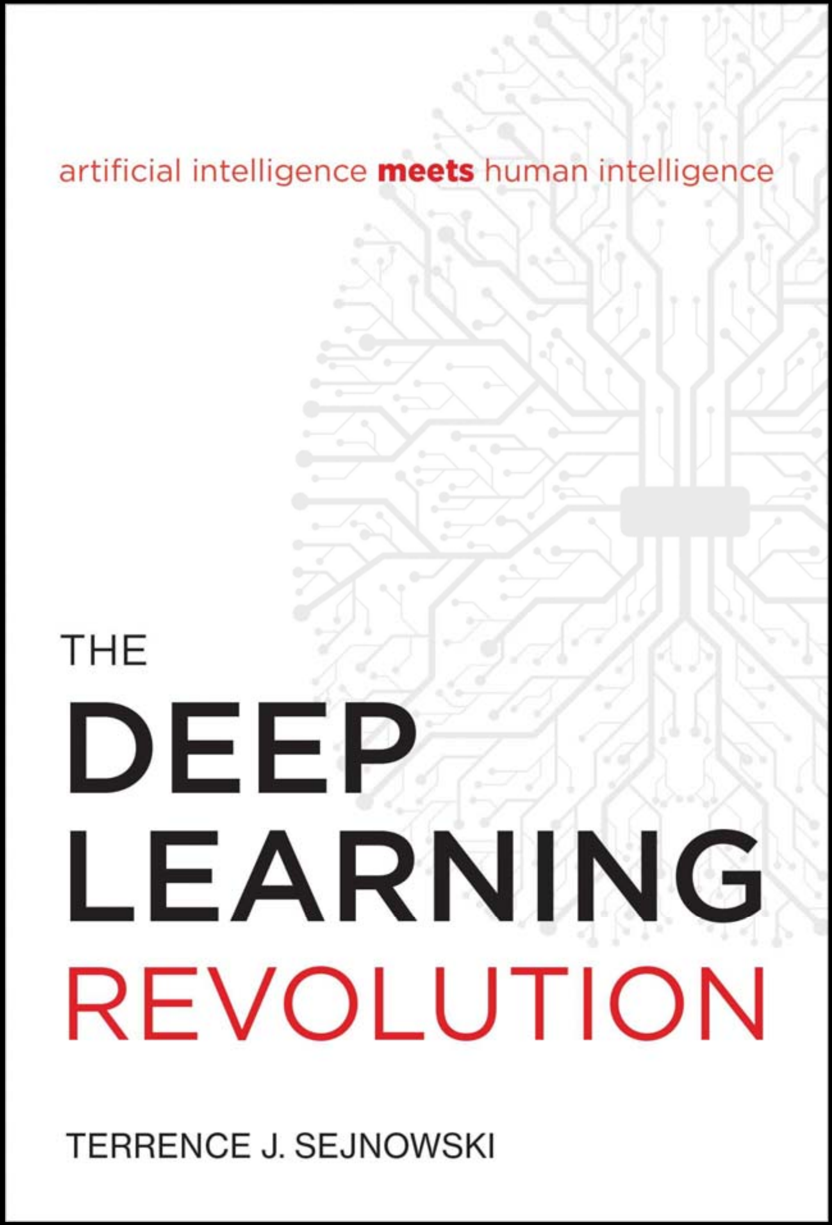

Artificial Intelligence Meets Human Intelligence
Terrence Sejnowski
Salk Institute
UC San Diego


Human Intelligence
Artificial Intelligence
Deep Learning

Tradeoff Between Learning and Programming



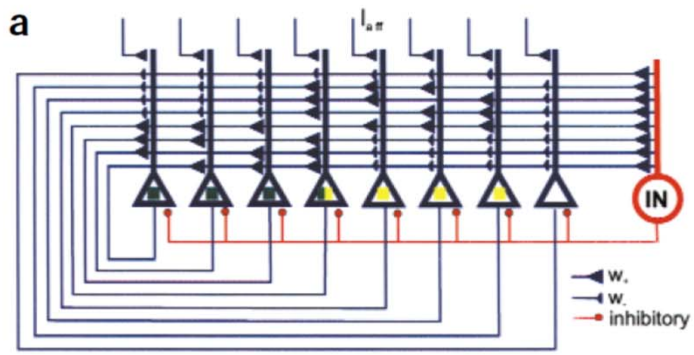


Sejnowski
PNAS, 2020

The Unreasonable Effectiveness of Deep Learning in Artificial Intelligence

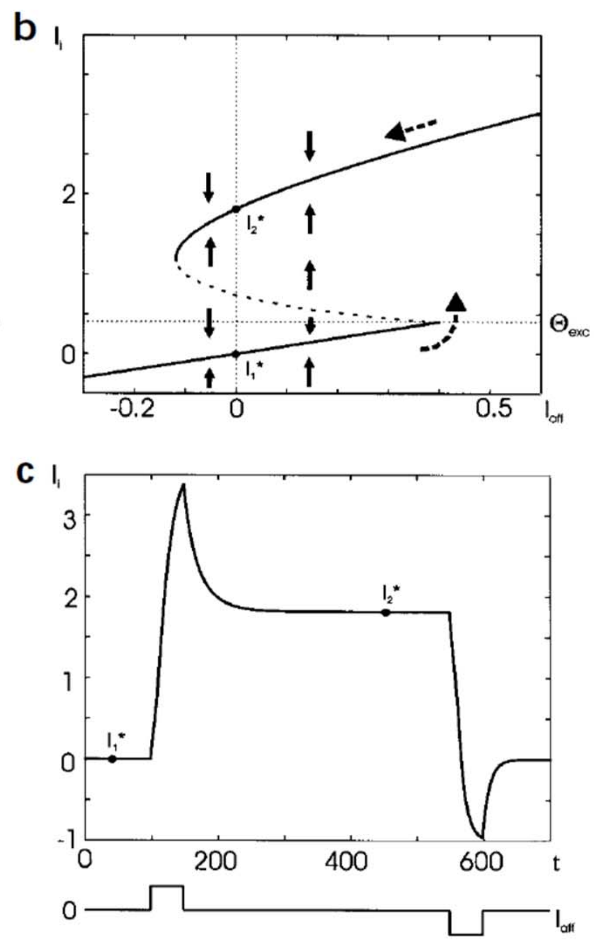
Attractor Model of Working Memory

Fig. 2. A firing rate model[28,30,31,34] of delay-period activity in networks of PFC neurons. (**a**) Structure of the network model. Two patterns ('cell assemblies,' green and yellow boxes) coding for two different objects are embedded in the symmetric synaptic weight matrix. Neurons within the same cell assembly are connected reciprocally by high synaptic weights (w_+), whereas neurons not belonging to the same assembly are connected by low synaptic weights (w_-). Single neurons might participate in more than one cell assembly (green/yellow box; this has also been observed experimentally[40,63]). In addition to these local recurrent excitatory connections, there is a global feedback inhibition (IN) driven by input from the excitatory neurons that allows only one pattern to stay active at a time, and an external afferent input (I_{aff}) to each unit in the network. Model neurons i are described by their total synaptic input current I_i, which evolves in time according to the leaky-integrator differential equation:



$$\tau_I \, dI_i/dt = -I_i + \Sigma\, w_{ij} R(I_j) - G[\Sigma R(I_j)] + I_{aff} \qquad (1)$$

where τ_I is the integration time constant, $w_{ij} \in \{w_-, w_+\}$ is the strength of the synaptic connection from unit j to unit i, $R(I_j)$ is the firing rate of neuron j, G is feedback inhibition that depends on the overall activity level in the network, and I_{aff} is the afferent input. The firing rate of a neuron is assumed to be $R(I_i) = 0$ as long as I_i stays below some firing threshold θ_{exc}, and $R(I_i) = \ln(I_i/\theta_{exc})$ as soon as I_i crosses the threshold. (Other choices for R are possible.) Self-sustaining states (fixed points) of the system are given by the condition that



Durstewitz, D. Seamans, J. K. Sejnowski, T. J. Neurocomputational Models of Working Memory, Nature Neuroscience Supplement, 3, 1184-1191, 2000

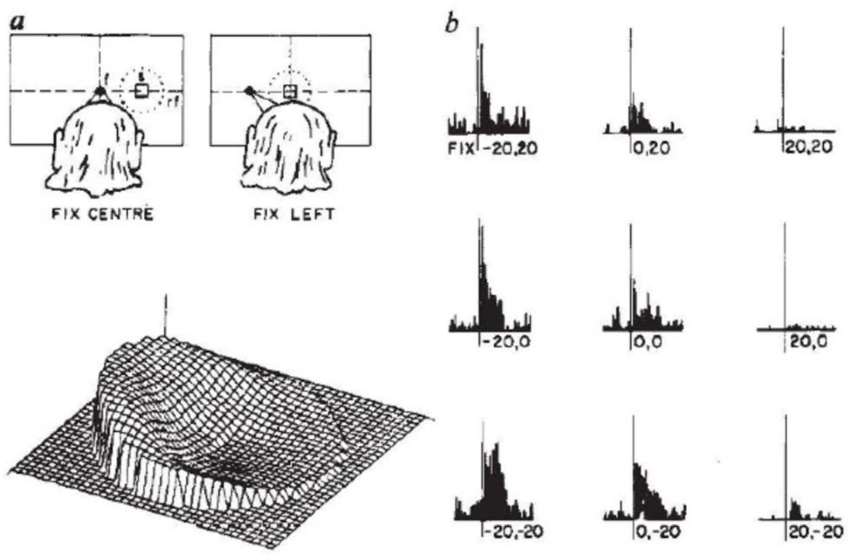
Nature (1988)

