

Session: The Impact of AI on Neuroscientific Discoveries

James DiCarlo MD, PhD

Director, MIT Quest for Intelligence

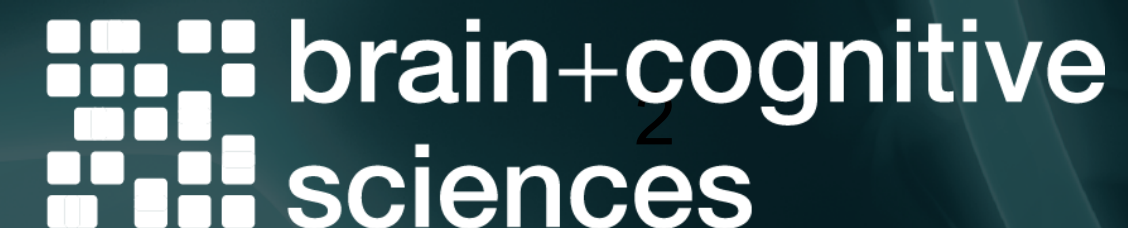
Professor of Systems & Computational Neuroscience

Department of Brain and Cognitive Sciences

Co-Director & Investigator, Center for Brains, Minds and Machines

Investigator, McGovern Institute for Brain Research

Massachusetts Institute of Technology



Financial disclosures:

Current research support:

- NSF, NIH
- US DoD/ONR
- DARPA/JUMP2.0
- Google, IBM
- McGovern Institute, Simons Foundation, Brain Research Foundation

Scientific Advisory Boards:

- Numenta Inc. (past, hold equity)
- Yale WuTasi Institute (current, compensated)
- Harvard Lefler Center (current, compensated)
- Columbia ARNI NSF Center (current, uncompensated)

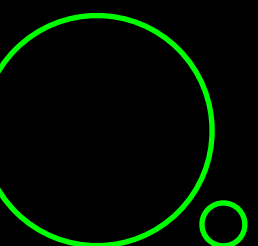
My spouse holds equity in:

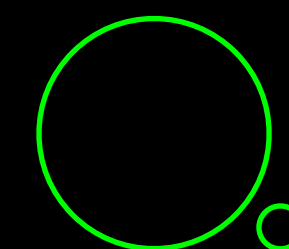
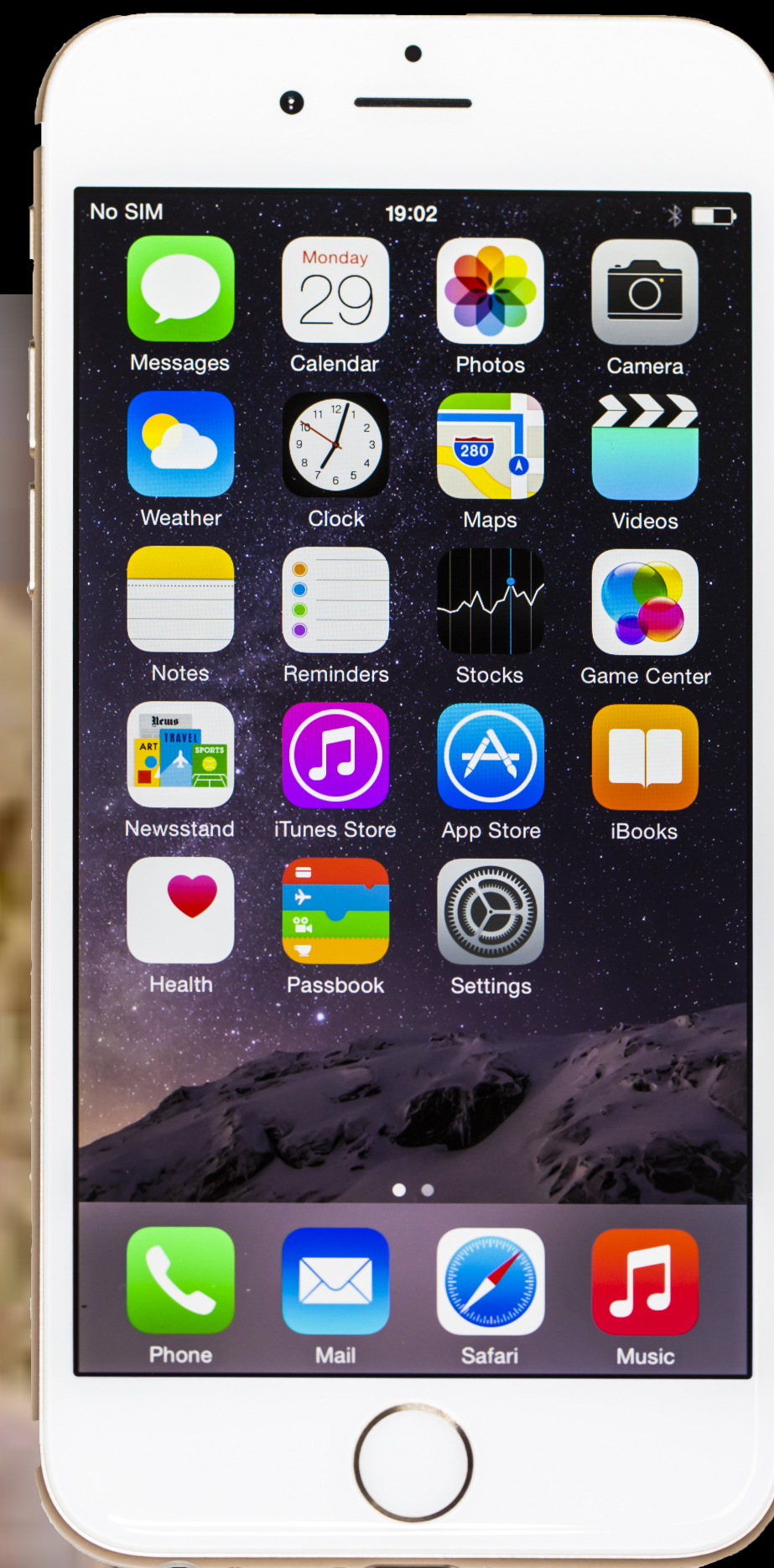
- Akili Interactive
- Rapport Therapeutics

Co-Inventor, US patent # 11813406: SOFTWARE AND METHODS FOR CONTROLLING NEURAL RESPONSES IN DEEP BRAIN REGIONS

Four takeaways:

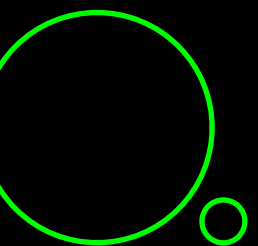
1. The goal of brain and cognitive science is necessarily interlocked with modern AI engineering.
2. Accurate, multi-scale, machine executable models of brain function are emerging from this AI-forward scientific strategy. This will accelerate.
3. The emerging new scientific models will unlock entirely new opportunities to improve human health and education.
4. Because AI engineering is core to brain and cognitive science (1-3), nations that do not appropriately embrace this will be left behind.





Human minds are products of machines and thus can be understood in engineering terms

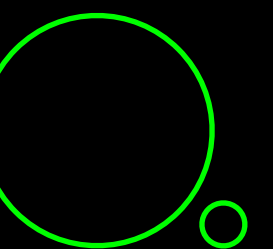
How does [the mind] work?



Human minds are products of machines and thus can be understood in engineering terms

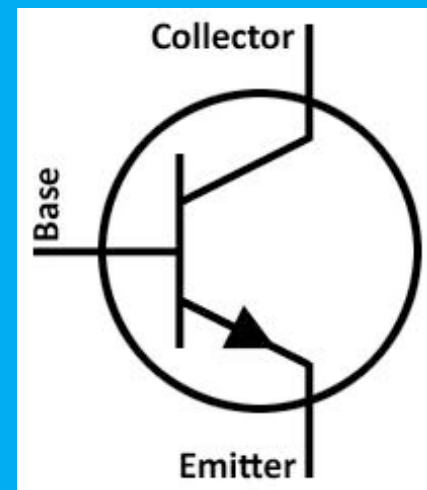


How does neural networks work?

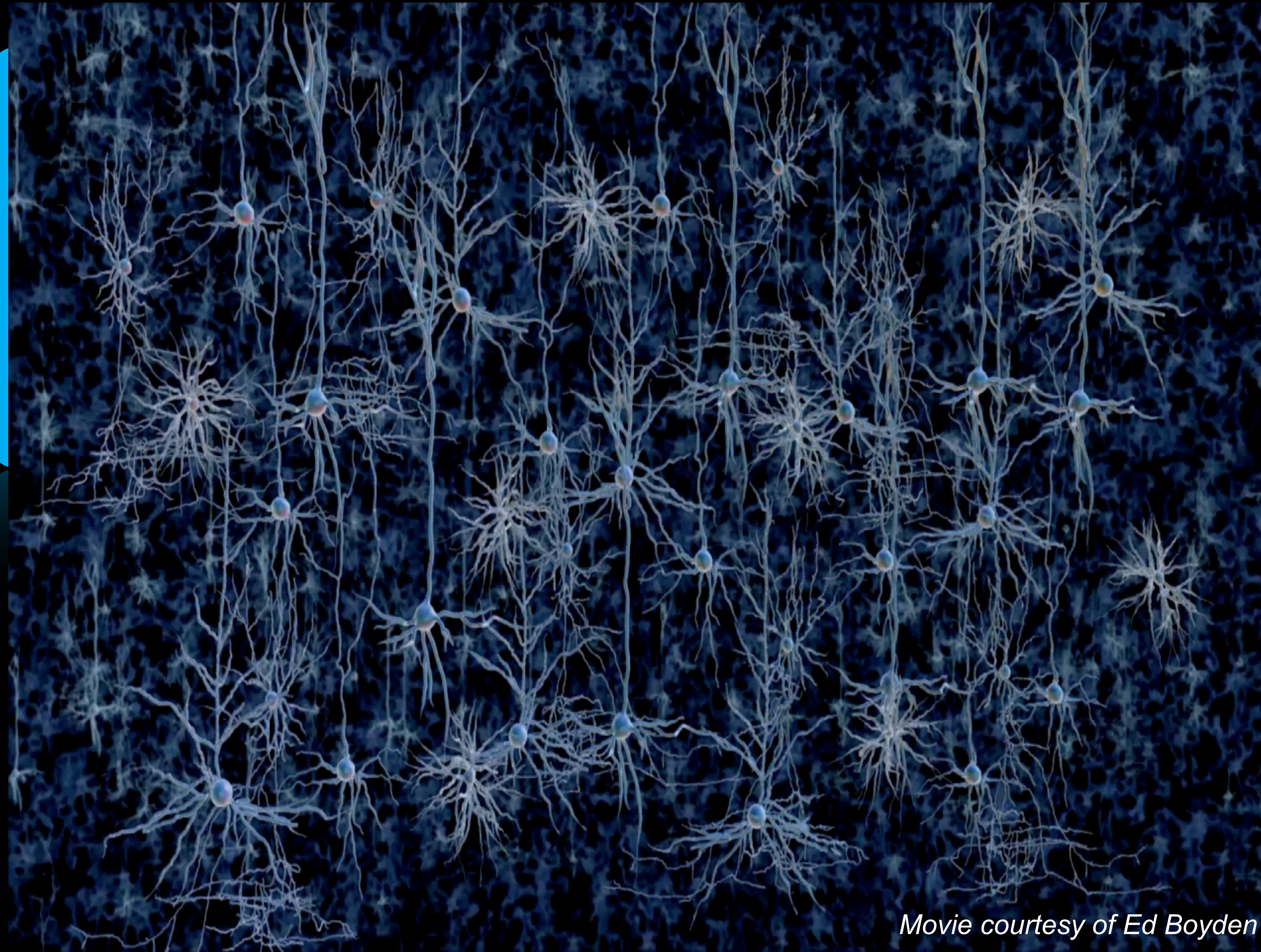
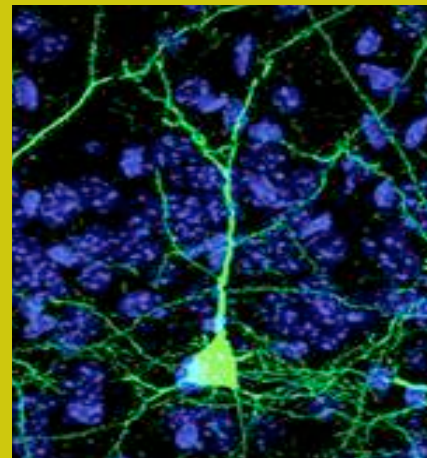


Human minds are products of machines and thus can be understood in engineering terms

Transistors



Neurons /
Synapses



How does
my phone
app work?

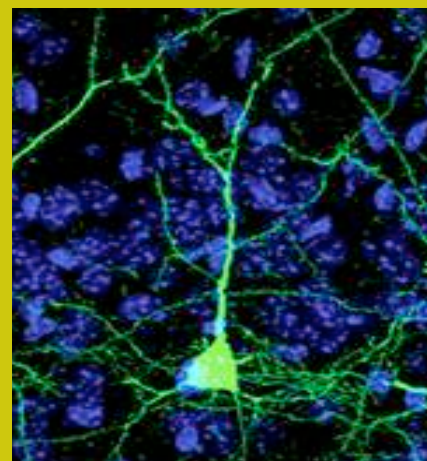


Movie courtesy of Ed Boyden

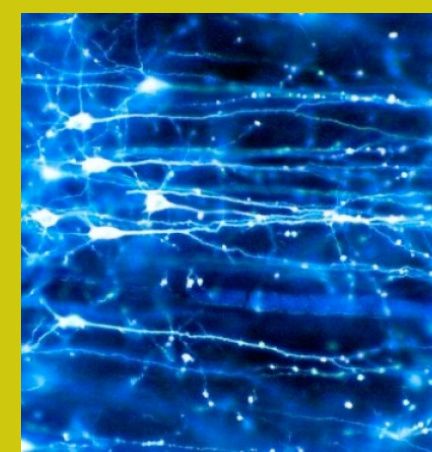
That is, our goal is to deliver **machine executable, multi-scale, predictively accurate, models of human capabilities.**

The goal of brain and cognitive science is to deliver **reproducible, accurate explanations** of how human capabilities (aka human intelligence, aka human minds, aka human cognition) **emerge from the biophysical interactions of the components of the brain** (aka “neural mechanisms”) and their interactions with the body.

Neurons /
Synapses



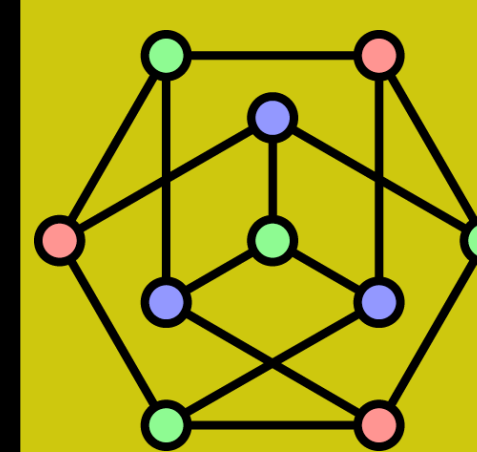
Neuronal
Circuits



Brain
Regions



States &
Algorithms



Behavior/
Cognition



That is, our goal is to deliver machine executable, multi-scale, predictively accurate, models of human capabilities.

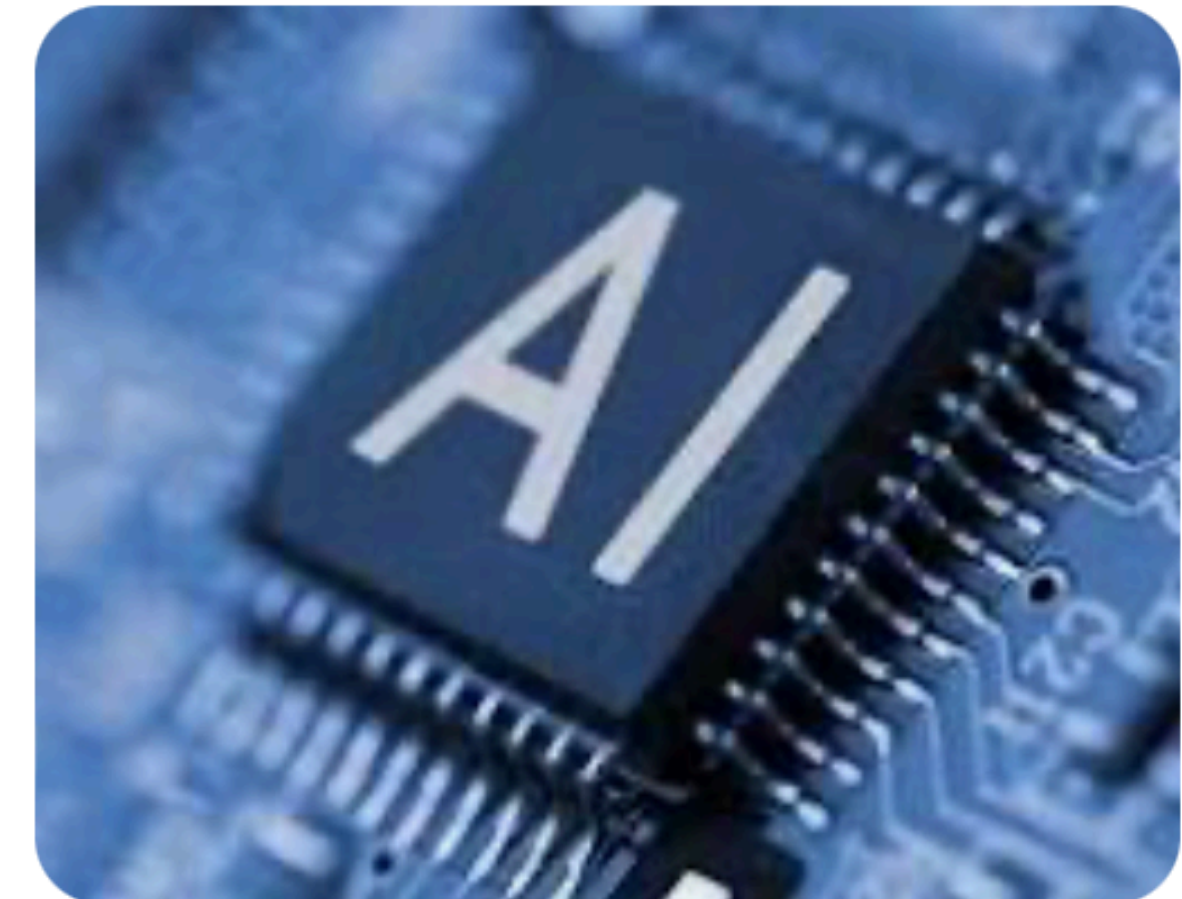


what is the goal of AI



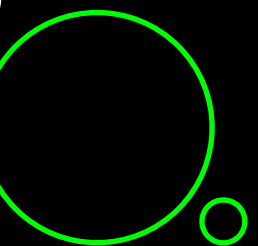
What is a goal of artificial intelligence?

