NeuroAI

Past, Present, (and Future?)

Panos Sapountzis, School of Medicine, University of Crete

What is neuroAI?

NeuroAI is defined by two research directions:

a) building more effective AI systems by drawing insights from brains and their circuits

b) applying AI models to understand how brain circuits function to drive perception and behavior

Richard Feynman: "That which I cannot create, I do not understand".

The origins of AI

- The first paper on neural networks by McCulloch and Pitts (1943) opens with the statement:

  “Because of the “all-or-none” character of nervous activity, neural events and the relations among them can be treated by means of propositional logic.”

Figure from McCulloch and Pitts showing how neuronal organization might reflect aspects of Boolean logic.

The origins of AI

- John von Neumann’s report on the first computer architecture (EDVAC) in 1945, dedicated an entire chapter discussing whether the proposed system was sufficiently brain-like.

> “the neurons of the higher animals… have all-or-none character, that is two states: Quiescent and excited… Following W. S. McCulloch and W. Pitts we ignore the more complicated aspects of neuron functioning: Thresholds, temporal summation, relative inhibition, changes of the threshold by after-effects of stimulation beyond the synaptic delay, etc.
>
> Since these tube arrangements are to handle numbers by means of their digits, it is natural to use a system of arithmetic in which the digits are also two valued. This suggests the use of the binary system. The analogs of human neurons are equally all-or-none elements. It will appear that they are quite useful for all preliminary, orienting, considerations of vacuum tube systems.”

- Von Neumann was justifying his choices about how to develop the structure and function of a computer by referring to a biological model.

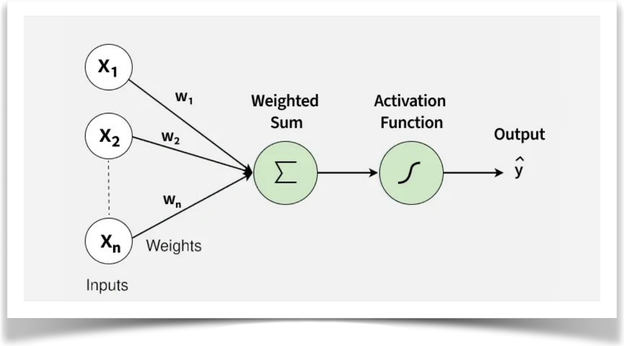
- At the moment of its birth, von Neumann's computer was seen as a brain!

The origins of AI

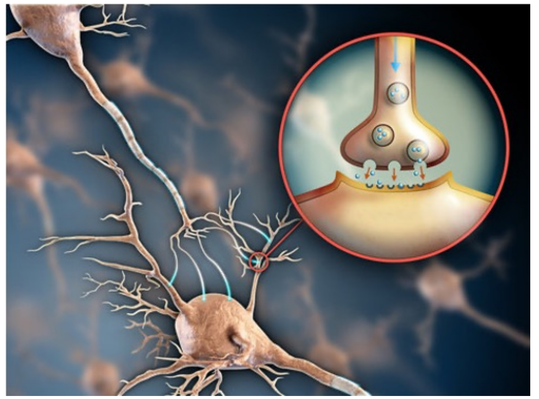
- Frank Rosenblatt's ***perceptron*** in 1958, established ***synaptic connections*** as the primary way of learning in artificial neural networks.

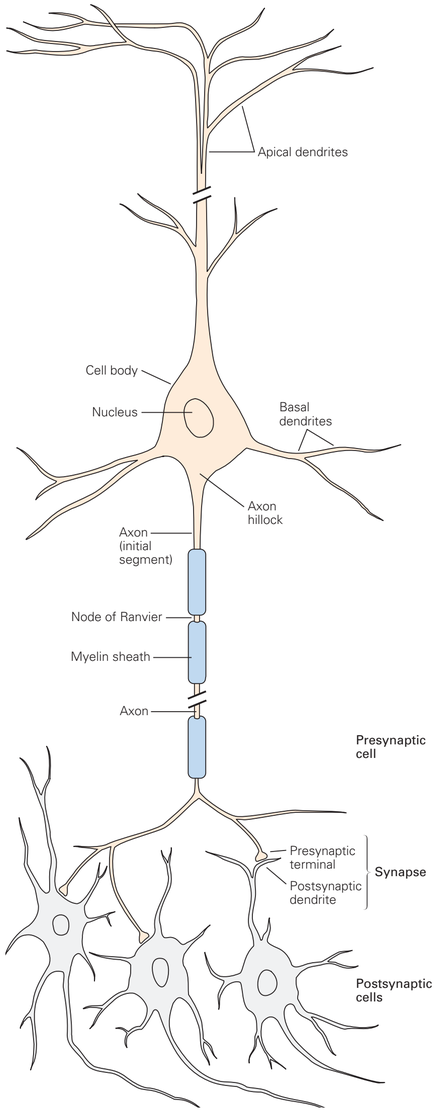
  An article was published in The New York Times in 1958 under the headline:
  "Electronic 'Brain' Teaches Itself"

- Rosenblatt was just as interested in developing a new technology (the perceptron) as he was in finding a theoretical explanation of brain function.



Synapses

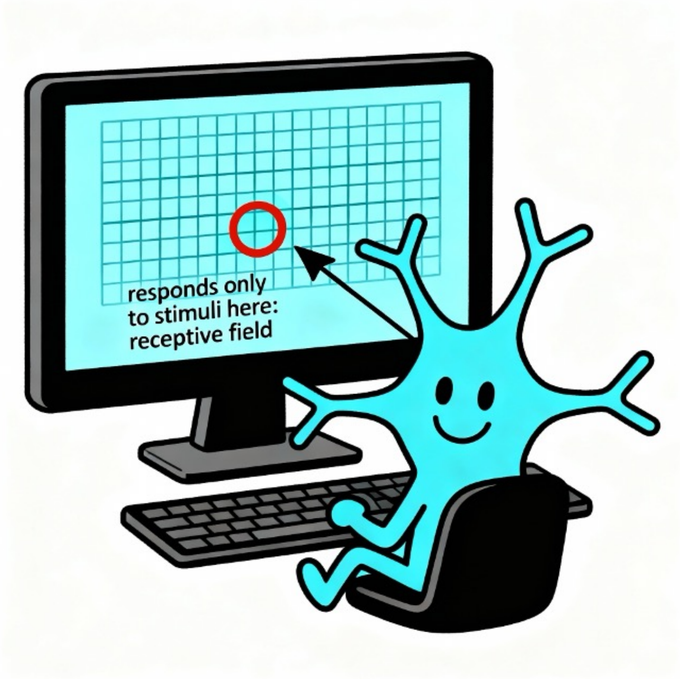




- The human brain contains about **86 billion neurons**.
- **Synapses** are the junctions between neurons—the points where nerve cells communicate with one another. There are roughly 100 trillion synapses in the brain.
- Neuroplasticity is the brain's ability to reorganize itself and change its connections and structure in response to experience, learning, or injury.
- This happens in several ways:
  - Synapses become more or less excitable, meaning they are more or less likely to pass a signal from one neuron to another.
  - Neurons form new synapses, while others are eliminated.
  - New dendrites grow, or existing ones shrink
- ·The neuroscience-inspired idea that synapses are the plastic elements, or free parameters, of a neural network has remained central to modern AI.

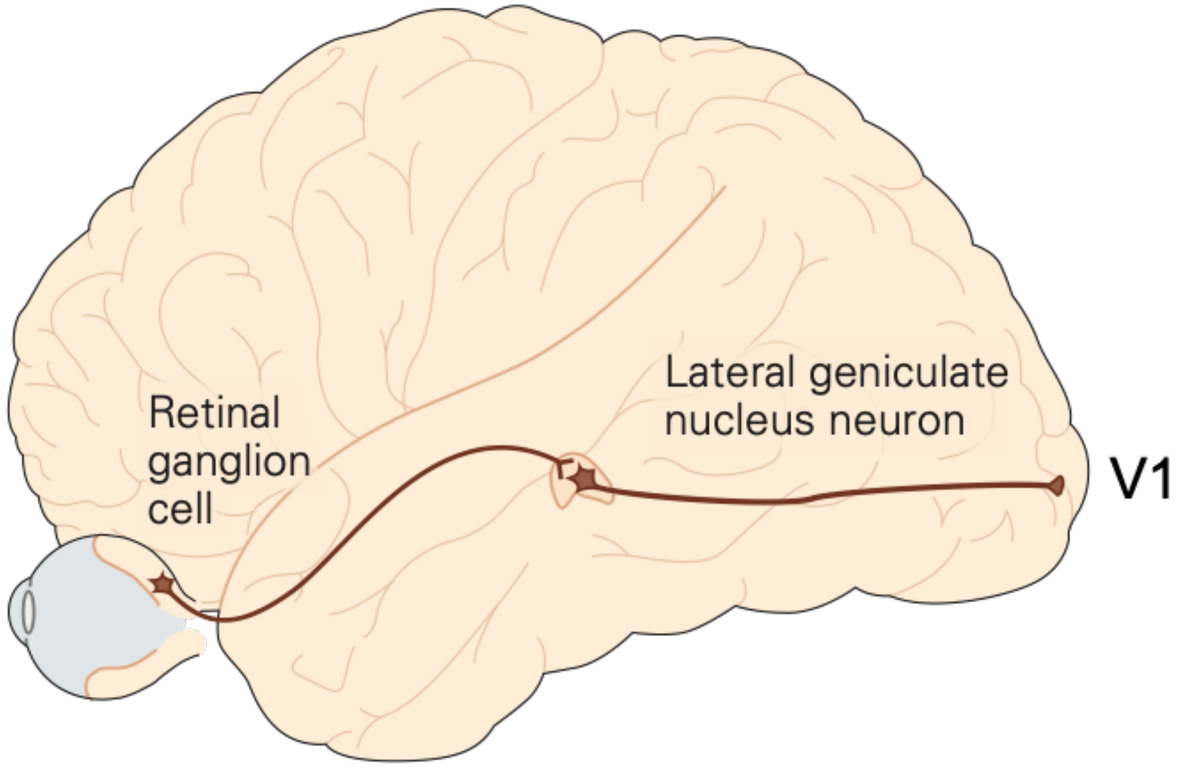
The receptive field

- In the visual system, a neuron's **receptive field** is the specific part of the visual field where visual stimulation influences the neuron's activity.
- In other words, the receptive field is the part of the visual scene that a neuron “sees”.



