Reverse engineering human visual object recognition

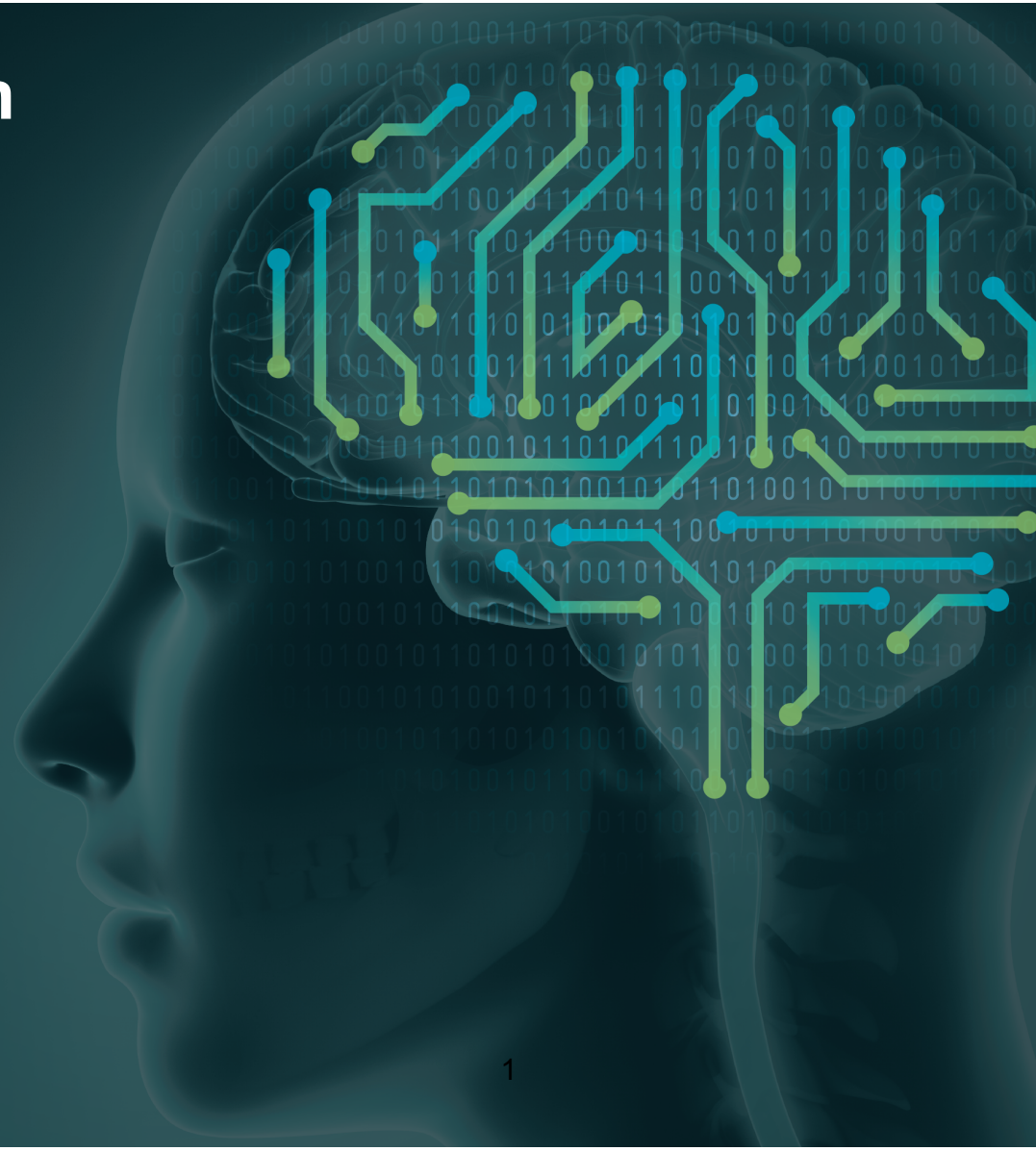
CBMM (Virtual) Summer School
August 2020

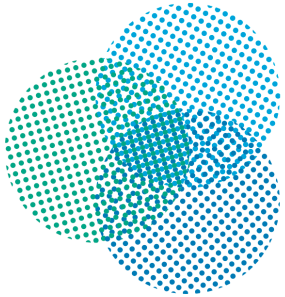
CENTER FOR
Brains
Minds+
Machines

James DiCarlo MD, PhD

Peter de Florez Professor of Neuroscience
Head, Department of Brain and Cognitive Sciences
Investigator, McGovern Institute for Brain Research
Investigator, Center for Brains, Minds and Machines
Massachusetts Institute of Technology

brain+cognitive
sciences





CENTER FOR
Brains
Minds+
Machines

Why are you here?



Prof. Hans-Lukas Teuber

Head, MIT Dept, of Psychology 1964-1977

Today:

brain+cognitive sciences

CENTER FOR Brains Minds+ Machines

~1974

Visual intelligence is astonishing!



Goal of AI researchers: silicon-based (computational) intelligence that meets or exceeds biological (human) intelligence

Some possible strategies:

1. *Ignore the brain sciences completely.*
2. *Ignore the brain sciences in your actual engineering, but talk about your systems as "brain inspired" or "brain-like" —> really good for advertising and PR!*
3. *Use human performance as a benchmark to report your engineering progress —> a bit better, and also good for advertising and PR!*
4. *Study simple, reduced neural systems, hope that "principles" are discovered that are scalable. (e.g. DNA science in viruses, etc.)*
5. *Forward engineer systems within wisely chosen brain science measurements; test deviations from such measurements as you go, and adjust system building direction if deviations are growing rather than*

"Reverse engineering"

Small bonus ;) greatest open science problem in the history of our species

Bonus: enable new advances in human health, education, and brain-machine interface

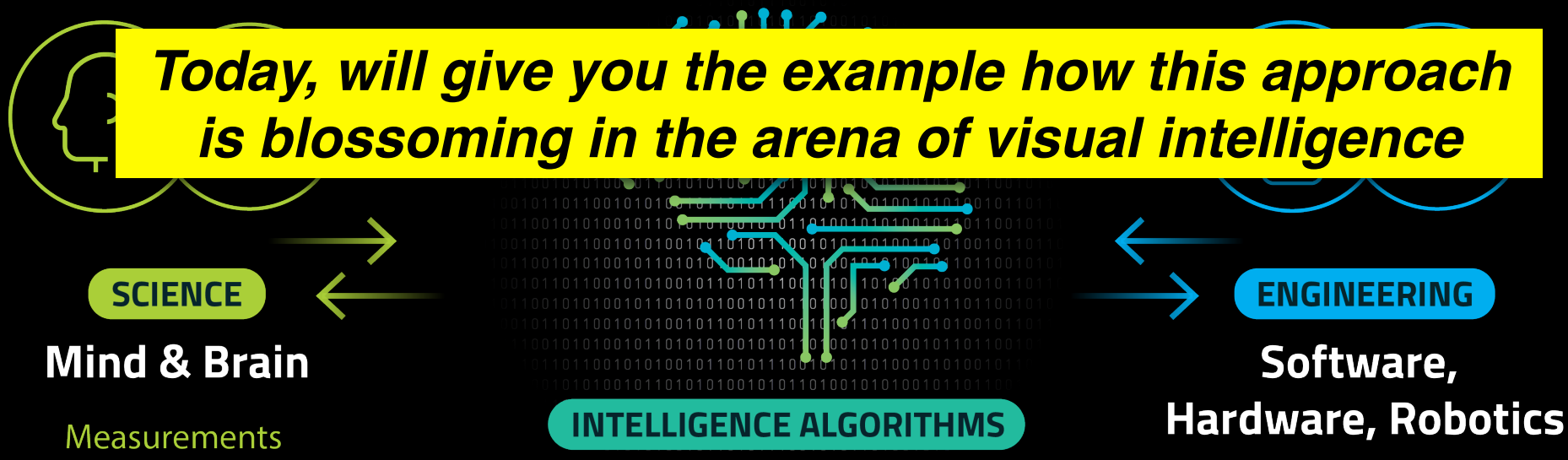
Reverse engineering:

Account for each ability of the mind (intelligent behavior)…

…using a system of components in the brain (connected neurons)…

…in the language of engineering (predictive, built systems).

Today, will give you the example how this approach is blossoming in the arena of visual intelligence



Reverse engineering team:

Current lab:

Robert Ajemian
Pouya Bashivan
Joel Dapello
Kamila Jozwik
Kohitij Kar
Hyodong Lee
Micheal Lee
Tiago Marques
Ratan Murty
Jon Prescott-Roy
Sachi Sanghavi
Kailyn Schmidt
Martin Schrimpf
Christopher Shay
Chris Stawarz



- *Simons Global Brain*
- *Office of Naval Research*
- *National Science Foundation (CBMM)*
- *IBM/MIT Watson Lab*
- *Semiconductor Research Corporation (SRC)/DARPA*

Alumni scientists:

Arash Afraz (=> Prof., NIH)
Paul Aparicio (=> NIH)
Charles Cadieu (=> CEO BayLabs)
David Cox (=> IBM, VP AI research)
Ha Hong (=> BayLabs)
Chou Hung (=> Army research)
Elias Issa (=> Prof., Columbia)
Xioaxuan Jia (=>Allen Institute)
Gabriel Kreiman (=> Prof., Harvard)
Nuo Li (=> Prof., Baylor College Med)
Rishi Rajalingham (=>MIT)
Najib Majaj (=> NYU)
Shay Ohayon (=> Google)
Nicolas Pinto (=> Apple)
Nicole Rust (=> Prof., U Penn)
Dan Yamins (=> Prof., Stanford)
Davide Zoccolan (=> Prof., SISSA)

Key collaborators:

Ed Boyden (MIT)
SueYeon Chung (Columbia)
David Cox (IBM)
Bob Desimone (MIT)
Danny Gutfreund (IBM)
Nancy Kanwisher (MIT)
Lynne Kiorpes (NYU)
J. Anthony Movshon (NYU)
Tomaso Poggio (MIT)
Kaushik Roy (Purdue)
Andreas Tolias (Baylor CM)
Dan Yamins (Stanford)

Starting ~2002: Can we reverse engineer human visual intelligence?



Constraint from brain and cognitive sciences: ~10 deg center of gaze



Constraint from brain and cognitive sciences: ~0.2 sec snapshots



Image adapted from MIT Street Scenes Database (Courtesy of Tommy Poggio)

Is a car here?

Is a person here?

Is a sign here?

What is the position of the car?

What is the pose of the car?

...

Foundational component of visual intelligence:

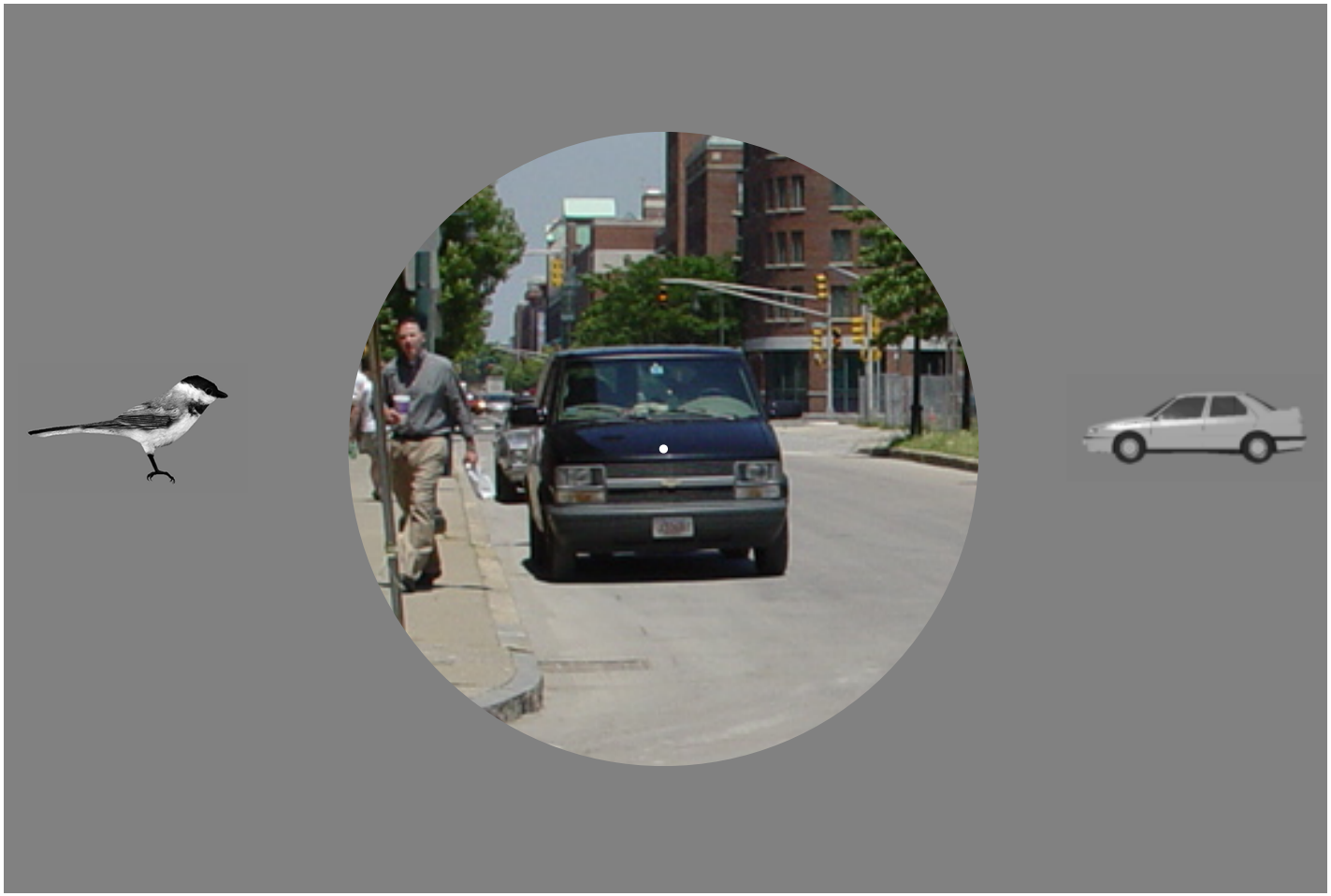
"Core object recognition"



* *Natural operating regime:*

- *center of gaze (~10 deg)*
- *viewing time (~1/5 of a second)*

Example behavioral test trial



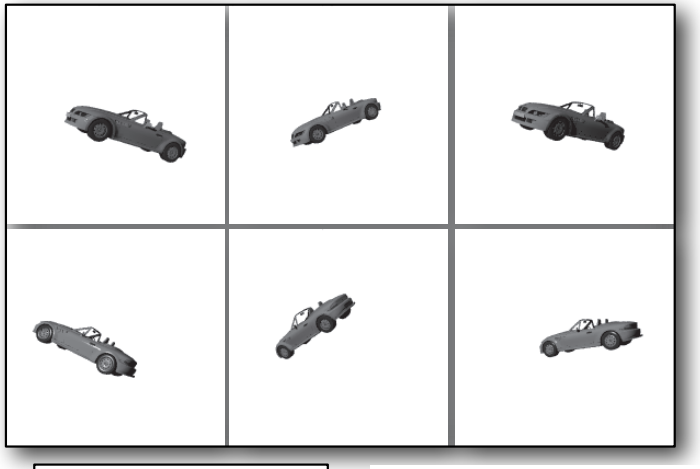
8 deg image at center of gaze, 100 msec viewing time

Why is object recognition computationally challenging?



Identity preserving image variation

View: position, size, pose, illumination





subordinate level variation



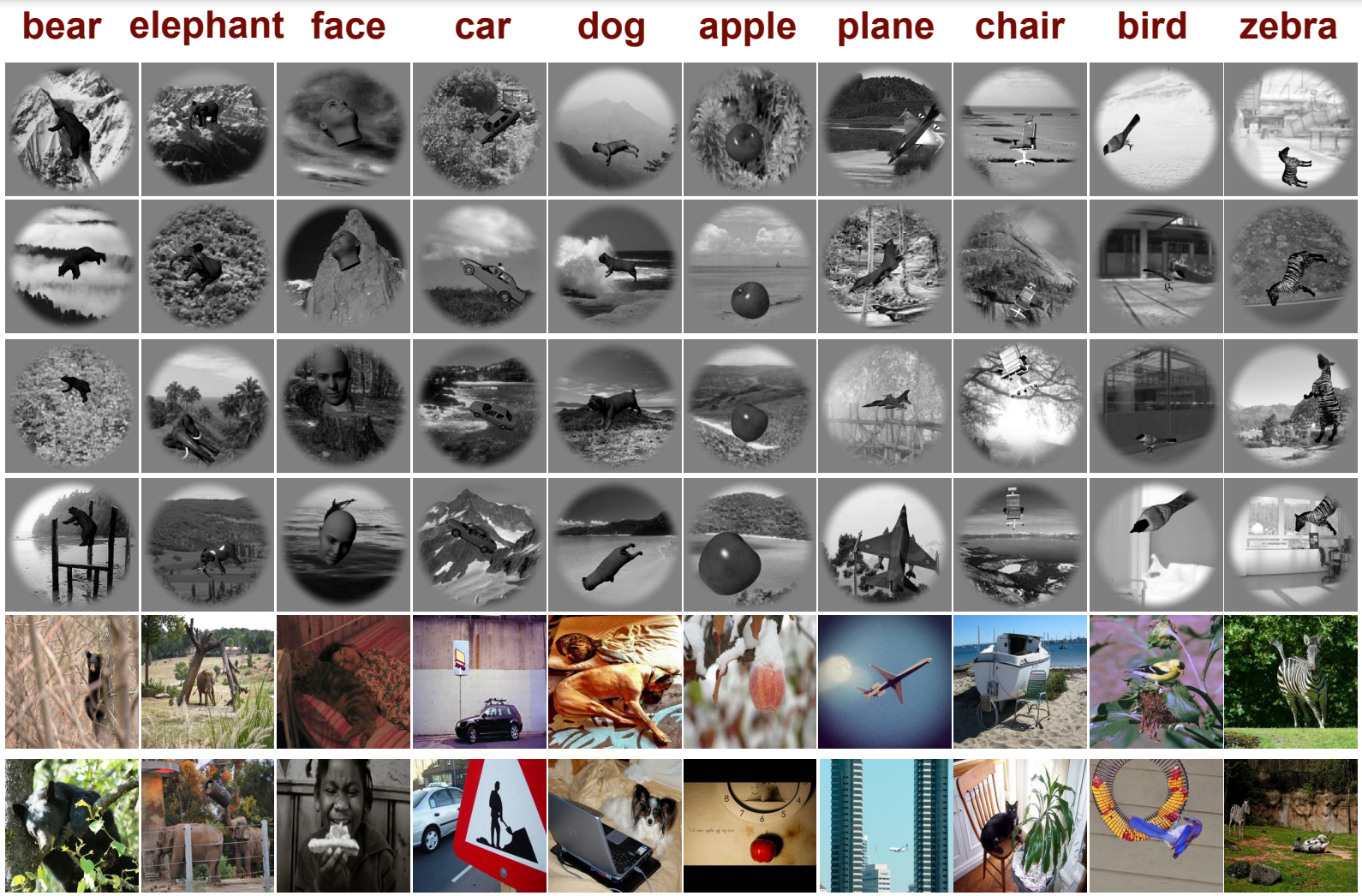


Clutter, occlusion

Poggio, Ullman, Grossberg, Edleman, Biederman, etc.

