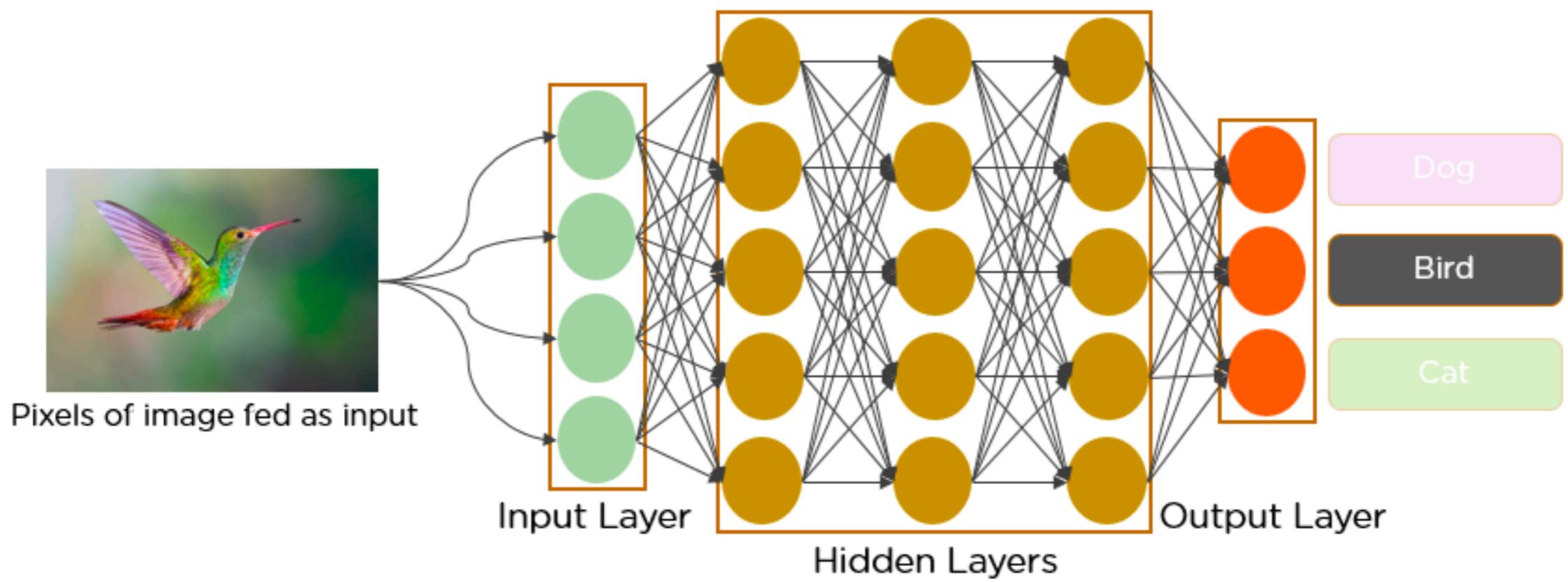


Multilayer Perceptron and deep neural network

Deep learning and back-propagation algorithm



The recent excitement about neural networks

Francis Crick

The remarkable properties of some recent computer algorithms for neural networks seemed to promise a fresh approach to understanding the computational properties of the brain. Unfortunately most of these neural nets are unrealistic in important respects.

THERE has been a lot of excitement recently about neural nets. A new algorithm has produced quite simple nets that perform surprisingly well. A thick two-volumed work, *Parallel Distributed Processing*¹, has been a best-seller, read enthusiastically by psychologists, computer designers and physicists. Even undergraduates are now designing new networks. The interested spectator may well wonder what it's all about. What are neural nets? How do they work? And what, if anything, do they tell us about the brain?

Neural nets are composed of 'units' that have some of the properties of real neurons². That is, each unit has many inputs, some of which excite it and some

having rather few inputs and capable of sending a pulse-coded message. Part of each message is the 'address' that indicates where a particular memory can be stored, and part is the information to be stored. The operation of the computer is largely serial.

The brain is different in almost every respect. Neurons are slow, operating in the millisecond time range, and typically have many hundreds or thousands of inputs. Although many of them produce action potentials or 'spikes' whose distribution in time is not completely random, there is no obvious sign of precise pulse-coded messages. Moreover the parts of the brain seem to be highly parallel in their operation⁴. How then can an assembly of

