



AI Explained -
Demystifying the Technology behind
ChatGPT to shape our Future

Prof. Benjamin F. Grewe
Institute of Neuroinformatics
5th. Sept. 2023, ISPF Lucerne

**institute of
neuroinformatics**



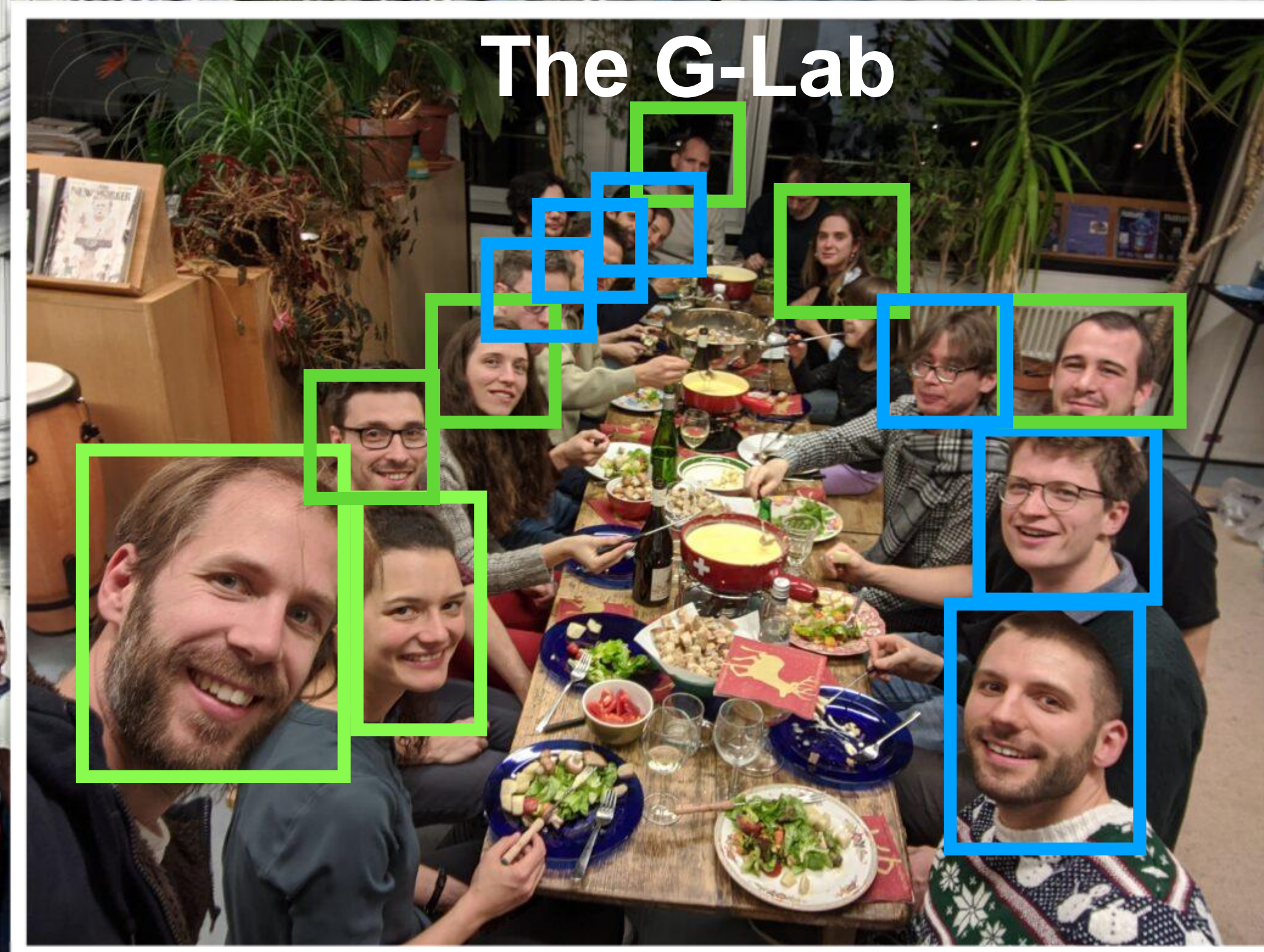
Mission of the Institute

The mission of the Institute is to discover the key principles by which brains work and to implement these in artificial systems that interact intelligently with the real world.




Mission of the Institute

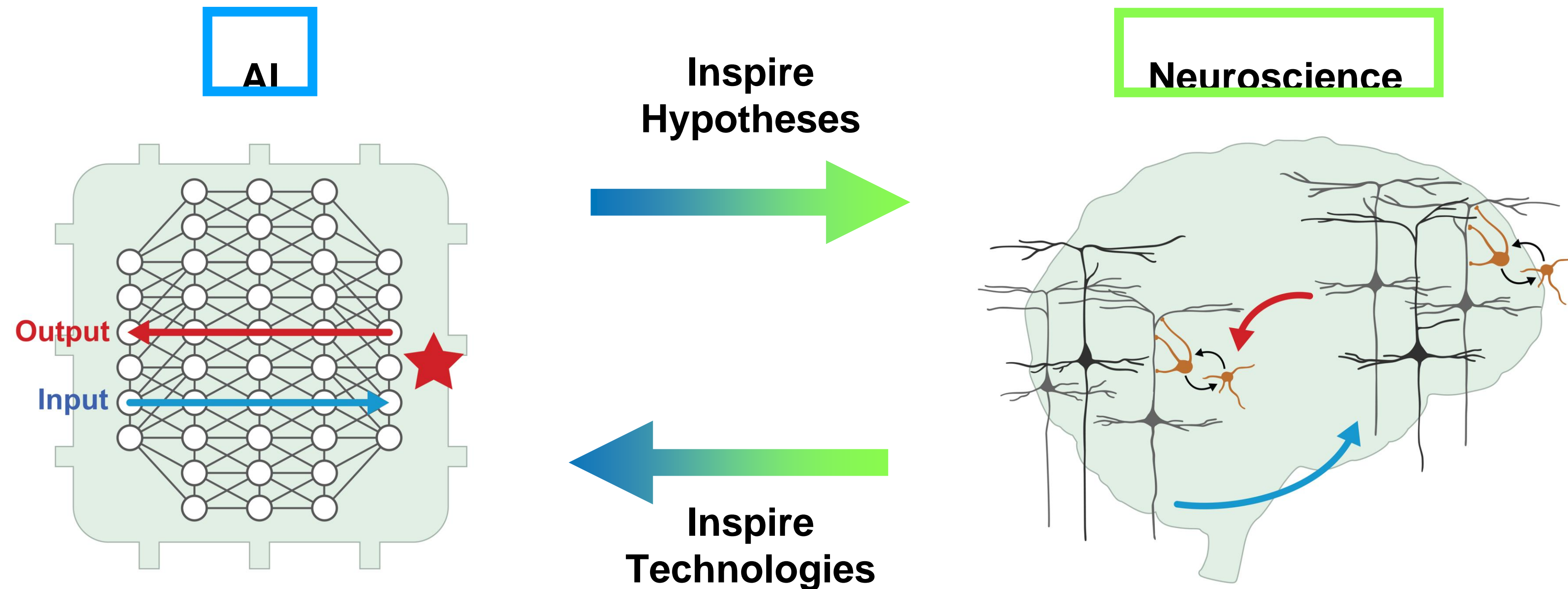
The mission of the Institute is to discover the key principles by which brains work and to implement these in artificial systems that interact intelligently with the real world.



Research Background

-  Computational Neuroscience
-  AI / Computational Neuroscience

G-Lab: Creating Synergies to Advance Neuroscience and AI



Artificial Intelligence

Deep Networks, Transformers

High energy consumption, hundreds of GPUs

Requires huge amounts training data

Natural Intelligence

Biological Neuronal Networks

Highly energy efficient (20W),

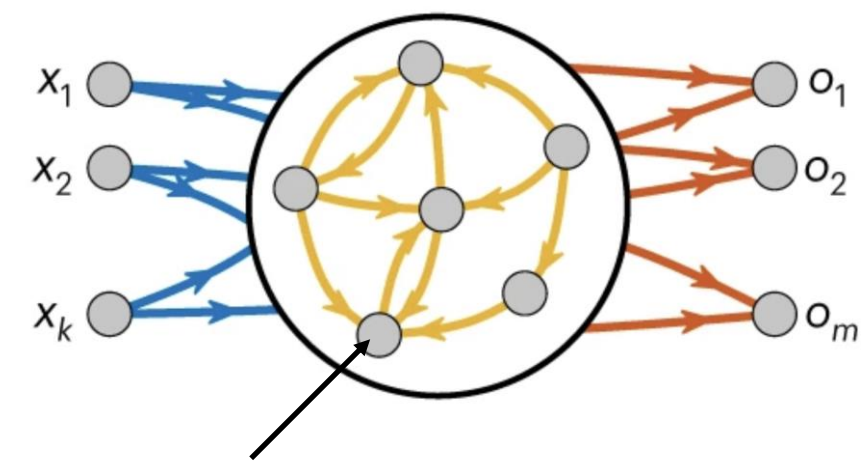
Learns extremely efficient

Part I: Learning in Hierarchical (Deep) Cortical Networks

AI



Neuroscience



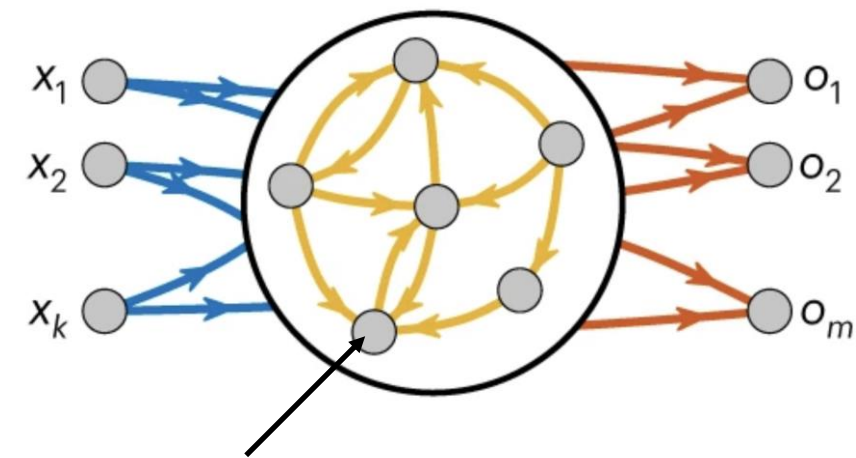
Scientific Question:
How does Credit Assignment
in the Brain work?

Part I: Learning in Hierarchical (Deep) Cortical Networks

AI



Neuroscience



Scientific Question:
How does Credit Assignment
in the Brain work?

Part II: Understanding Hierarchical Neuronal Representations in Brain

AI



Neuroscience

Scientific Question:
What are the neuronal representations of the
sensory input (e.g. image of bus) that allows our brain
to generate goal directed actions?

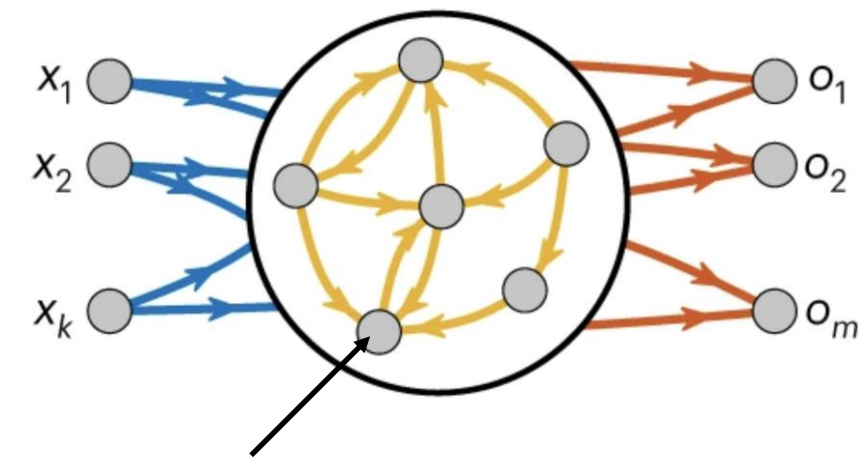


Part I: Learning in Hierarchical (Deep) Cortical Networks

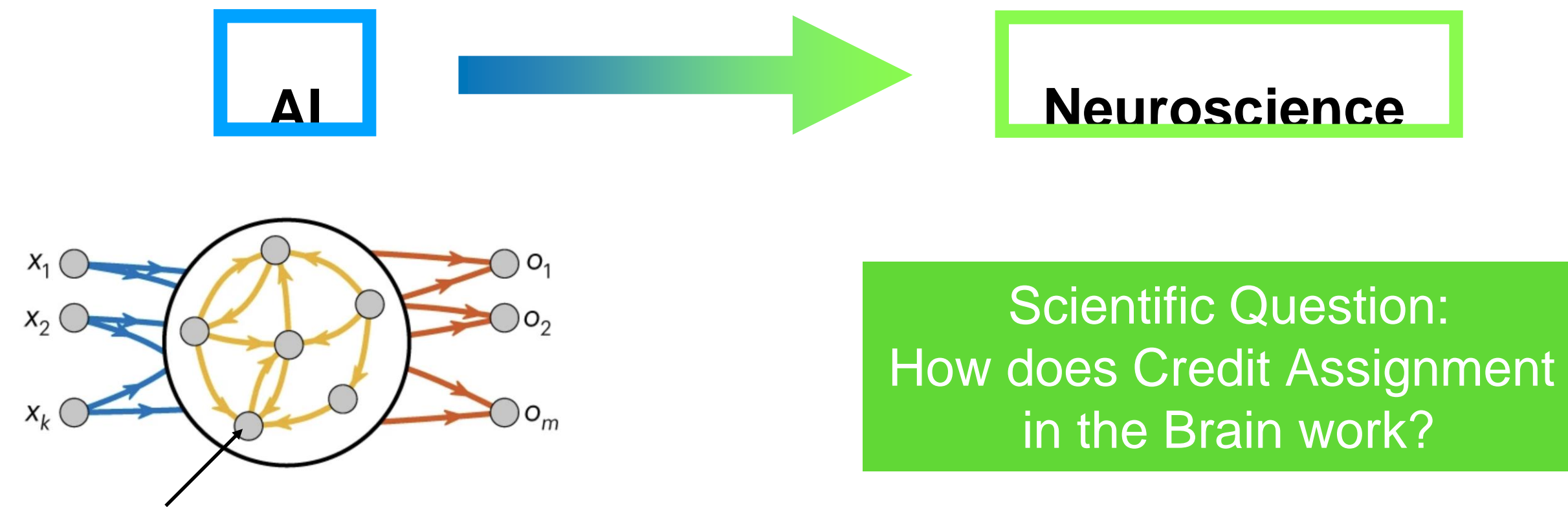
AI



Neuroscience

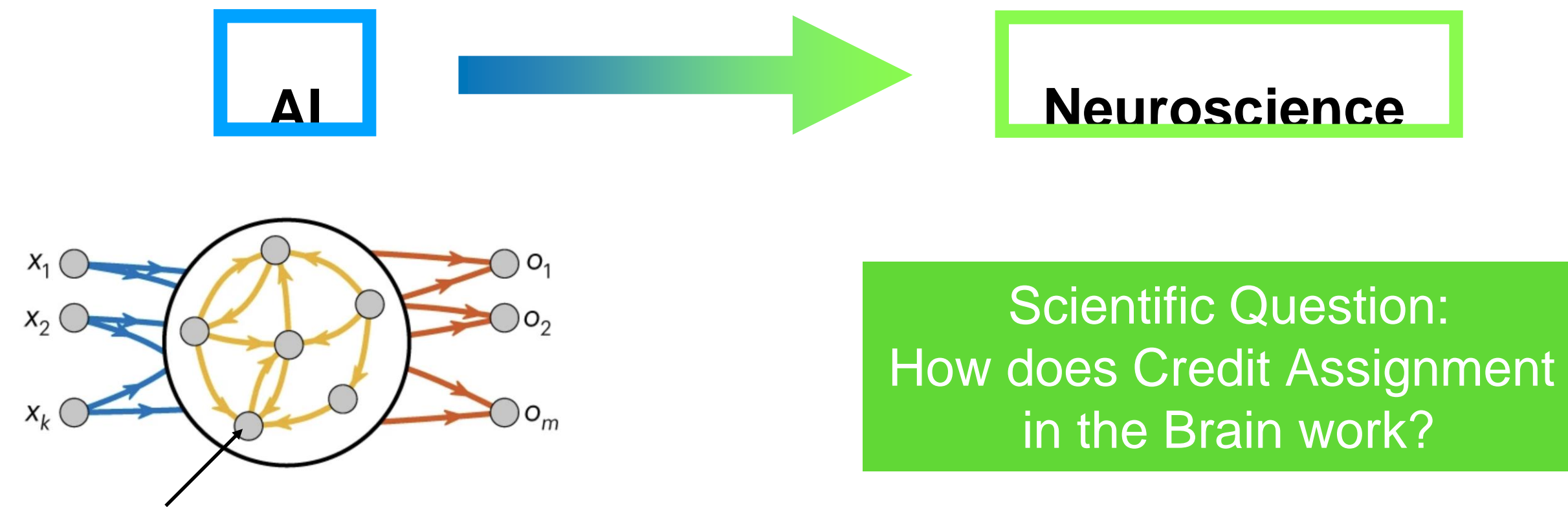


Scientific Question:
How does Credit Assignment
in the Brain work?

Part I: Learning in Hierarchical (Deep) Cortical Networks

Our Approach: Make Deep Network Learning more Biologically Plausible.

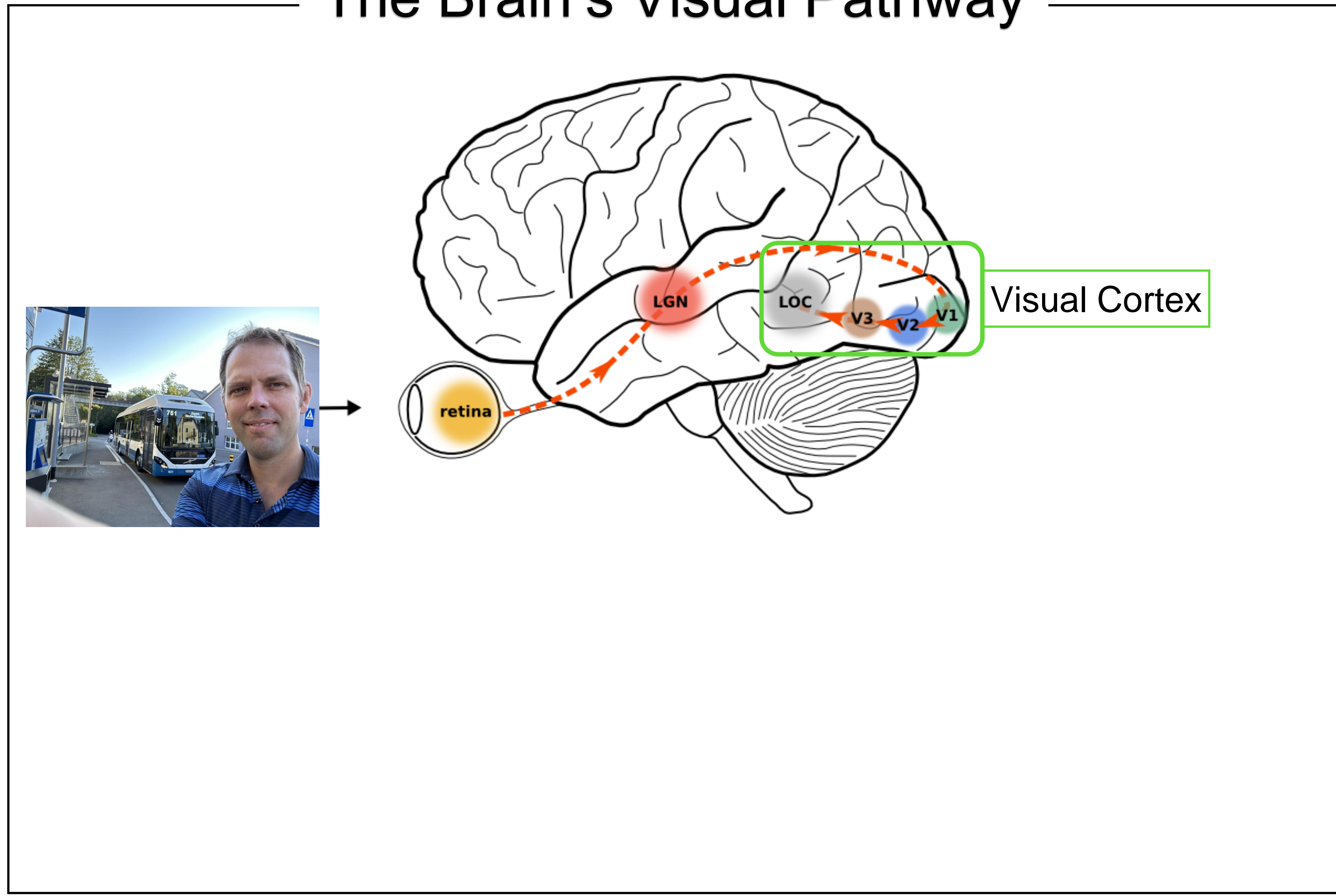
Why?

Part I: Learning in Hierarchical (Deep) Cortical Networks**Engineering:**

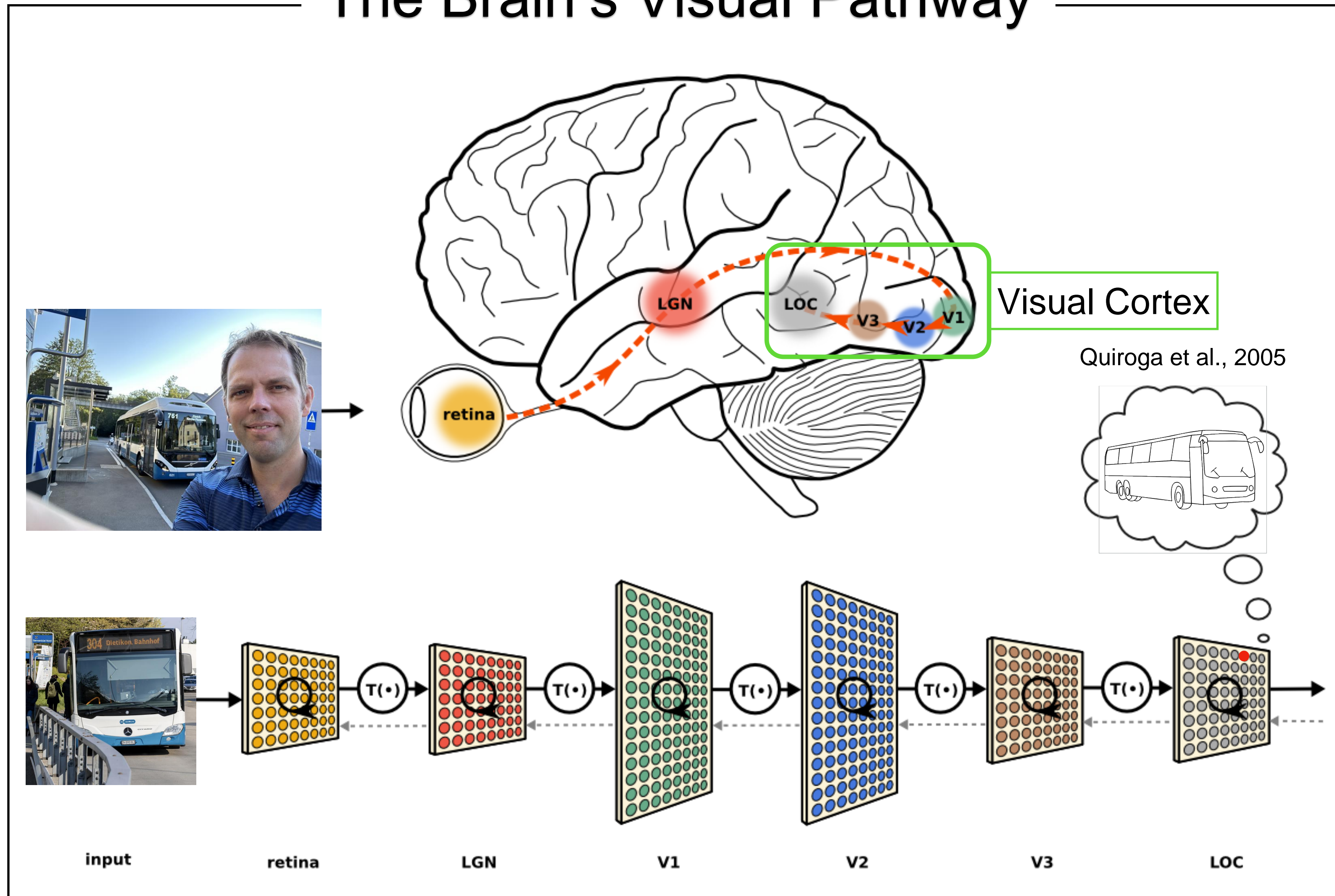
- To develop Deep Networks that are more energy efficient.
- To train Deep Networks with less training data.
- To enable Deep Continual Learning.
- To train Deep Networks with dynamic data.

Neuroscience:

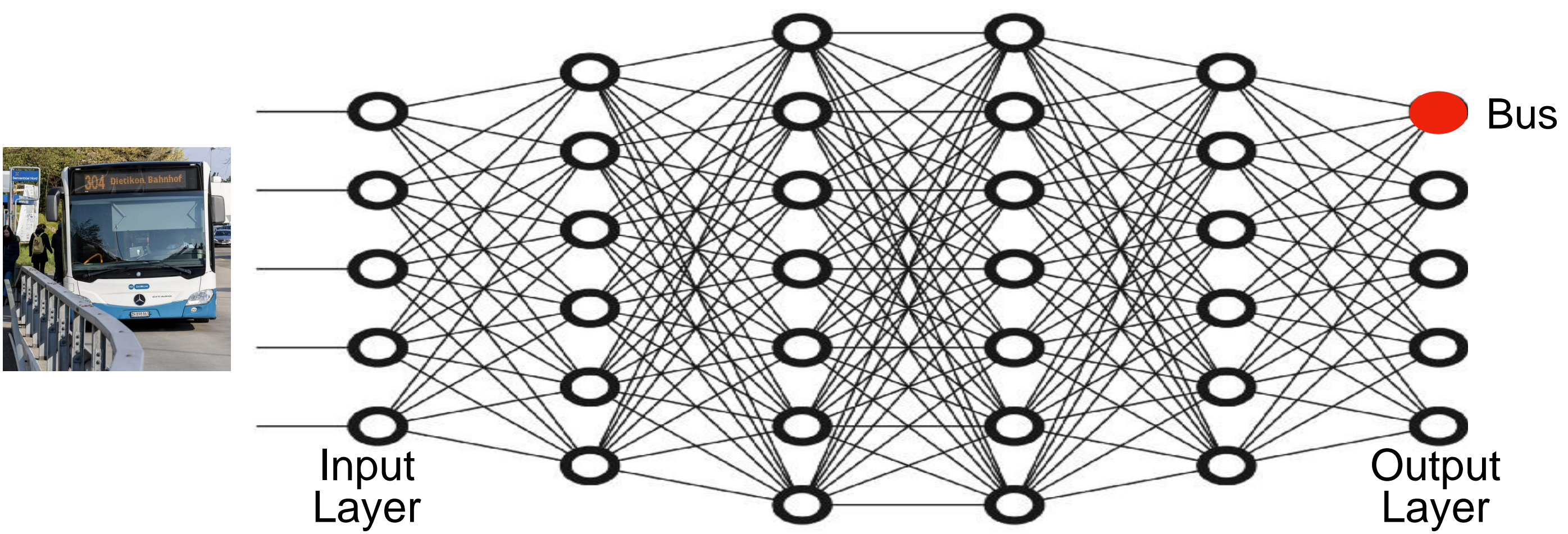
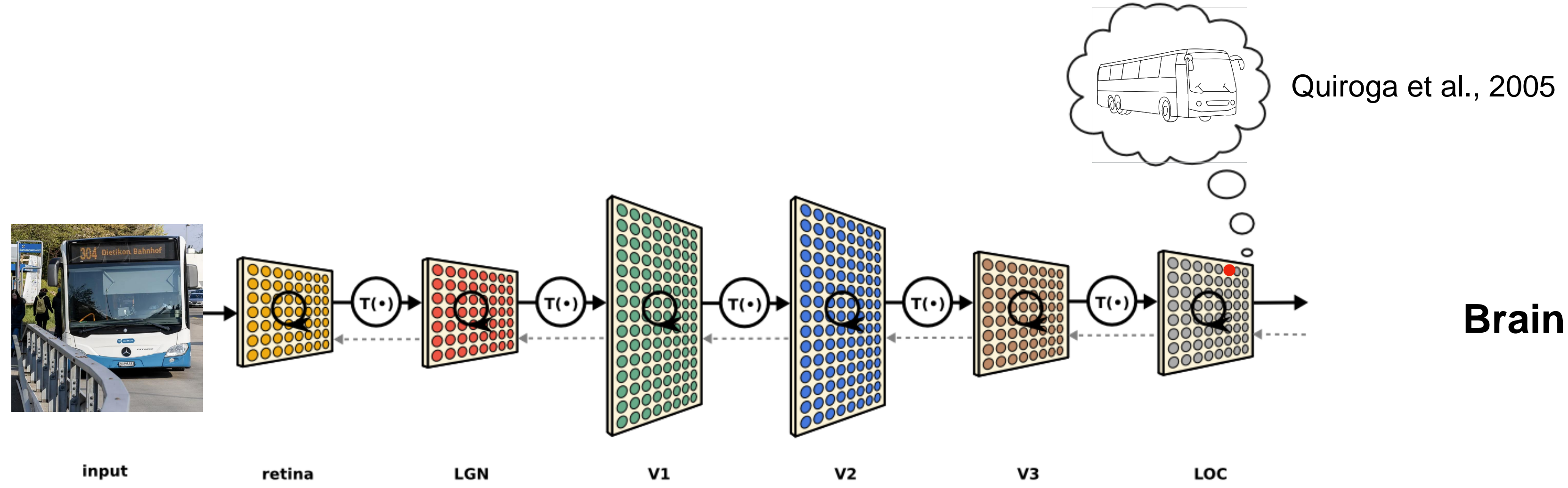
- We currently cannot use deep learning algorithms to explain how credit assignment in the brain works.

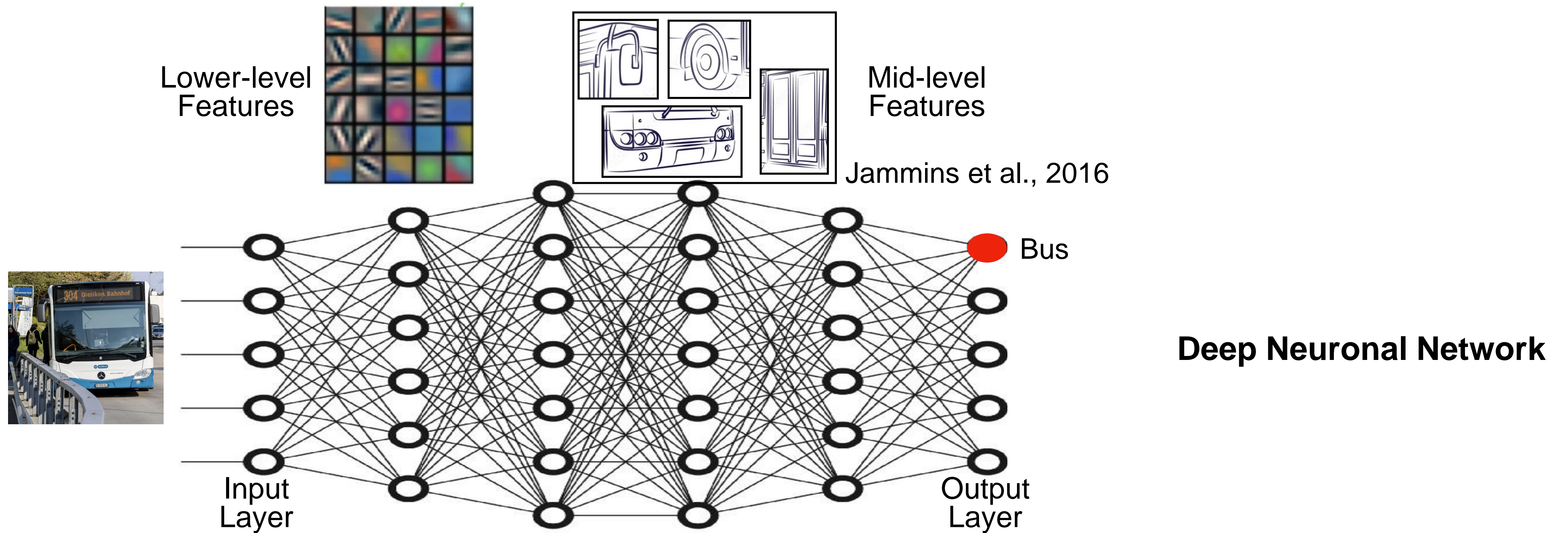


LGN: Lateral Geniculate Nucleus
V1: Primary Visual Cortex
V2: Secondary Visual Cortex
V3: Third Visual Cortex
LOC: Lateral Occipital Cortex

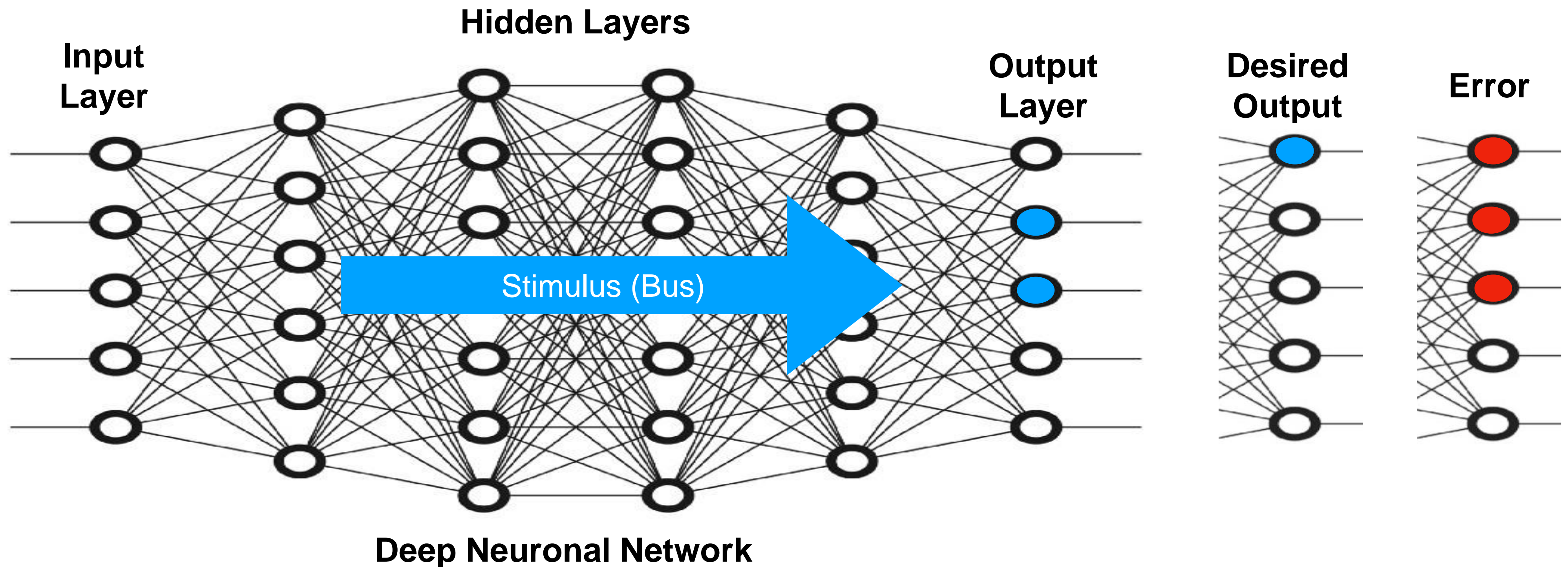


LGN: Lateral Geniculate Nucleus
 V1: Primary Visual Cortex
 V2: Secondary Visual Cortex
 V3: Third Visual Cortex
 LOC: Lateral Occipital Cortex





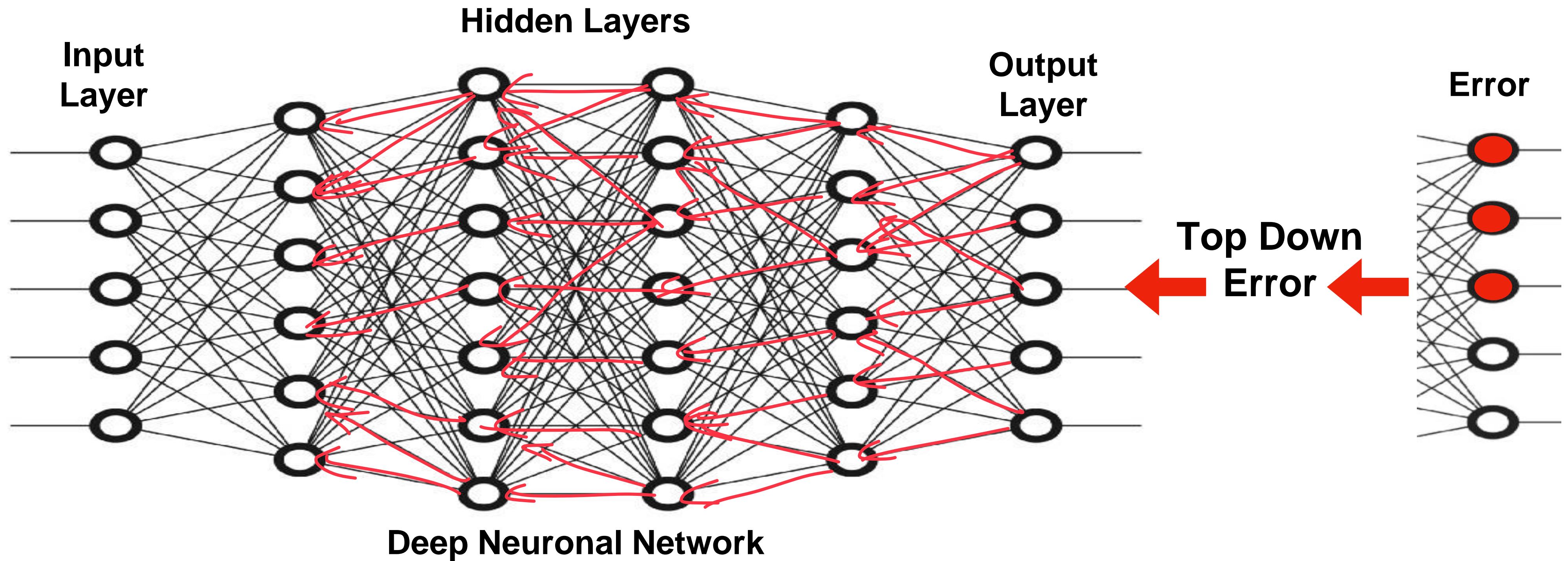
1. The Forward Pass



Background and Motivation - Part I

The Error Backpropagation Method

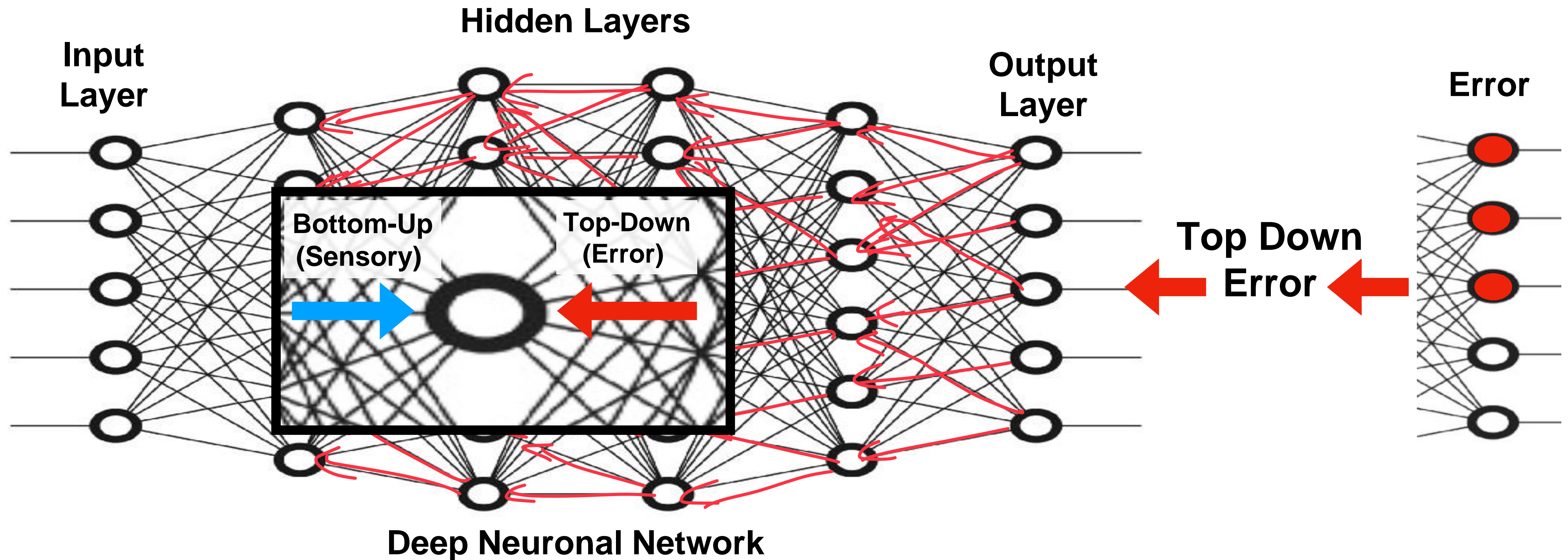
1. The Forward Pass
2. The Backward (Error) Pass



Background and Motivation - Part I

The Error Backpropagation Method

1. The Forward Pass
2. The Backward (Error) Pass

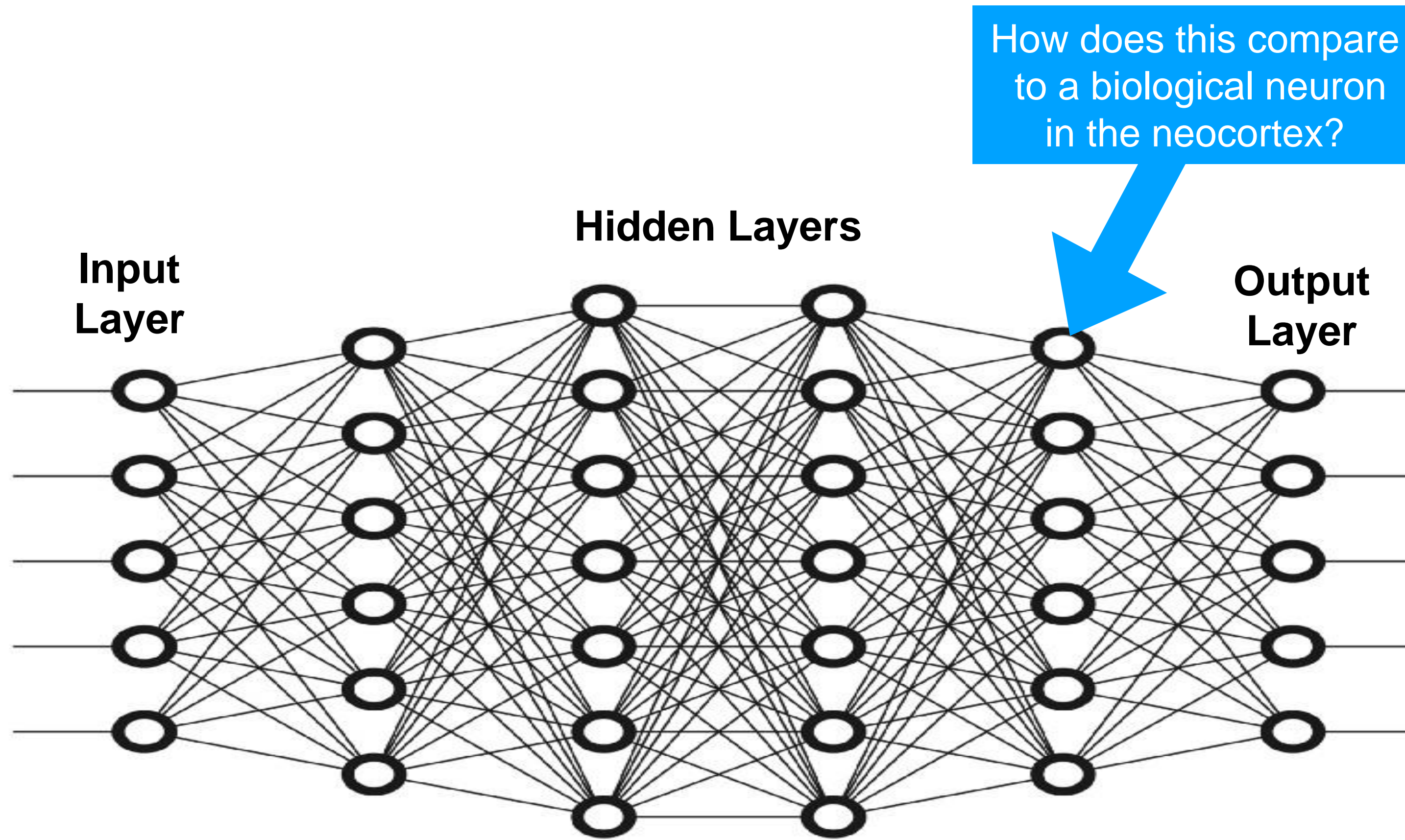


Why does Backpropagation not explain how credit assignment in biology works?