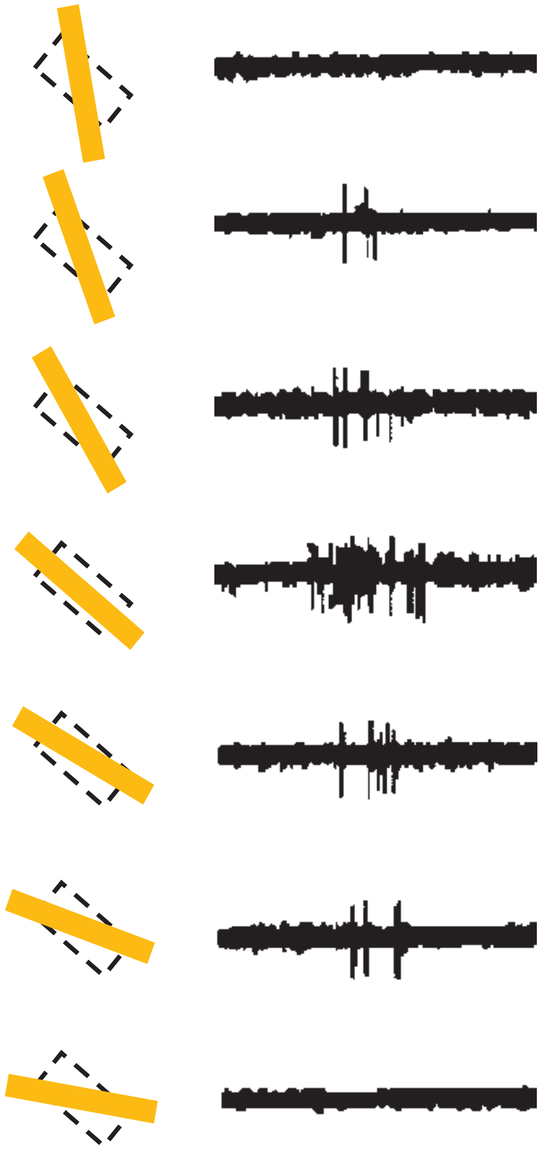
Area V1

- Area V1 is located in the occipital lobe, at the back of the head.
- It is the largest of the visual areas.
- ***Neuronal selectivity*** means that a neuron is more responsive to one kind of input than to others. It is like a specialist cell: one neuron may respond strongly to a certain shape, direction of motion, face, or other feature, while ignoring most others.



Orientation selectivity in V1

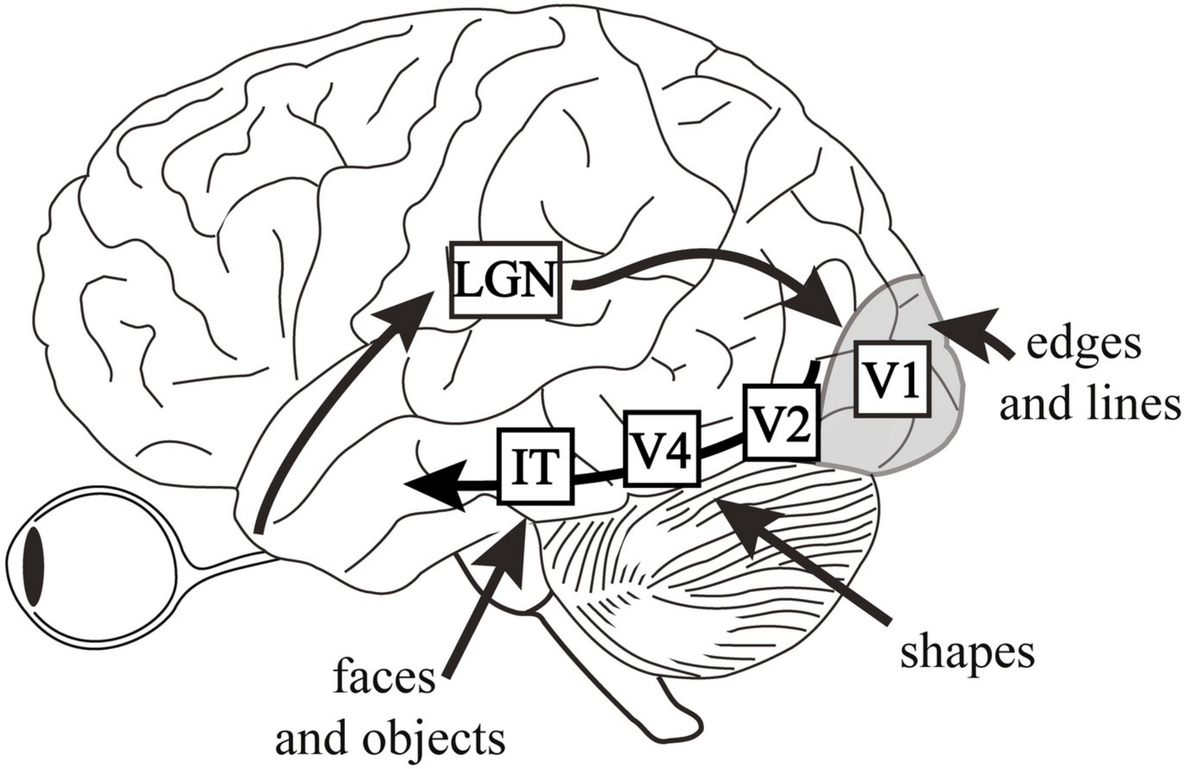
- Hubel and Wiesel discovered that the critical stimulus for V1 cells is elongated lines.
- Moreover, V1 cells responded to ***lines of particular orientation*** (e.g., one cell might respond best to horizontal lines, whereas another cell might prefer vertical lines).
- This specialization of V1 neurons to respond to specific orientations of edges or lines within their receptive fields is called ***orientation selectivity.***

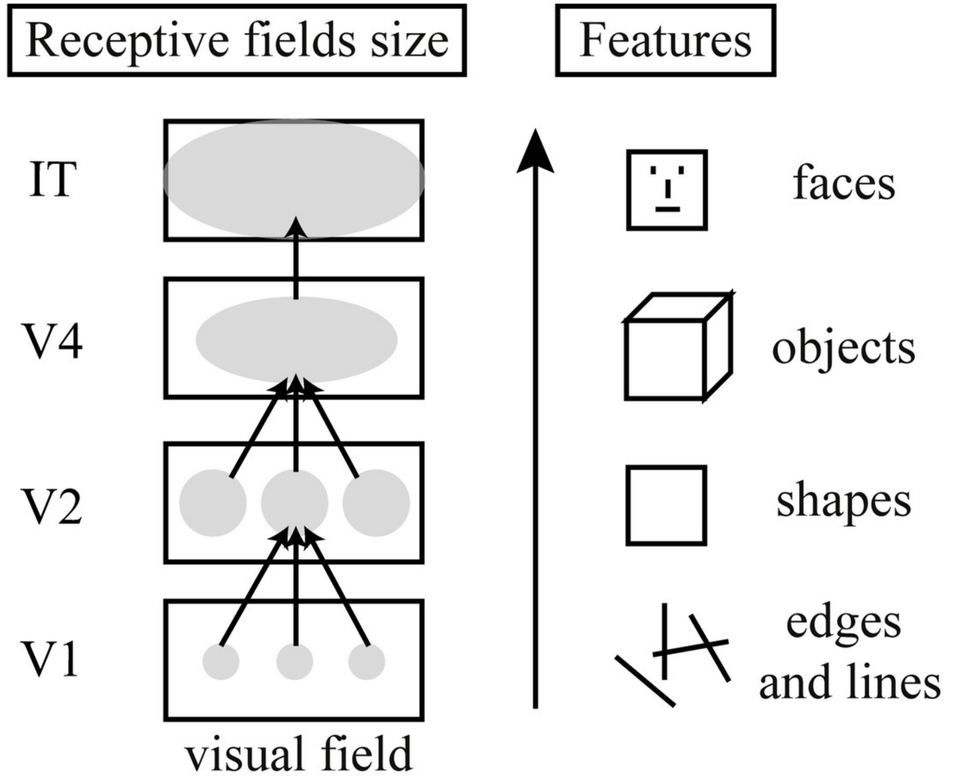


Hubel and Wiesel (Nobel Prize 1981)

- An image after being filtered by receptive fields similar to those in V1.
- Note that each neuron signals only the ***contours*** at its preferred orientation.
- Orientation selectivity is crucial for detecting ***edges and contours*** in a visual scene. V1 neurons help to define the shapes and boundaries of objects, contributing to our perception of form.

The visual hierarchy

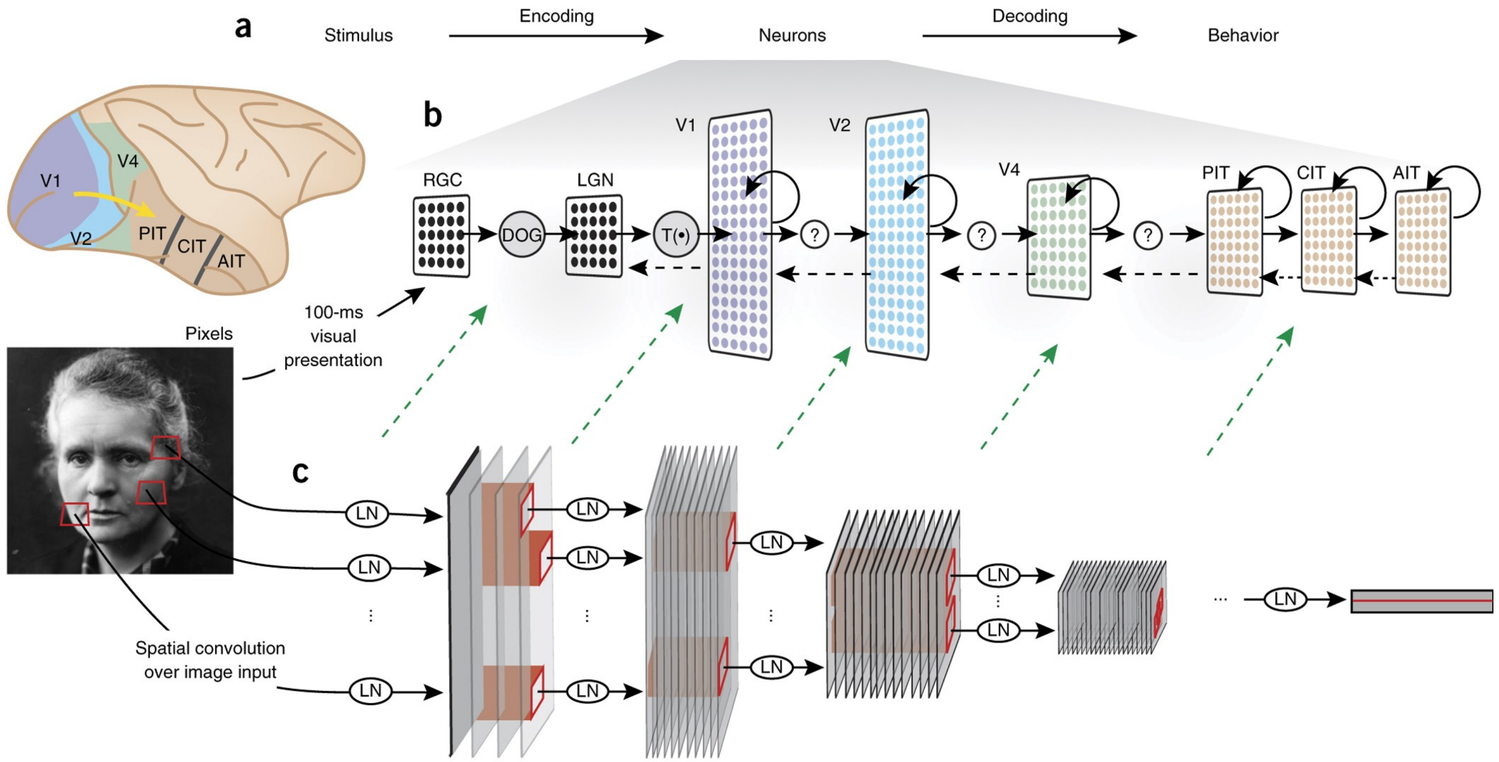




- As you move up the visual hierarchy, neurons gradually have larger receptive fields and respond to more complex stimuli.