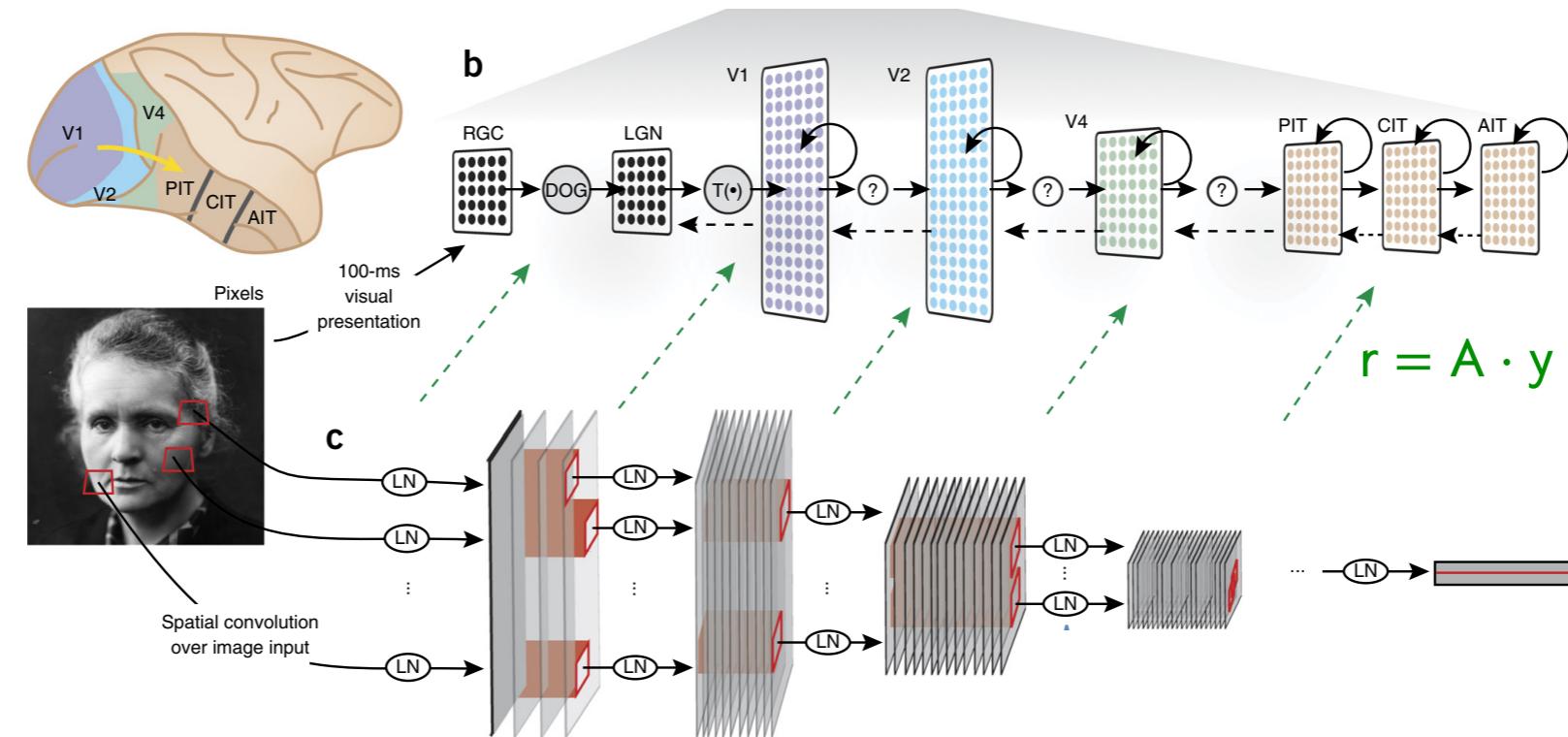
altering a few synapses degrades the performance very little⁵.

Such nets are usually simple, having a single layer of units. Moreover they are usually unsupervised. That is, there is no teacher to tell the net how to adjust its output to make it resemble the desired one. The net learns by using an algorithm based on an idea of Hebb⁶. This is a local algorithm as it depends only on the activity near that particular synapse: roughly, the synapse is strengthened if it receives an input signal on the presynaptic side, together with some indication of activity, such as the unit firing, on the postsynaptic side.