At its core, the primary goal of Artificial Intelligence is to simulate human-like intelligence in machines,



Jul 27, 2023

Takeaway #1: The goal of brain and cognitive sciences is necessarily interlocked with modern AI engineering.

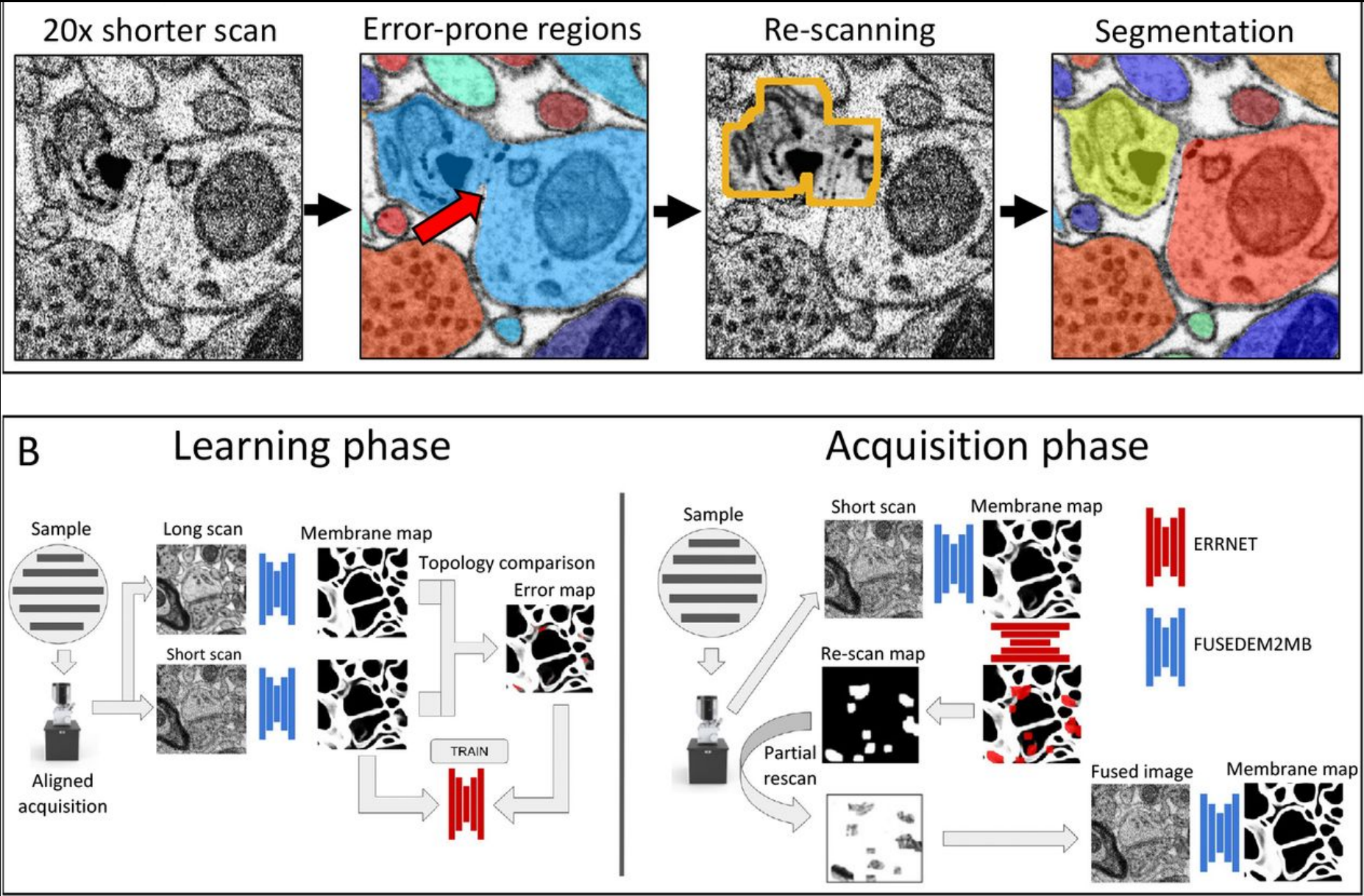


Session: The Impact of AI on Neuroscientific Discoveries

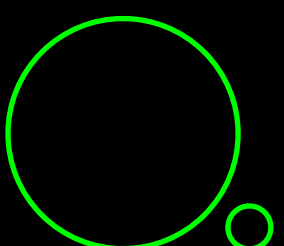
AI methods as scientific tools



Nath, Mathis, Chen, Patel, Bethge & Mathis, ***Nature Protocols***



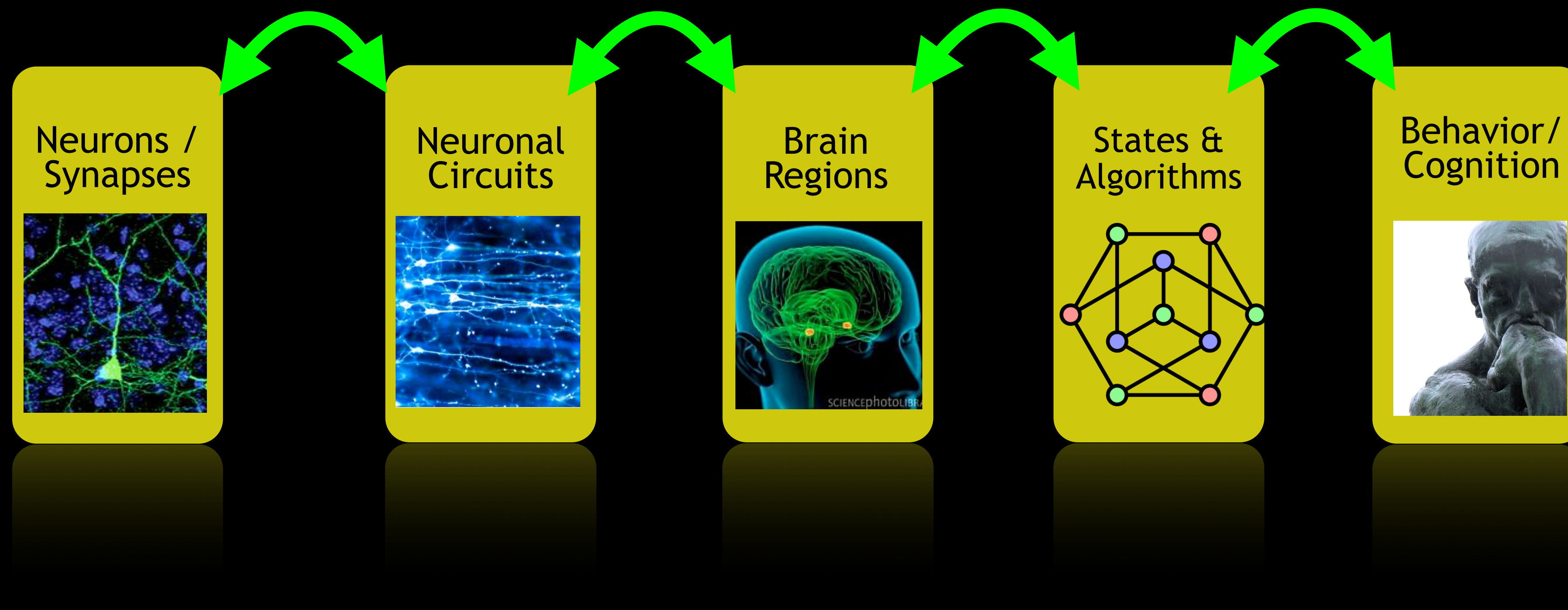
Meirovitch, Park, Mi, Potocek, Sawmya, Li, Wu, Schalek, Pfister, Schoenmakers, Peemen, Lichtman, Samuel & Shavit ***bioRxiv*** (2023).



Session: The Impact of AI on Neuroscientific Discoveries

AI methods as scientific tools

AI-engineered systems as scientific models

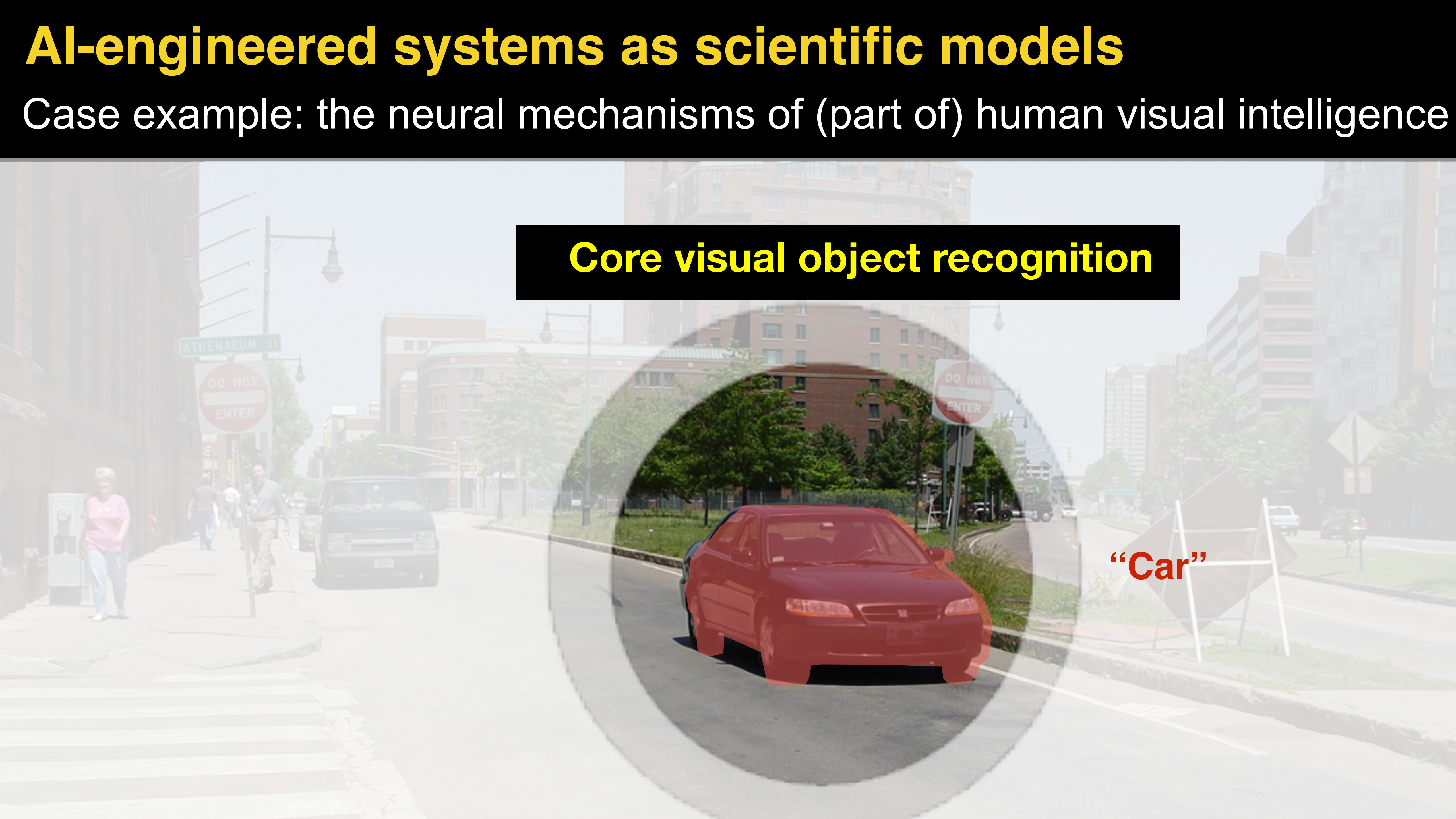


AI-engineered systems as scientific models

Case example: the neural mechanisms of (part of) human visual intelligence

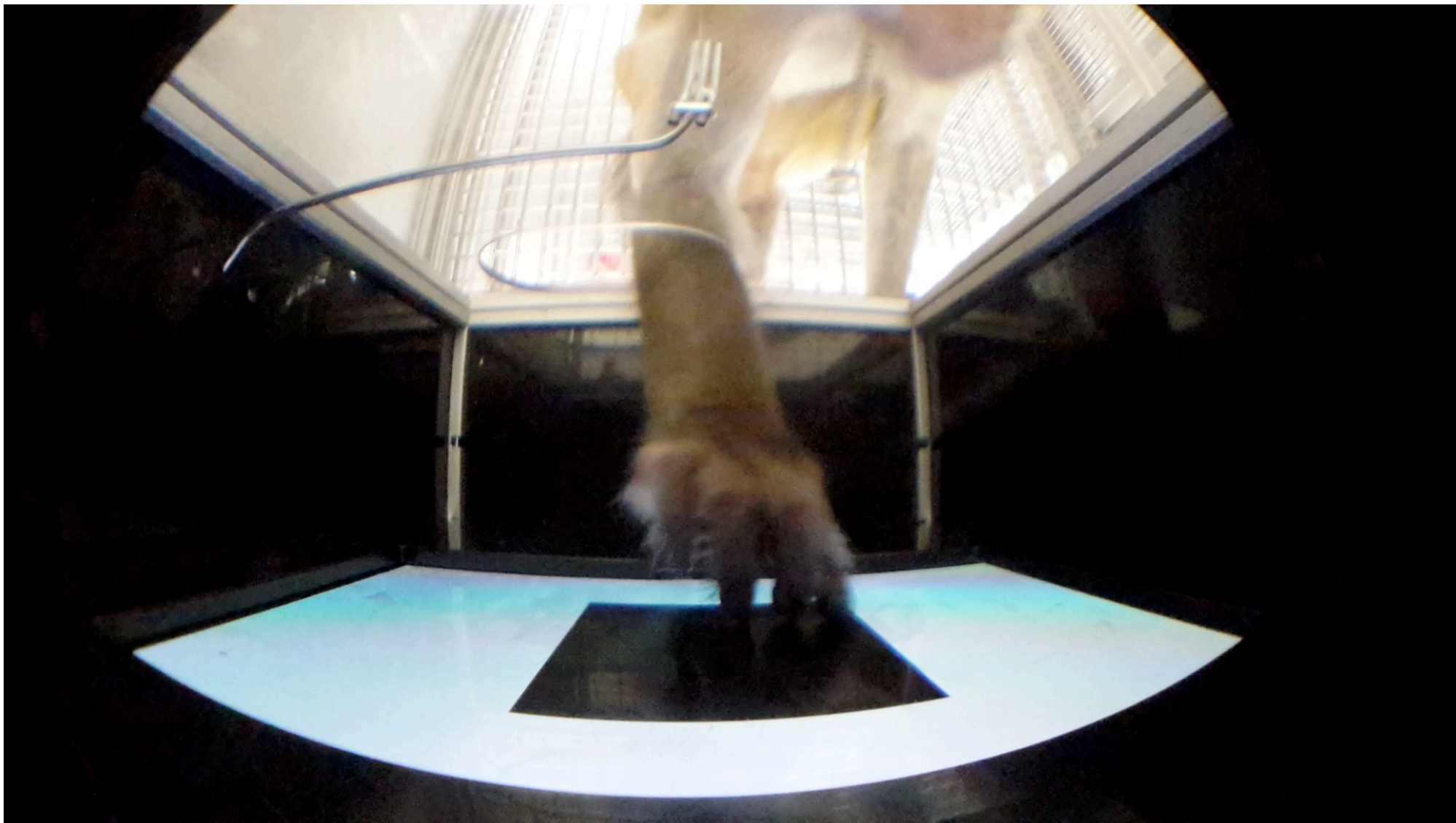
Core visual object recognition

“Car”

A street scene with a red car highlighted by a circular magnifying glass. The background shows a city street with buildings, trees, and pedestrians. A street sign for 'ATHENAEUM ST' and a 'DO NOT ENTER' sign are visible. The red car is a sedan, and the magnifying glass focuses on it, making it the central subject of the image.