Herzog & Clarke, Front Comput Neurosc., 2014

Convolutional Neural Networks



- The activity of artificial neurons in intermediate layers can be used to predict the firing patterns of biological neurons in the primate ventral visual stream.

- Because these models predict neural responses so well, they serve as excellent in silico models of the brain.

Yamins & DiCarlo, Nat Neurosci, 2016

Reinforcement learning

- ***Temporal-difference (TD)*** methods were inspired by research into animal behavior in conditioning experiments.
- ***Second-Order Conditioning*** explains how value "travels backward" through time. An animal learns to value a Light that predicts a Bell, even if the light is never paired with Food.
- In TD learning, the Bell becomes a reward in itself because it predicts future food. This allows both animals and AI to master long, ***complex sequences of behavior*** (Sutton and Barto, 1981).
- This allowed for the development of "brain-like" agents that can learn from sparse rewards. Agents do not have to wait until the end of a task to learn; they update their guess based on the very next step.
- The brain's dopamine system is modeled as a biological TD Error signal. It calculates the difference between expected and actual rewards.

The present

- Although some of the major advances that gave rise to modern ANNs were inspired by neuroscience, the fields have gradually drifted apart.
- Current state of the art architectures, like Transformers, do not resemble biological neurons or known brain structures.
- Biological plausibility is not necessary to give you results. Maybe, massive data and compute is all you need.
- The direction has shifted: although historically neuroscience shaped AI, it is now neuroscientists who use ANN models to help them understand the brain.

The case of DeepMind

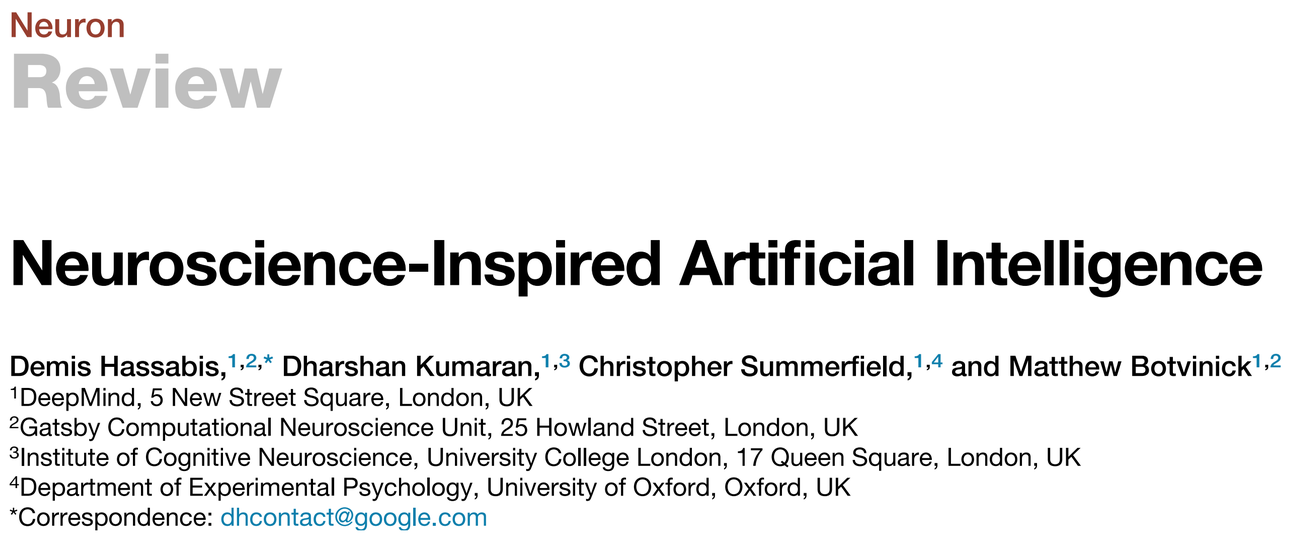
- Hassabis and Legg met at the Gatsby Computational Neuroscience Unit at UCL. The original spirit was that if we want to build Artificial General Intelligence (AGI), we should look at the only proof of concept we have: the human brain.

- DeepMind's early fame came from Reinforcement Learning. This was inspired by how the brain's reward system (dopamine) works.

- In the brain, the hippocampus replays events during sleep to consolidate memories. DeepMind applied this to their deep-Q network (DQN), an algorithm that learnt to master a range of Atari 2600 games. DQN mimics "experience replay", by storing a subset of training data that it reviews "offline", allowing it to learn from successes or failures that occurred in the past.

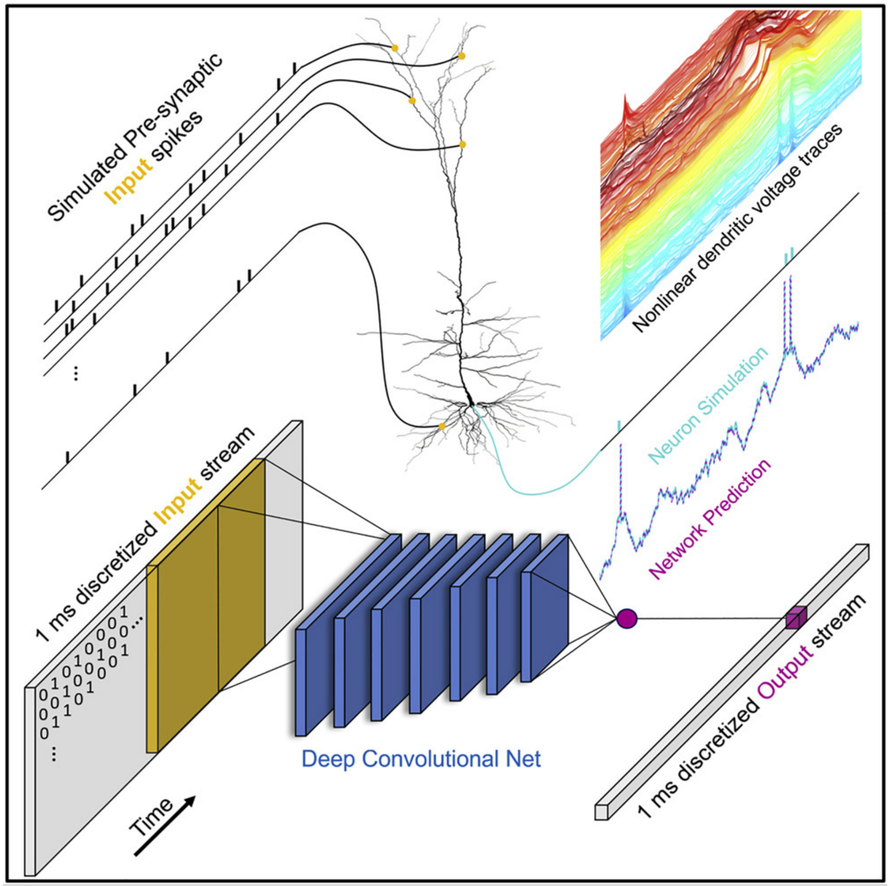
- In the early days, DeepMind had a more distinct Neuroscience Team that operated somewhat independently. Following the 2023 merger with Google, this work was integrated into the broader Science and Research divisions and the company shifted towards commercial AI applications.

- Although their priorities have shifted, they still produce relevant work.



Biological neurons are complex

- The basic unit of ANNs is an oversimplification.
- Biological cells are not simple "sum-and-threshold" units. They act as sophisticated nonlinear processors where the dendritic tree performs complex computations on inputs before a signal ever reaches the soma.
- Incorporating the nonlinear dynamics of NMDA receptors (which allow for dendritic spikes) dramatically increases computational power.
- ***A single biological cortical neuron is best approximated by a deep neural network with 5 to 8 layers***.



Beniaguev et al., Neuron, 2021

Non-invasive methods to study brain activity

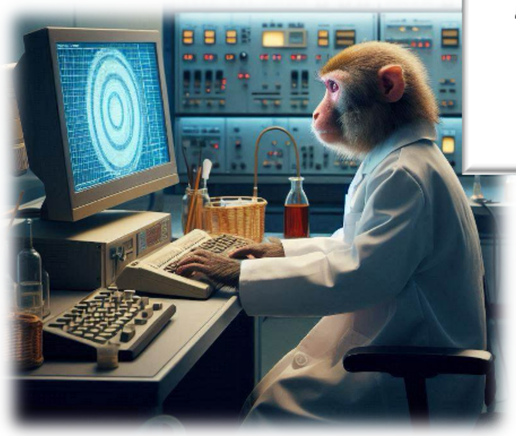
- ***Electroencephalography (EEG)*** is one of the oldest and most widely used methods for the investigation of electrical activity in the brain using electrodes placed on the scalp (non-invasive).

- Under most conditions, ***it has little relationship with the firing patterns of individual neurons***; this is due to the distorting and attenuating effects of the tissues between the current source and the recording electrode.

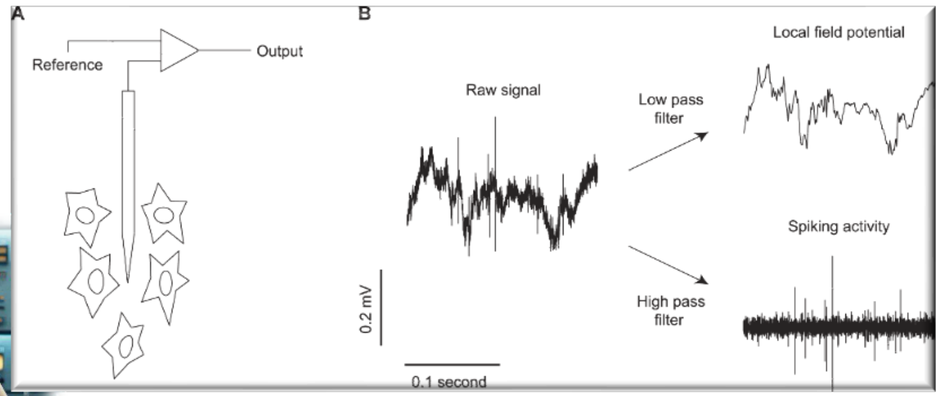
- Other non-invasive methods such as magnetoencephalography (MEG) or functional MRI (fMRI) also suffer from limitations, including poor spatial and/or temporal resolution.