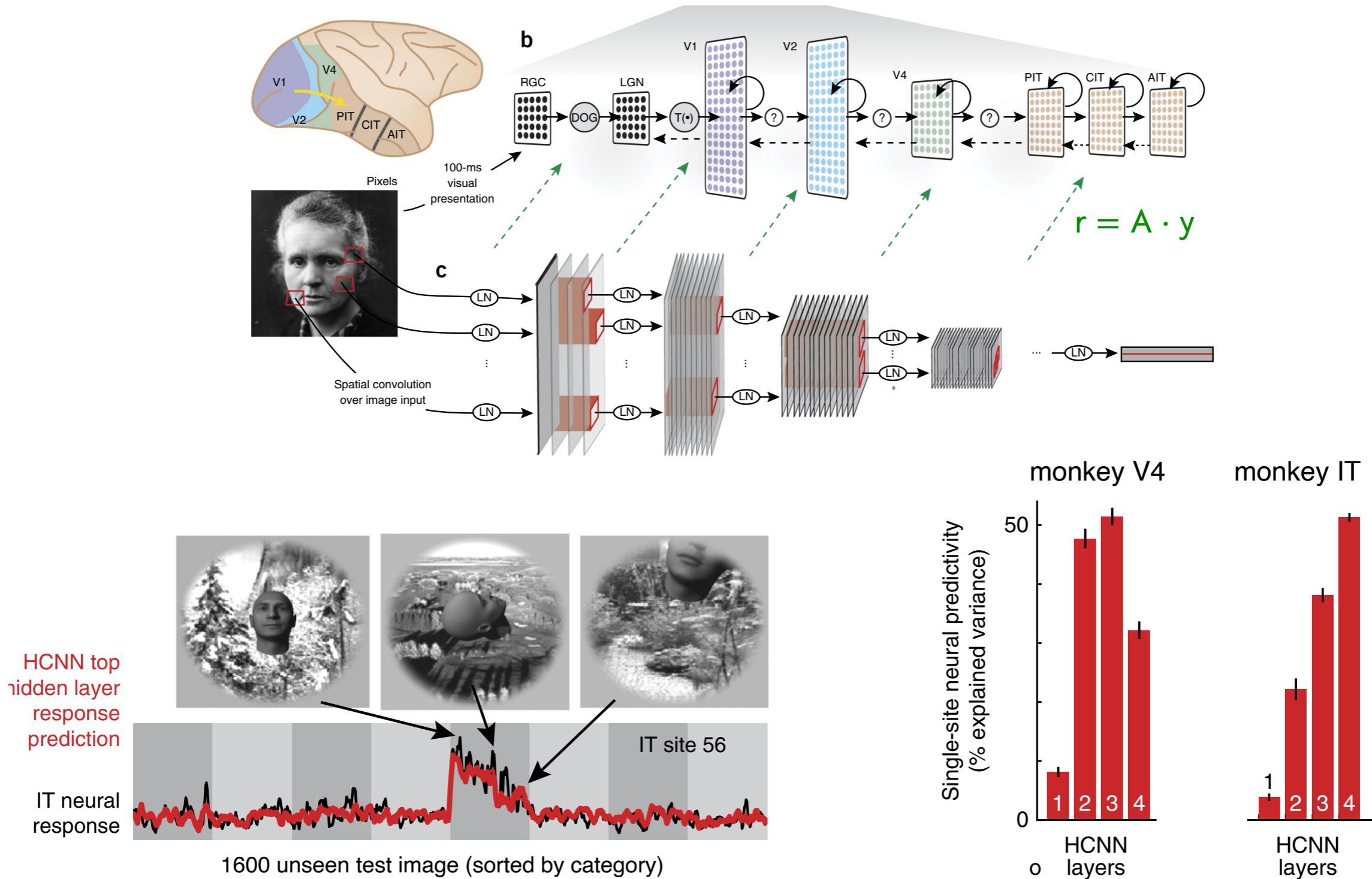
Back propagation

If Francis were alive today, what would he say?