DiCarlo and Cox, ***TICS*** *(2007), Pinto, Cox, and DiCarlo,* ***PLoS Comp Bio*** *(2008), DiCarlo, Zoccolan and Rust,* ***Neuron*** *(2012)*

We typically test thousands of synthetic images and photographs

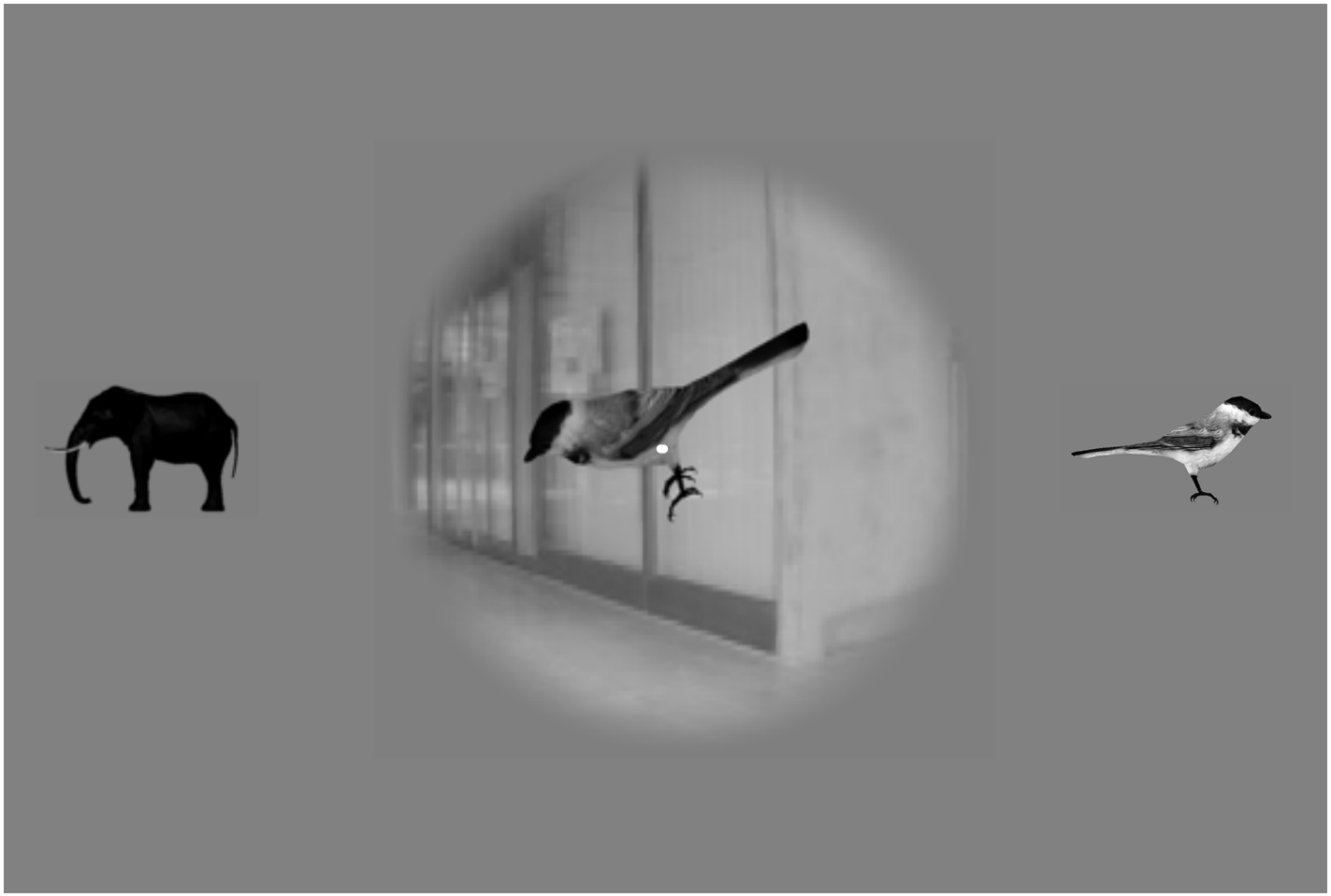


Example behavioral test trial



8 deg image at center of gaze, 100 msec viewing time

Example behavioral test trial



8 deg image at center of gaze, 100 msec viewing time

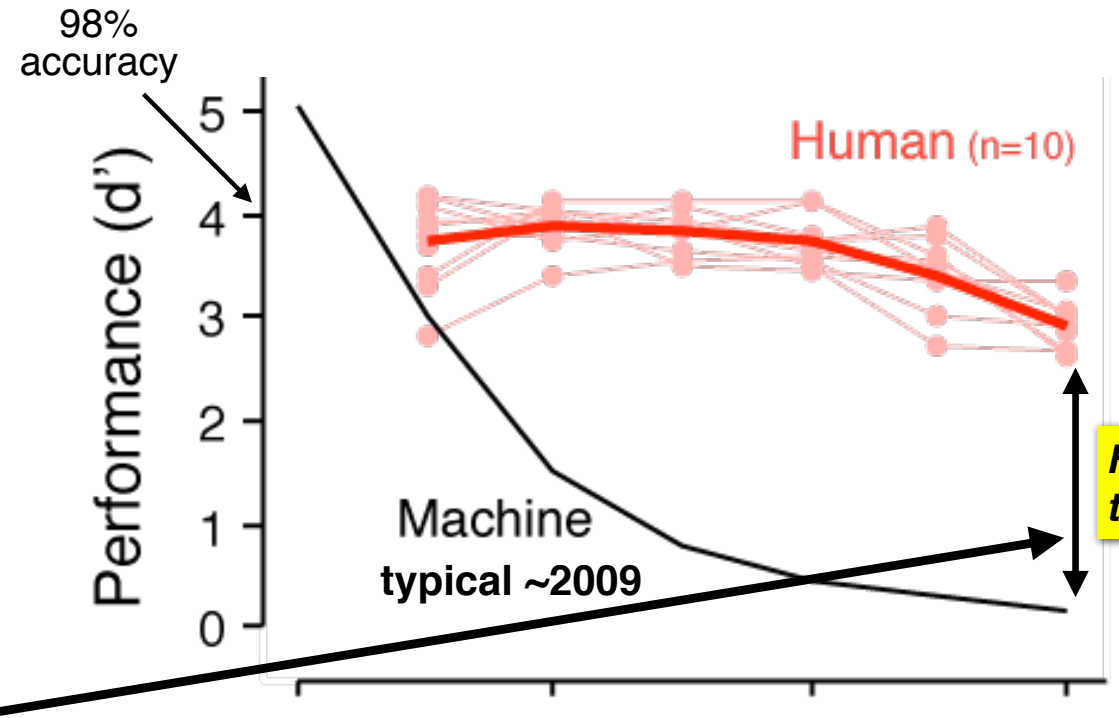
Example behavioral test trial



8 deg image at center of gaze, 100 msec viewing time

Primate: Human

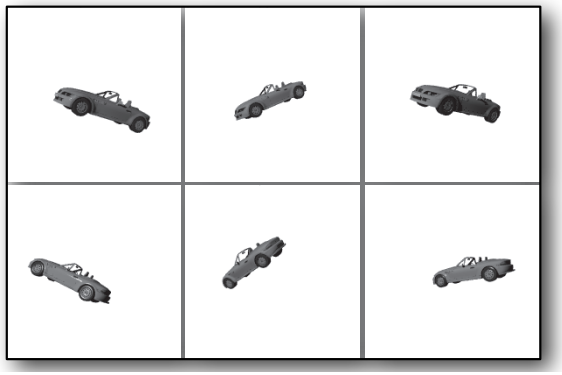




Core object recognition

Low uncertainty in object view (centered, frontal)

High uncertainty in object view

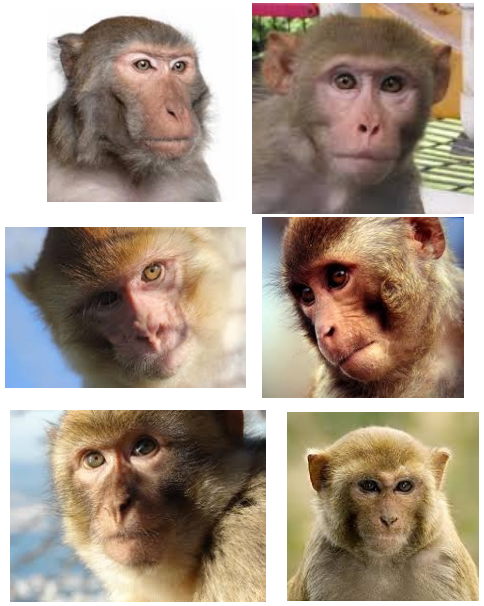


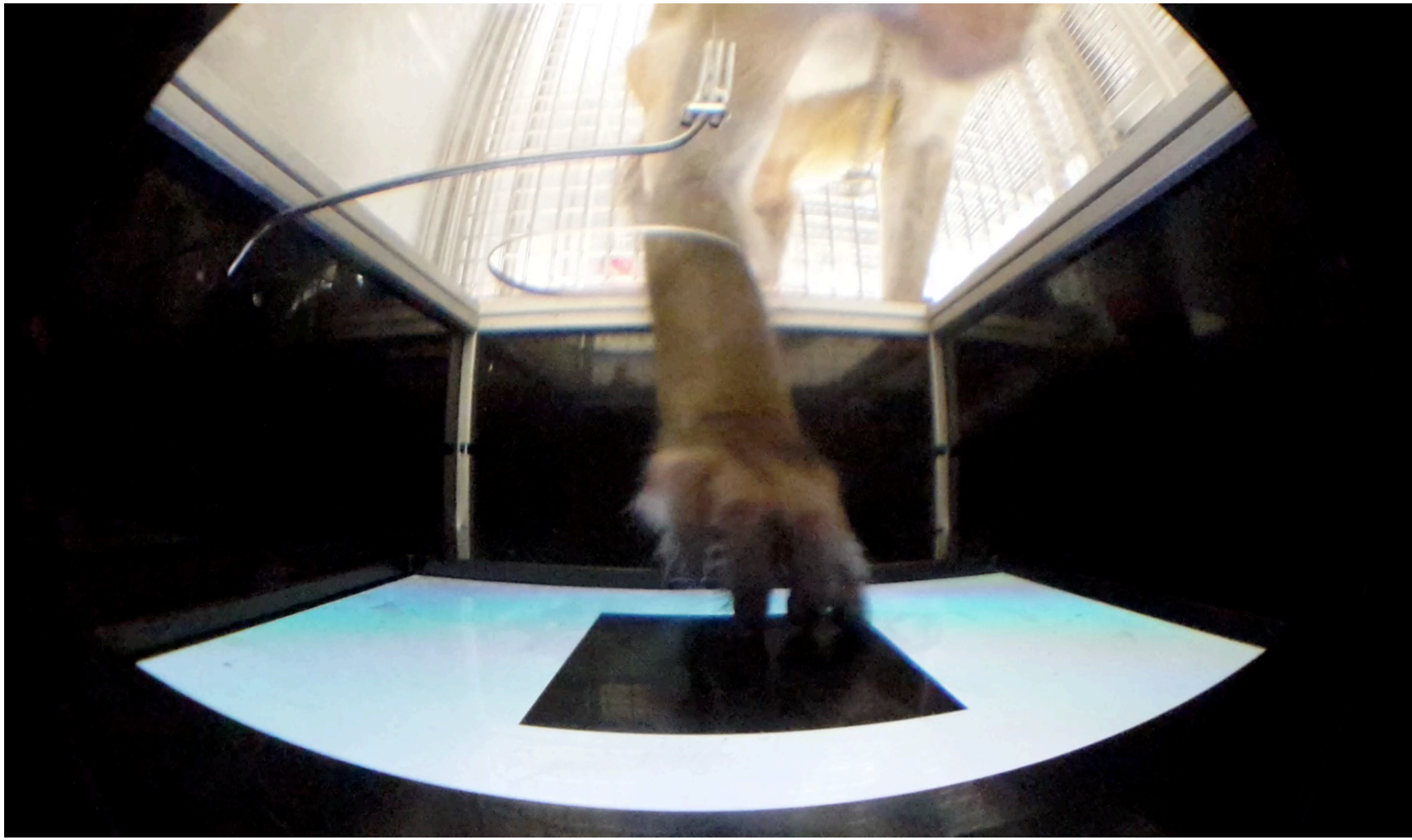
Adapted from Pinto, Cox, and DiCarlo, ***PLoS Comp Bio*** *(2008)*

Primate: Human



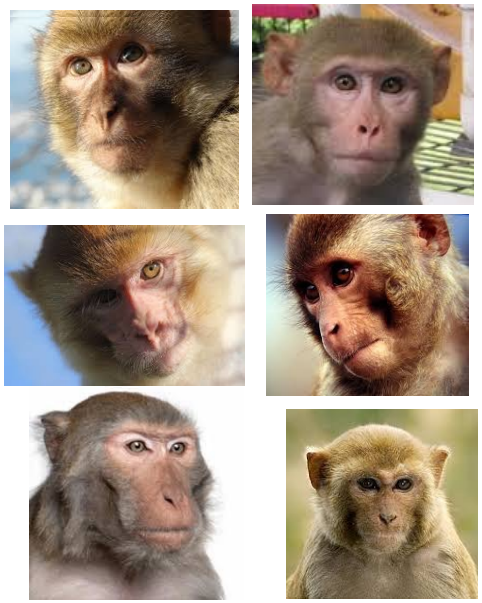
Primate: Rhesus monkey







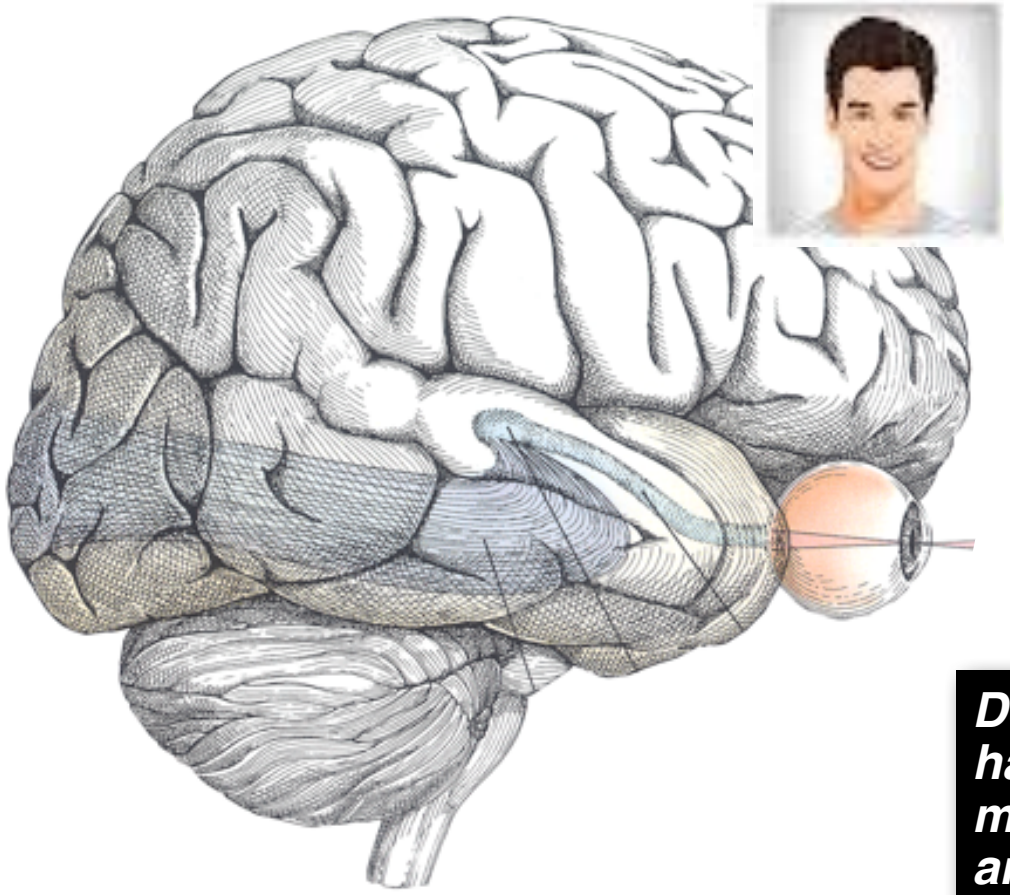


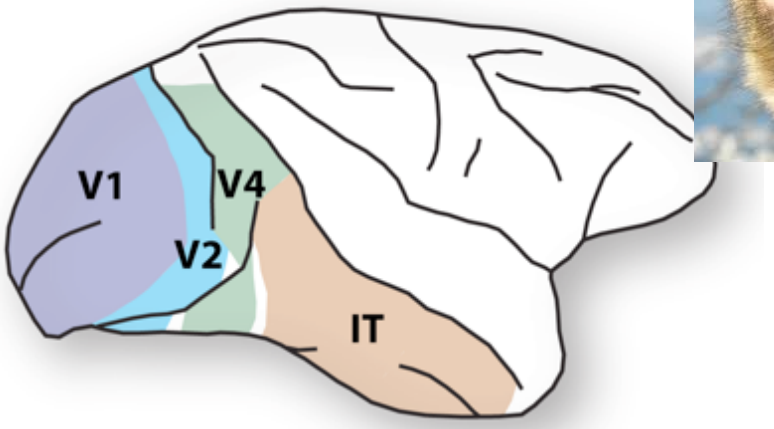


Computer vision systems < [**primate, Homo sapien** = **primate, rhesus monkey**]

Intelligence test: Core Visual Object Recognition

Reverse engineering:

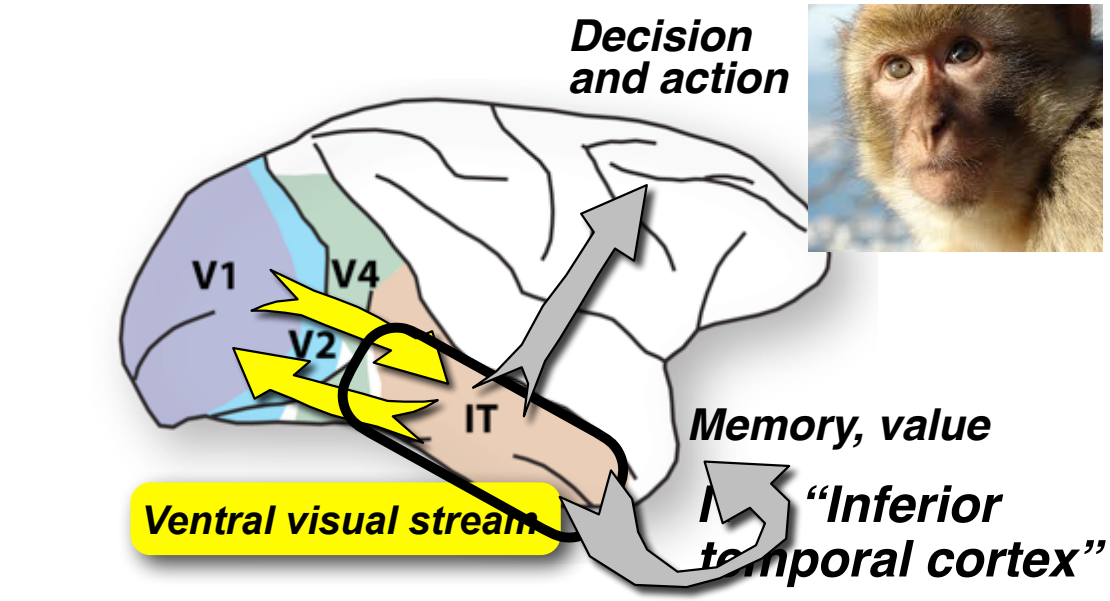






Decades of neuroscience have provided measurements of macro- and meso- architecture

Reverse engineering:

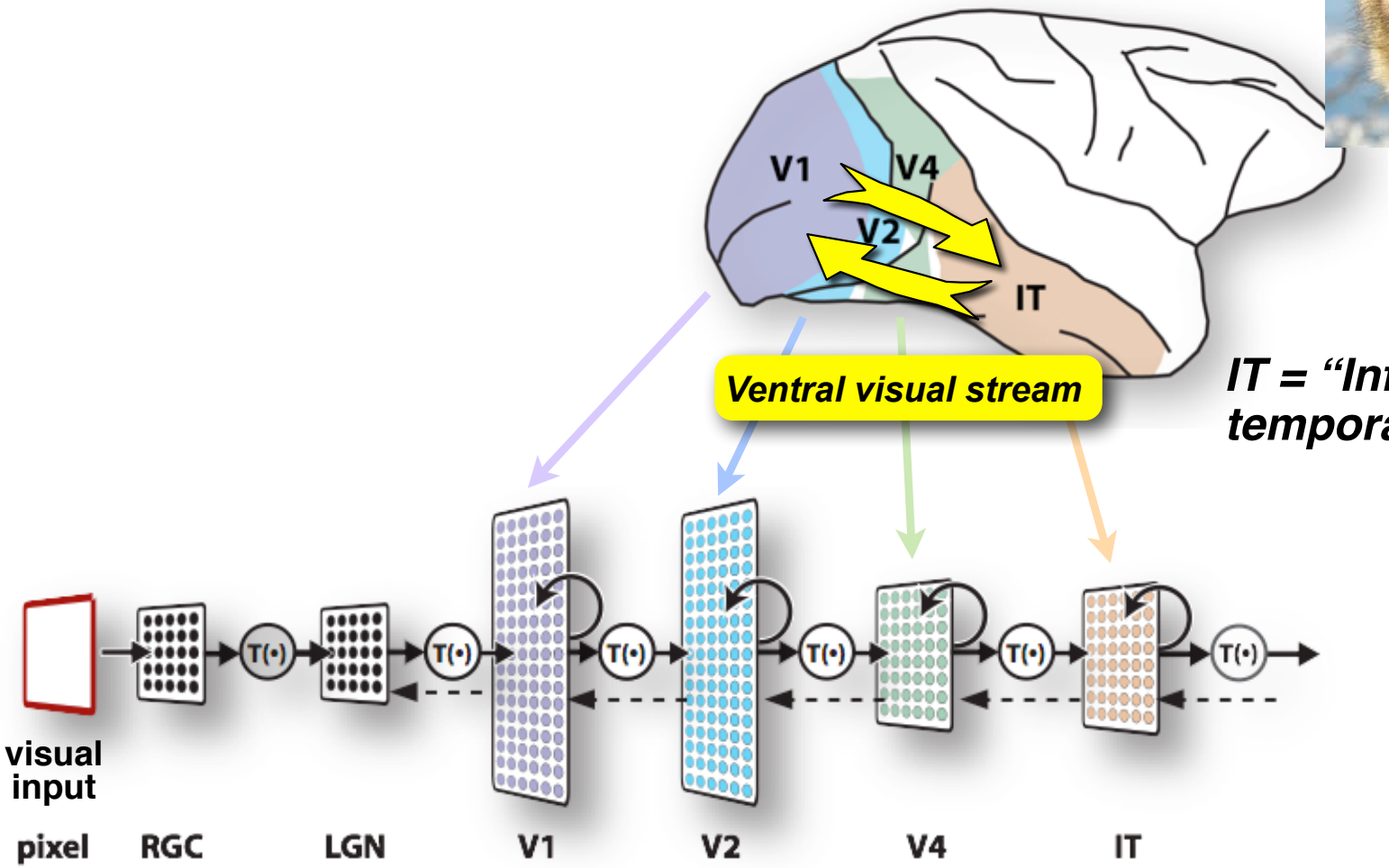


Decades of neuroscience have provided measurements of macro- and meso- architecture