Extracellular recordings in behaving NHPs

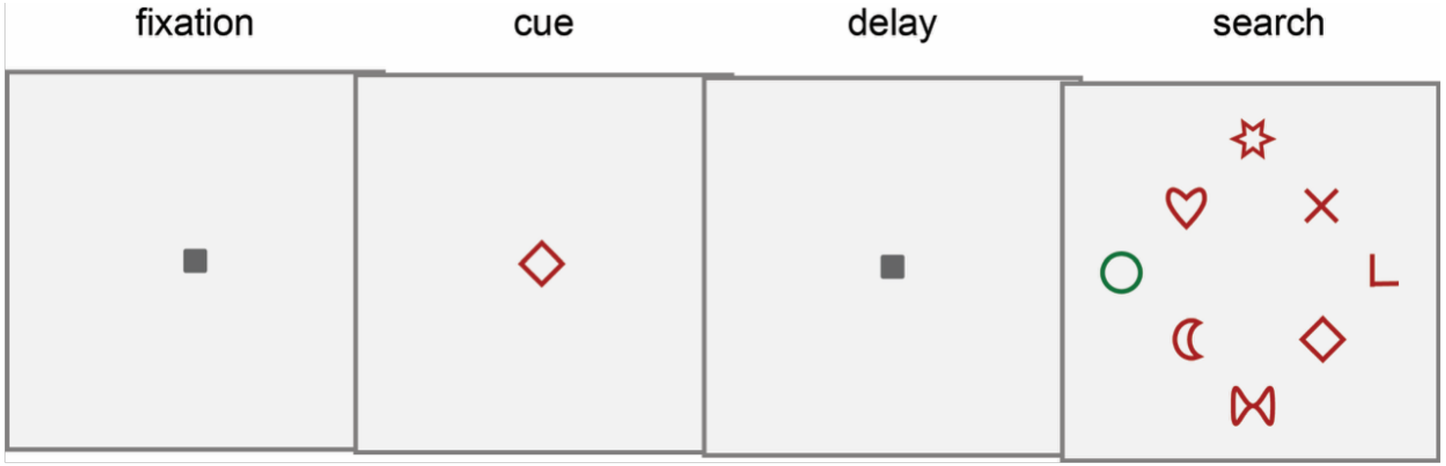
- Monkeys are trained to perform cognitively demanding tasks.
- We conduct invasive extracellular recordings using multi-electrode arrays inserted into the brain.
- We extract ***single-neuron spiking activity*** and ***local field potentials (LFPs)***.