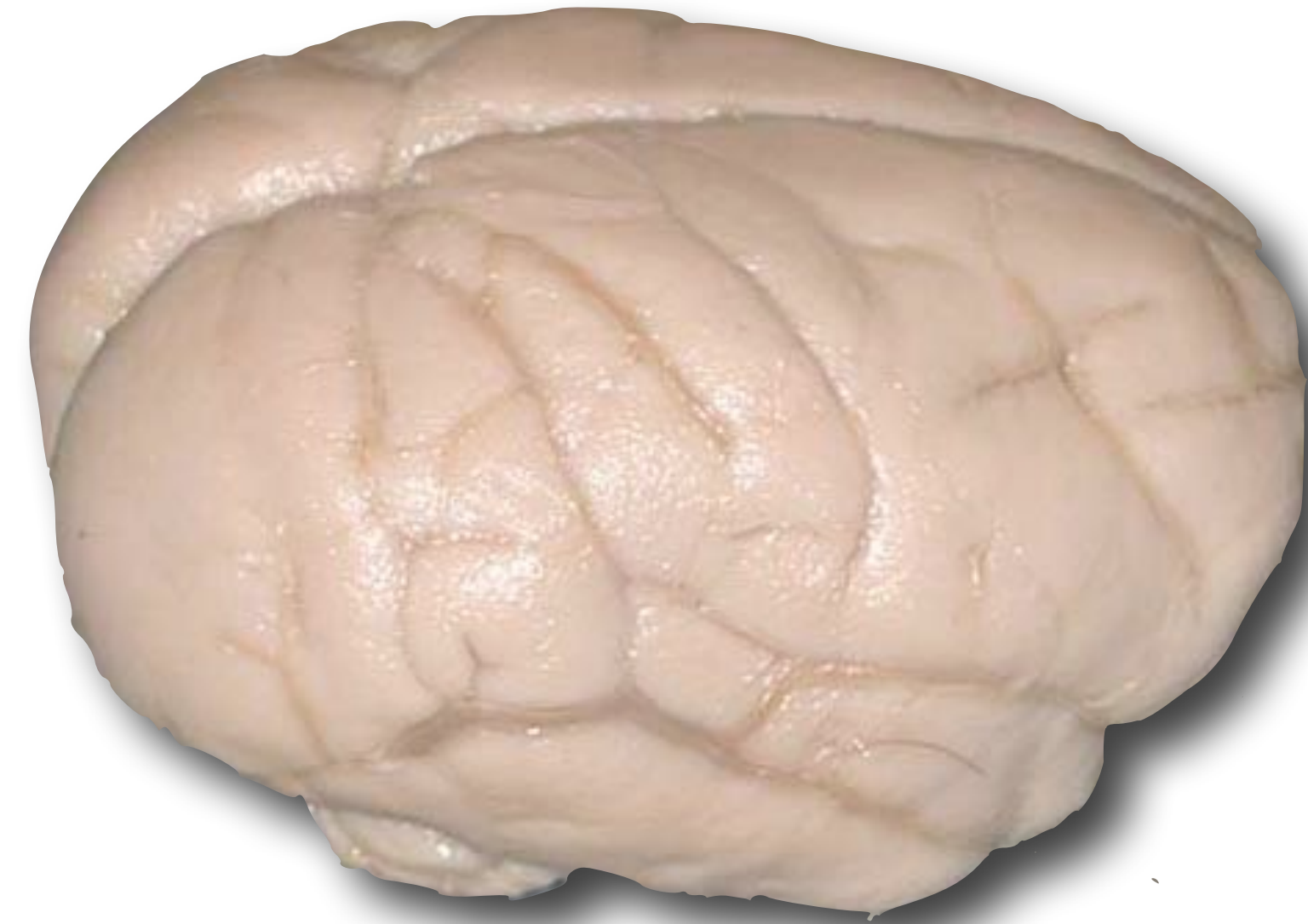
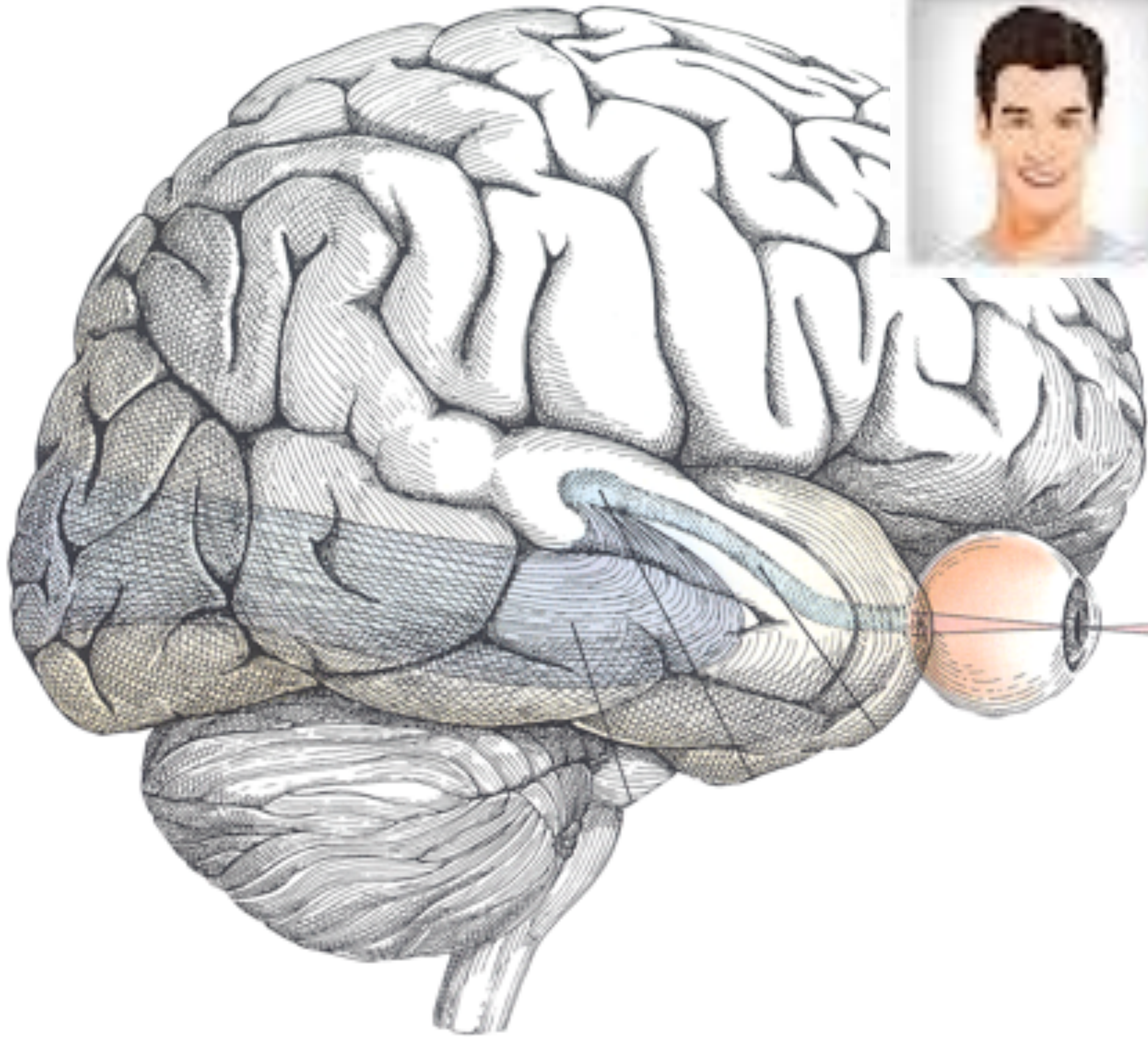
Non-human primates and humans have virtually identical visual capabilities

Primates



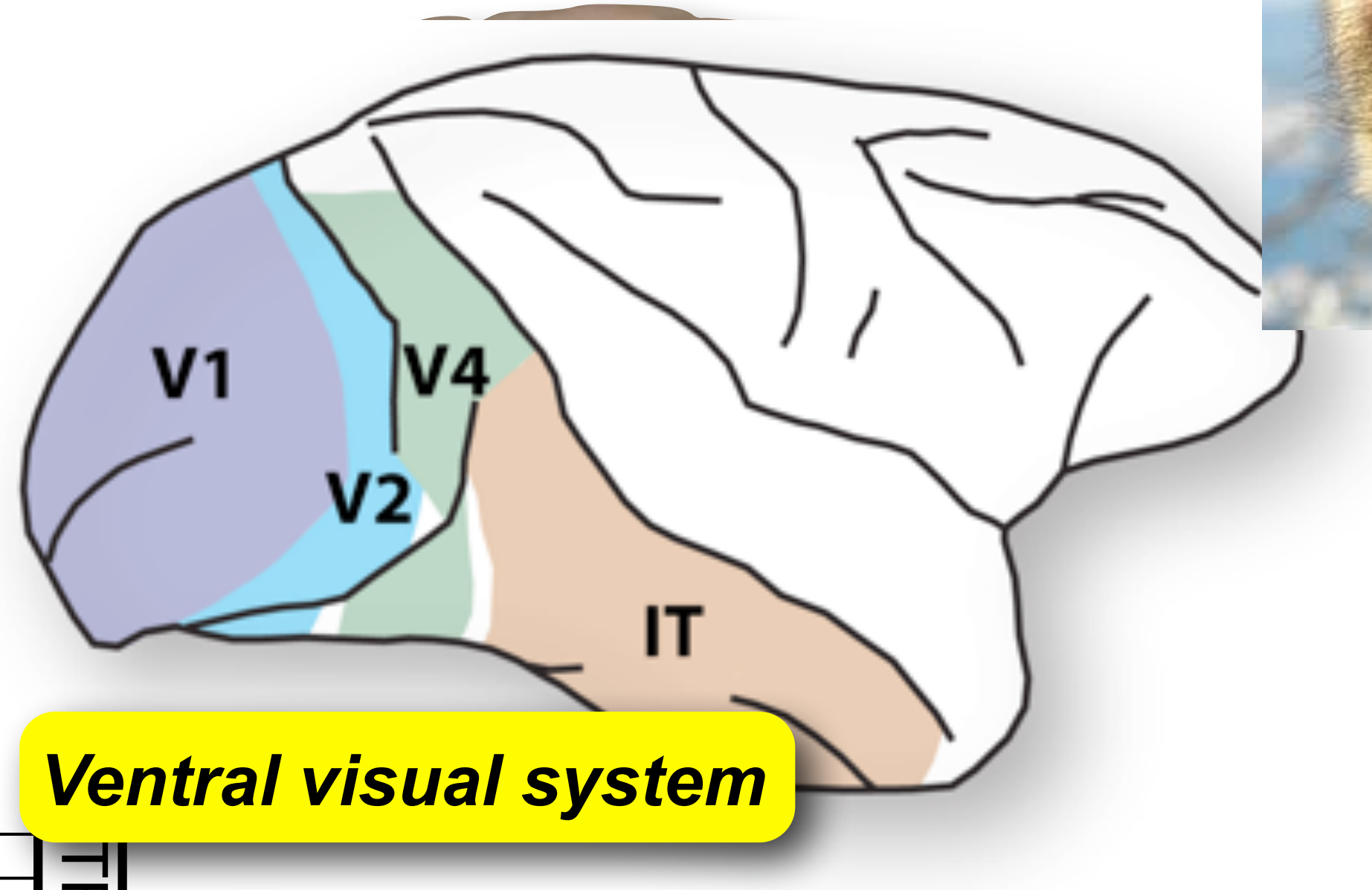
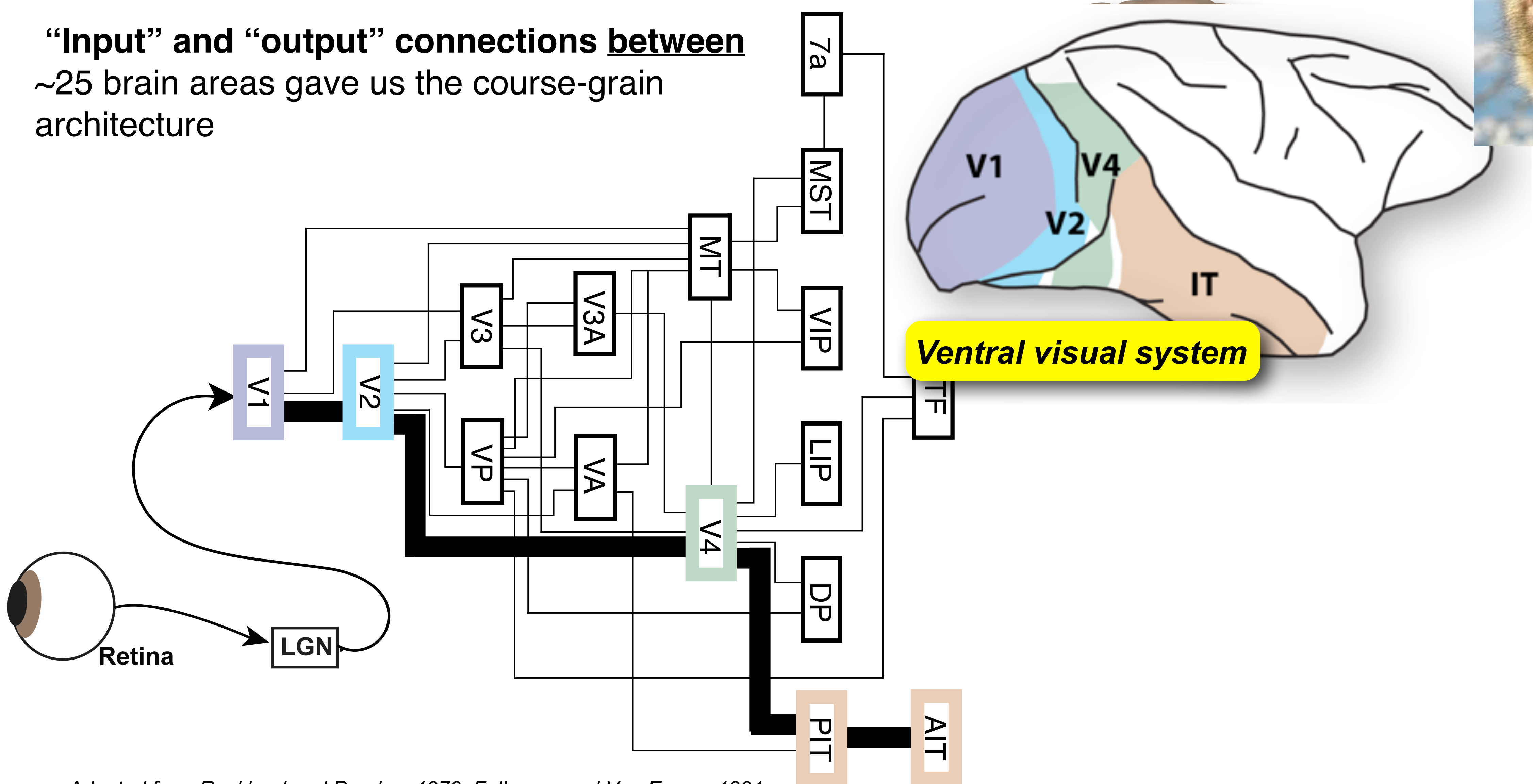
[primate, Homo sapien = primate, rhesus monkey]

Neuroanatomy provided ~ “wiring diagram” of the monkey visual system



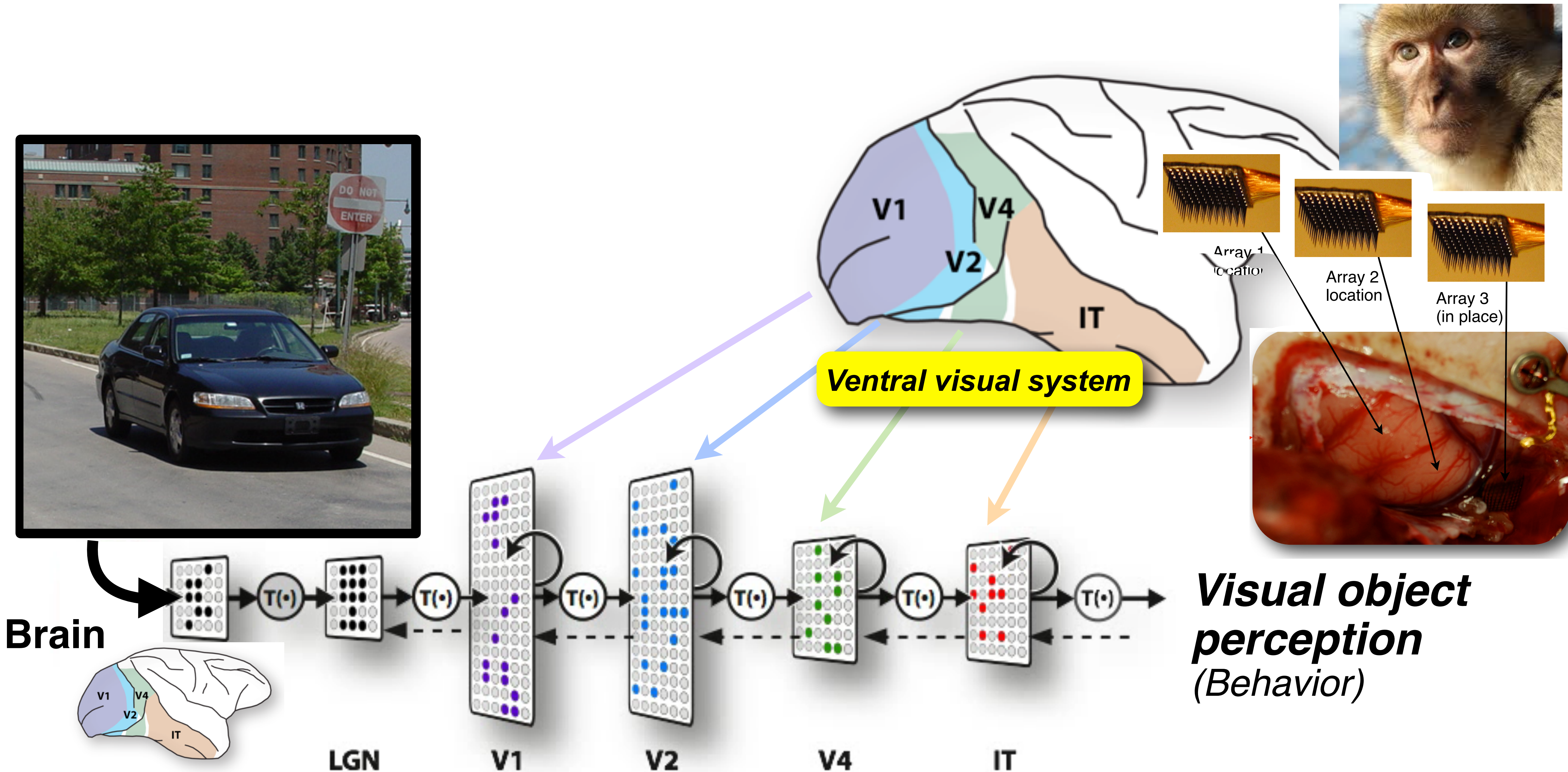
Neuroanatomy provided ~ “wiring diagram” of the monkey visual system

“Input” and “output” connections between
~25 brain areas gave us the course-grain
architecture