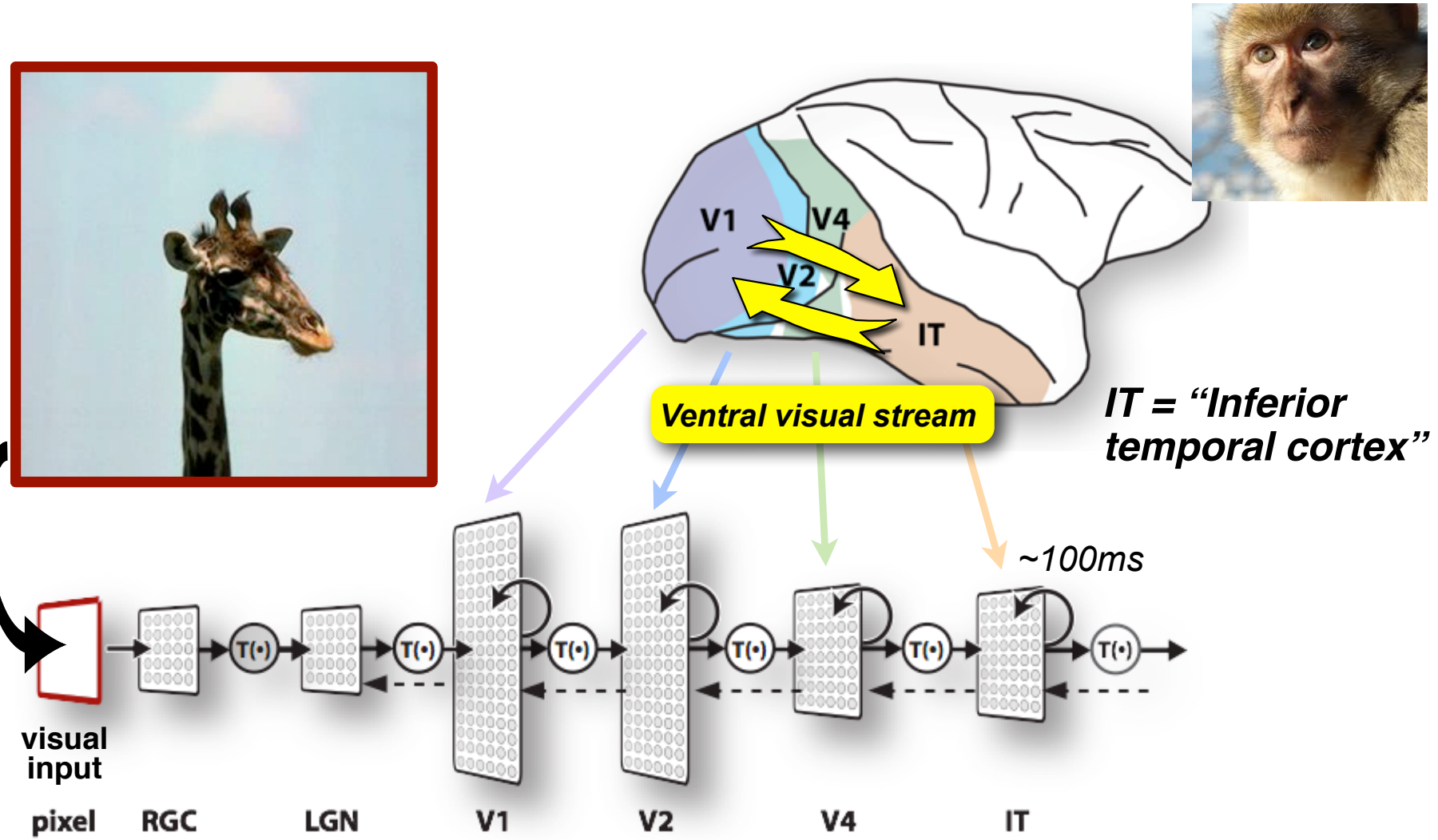
Lesions result in deficits in object recognition.

Reverse engineering:

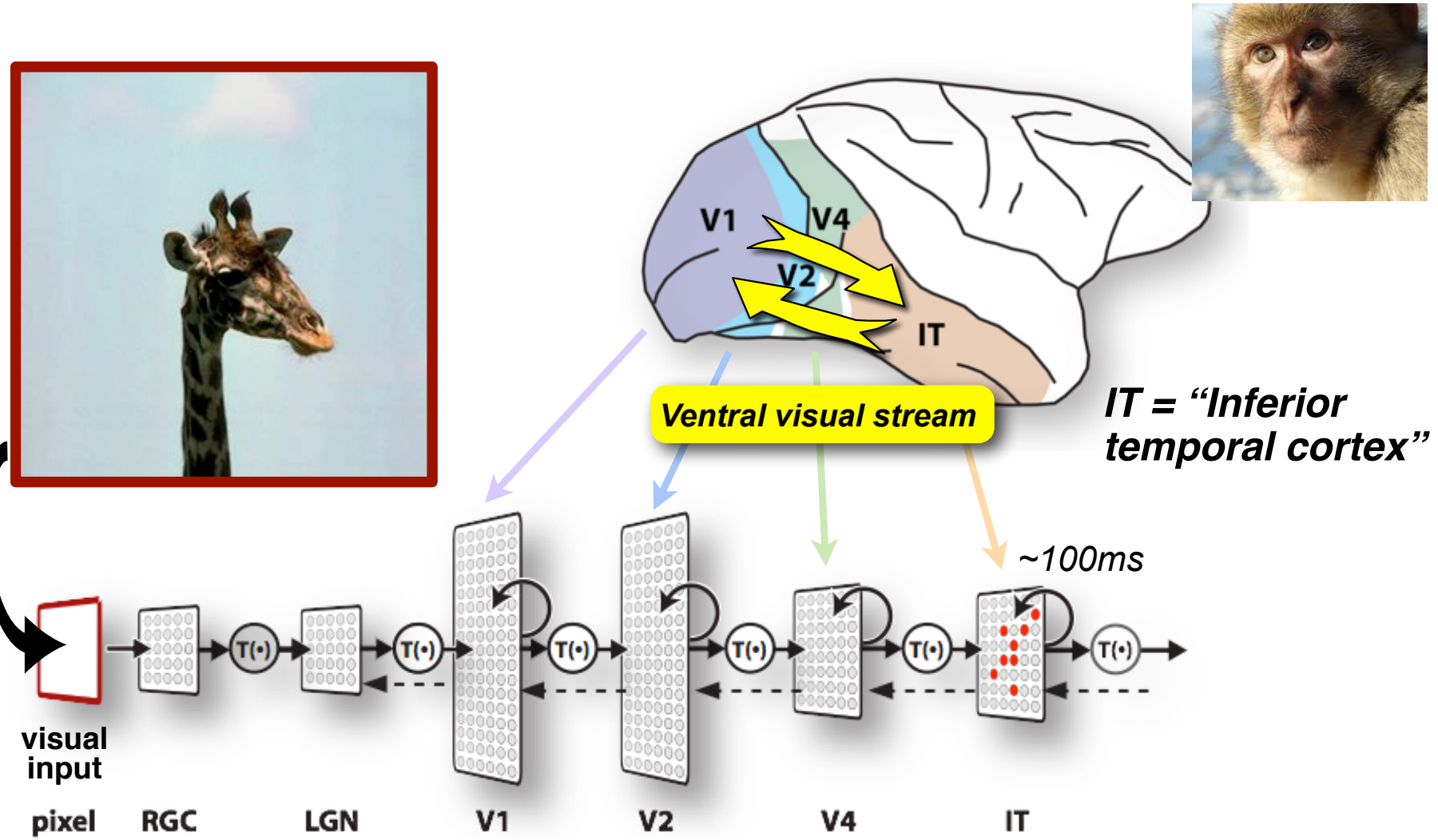




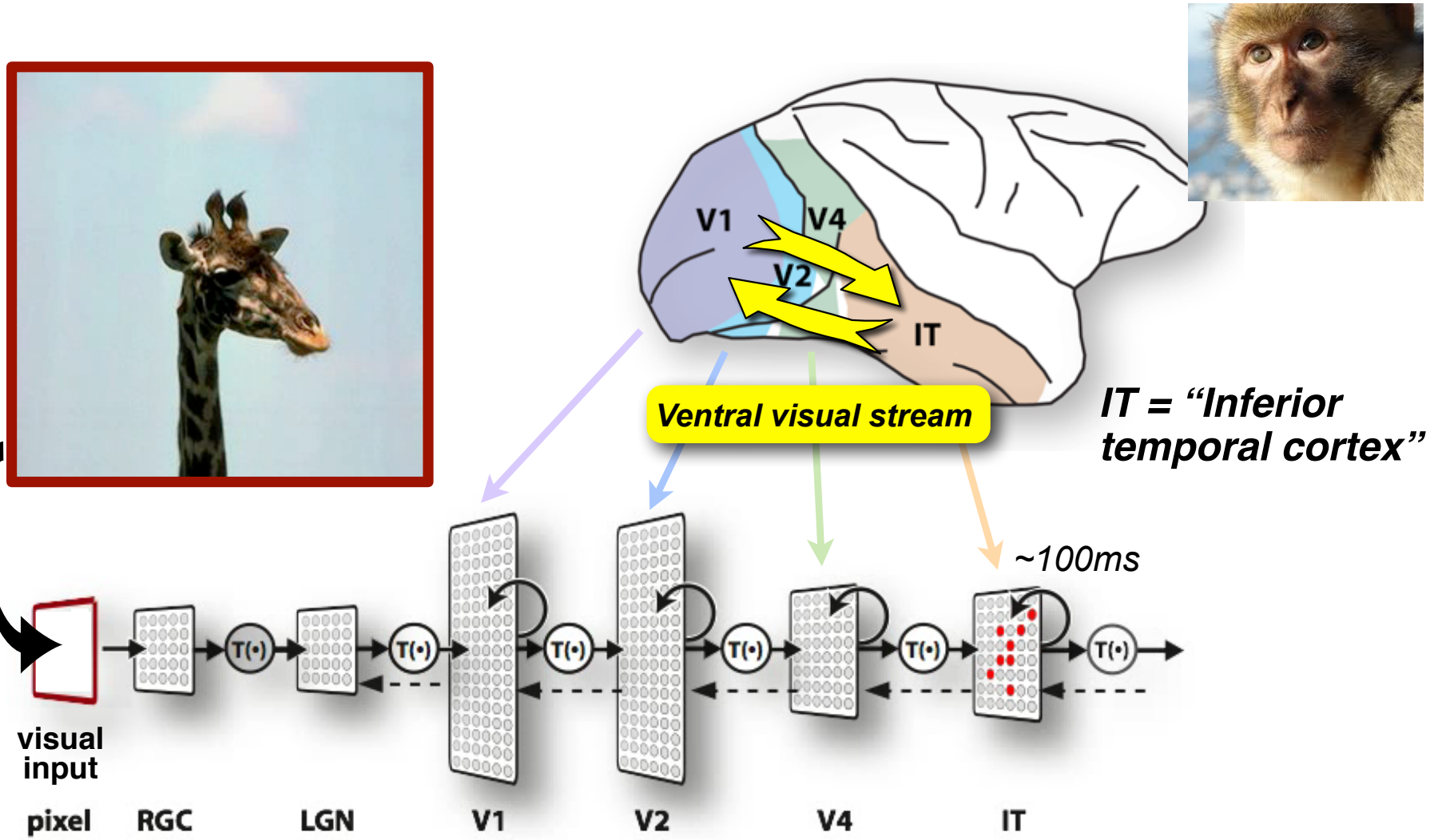
Reverse engineering:



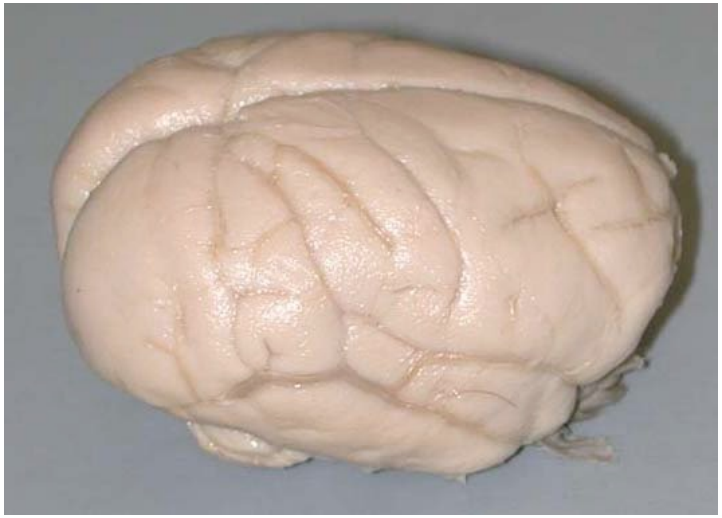
Reverse engineering:

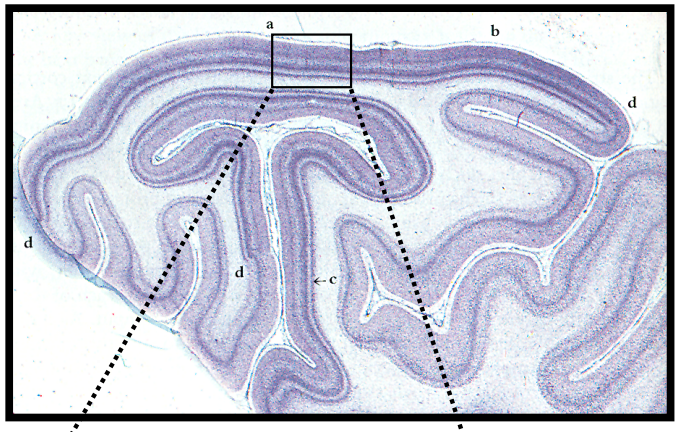


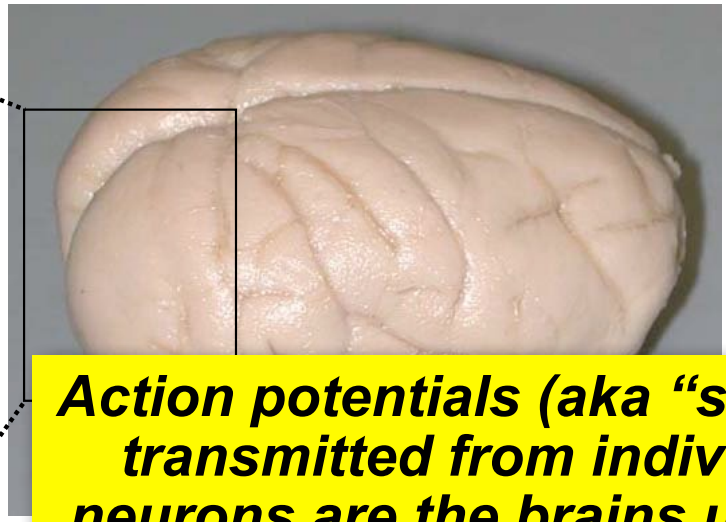
Reverse engineering:







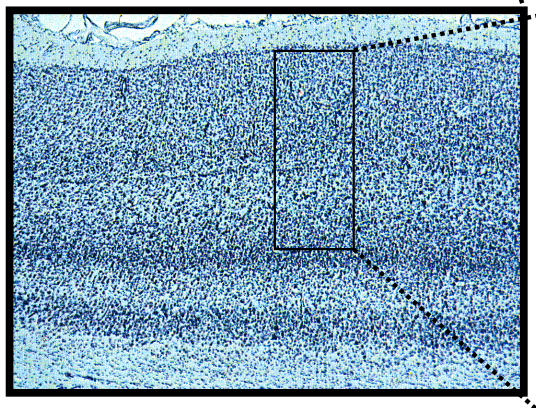


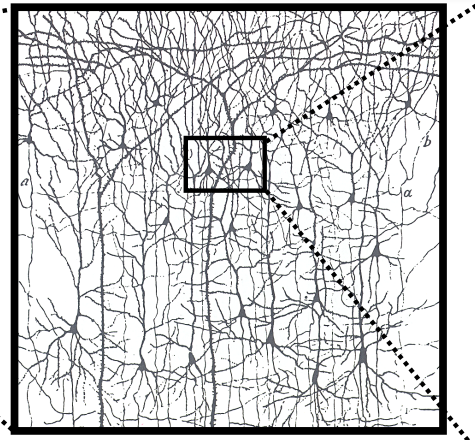


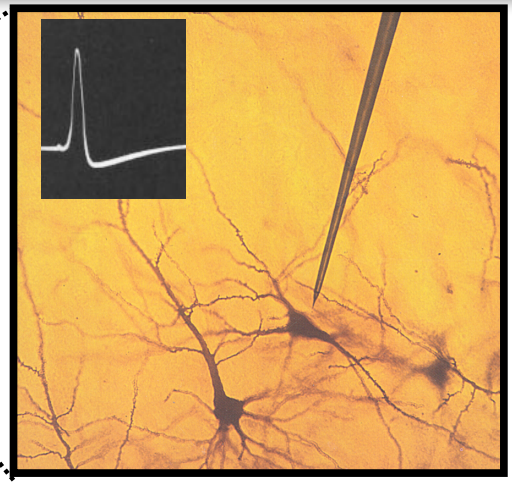
Action potentials (aka "spikes") transmitted from individual neurons are the brains units of online information transmission