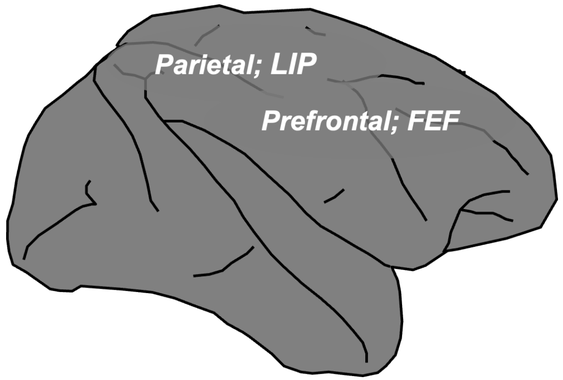


Created with Microsoft Copilot

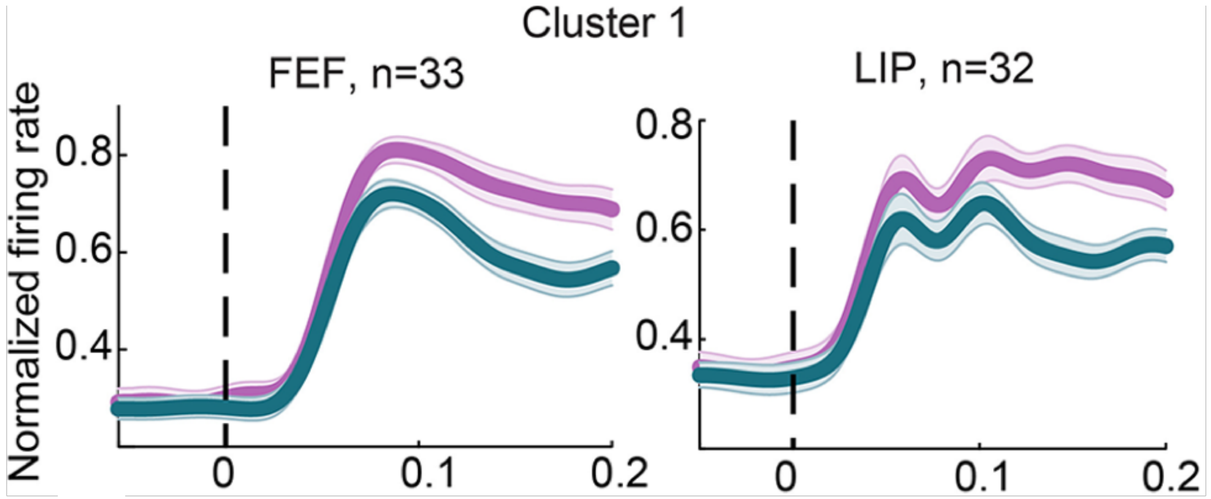


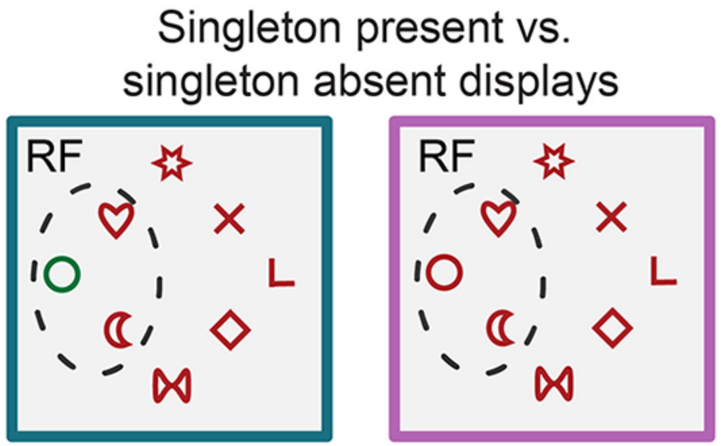
Jia & Kohn, PLOS Biology, 2011



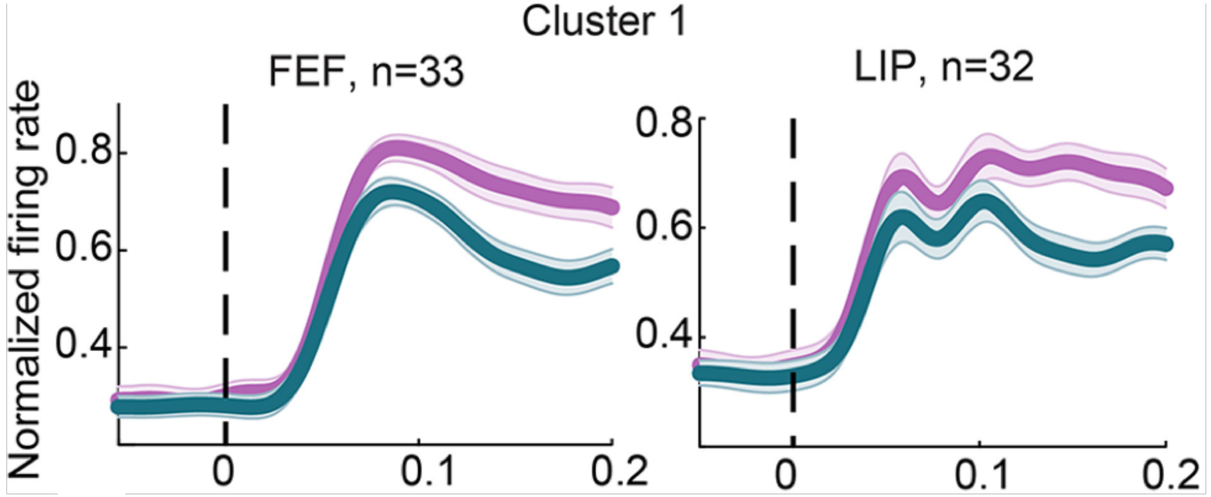


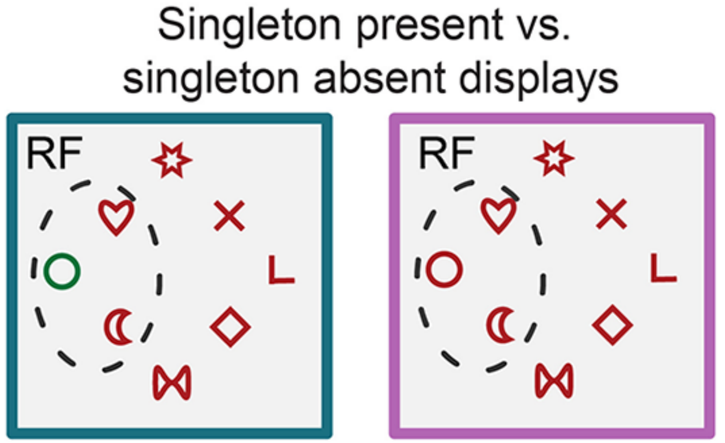
Sapountzis et al., PLOS Biology , 2025

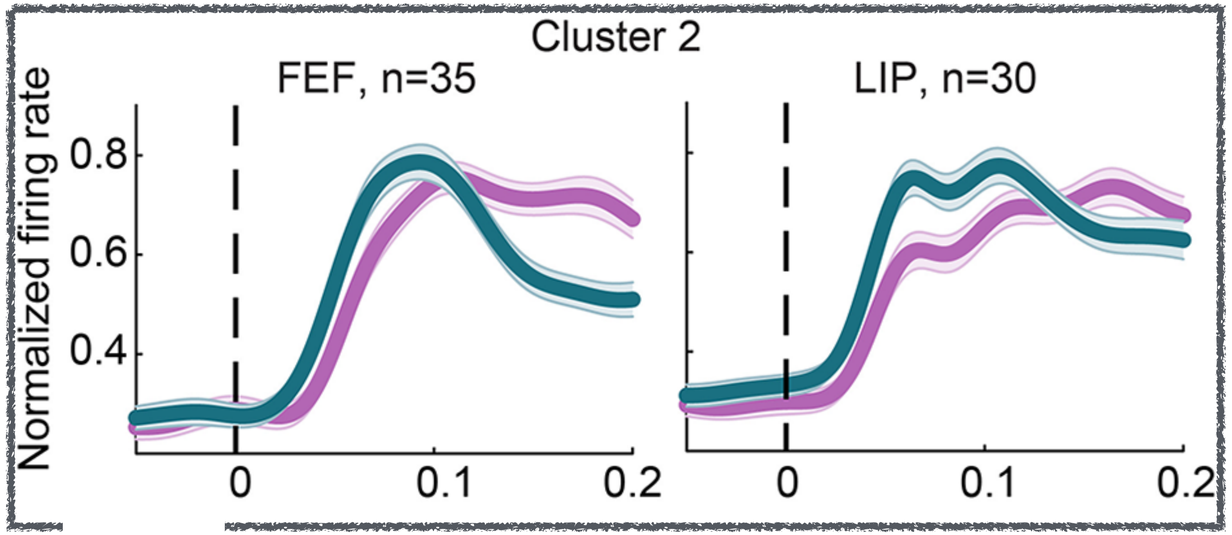




Sapountzis et al., PLOS Biology , 2025

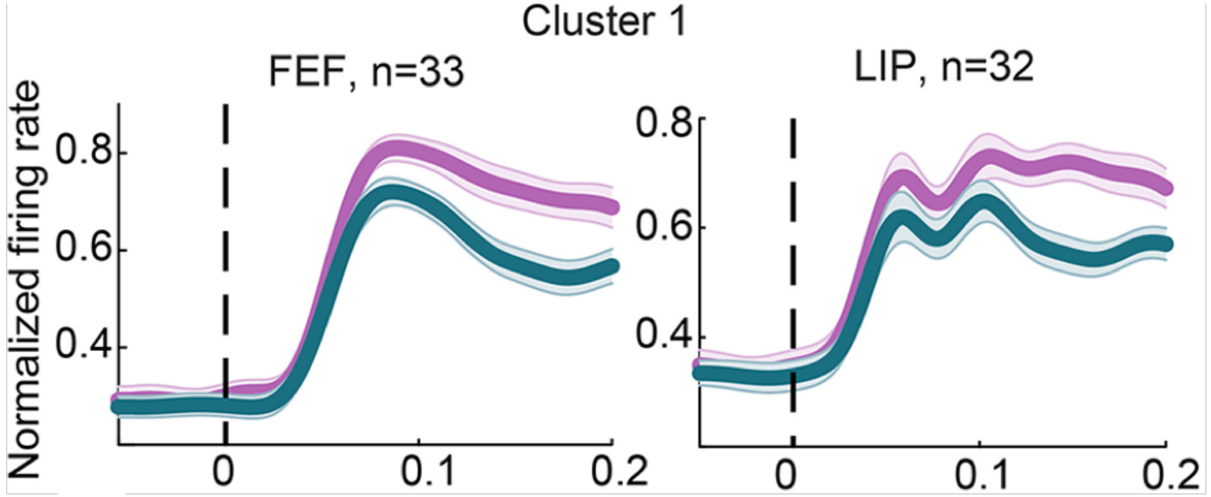


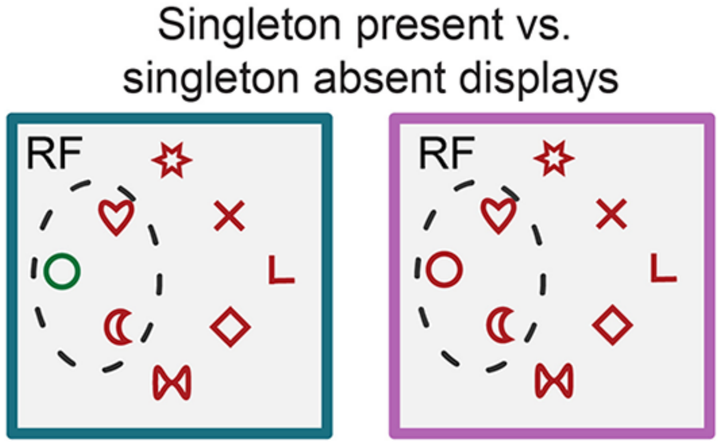


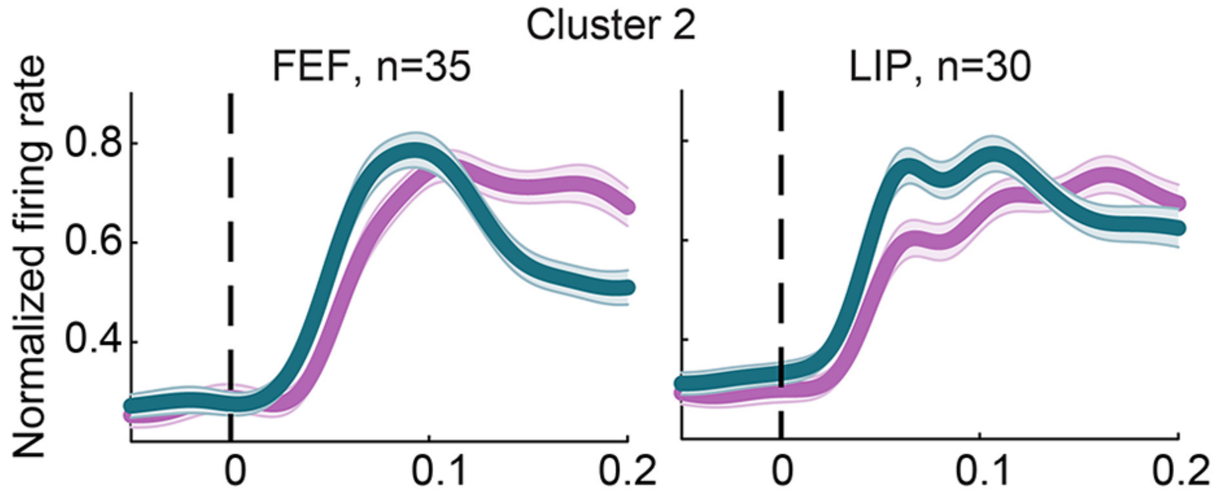


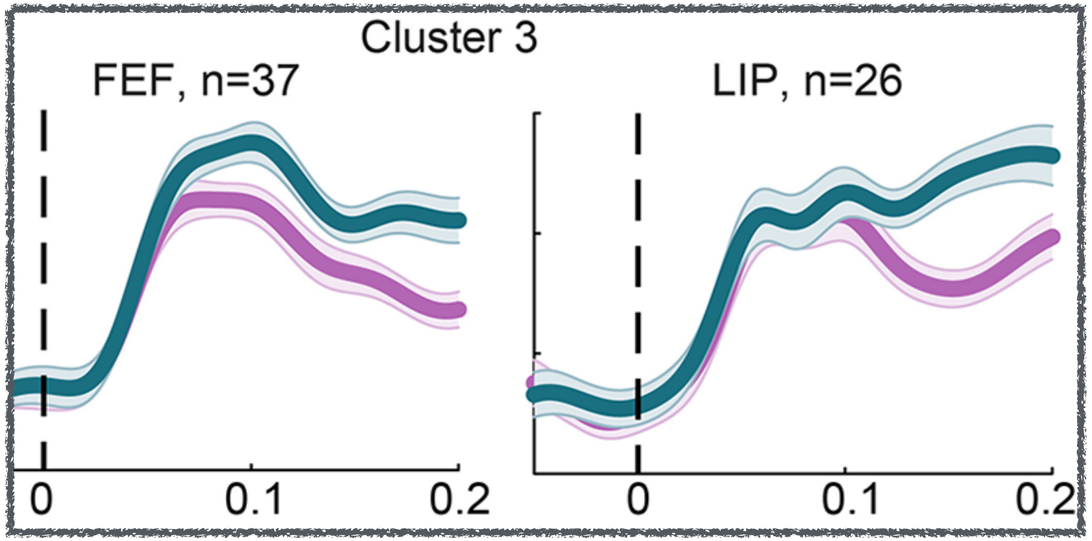
Time (s)

Sapountzis et al., PLOS Biology , 2025





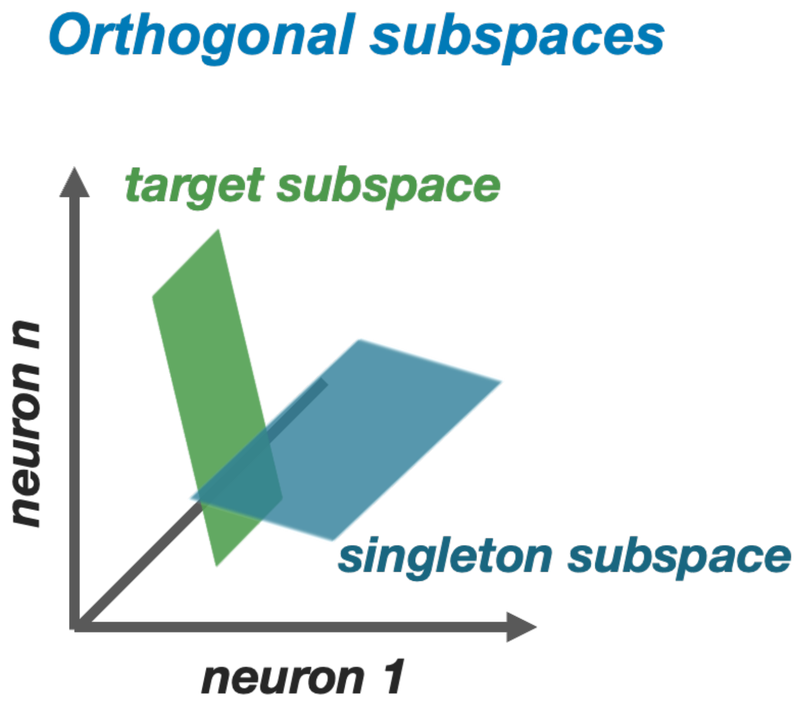




Time (s)