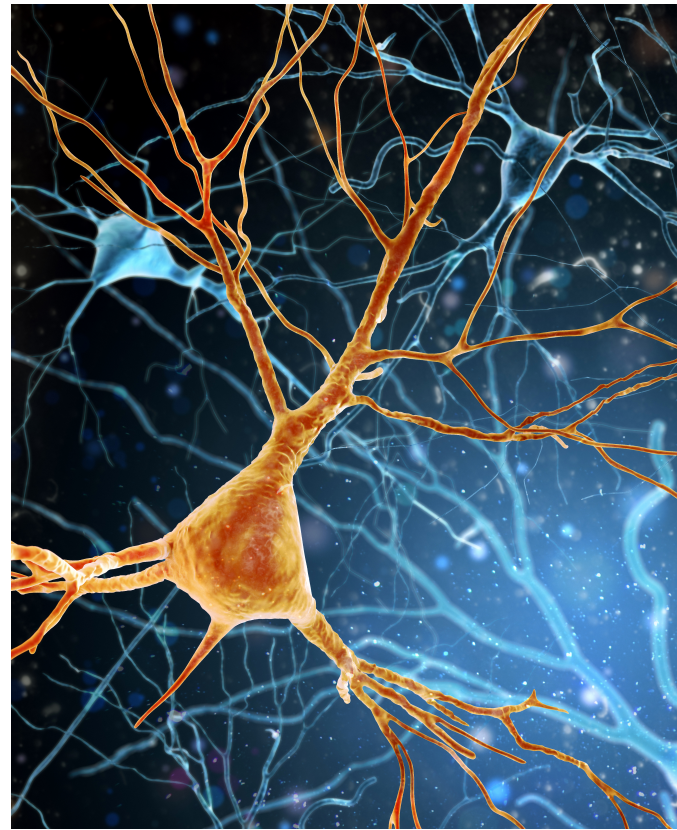


Adapted from Rockland and Pandya, 1979; Felleman and Van Essen, 1991

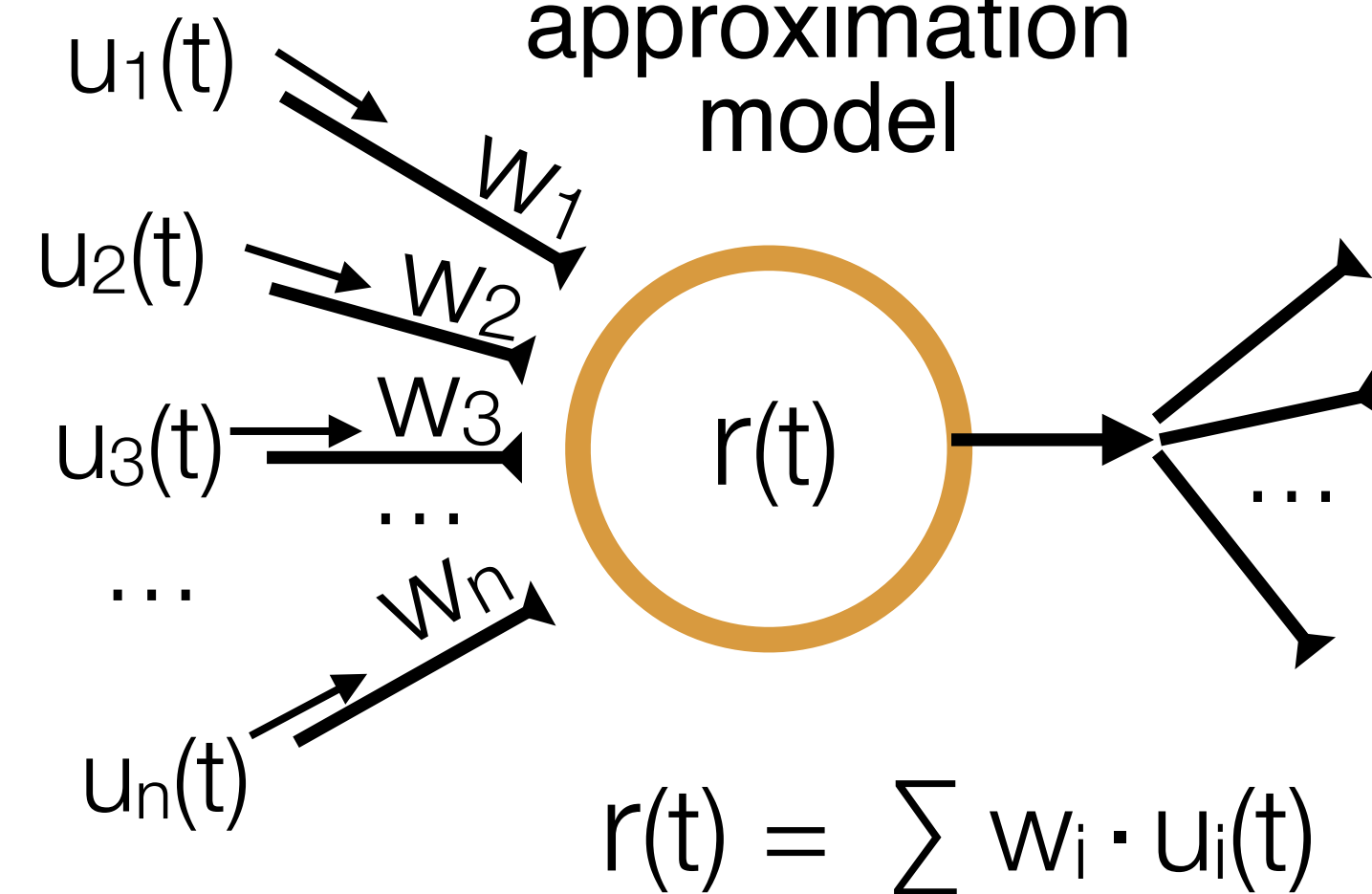
Neurophysiology provided access to the system's information processing



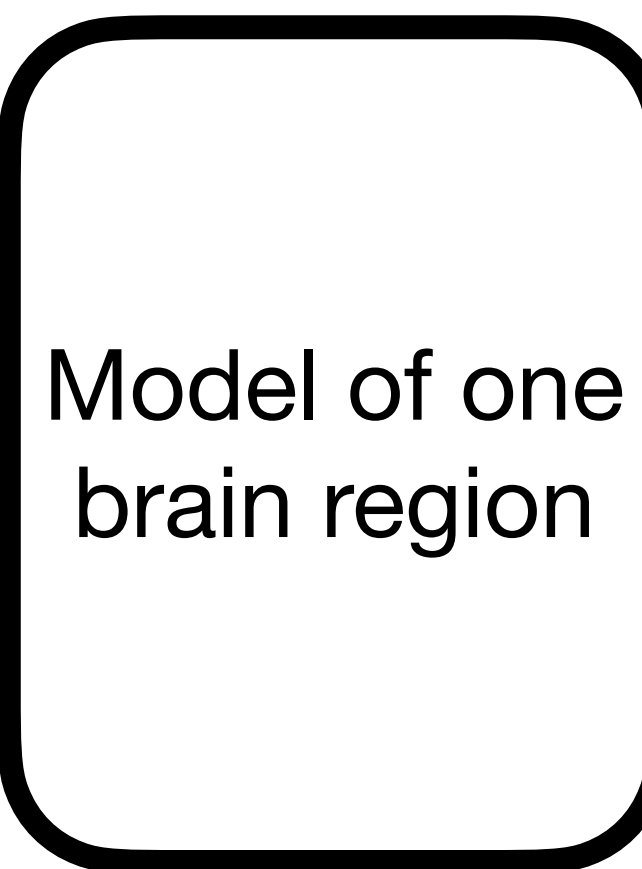
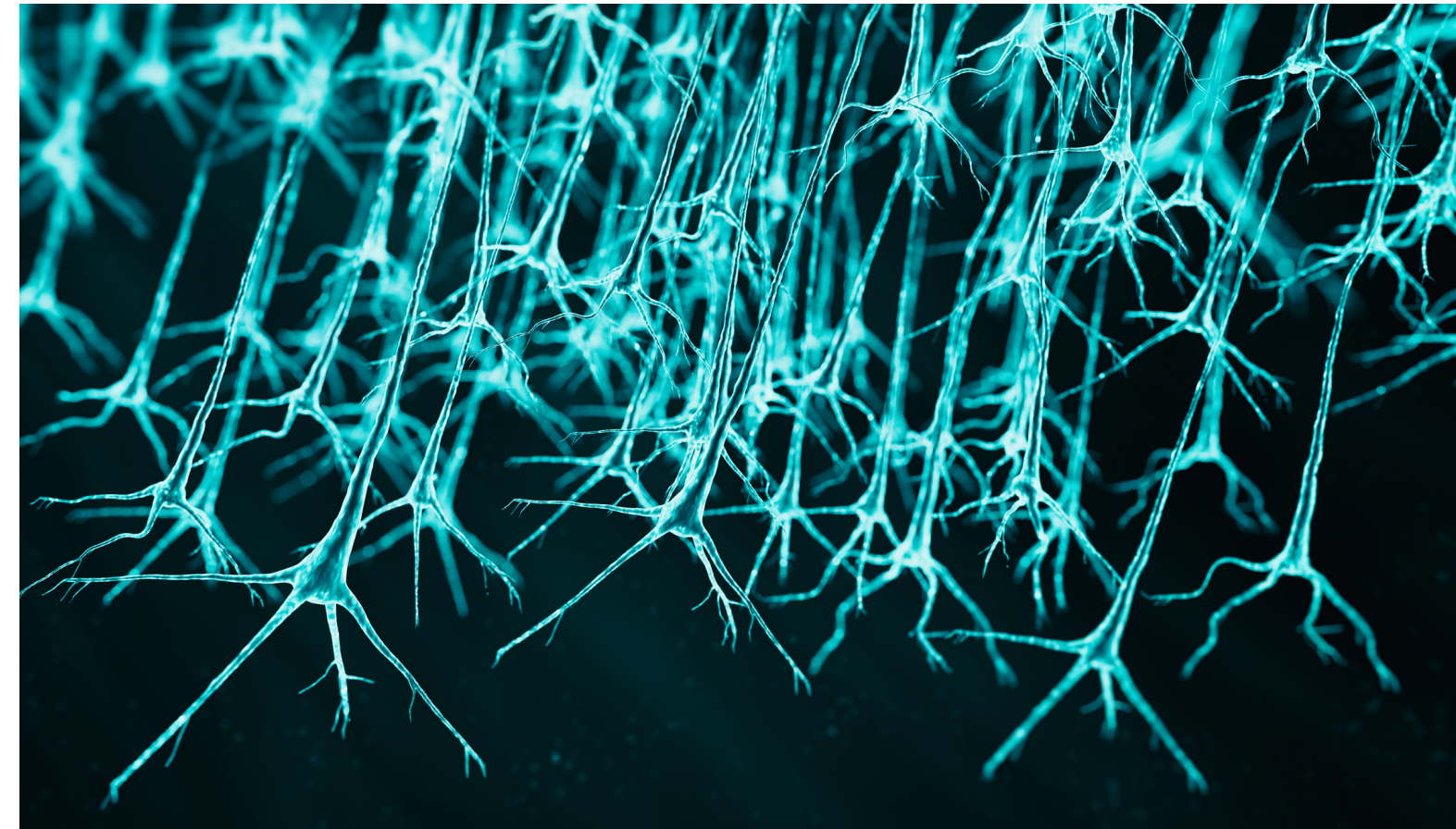
Explaining primate visual processing by building machine-executable, multi-scale models of the integrated set of neural mechanisms



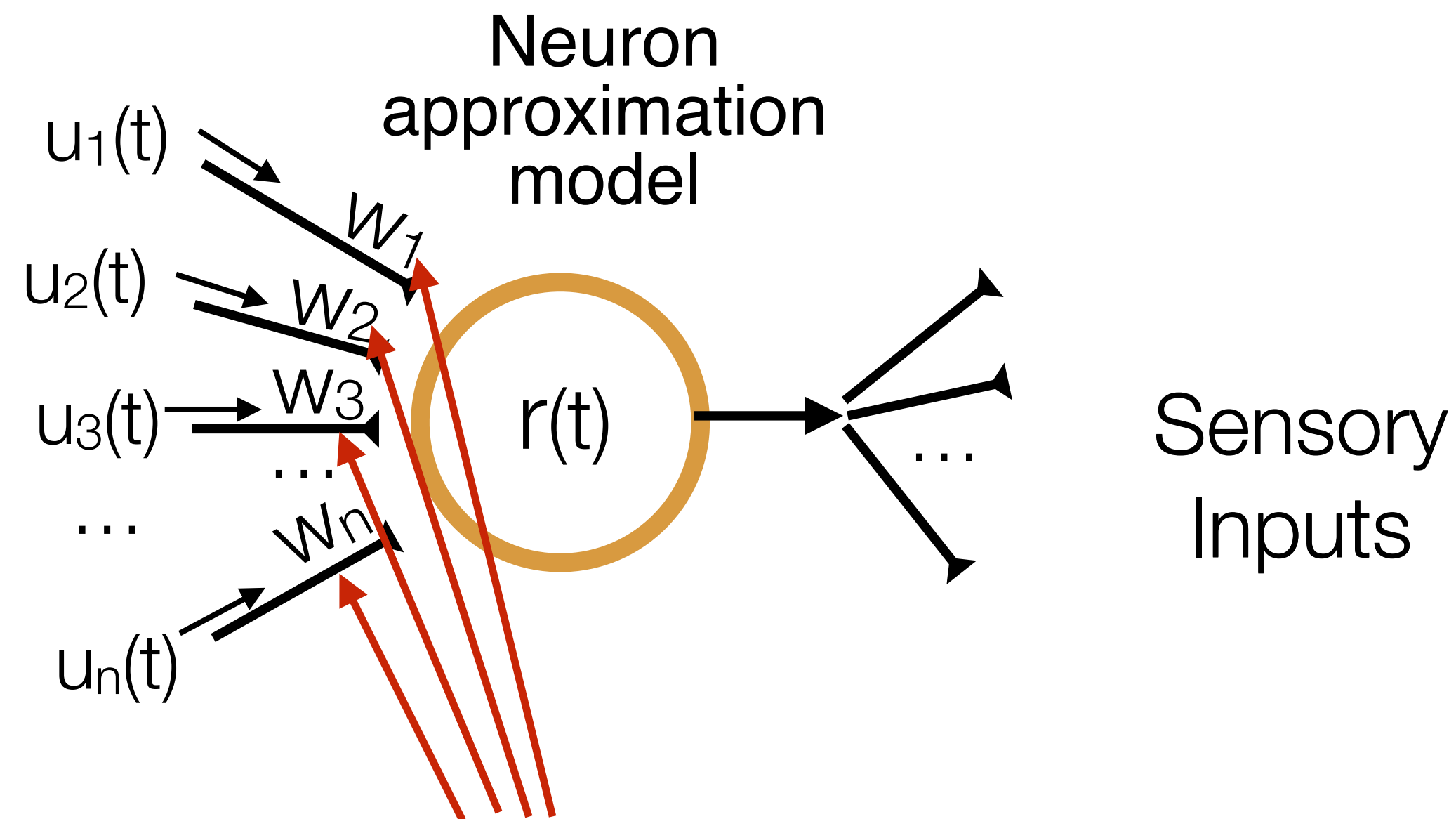
Neuron
approximation
model



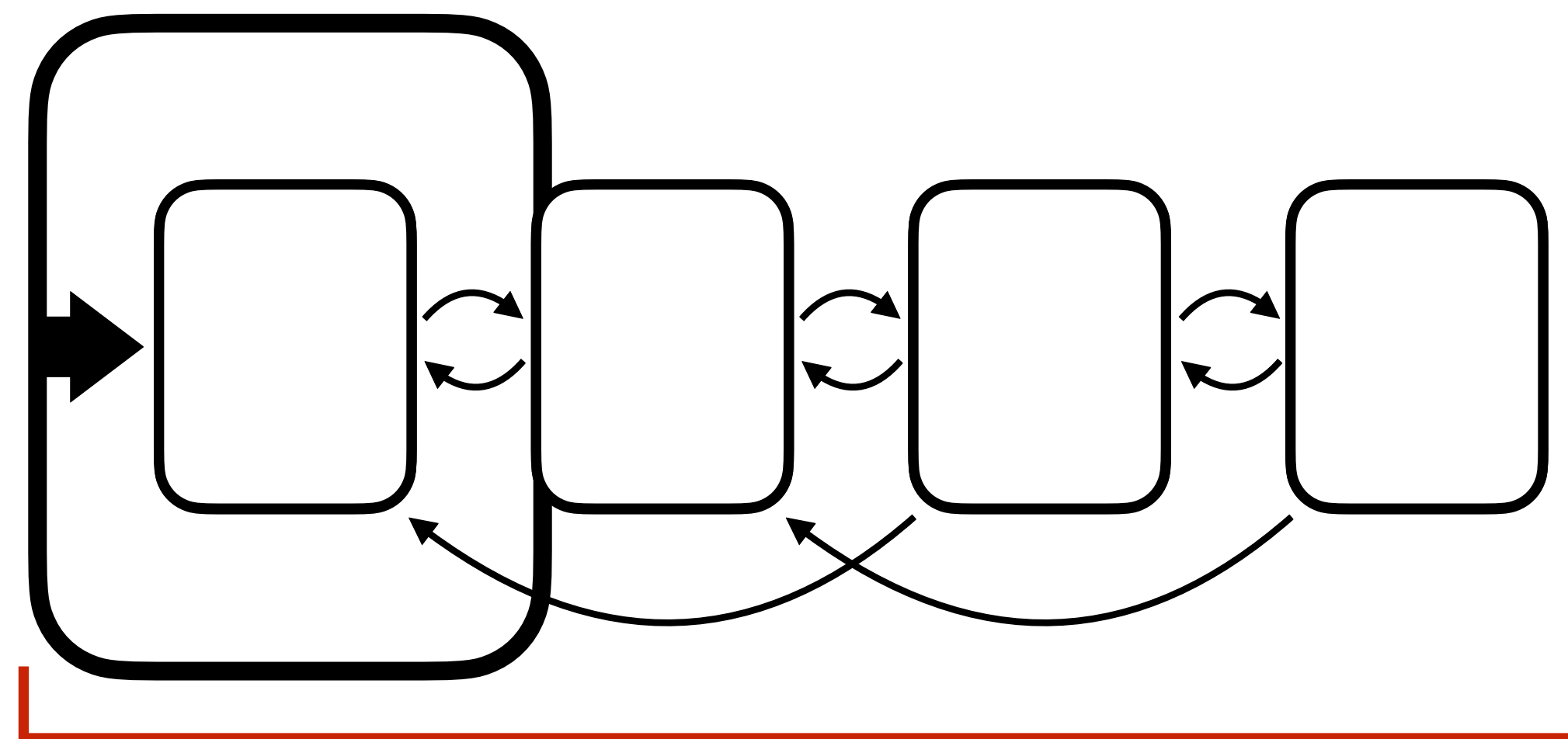
If $r(t) > 0$, neuron response is $r(t)$
else, neuron response is 0



Explaining primate visual processing by building machine-executable, multi-scale models of the integrated set of neural mechanisms