Within EVERY cortical area*

Layer 1
Layer 2/3
Layer 4
Layer 5
Layer 6

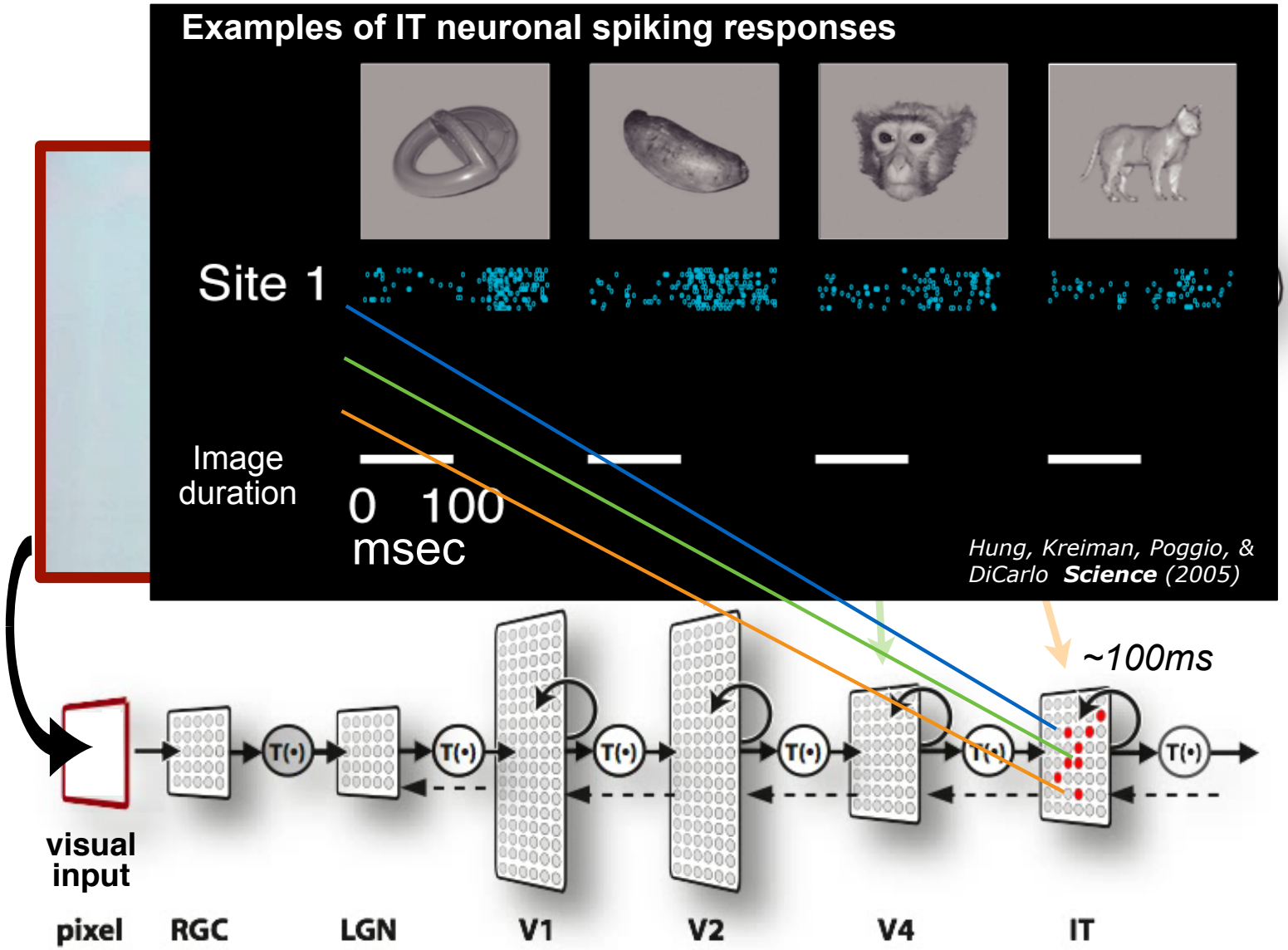


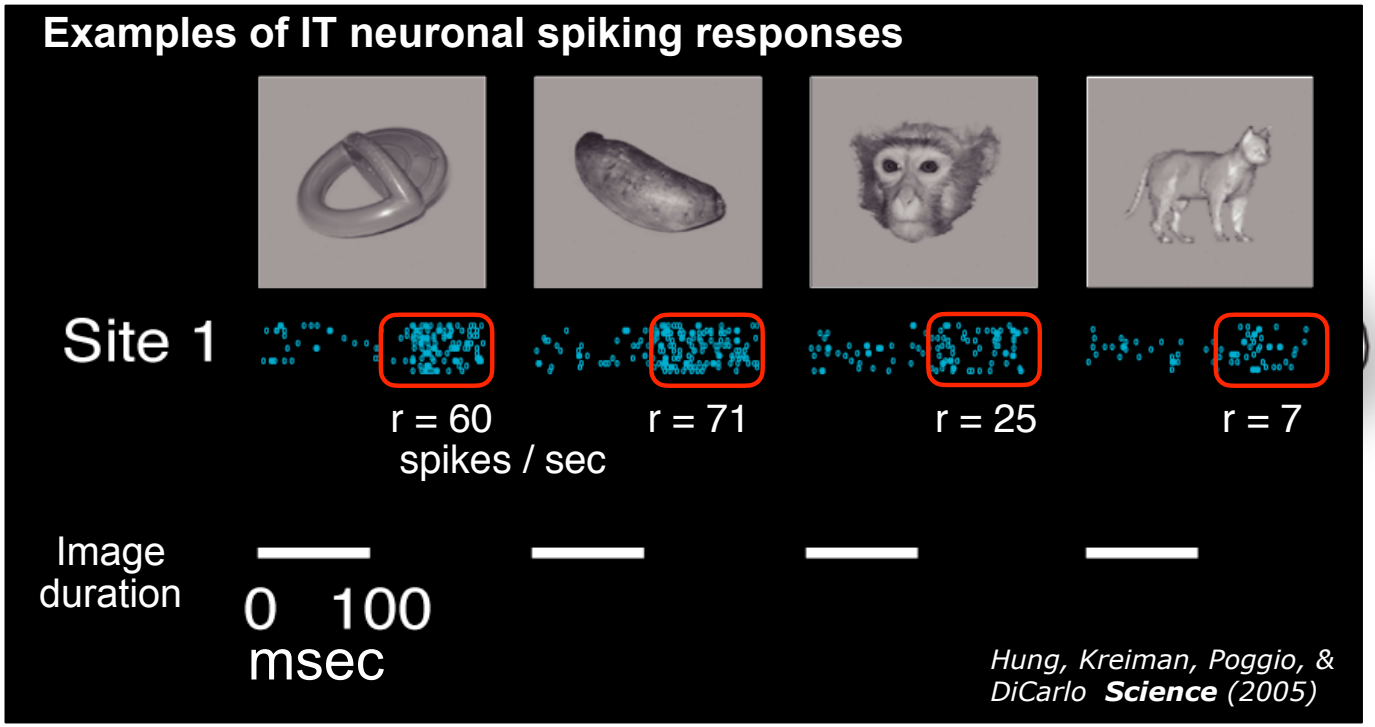




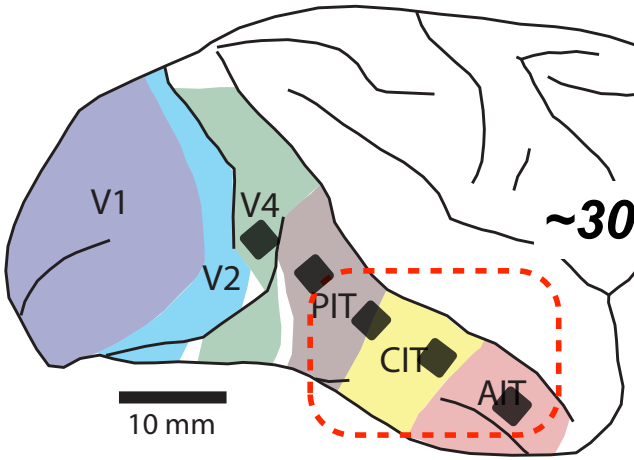
*neuroscience "area" == "layer" in current ventral stream models

Adapted from Hubel, 1988

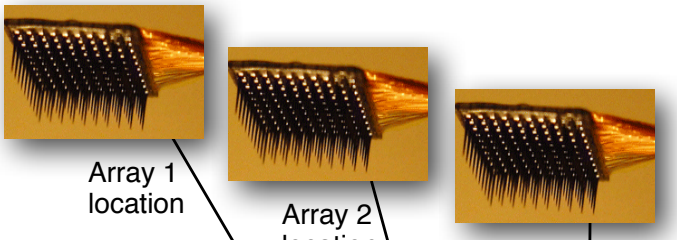

Examples of IT neuronal spiking responses
Site 1
Image duration
0 100
msec
Hung, Kreiman, Poggio, & DiCarlo Science (2005)
~100ms
T(•)
visual input
pixel
RGC
LGN
V1
V2
V4
IT


Examples of IT neuronal spiking responses
Site 1
r = 60
spikes / sec
r = 71
r = 25
r = 7
Image
duration
0 100
msec
Hung, Kreiman, Poggio, &
DiCarlo Science (2005)

We can estimate the neural population code by recording many neurons



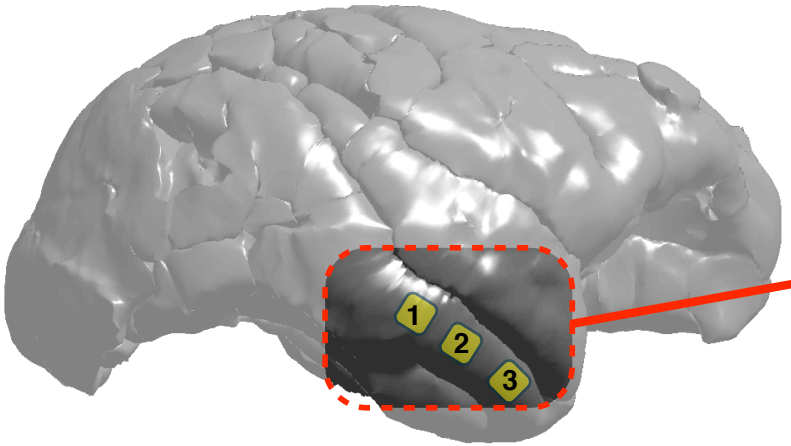
~300 simultaneous spiking sites

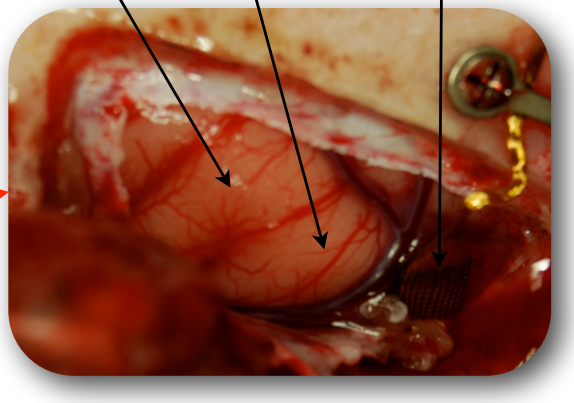


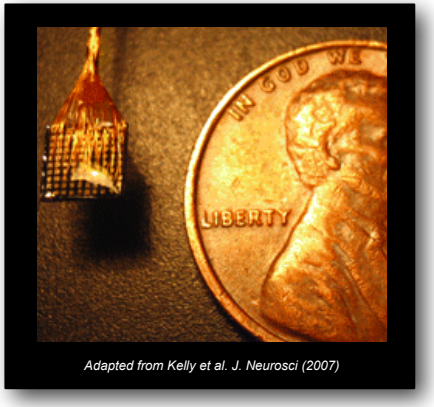
Array 1 location

Array 2 location

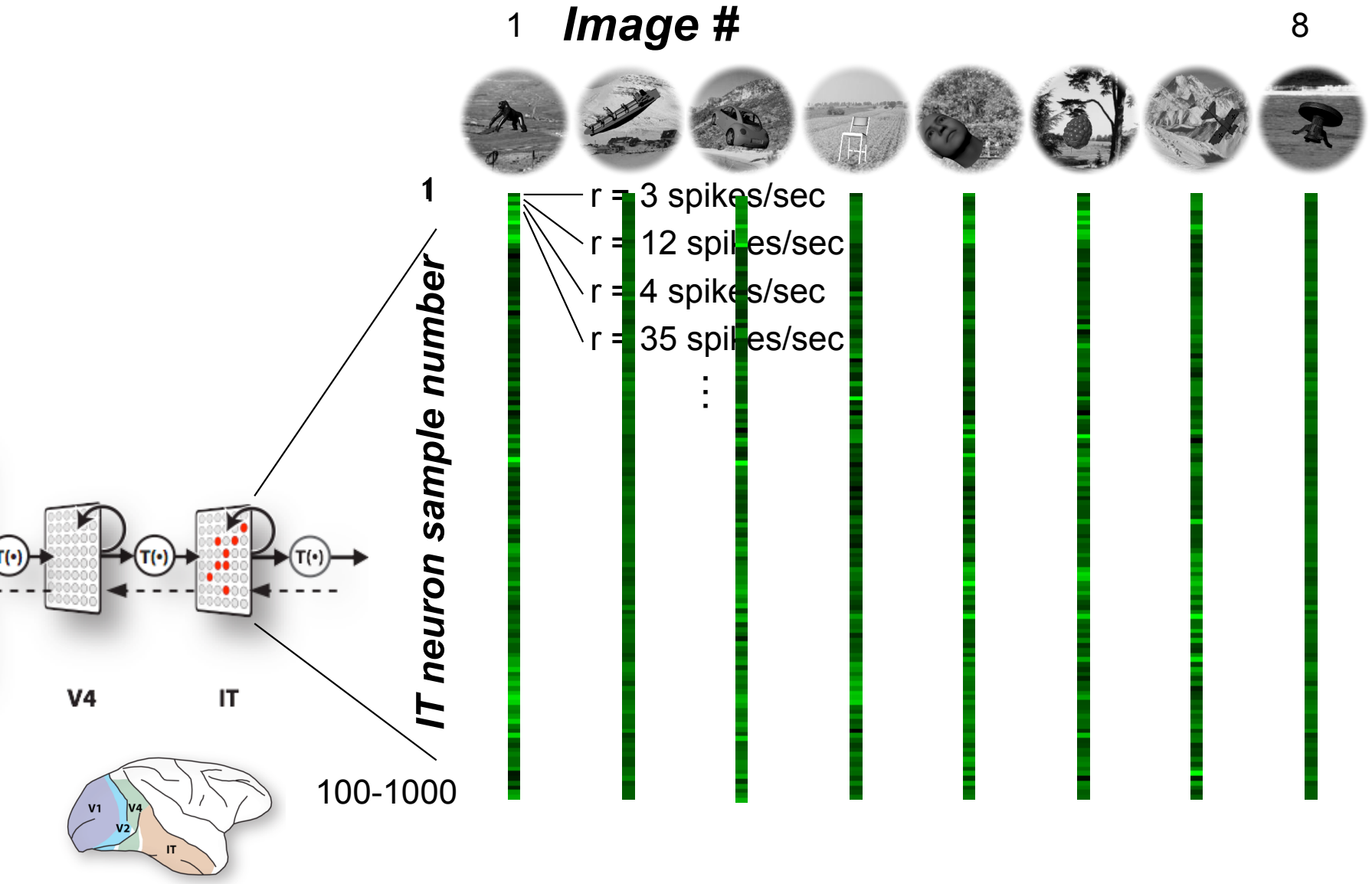
Array 3 (in place)







Adapted from Kelly et al. J. Neurosci (2007)


1
Image #
8
1
r = 3 spikes/sec
r = 12 spikes/sec
r = 4 spikes/sec
r = 35 spikes/sec
IT neuron sample number
100-1000
T(•)
T(•)
T(•)
V4
IT
V1
V4
V2
IT

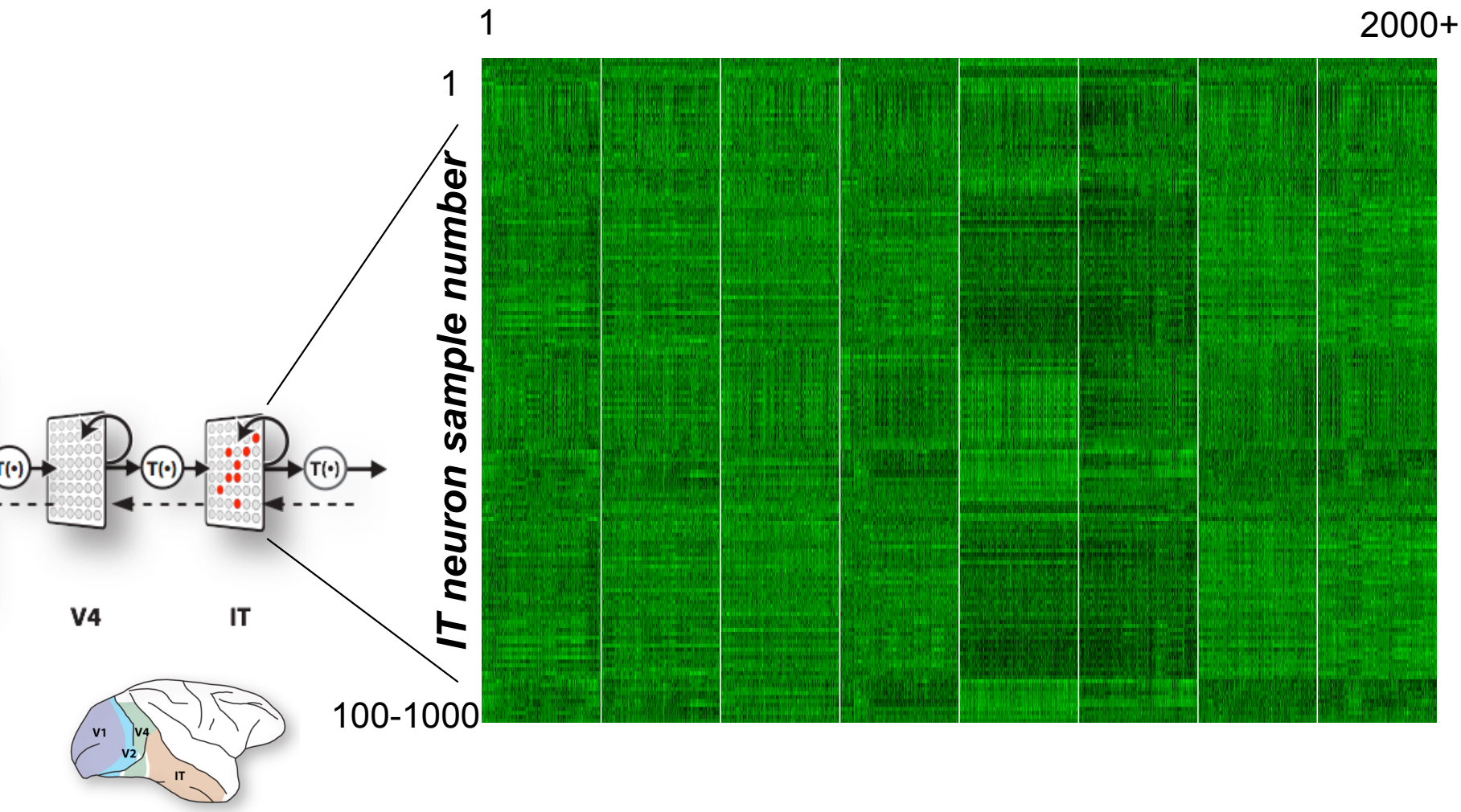
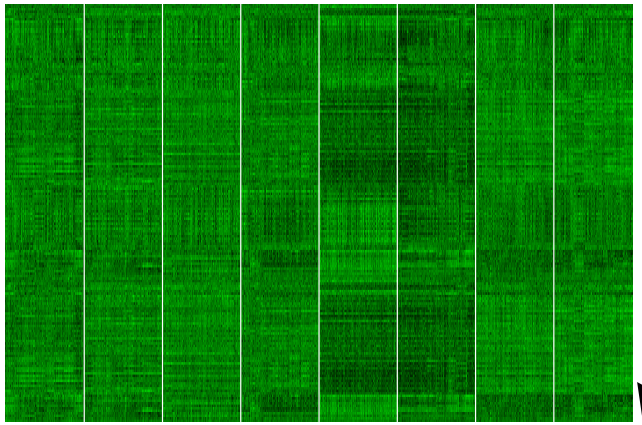


Image #
1
2000+
1
IT neuron sample number
100-1000
T(•)
T(•)
T(•)
V4
IT
V1
V4
V2
IT

Hung, Krieman, Poggio, & DiCarlo ***Science*** *(2005)*

Majaj, Hong, Soloman, & DiCarlo, ***J Neuro*** *(2015)*

Brain (IT) population patterns of activity evoked by each image

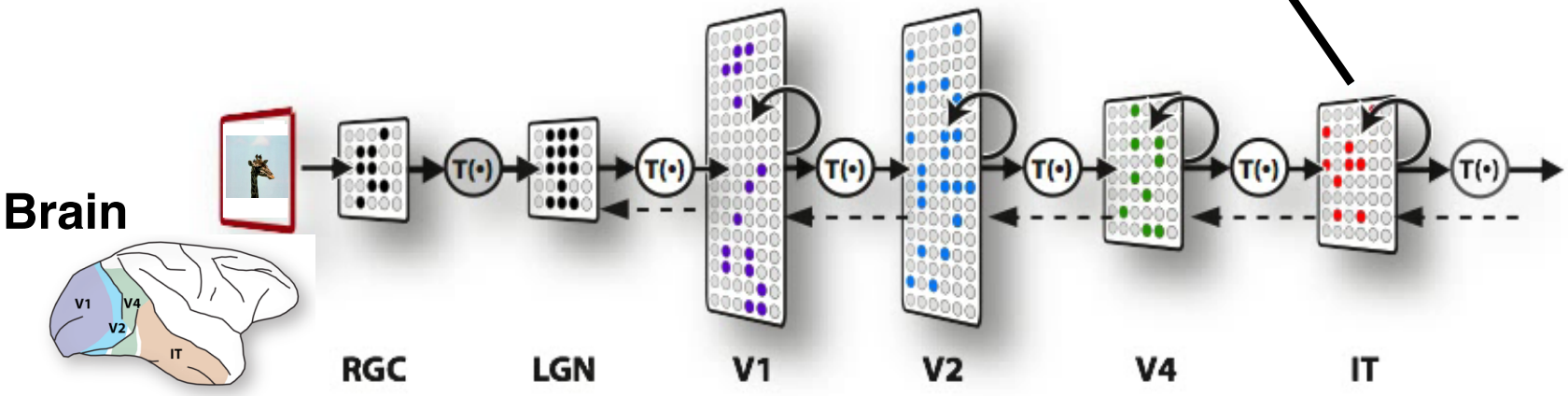


~Linear (decoder) mechanism explains!

Core object recognition behavior



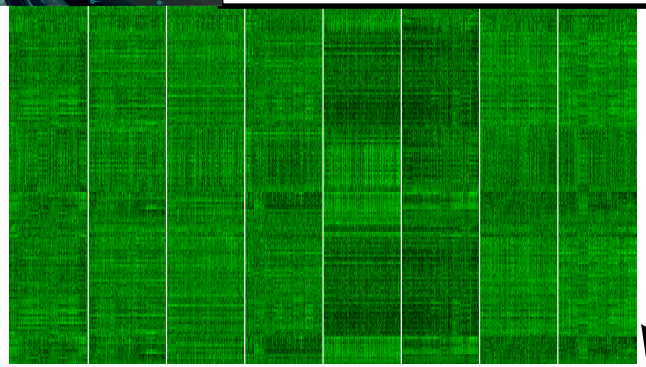






AI/CV relevance: Primates better than computer vision systems by virtue of having this IT code.