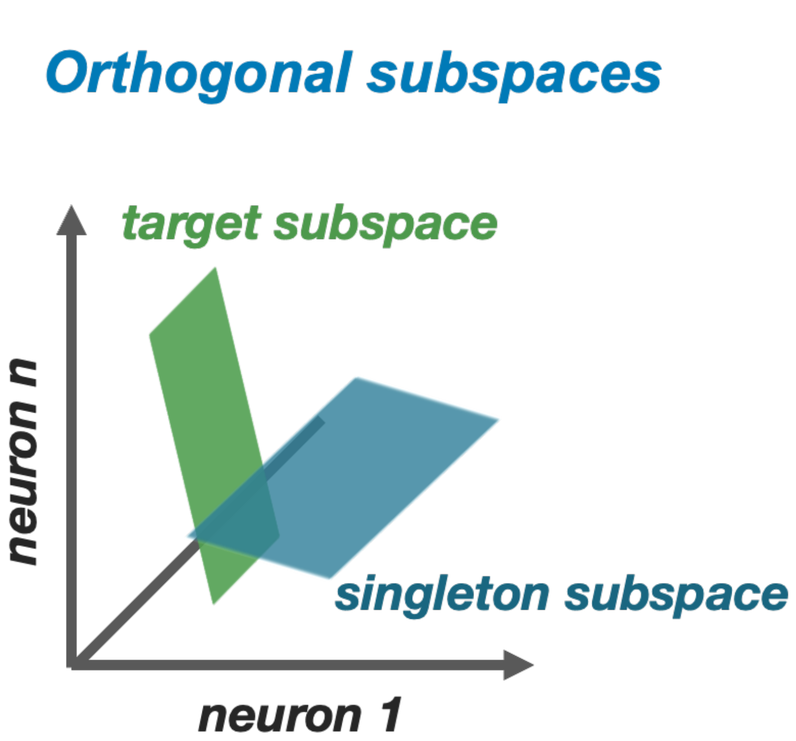
Sapountzis et al., PLOS Biology , 2025

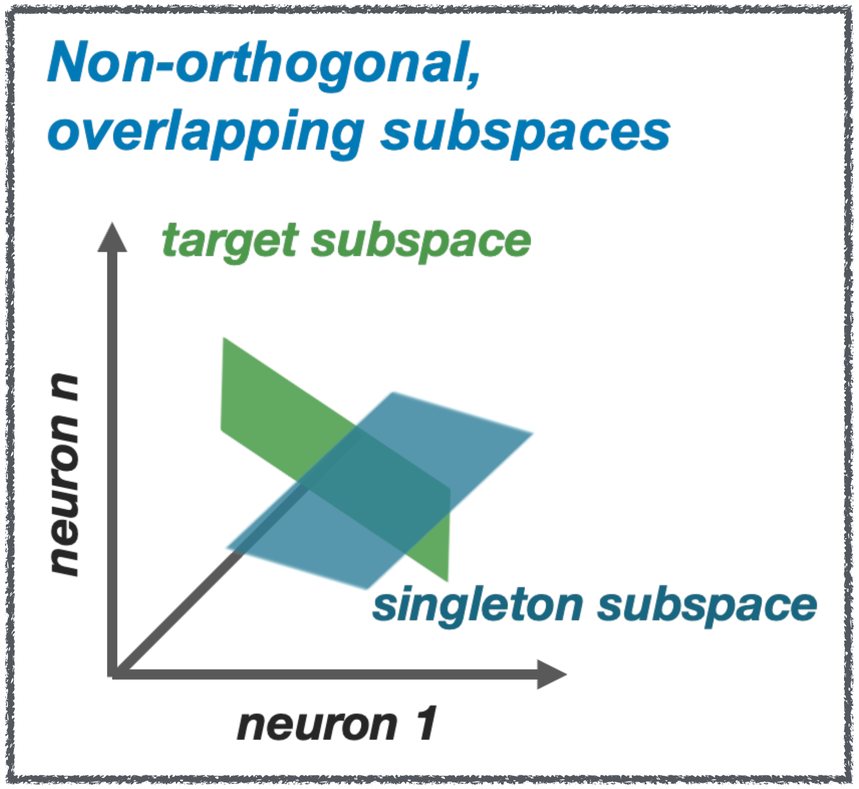
Population representations of target and salient distractor



Sapountzis et al., PLOS Biology , 2025

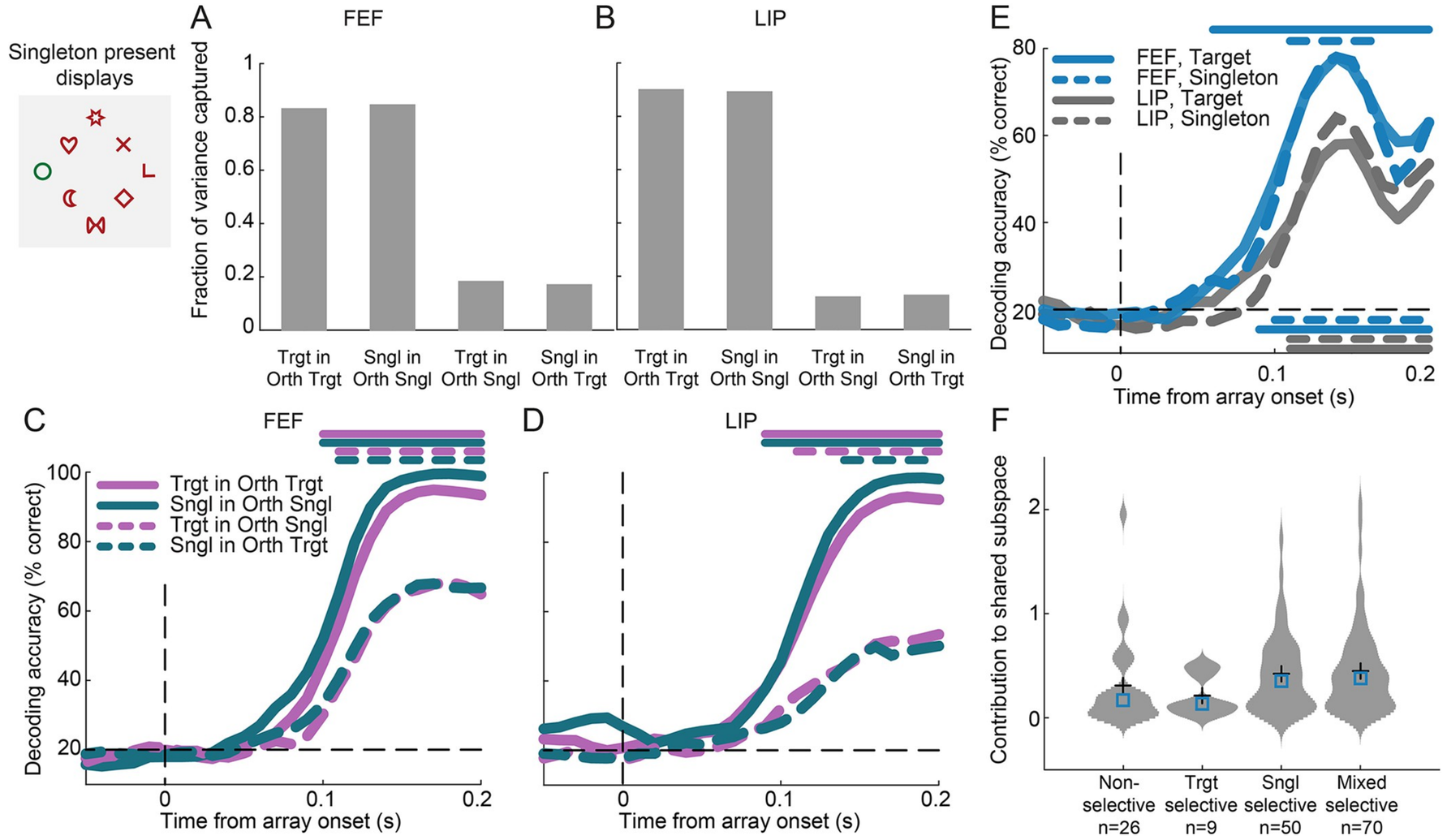
Population representations of target and salient distractor





Sapountzis et al., PLOS Biology , 2025

Population representations of target and salient distractor



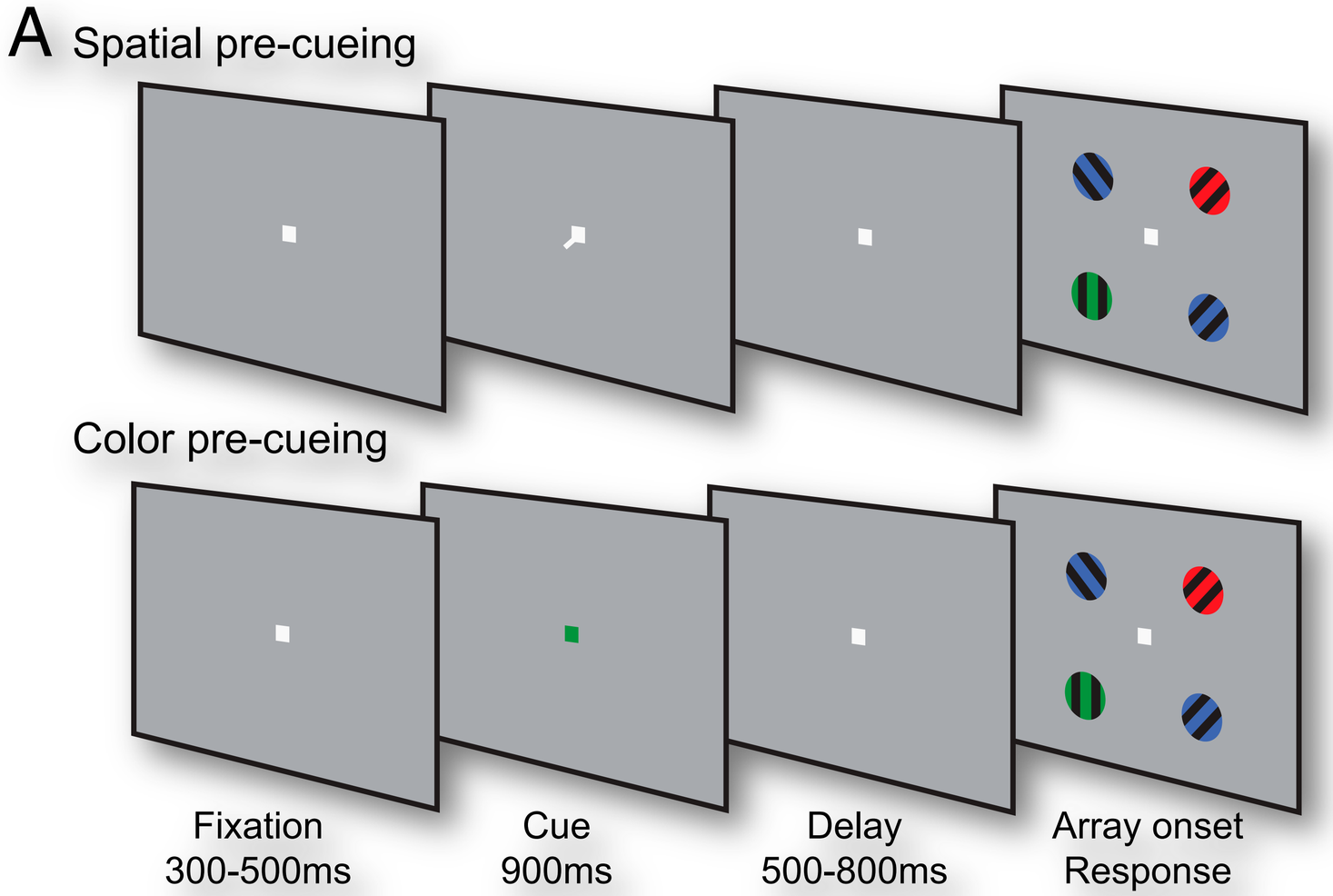
Sapountzis et al., PLOS Biology , 2025

Picsart AI Image Generator (2025)

Scenario 1: She knows about the topics and questions in advance, so she has time to prepare.