Deep neural network and biological neurons

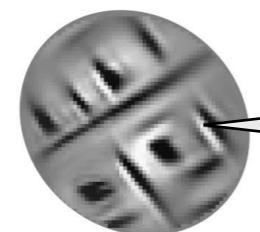


Deep neural network and biological neurons

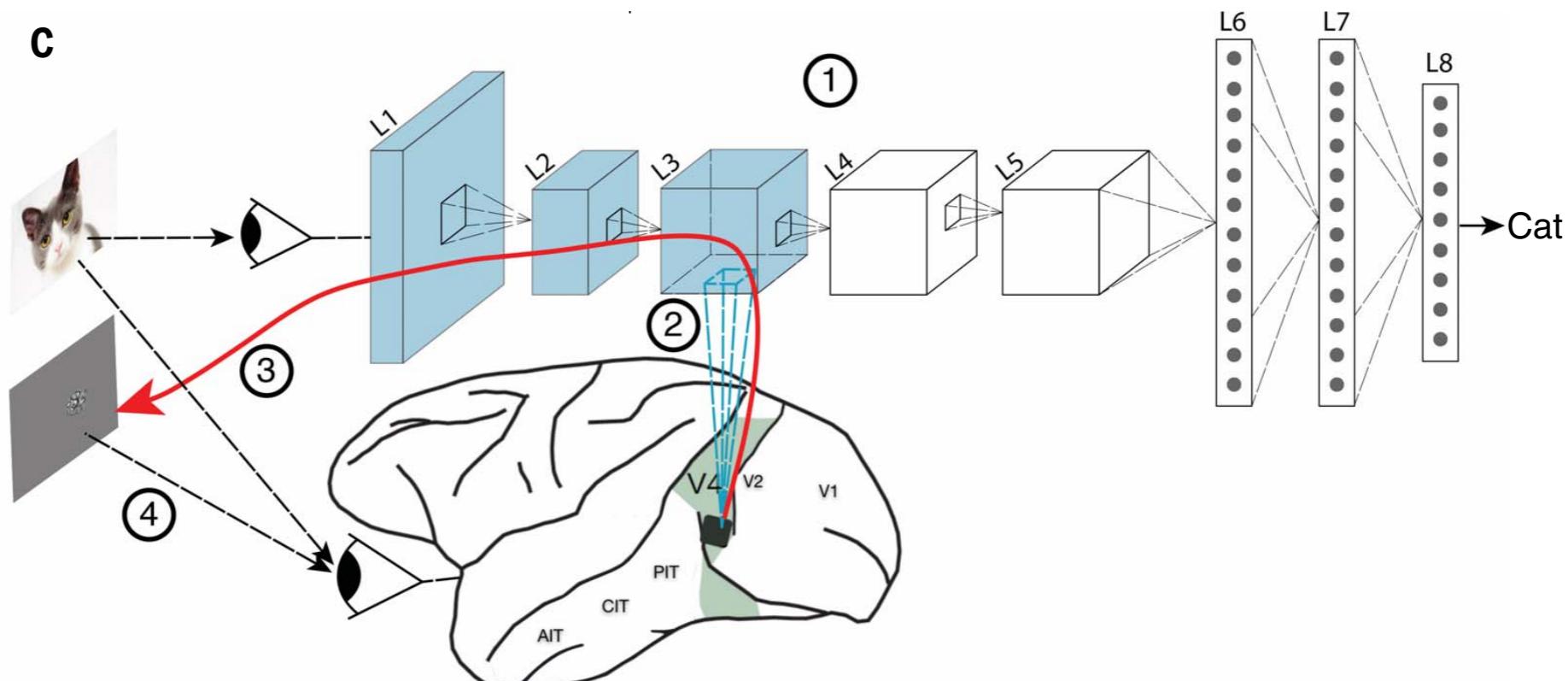


Test deep neural networks on V4 neurons

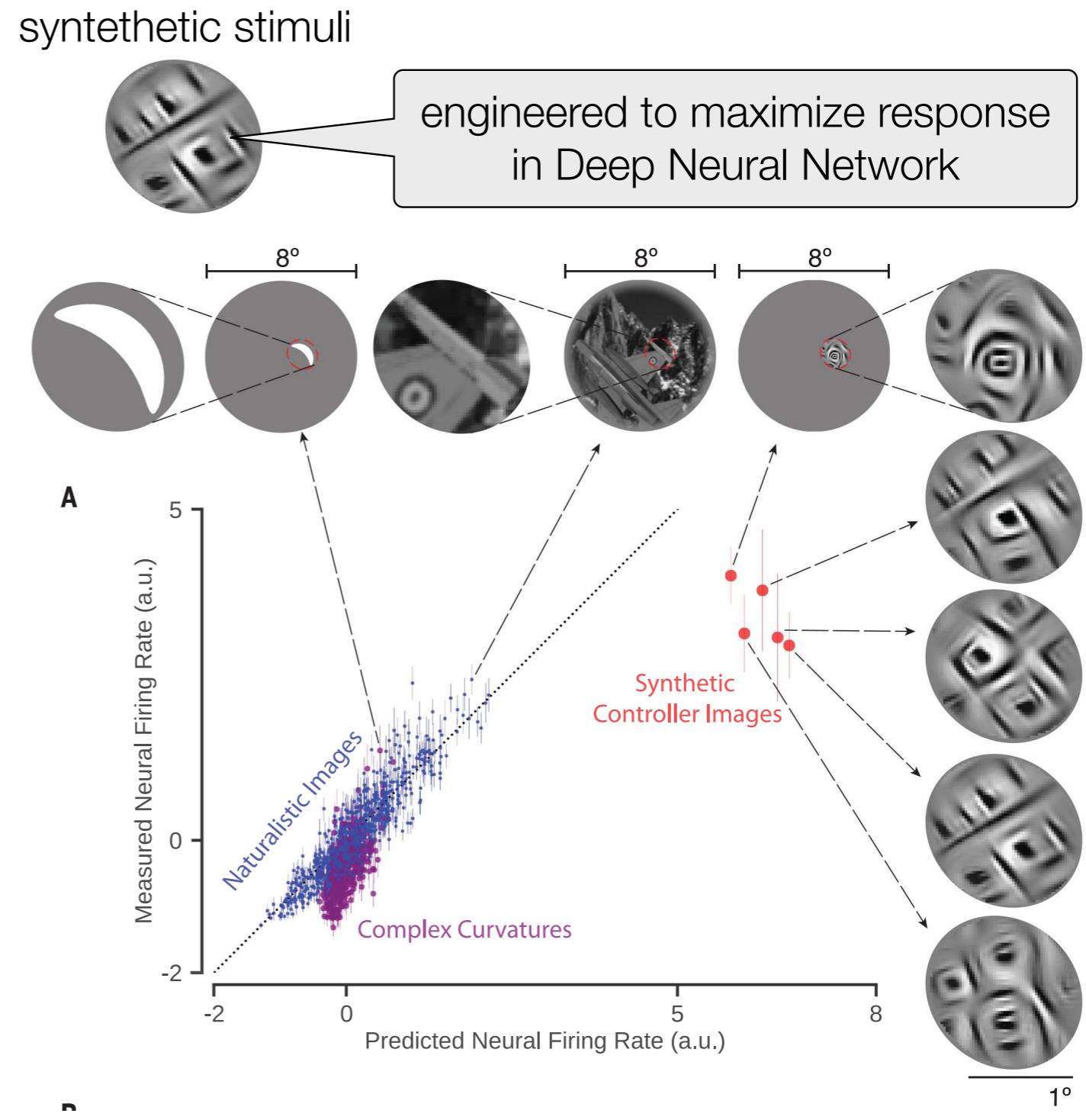
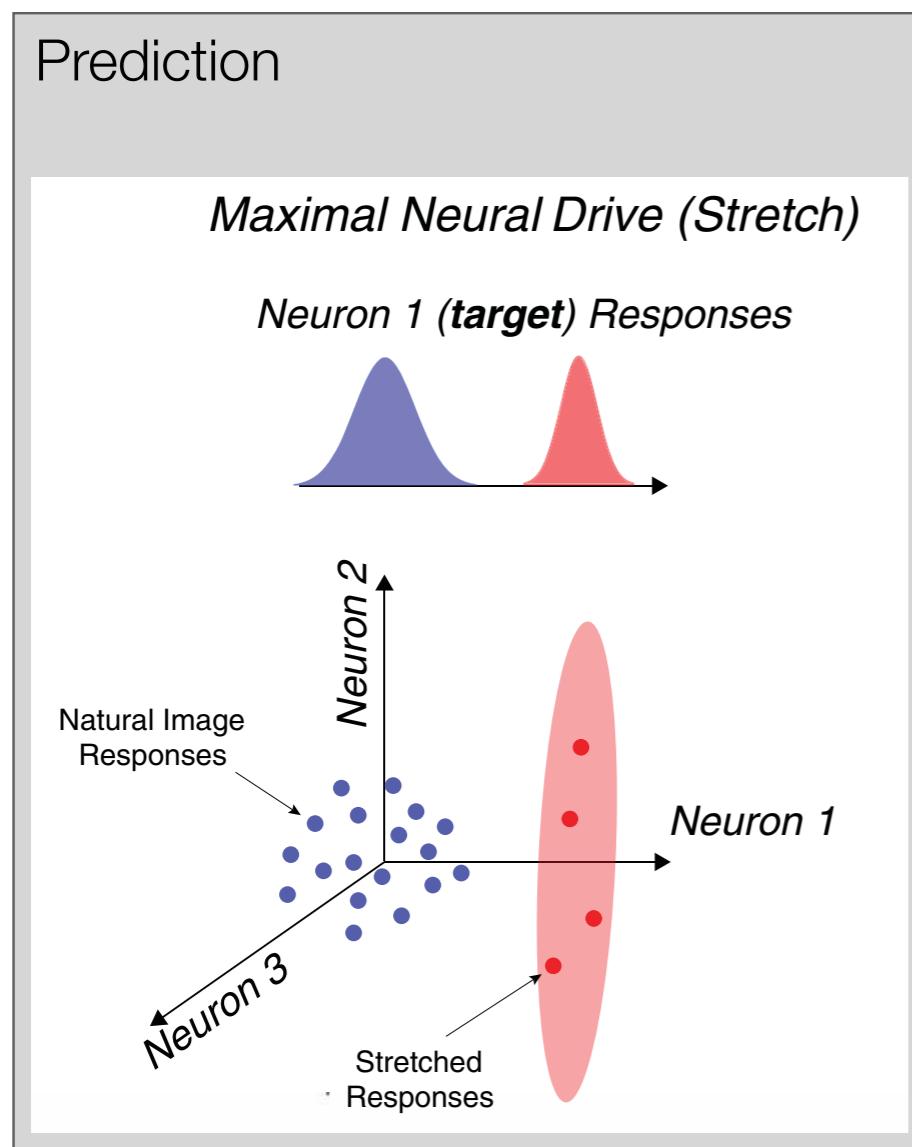
synthetic stimuli