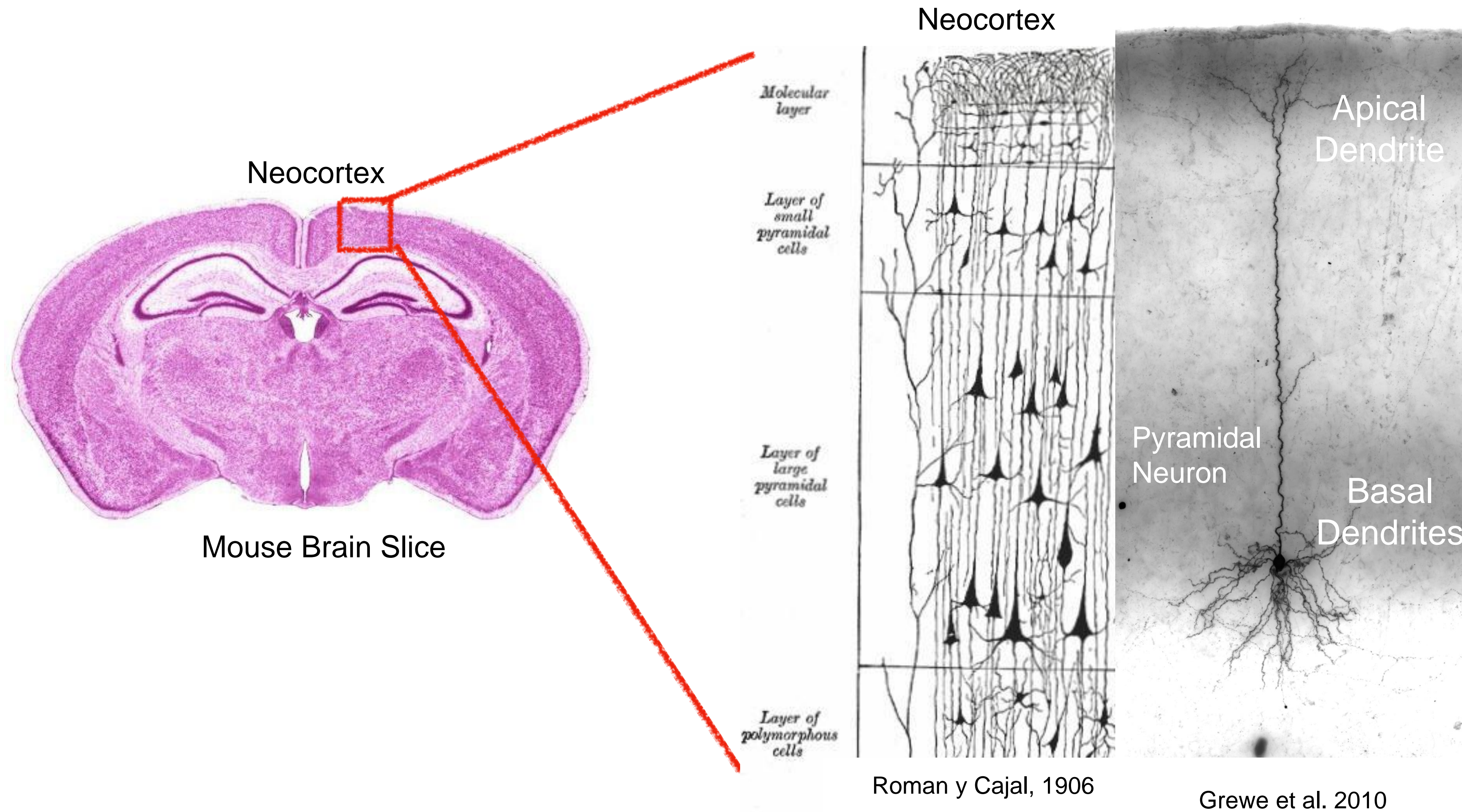
- Requires separate forward and backward phases.
- Sends sensory information forward, but error backward.
- Sends errors backward through the SAME weights.
- Is based on discrete stepwise computations.
- Neuron update/plasticity not biologically observed.

Background and Motivation - Part I

Seeking Inspiration from Biology



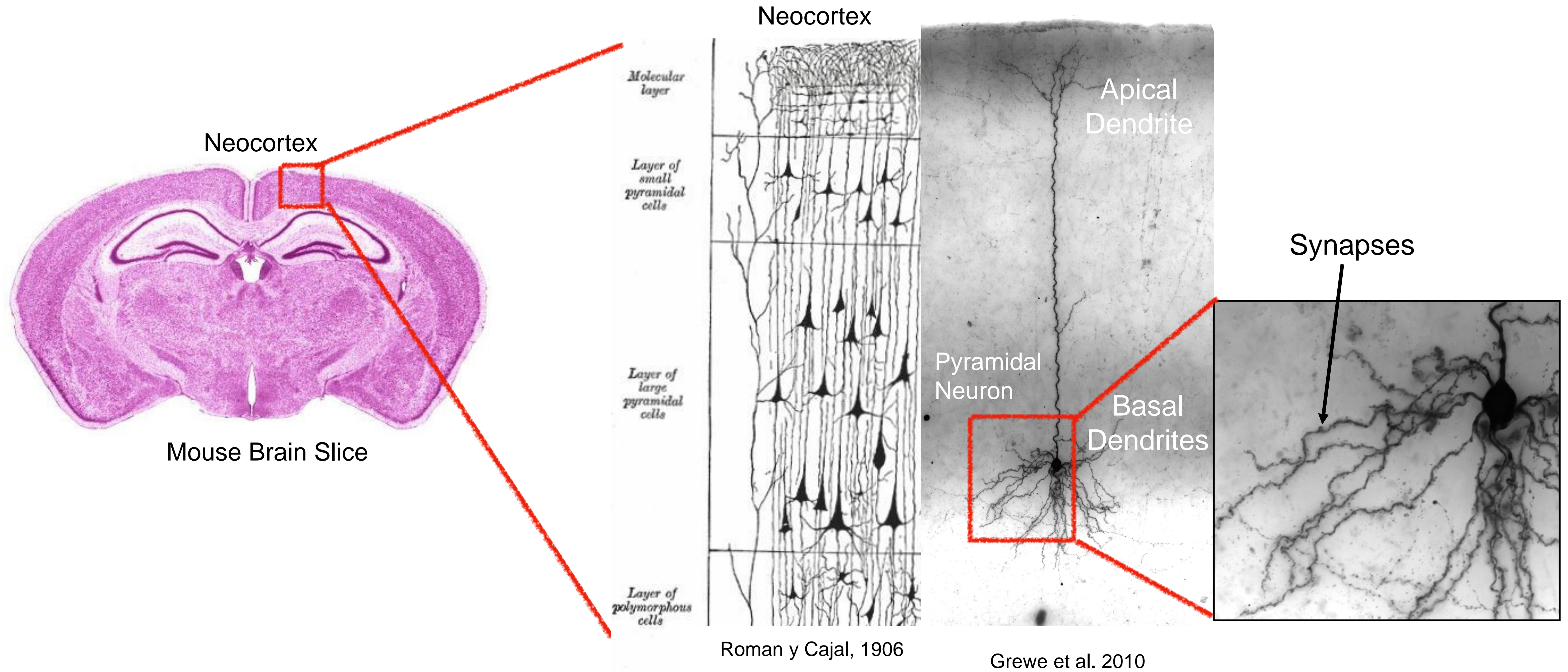
Background and Motivation - Part I Seeking Inspiration from Biology



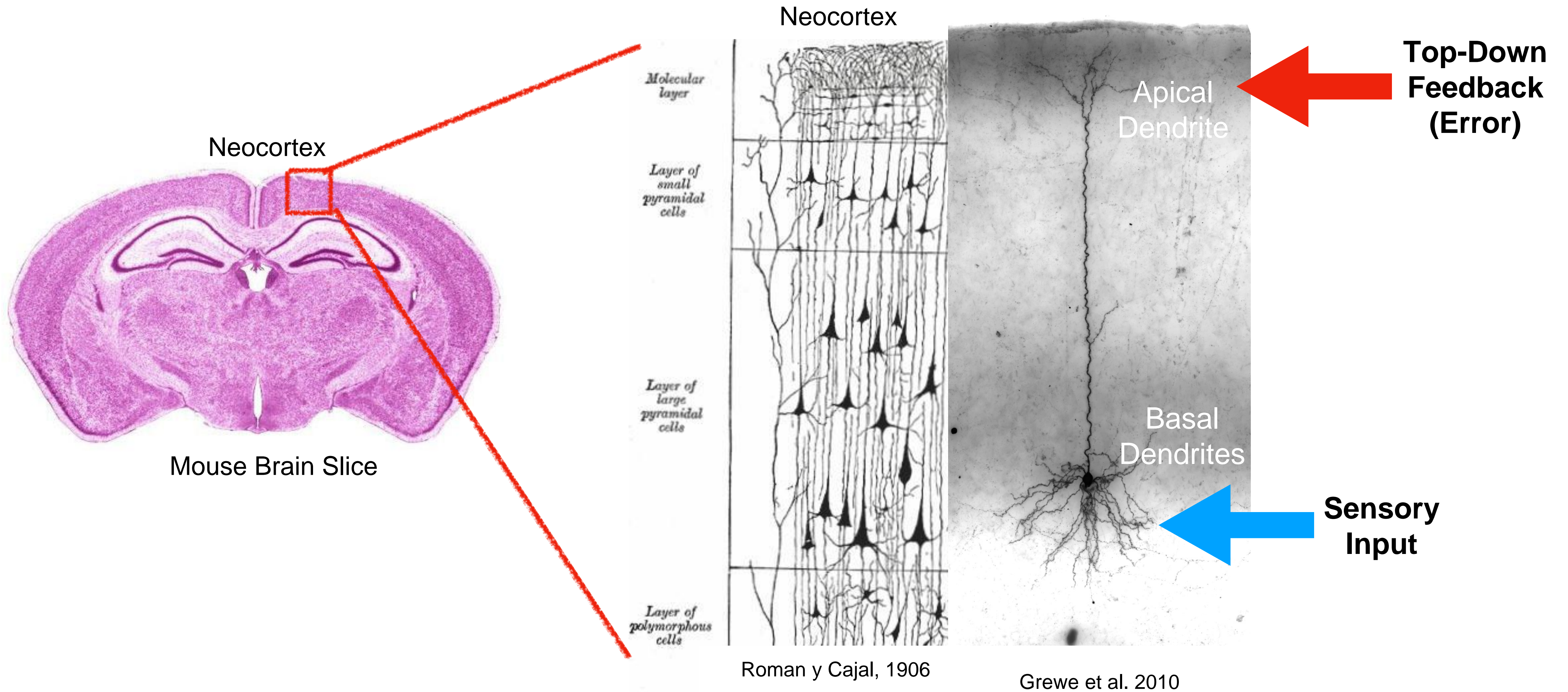
Roman y Cajal, 1906

Grewe et al. 2010

Background and Motivation - Part I Seeking Inspiration from Biology



Background and Motivation - Part I Seeking Inspiration from Biology

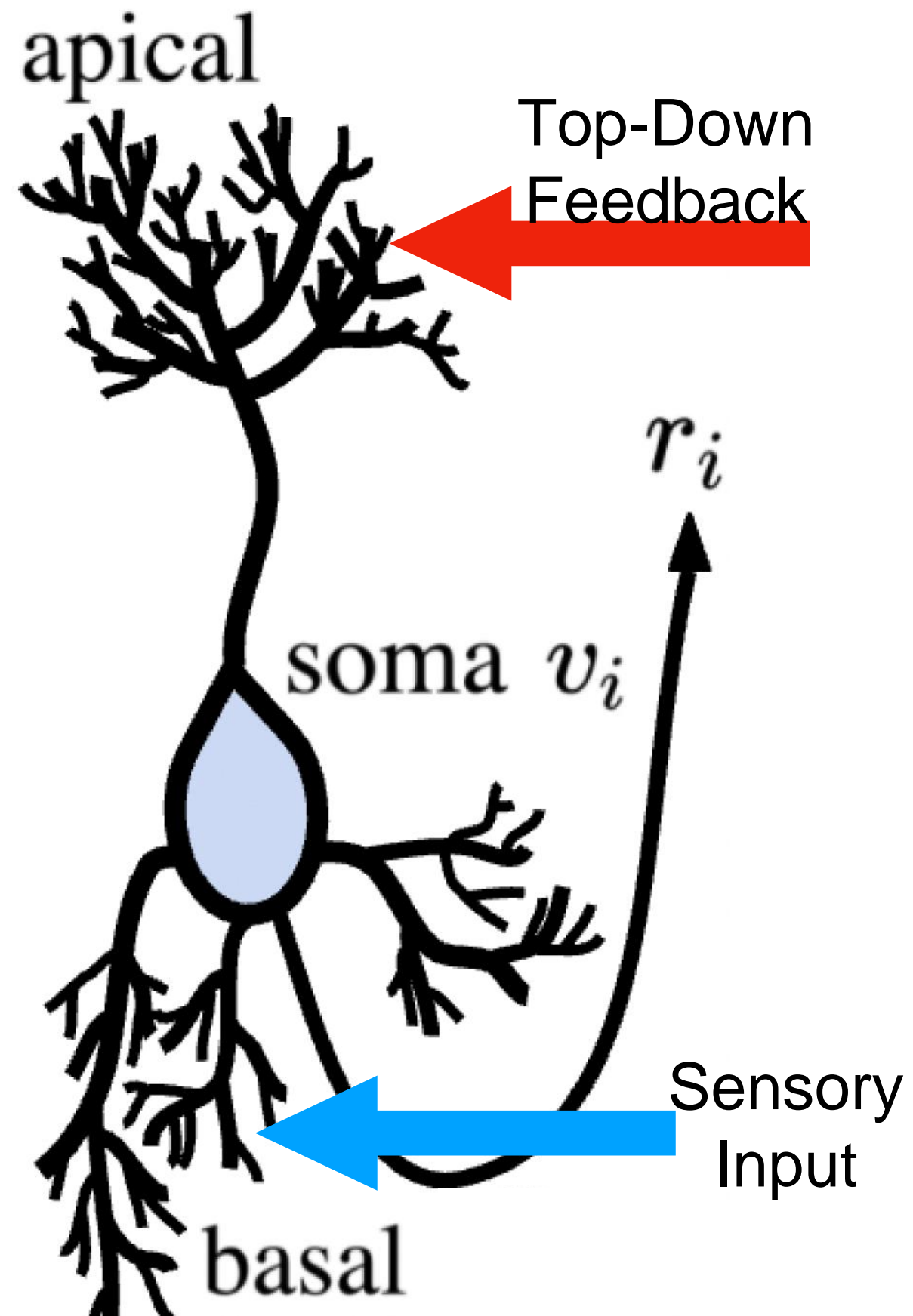




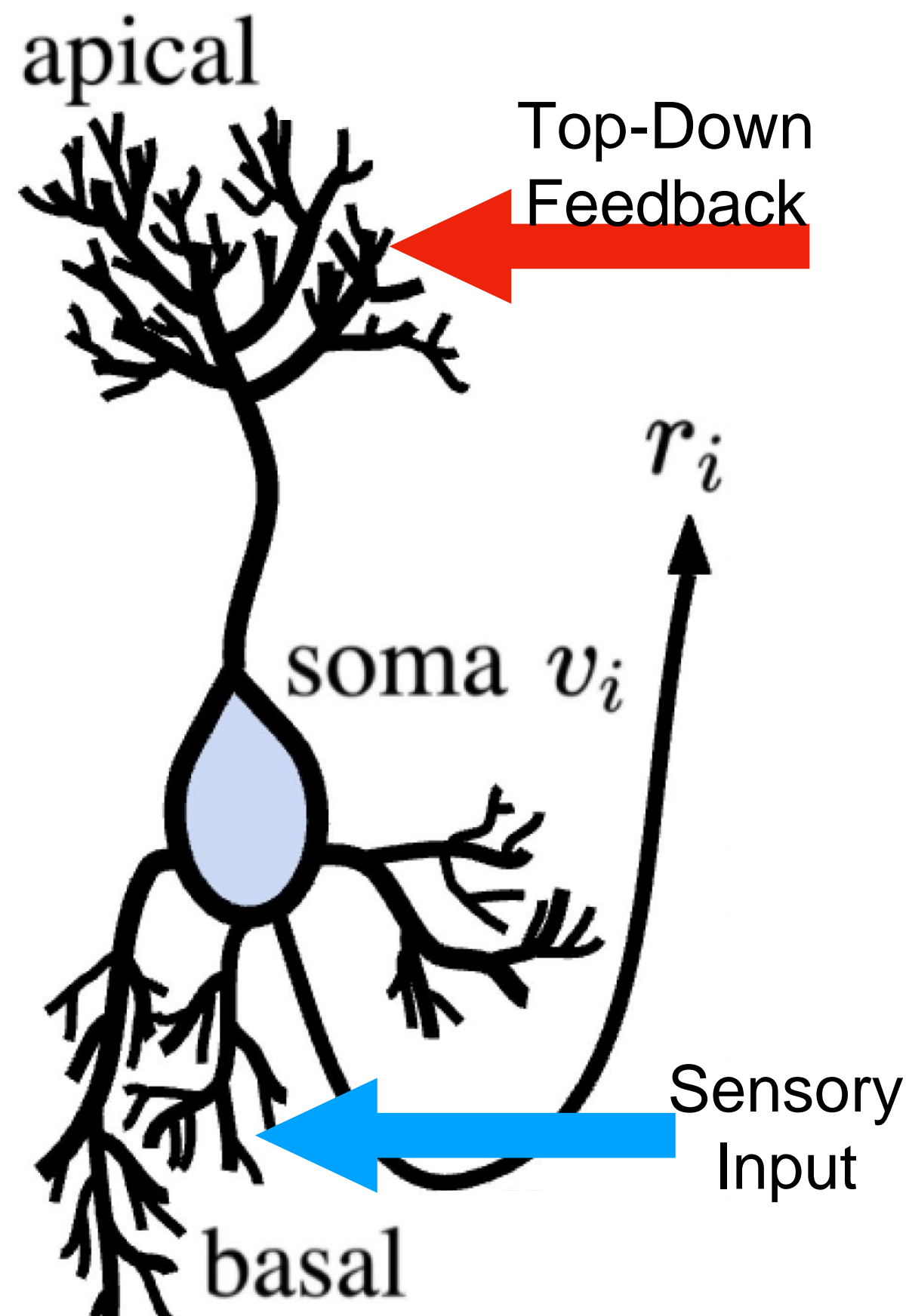
Matilde T. Farinha



Alexander Meulemans



v_i membrane potential
 r_i neuron firing rate

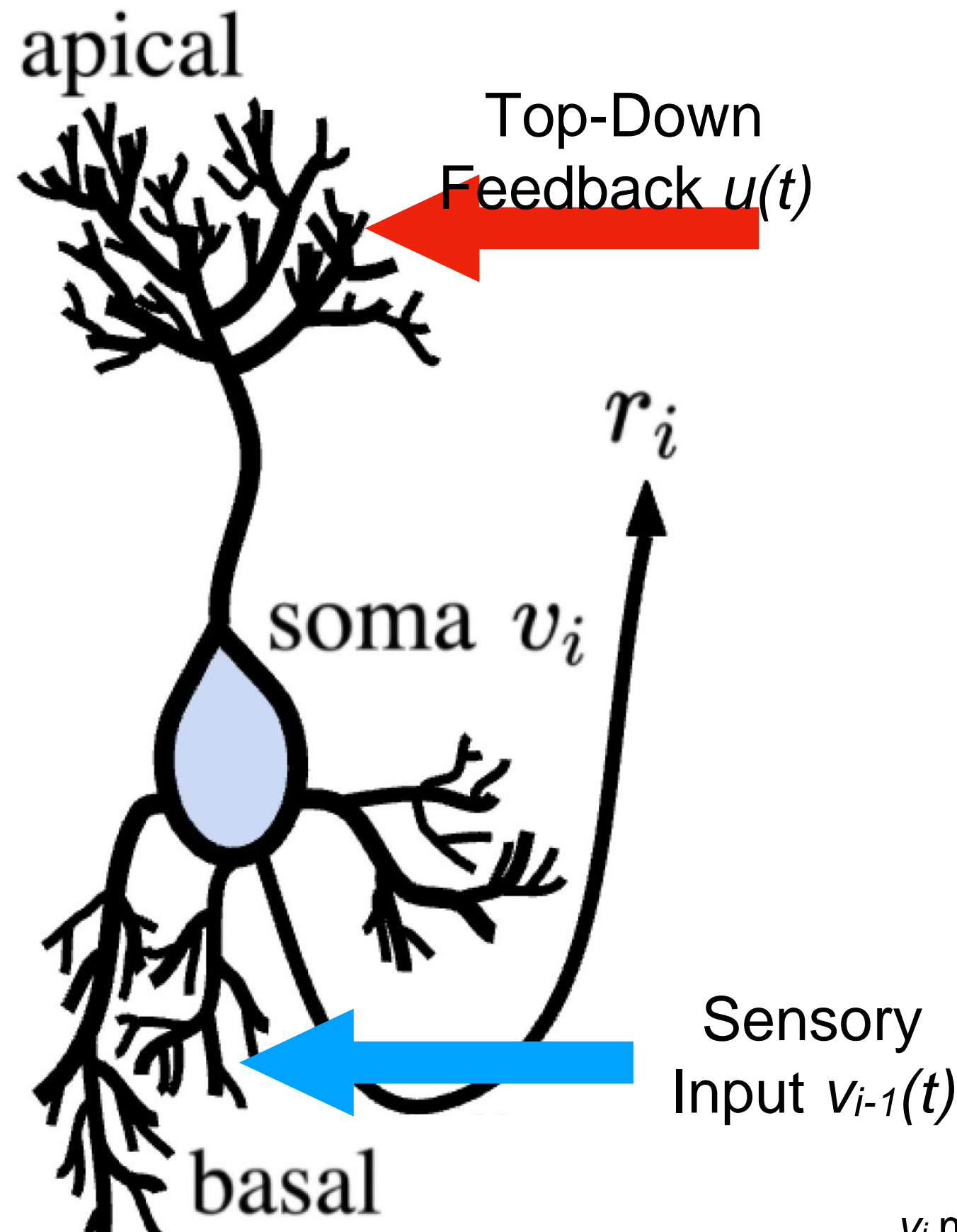


Neuron Output /
Firing Rate

$$r_i = \phi(v_i)$$

Non.
Linearity

v_i membrane potential
 r_i neuron firing rate



Neuron Output / Firing Rate

$$\mathbf{r}_i = \phi(\mathbf{v}_i)$$

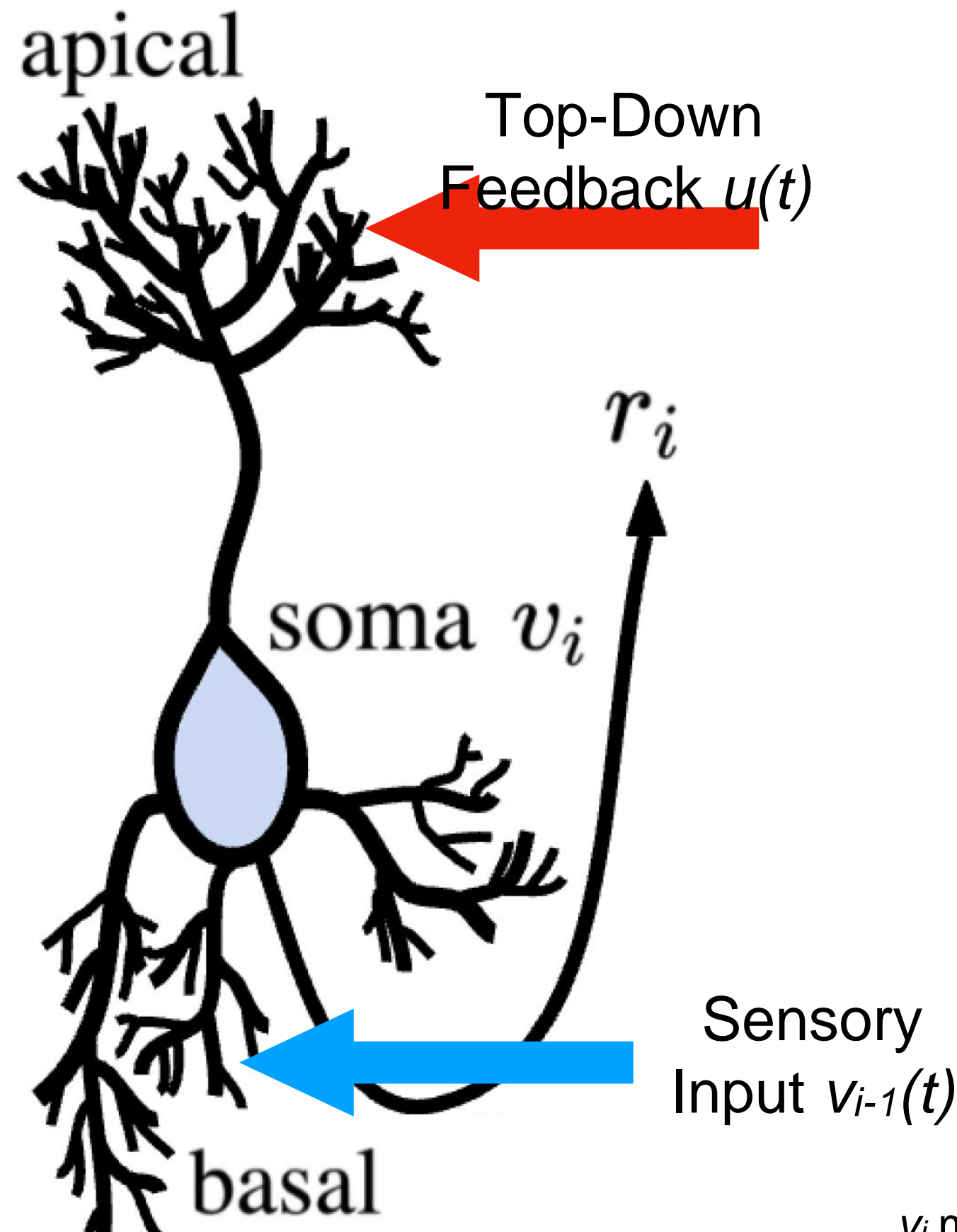
Non-Linearity

Membrane Potential Dynamics

$$\tau_v \frac{d}{dt} \mathbf{v}_i(t) = -\mathbf{v}_i(t) + \underbrace{W_i \phi(\mathbf{v}_{i-1}(t))}_{\text{Sensory}} + \underbrace{Q_i \mathbf{u}(t)}_{\text{Feedback}}$$

v_i membrane potential
 r_i neuron firing rate

$u(t)$ Feedback signal
 Q_i Feedback weights



Neuron Output / Firing Rate

$$\mathbf{r}_i = \phi(\mathbf{v}_i)$$

Non-Linearity

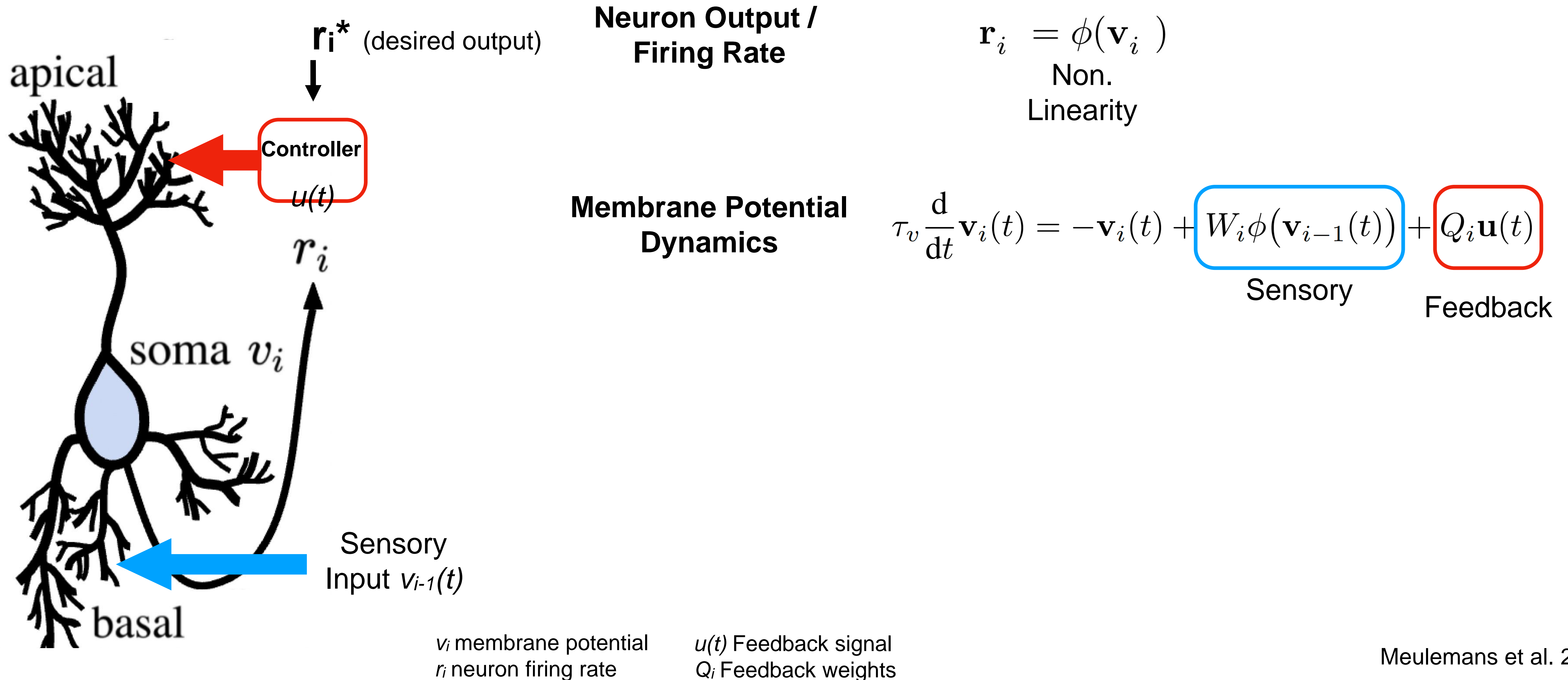
Membrane Potential Dynamics

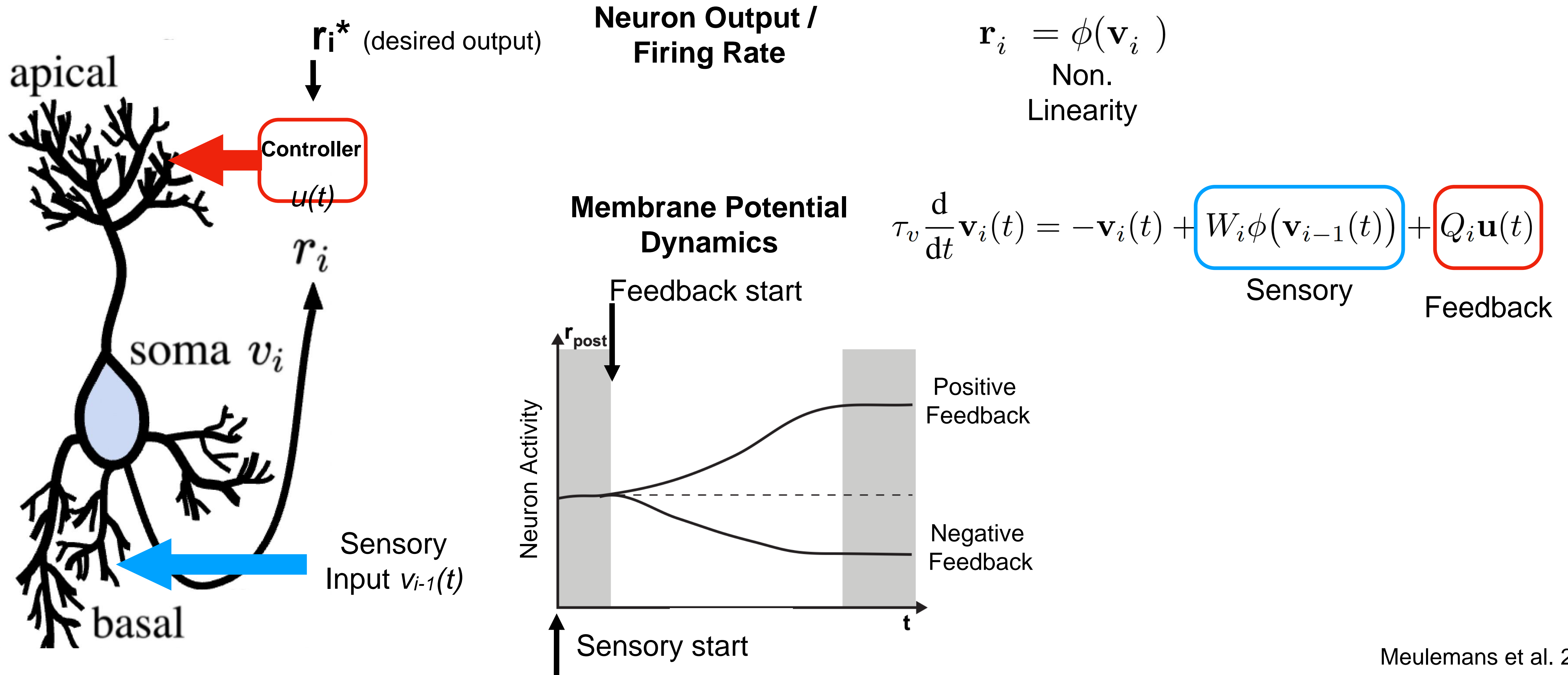
$$\tau_v \frac{d}{dt} \mathbf{v}_i(t) = -\mathbf{v}_i(t) + \underbrace{W_i \phi(\mathbf{v}_{i-1}(t))}_{\text{Sensory}} + \underbrace{Q_i \mathbf{u}(t)}_{\text{Feedback}}$$

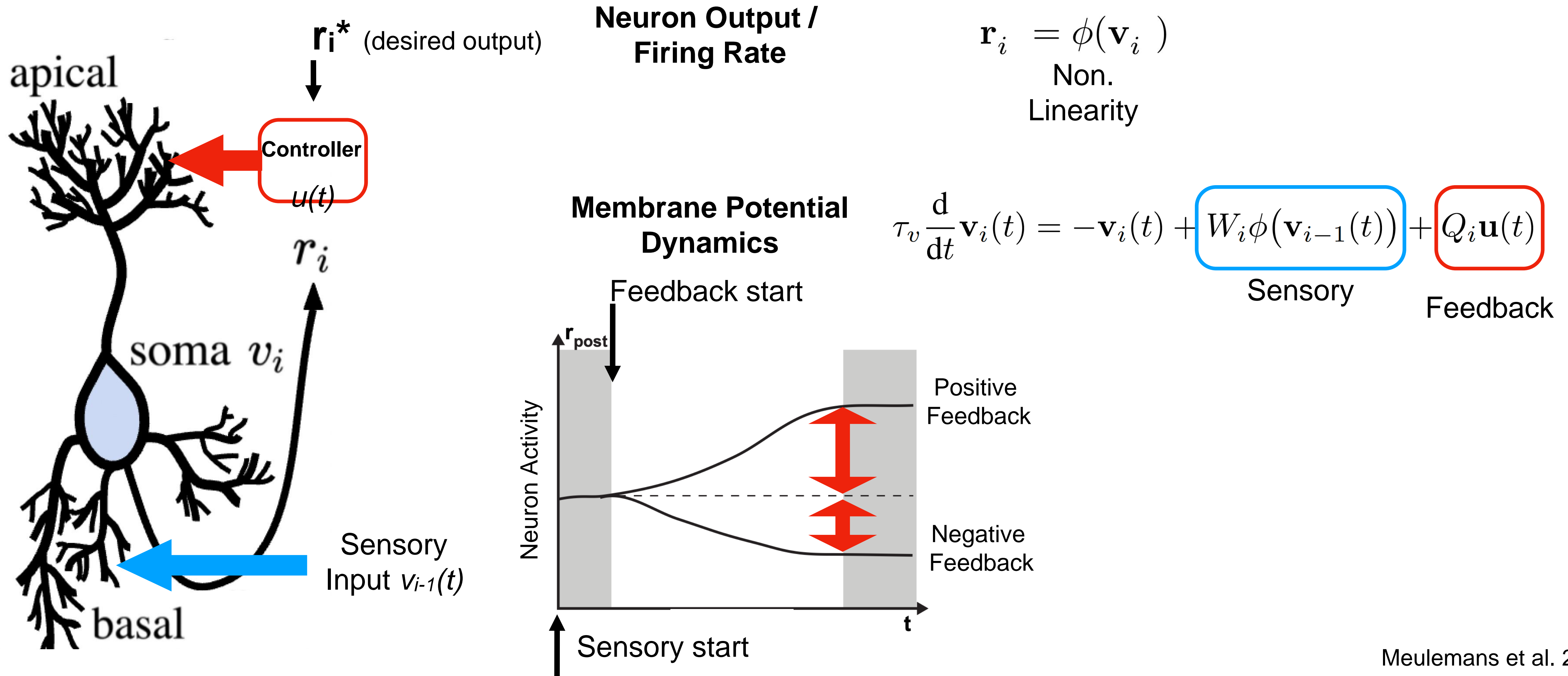
How can we train this single neuron to detect a specific sensory input pattern?