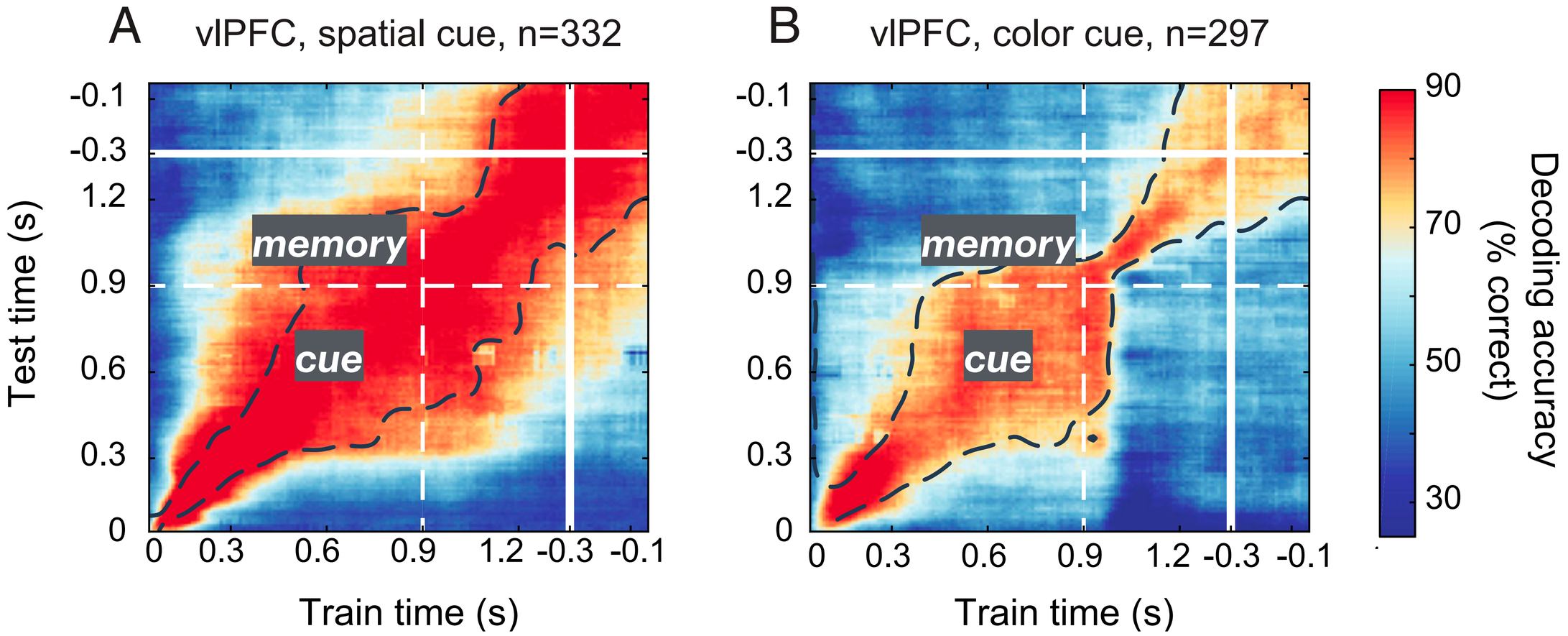
Scenario 2: Questions are not known. She must apply her expertise to address each question on the spot.

Working memory task



Sapountzis et al., PNAS, 2022

Static and dynamic population coding in PFC



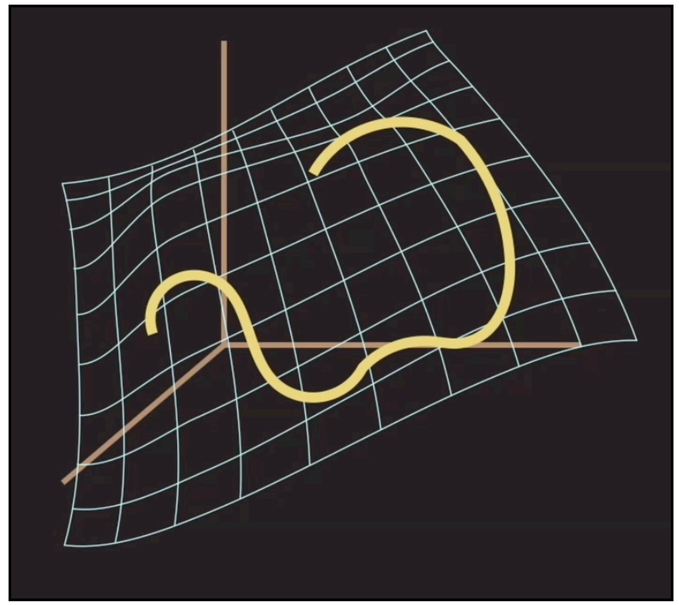
Sapountzis et al., PNAS, 2022

The PFC as a dynamical system

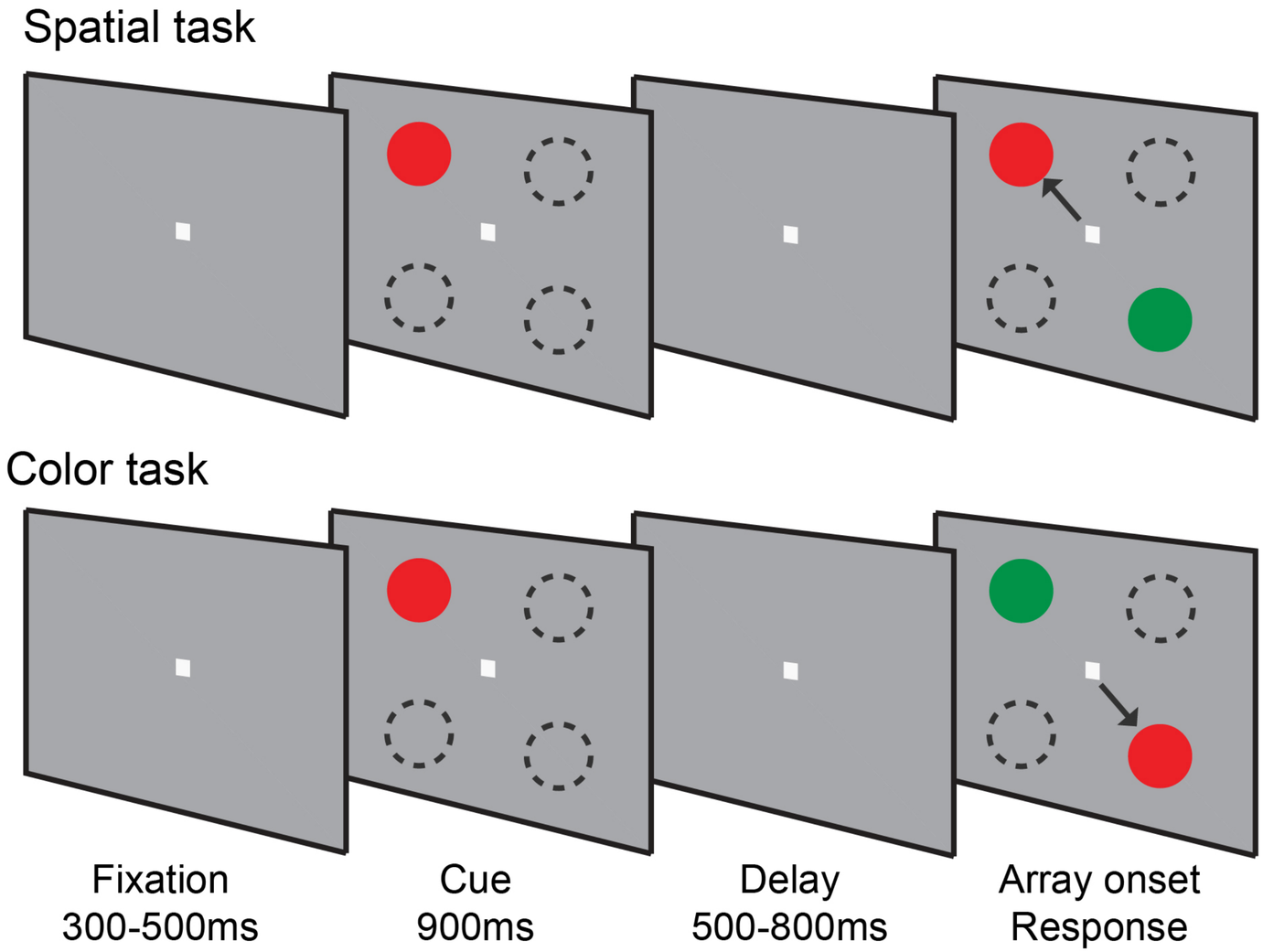
- The PFC exhibits high computational flexibility. Unlike the visual cortex, PFC neurons demonstrate ***mixed selectivity;*** a single neuron encodes multiple, task variables (such as color in one context and location in another).

- As dynamical systems, RNNs do not merely perform static input-output associations; they evolve along ***state-space trajectories***.

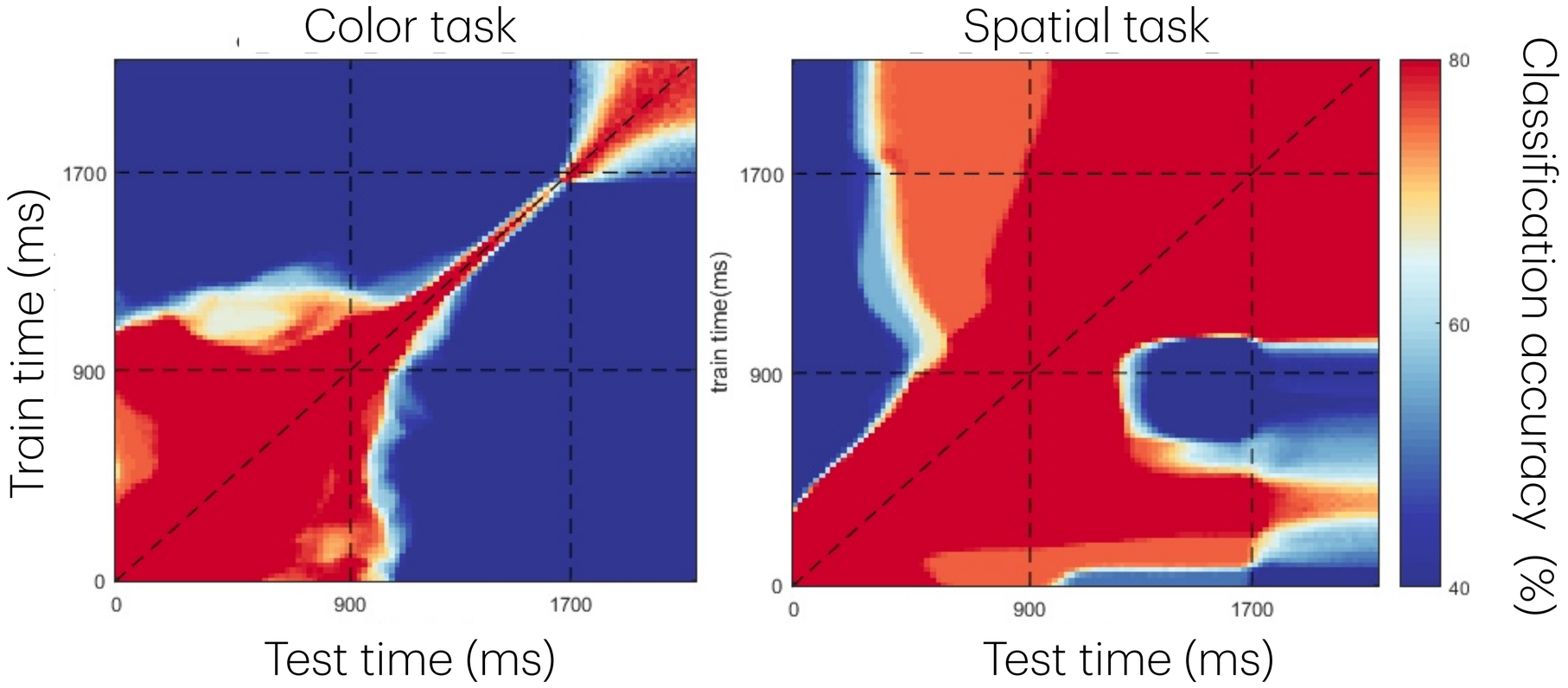
- A trained RNN can effectively re-configure its internal latent states. This enables the model to solve distinct phases of a complex task, closely mirroring the dynamic coding observed in the biological brain during adaptive behavior.