A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons

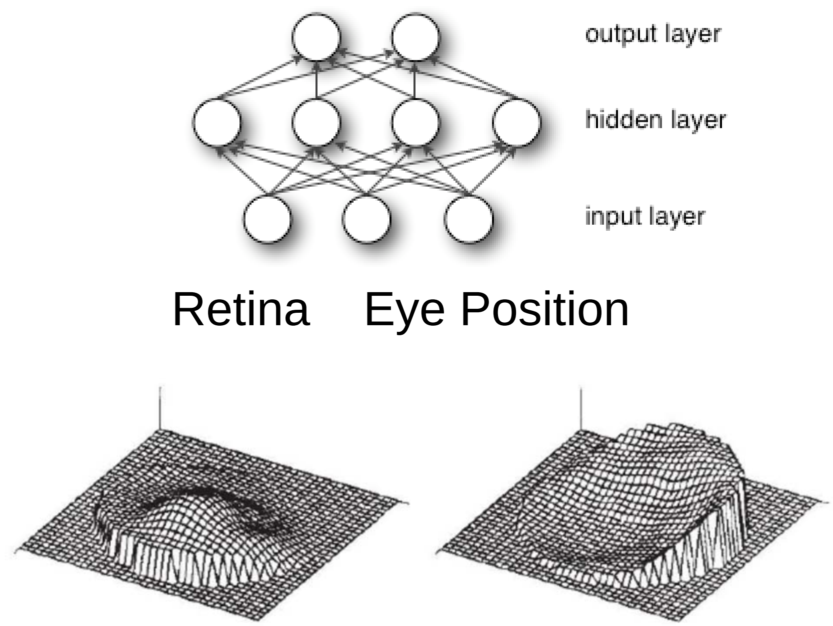
David Zipser* & Richard A. Andersen†

* Institute for Cognitive Science, University of California, San Diego, La Jolla, California 92093, USA
† Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Gain Field

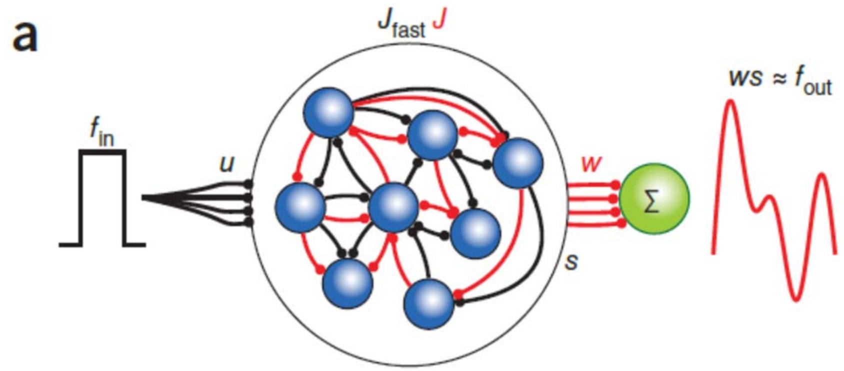


Head Position



Retina Eye Position

Transfer from Rate to Spiking RNNs



Continuous RNN

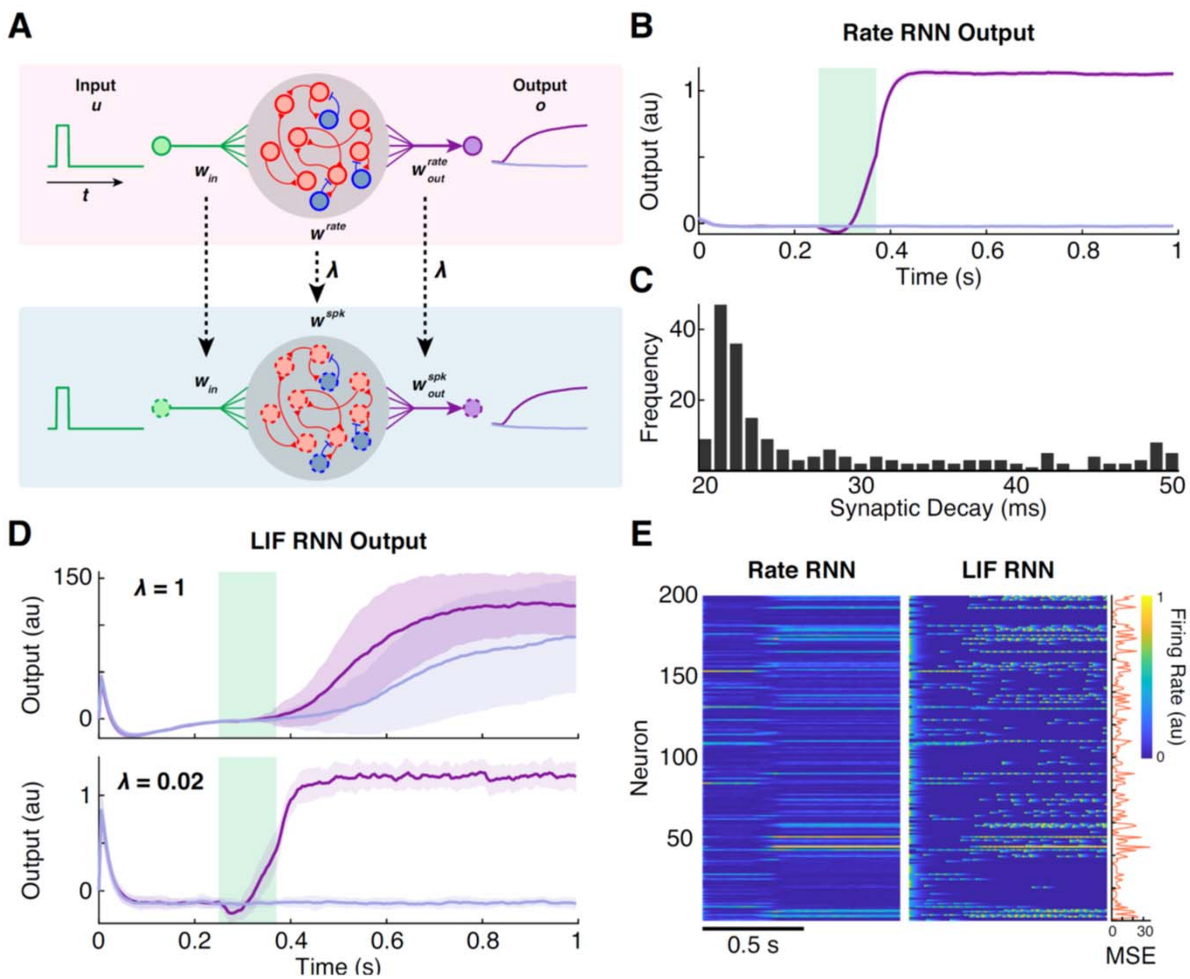
$$\tau_m \frac{dv_i}{dt} = -v_i + (x_i + I_{ext})R$$

LIF RNN

$$\tau_i^s \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} w_{ij} \sum_{t_{jk}<t} \delta(t - t_{jk})$$

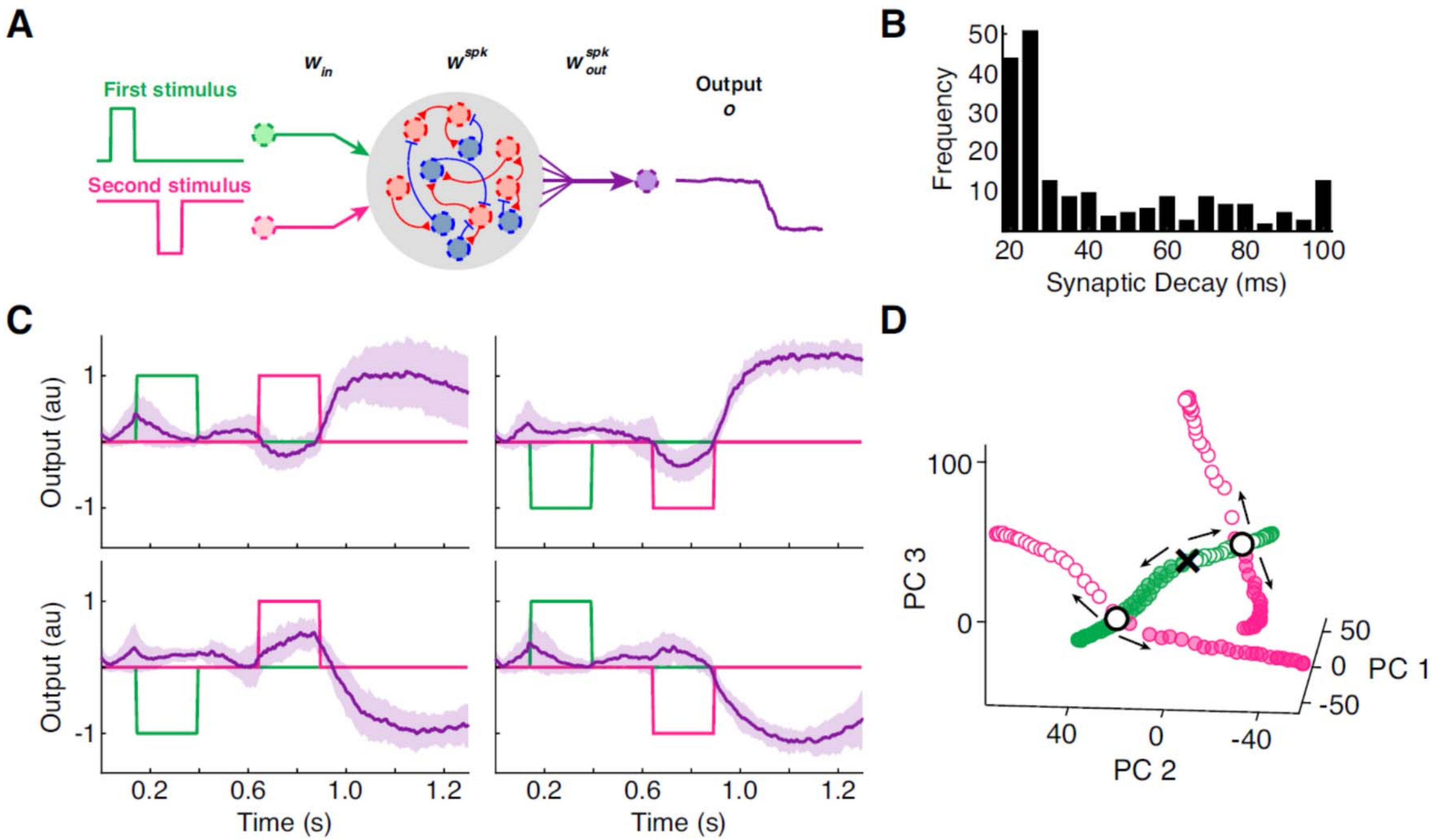
τ_m (membrane time constant; 10 ms), v_i (membrane voltage), x_i (synaptic input current), I_{ext} (external input current), R (leak resistance; 1), τ_i^s (synaptic decay time constant of unit i), w_{ij} (synaptic connection weight from unit j to unit i)