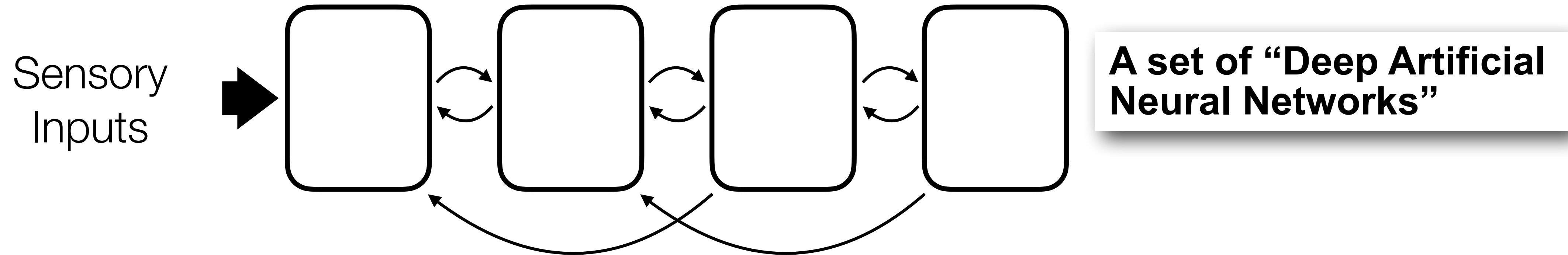
Parameters to specify connection strengths inside that architecture



Course-grain architecture was known

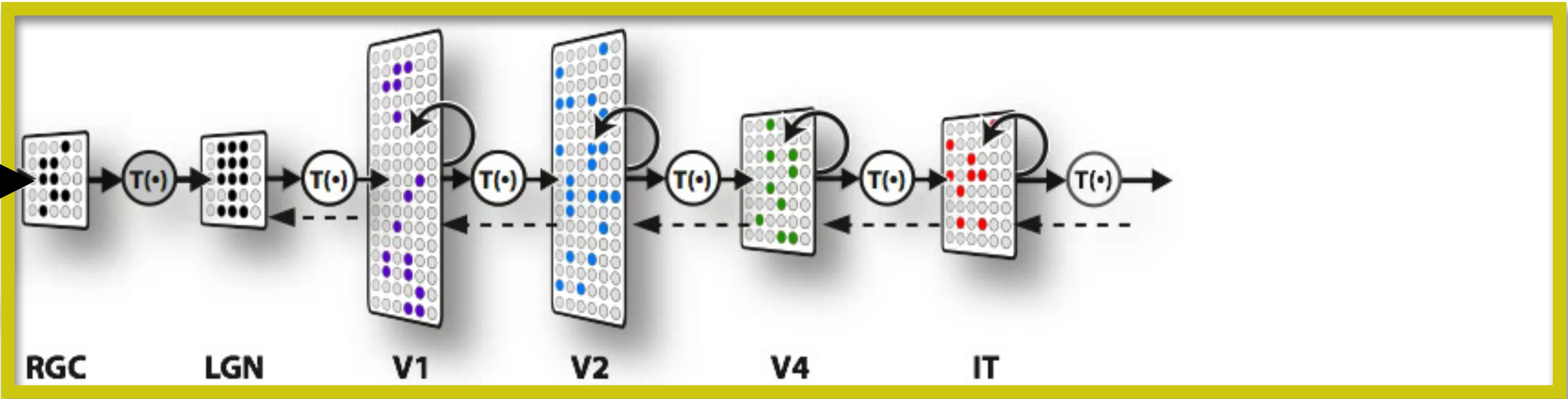
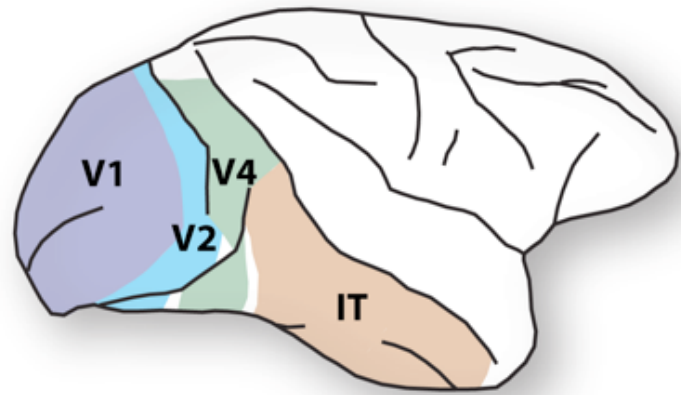
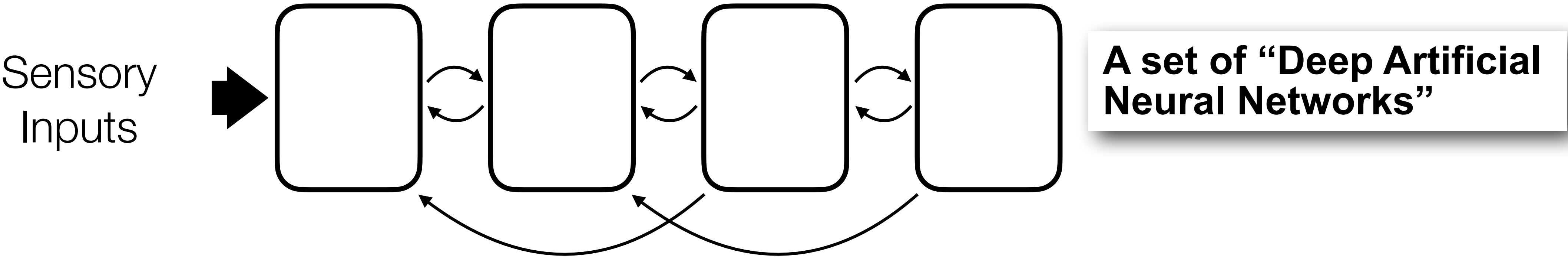
Conceptual Model

A set of possible descriptions of how the (ventral) visual system works



Conceptual Model

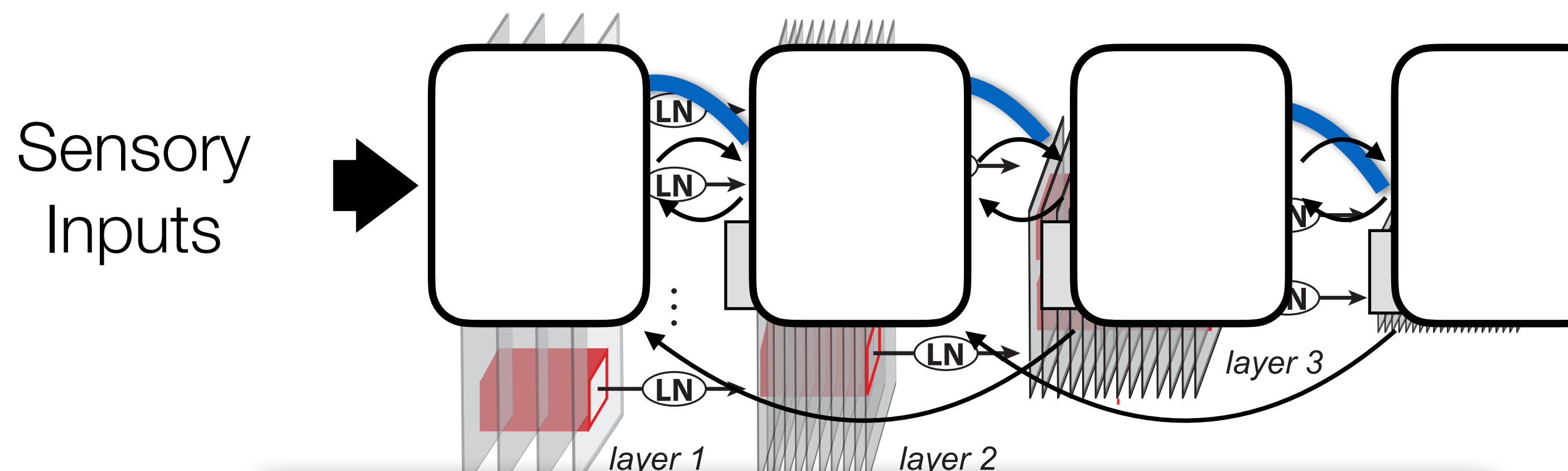
A set of possible descriptions of how the (ventral) visual system works



Brain

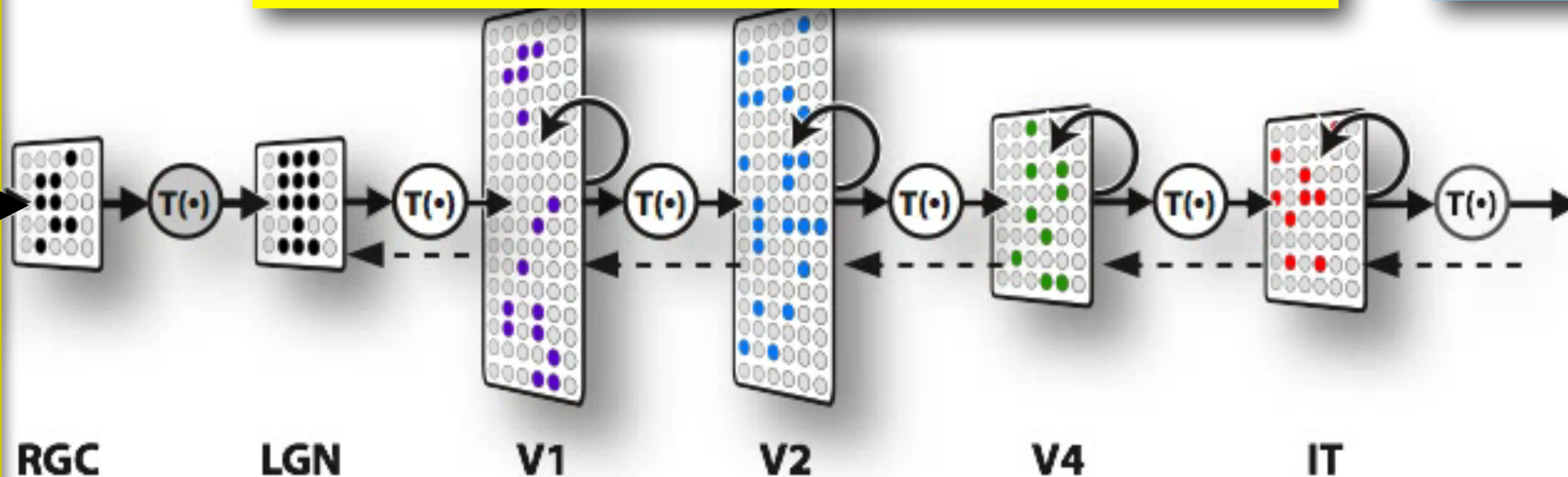
Implemented Model

A matrix of possible descriptions of models between (traditional) visual systems



The breakthrough came in ~2013, when we used AI engineering tools to set the model parameters so that our computer model of the visual system succeeded in object recognition tasks

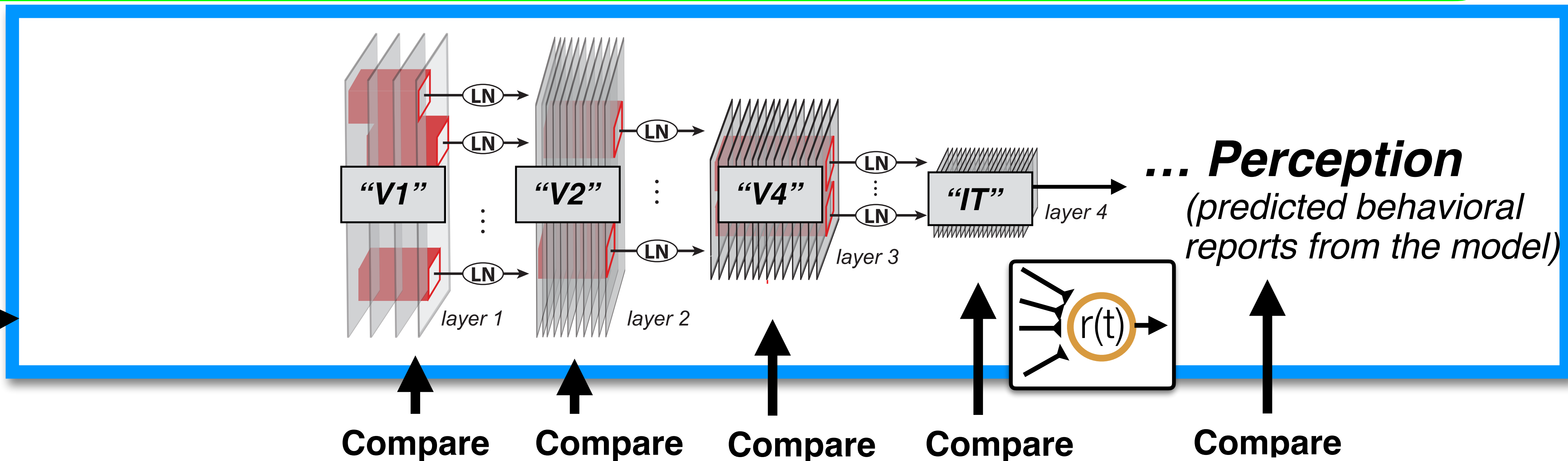
Before ~2013, our field was stuck: the models of the visual system were failing to explain the measured internal information processing



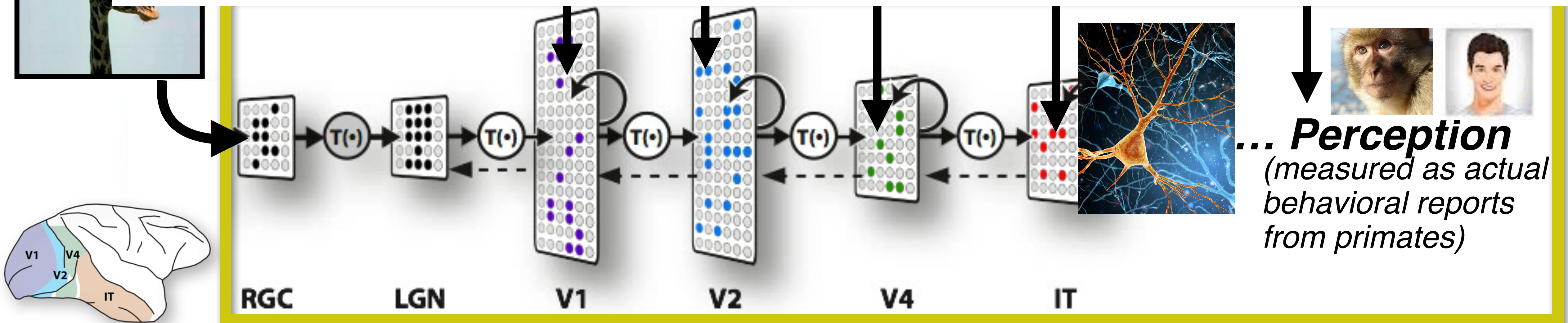
Brain

Implemented Model

A machine executable, multi-scale model of the (ventral) visual system



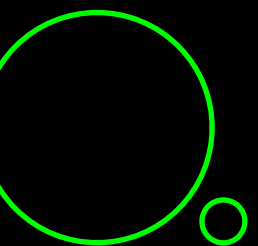
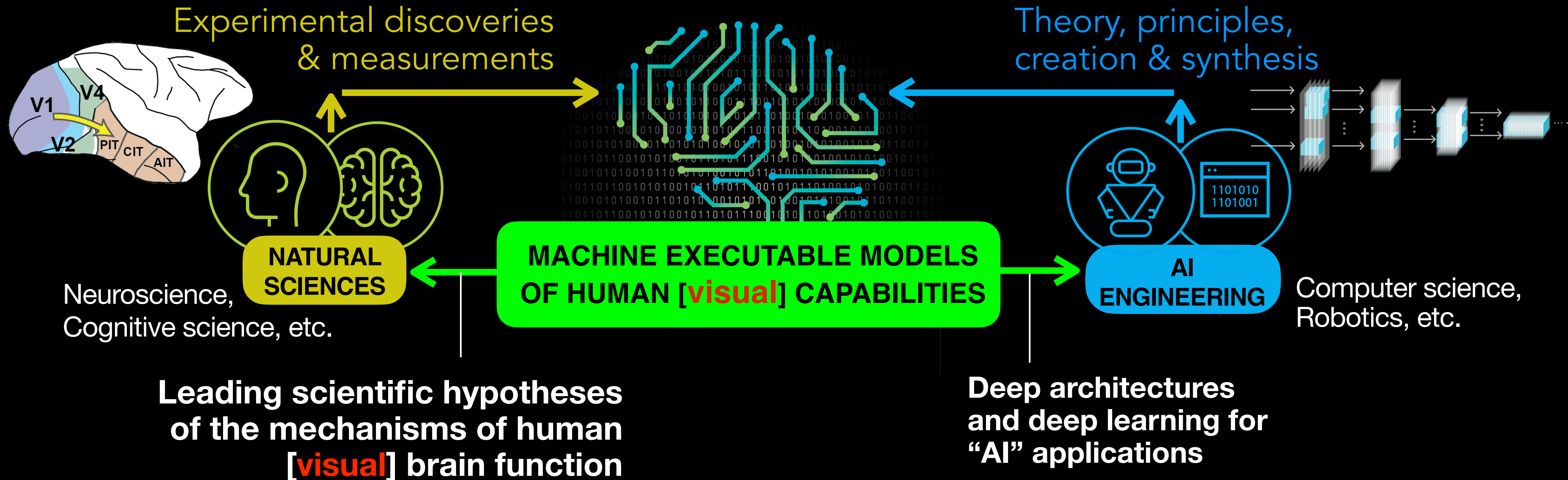
Over the past decade, we and others established precise functional alignments!