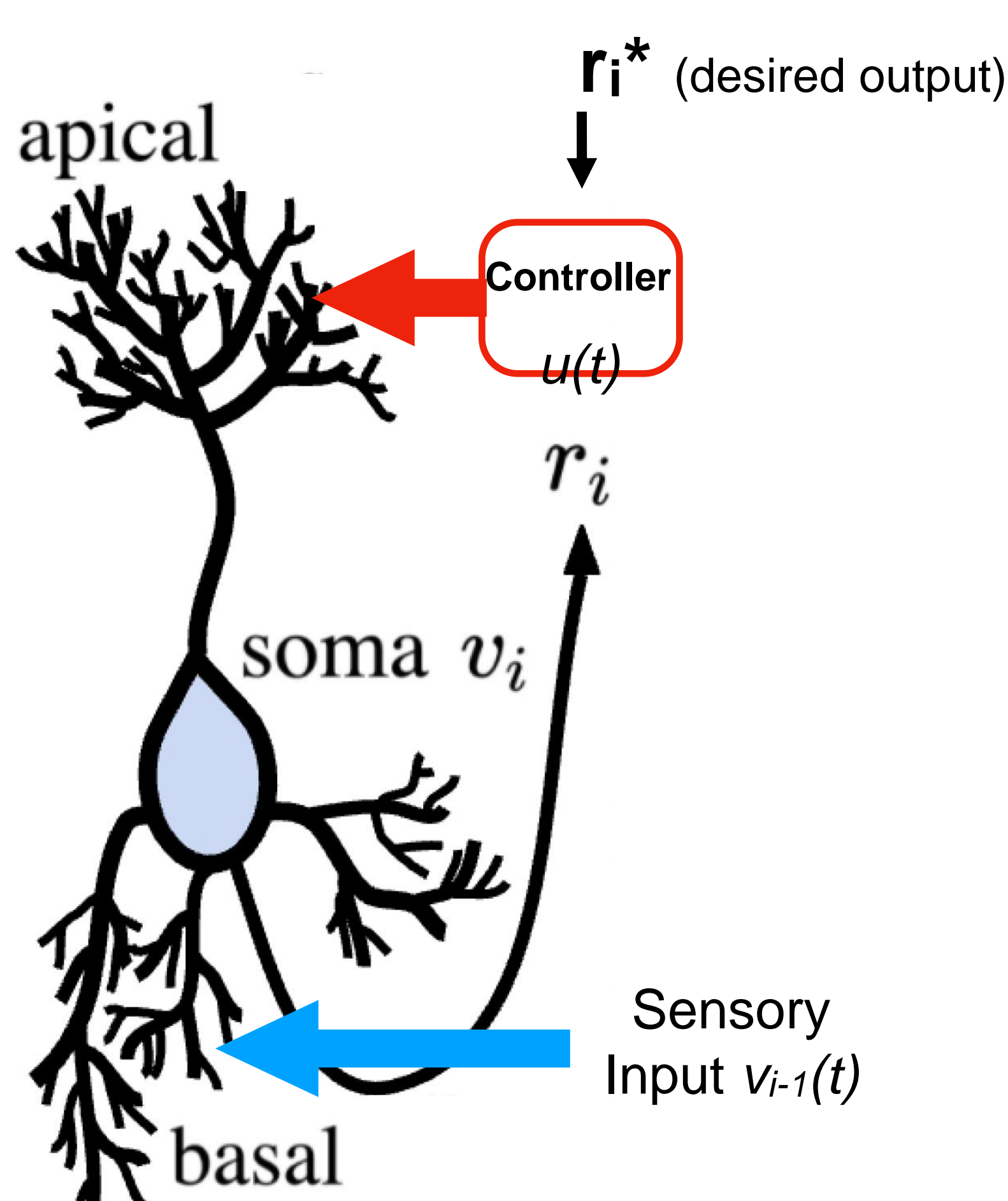
v_i membrane potential
 r_i neuron firing rate

$u(t)$ Feedback signal
 Q_i Feedback weights









Neuron Output / Firing Rate

$$\mathbf{r}_i = \phi(\mathbf{v}_i)$$

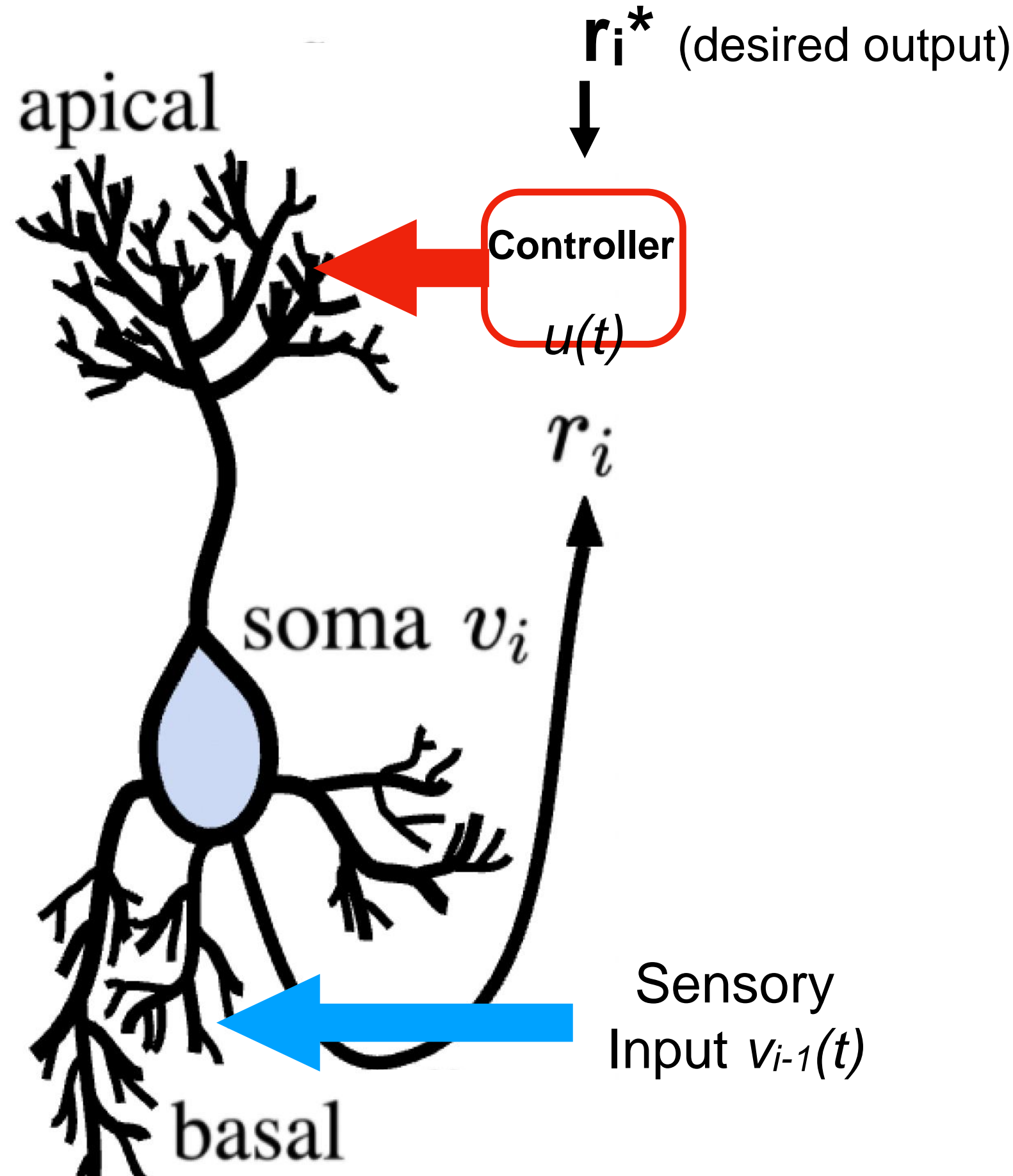
Non-Linearity

Membrane Potential Dynamics

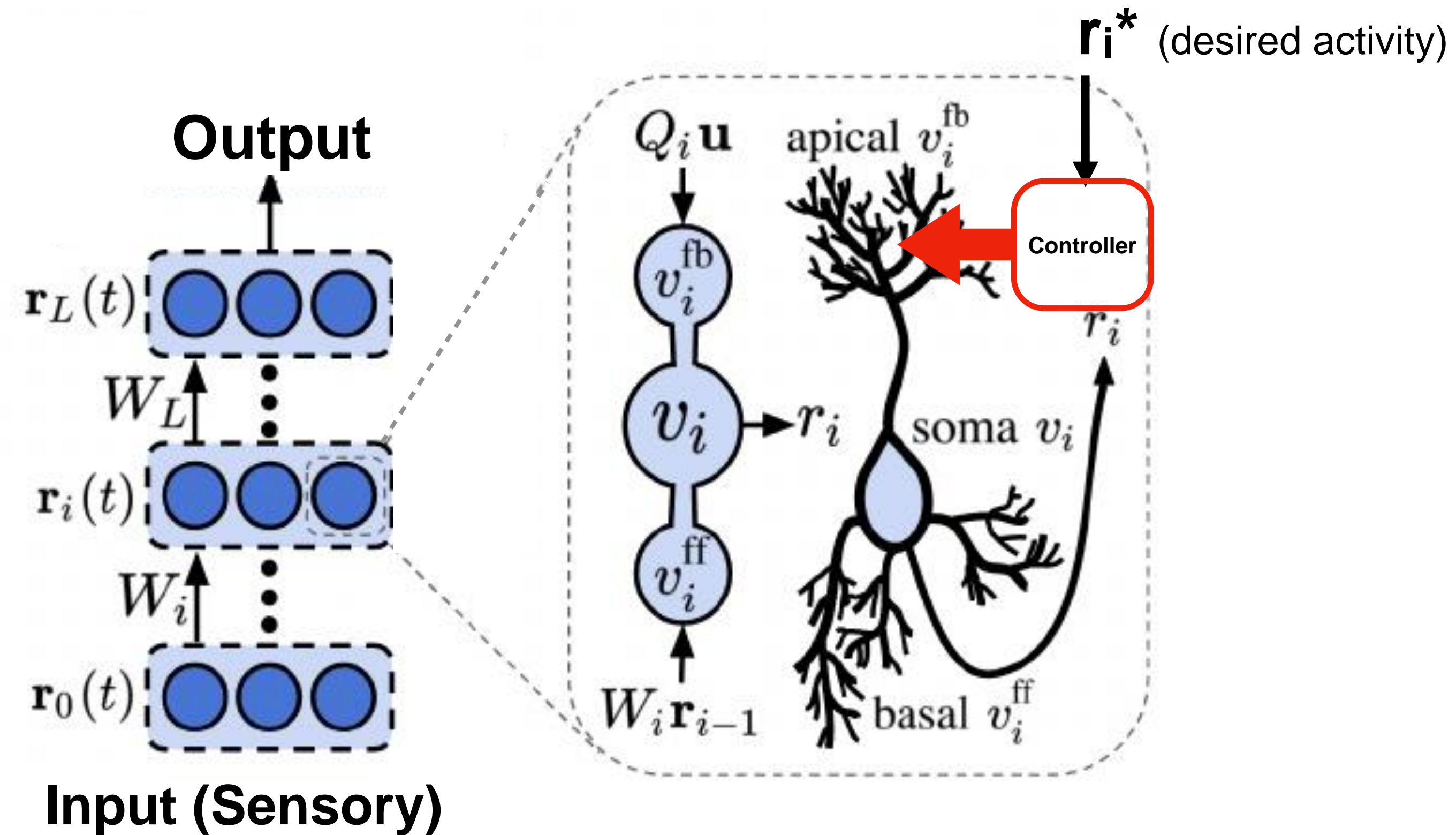
$$\tau_v \frac{d}{dt} \mathbf{v}_i(t) = -\mathbf{v}_i(t) + \underbrace{W_i \phi(\mathbf{v}_{i-1}(t))}_{\text{Sensory}} + \underbrace{Q_i \mathbf{u}(t)}_{\text{Feedback}}$$

Synaptic Weight Update

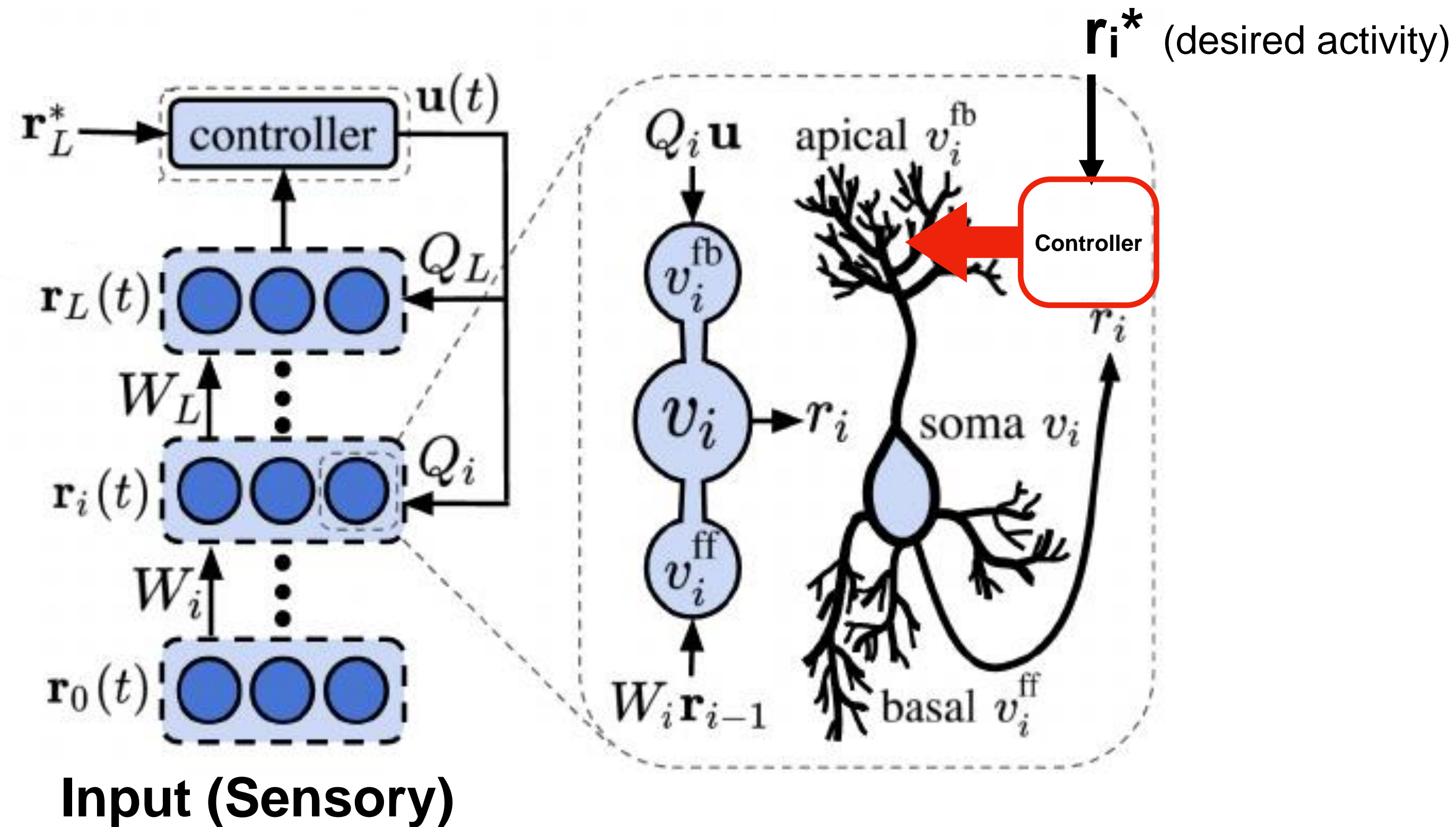
$$\tau_W \frac{d}{dt} W_i(t) = \underbrace{(\phi(\mathbf{v}_i(t)))}_{\text{Activity w/o Feedback}} - \underbrace{\phi(W_i \mathbf{r}_{i-1}(t))}_{\text{Activity with Feedback}} \underbrace{\mathbf{r}_{i-1}(t)^T}_{\text{Presynaptic Term}}$$

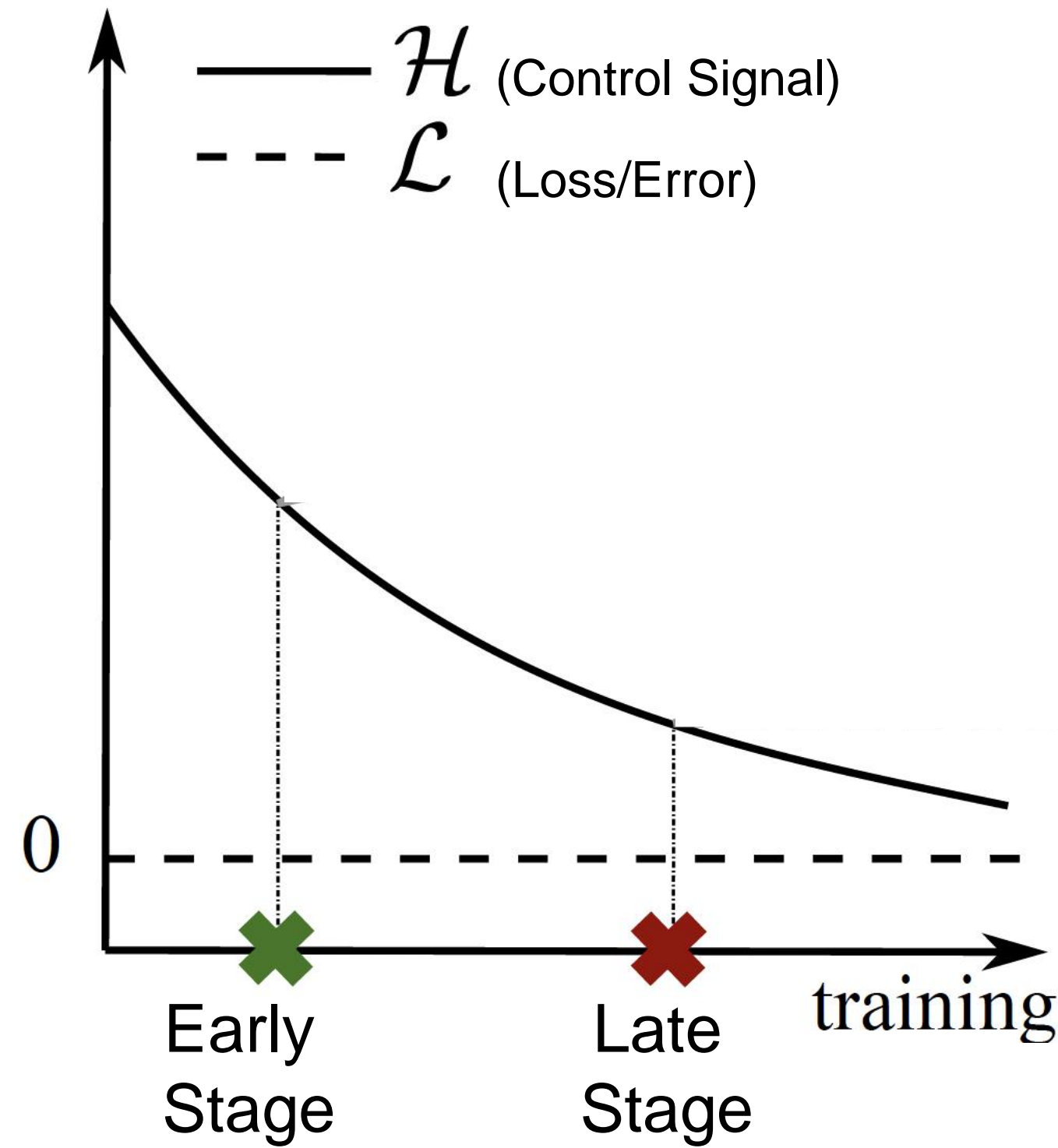
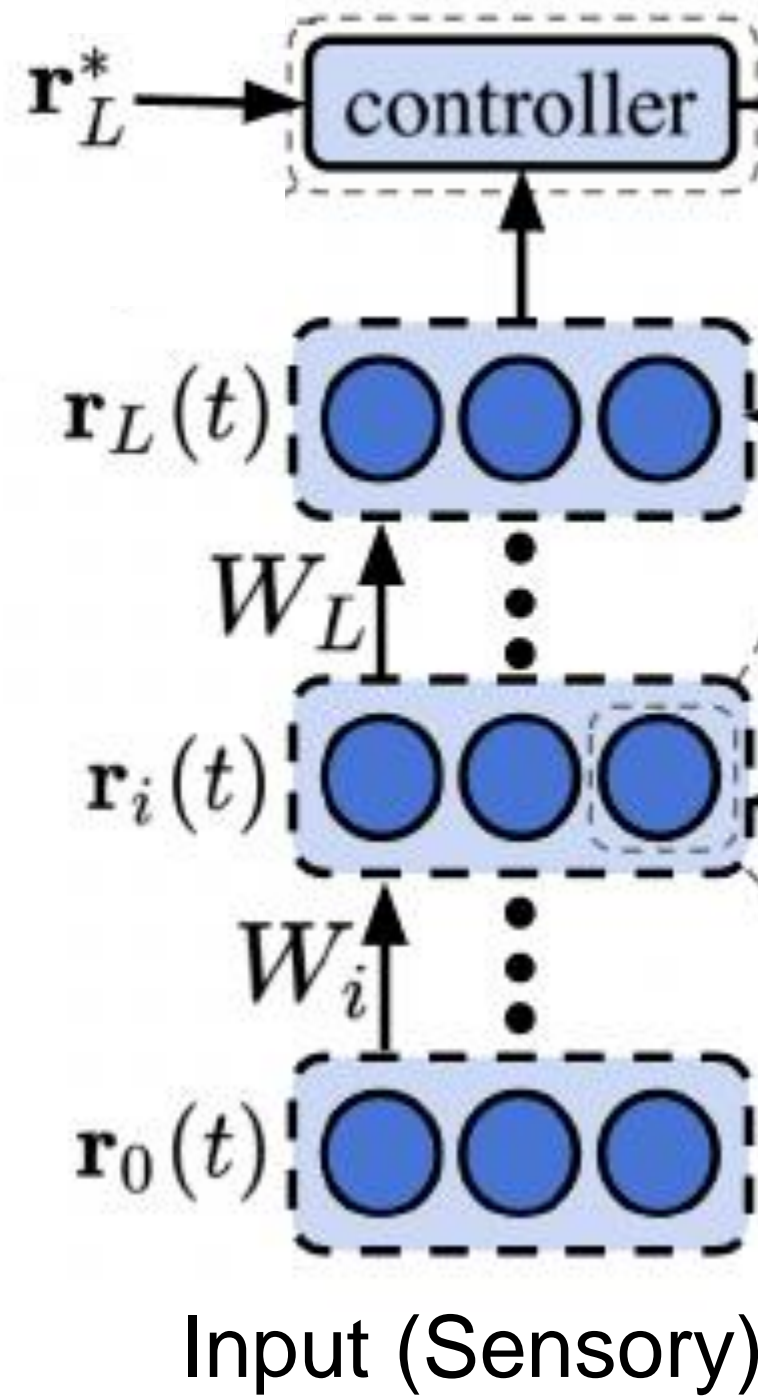


<p>Neuron Output / Firing Rate</p>	$\mathbf{r}_i = \phi(\mathbf{v}_i)$ <p>Non-Linearity</p>	<p>Neuron Activity</p>
<p>Membrane Potential Dynamics</p>	$\tau_v \frac{d}{dt} \mathbf{v}_i(t) = -\mathbf{v}_i(t) + \underbrace{W_i \phi(\mathbf{v}_{i-1}(t))}_{\text{Sensory}} + \underbrace{Q_i \mathbf{u}(t)}_{\text{Feedback}}$	
<p>Synaptic Weight Update</p> <p>Neuron Plasticity</p>	$\tau_W \frac{d}{dt} W_i(t) = \underbrace{(\phi(\mathbf{v}_i(t)))}_{\text{Activity w/o Feedback}} - \underbrace{\phi(W_i \mathbf{r}_{i-1}(t))}_{\text{Activity with Feedback}} \underbrace{\mathbf{r}_{i-1}(t)^T}_{\text{Presynaptic Term}}$	



We solve a complex control problem!



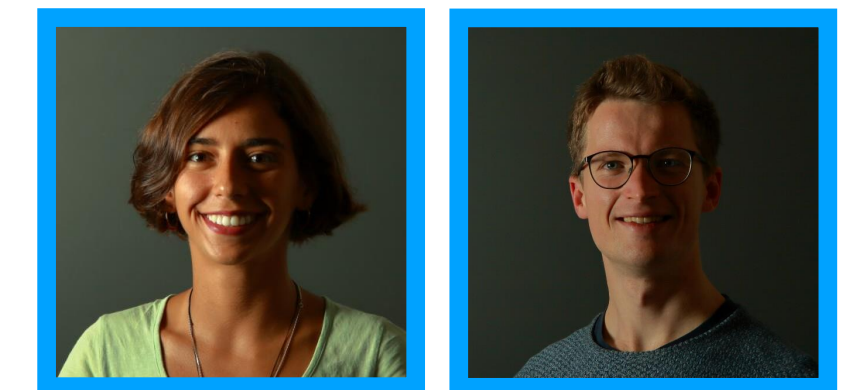


Learning = Reducing Help!



Checkpoint: How Bio-Plausible are we?

1. We no longer use separate forward and backward phases. ✓
2. We don't send sensory information forward and errors backward. ✓
3. We don't send feedback signals though the SAME weights. ✓
4. We allow continuous (in time) computation. ✓
5. Our update/plasticity rule still not biologically plausible. ✗

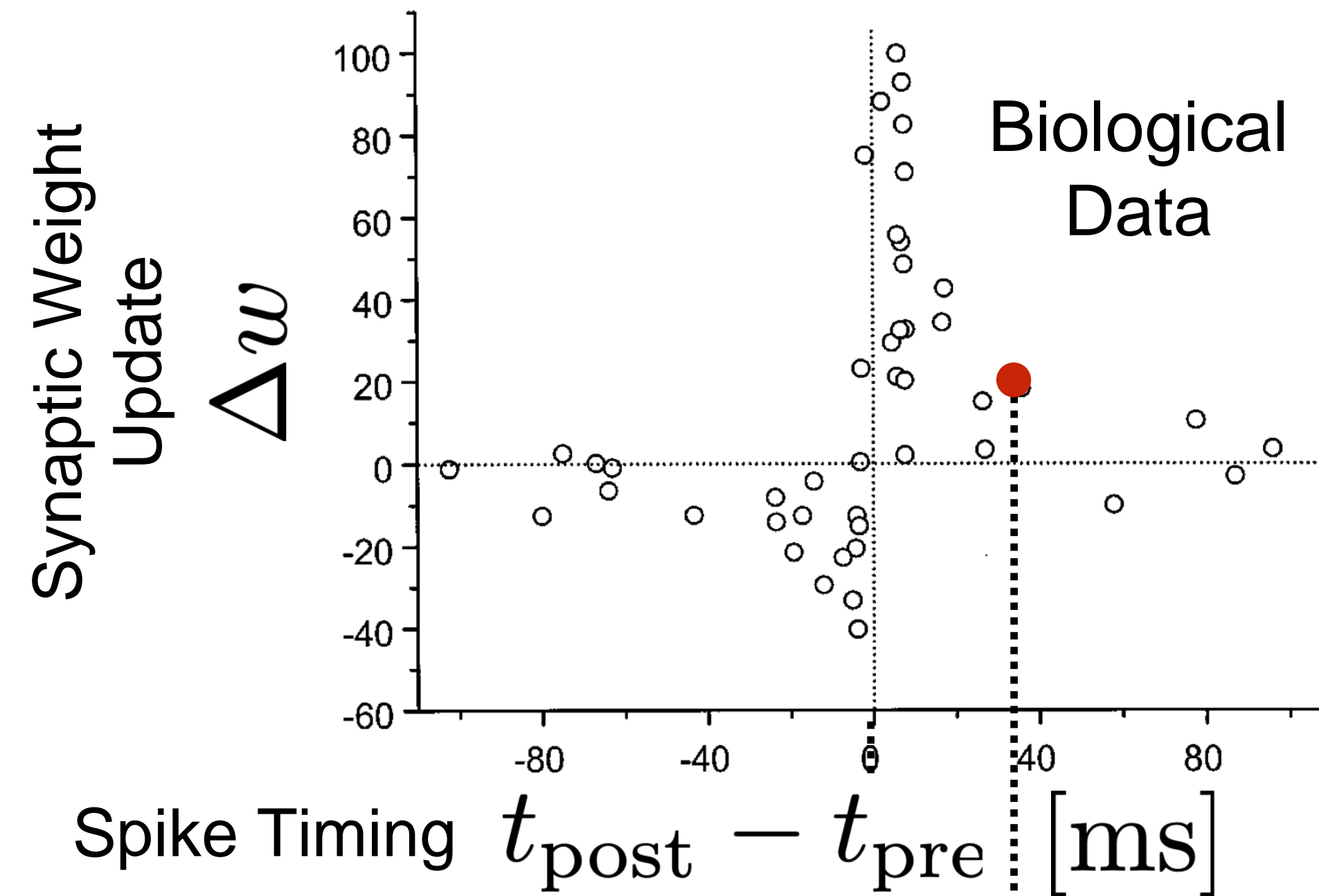


Spotlight Awards
(Best 1-2%)
* *

Meulemans et al. 2020, 2021, 2022

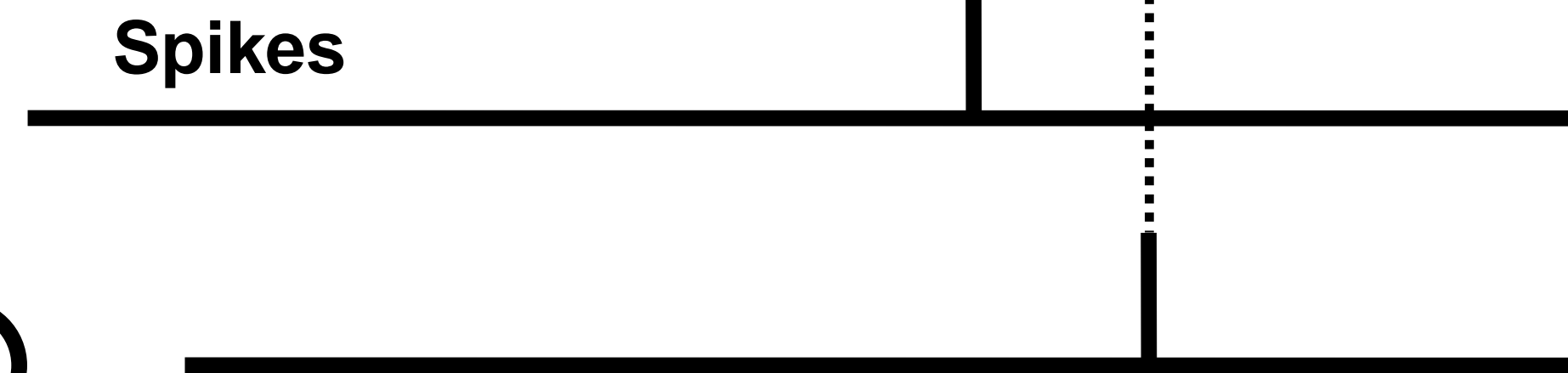
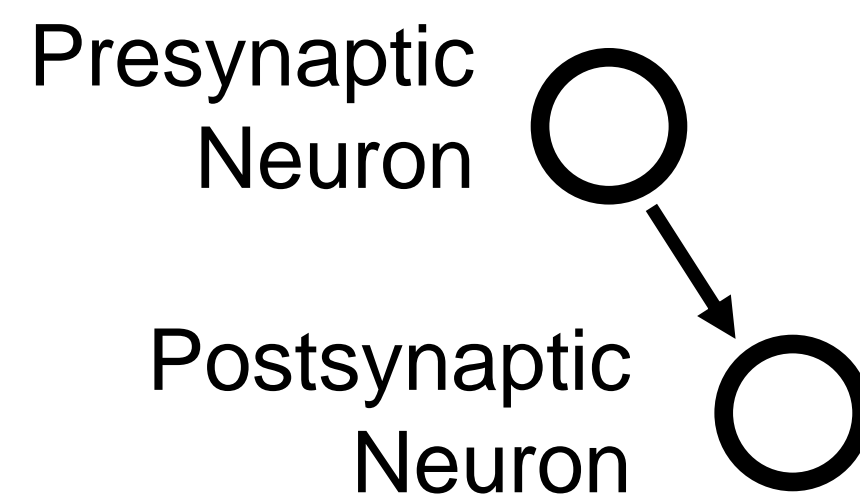
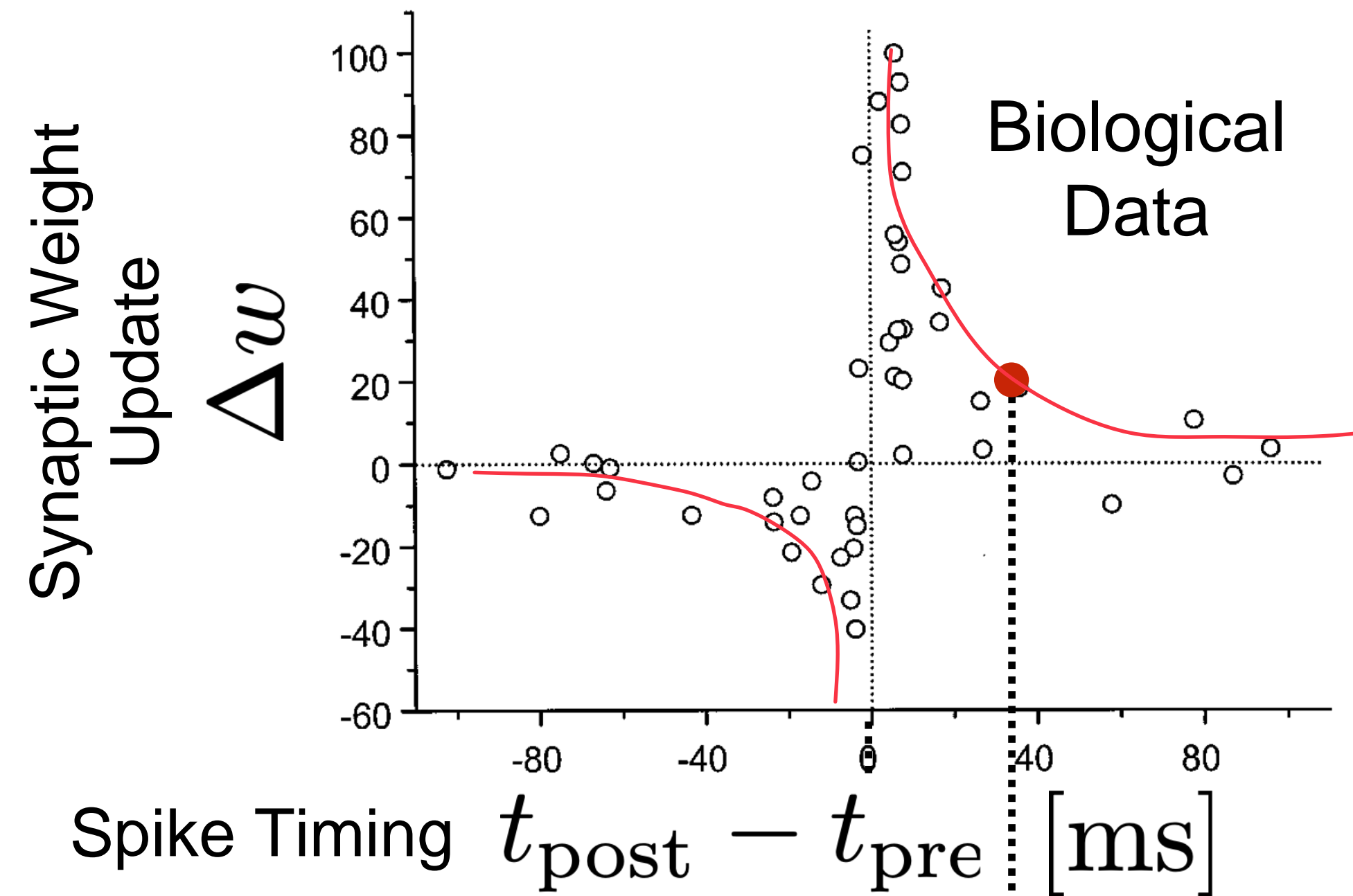
Neuronal Plasticity in Biology

Spike Timing Dependent Plasticity (STDP)



Neuronal Plasticity in Biology

Spike Timing Dependent Plasticity (STDP)