Transfer from Rate to Spiking RNNs



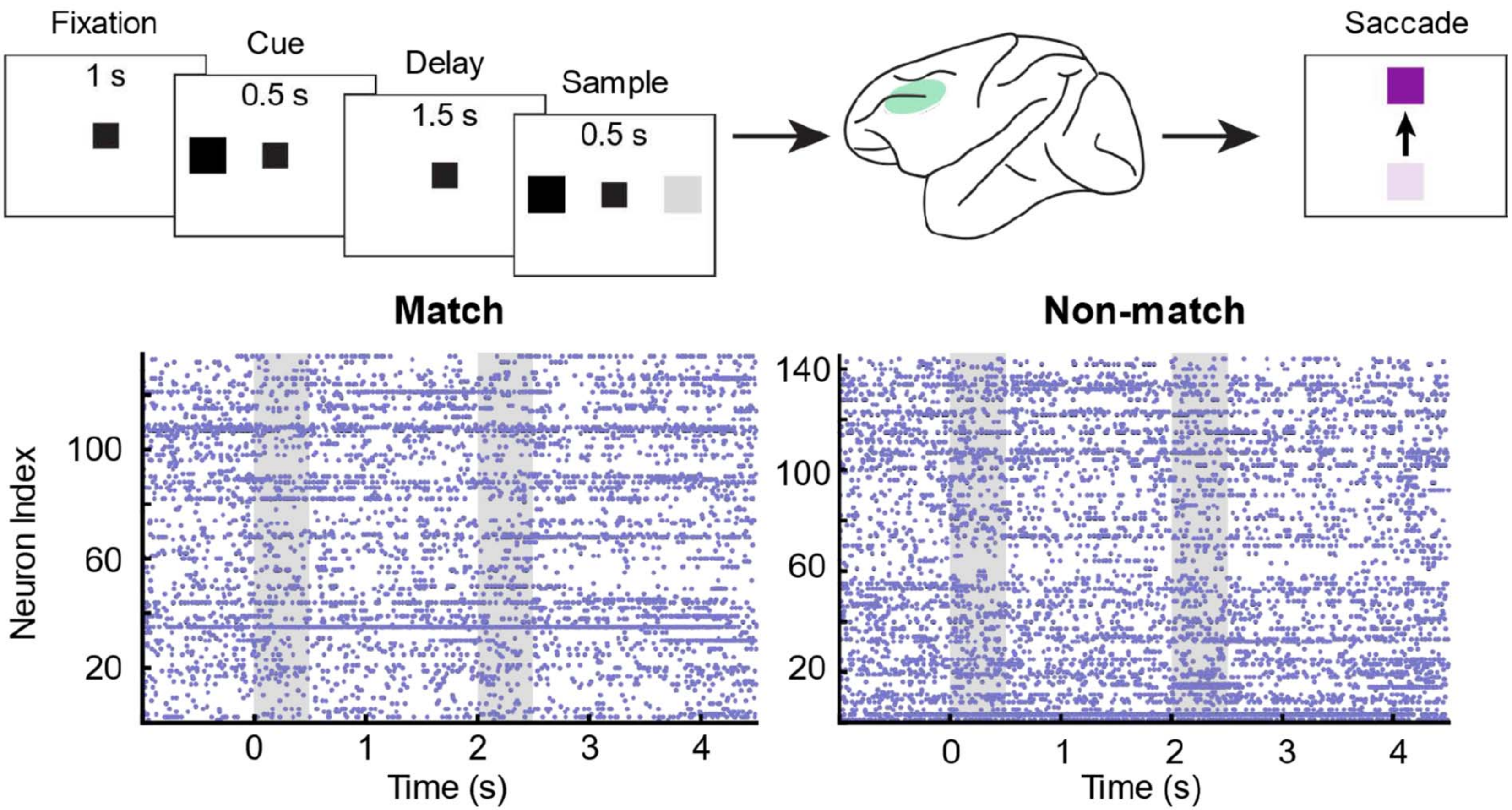
Kim, R., Li, Y. and Sejnowski, T. J. Simple framework for constructing functional spiking recurrent neural networks. PNAS,116: 22811-22820 (2019).

Temporal XOR



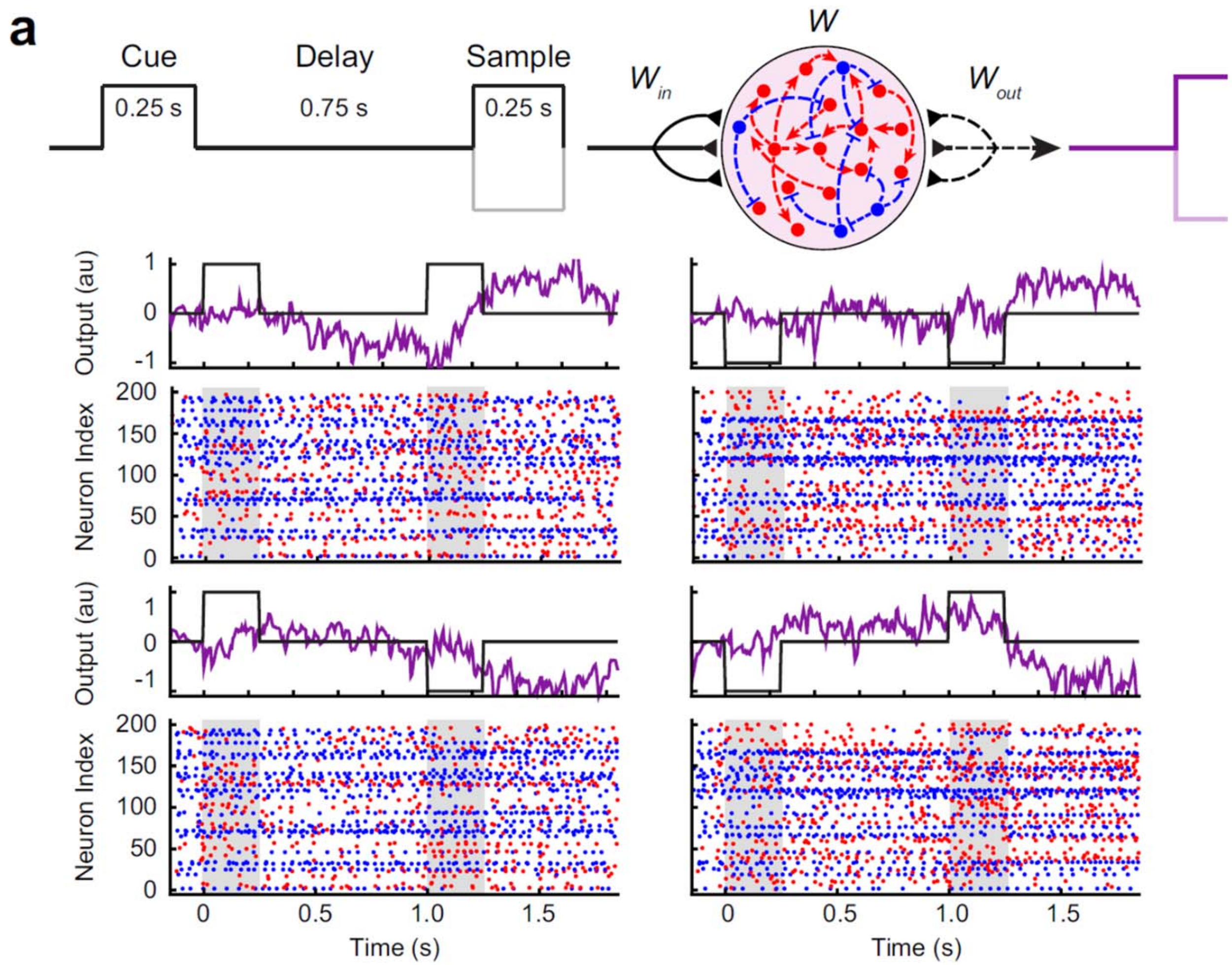
Kim, R., Li, Y. and Sejnowski, T. J. Simple framework for constructing functional Spiking recurrent neural networks. PNAS,116: 22811-22820 (2019).

Delayed Match-to-Sample Task



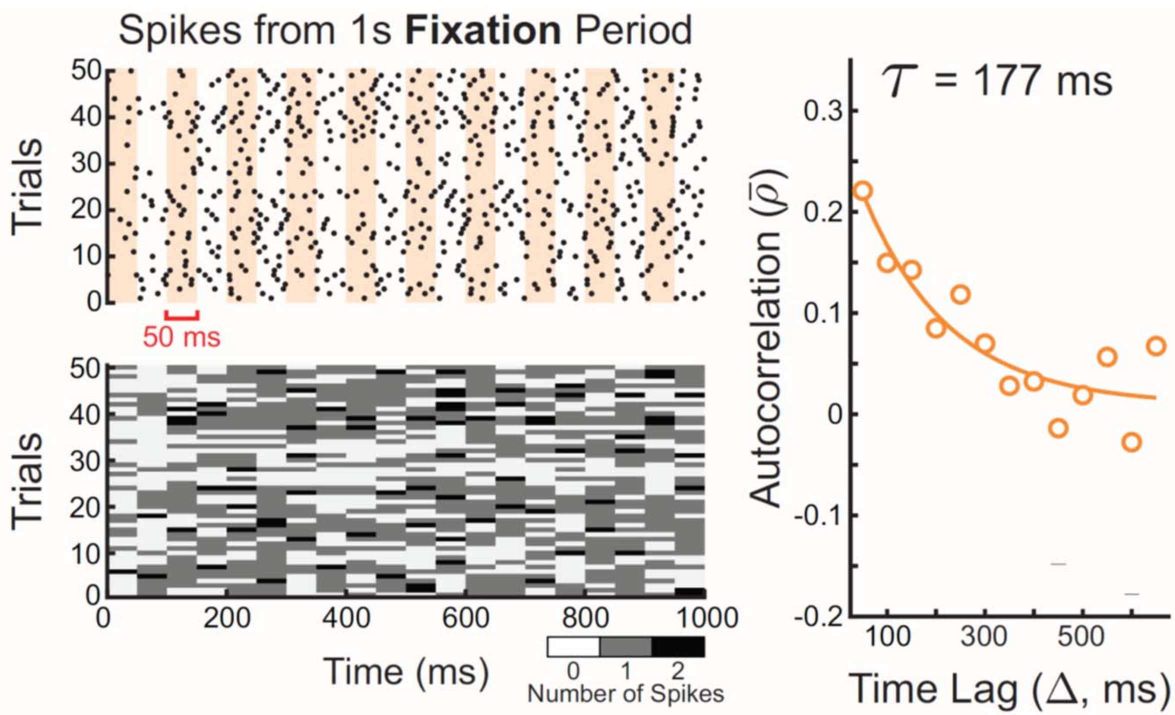
Constantinidis, C., Qi, X.-L. & Meyer, T. Single-neuron spike train recordings from macaque prefrontal cortex during a visual working memory task before and after training. Crcns.org389 (2016).