Brain

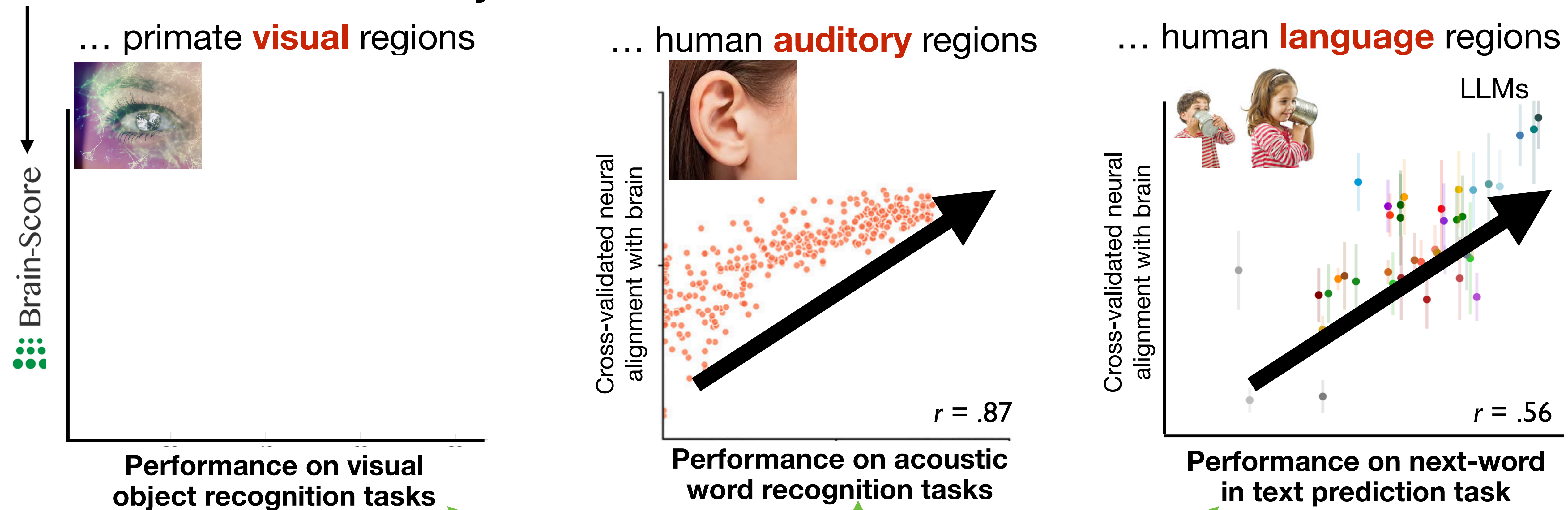
Yamins et al. **NeurIPS** (2013); Yamins, Hong, Solomon, Seibert and DiCarlo **PNAS** (2014)

The AI-forward strategy behind this vision case example is being successfully generalized to other brain and cognitive science domains



In each domain, more capable AI systems turn out to be better multi-scale, neuroscientific models of brain function

Goodness of different AI systems as neuroscientific models of ...



Adapted from Yamins, Hong, Solomon, Seibert and DiCarlo *PNAS* (2019) and Kubilius, Schrimpf, Nayebi, and DiCarlo. *bioRxiv* (2019)

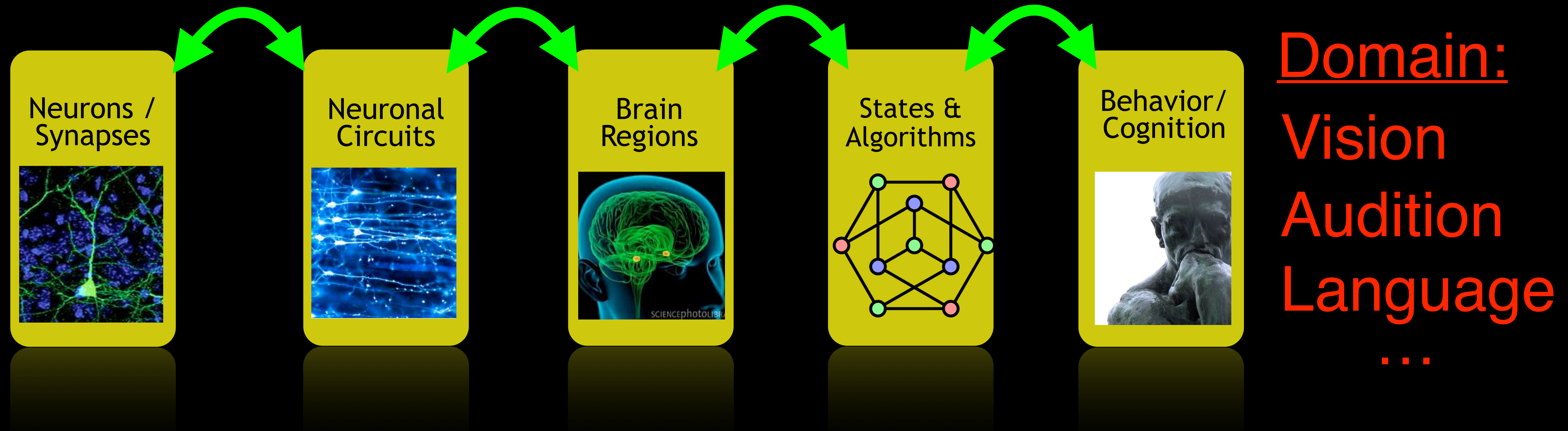
Adapted from Kell, Yamins, Schuck, Norman-

Adapted from Schrimpf, Blank, Tuckute & (2021)

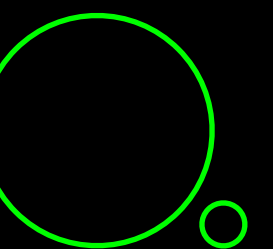
Each domain features a shared cognitive science and AI goal

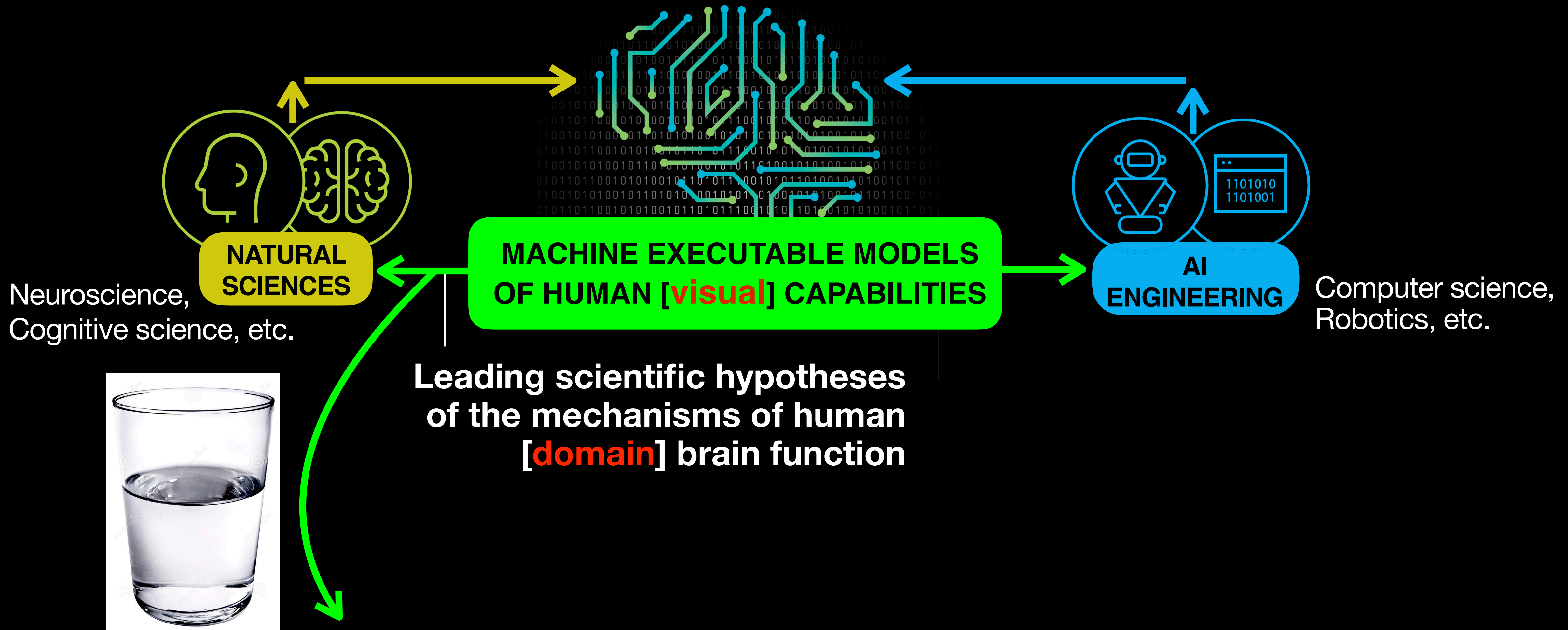
In each domain, new networks are optimized for this goal by AI-engineering

AI-engineered systems as neuroscientific models

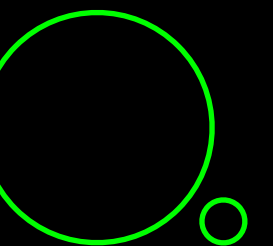


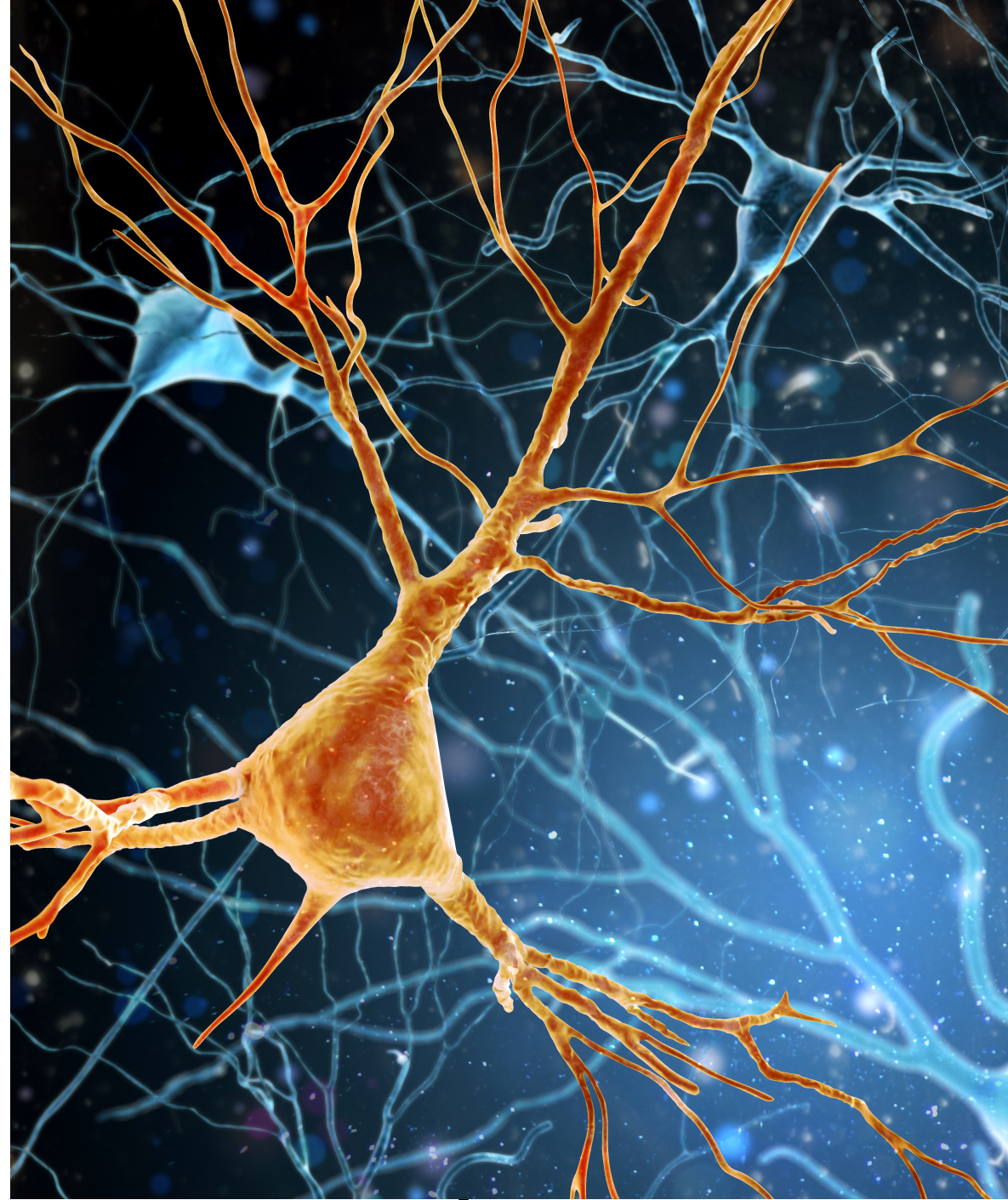
Takeaway #2: Accurate, multi-scale, machine executable models of brain function are emerging from this AI-forward scientific strategy. This will accelerate.



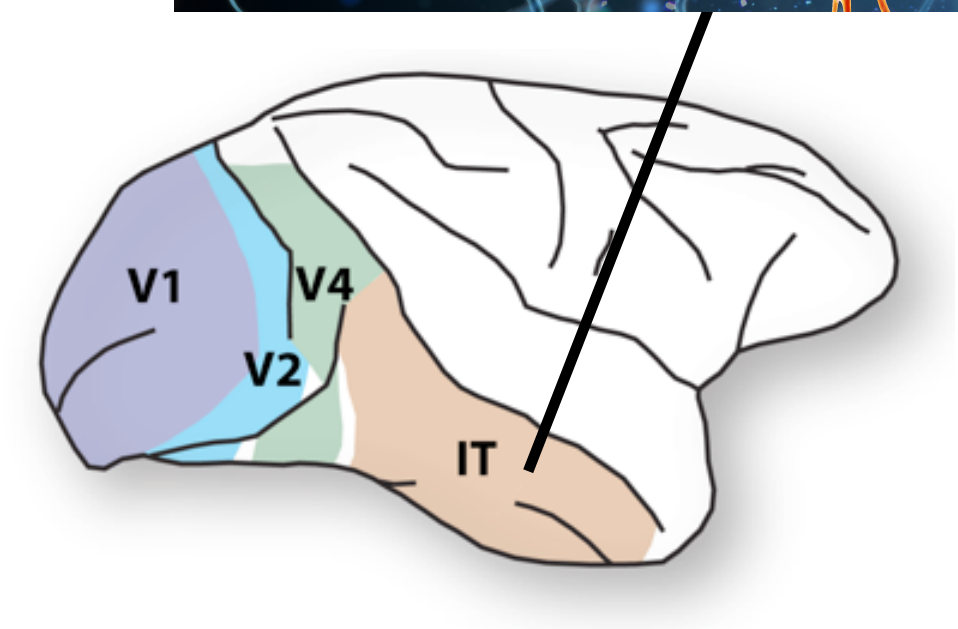


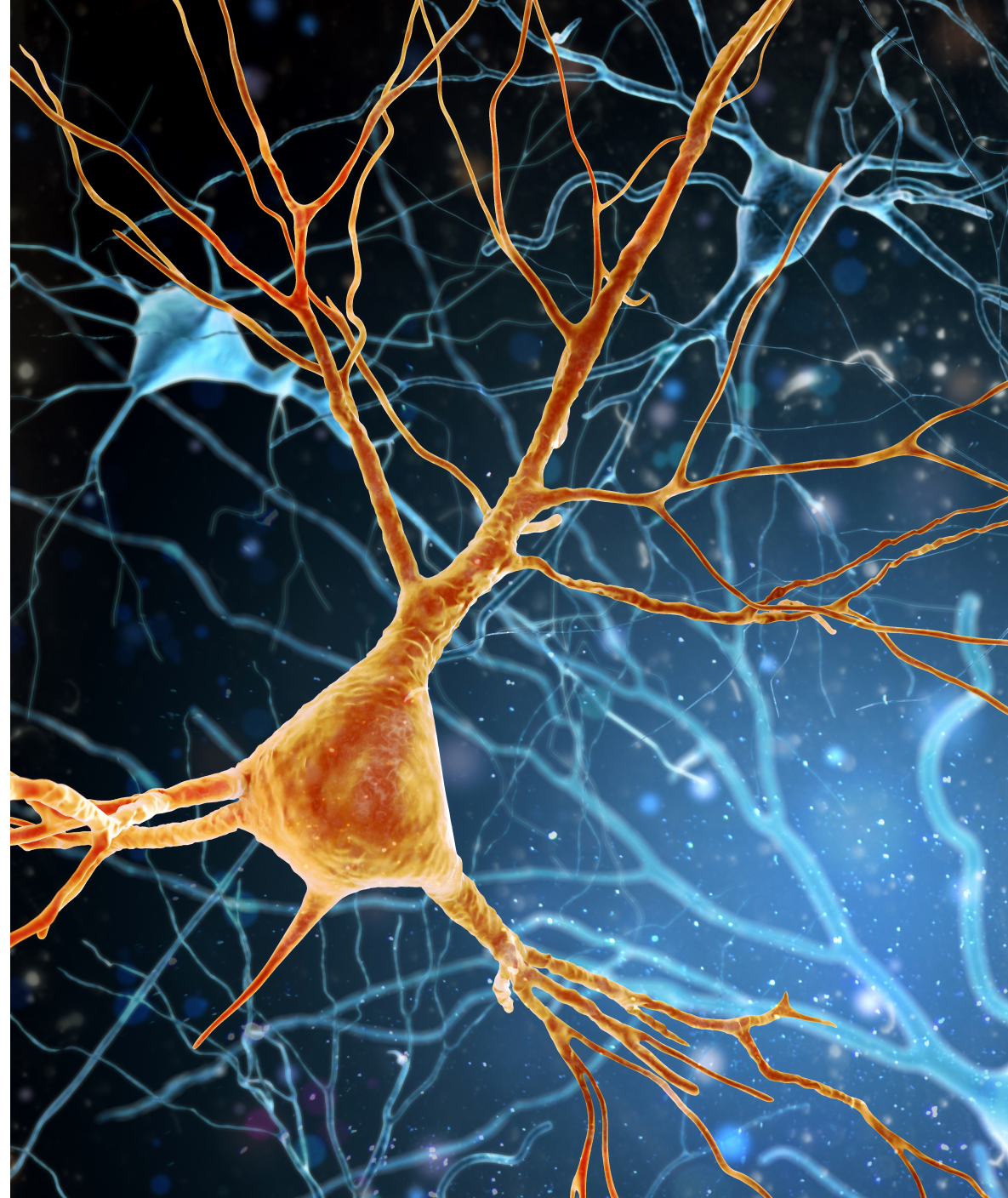
Takeaway #3: The emerging new scientific models will unlock entirely new opportunities to improve human health and education.