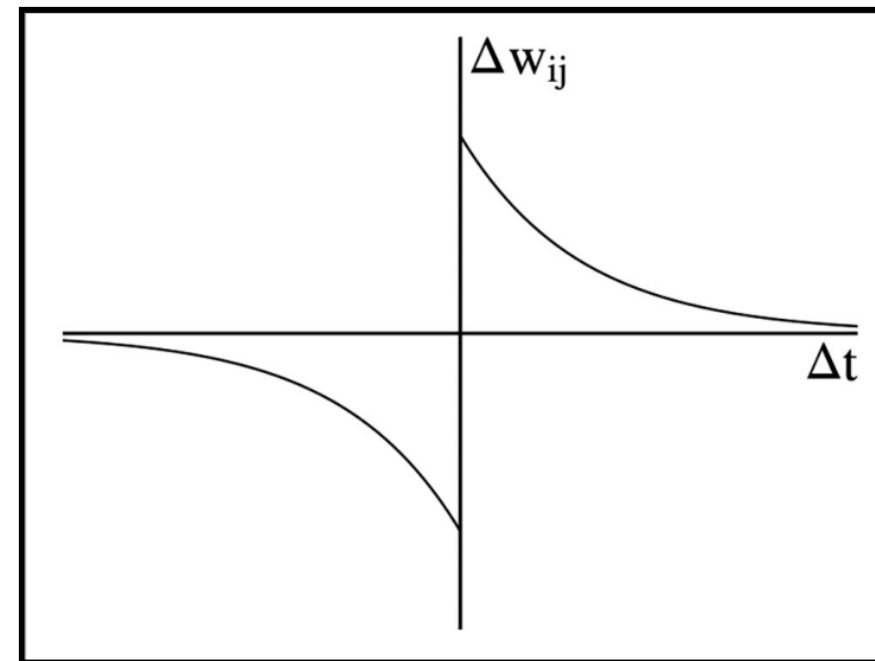


Mimicking Biological Plasticity

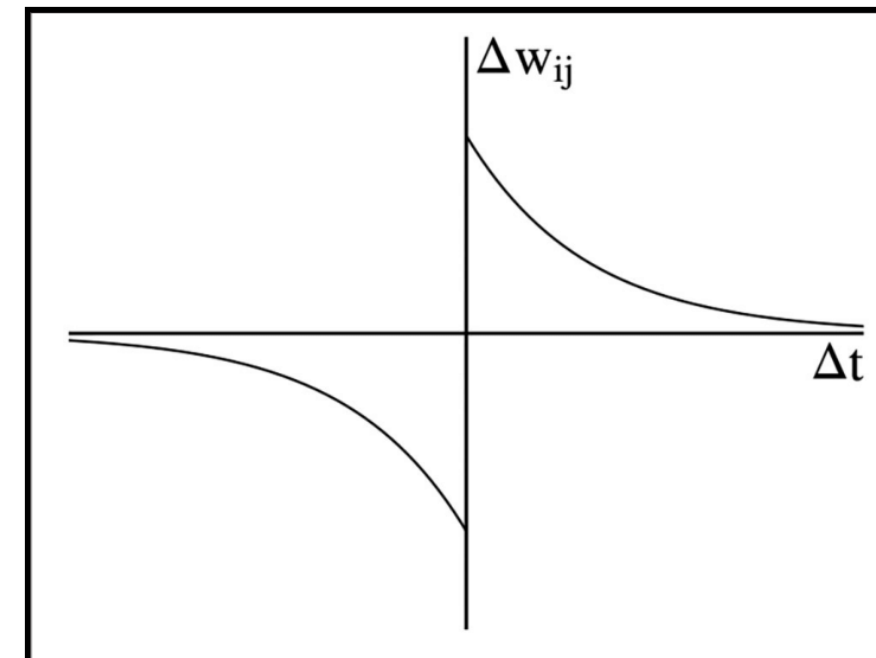


Pau Aceituno

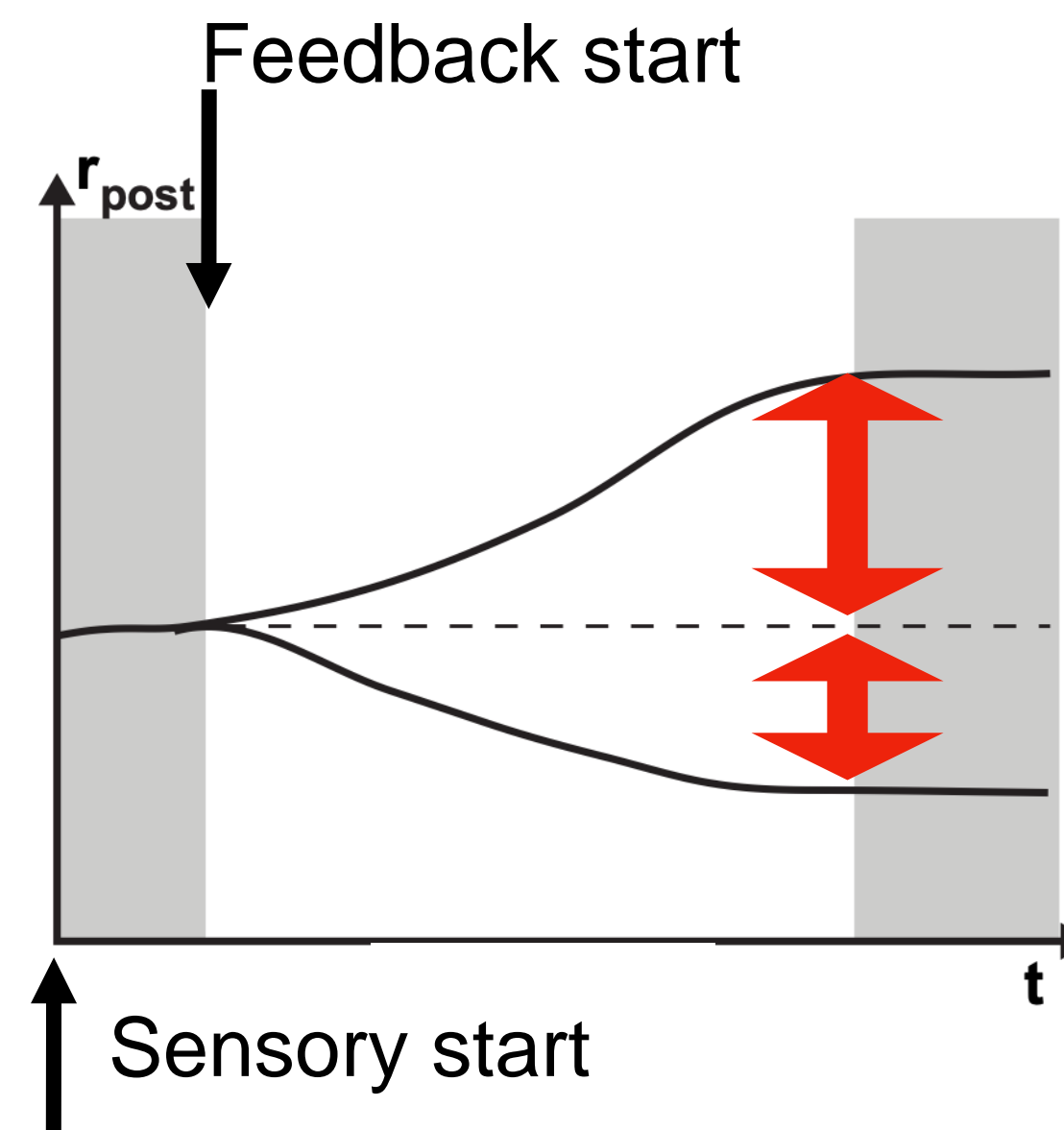
Biology:
Spike Timing
Dependent Plasticity
(STDP)



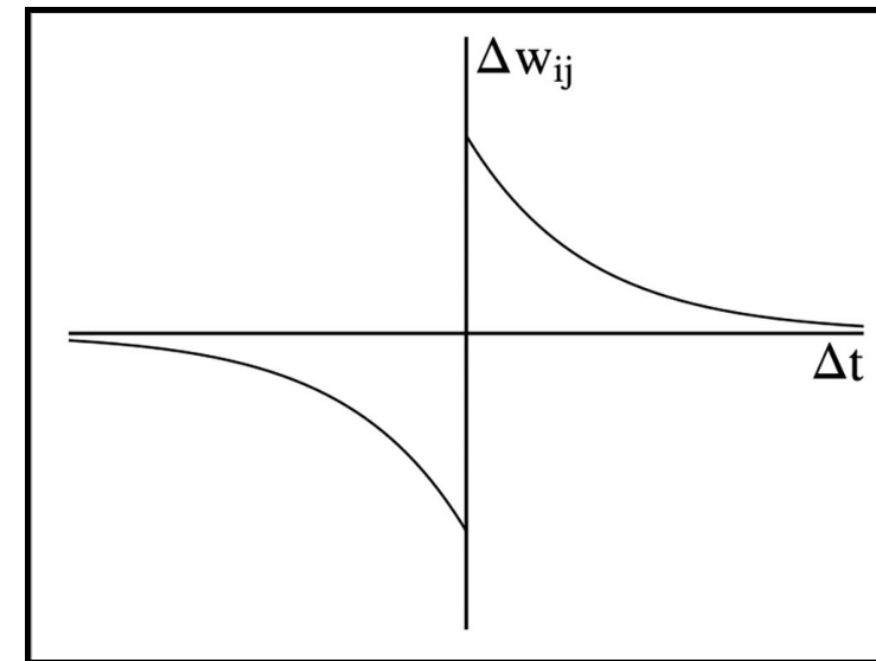
Biology:
Spike Timing
Dependent Plasticity
(STDP)



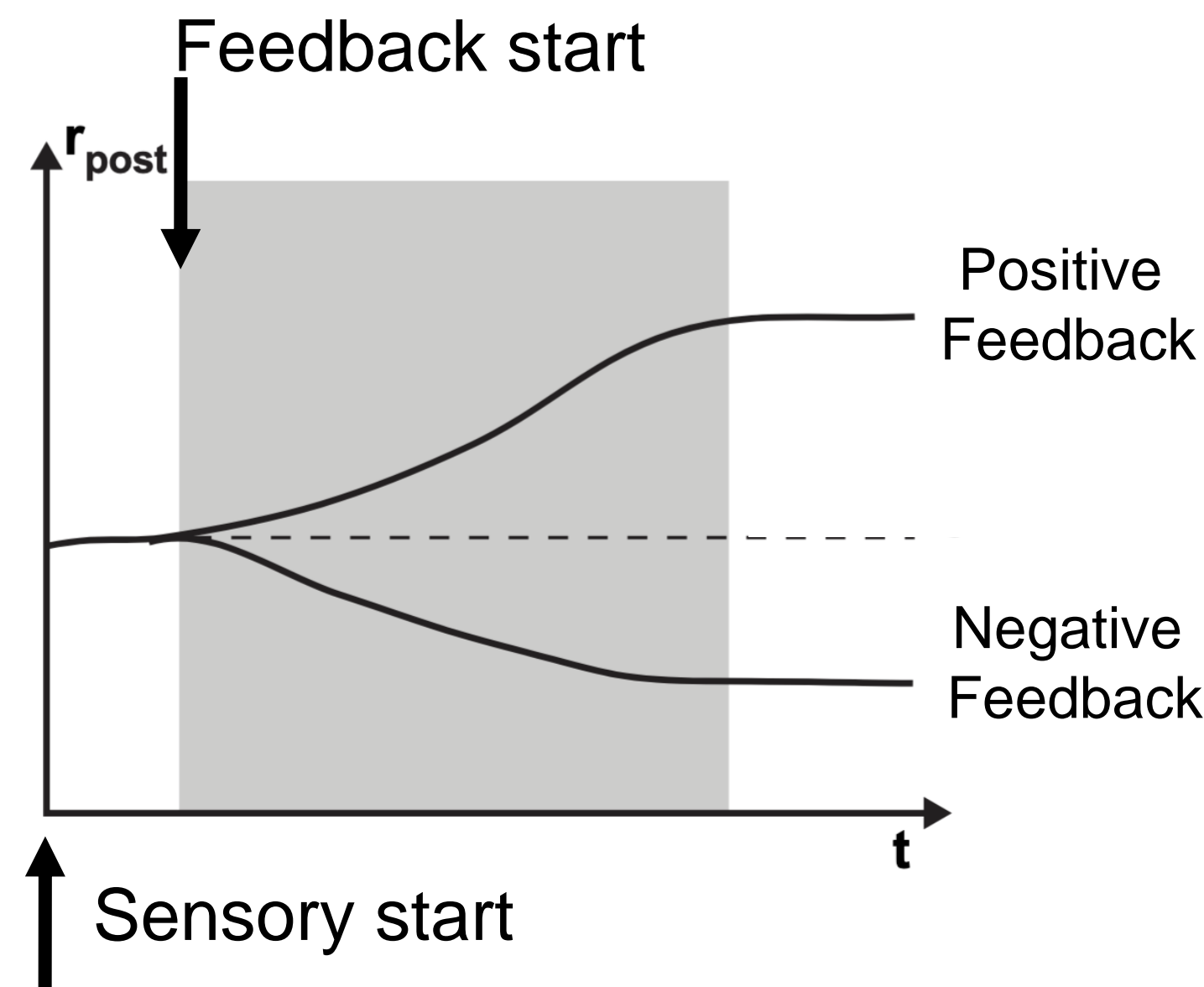
Learning Through Control
Differential Hebbian +
Feedback Control



Biology:
Spike Timing
Dependent Plasticity
(STDP)



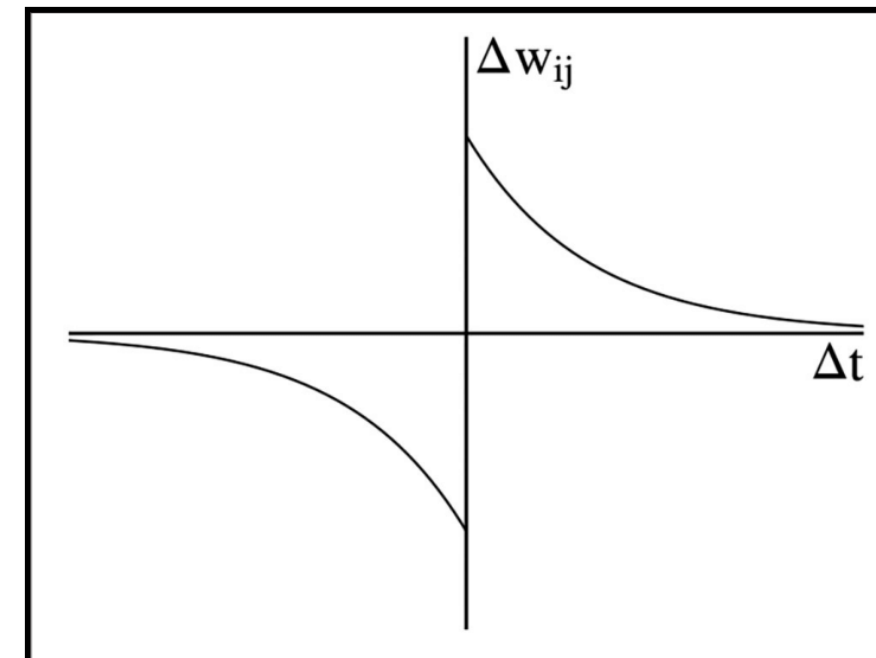
Learning Through Control
Differential Hebbian +
Feedback Control



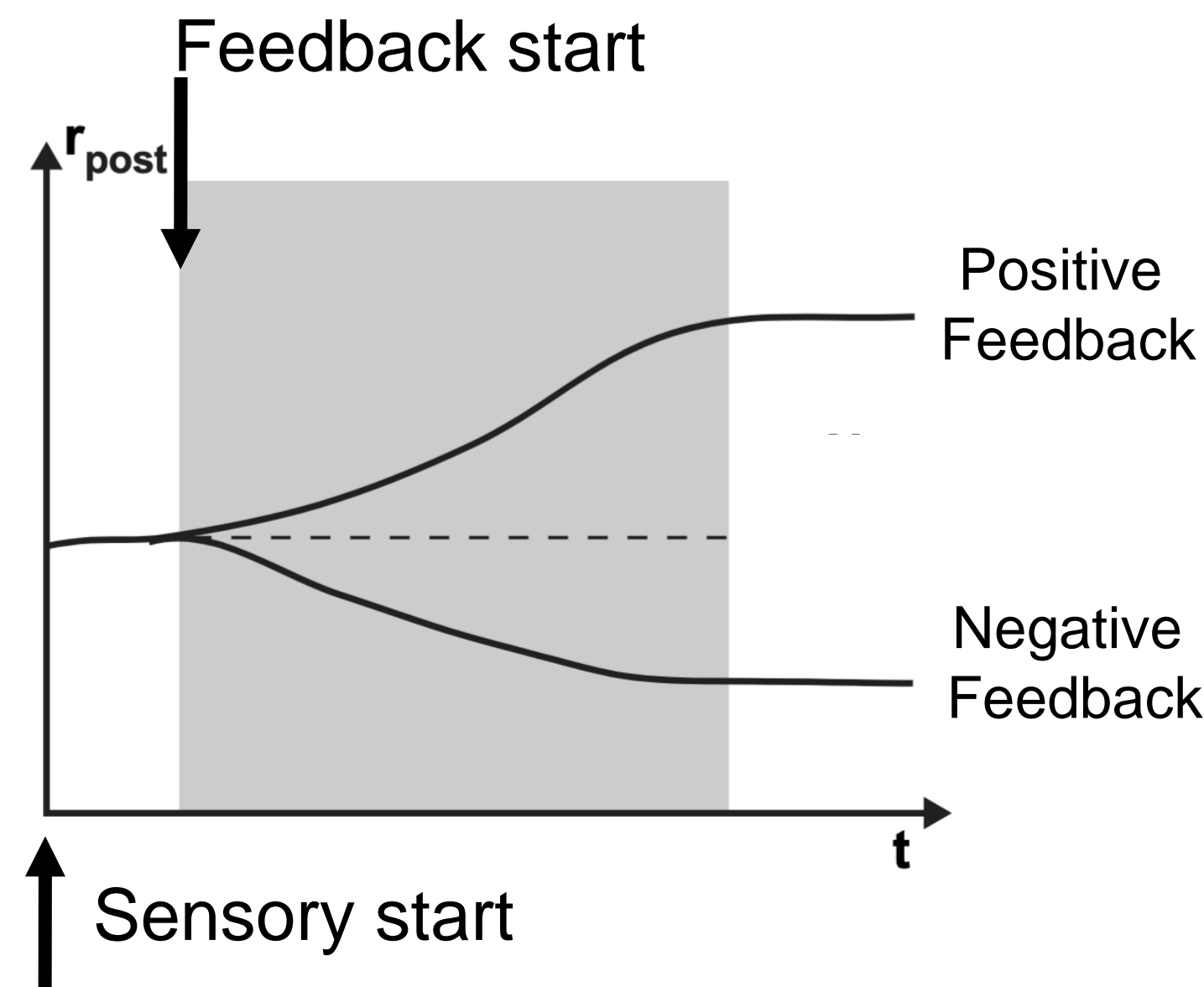
Weight Update

$$\Delta w \propto \int r_{pre}(t) \dot{r}_{post}(t) dt$$

Biology:
Spike Timing
Dependent Plasticity
(STDP)



Learning Through Control
Differential Hebbian +
Feedback Control



Weight Update

$$\Delta w \propto \int r_{pre}(t) \dot{r}_{post}(t) dt$$

Temporal Hebbian Updates Allow to Learn Cortical Hierarchies

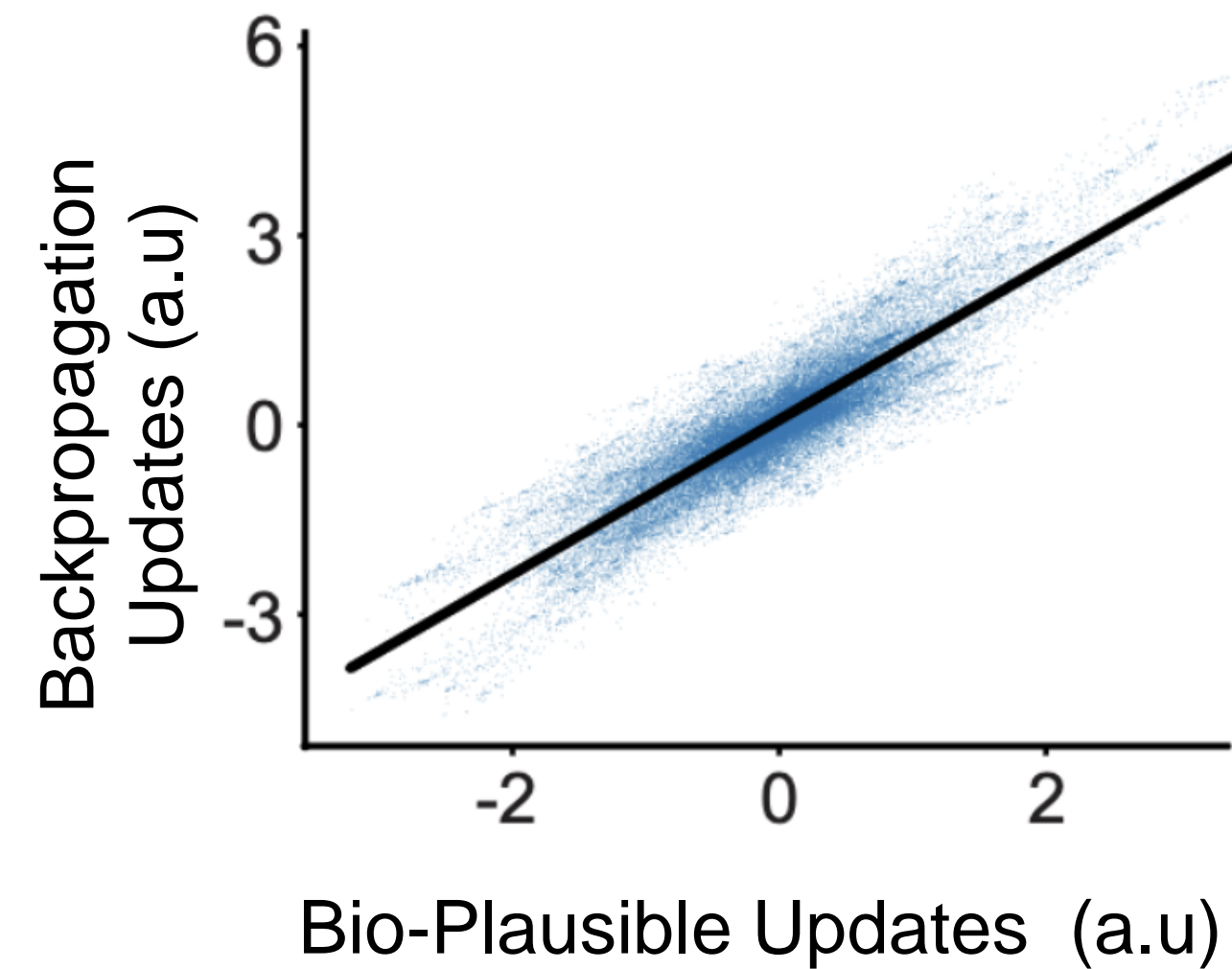
MNIST Dataset

(standard benchmark for
handwritten digit recognition)



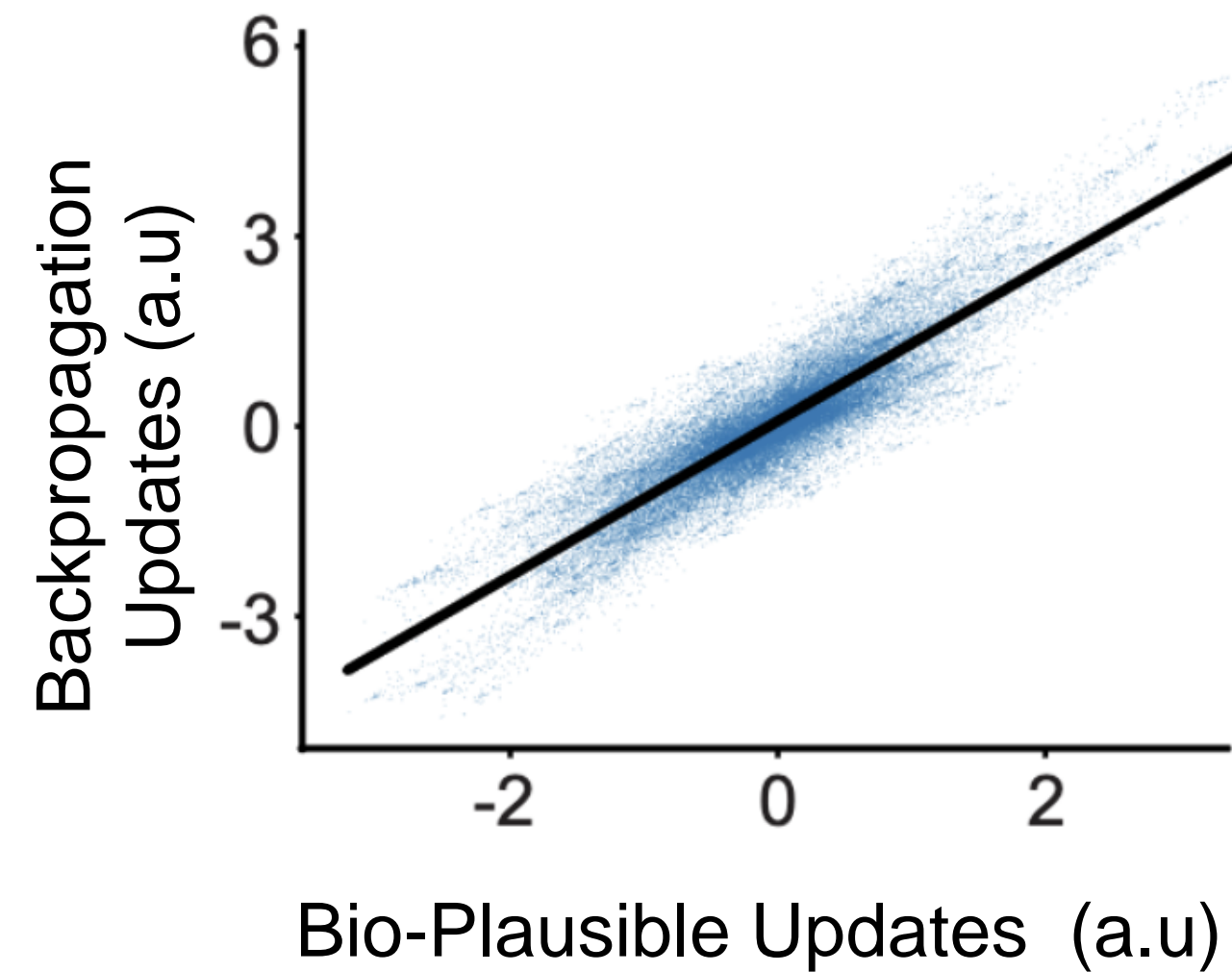
Temporal Hebbian Updates Allow to Learn Cortical Hierarchies

MNIST Dataset
(standard benchmark for
handwritten digit recognition)



Temporal Hebbian Updates Allow to Learn Cortical Hierarchies

MNIST Dataset
(standard benchmark for
handwritten digit recognition)

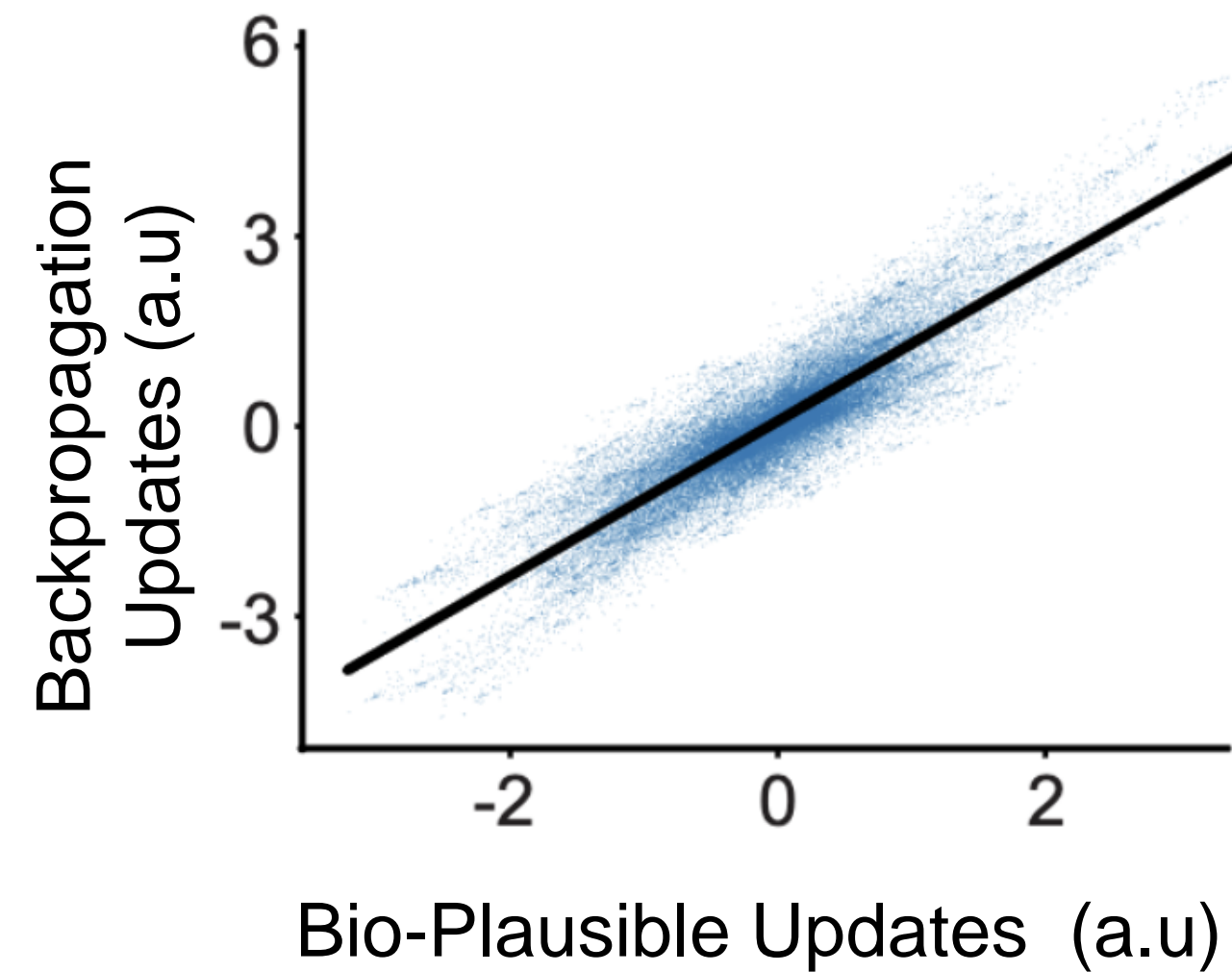


Classification Error

BP	$1.74 \pm 0.10\%$
DFC	$1.98 \pm 0.05\%$
Bio-DFC	$1.89 \pm 0.15\%$

Temporal Hebbian Updates Allow to Learn Cortical Hierarchies

MNIST Dataset
(standard benchmark for
handwritten digit recognition)



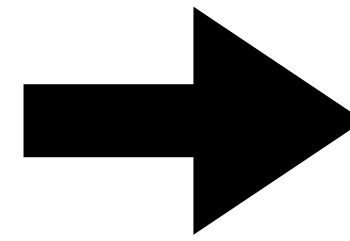
Classification Error

BP	$1.74 \pm 0.10\%$
DFC	$1.98 \pm 0.05\%$
Bio-DFC	$1.89 \pm 0.15\%$

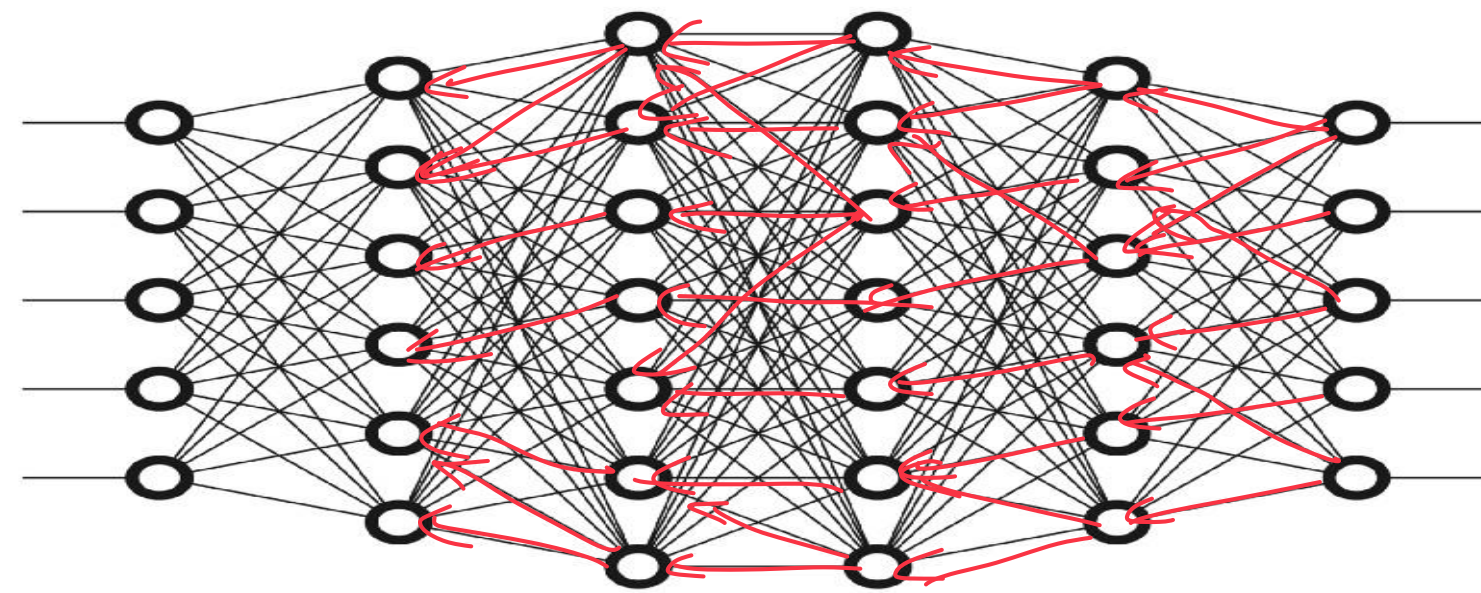
Classification Error

Hierarchical (Deep) Learning through Control

Backpropagation
of the Error



Learning Through Controlling
a Complex Dynamic System



**Learning =
Reducing Control
Feedback!**

Advantages:

- Continuous / asynchronous updates and data processing.
- Absence of phases or back-propagation errors (+ uses less memory).
- Naturally works with time series data.
- Very simple learning rule that is local in space and time.

Hierarchical (Deep) Network Learning through Control

Home > [Biological Cybernetics](#) > Article

Bio-inspired, task-free continual learning through activity regularization

Original Article | [Open Access](#) | [Published: 17 August 2023](#) | (2023)



 What are your technology predictions for 2023? 

 I'm sorry, but I am not able to browse the internet or access current information, so I am unable to provide you with any technology predictions for the year 2023. My training data only goes up until 2021, and I do not have the ability to browse the internet or access current information. Is there anything else I can help you with?  

- Enables Continual Learning when neuronal activity is sparse.

(Deep) Learning through Control in Hardware



**Indiveri
Group
@ INI**

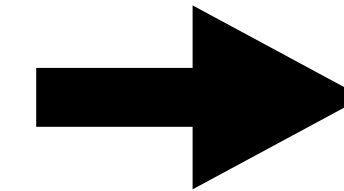


Matteo Saponati
(Grewe lab)

Neuromorphic Devices

- Ultra low-power
- Scalable to 3D

Learning by
Controlling a complex
Dynamic System



Building a
Neuromorphic
Device for low-power
Deep Learning

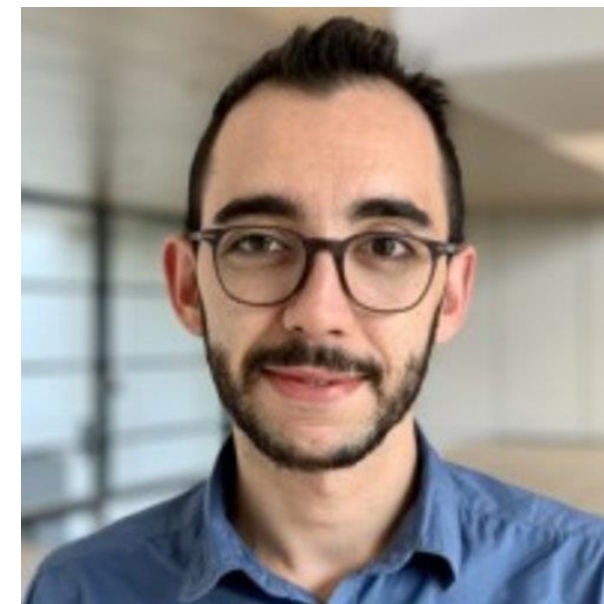
can do Continual Learning which neuronal activity is sparse.

- Is ideally suited for low-power deep learning on neuromorphic processors.