RNN task

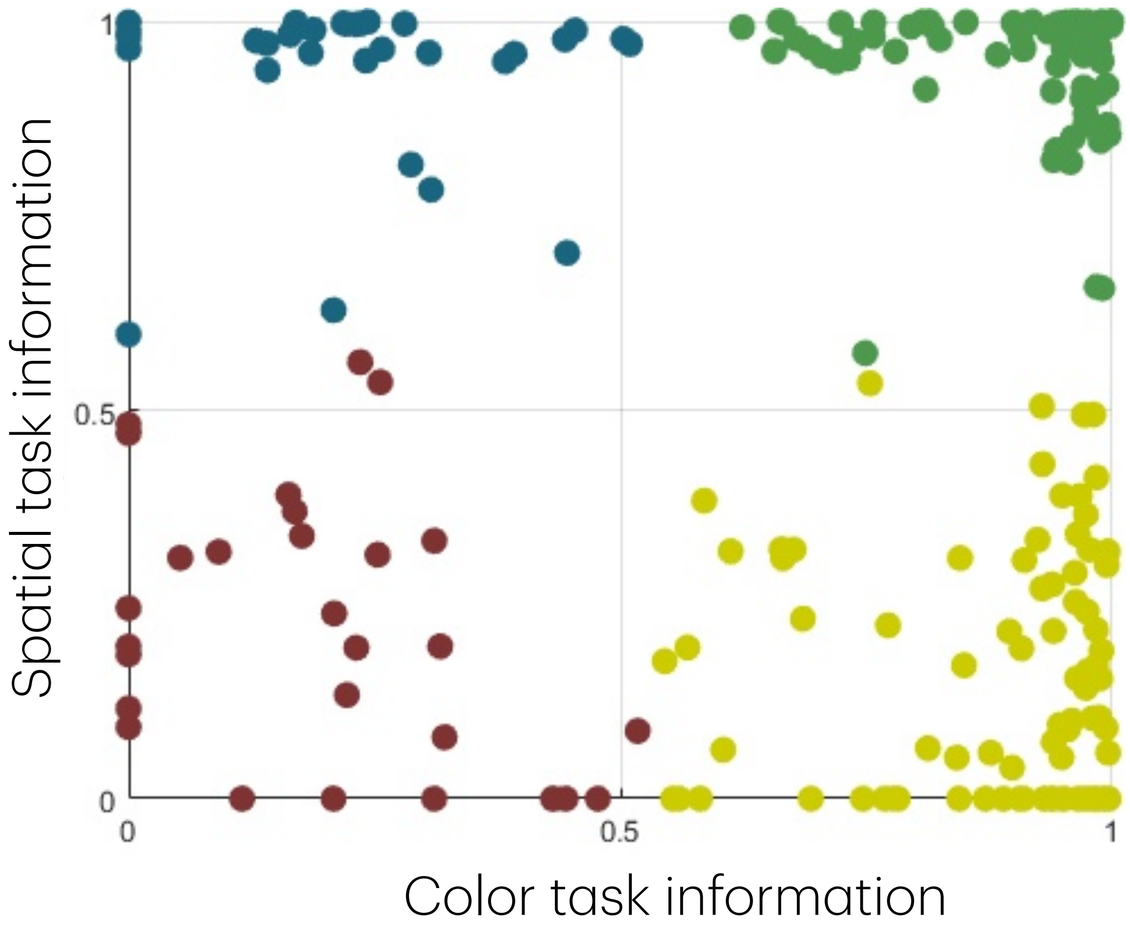

Spatial task
Color task
Fixation
300-500ms
Cue
900ms
Delay
500-800ms
Array onset
Response

Dynamic and static coding in RNNs



Mixed selectivity

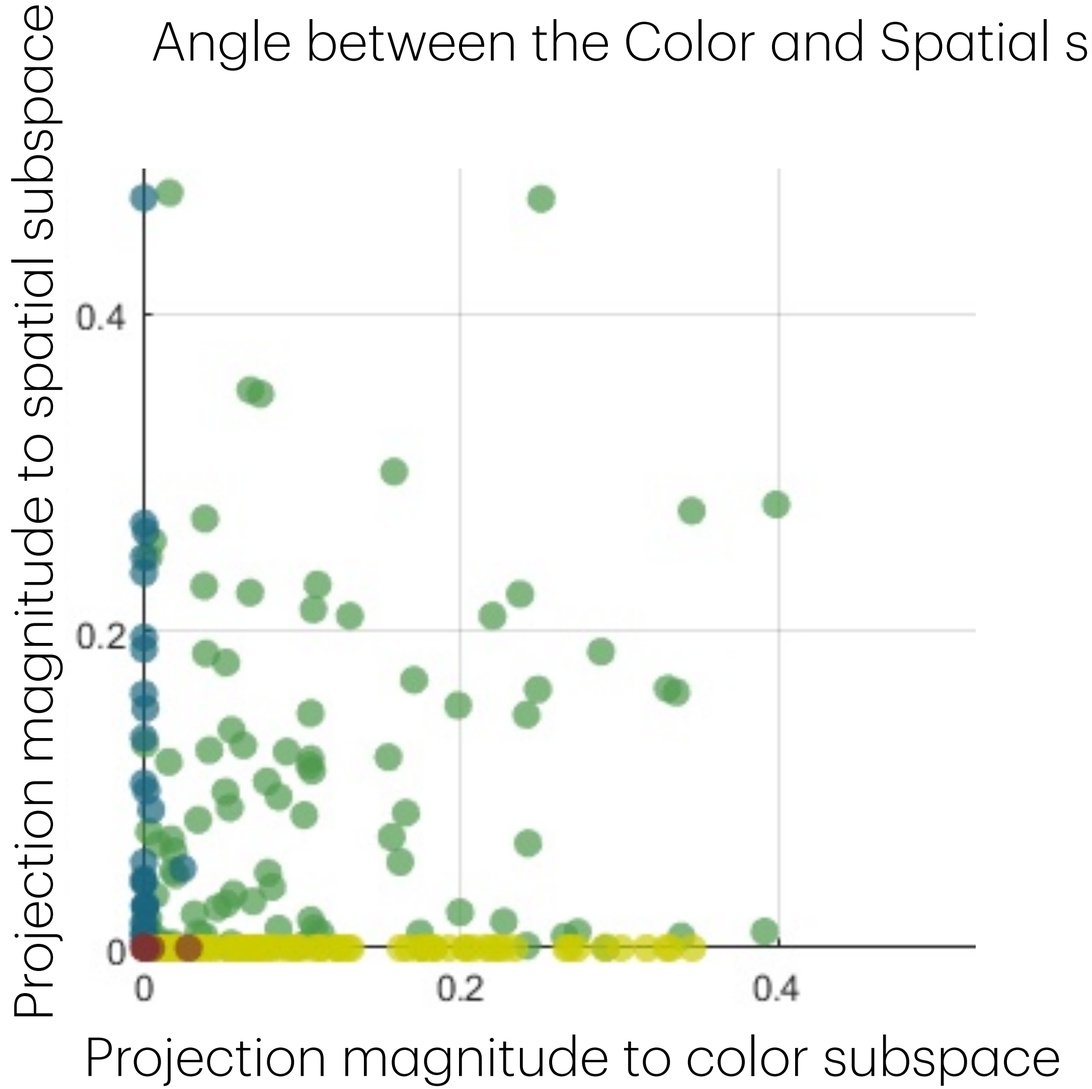
Lesioning mixed selective units impairs the RNN's task performance.



Mixed-selective
Spatial-selective
Color-selective
Non-selective

Mixed-selective units contribute to both subspaces

Angle between the Color and Spatial subspaces is 78 degs



Mixed-selective

Spatial-selective

Color-selective

Non-selective

nature machine intelligence
REVIEW ARTICLE
https://doi.org/10.1038/s42256-019-0025-4

Reinforcement learning in artificial and biological systems

https://1drv.ms/b/c/a9d6a37ab6da0ae2/IQDiCtq2eqPWIICpXz0AAAAAAUz7HTCco_0R61wWR-cvTAA?e=sb1VDe

nature machine intelligence

Perspective
https://doi.org/10.1038/s42256-025-01146-z

What neuroscience can tell AI about learning in continuously changing environments

https://1drv.ms/b/c/a9d6a37ab6da0ae2/IQDC2AhlwUTJQ65rvy4PAe64AduilwnBMRAGvoLt2V6nBnc?e=apkoiX

Neuron
Review

Neuroscience-Inspired Artificial Intelligence

Demis Hassabis,[1,2,*] Dharshan Kumaran,[1,3] Christopher Summerfield,[1,4] and Matthew Botvinick[1,2]

https://www.sciencedirect.com/science/article/pii/S0896627317305093

Backpropagation and the brain

Timothy P. Lillicrap, Adam Santoro, Luke Marris, Colin J. Akerman and Geoffrey Hinton

https://1drv.ms/b/c/a9d6a37ab6da0ae2/IQAc_gpT-onlRY5i6j4fMR6aAWAq2tOVslVNWMXmOD5er7I?e=oZ0GIr

nature communications

Perspective
https://doi.org/10.1038/s41467-023-37180-x

Catalyzing next-generation Artificial Intelligence through NeuroAI

https://www.nature.com/articles/s41467-023-37180-x

Spike-based neuromorphic computing: An overview from bio-inspiration to hardware architectures and learning mechanisms☆

https://www.sciencedirect.com/science/article/pii/S0141933125001073

Physiology of Cognitive Functions Lab

Georgia Gregoriou

Sophie Paneri

Alexandra Antoniadou

Now a PhD Candidate
@ Universitat Autònoma de Barcelona

Sotirios Papadopoulos

Now a PhD Candidate
@ Claude Bernard University Lyon 1