Delayed Match-to-Sample Task



Kim and Sejnowski, Strong inhibitory signaling underlies stable temporal dynamics and working memory in spiking neural networks. Nature Neuroscience (in press)

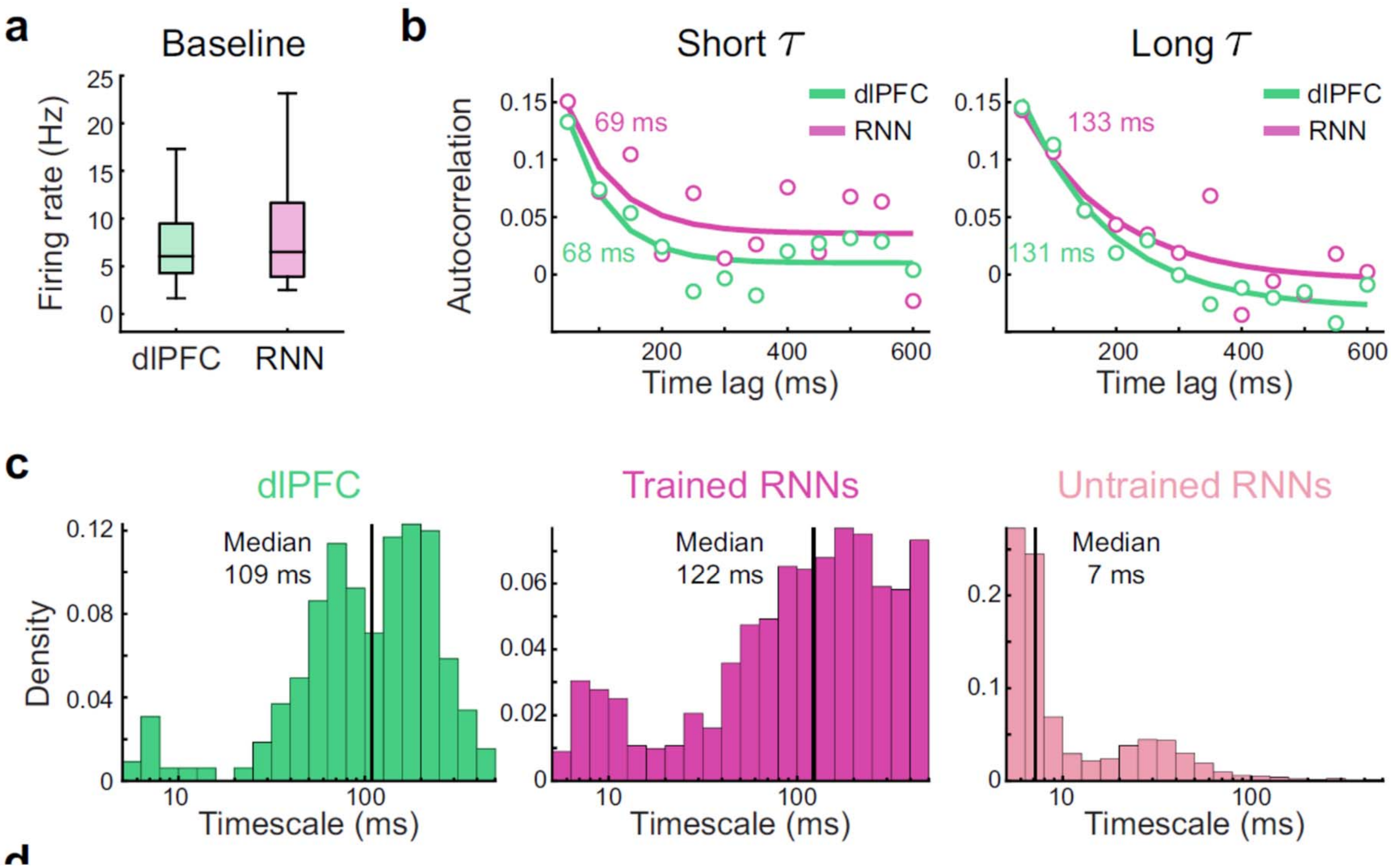
Binned Spike Autocorrelation



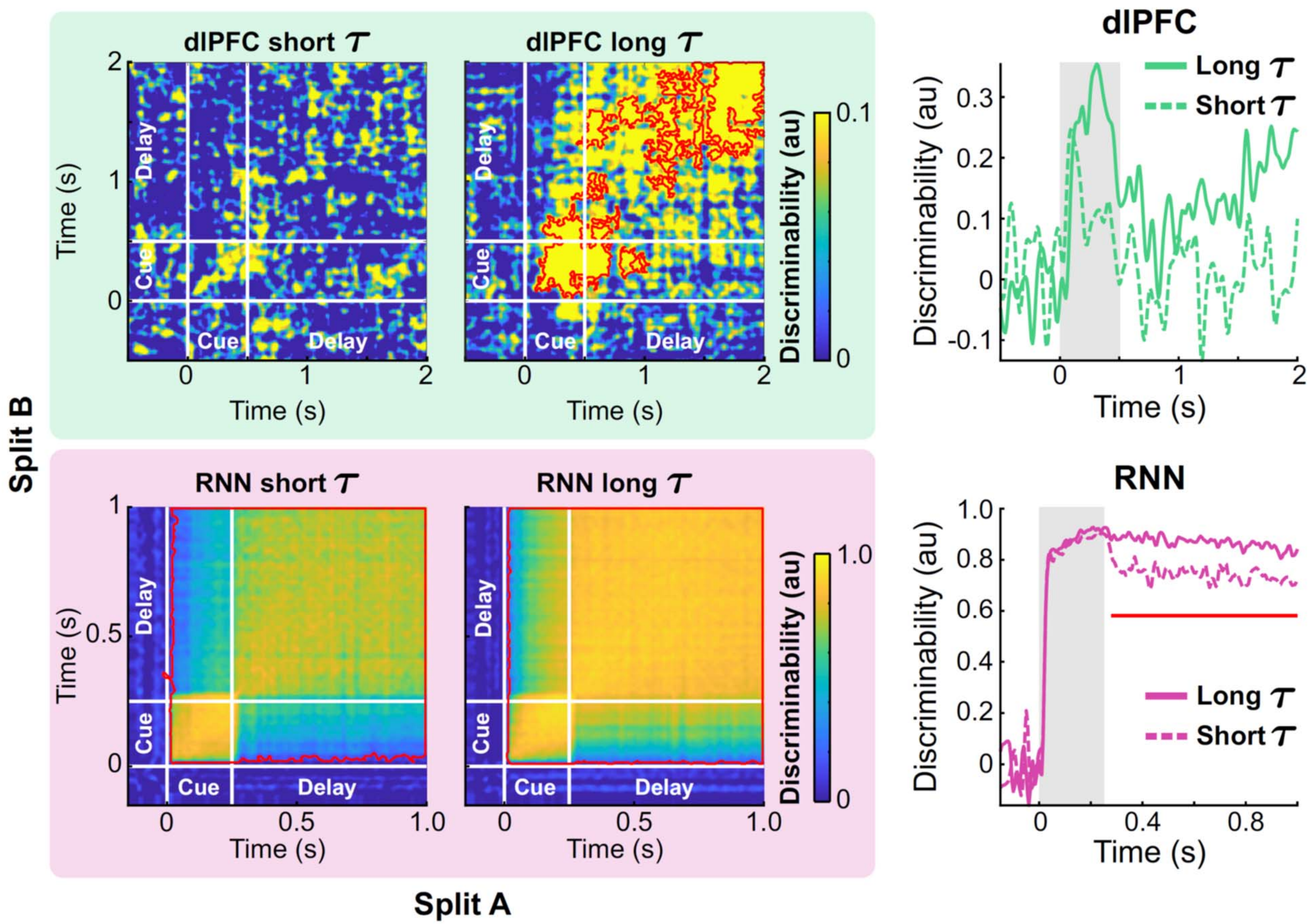
$$\bar{\rho}(\Delta) = A\left(\exp\left(-\frac{\Delta}{\tau}\right) + B\right)$$