Brain (IT) population patterns of activity evoked by each image



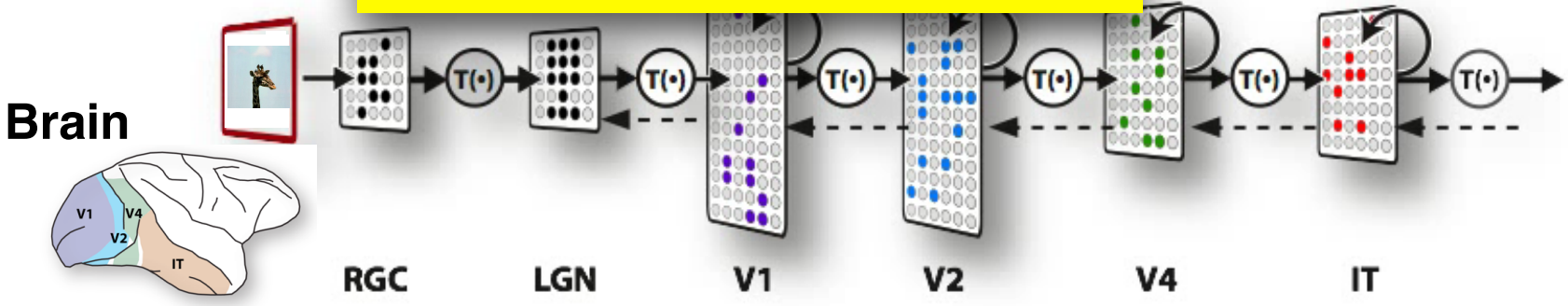
~Linear (decoder) mechanism explains!

Core object recognition behavior





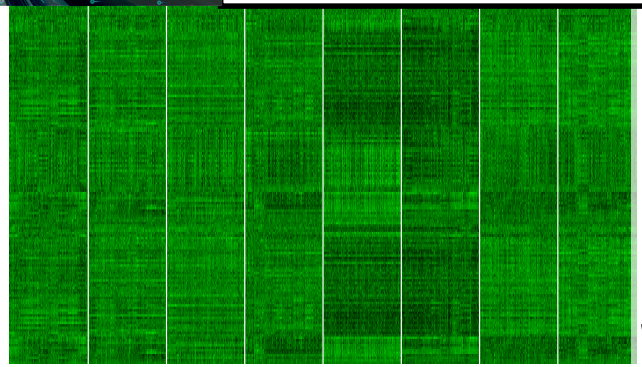
Pause for Questions





AI/CV relevance: Primates better than computer vision systems by virtue of having this IT code.

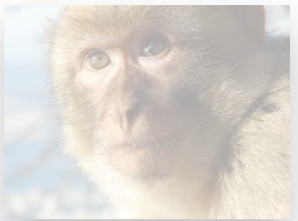
Brain (IT) population patterns of activity evoked by each image



~Linear (decoder) mechanism explains!

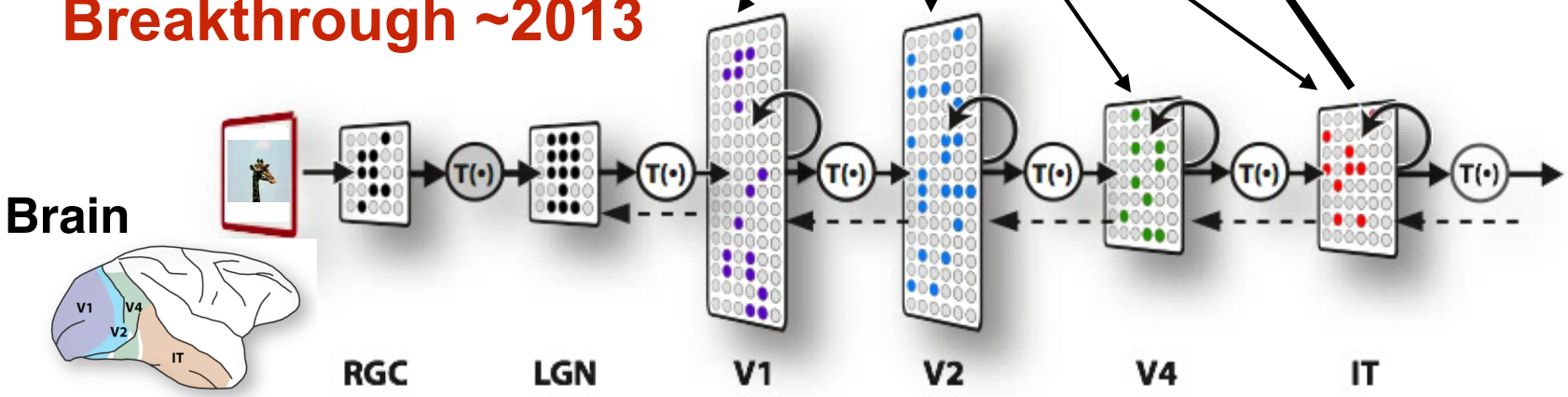
Core object recognition behavior



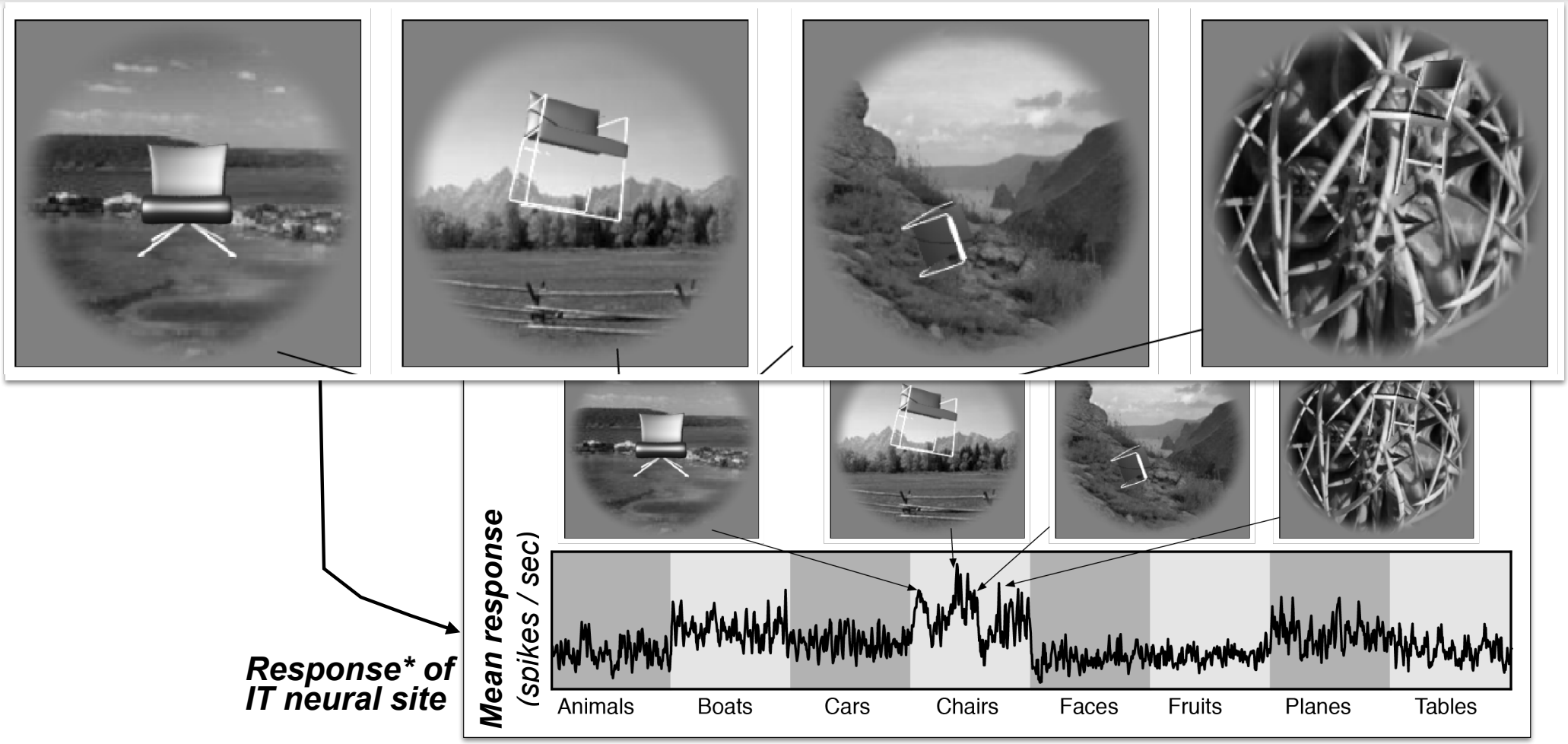


But how is the IT code computed from the image?
Similarly: what are the intermediate codes?

Breakthrough ~2013

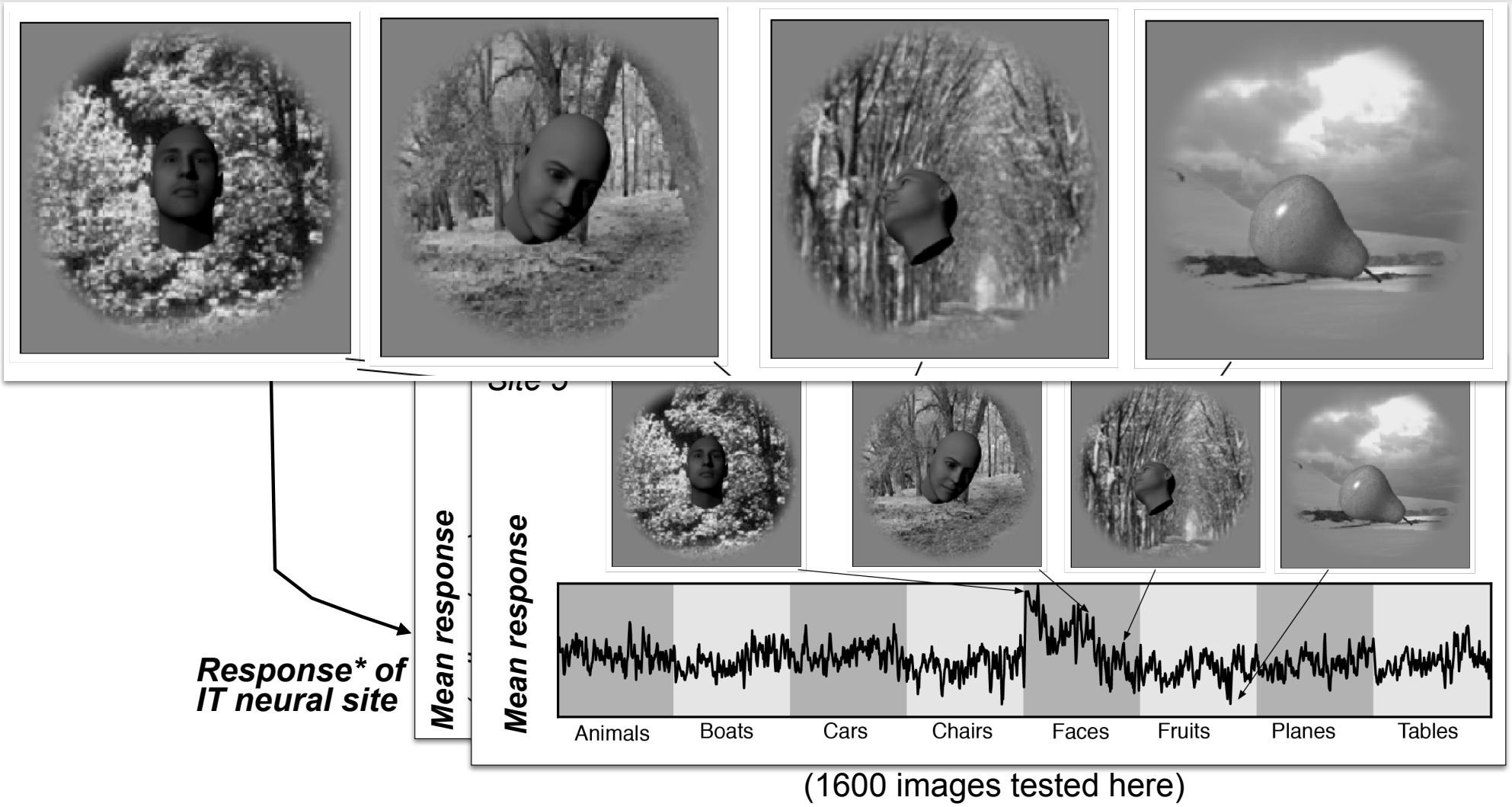


In 2013, the responses of mid (V4) and high (IT) level ventral stream neurons were a mystery.



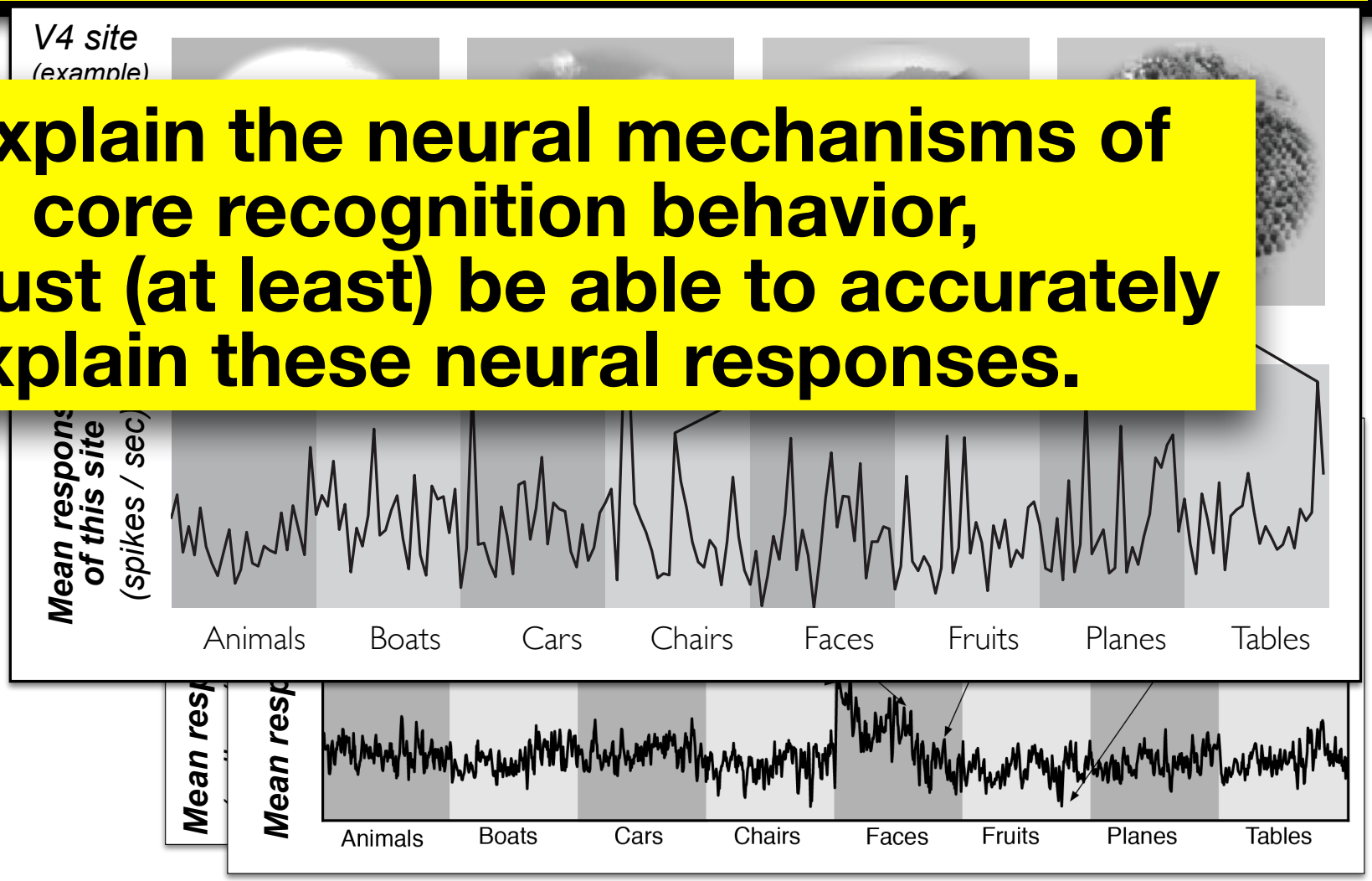
(1600 images tested here)

In 2013, the responses of mid (V4) and high (IT) level ventral stream neurons were a mystery.



In 2013, the responses of mid (V4) and high (IT) level ventral stream neurons were a mystery.

To explain the neural mechanisms of core recognition behavior, we must (at least) be able to accurately explain these neural responses.

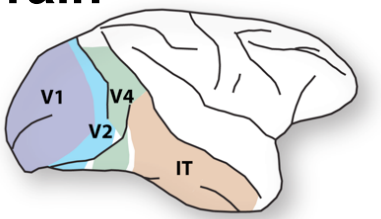


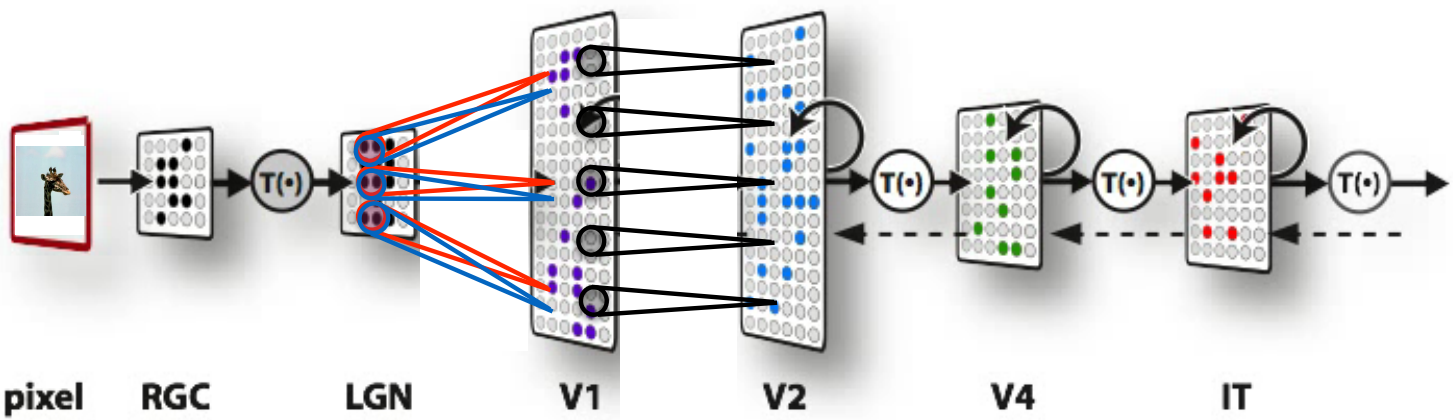
(1600 images tested here)

GUIDANCE FROM NEUROSCIENCE:

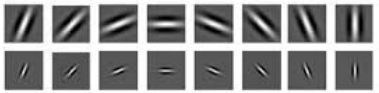


Brain



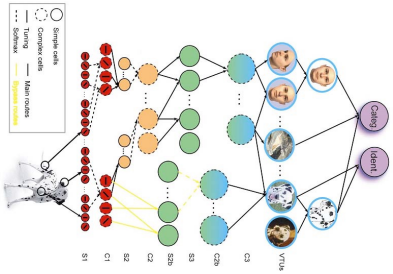


Hypotheses (computer models):

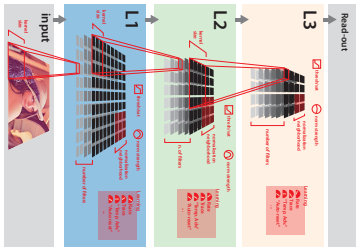


"V1-like" models

"V2-like" models



HMAX models



PLoS09 models

~1980-2010

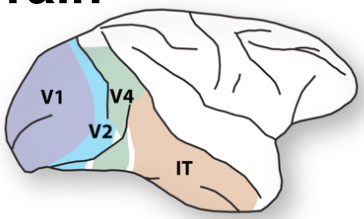
GUIDANCE FROM NEUROSCIENCE:

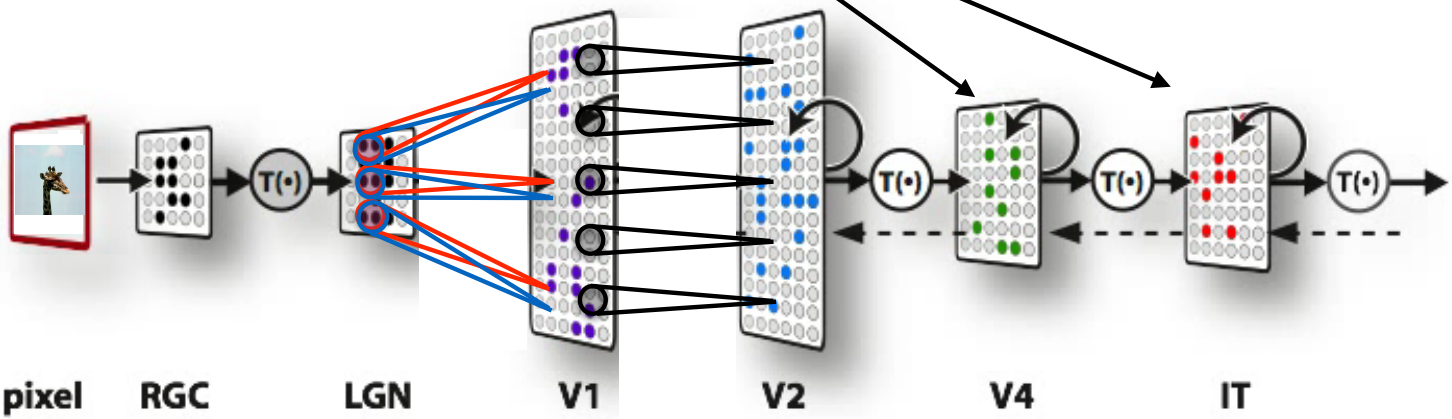


Unfortunately, <u>all</u> of these scientific hypotheses were inadequate.