engineered to maximize response
in Deep Neural Network

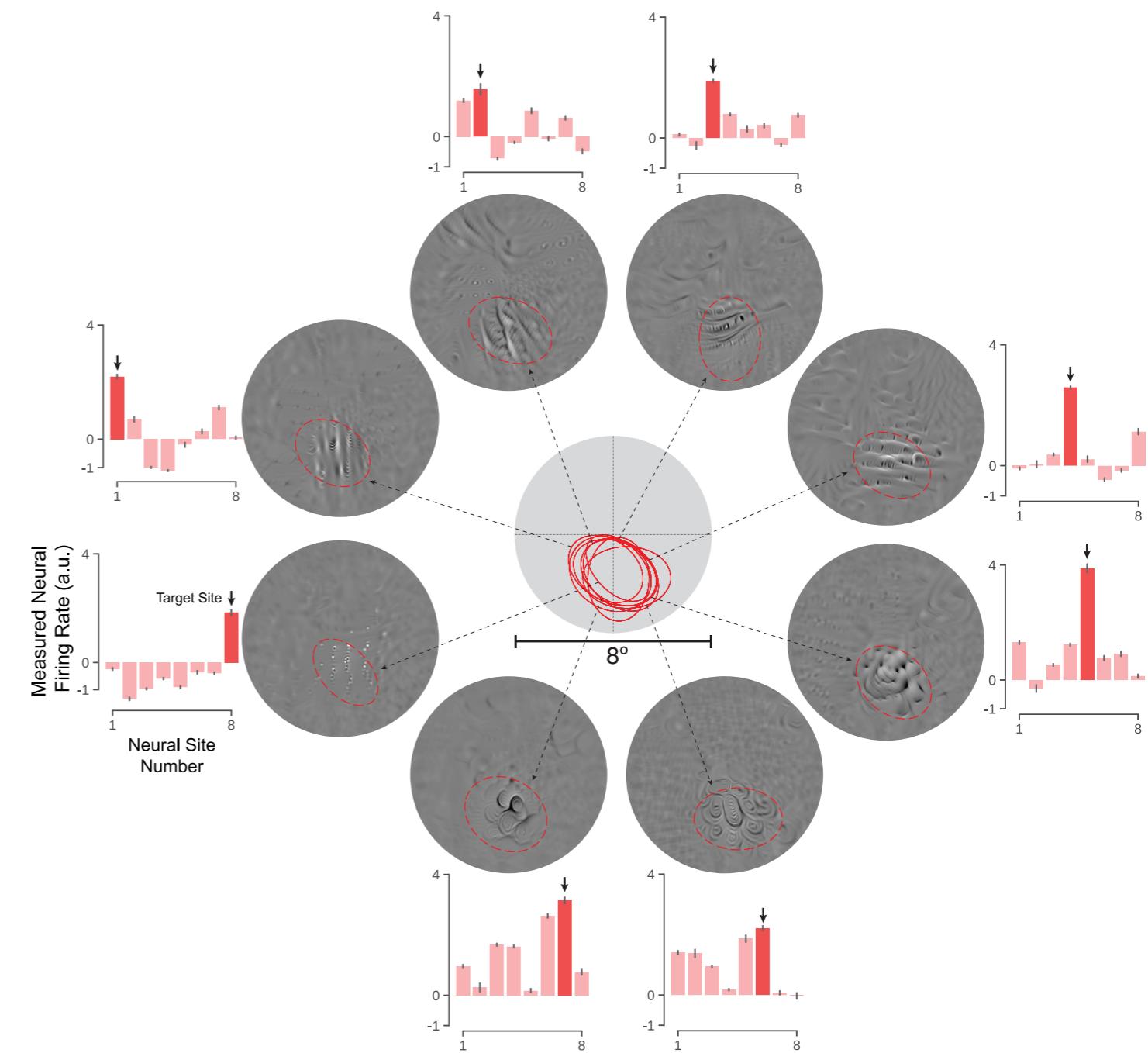
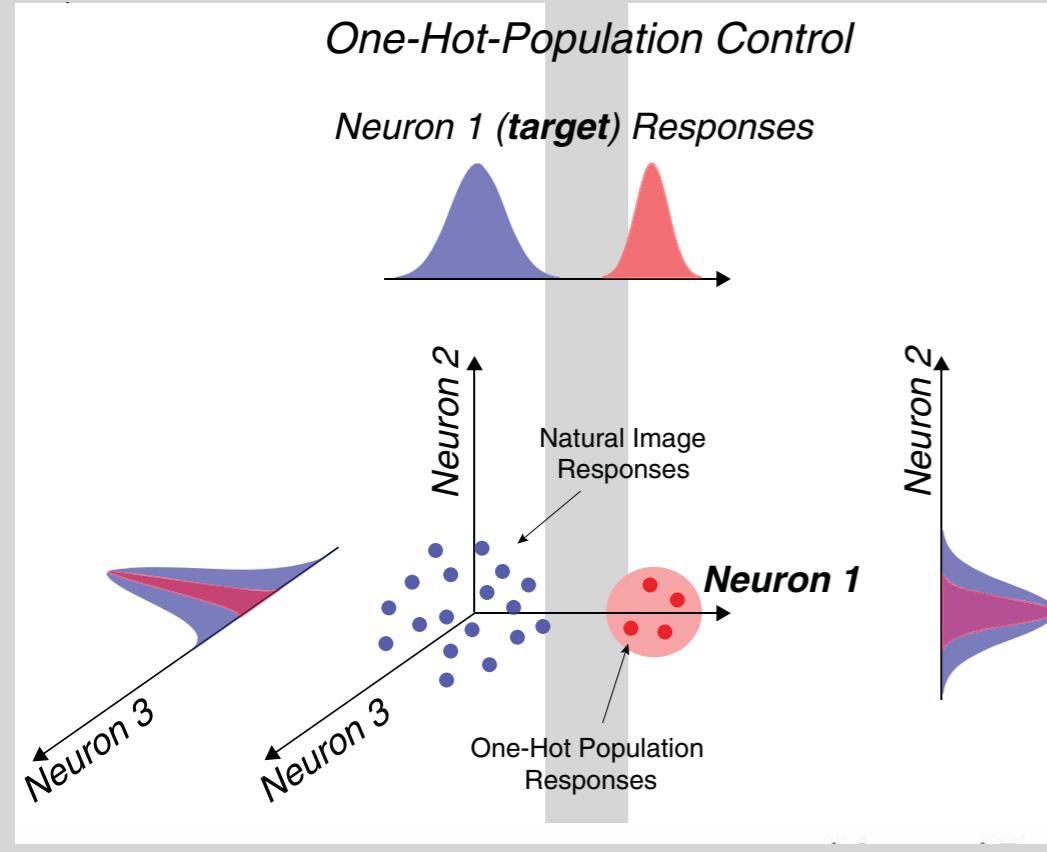