A (amplitude) and B (offset) of the exponential decay fit

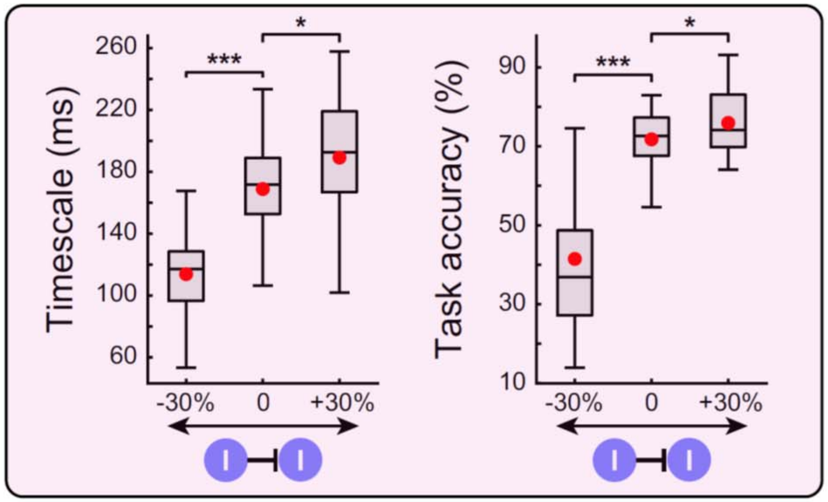
Comparing PFC Neurons to RNN Units

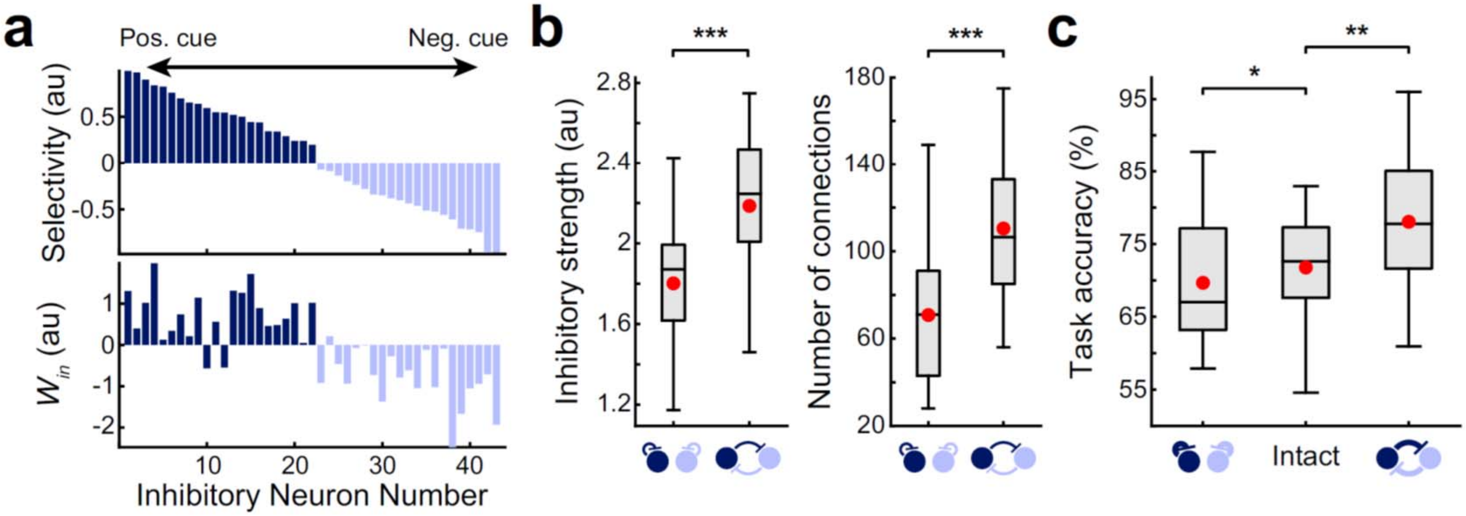


Comparing PFC and RNN Discriminability

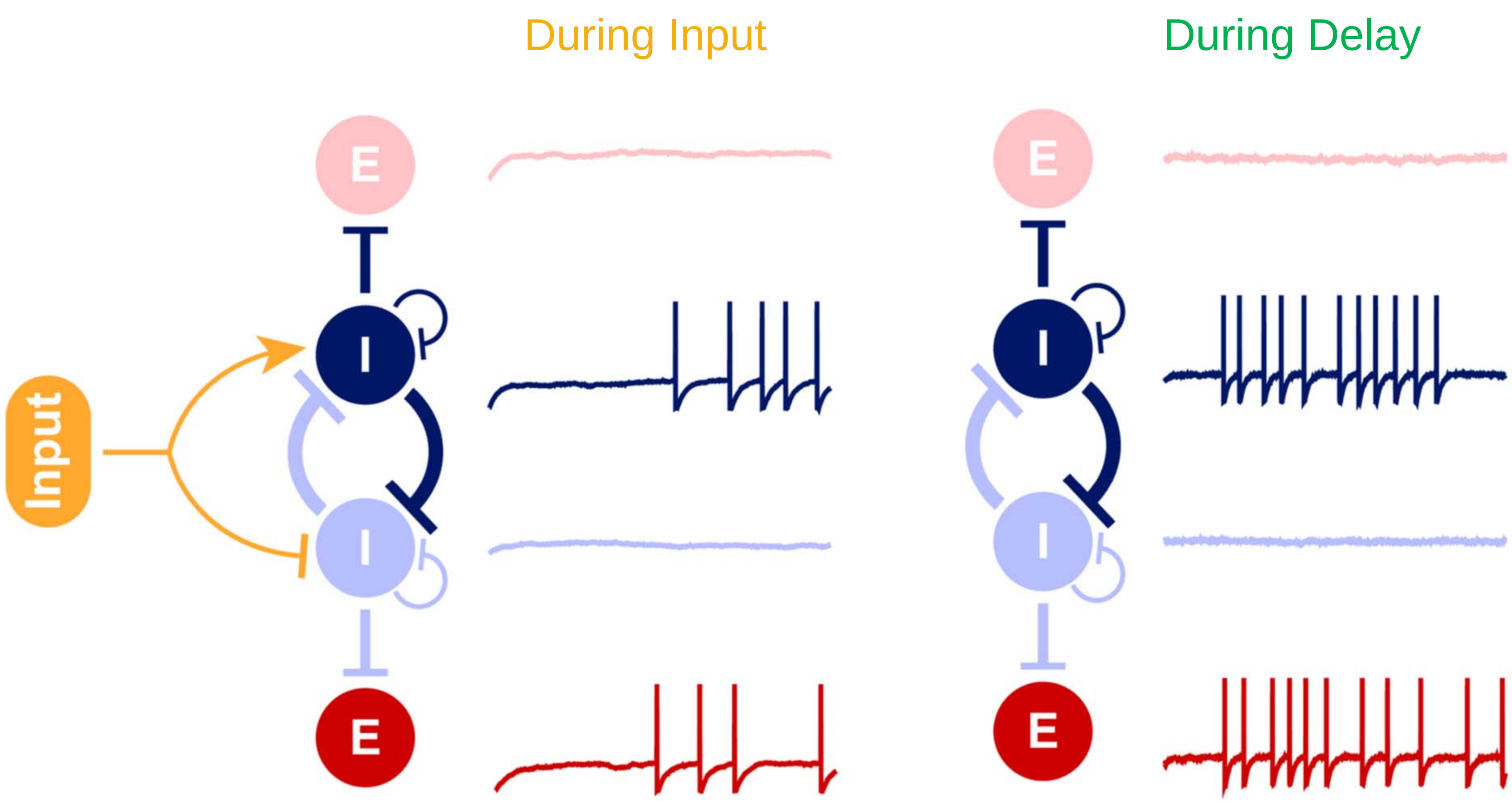


Inhibitory Subgroups Key to Decision?