… they all failed to accurately explain and predict the internal neural responses.

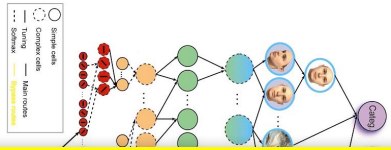
Brain

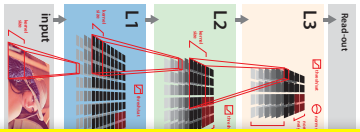




Hypotheses (computer models):







The underlying problem: most of the parameters of these hypotheses (models) are not determined by existing science results !

"V

"V2-like" models

HMAX models

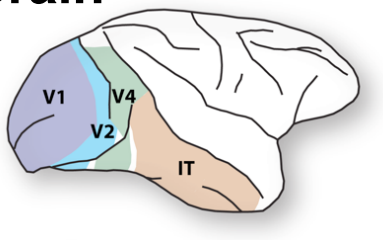
PLoS09 models

~1980-2010

GUIDANCE FROM NEUROSCIENCE:



Brain



Maximum possible (depends on recording SNR) → 100%

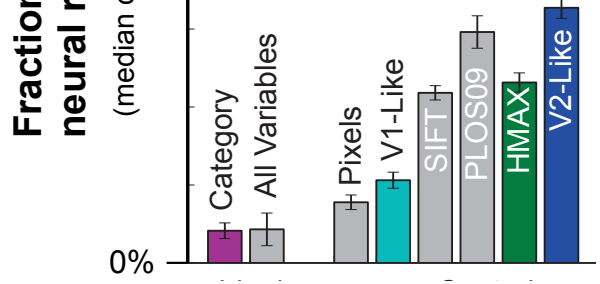
For us, "explained" means predicted for new images

Fraction of visually driven neural response explained

(median over all recorded neurons)

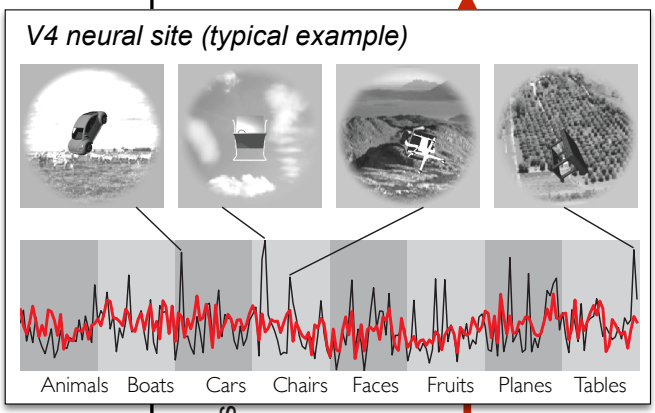
Cortical area V4

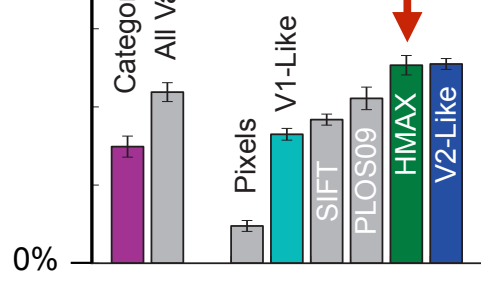
Failure!



100%

V4 neural site (typical example)

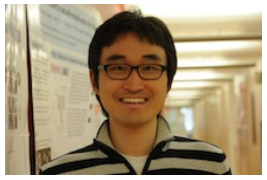




~2013: Breakthrough



Dan Yamins

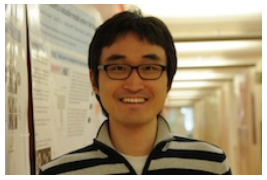


Ha Hong

Yamins, Hong, Solomon, Seibert and DiCarlo ***NIPS (2013), PNAS (2014)***

The underlying problem: most of the parameters of these hypotheses (models) are not determined by existing science results !

~2013: Breakthrough

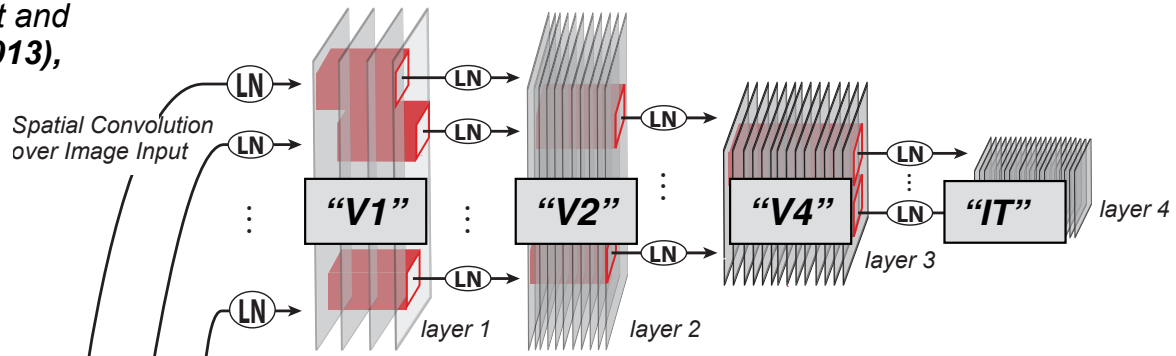
Dan Yamins ***Ha Hong***

Cognitive science guided the task

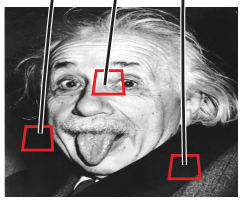
The underlying problem: most of the parameters of these hypotheses (models) are not determined by existing science results !

Yamins, Hong, Solomon, Seibert and DiCarlo ***NIPS (2013), PNAS (2014)***

Artificial neural network (ANN)

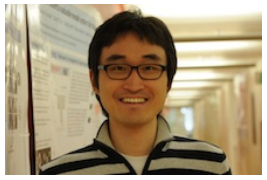


Neuroscience guided the choice of system architecture



Example test image (one of many)

~2013: Breakthrough

Dan Yamins ***Ha Hong***

Yamins, Hong, Solomon, Seibert and DiCarlo ***NIPS (2013), PNAS (2014)***

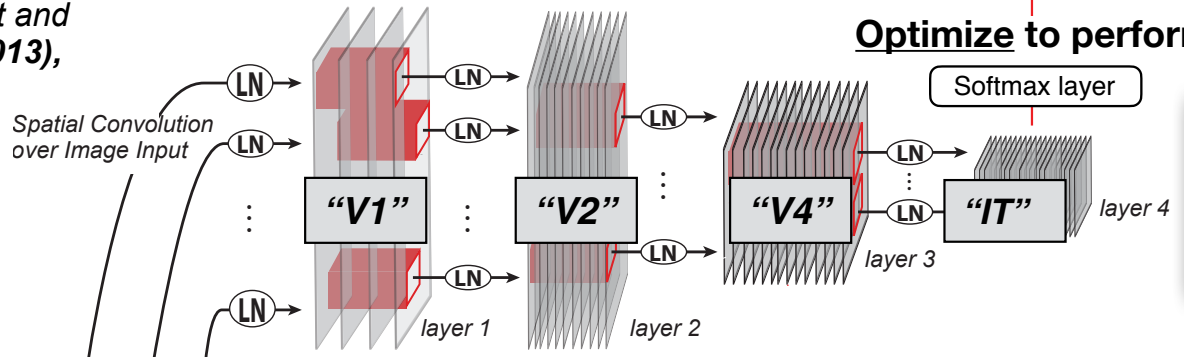
Cognitive science guided the task

Core object recognition



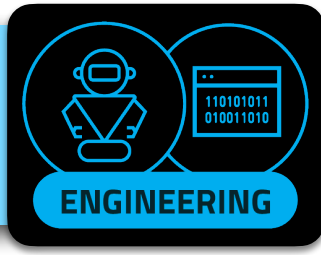
Optimize to perform this task!

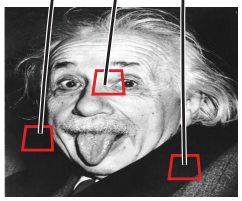
Artificial neural network (ANN)



Neuroscience guided the choice of system architecture

Engineering tools to tune the neural network parameters to perform well !



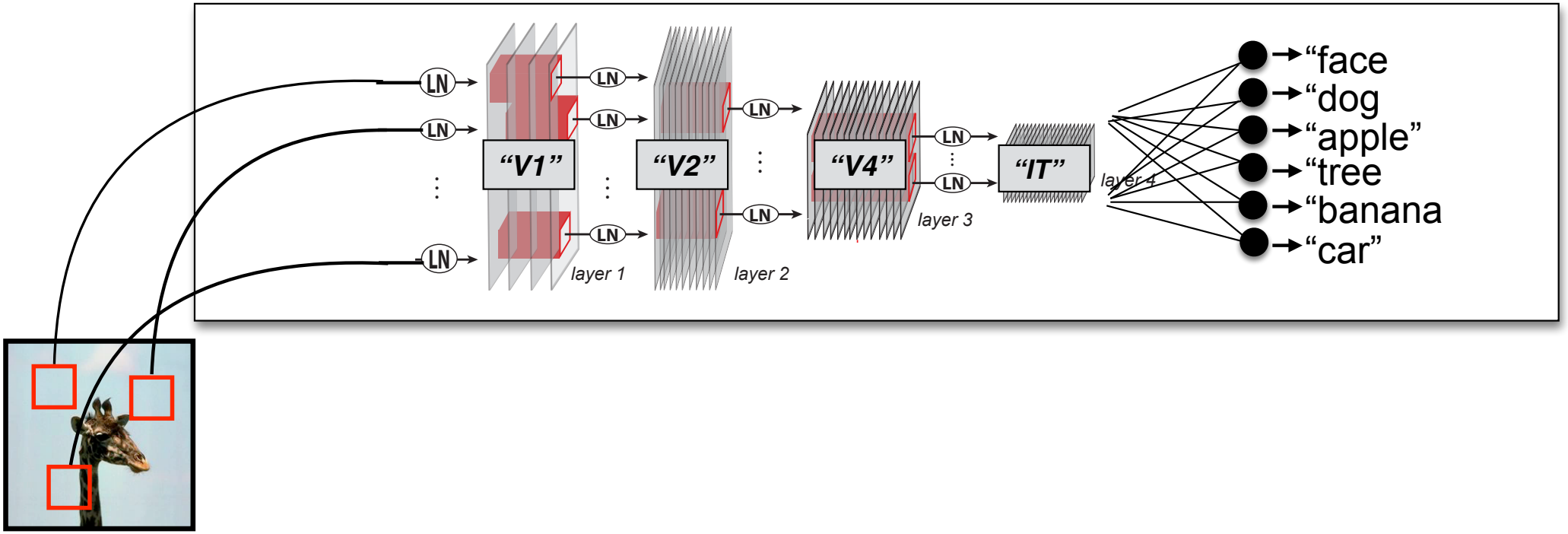


Example test image (one of many)

Discovery: the individual "neurons" inside these new models were now behaving very much like the individual neurons that we and others were recording in the monkey brain !

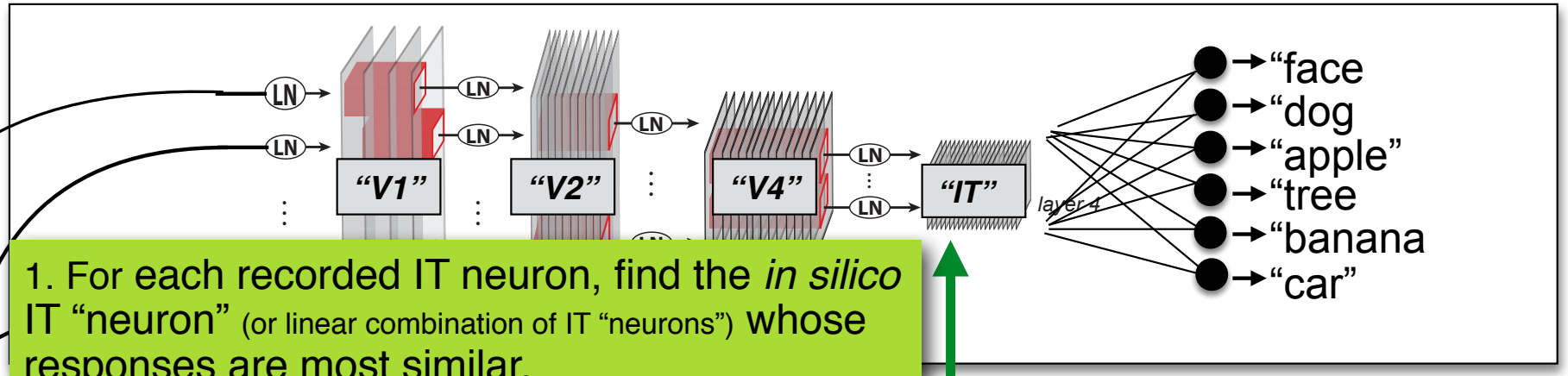
*Yamins, Hong, Solomon, Seibert and DiCarlo **NIPS (2013), PNAS (2014**)*

A specific deep ANN ("evolved" to perform core recognition)



*Yamins, Hong, Solomon, Seibert and DiCarlo **NIPS (2013), PNAS (2014)***

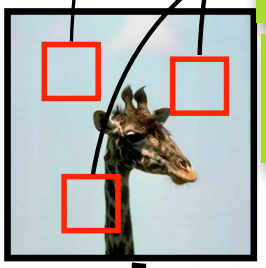
A specific deep ANN ("evolved" to perform core recognition)



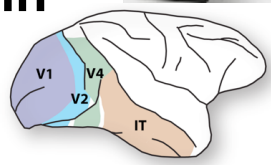
1. For each recorded IT neuron, find the *in silico* IT "neuron" (or linear combination of IT "neurons") whose responses are most similar.

2. Ask: how well does that in silico "neuron" explain (i.e. **predict**) the IT neuron's responses?

Can compare!



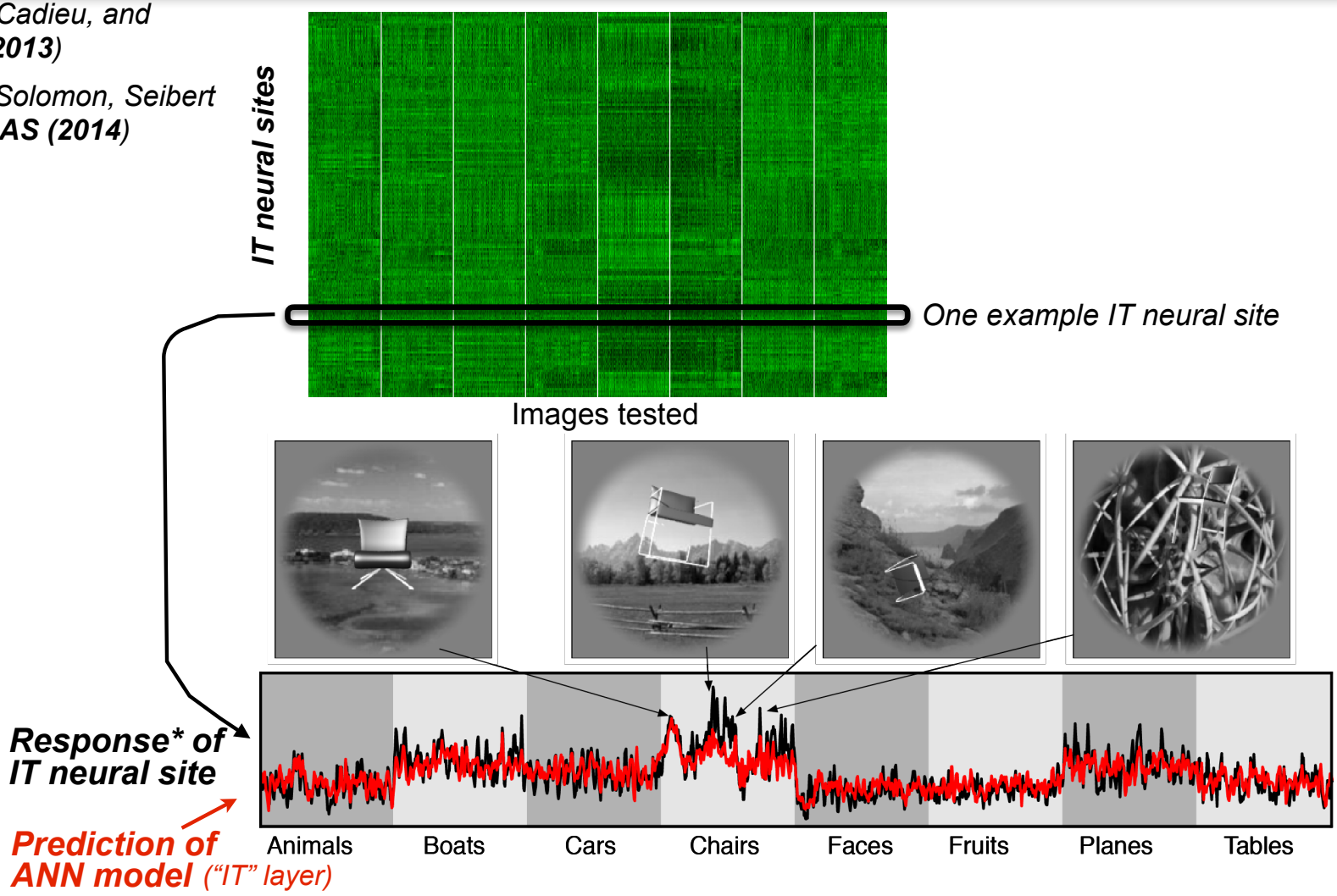
Brain



Result: the evolved model "hidden neurons" ~ match the brain's "hidden neurons"