Test deep neural networks on V4 neurons



Test deep neural networks on V4 neurons

Prediction



Adversarial example

Q: ?



A: "schoolbus"

Adversarial example

Q: ?



+



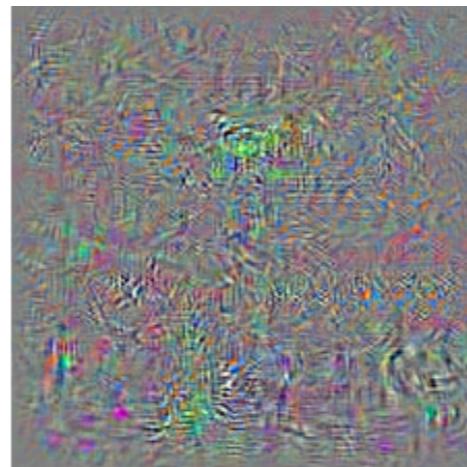
A: “schoolbus”

Adversarial example

Q: ?



+



=



A: “schoolbus”

Q: ?

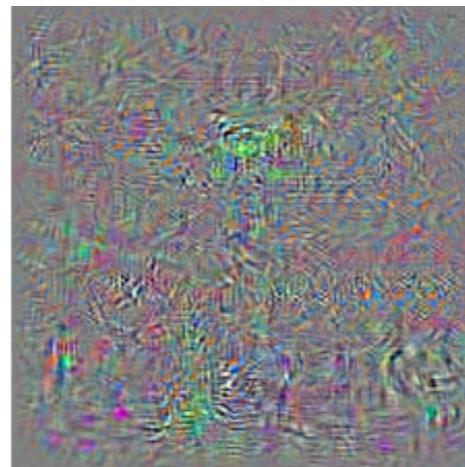
A: “ostrich”

Adversarial example

Q: ?



+



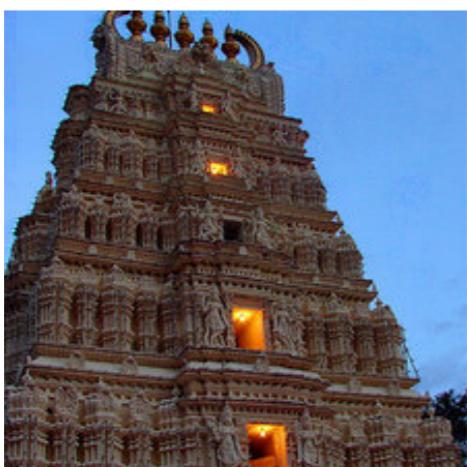
=



Q: ?

A: “schoolbus”

A: “ostrich”



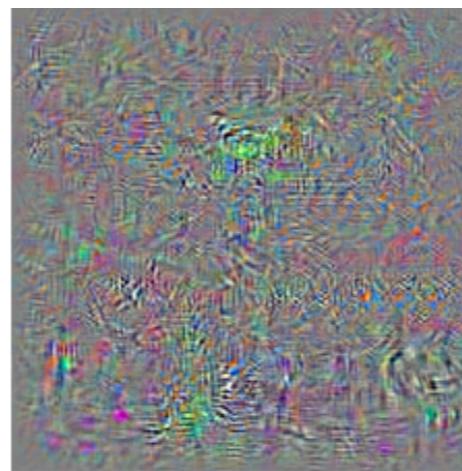
A: “pyramid”

Adversarial example

Q: ?