(Deep) Learning through Control in Hardware



Indiveri Group @ INI

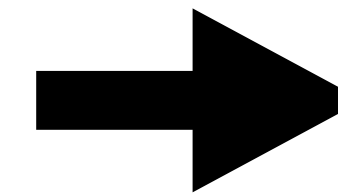


Matteo Saponati (Grewe lab)

Neuromorphic Devices

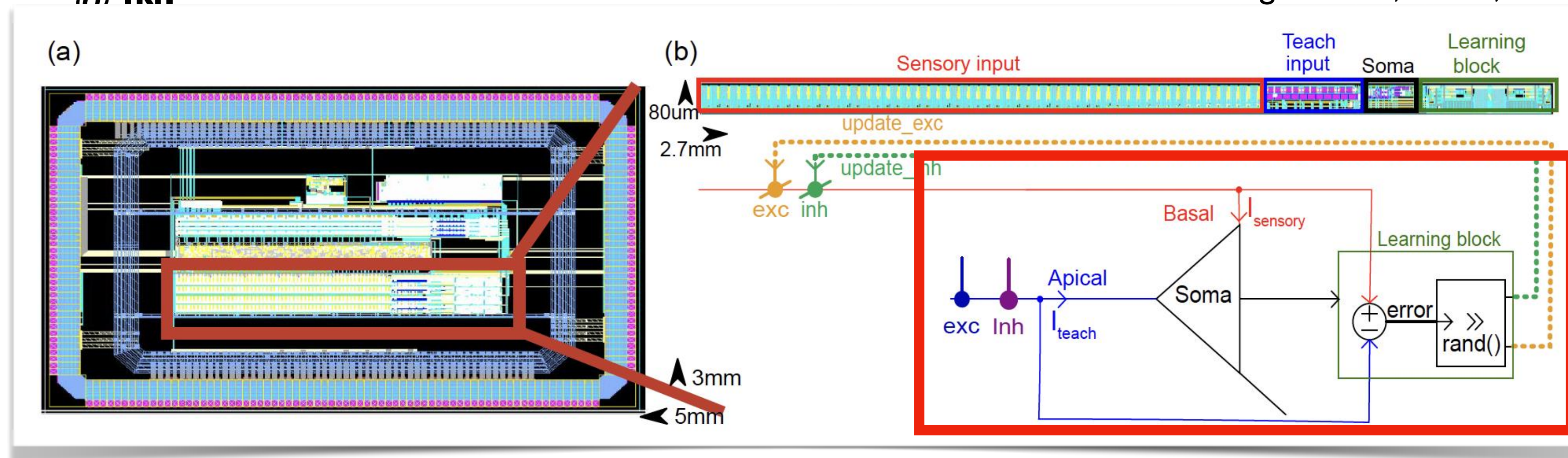
- Ultra low-power
- Scalable to 3D

Learning by Controlling a complex Dynamic System



Building a Neuromorphic Device for low-power Deep Learning

Cartiglia et al, 2022, IEEE

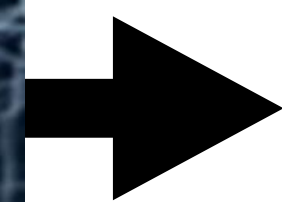
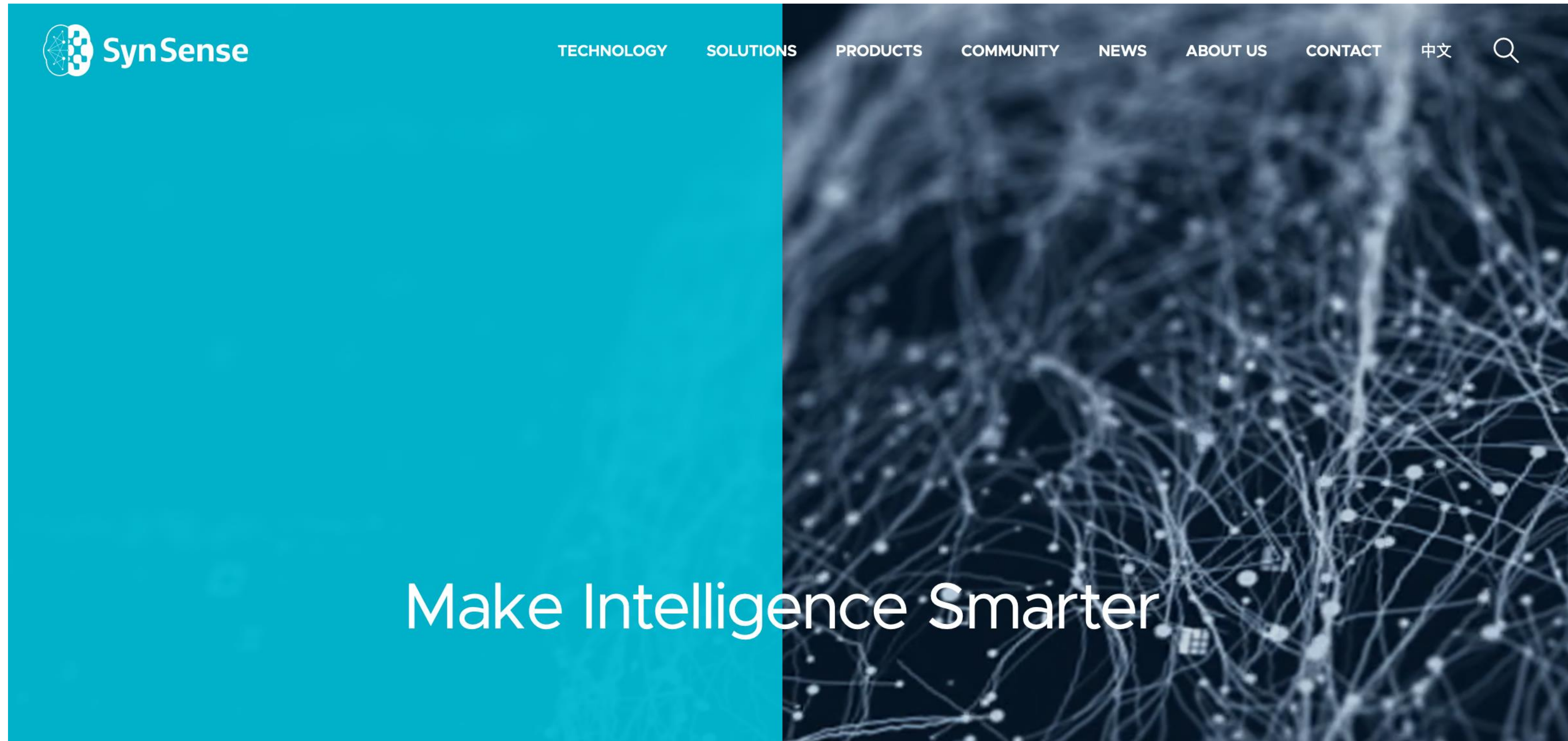


Pyramidal Neuron In Silico

can do continual learning with neuronal activity is sparse.

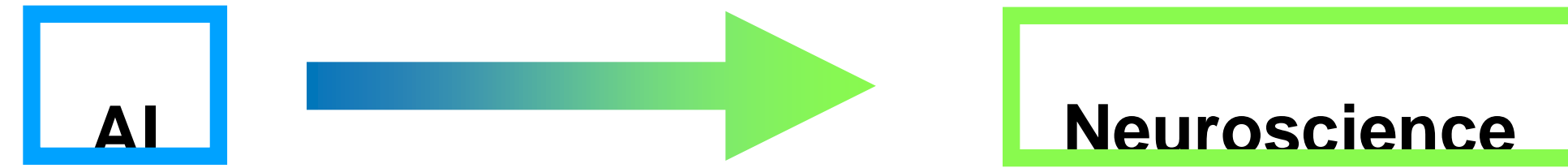
- Is ideally suited for low-power deep learning on neuromorphic processors.

(Deep) Learning through Control in Hardware



**Building a
Neuromorphic
Device for
Deep Learning**

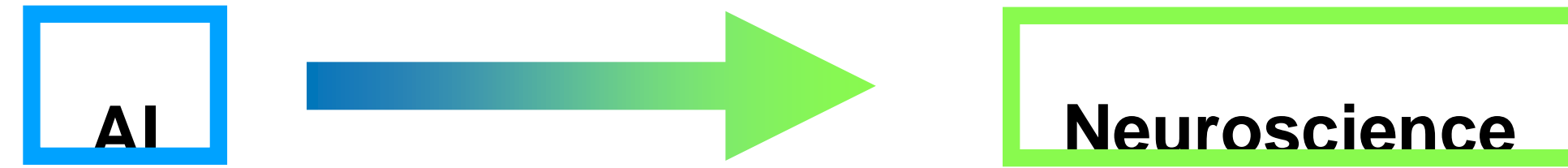
- Is ideally suited for low-power neuromorphic processors.



Weight Update

$$\Delta w \propto \int r_{\text{pre}}(t) \dot{r}_{\text{post}}(t) dt$$

Plasticity is only 'ON' when Feedback is active!



Weight Update

$$\Delta w \propto \int r_{\text{pre}}(t) \dot{r}_{\text{post}}(t) dt$$

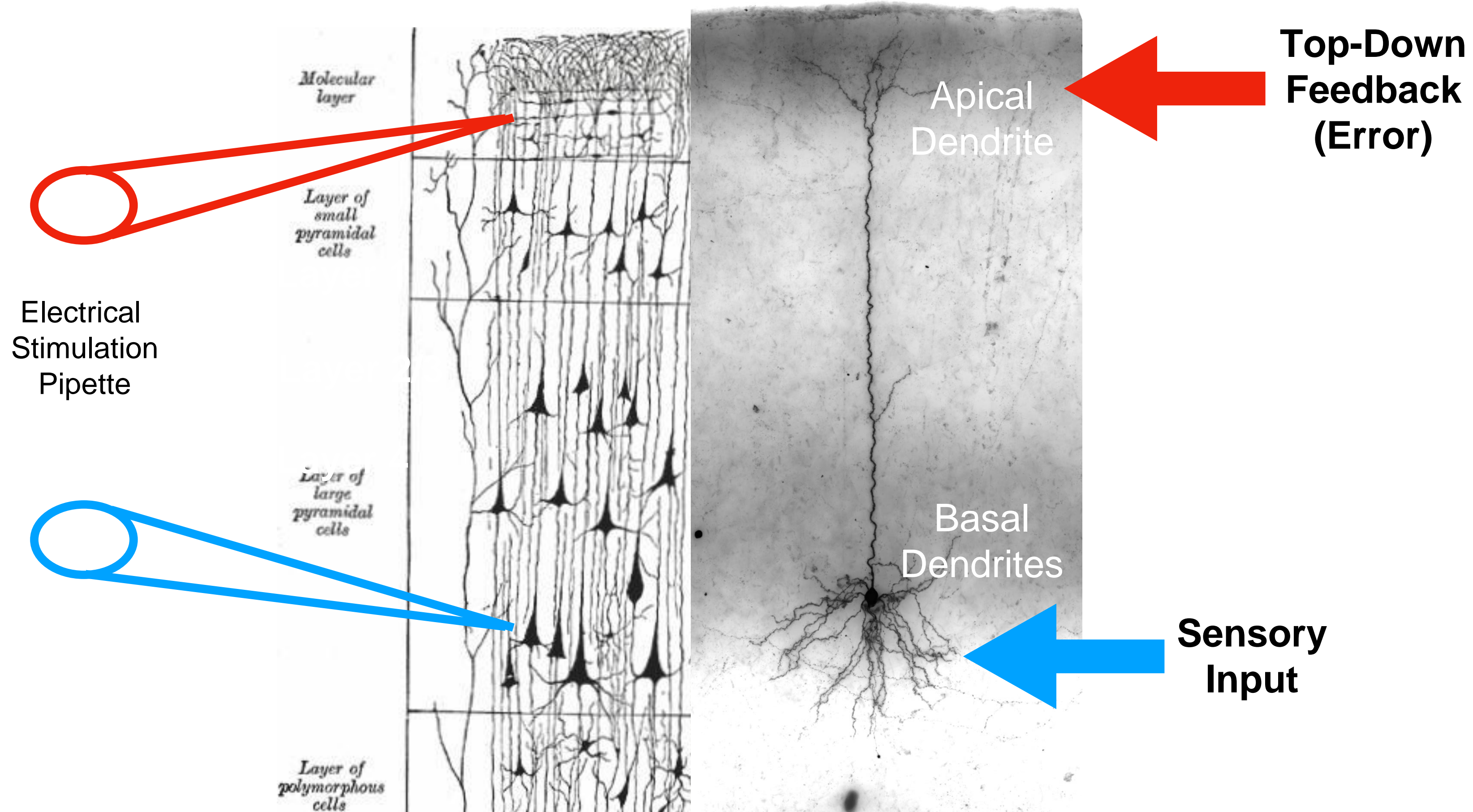
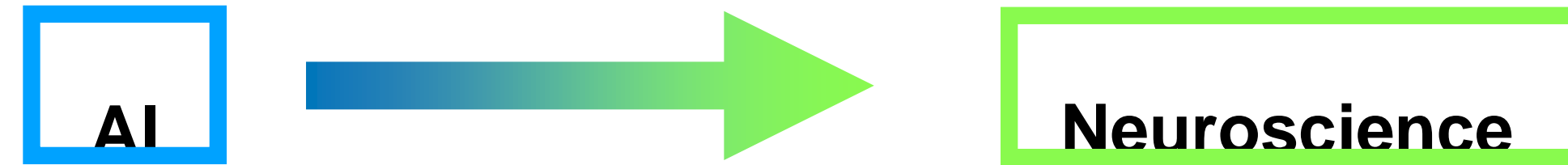


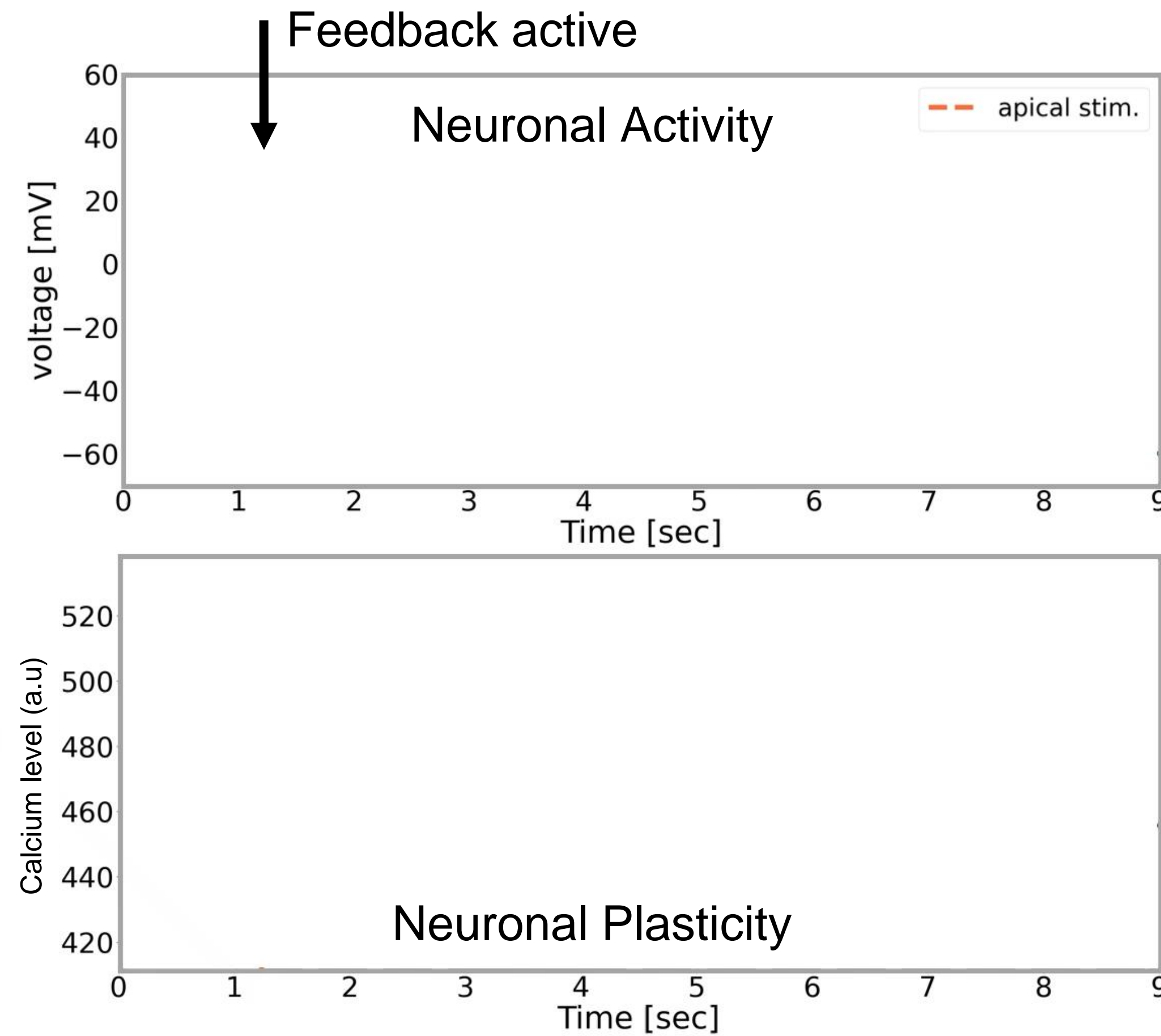
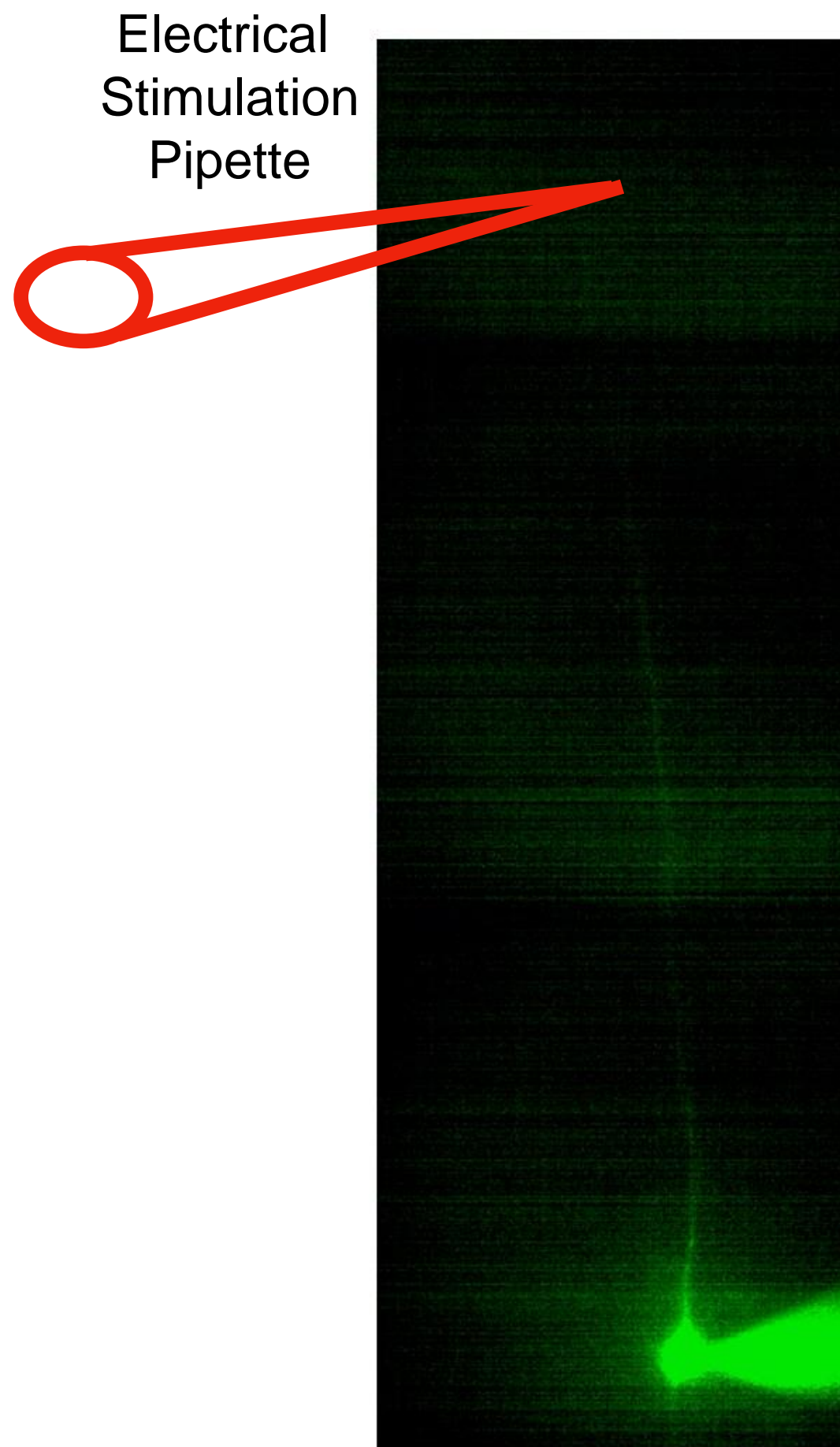
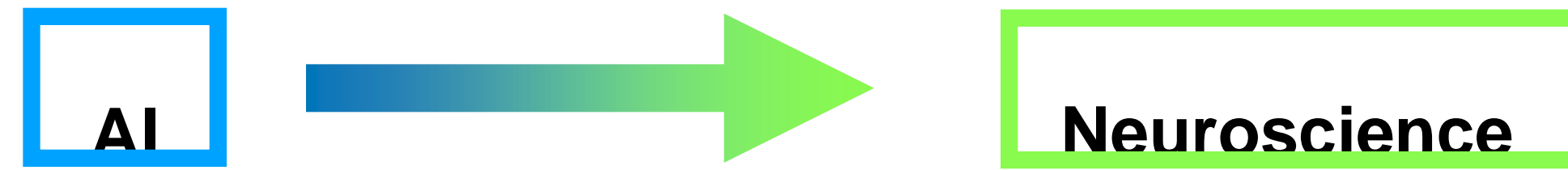
Plasticity is only 'ON' when Feedback is active!

In Vitro Measurement of
Neuronal Activity & Plasticity

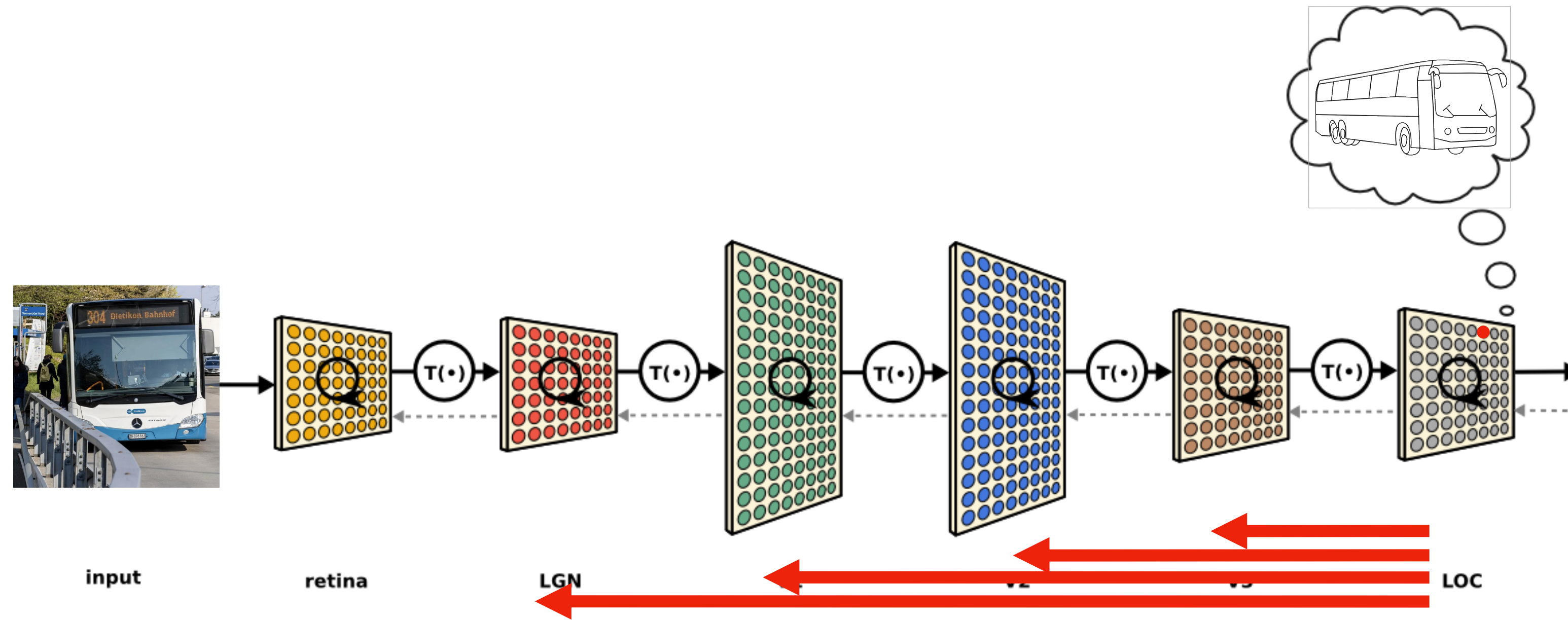


We are testing this theoretical prediction in biological neurons!

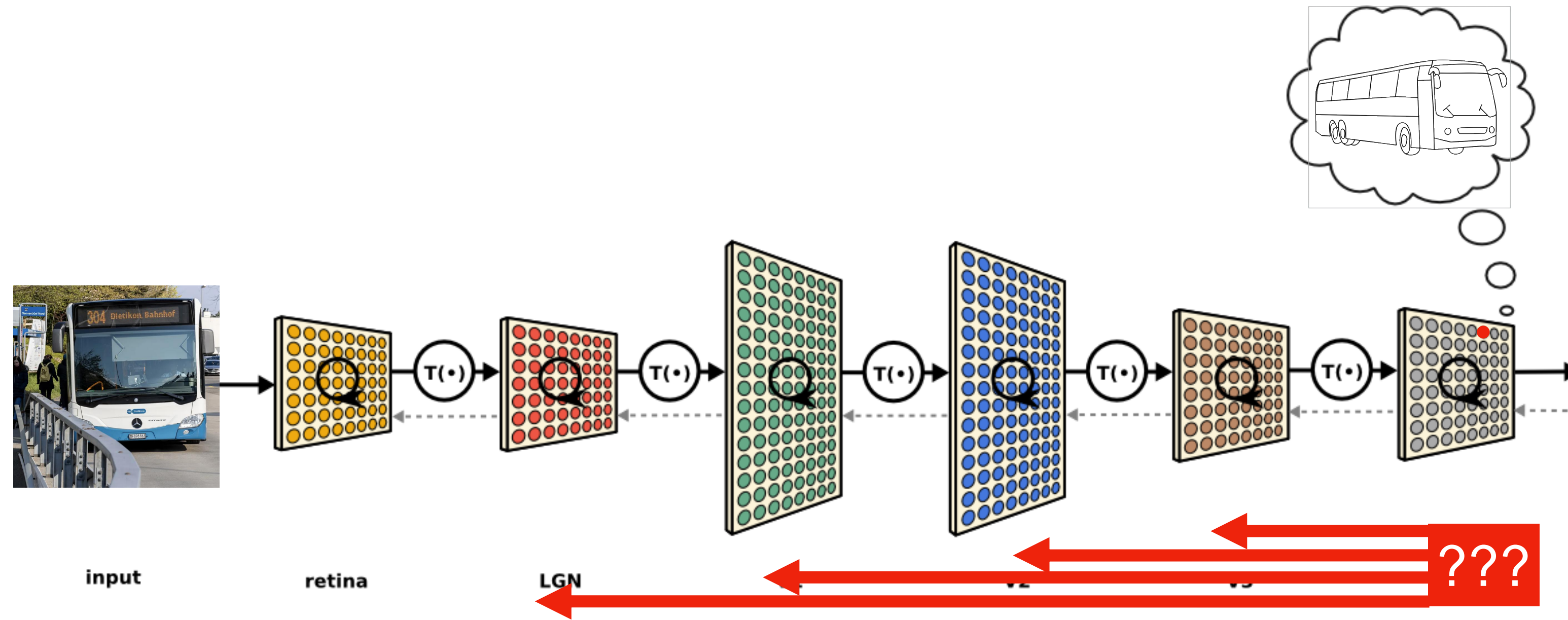


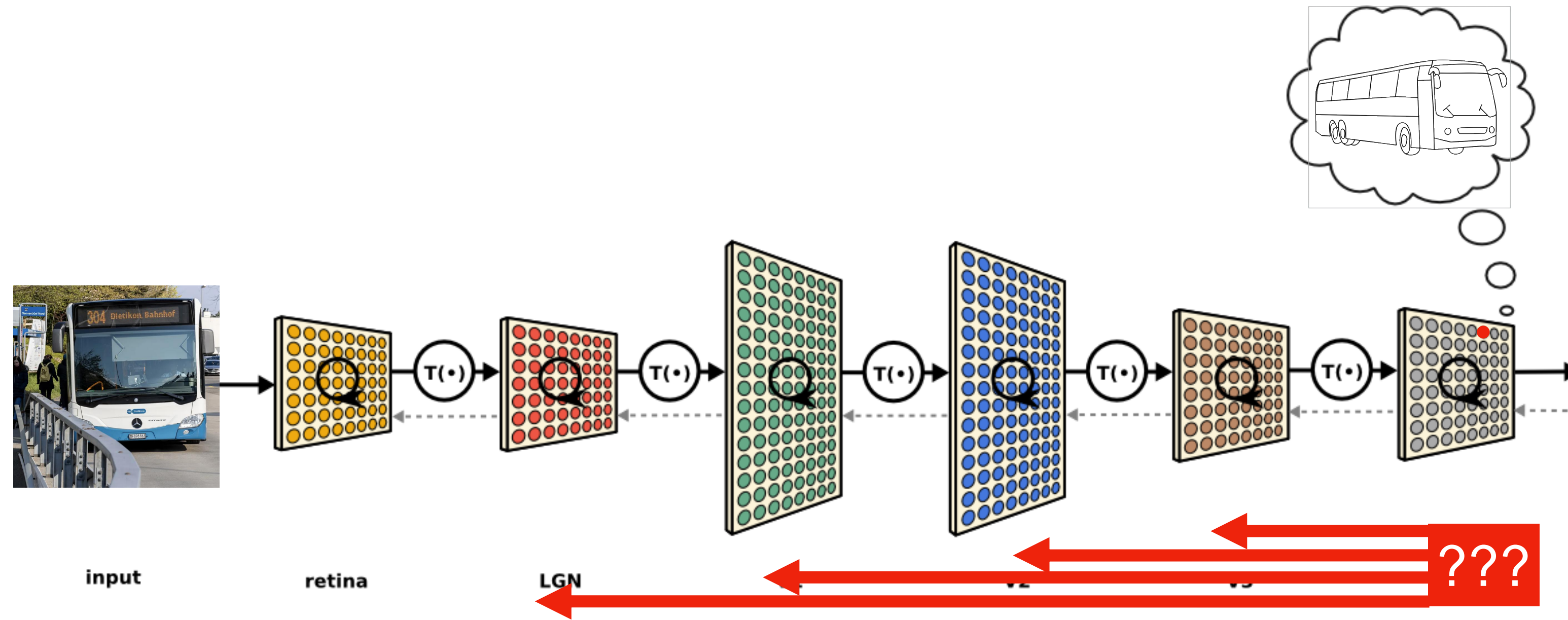


Summary Part 1:



Summary Part 1:





Part II: Understanding Hierarchical Neuronal Representations in Brain

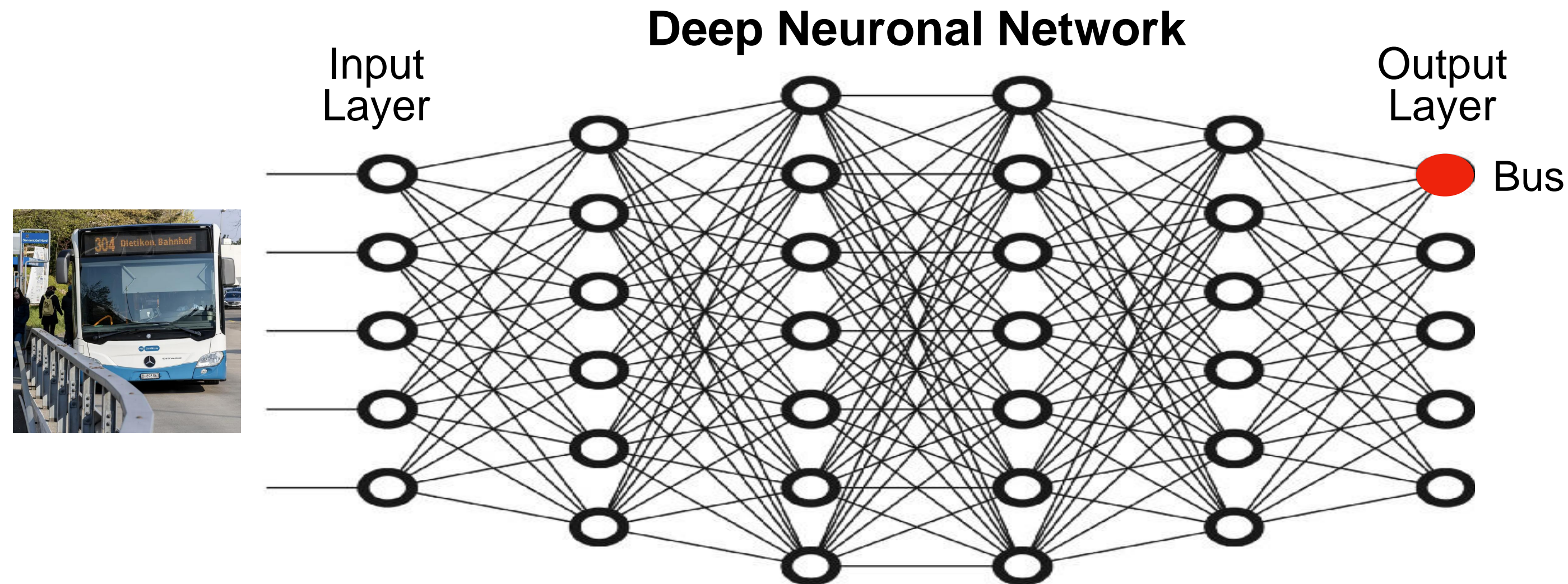
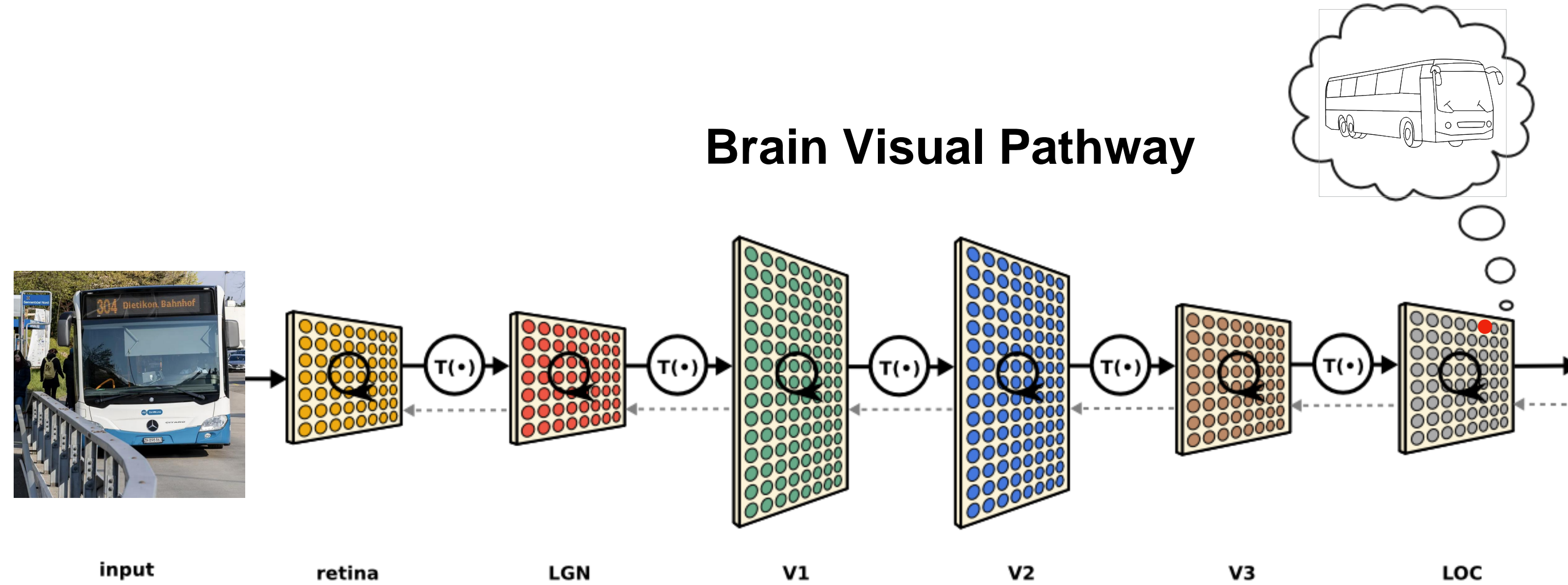
AI

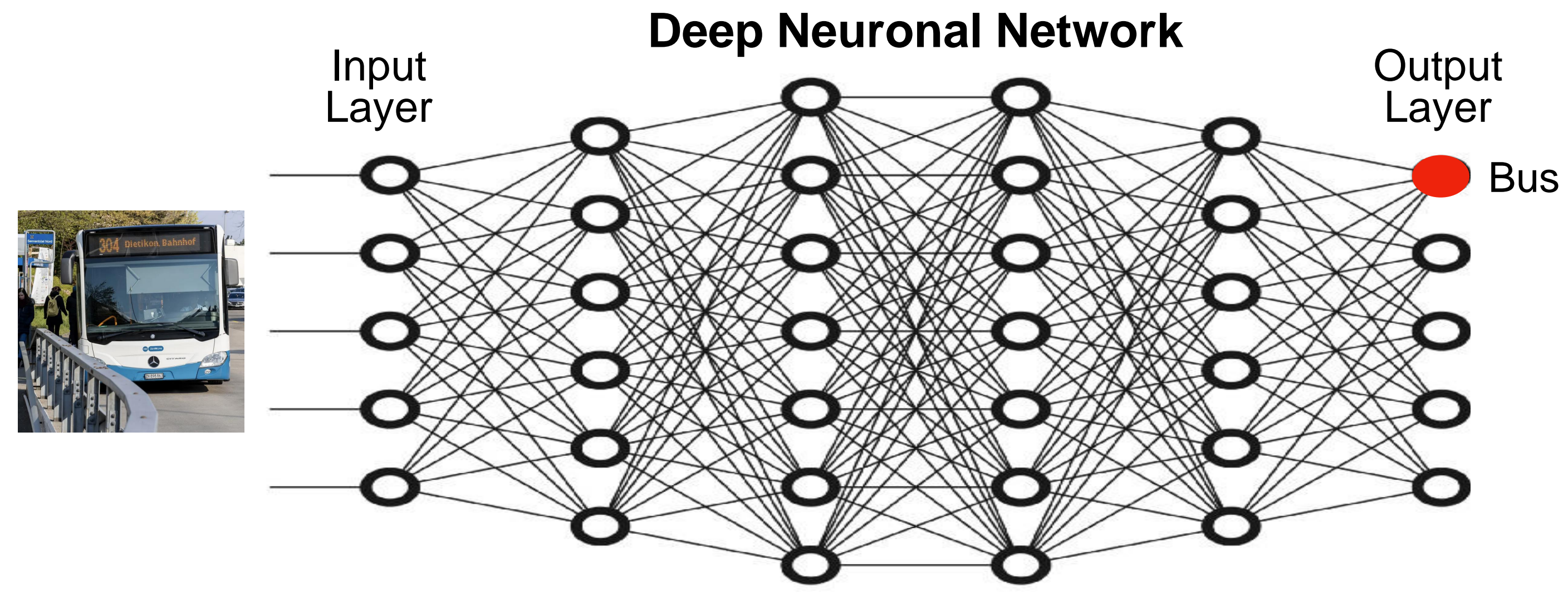
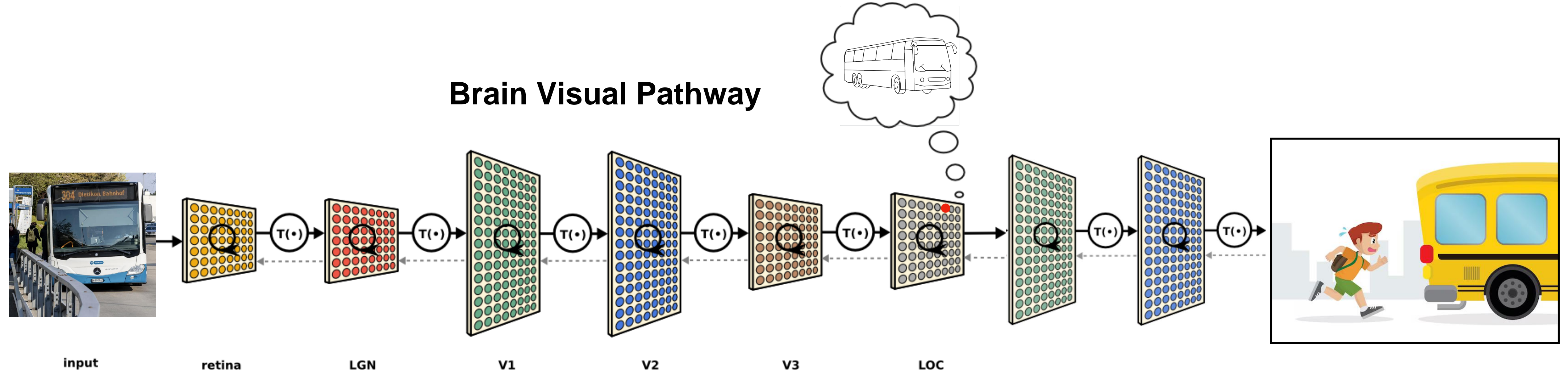


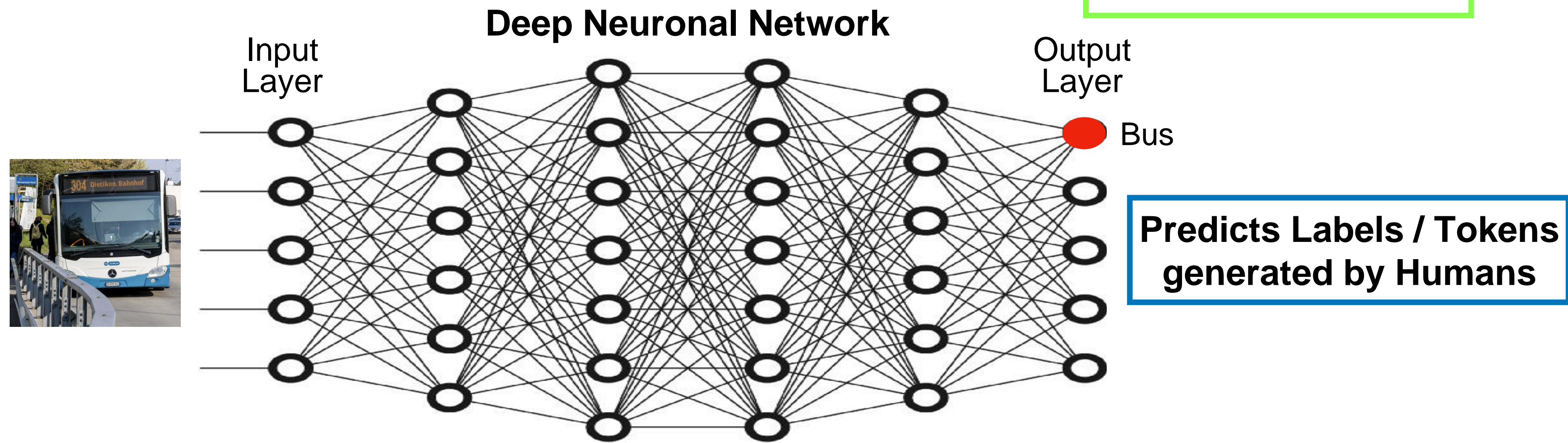
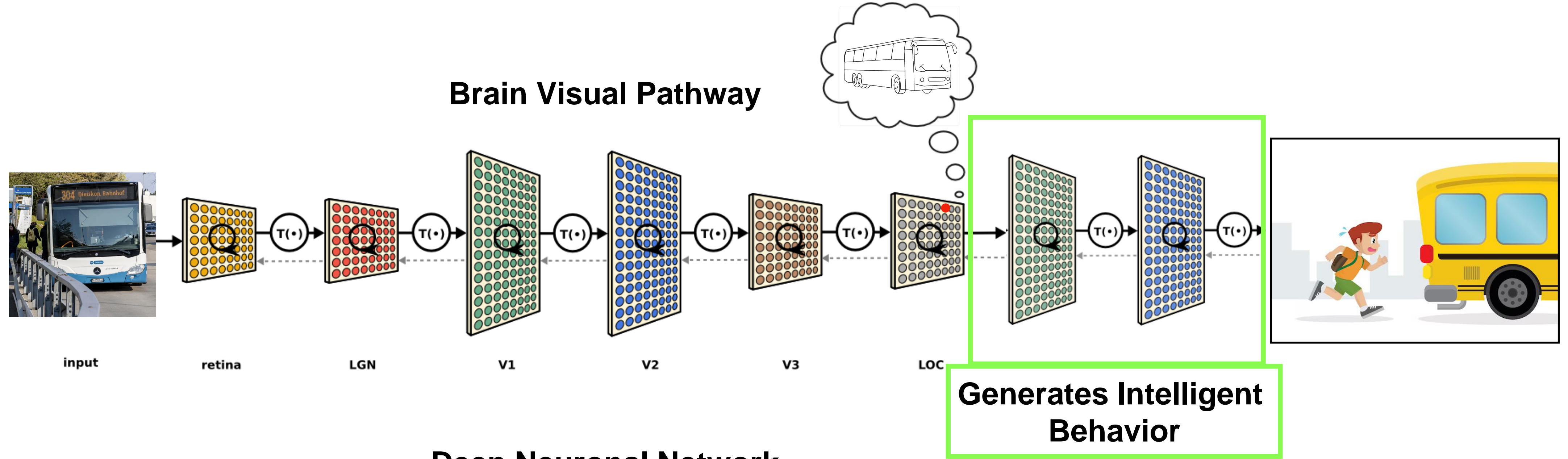
Neuroscience

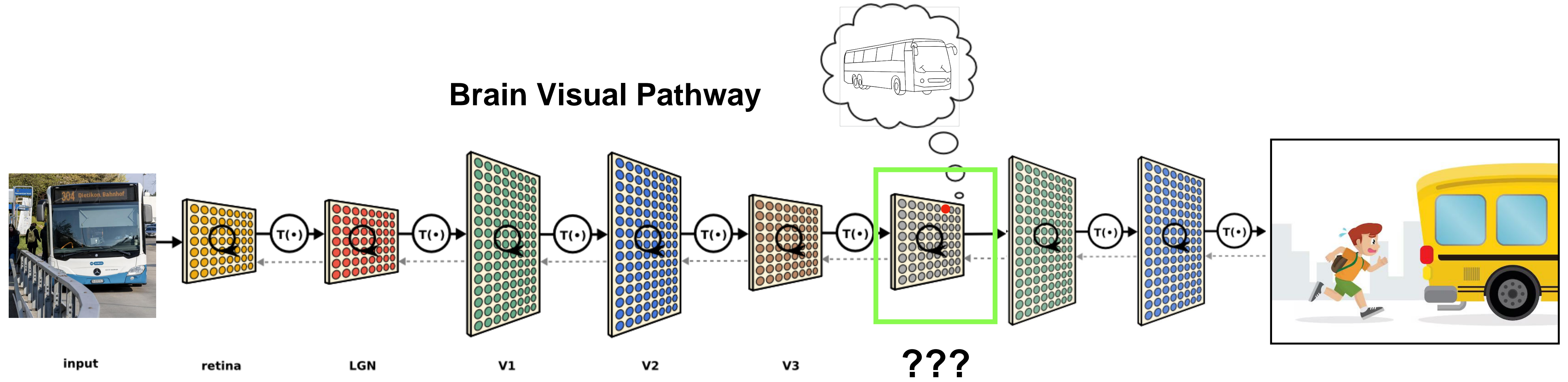
Scientific Question:
What are the neuronal representations of the
sensory input (e.g. image of bus) that allows our brain
to generate goal directed actions?