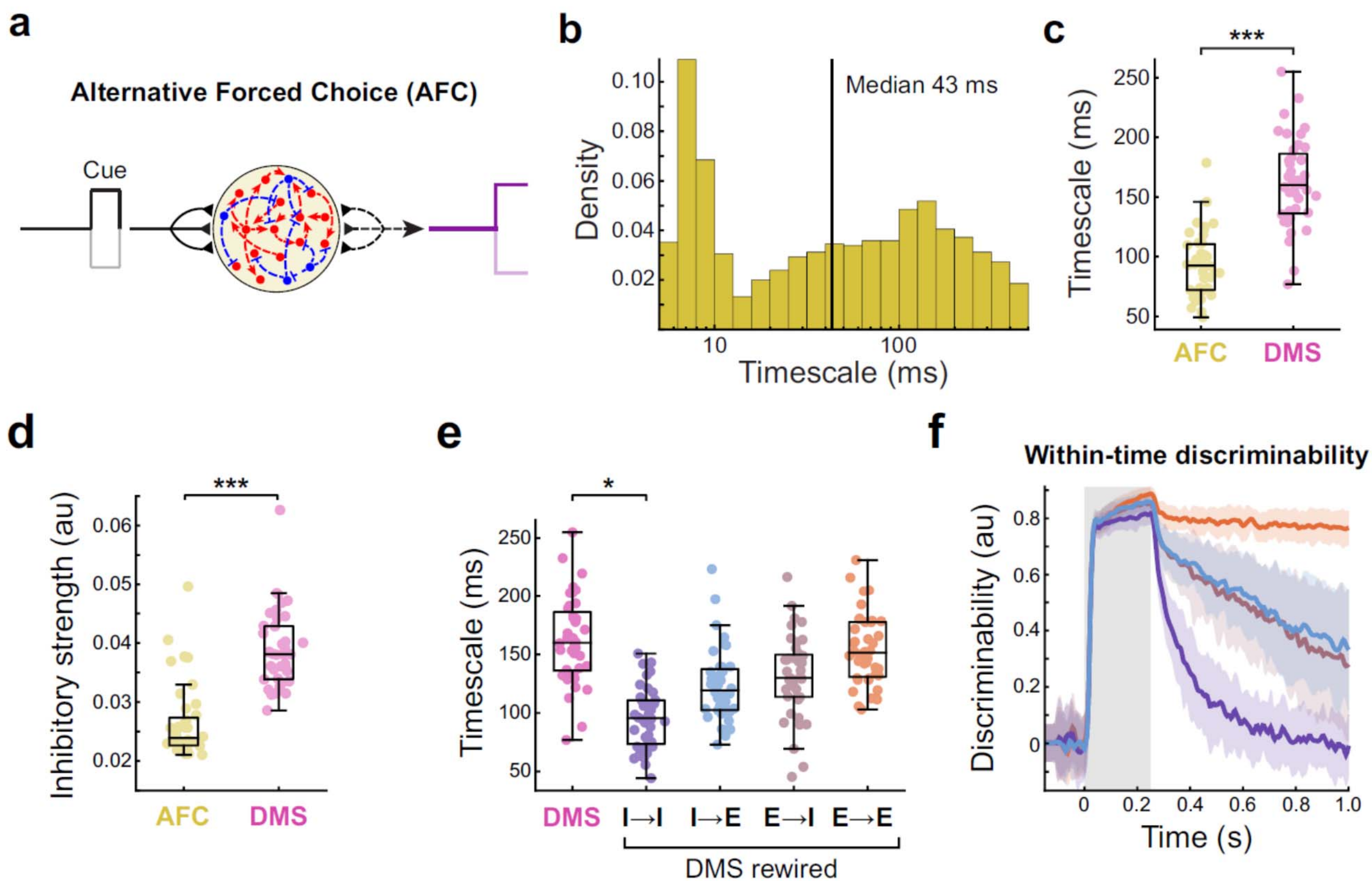




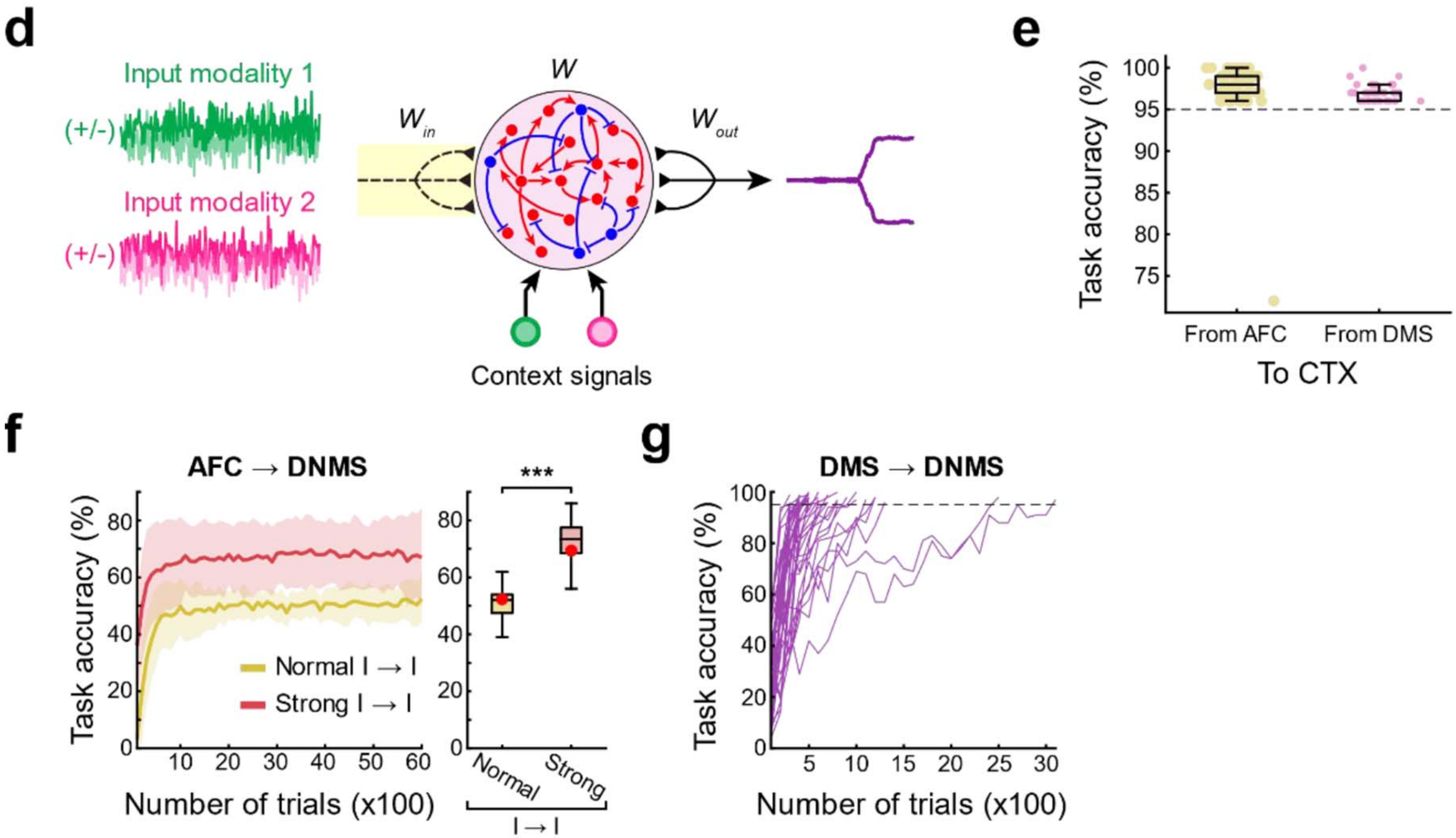
Disinhibition Model of Working Memory



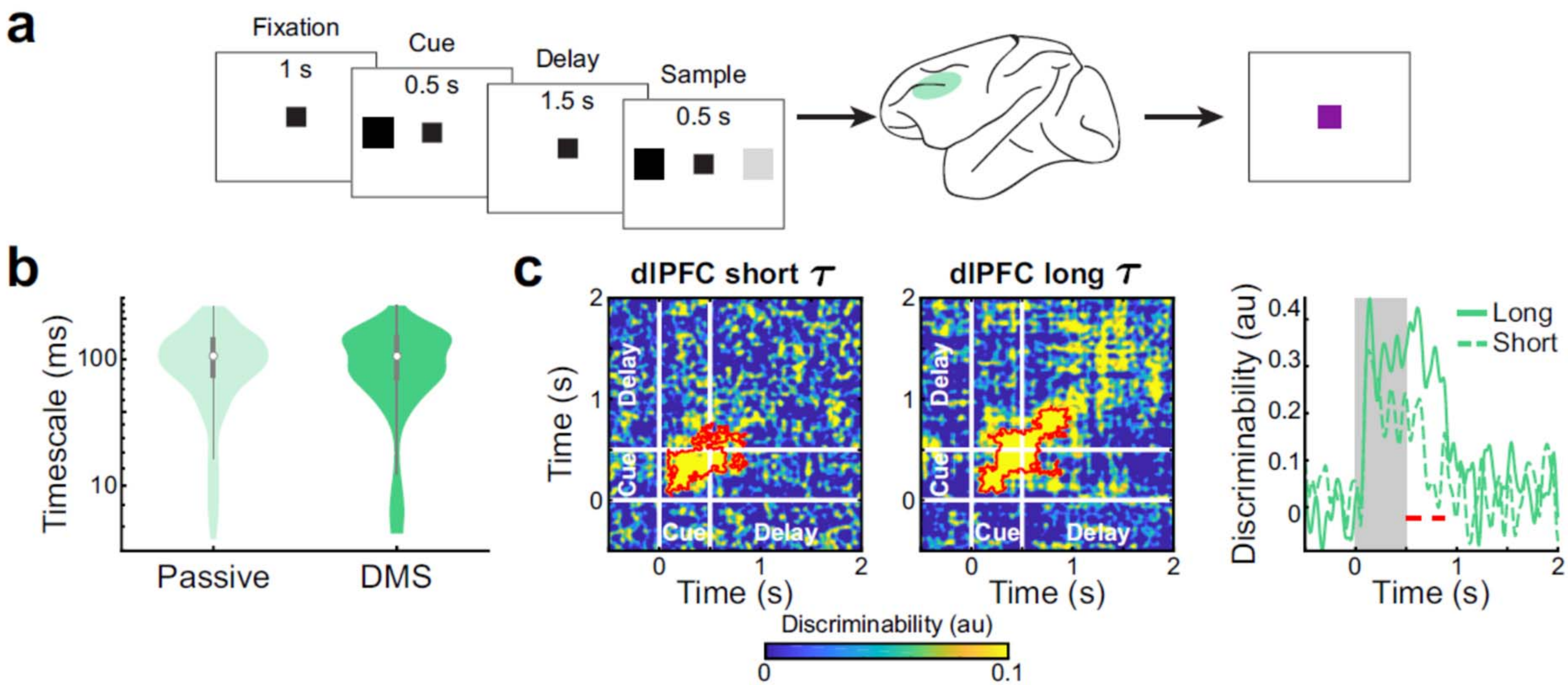
Two-Alternative Forced Choice (AFC) Task