+



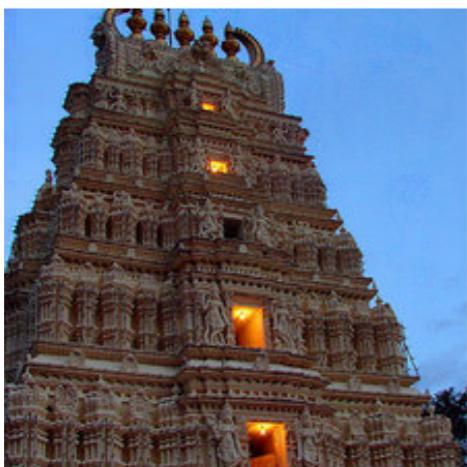
=



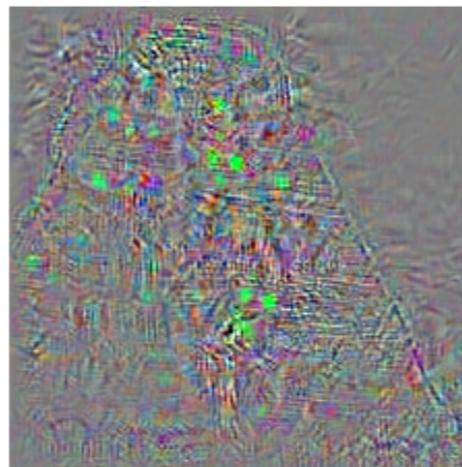
Q: ?

A: “schoolbus”

A: “ostrich”



+



A: “pyramid”

Adversarial example

Q: ?



+

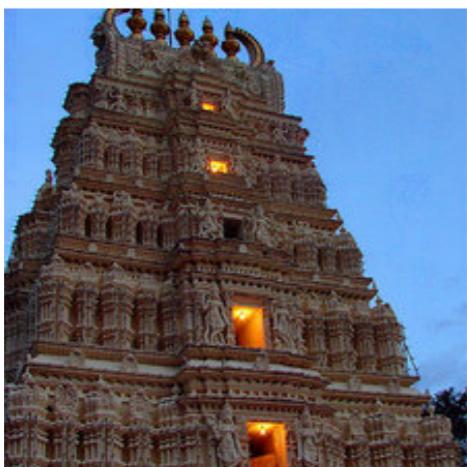


=

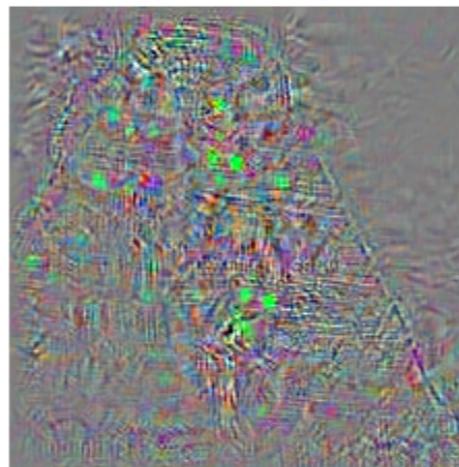


A: “ostrich”

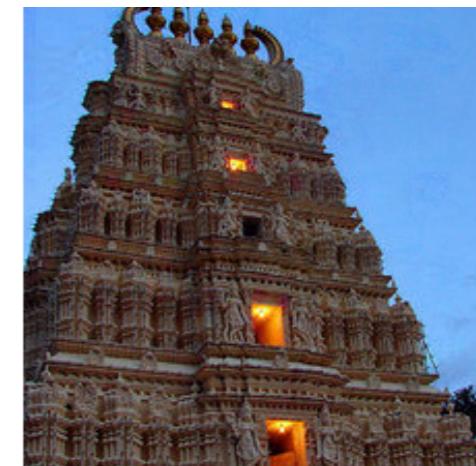
A: “schoolbus”