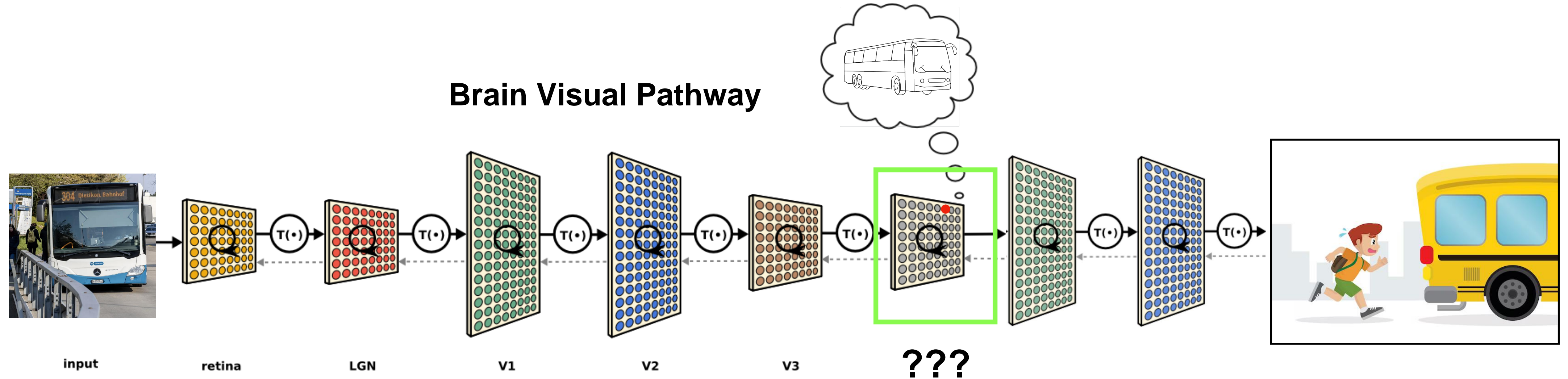








Scientific Question:
 What are the neuronal representations of the sensory input (e.g. image of bus) that allows our brain to generate goal directed actions?



Prof. Dr. Jean Piaget

Swiss psychologist
Neuchatel. 1896-1980

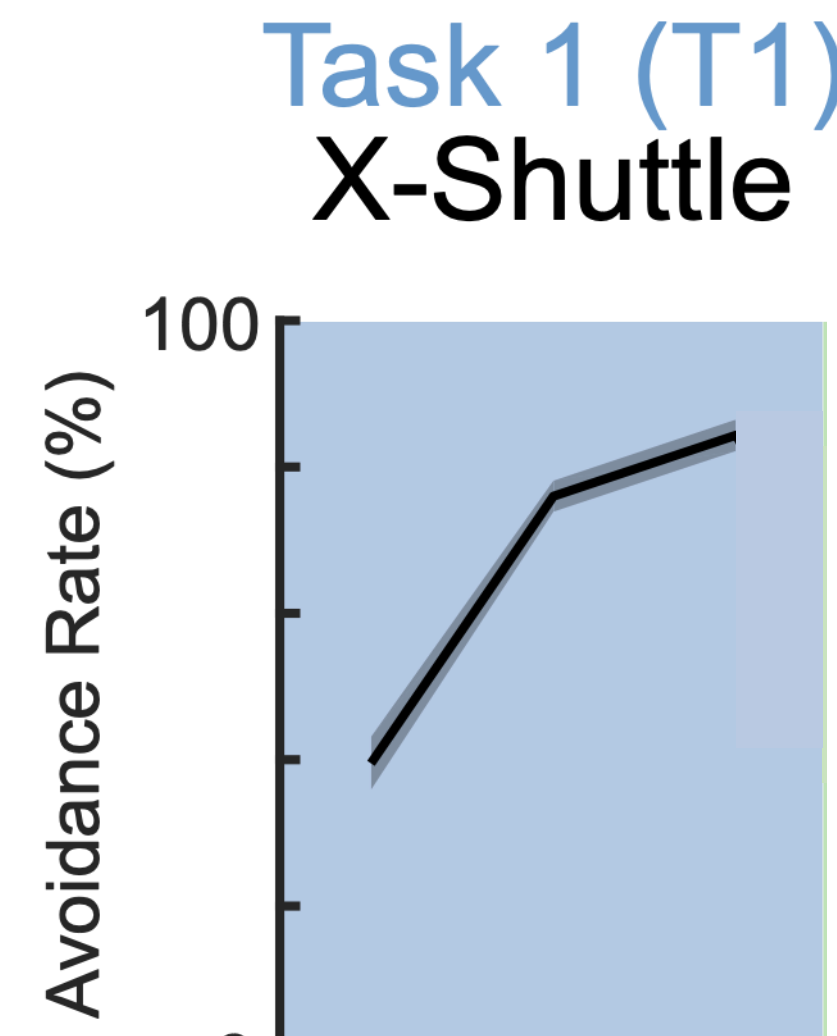
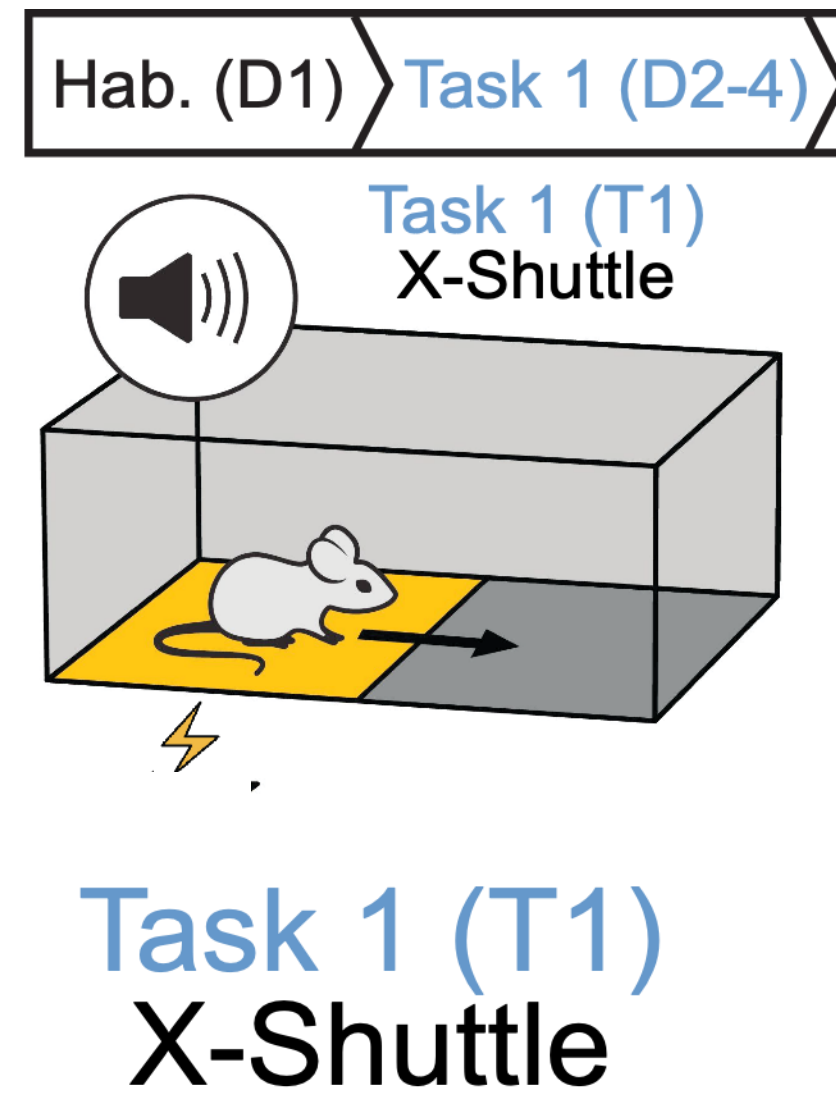


The Concept of Affordance in Psychology.

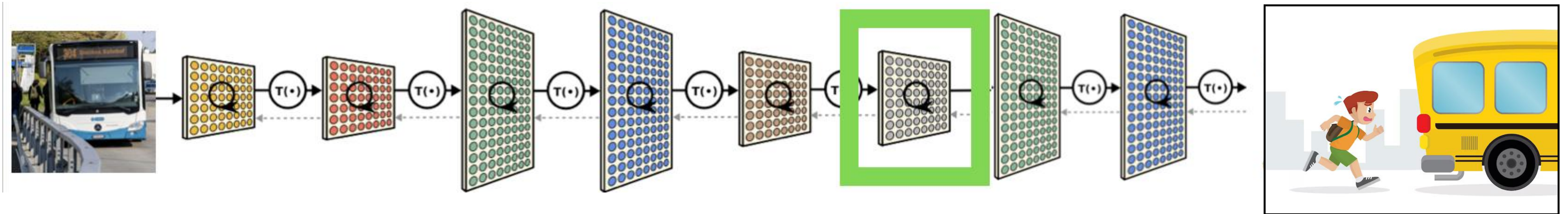
Affordance alludes to the qualities of an object or situation that define its possible use or make clear how it can or should be used.

Affordance adheres to the idea that perception and action are inseparable (Principles of Genetic Epistemology, Jean Piaget).

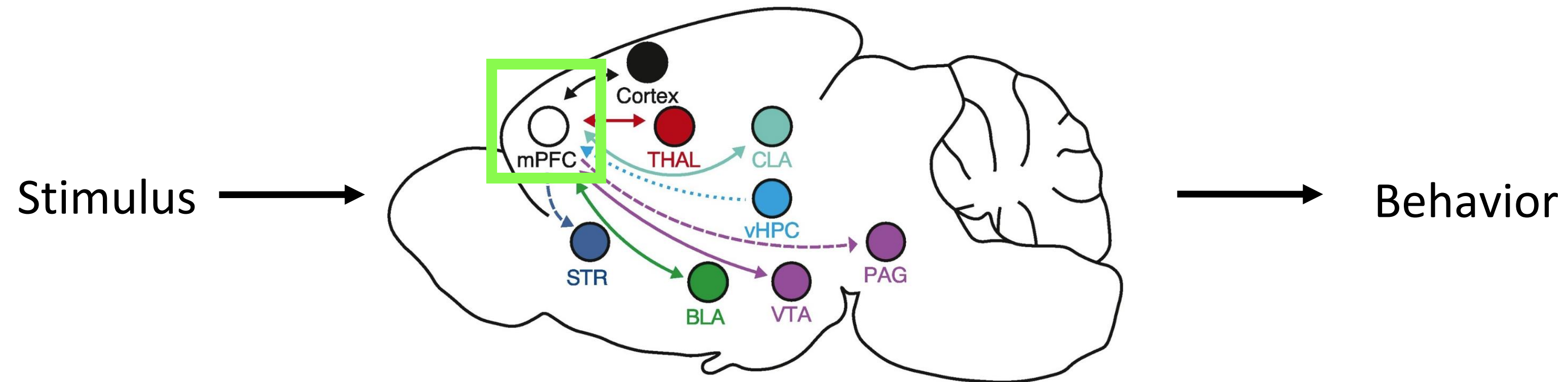
Investigating Abstract Stimulus Representations in mPFC



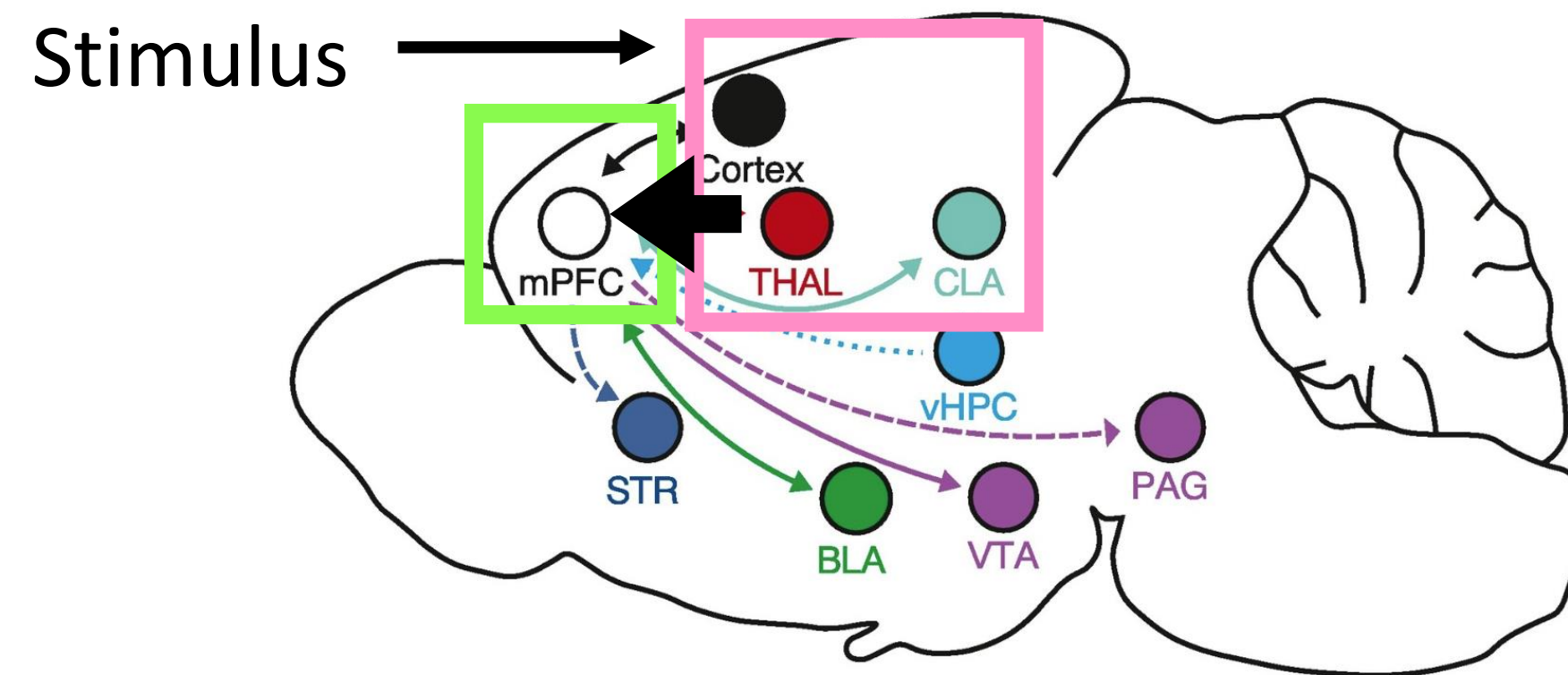
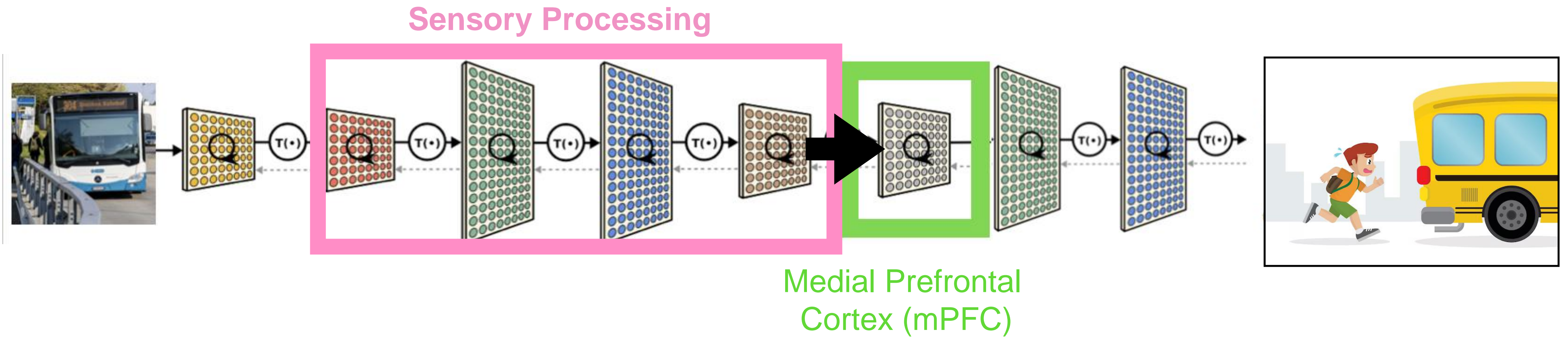
Investigating Abstract Stimulus Representations in mPFC



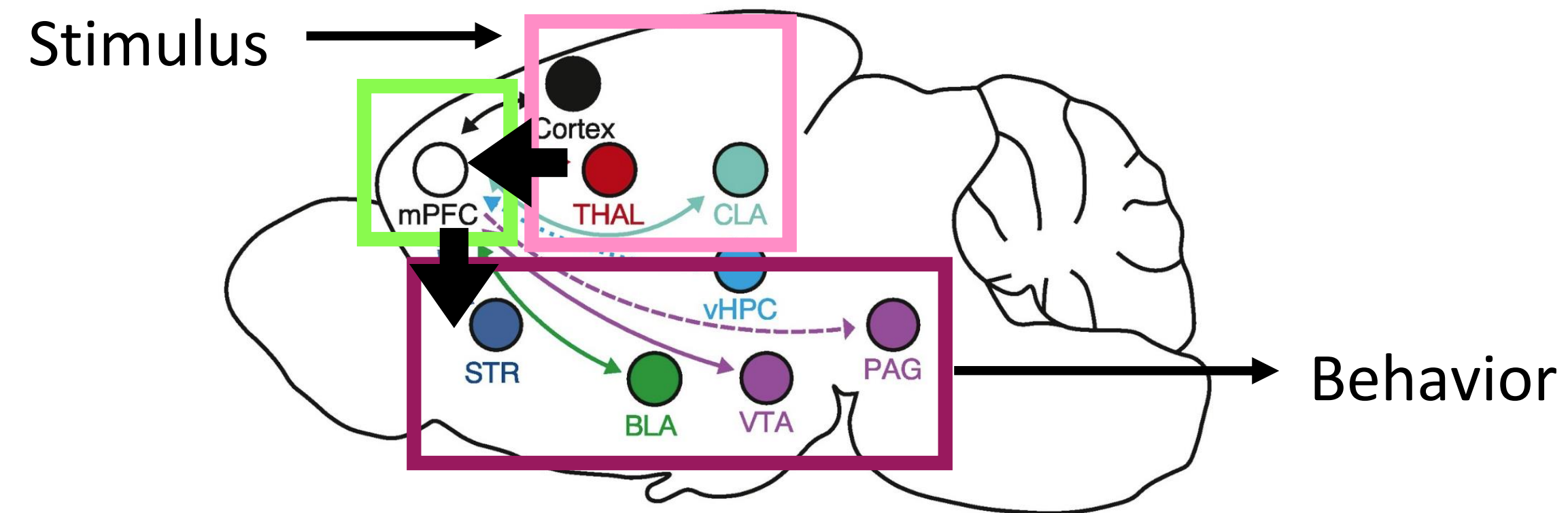
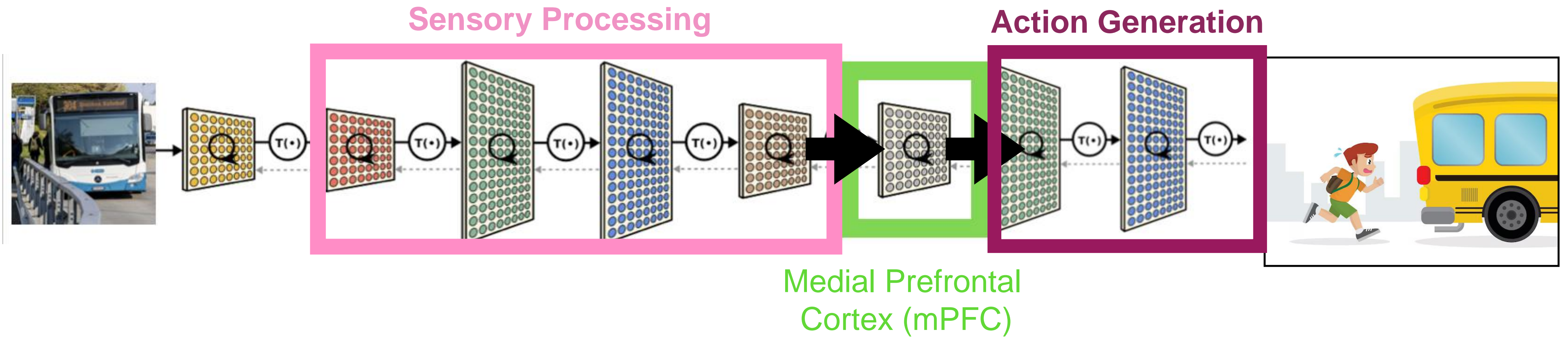
Medial Prefrontal
Cortex (mPFC)



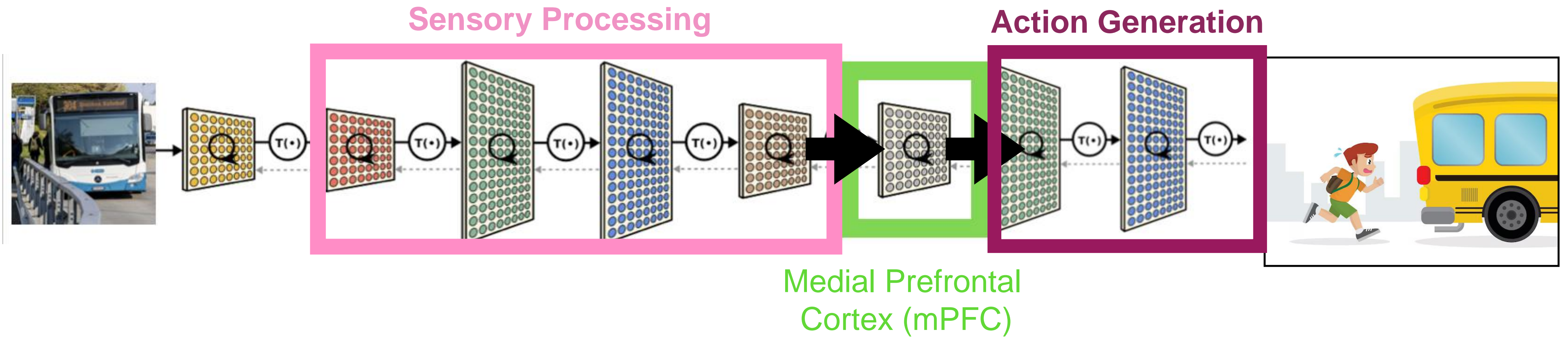
Investigating Abstract Stimulus Representations in mPFC



Investigating Abstract Stimulus Representations in mPFC

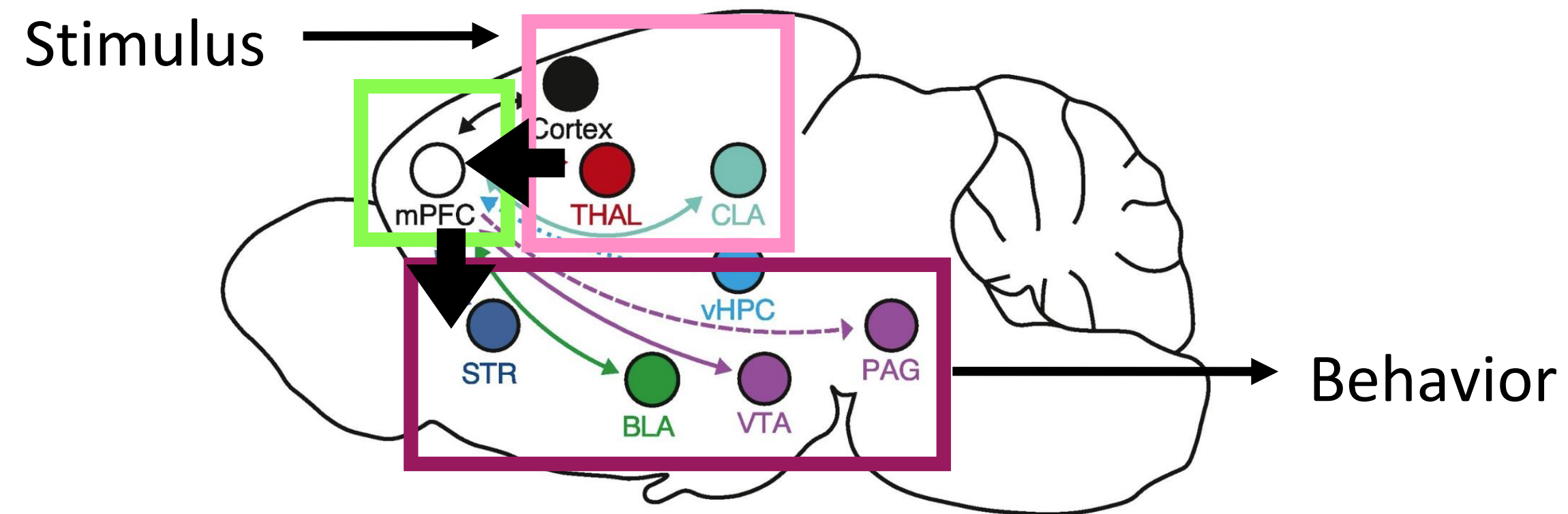


Investigating Abstract Stimulus Representations in mPFC

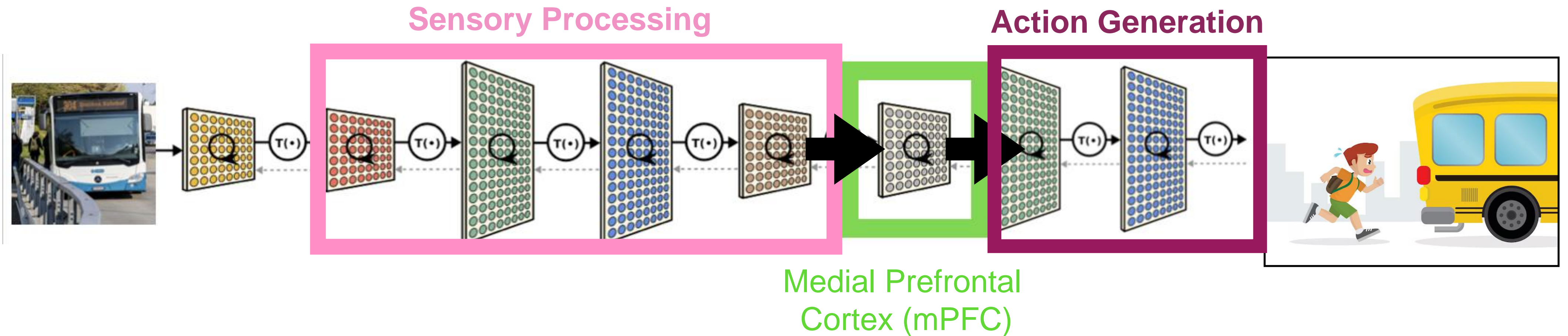


‘Learning induces representations of behaviorally relevant stimuli’

Burgos-Robles et al. 2009
Le Merre et al. 2018
Otis et al. 2017

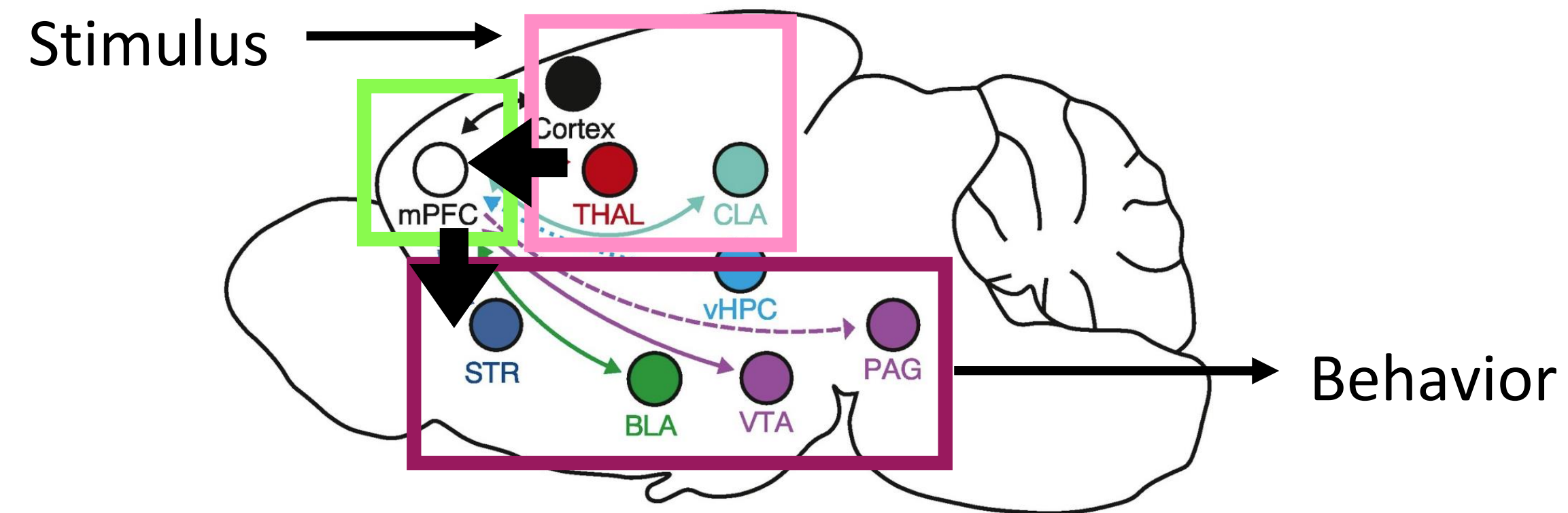


Investigating Abstract Stimulus Representations in mPFC



‘Learning induces representations of behaviorally relevant stimuli’

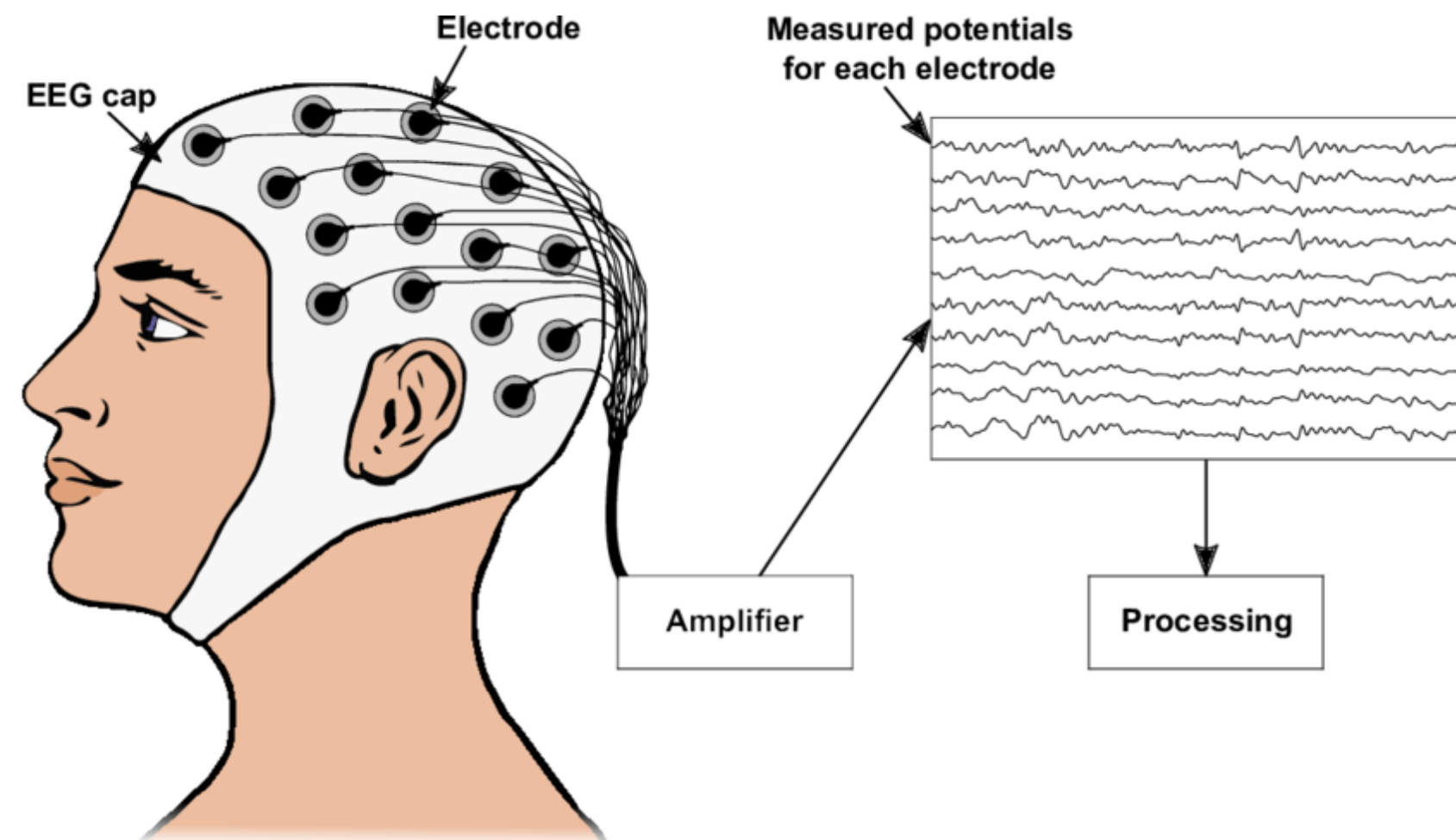
Burgos-Robles et al. 2009
Le Merre et al. 2018
Otis et al. 2017