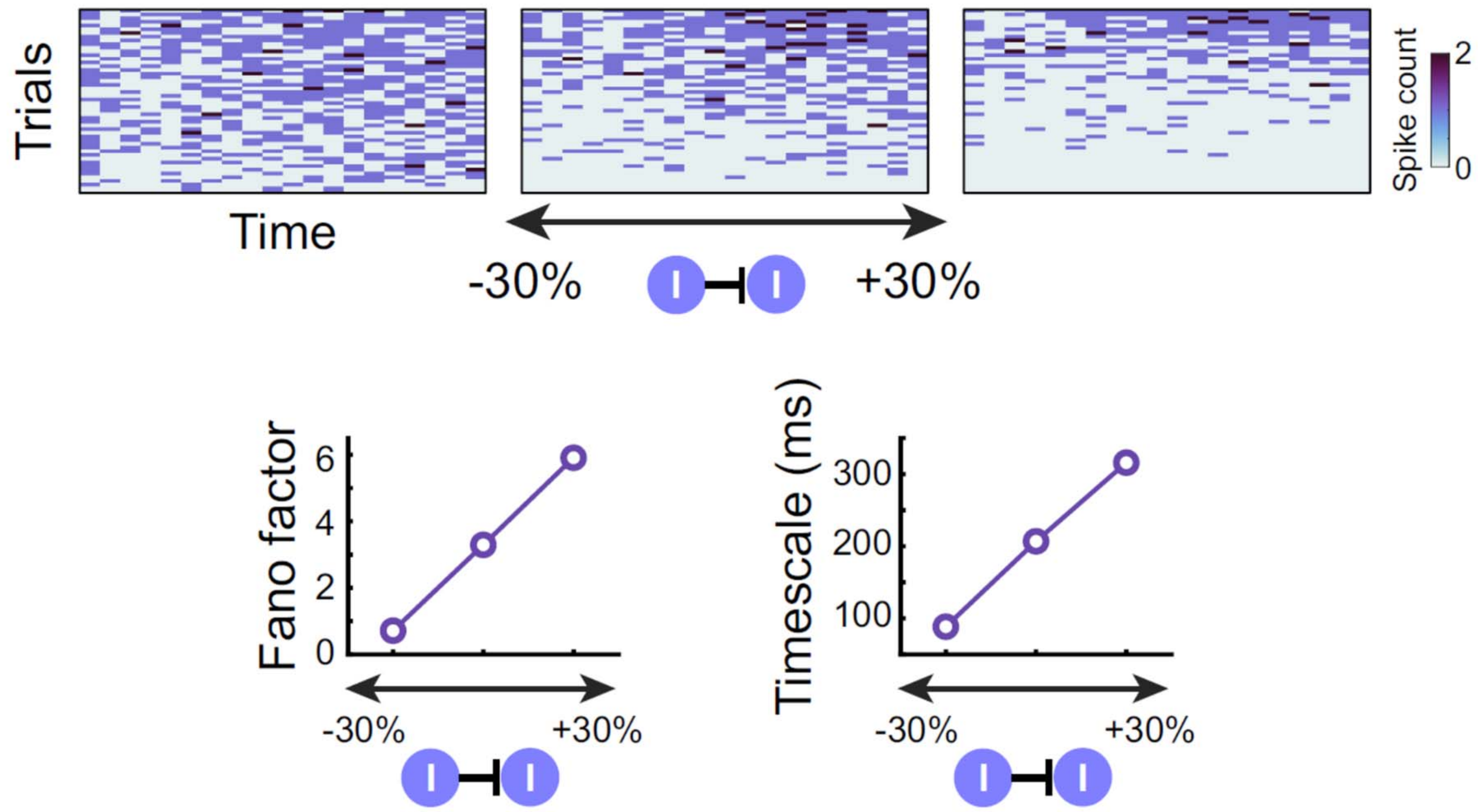
RNN Cross Training



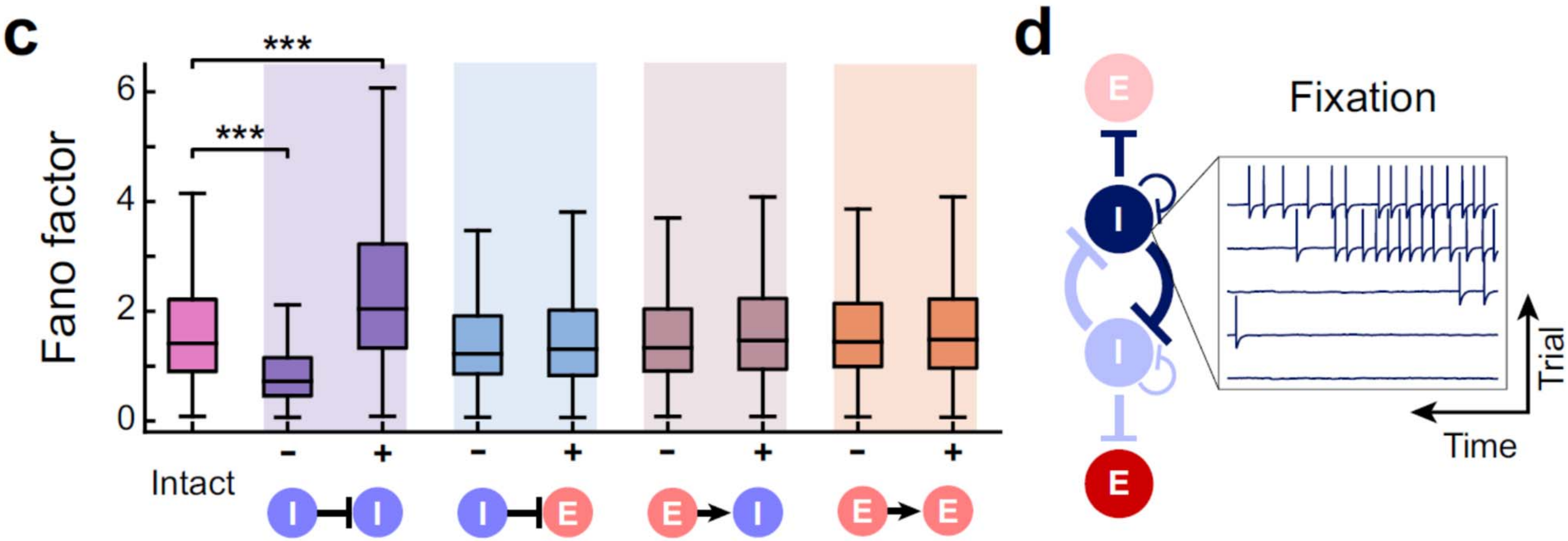
PFC Timescales for a Passive Task



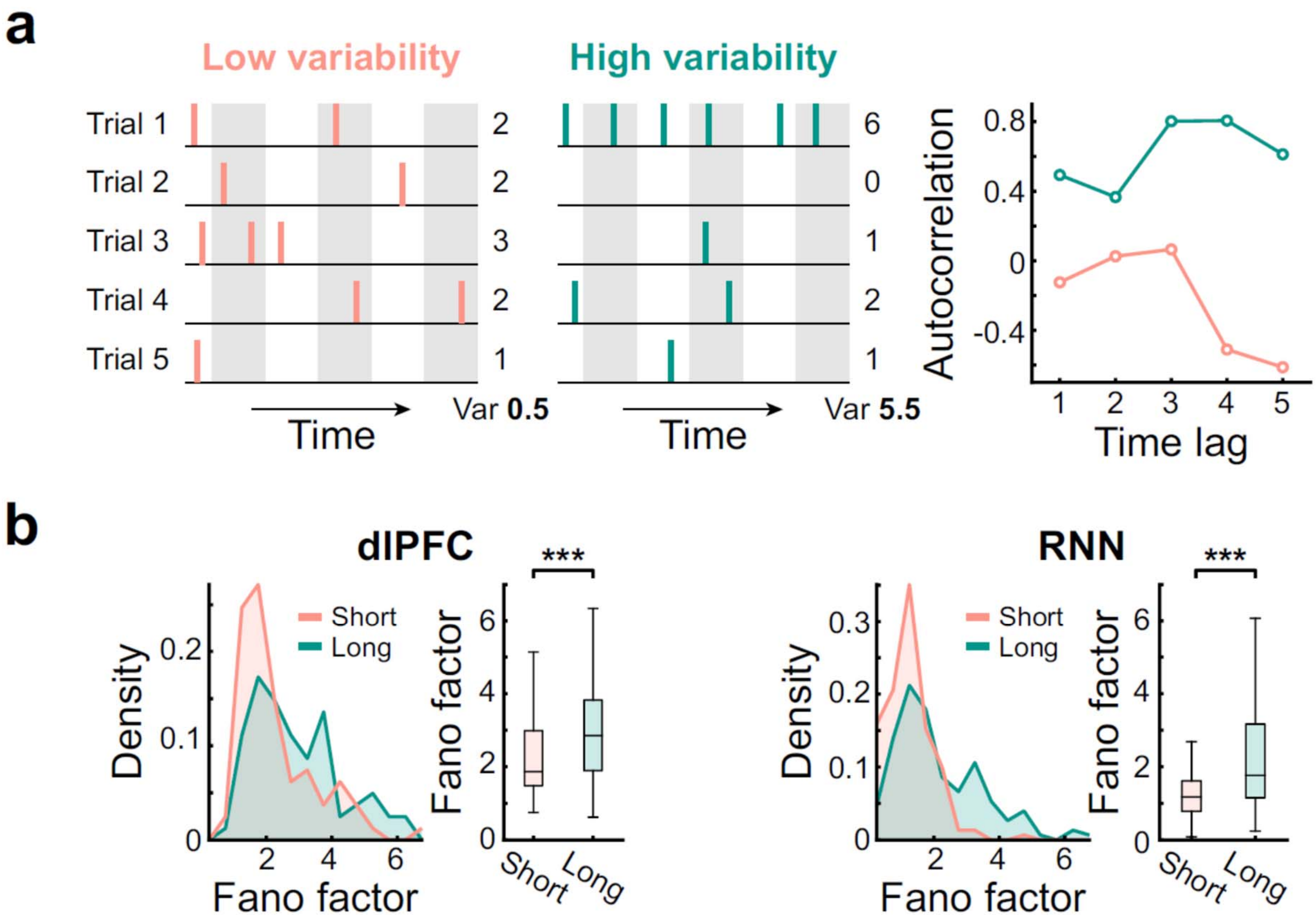
Fano Factor Increases with Strength of Inhibition