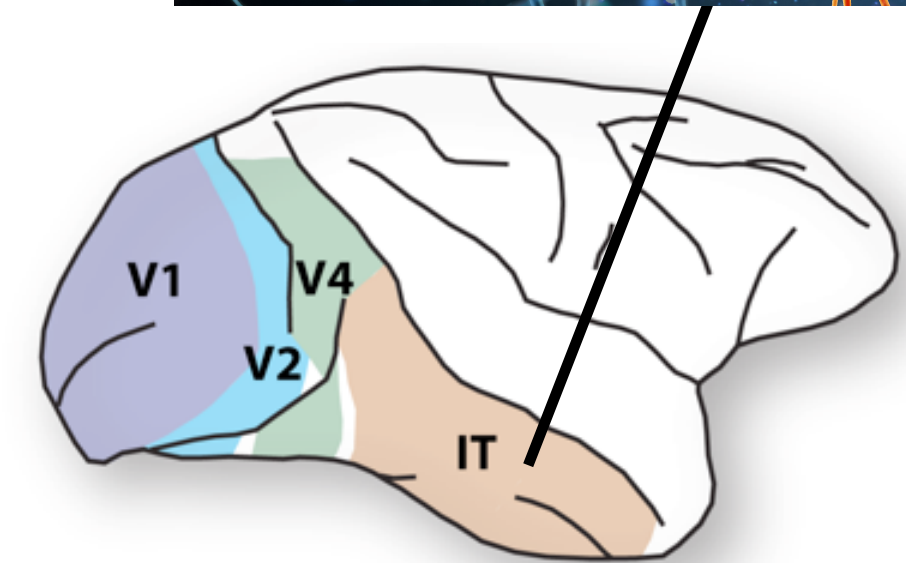


These new (AI-derived) neuroscientific models have already unlocked new, non-invasive brain modulation strategies

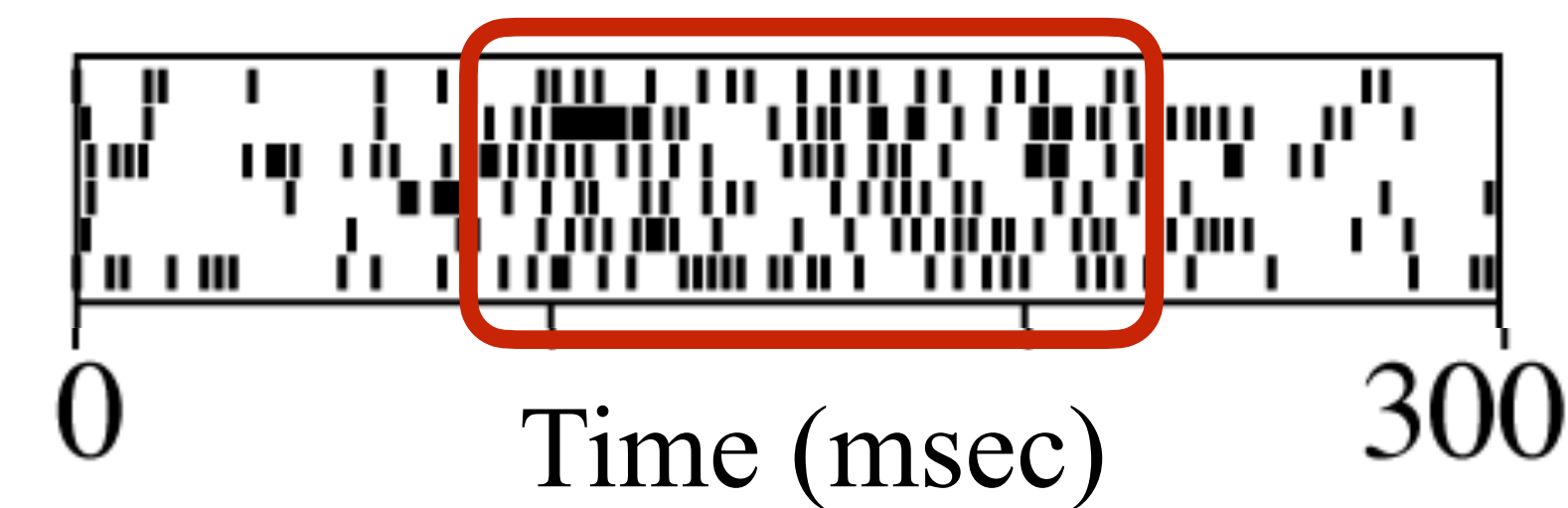
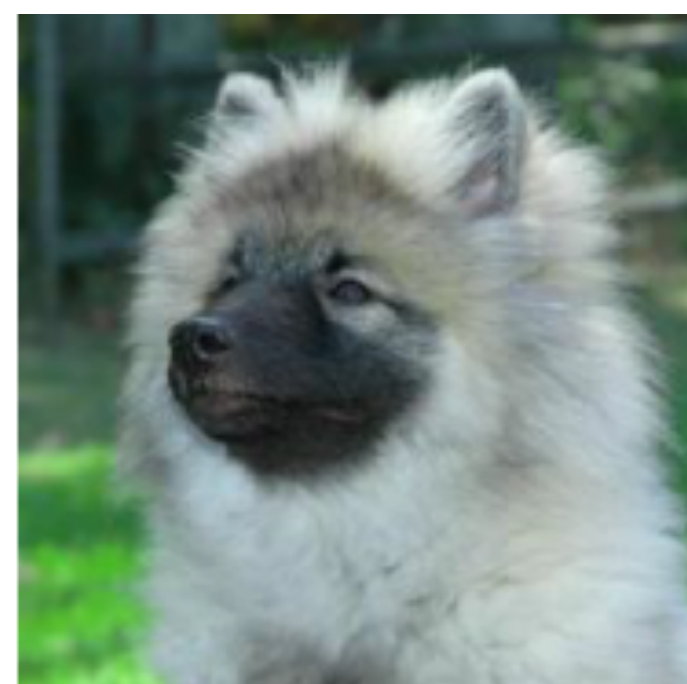
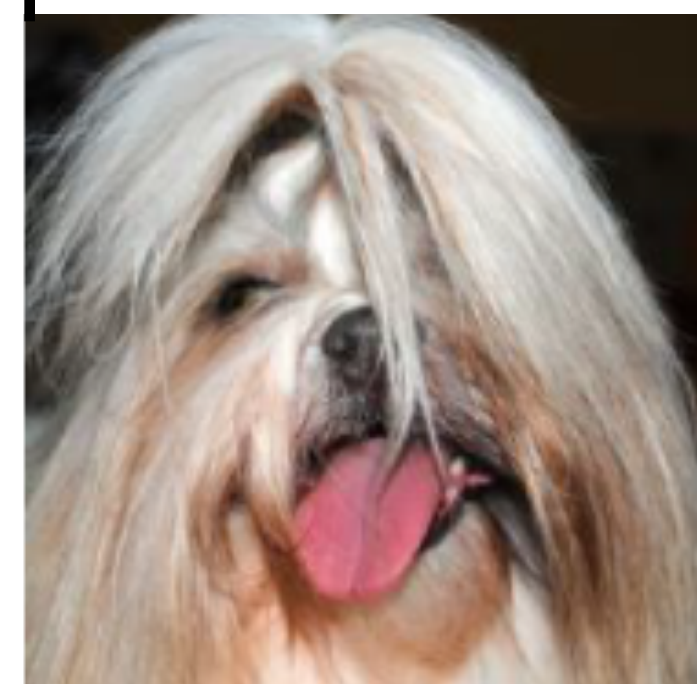
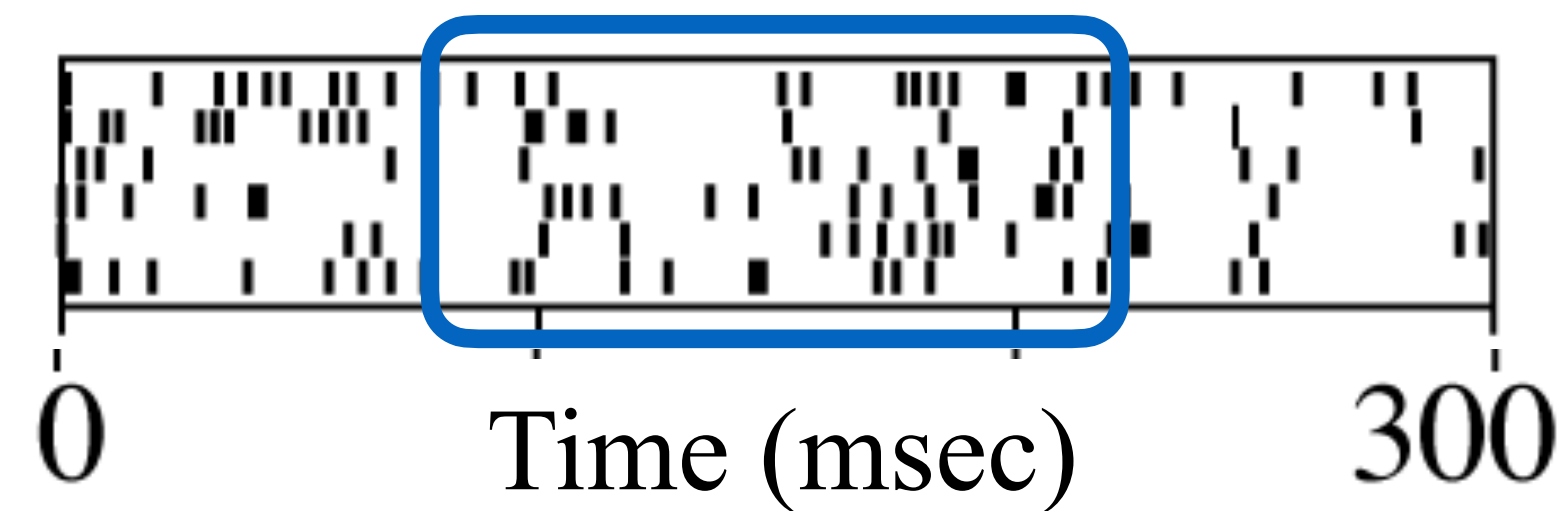




These new (AI-derived) neuroscientific models have already unlocked new, non-invasive brain modulation strategies

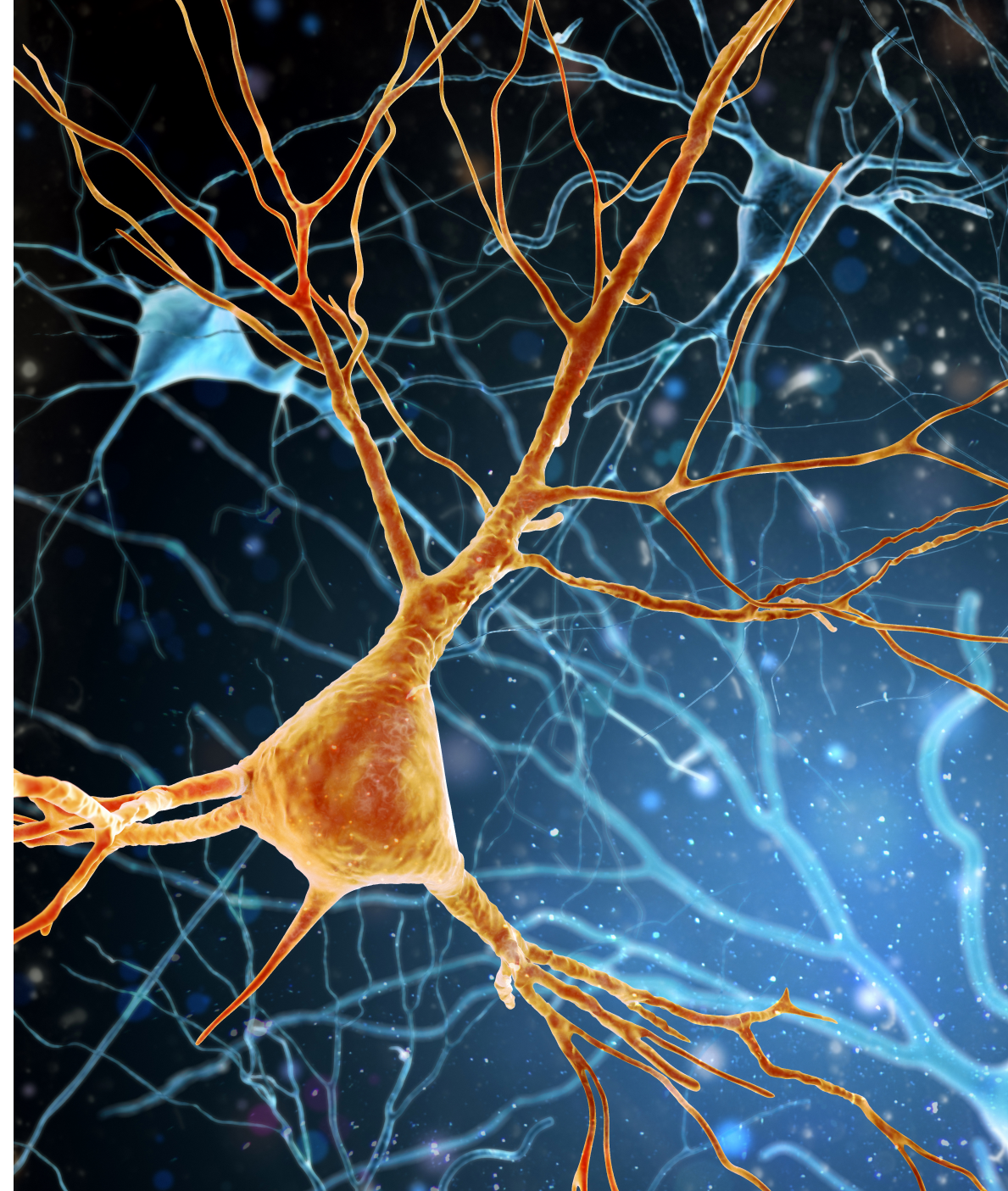


Non-preferred perceptual category
 (“dog”)

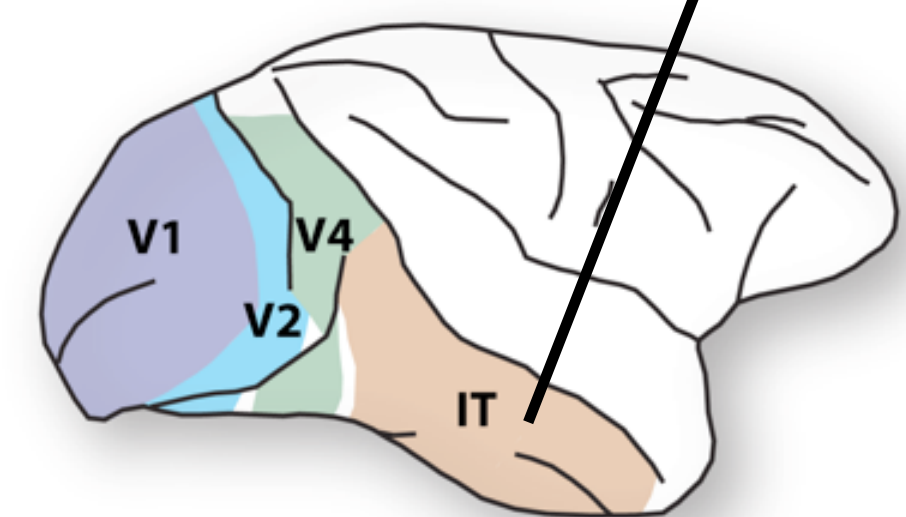
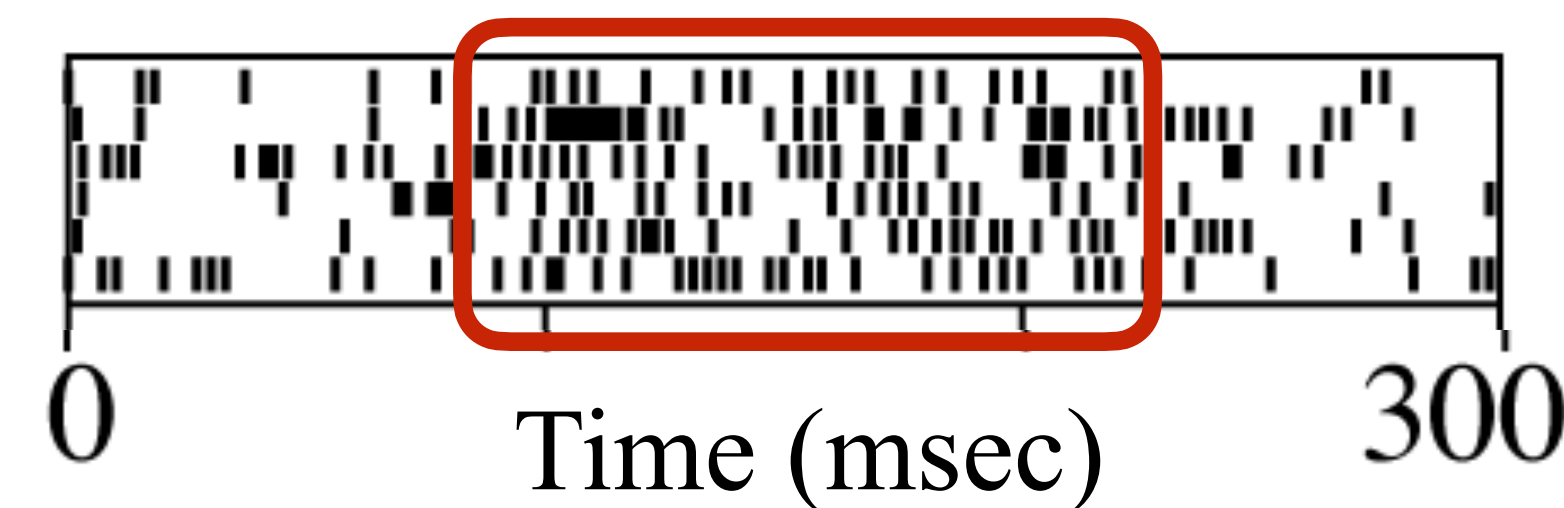
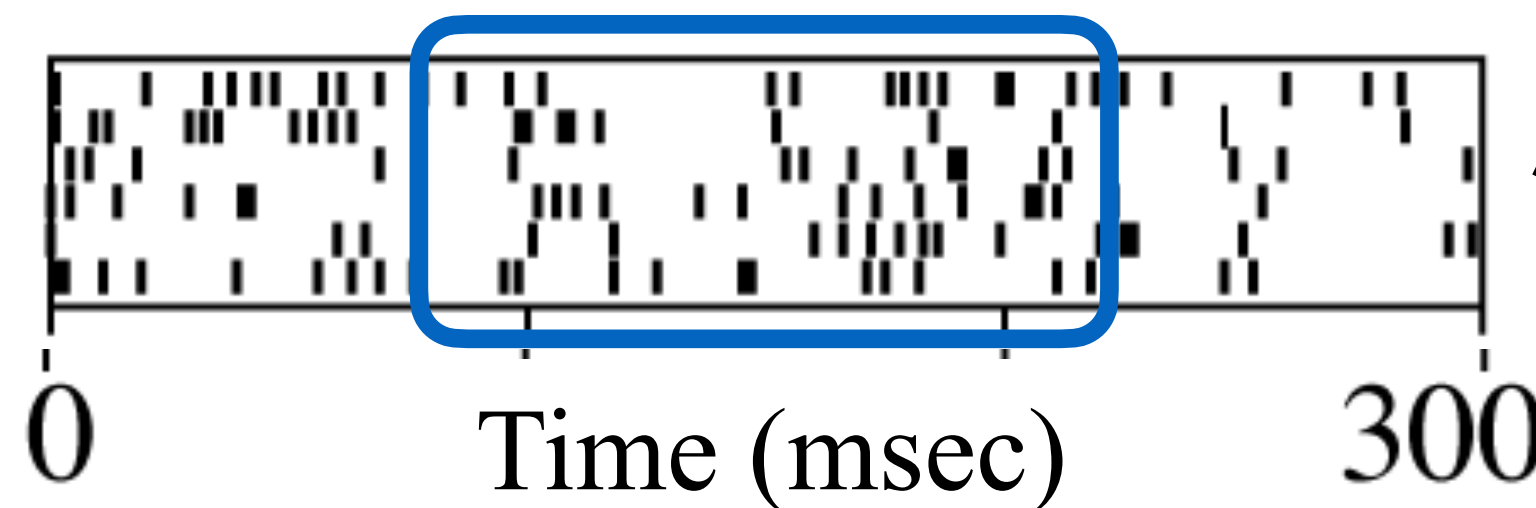


Preferred perceptual category
 (“gauge”)





Specific, tiny change to the visual input!