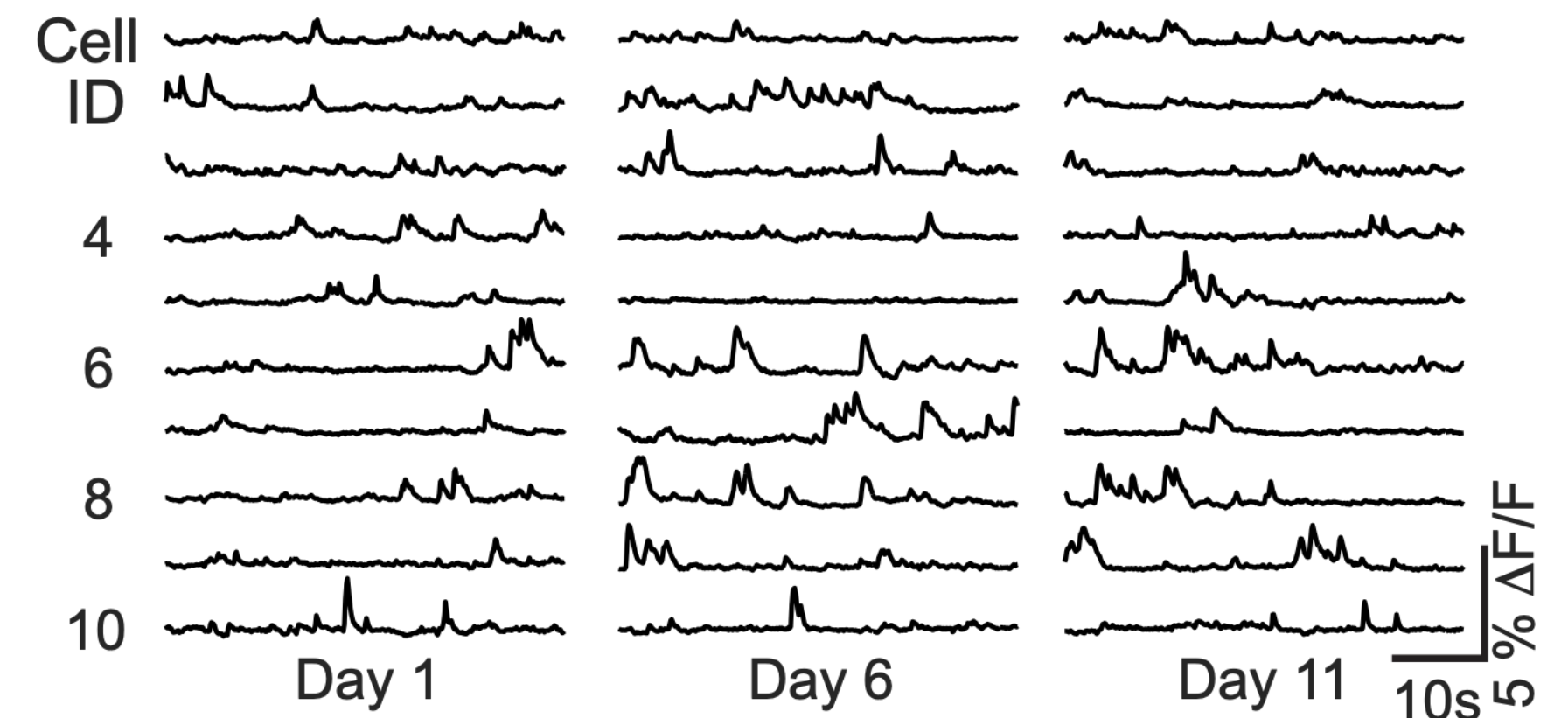
‘mPFC activity guides/ alters behavior’

Murugan et al. 2017
Otis et al. 2017
Rozeske et al. 2018
Diehl et al. 2020

Investigating Abstract Stimulus Representations in mPFC

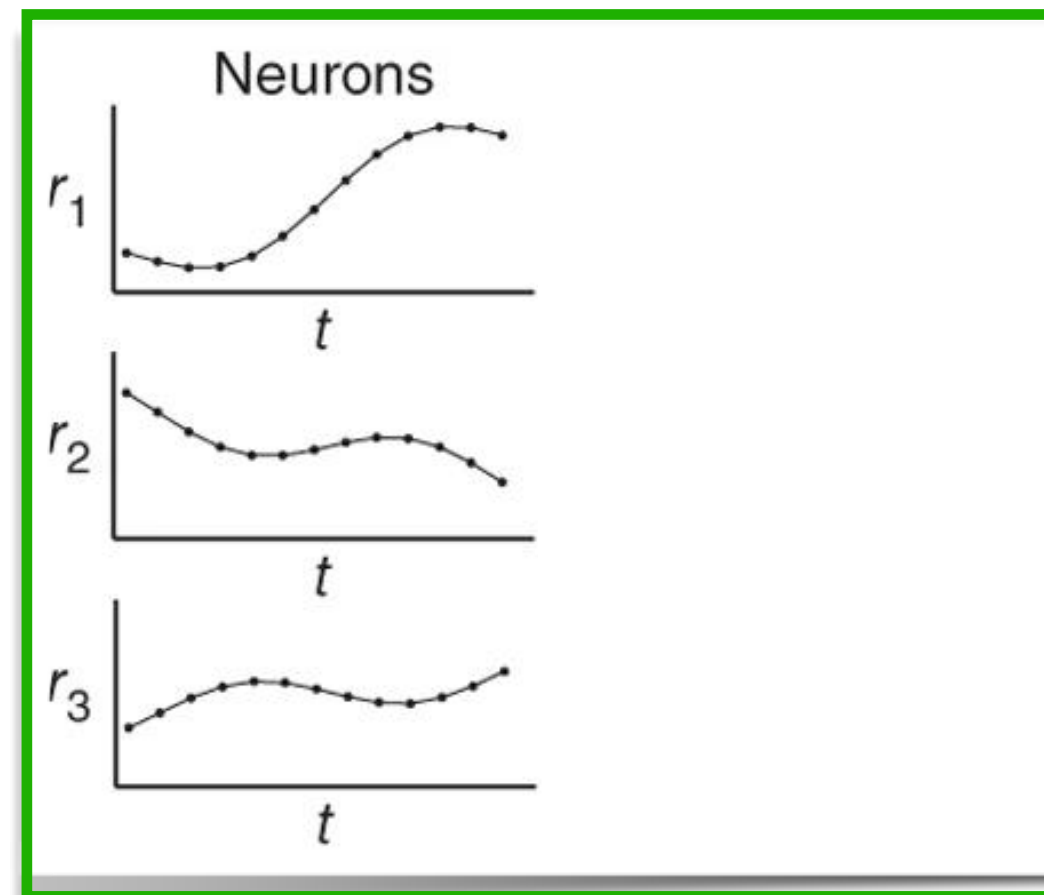


Measuring Neuronal Activity in the Mouse Brain



Investigating Abstract Stimulus Representations in mPFC

Data

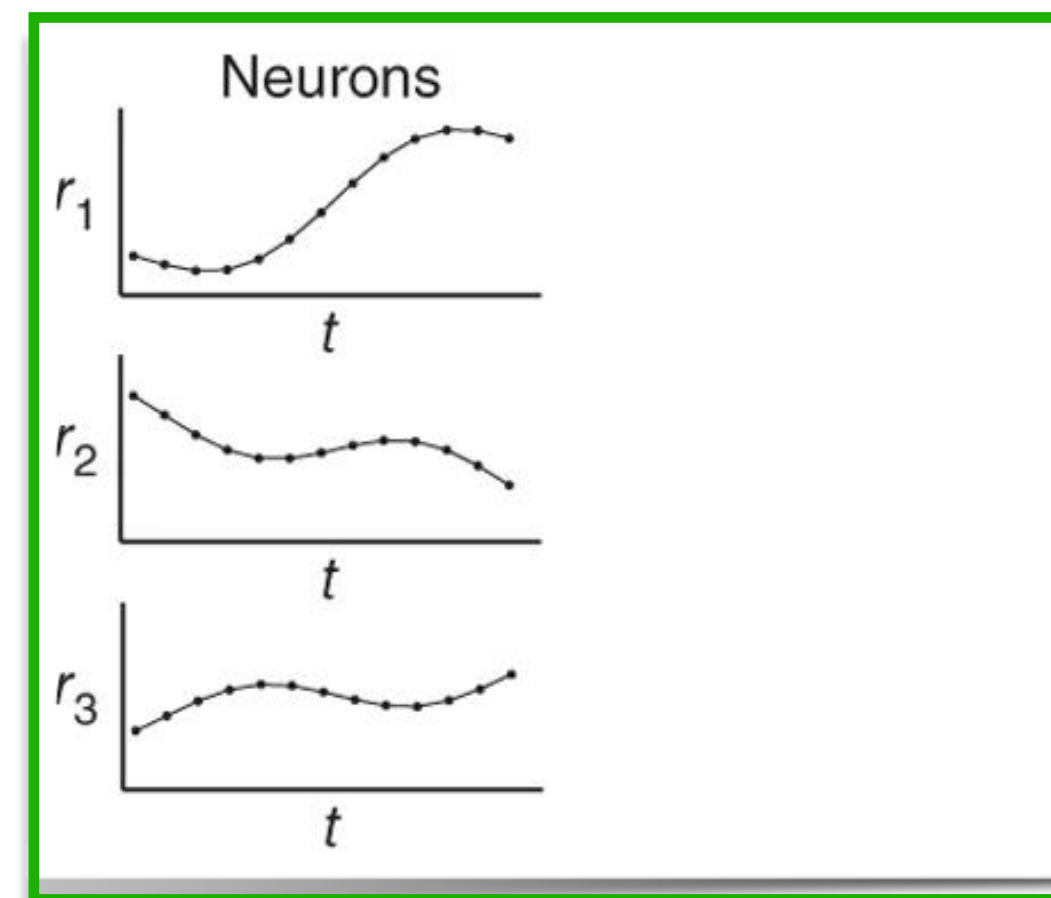


n=12 mice,
3395 neurons

**The bus neuron is actually
100 or 1000 neuron's. The bus is encoded
as a pattern of activity.**

Investigating Abstract Stimulus Representations in mPFC

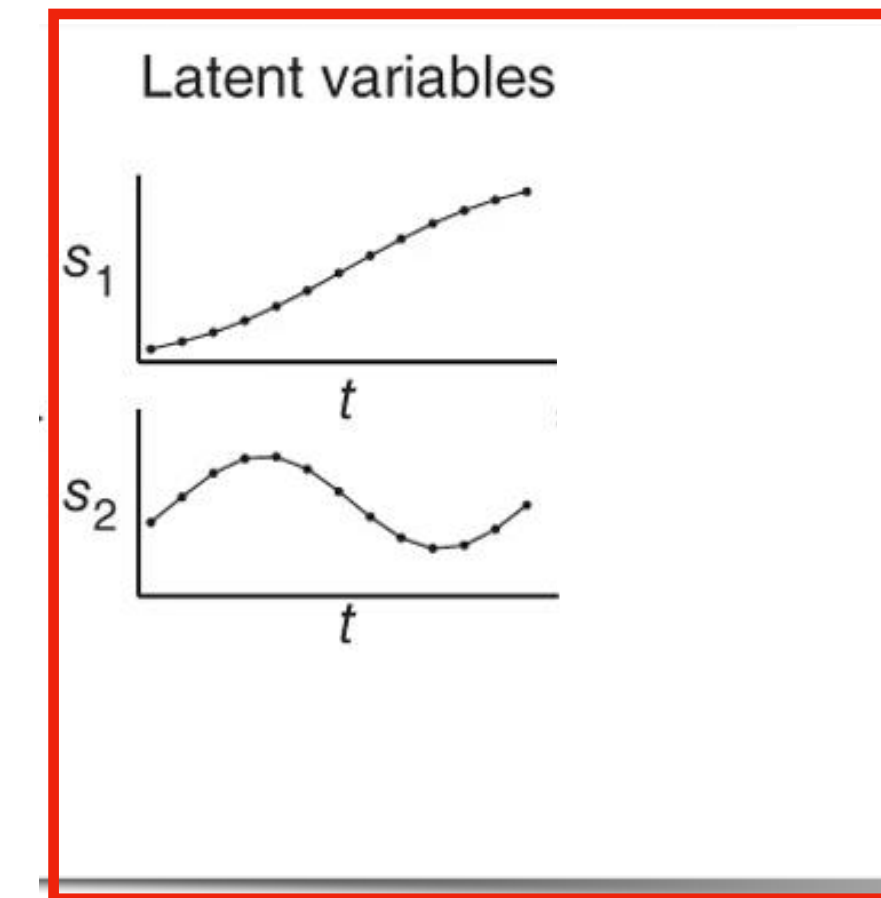
Data



n=12 mice,
3395 neurons

**The bus neuron is actually
100 or 1000 neuron's. The bus is encoded
as a pattern of activity.**

Task Variables

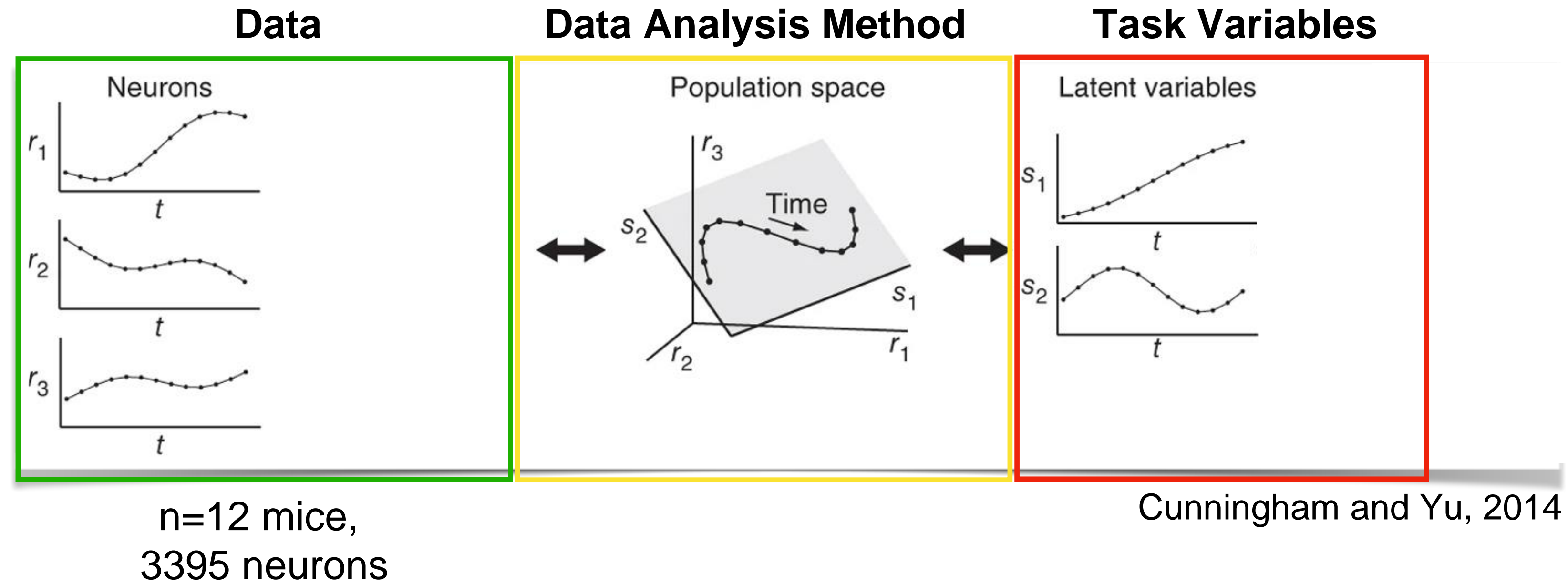


Cunningham and Yu, 2014

Interesting latent/task variables for us:

- Shuttle motion
- Auditory tone stimulus
- Direction of shuttle motion

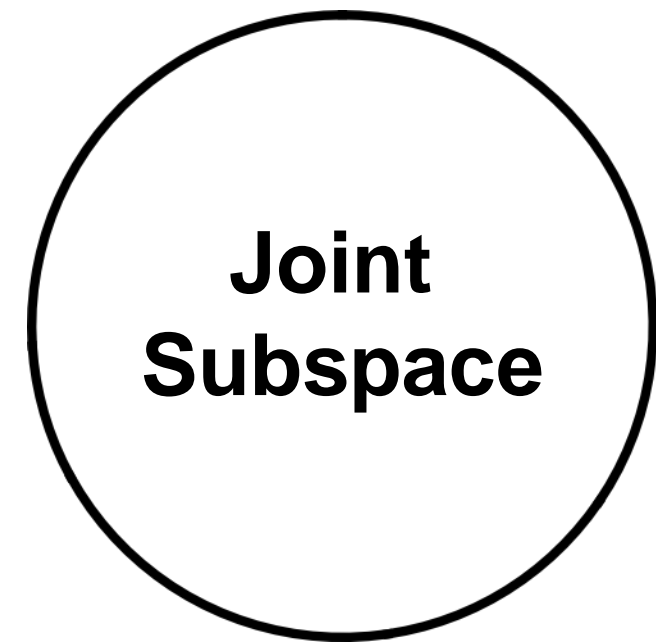
Investigating Abstract Stimulus Representations in mPFC



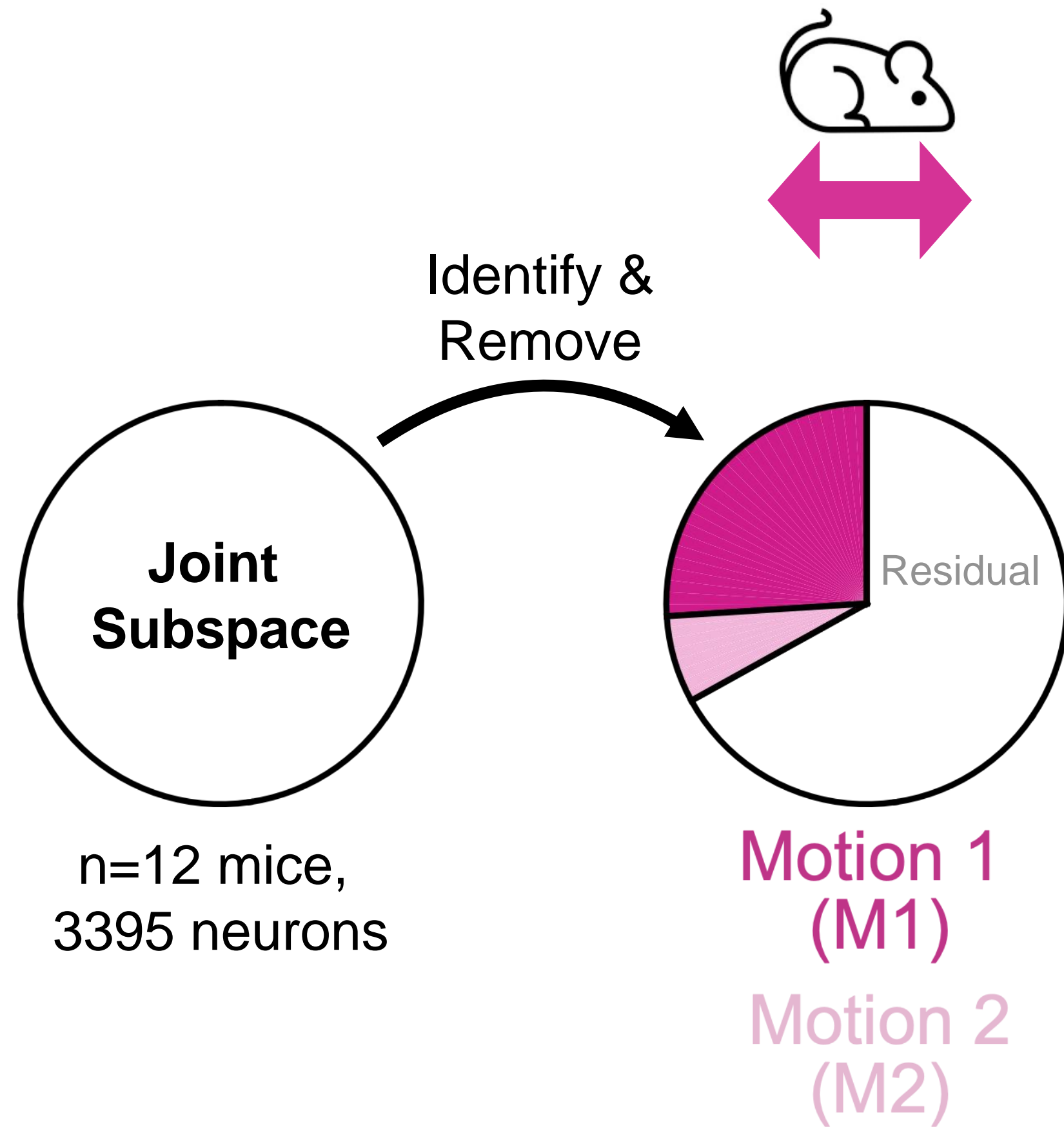
The bus neuron is actually 100 or 1000 neuron's. The bus is encoded as a pattern of activity.

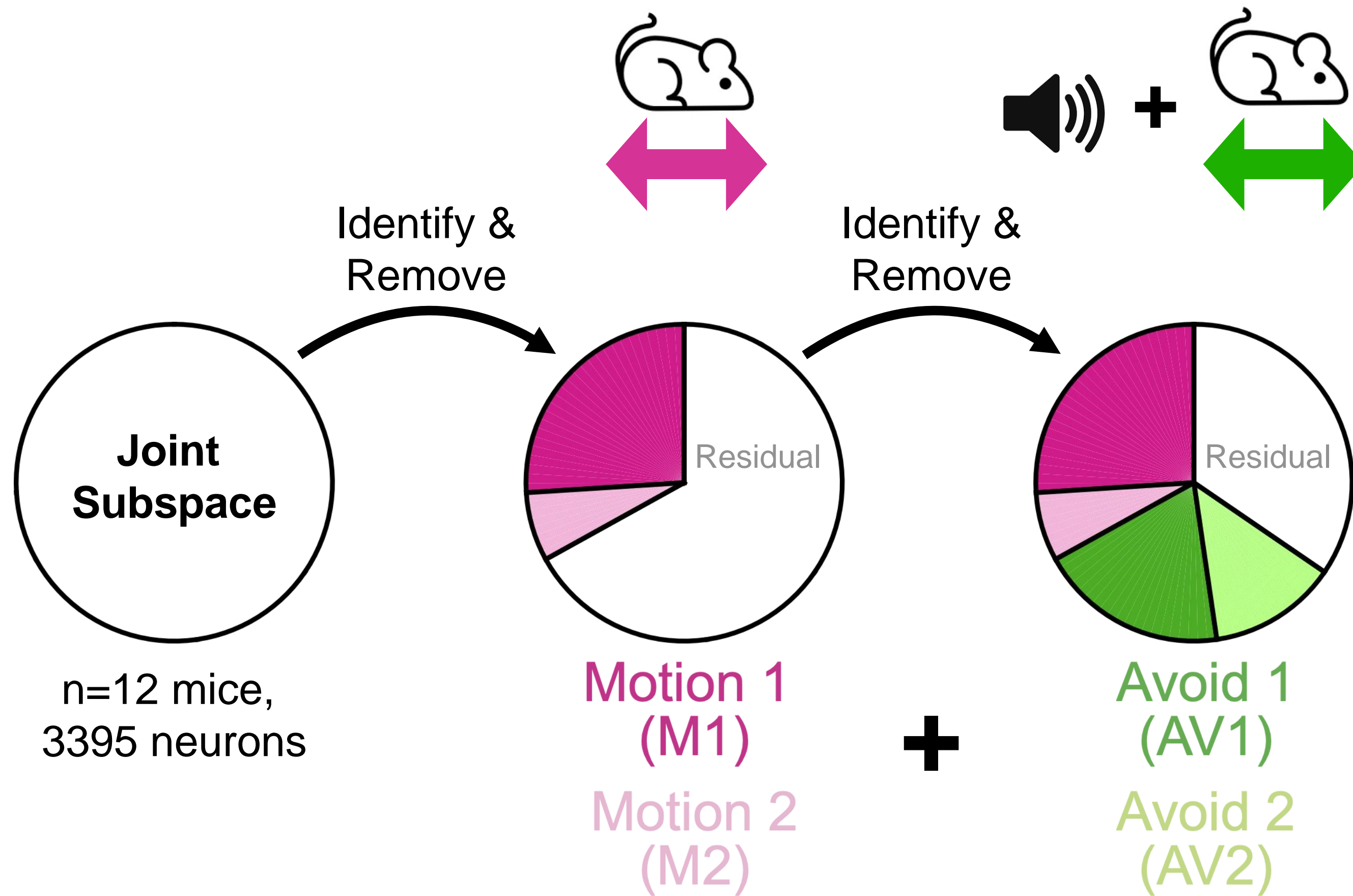
Interesting latent/task variables for us:

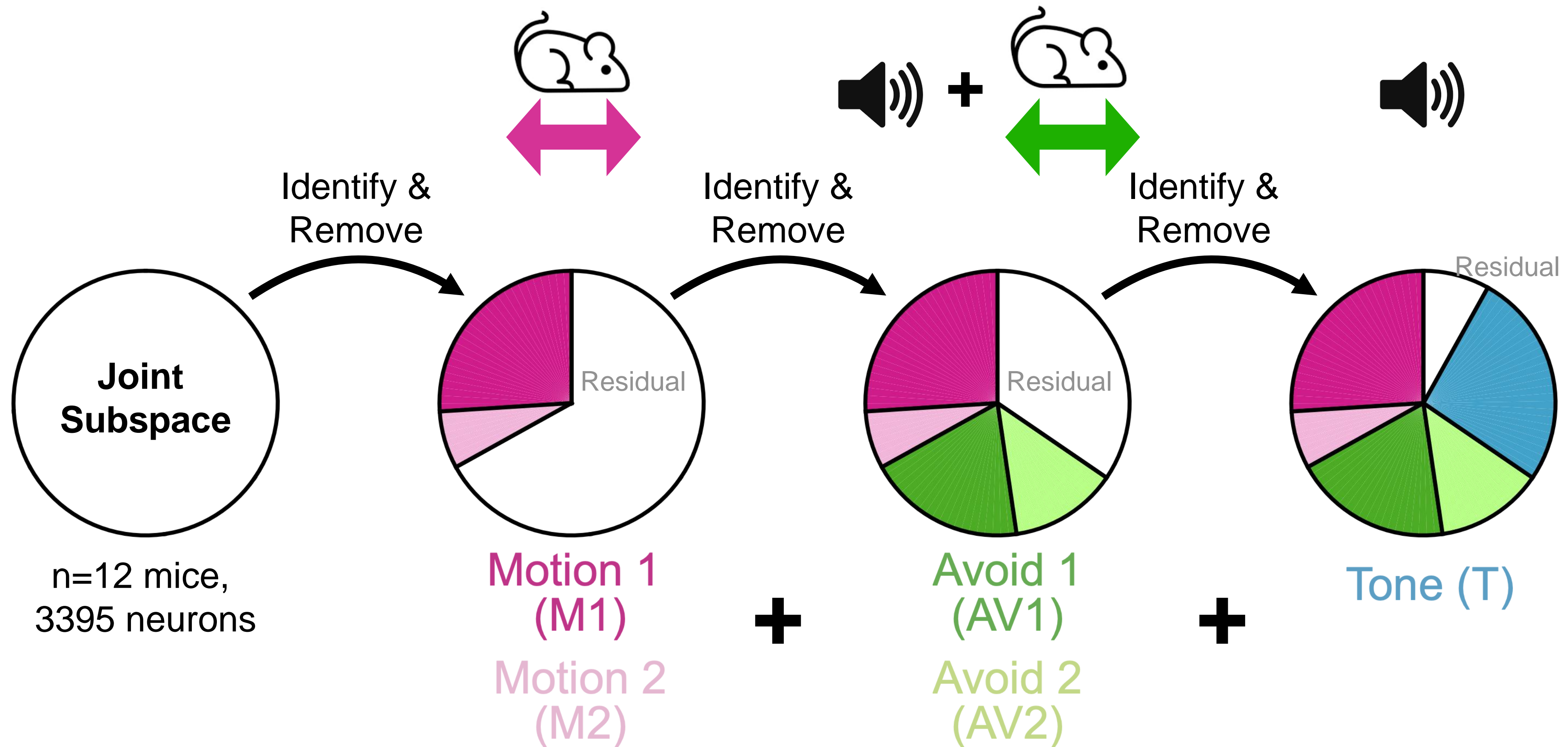
- Shuttle motion
- Auditory tone stimulus
- Direction of shuttle motion

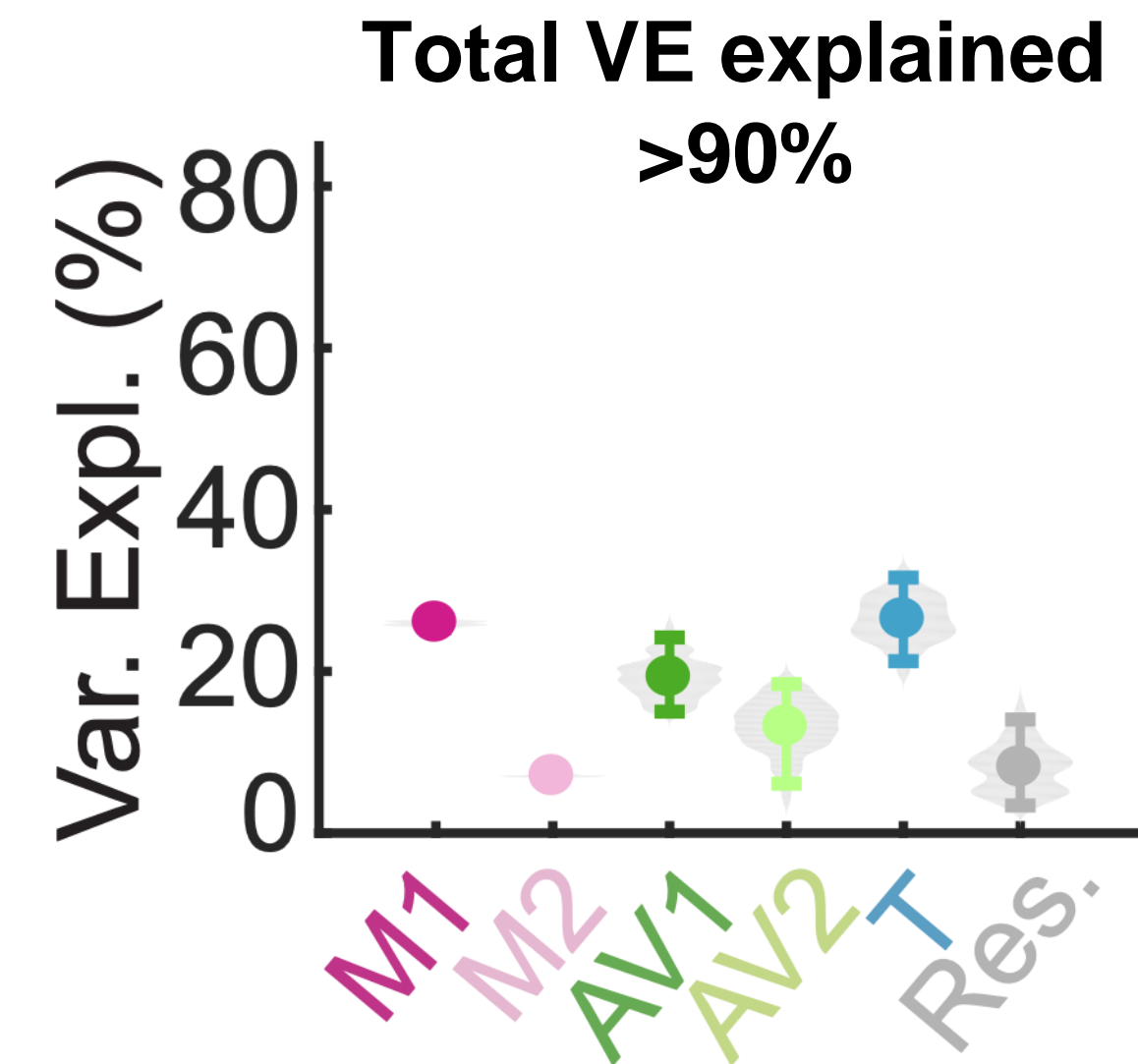
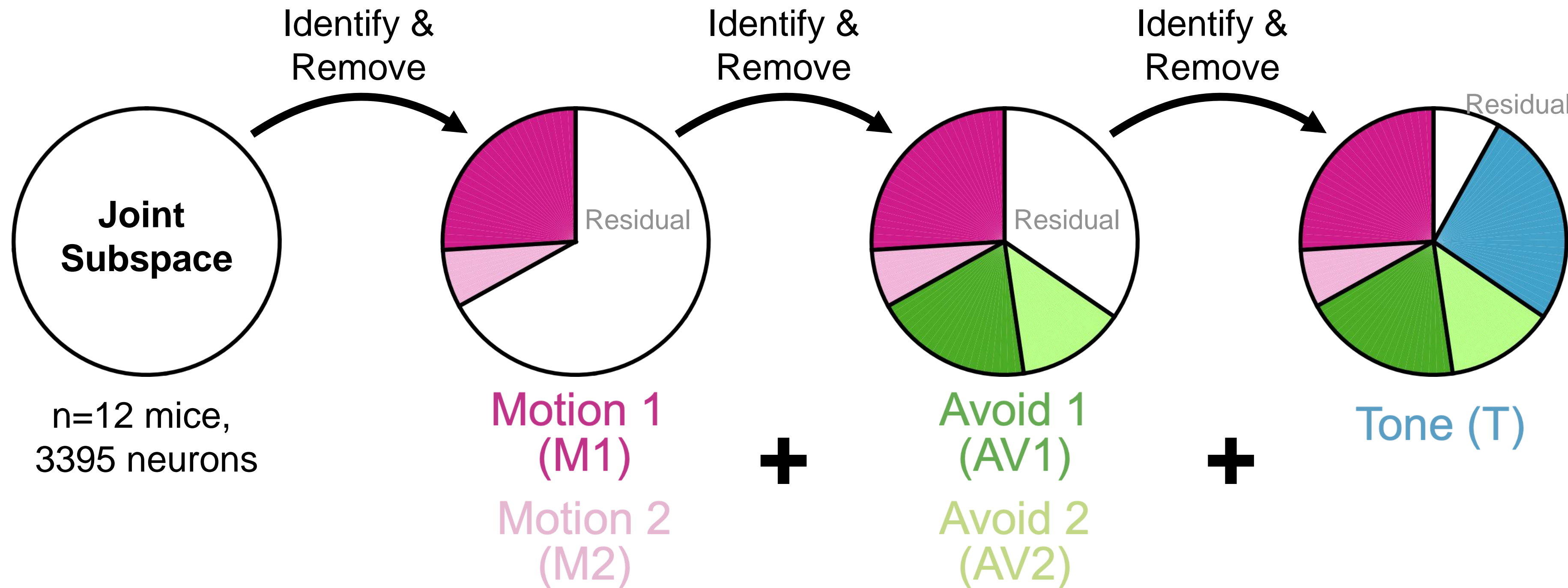


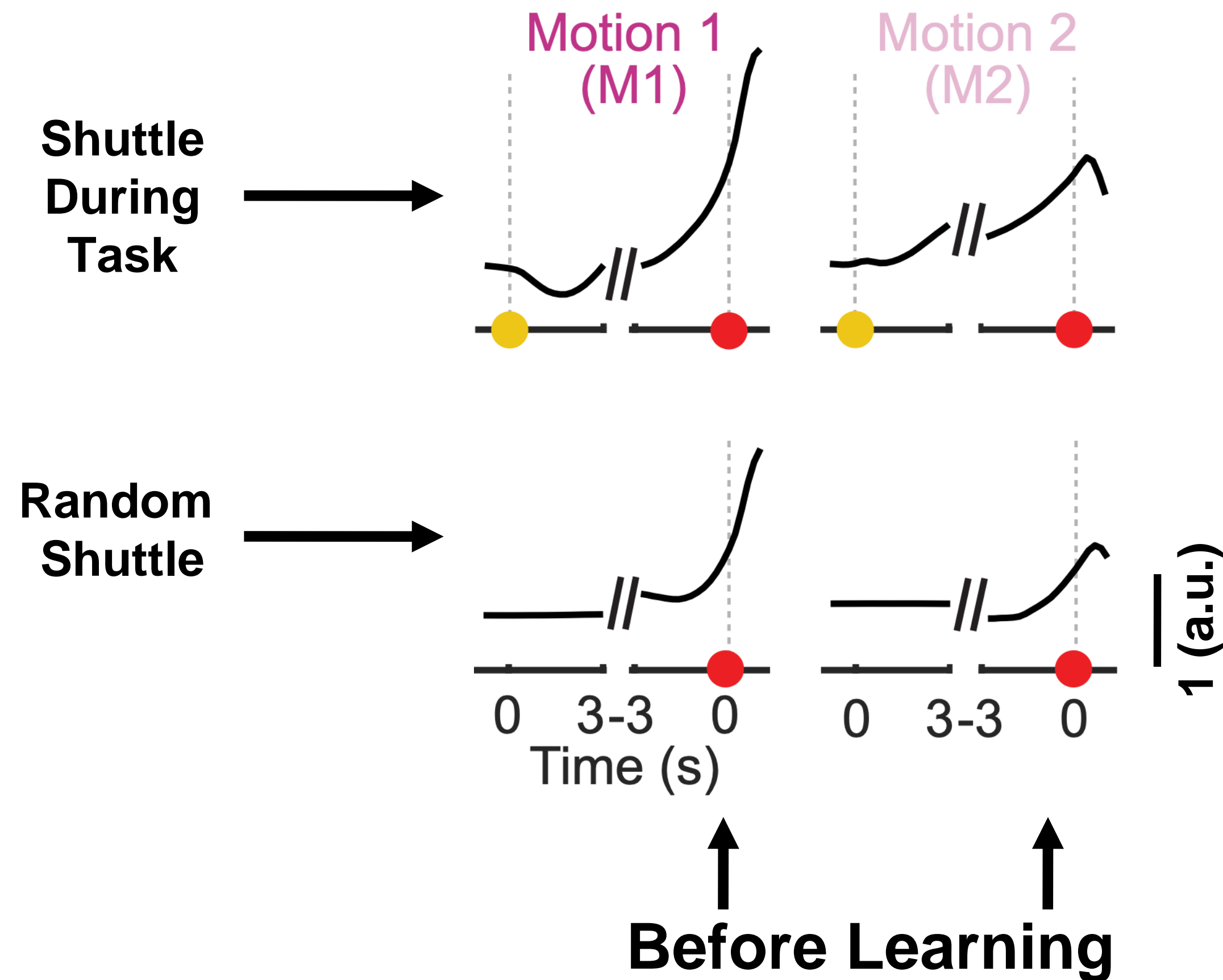
n=12 mice,
3395 neurons



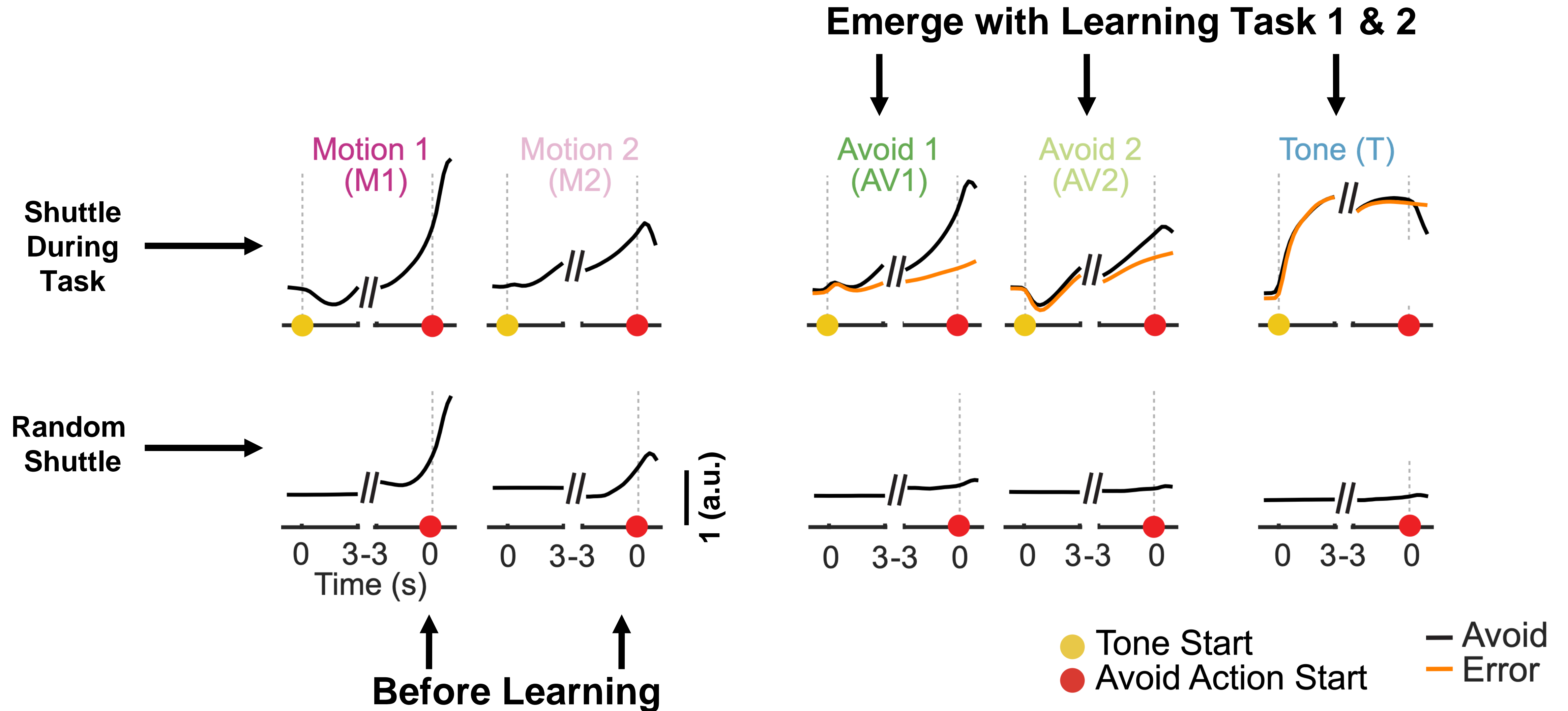


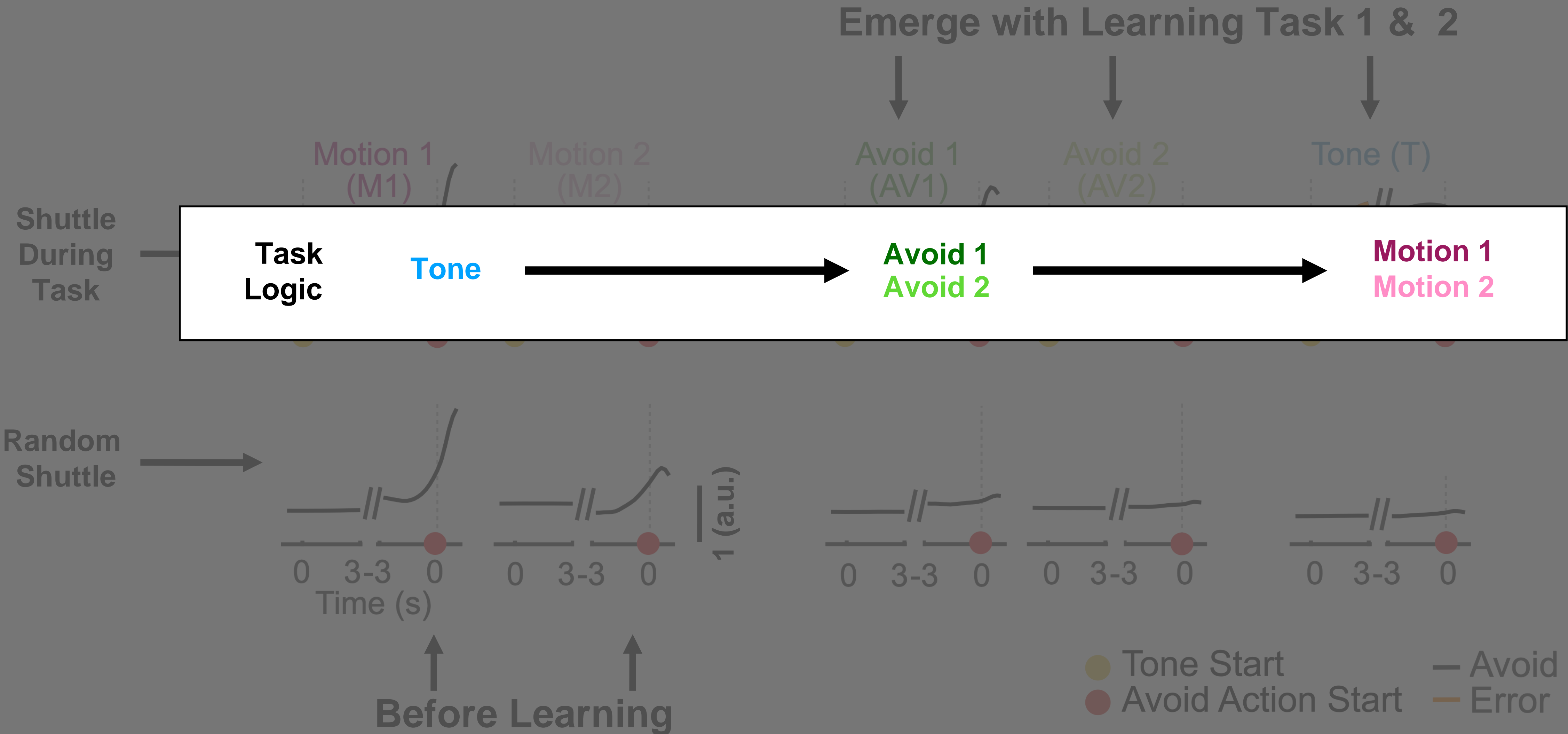






- Tone Start
- Avoid Action Start





Emerge with Learning Task 1 & 2

Emerge with Learning Task 1 & 2

Motion 1 (M1)