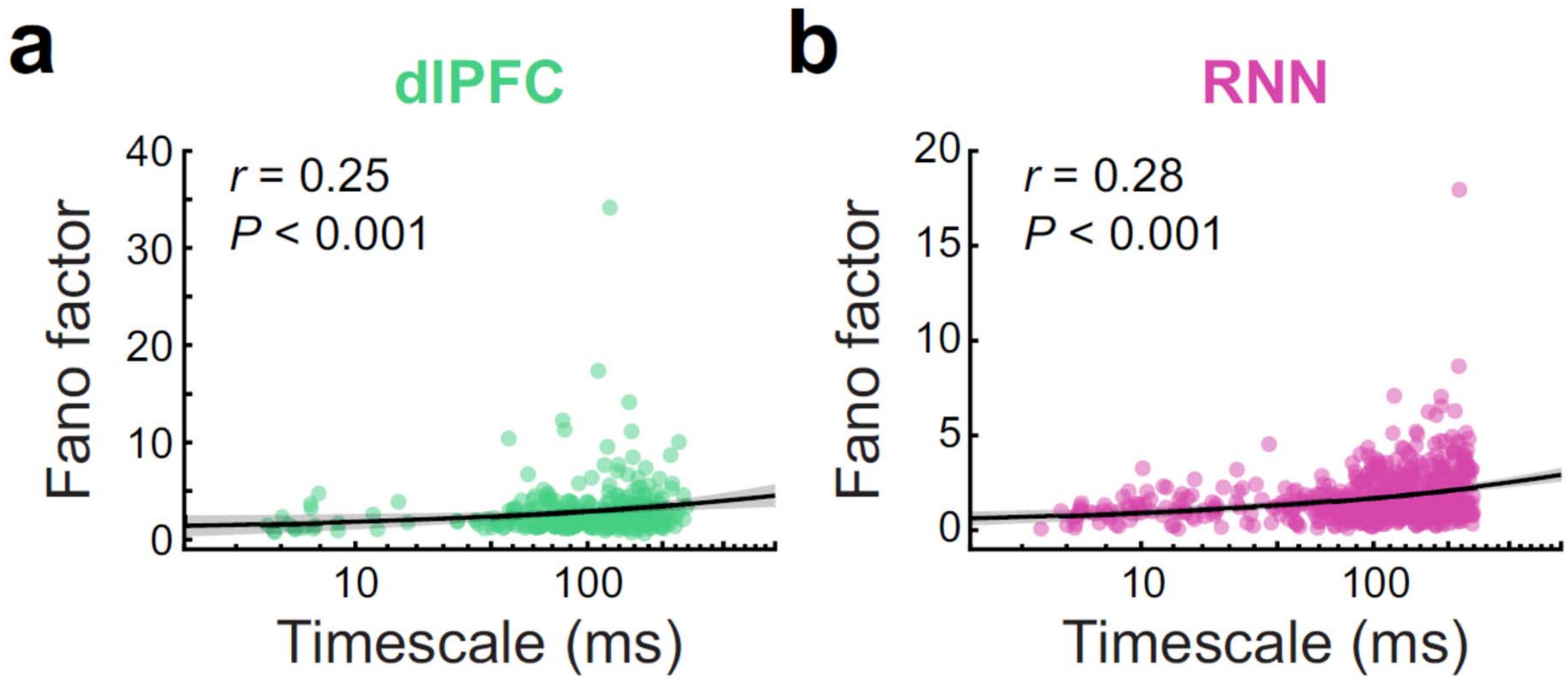
Variability Is Driven by Inhibition



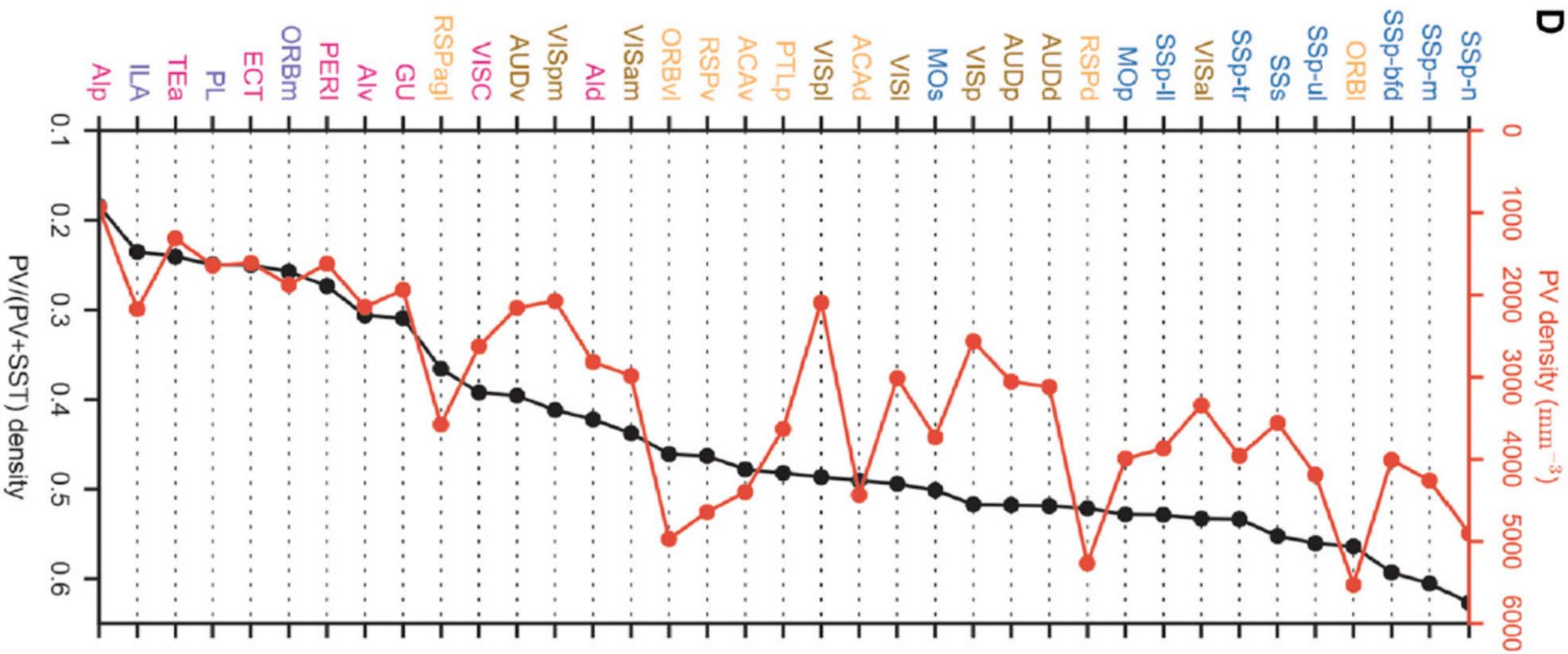
Trial-to-Trial Variability and Autocorrelation



Fano Factor Increases with Timescale



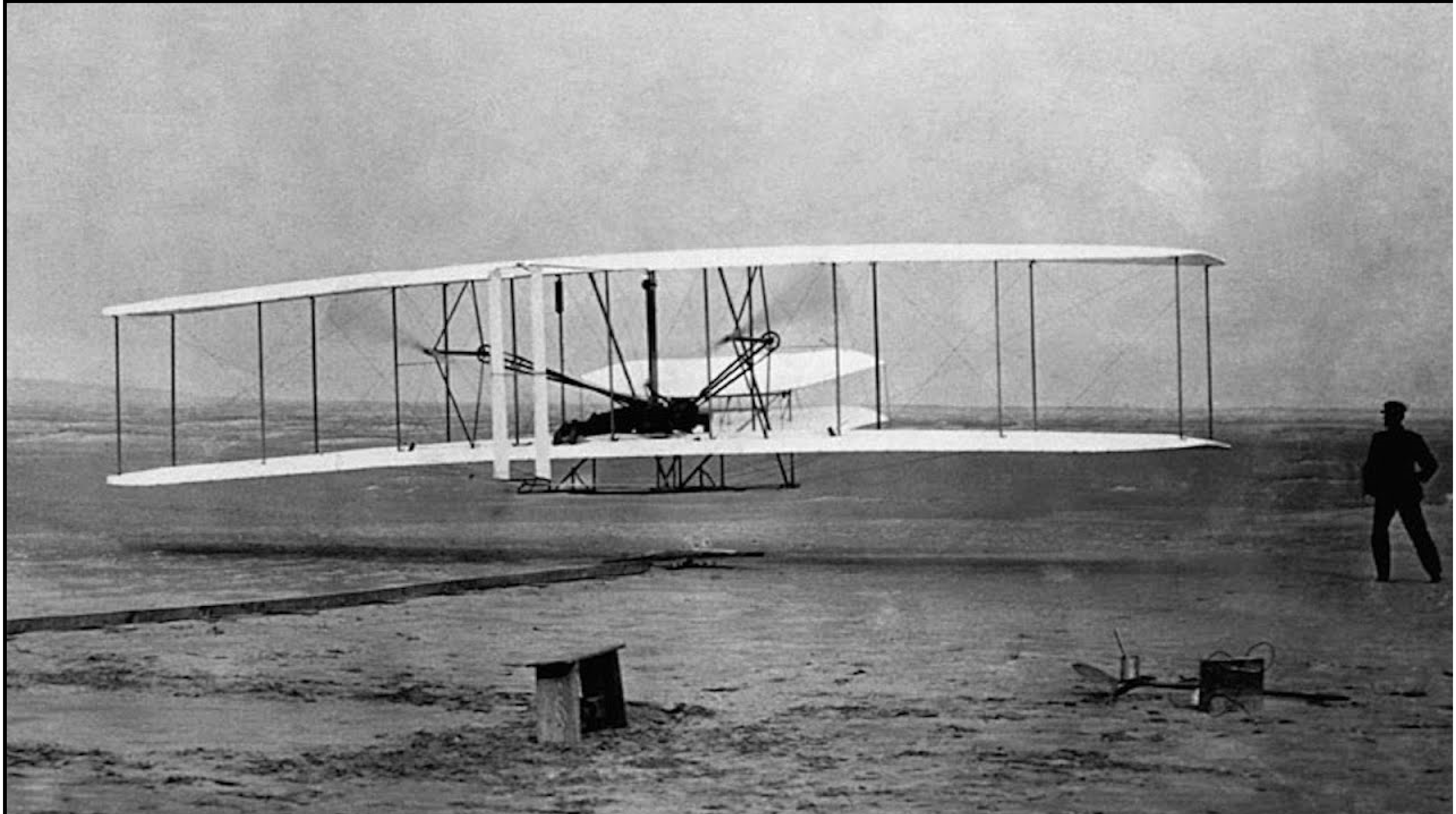
PV and SST Interneurons in Hierarchical Gradients



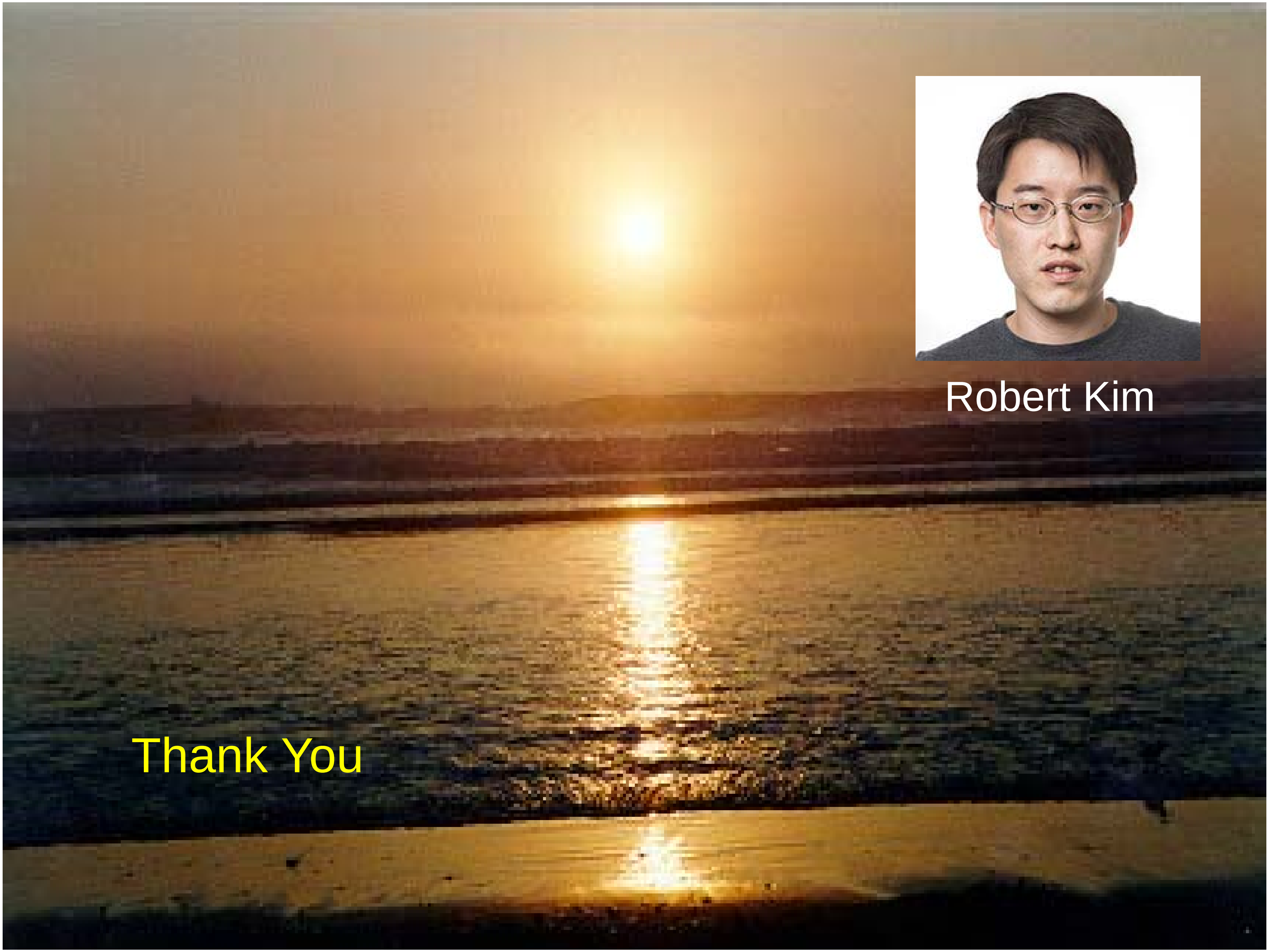
Cortical Hierarchy

Kim et al., Cell 2017

Wright Brothers at Kitty Hawk



The Wright Brothers – David McCullough – 2016


Robert Kim
Thank You