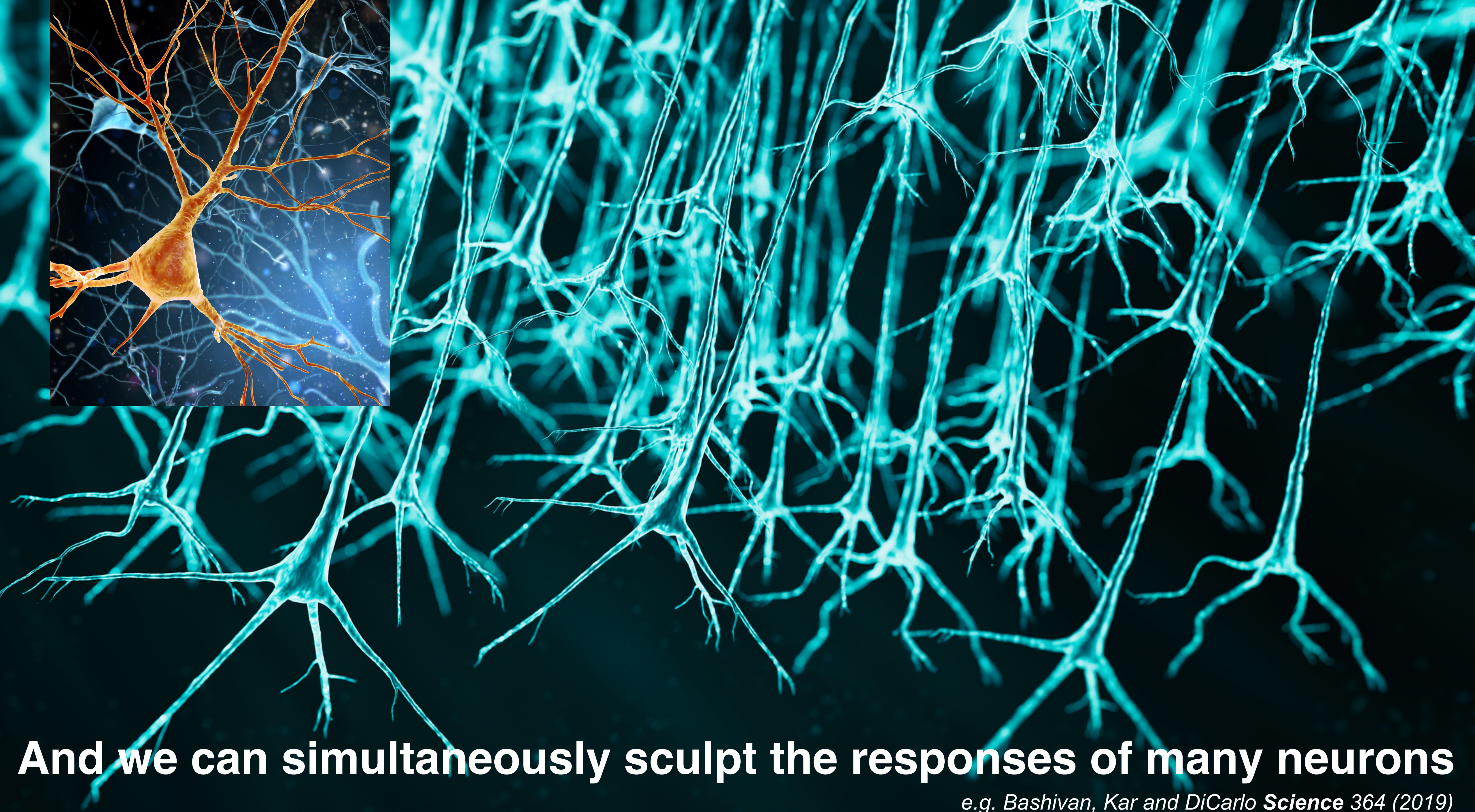


**Non-preferred
perceptual
category
("dog")**



**Preferred
perceptual
category
("gauge")**



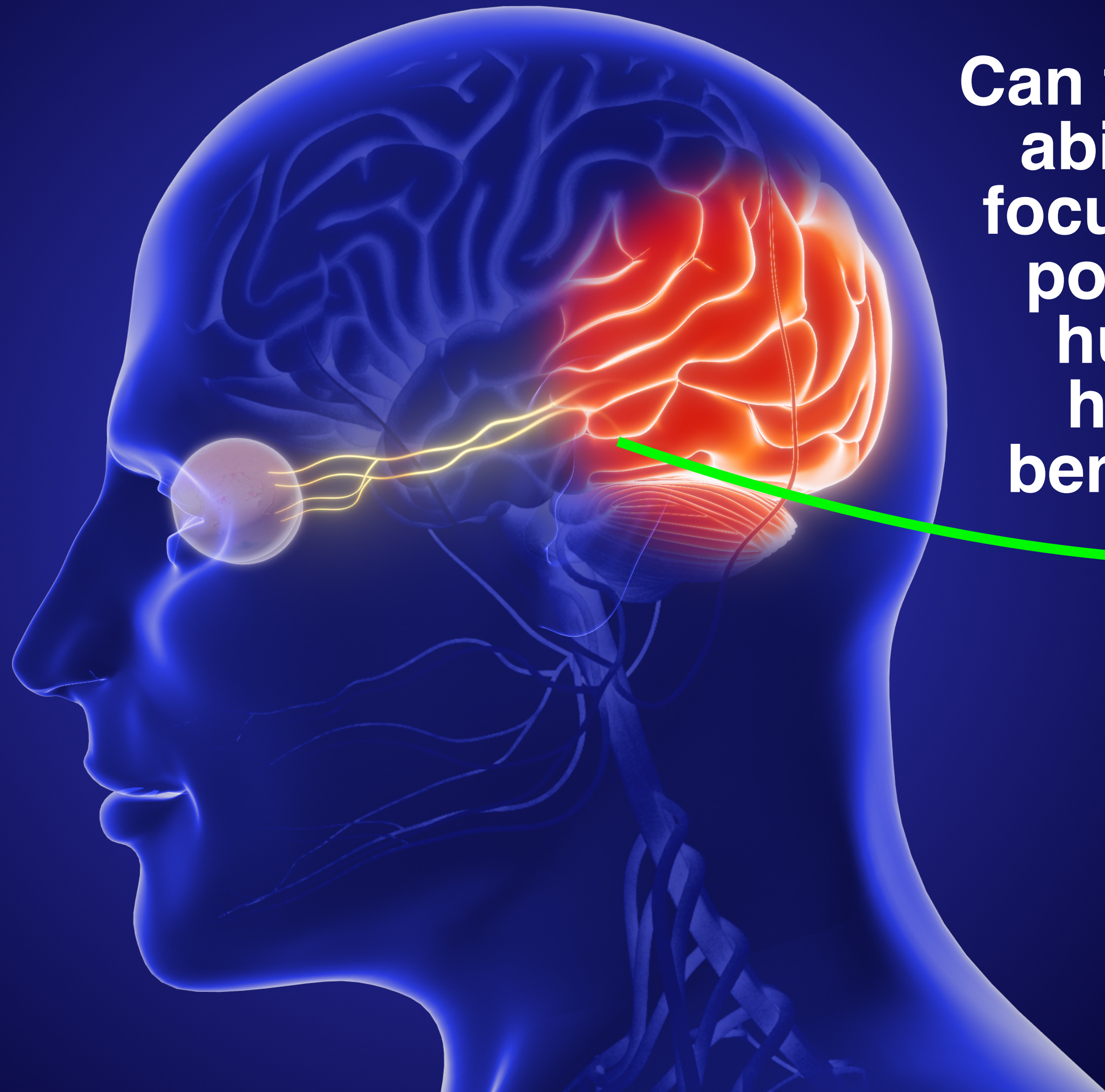


And we can simultaneously sculpt the responses of many neurons

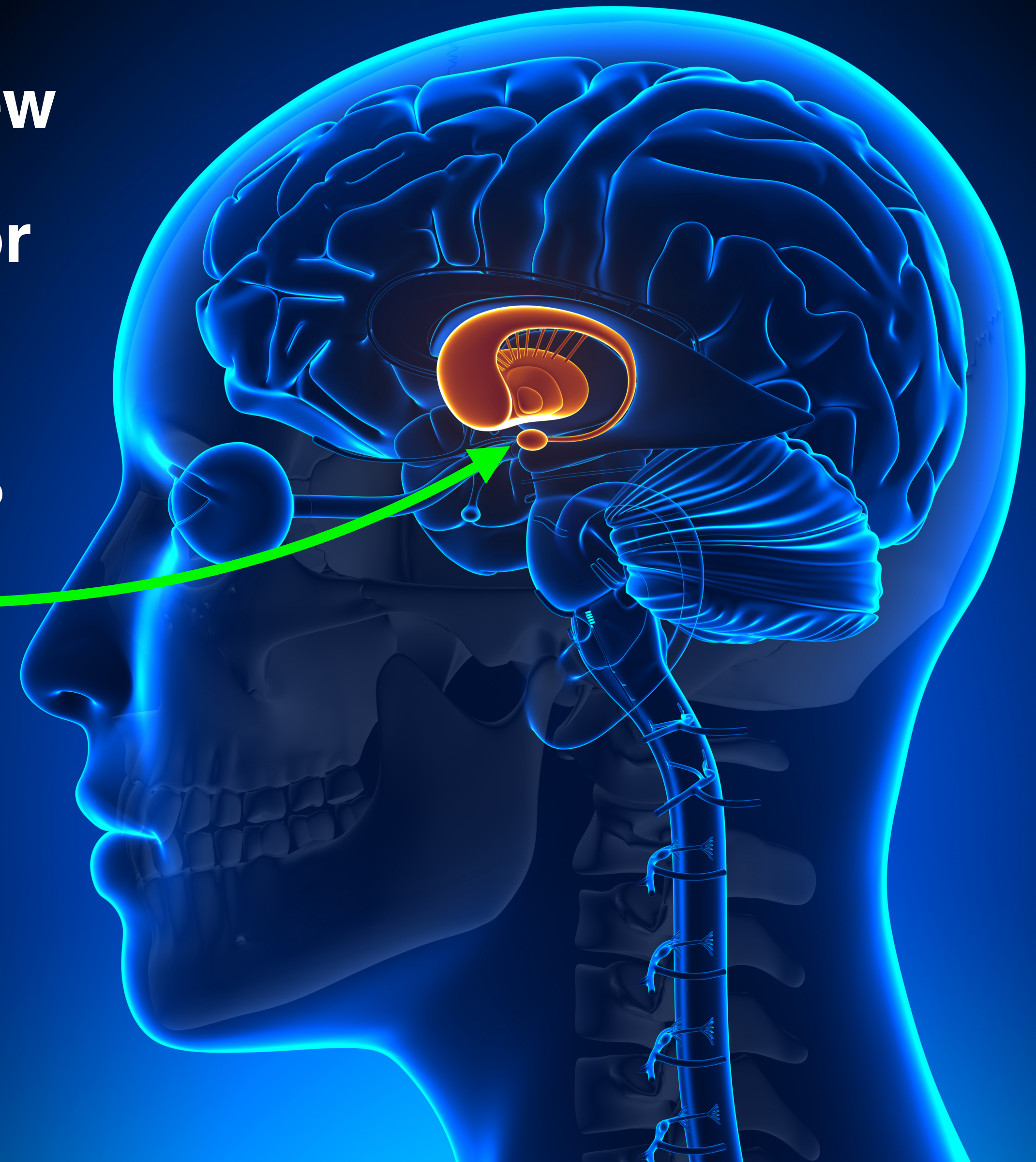
*e.g. Bashivan, Kar and DiCarlo **Science** 364 (2019)*

Visual system

Limbic system



Can this new
ability be
focused for
potential
human
health
benefits ?

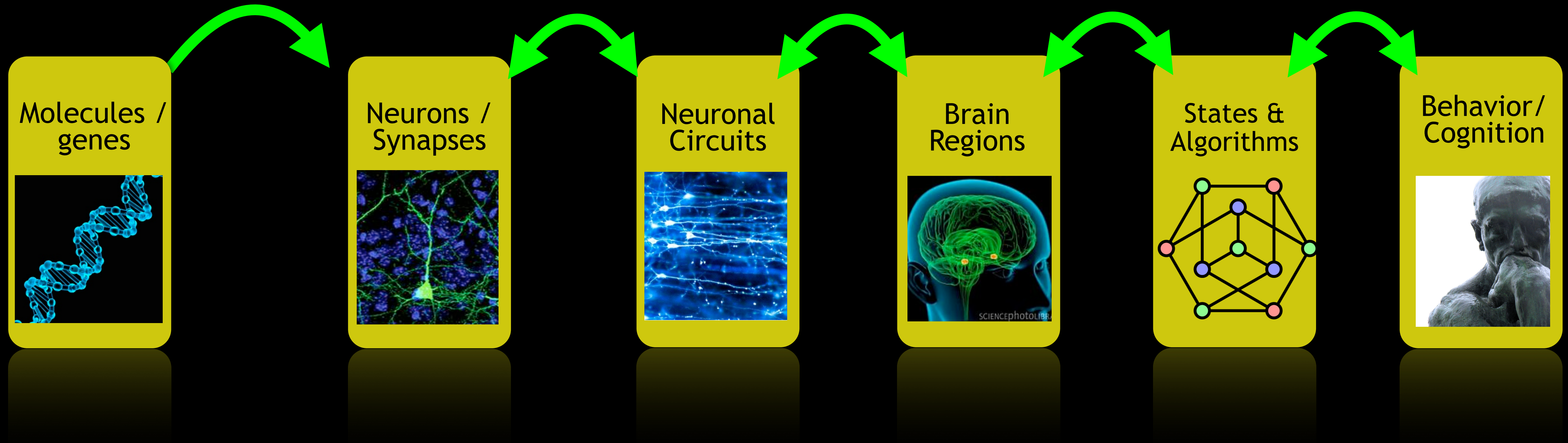


Experiments to test this...

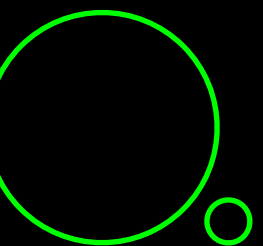
AI-derived models of brain processing could guide us to beneficial, noninvasive, precision mental state interventions



AI-engineered systems as neuroscientific models

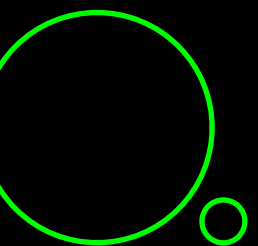


Takeaway #3: The emerging new scientific models will unlock entirely new opportunities to improve human health and education.



Takeaways:

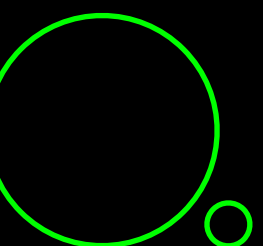
1. The goal of brain and cognitive science is necessarily interlocked with modern AI engineering.
2. Accurate, multi-scale, machine executable models of brain function (aka “digital twins”) are emerging from this AI-forward scientific strategy. This will accelerate.
3. The emerging new scientific models will unlock entirely new opportunities to improve human health and education.
4. **Because AI engineering is core to brain and cognitive science (1-3), nations that do not appropriately embrace this will be left behind.**



Key ways to appropriately embrace:

- Support for medium to large research teams of brain and cognitive scientists tightly partnered with AI-engineers
- The AI capabilities matter: progress requires access to high capability systems as scientific models.
- Be careful and thoughtful about AI regulation.
- Open, integrated benchmarking platforms in brain and cognitive sciences are critical to the virtuous loop of model discovery, improvement and dissemination. (e.g. www.Brain-Score.org)

4. Because AI engineering is core to brain and cognitive science (1-3), nations that do not appropriately embrace this will be left behind.



dicarlo@mit.edu

James DiCarlo MD, PhD

Director, MIT Quest for Intelligence

Professor of Systems & Computational Neuroscience

Department of Brain and Cognitive Sciences

Co-Director & Investigator, Center for Brains, Minds and Machines

Investigator, McGovern Institute for Brain Research

Massachusetts Institute of Technology