Motion 2 (M2)

Avoid 1 (AV1)

Avoid 2 (AV2)

Tone (T)

Shuttle During Task

Task Logic

Tone

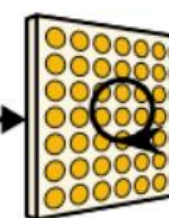
Avoid 1
Avoid 2

Motion 1
Motion 2

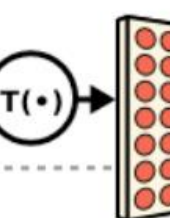
Random Shuttle



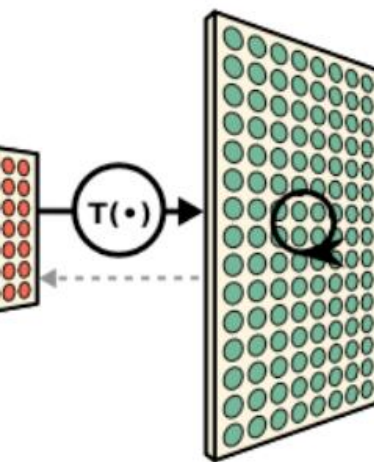
input



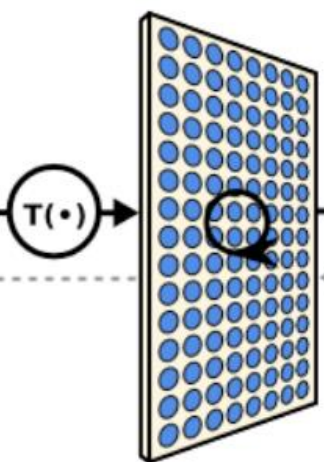
retina



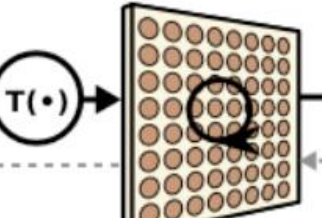
LGN



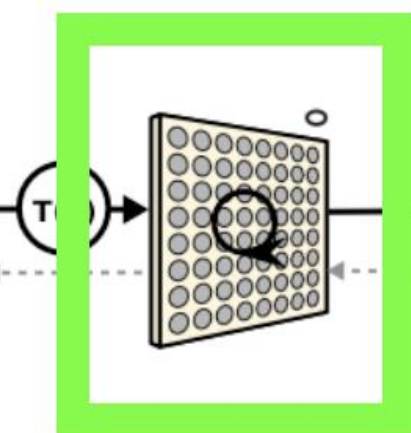
V1



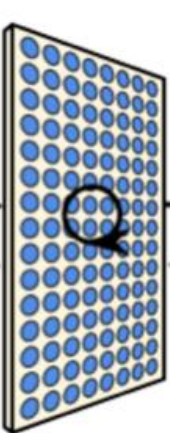
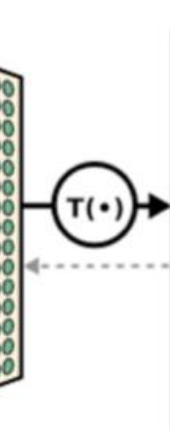
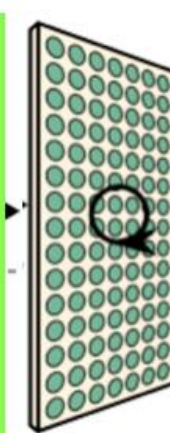
V2



V3



Medial Prefrontal Cortex (mPFC)



Time (s)

Before Learning

- Tone Start
- Avoid Action Start

A candidate neural representation for 'Affordance'.

Prof. Dr. Jean Piaget

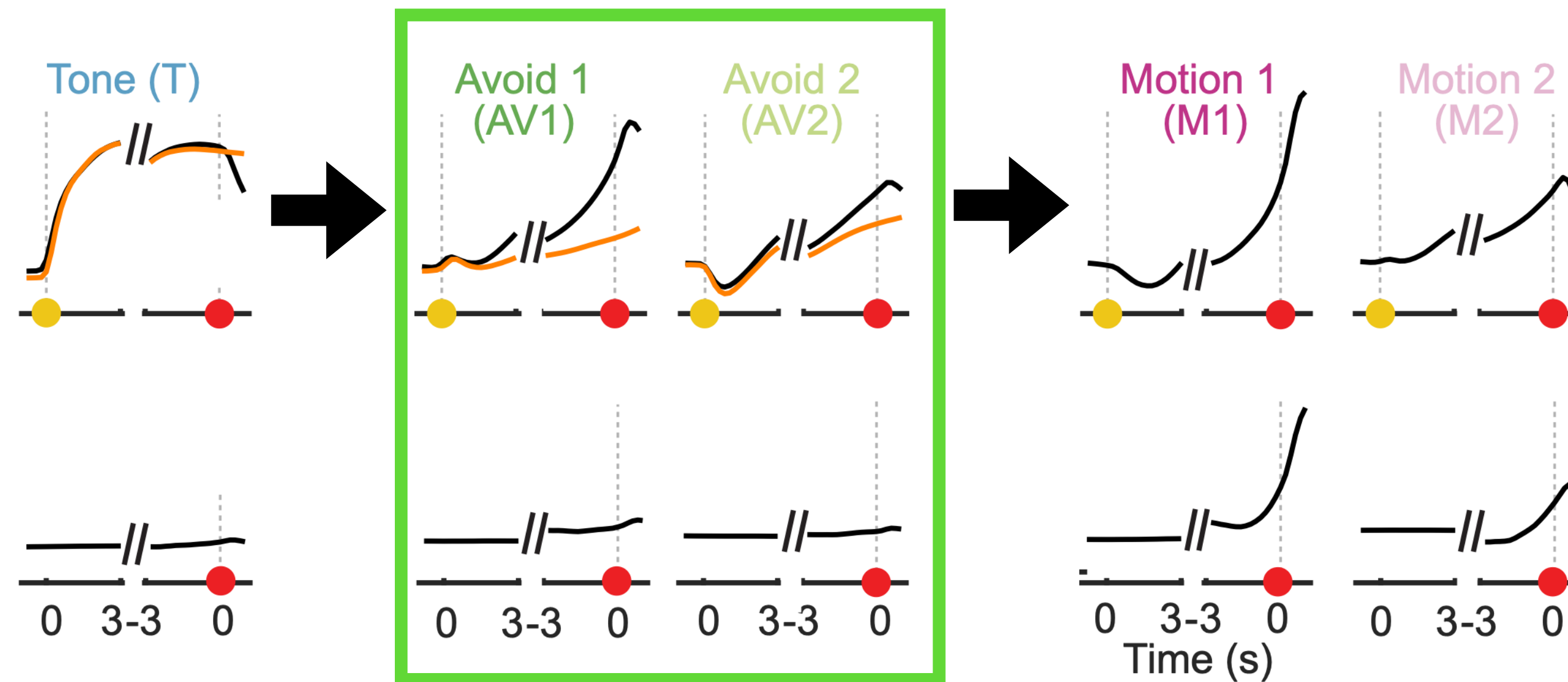
Swiss psychologist and pioneer
Neuchatel. 1896-1980



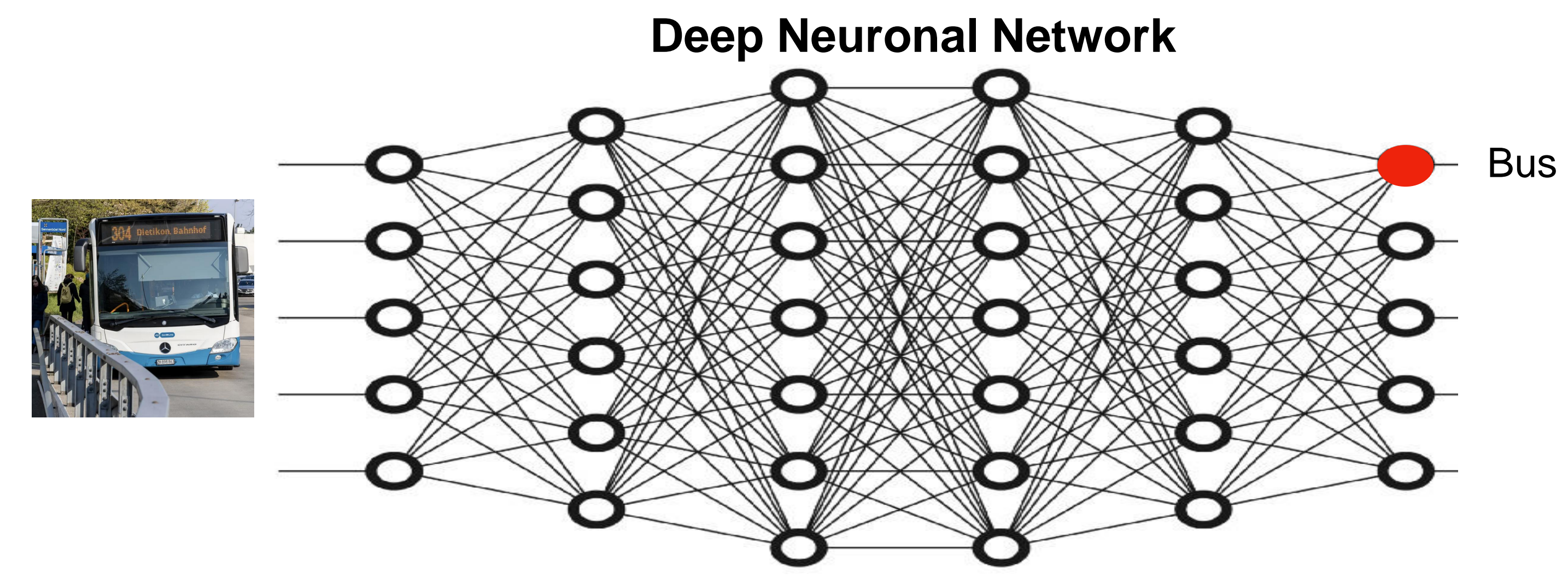
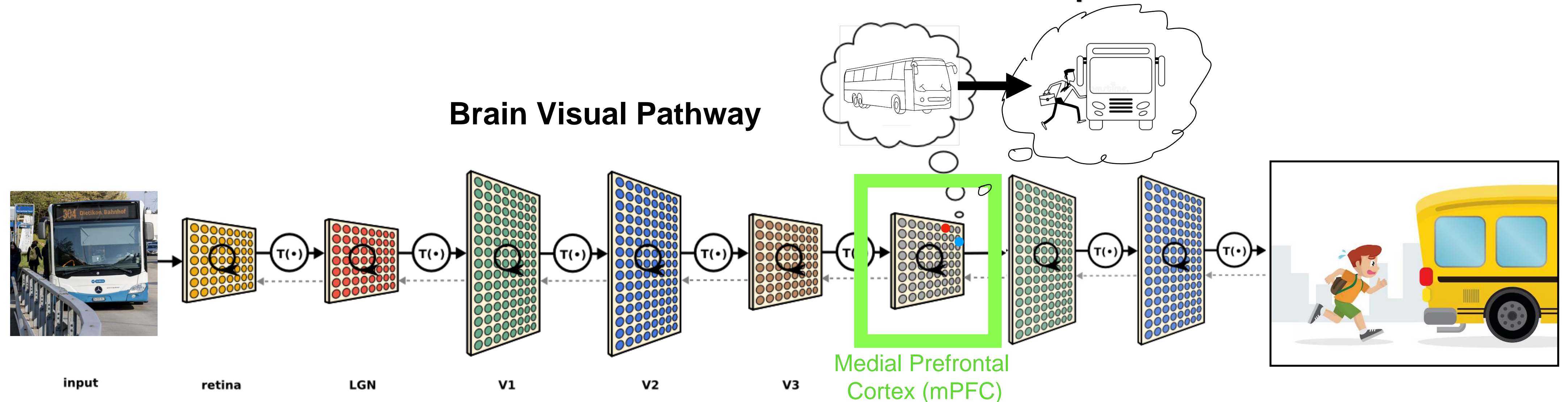
The Concept of 'Affordance' in Psychology

Affordance alludes to the qualities of an object or situation that define its possible use or make clear how it can or should be used.

Affordance adheres to the idea that perception and action are inseparable (Principles of Genetic Epistemology, Piaget).

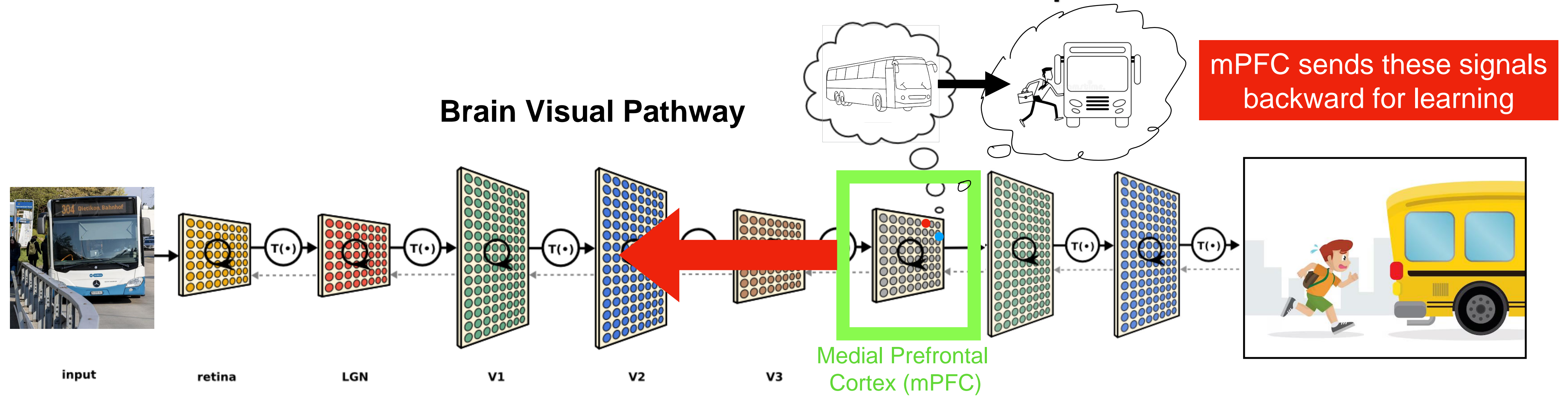


Conclusions Part II: The Nature of Semantic Representations

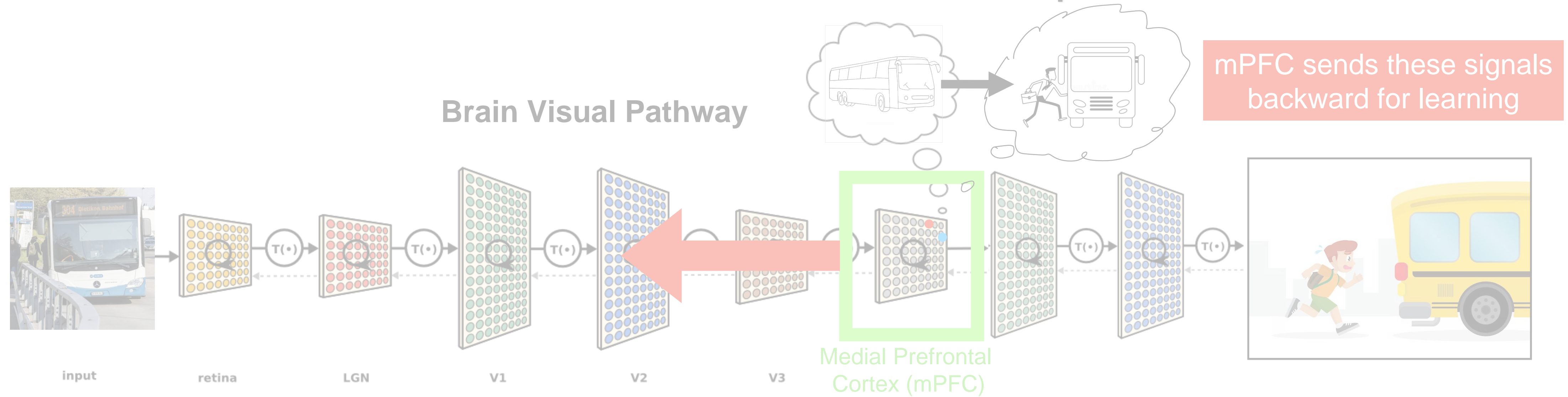


Ehret et al. final revisions @ Nature Neurosc.

Conclusions Part II: The Nature of Semantic Representations



Conclusions Part II: The Nature of Semantic Representations

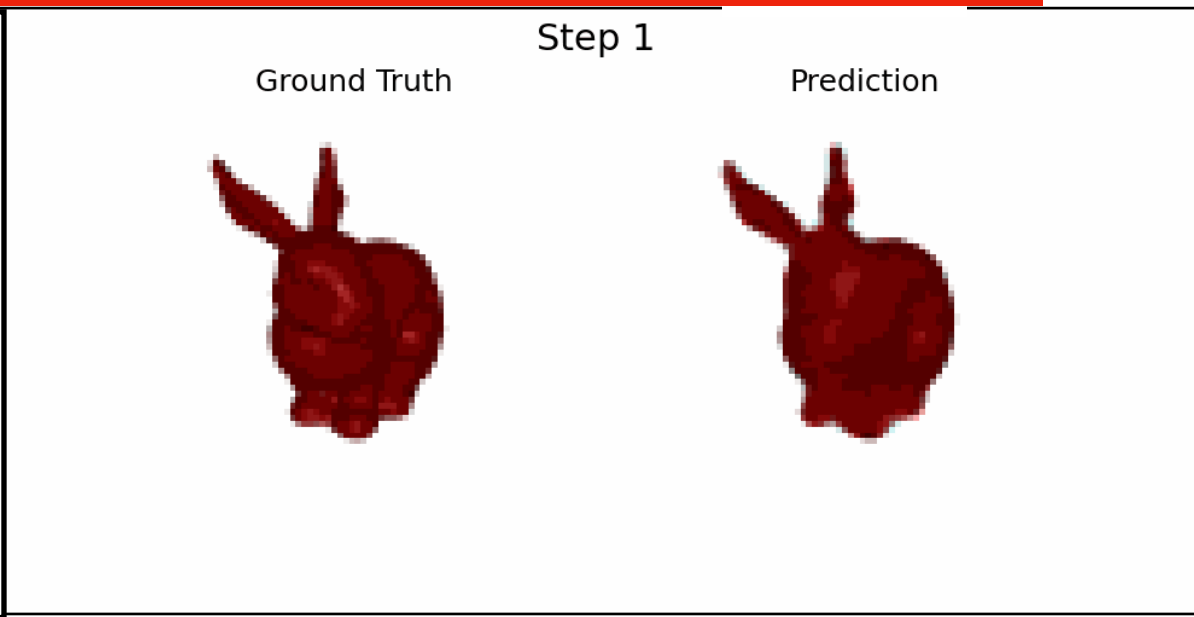


We are using motor signals to supervise the hierarchical learning of objects.

ZUKUNFTSBLOG · DIGITALISIERUNG

KI muss lernen wie ein Kind

Neue KI-Systeme verblüffen, doch Menschen reichen sie nicht das Wasser. Benjamin Grewe plädiert deshalb dafür, dass intelligente Maschinen von morgen so lernen wie kleine Kinder.

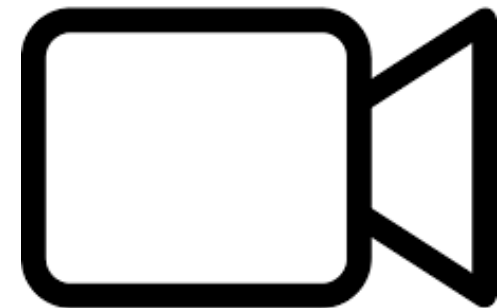


Hamza Keurti

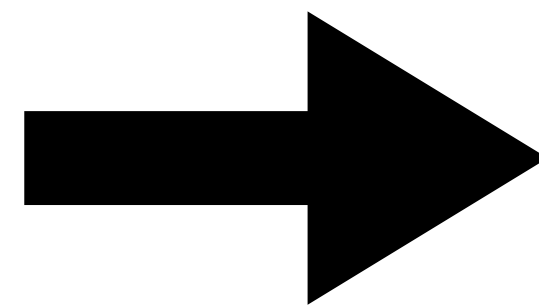
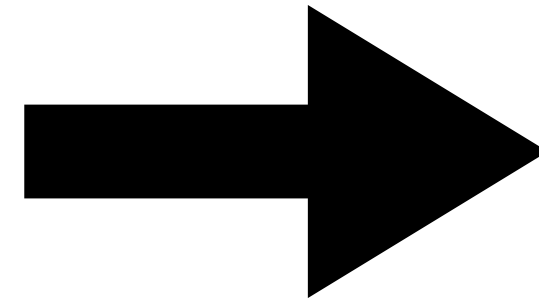
Keurti et al., 2023, ICLM

Prompt:

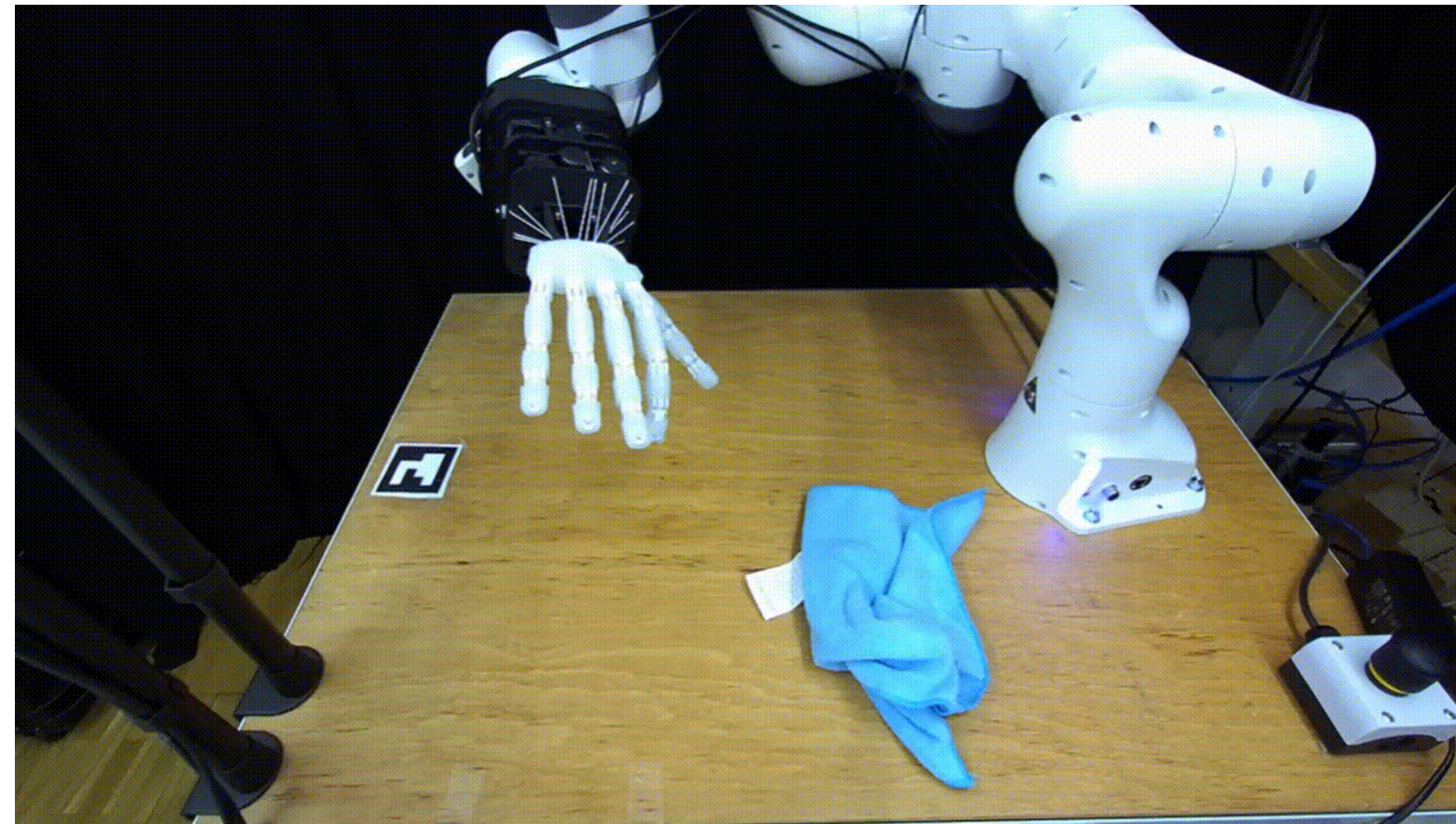
- Wipe the table with thin cloth.
- Pick cube and place in bowl.
- Pick lying bottle from bowl.
- Pick coffee cup from pedestal
- Push bowl around the table.
- Pick standing bottle from pedestal.
- Pick cube and place in bowl, with distractors.



Camera Image
Table Scence



Robotic Action Transformer (RAT)



Movie Credit Elvis Nava



Elvis Nava



Robert Katzschamn

Developed @



ETH AI CENTER

ChatGPT:	Prompt to Prompt
Dalle:	Prompt to Image
RAT:	Prompt to Action



Industry Partner of



ETH AI CENTER

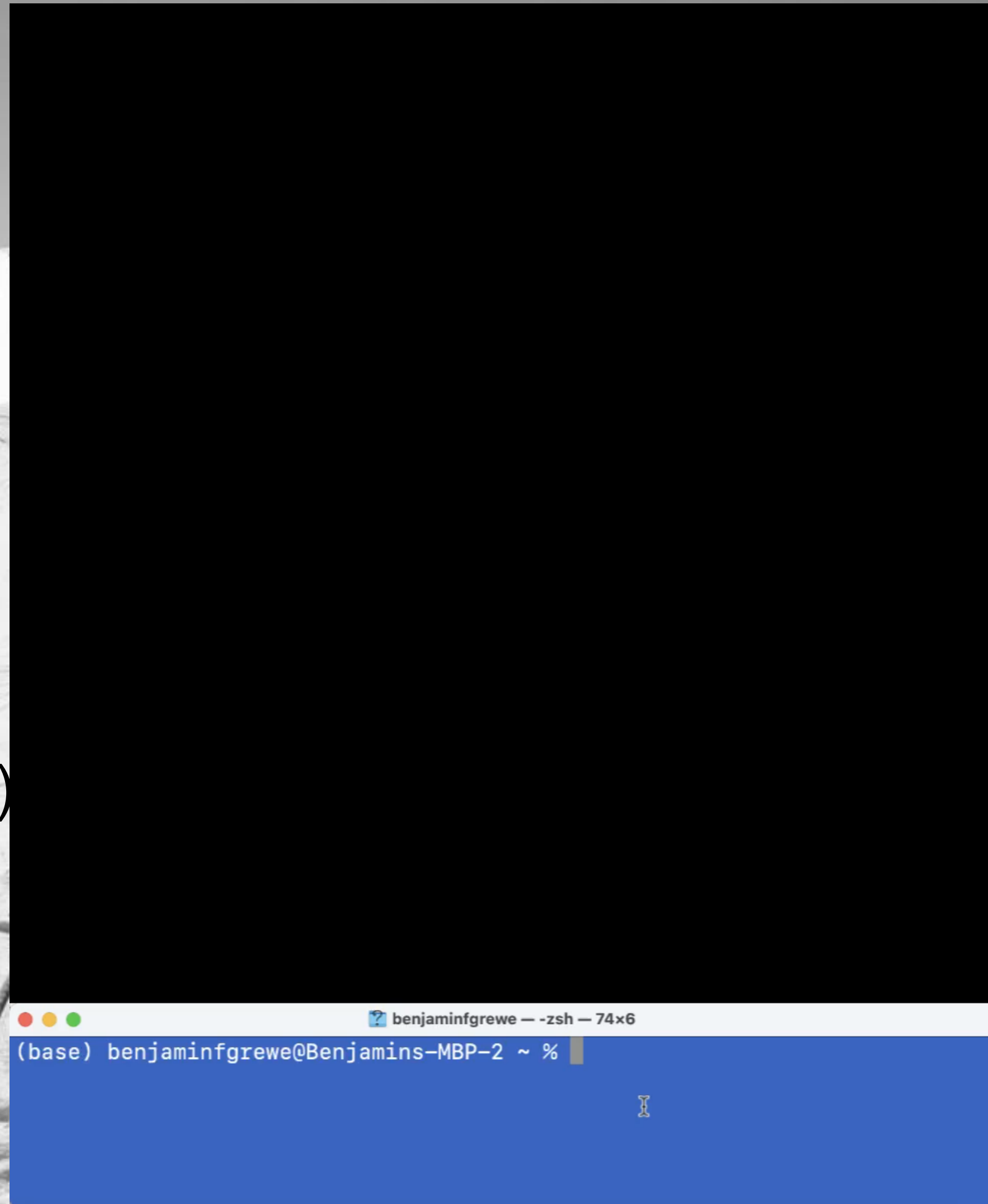


Virtual Action
Transformer (VAT)

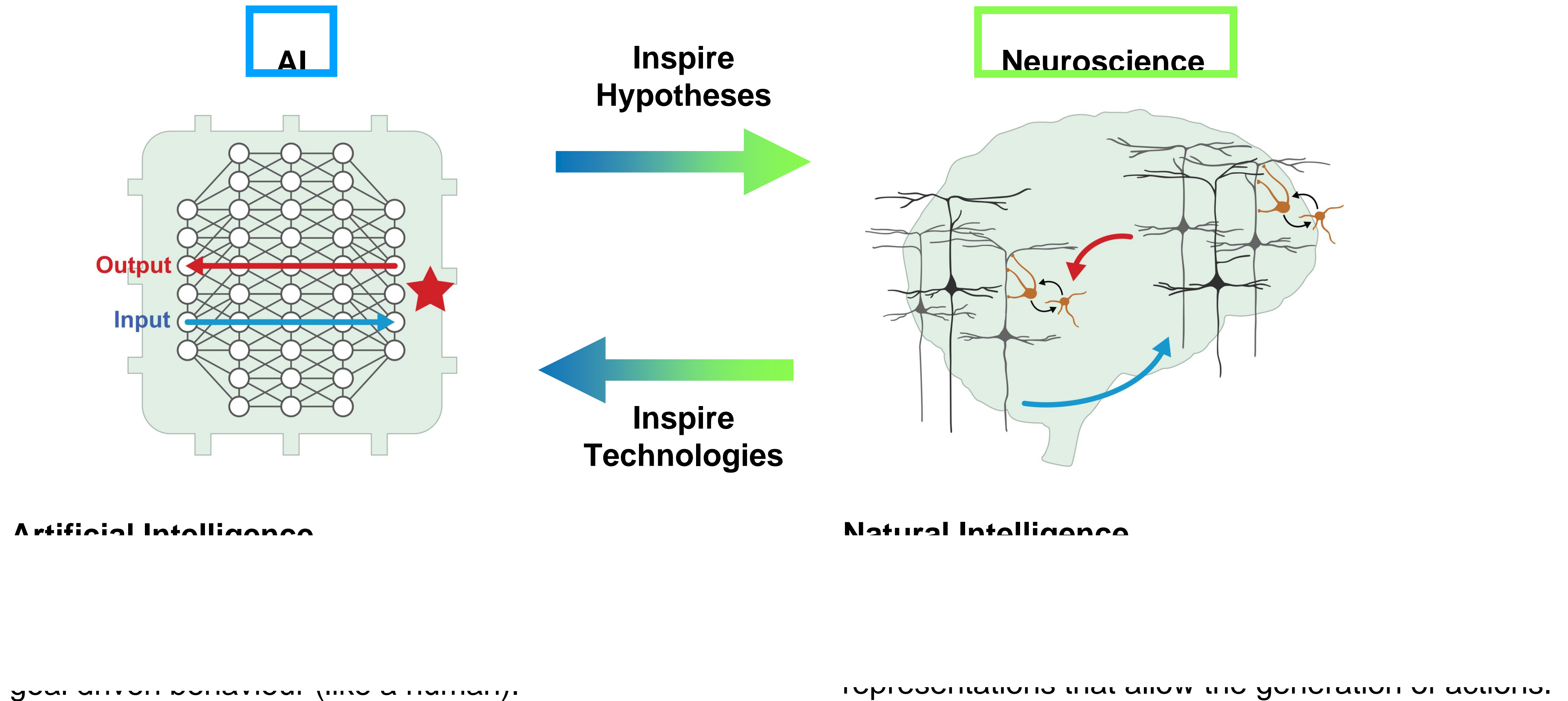
Industry Partner of



ETH AI CENTER



Summary: Advancing Neuroscience and AI

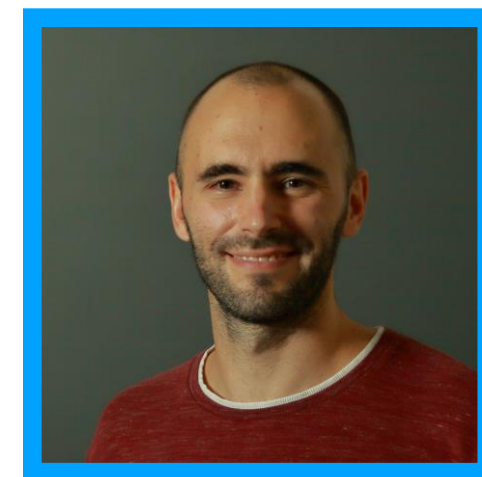


Part I: Learning in Hierarchical (Deep) Cortical Networks

AI



Neuroscience



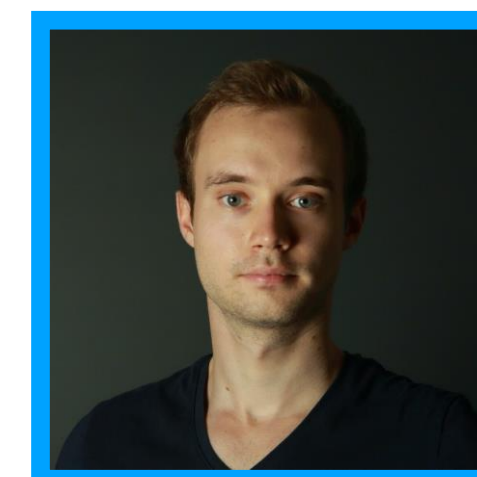
Pau Aceituno



Matilde T. Farinha



Alexander Meulemans



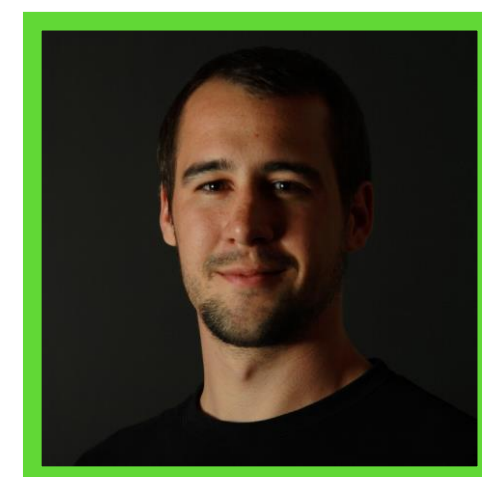
Francesco Laessig

Part II: Understanding Hierarchical Neuronal Representations in Brain

AI



Neuroscience



Benjamin Ehret



Liz Ann Amadei



Roman Boehringer



Valerio Mante