Yamins, Hong, Cadieu, and DiCarlo ***NIPS (2013)***

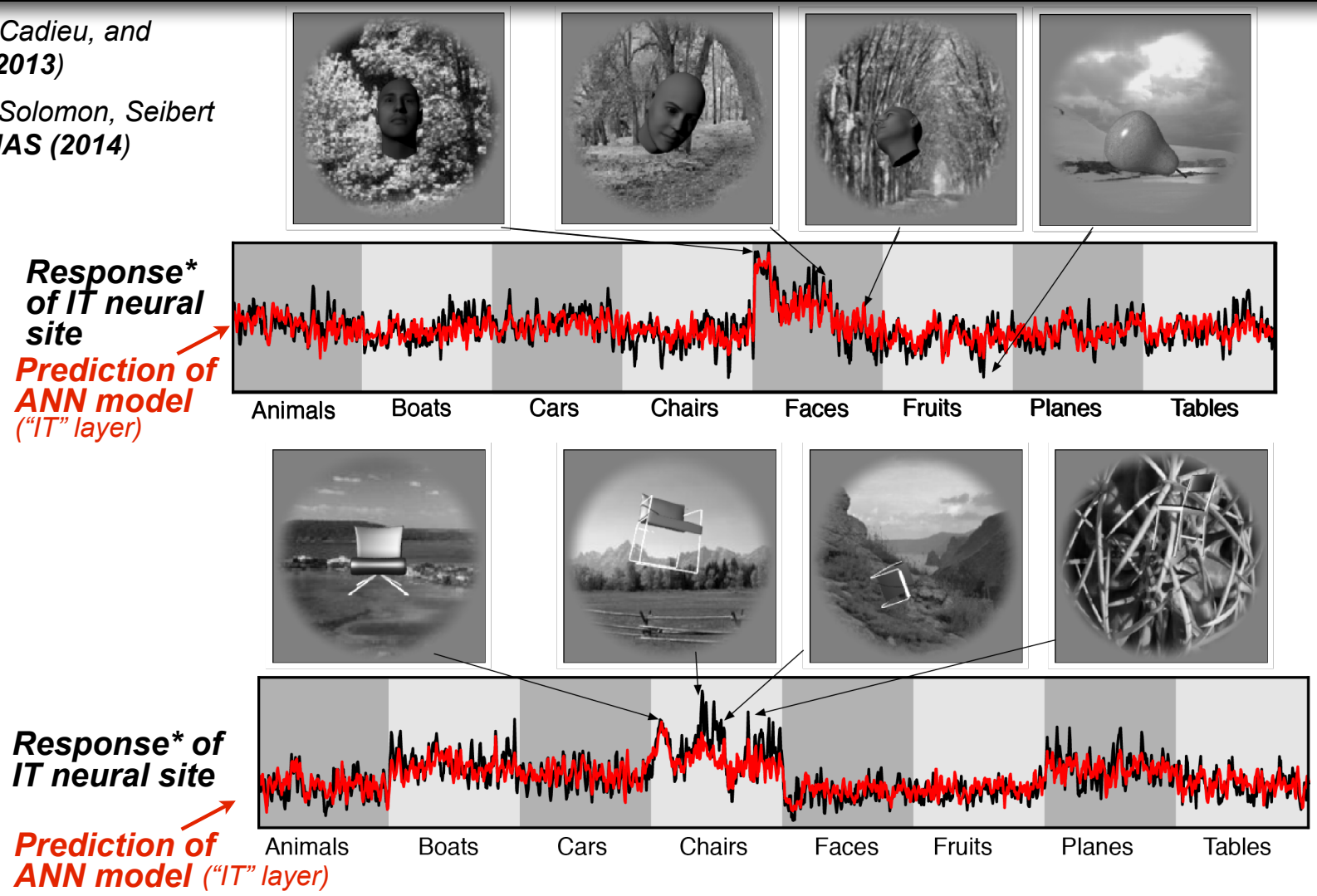
Yamins, Hong, Solomon, Seibert and DiCarlo ***PNAS (2014)***

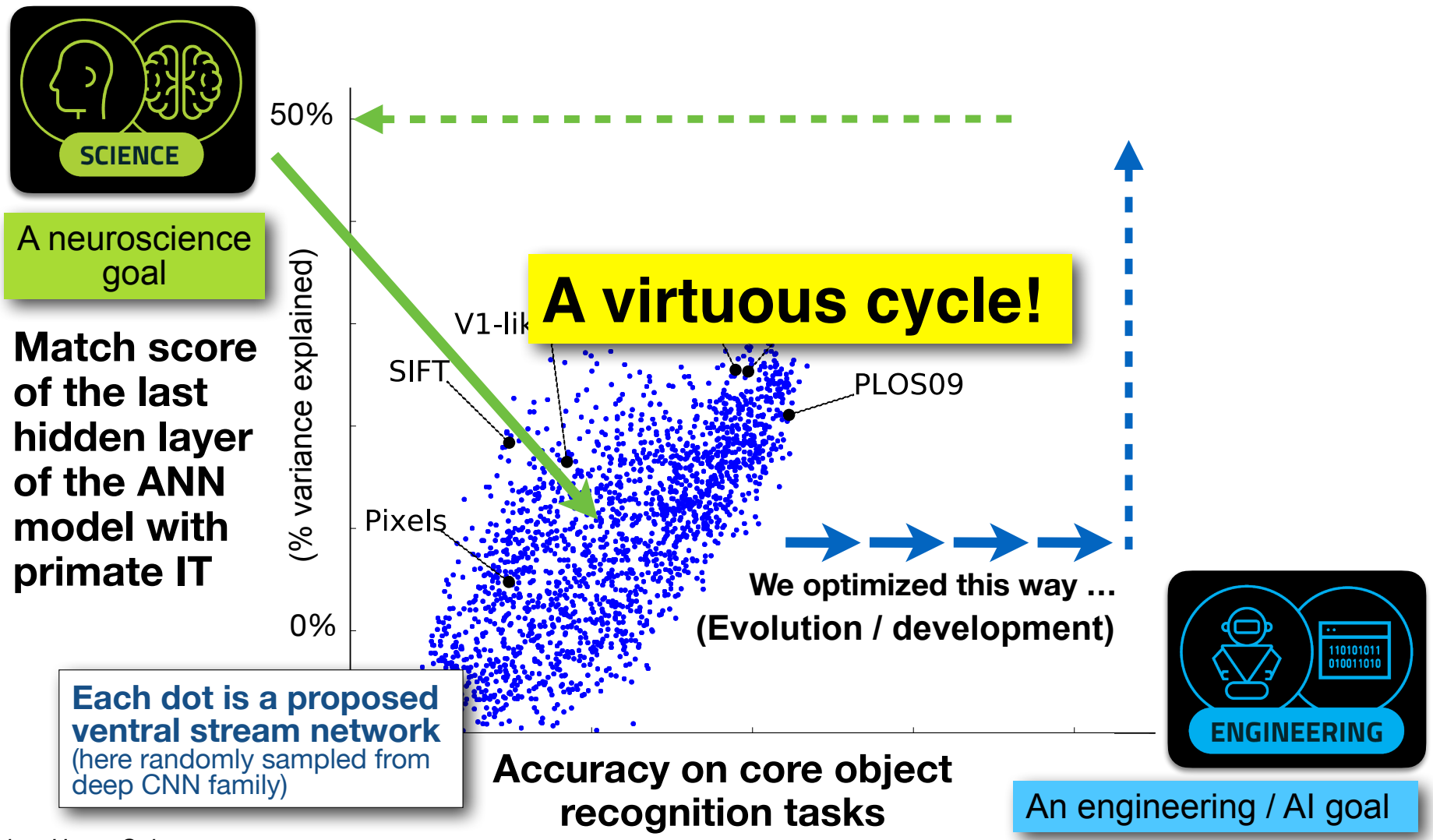


Result: the evolved model "hidden neurons" ~ match the brain's "hidden neurons"

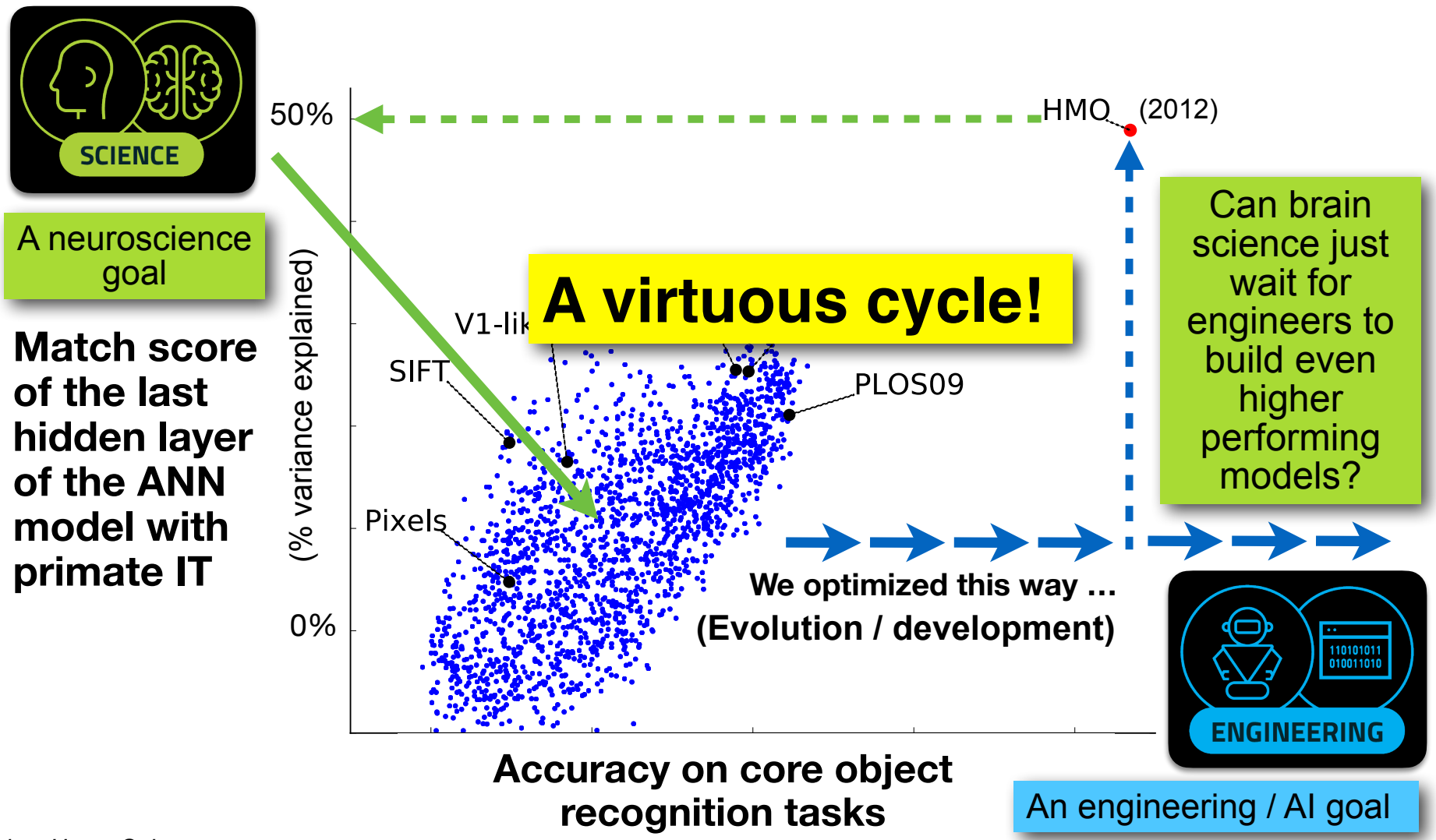
Yamins, Hong, Cadieu, and DiCarlo ***NIPS (2013)***

Yamins, Hong, Solomon, Seibert and DiCarlo ***PNAS (2014)***





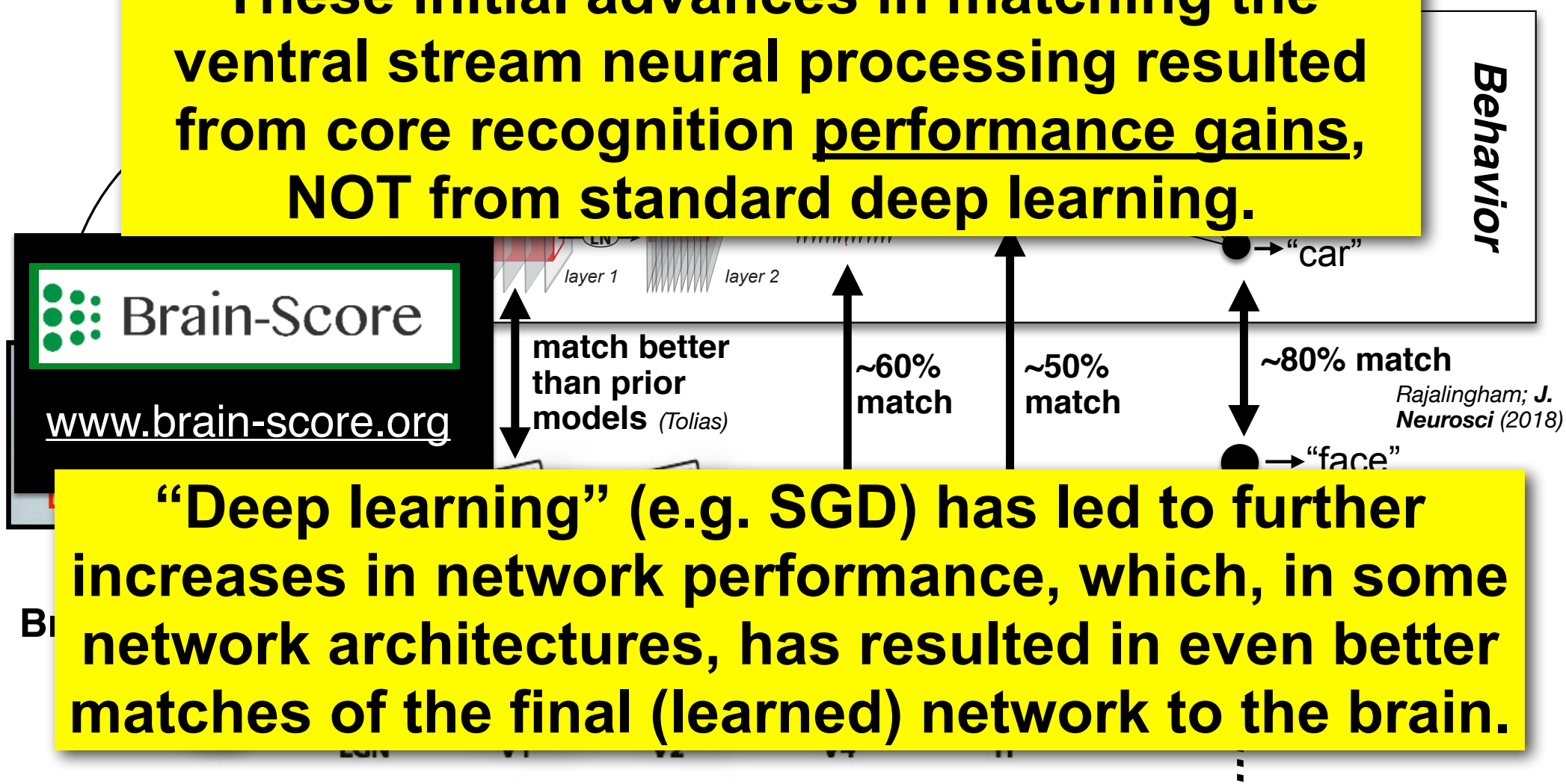
*Adapted from Yamins, Hong, Solomon, Seibert and DiCarlo **PNAS (2014)***



Adapted from Yamins, Hong, Solomon, Seibert and DiCarlo ***PNAS (2014)***

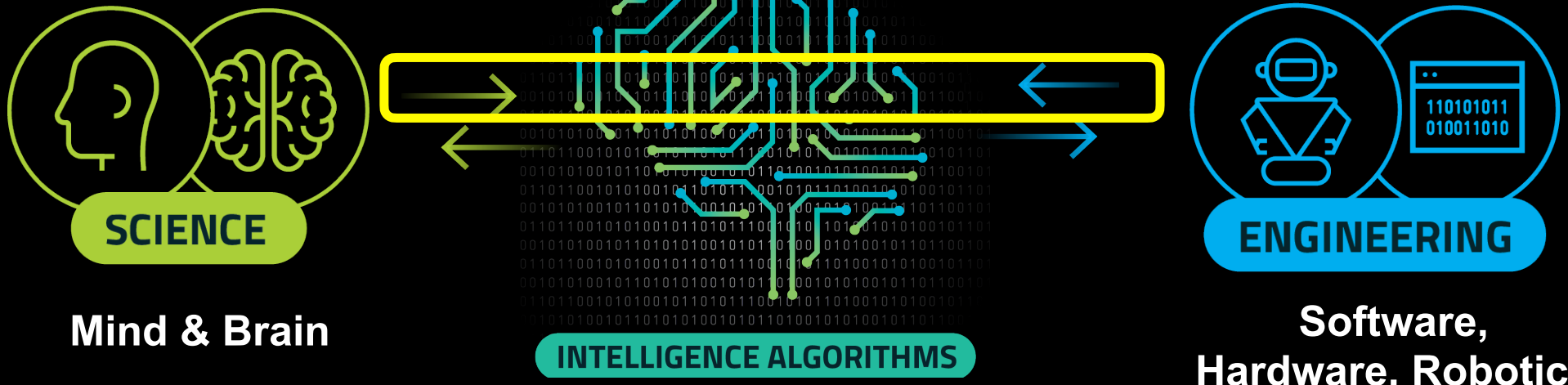
A typical"evolved" artificial ventral visual stream

These initial advances in matching the ventral stream neural processing resulted from core recognition performance gains, NOT from standard deep learning.



"Deep learning" (e.g. SGD) has led to further increases in network performance, which, in some network architectures, has resulted in even better matches of the final (learned) network to the brain.

An implicit collaboration!



Mind & Brain

In silico **neural networks** (“ANNs”)

Software, Hardware, Robotics

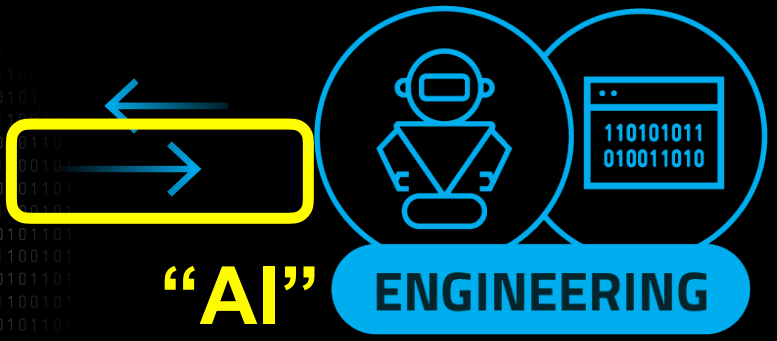
An implicit collaboration!
To the benefit of both fields

The New York Times

Scientists See Promise in Deep-Learning Programs



Using an artificial intelligence technique inspired by theories about how the brain recognizes patterns, technology companies are reporting startling gains in fields as diverse as computer vision, speech recognition and the identification of promising new molecules for designing drugs.



"AI"

ENGINEERING

Software,
Hardware, Robotics

HMS

orks ("ANNs")



Pouya Bashivan



Kohitij Kar

RESEARCH ARTICLE SUMMARY

SCIENCE

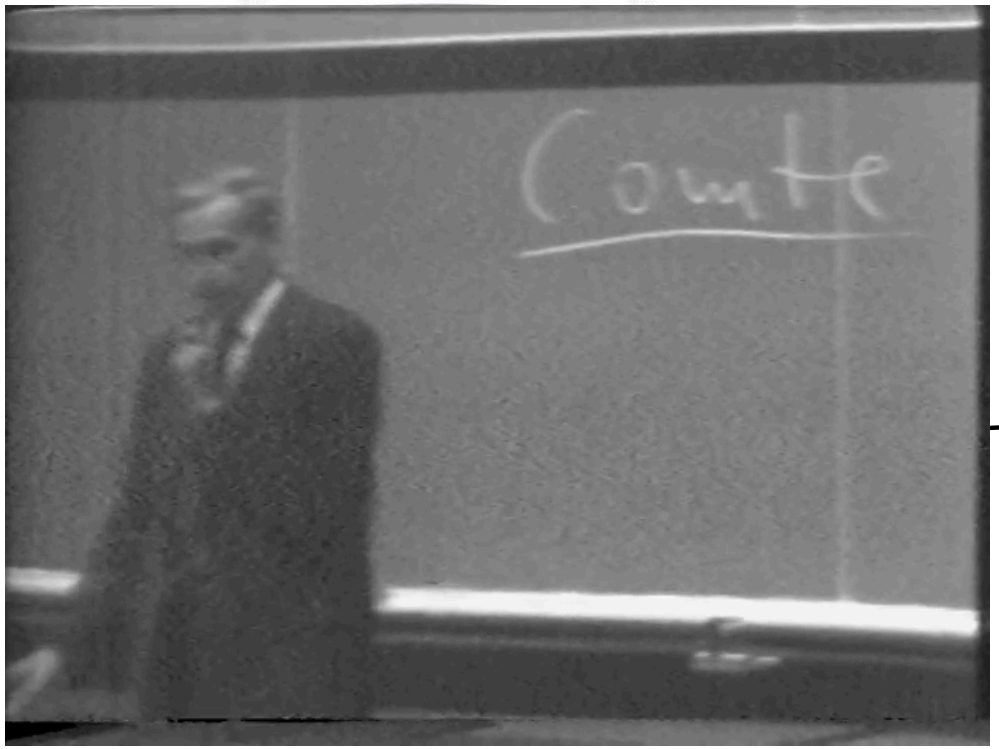
2019

NEUROSCIENCE

Neural population control via deep image synthesis

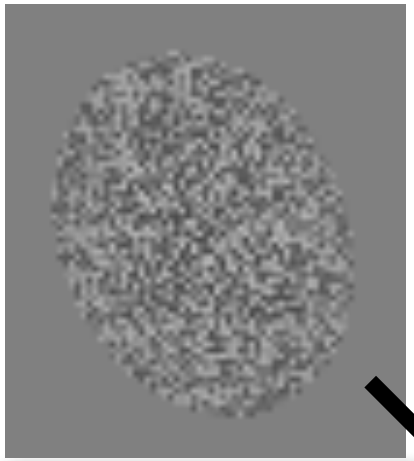
Pouya Bashivan*, Kohitij Kar, James J. DiCarlo

Hans-Lukas Teuber ~1974



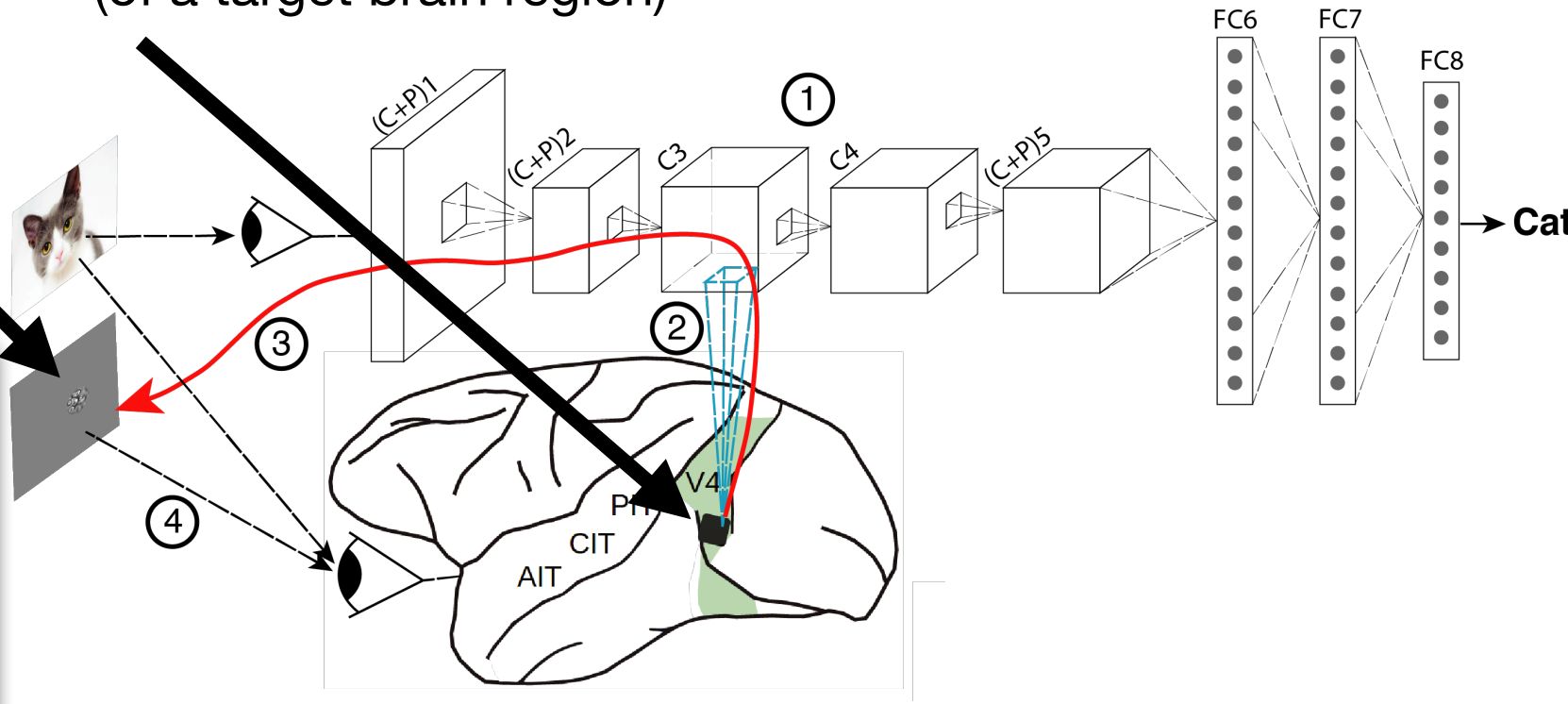
Are we [illegible] new "understanding"?

Particular ANNs can now reasonably accurately **explain** / **predict** the (spiking-level) workings of the ventral visual stream



We used our in silico ventral stream models to design precise patterns of light energy on the eyes to try to "set" different internal states of the brain.

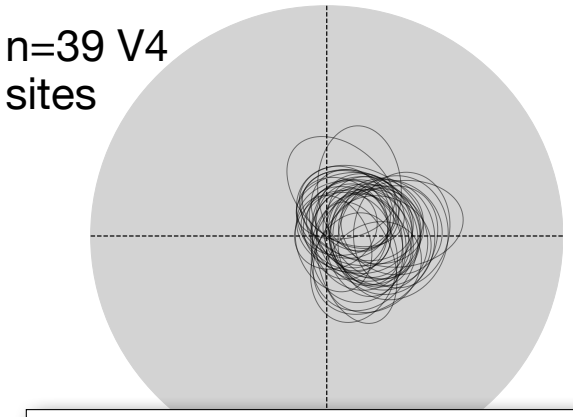
A control goal: set a desired neural state
(of a target brain region)



We demonstrated that we could use these non-invasive "digital treatments" to successfully set neural activity states deep in the brain !

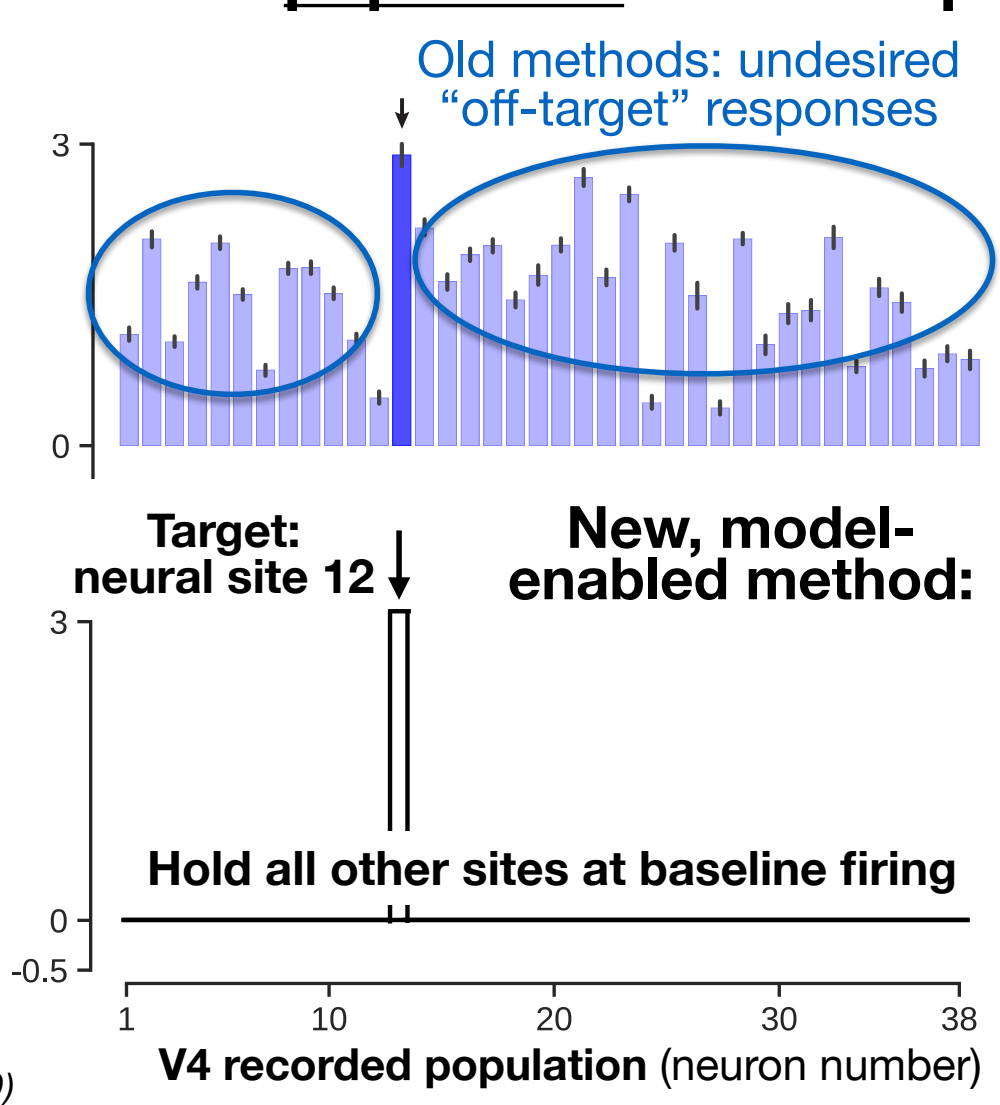
Bashivan, Kar and DiCarlo ***CCN*** *(2018),* ***Science*** *364 (2019)*

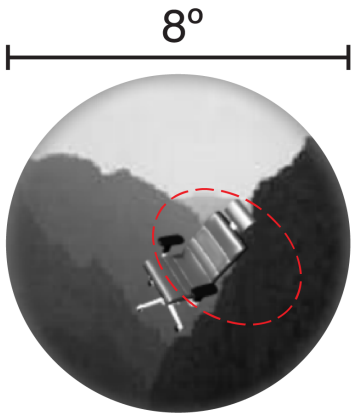
Application: control the neural population state deep in the brain

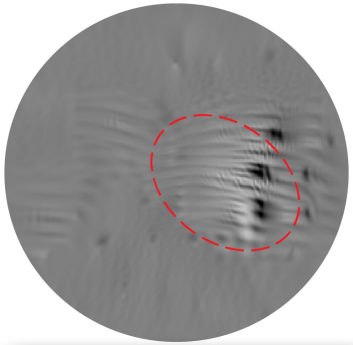


All respond to stimuli in the same part of the visual field

Example goal:
drive one target neural site to high activity, while driving all other neural sites low.







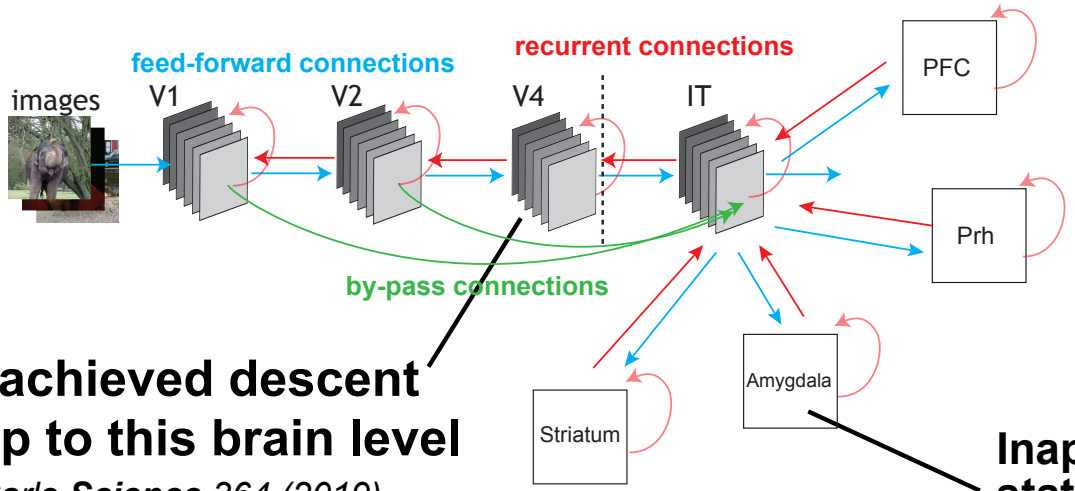
synthetic "controller image"

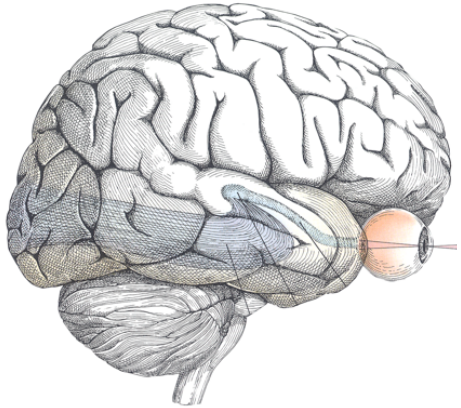
Bashivan, Kar and DiCarlo ***Science*** *364 (2019)*

A new application superpower for scientists?:

The ability to control patterns of neural activity deep in the brain by precisely designing the visual input (images, movies)

Images, movies, or video games





So far, we have achieved descent neural control up to this brain level

Bashivan, Kar and DiCarlo ***Science*** *364 (2019)*

Inappropriate neural states in this part of the brain implicated in mood disorders

As we further improve our models, this superpower will further improve!

Health applications?

Another key CBMM-supported finding:



Kohitij Kar

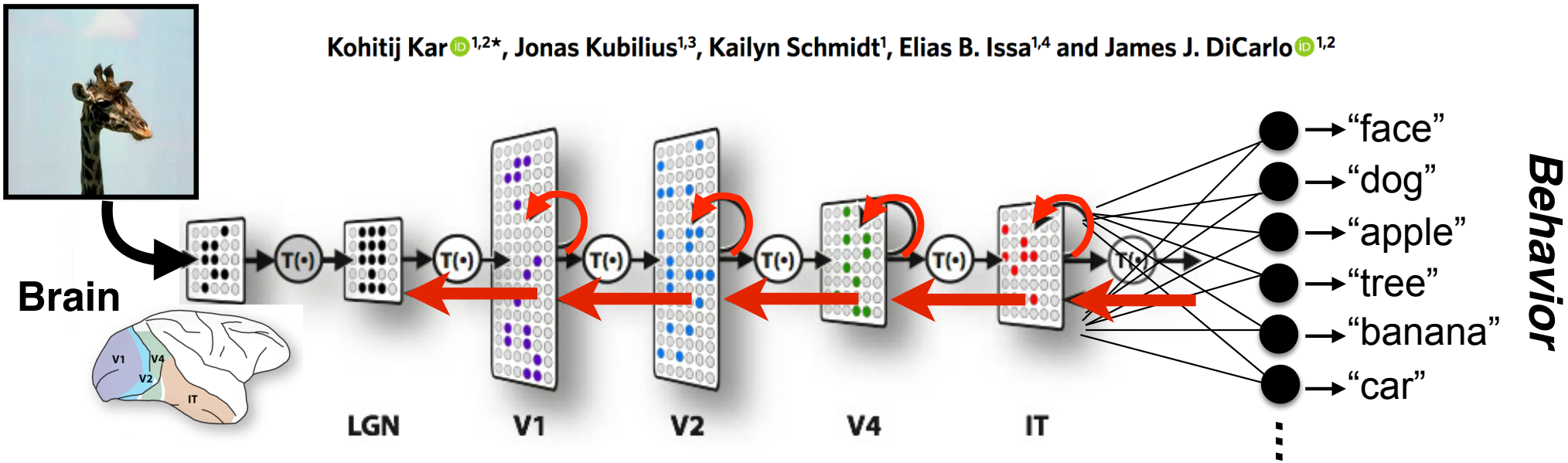
nature neuroscience

ARTICLES

https://doi.org/10.1038/s41593-019-0392-5

Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior

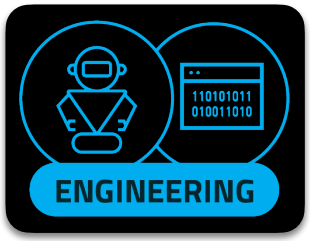
Kohitij Kar[1,2*], Jonas Kubilius[1,3], Kailyn Schmidt[1], Elias B. Issa[1,4] and James J. DiCarlo[1,2]



Ongoing…

Our goal is to discover even more accurate computer models of the human visual system.





break current models

Use those breaks to build better models

Brain-Score

brain-score.org

Current best ventral stream models

AI

Health

Education

…

Summary take home messages

1. **Background: The ventral visual stream produces an IT neural population representation that carries linearly decodable, image generalizable solutions for all (tested) core object recognition tasks.**

2. **Optimizing deep artificial neural network (ANN) architectures for core recognition tasks leads to internal neural representations in those ANNs that are remarkably similar to the internal neural representations of the ventral visual stream.** (*NIPS* 2013, *PNAS* 2014)

   - **This result (above) includes IT "face neurons".**
   - **This result (above) is consistent with, but does <u>not</u> imply, that the brain learns by classical backpropagation. My preferred interpretation is that this is a demonstration of convergent evolution/development**
     Species A: primate vision optimized by evolution / post-natal development.
     Species B: deep ANNs optimized by engineers.

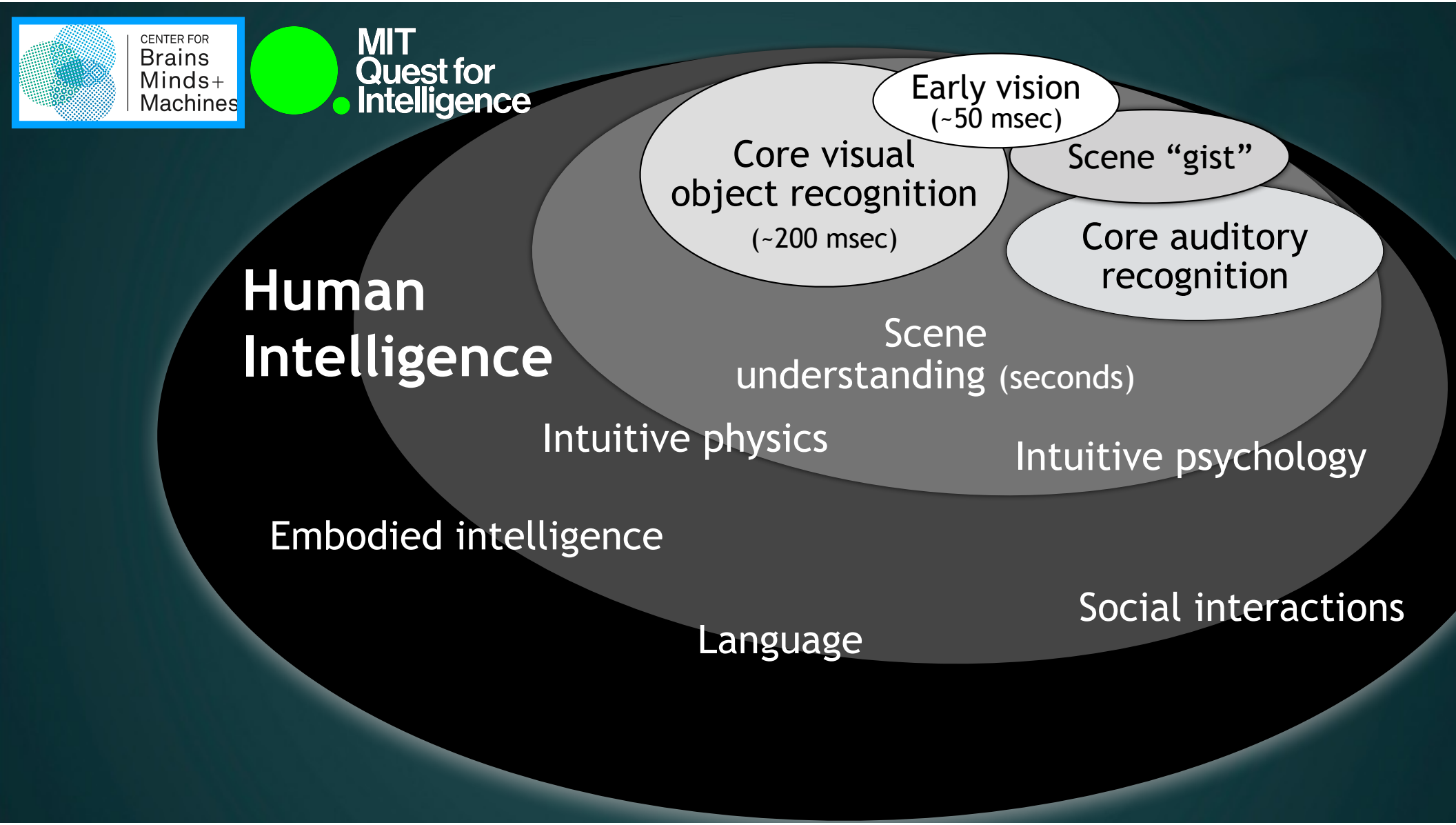
Summary take home messages

3. **These same (optimized) ANN models can be used to guide the construction of novel synthetic images to super-activate ventral stream neurons and control sub-populations of neurons** (*CCN*, 2018; *Science*, 2019)

4. **Nevertheless, these same (optimized) ANNs are <u>not</u> yet functionally <u>identical</u> to the ventral visual stream.** (*e.g. J Neuroscience*, 2018; "Brain-Score" *bioRxiv* 2018)

5. **One difference is the lack of recurrent circuits, and recent IT neurophysiology suggests that fast-acting, automatically-evoked recurrent circuits enable the ventral stream's superior performance on many images** (*Nature Neuro.*, 2019).

6. **We and our collaborators are building a series of new models that incorporate more biological constraints — thus far, these model show computer vision gains in efficiency (depth) and gains in robustness to image perturbation.**


CENTER FOR
Brains
Minds+
Machines
MIT
Quest for
Intelligence
Human
Intelligence
Early vision
(~50 msec)
Core visual
object recognition
(~200 msec)
Scene "gist"
Core auditory
recognition
Scene
understanding (seconds)
Intuitive physics
Intuitive psychology
Embodied intelligence
Social interactions
Language

CENTER FOR
Brains
Minds+
Machines

MIT
Quest for
Intelligence



dicarlo@mit.edu