+



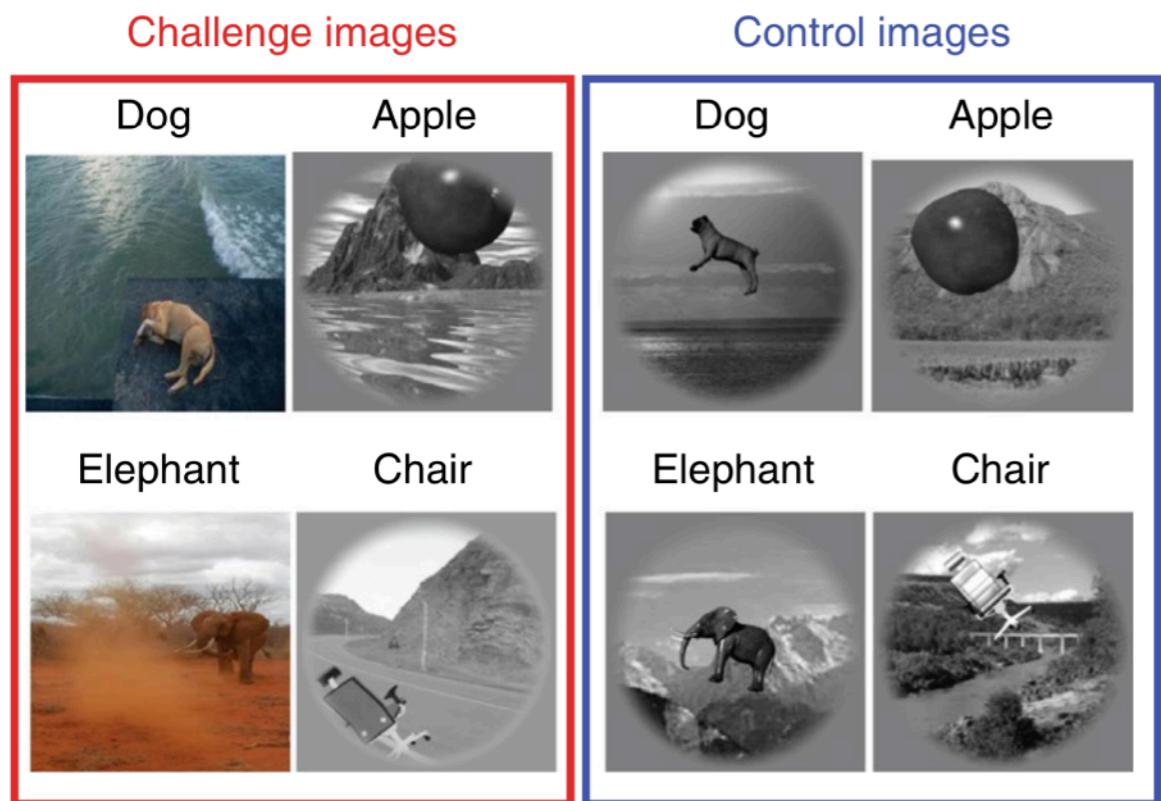
=



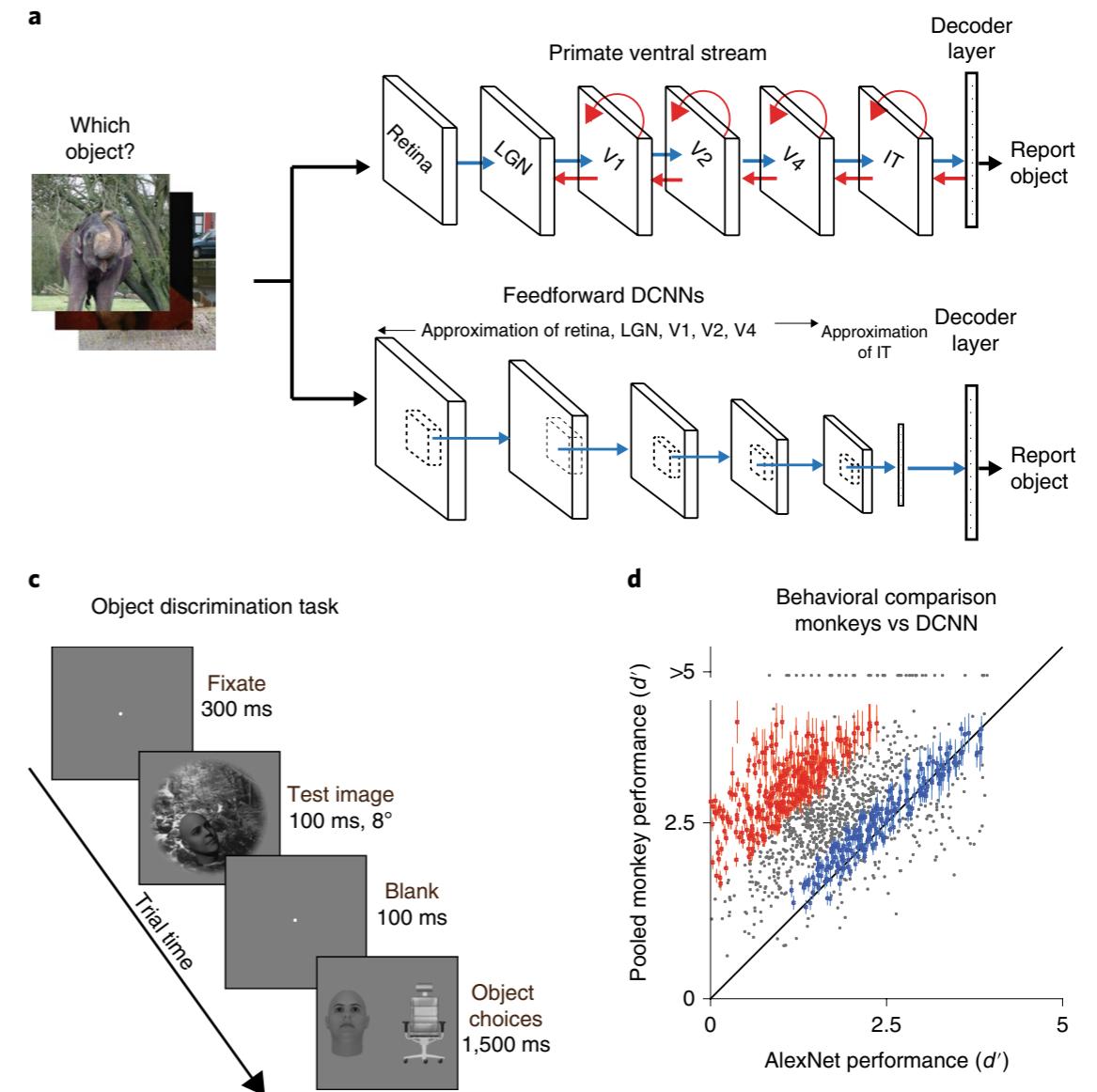
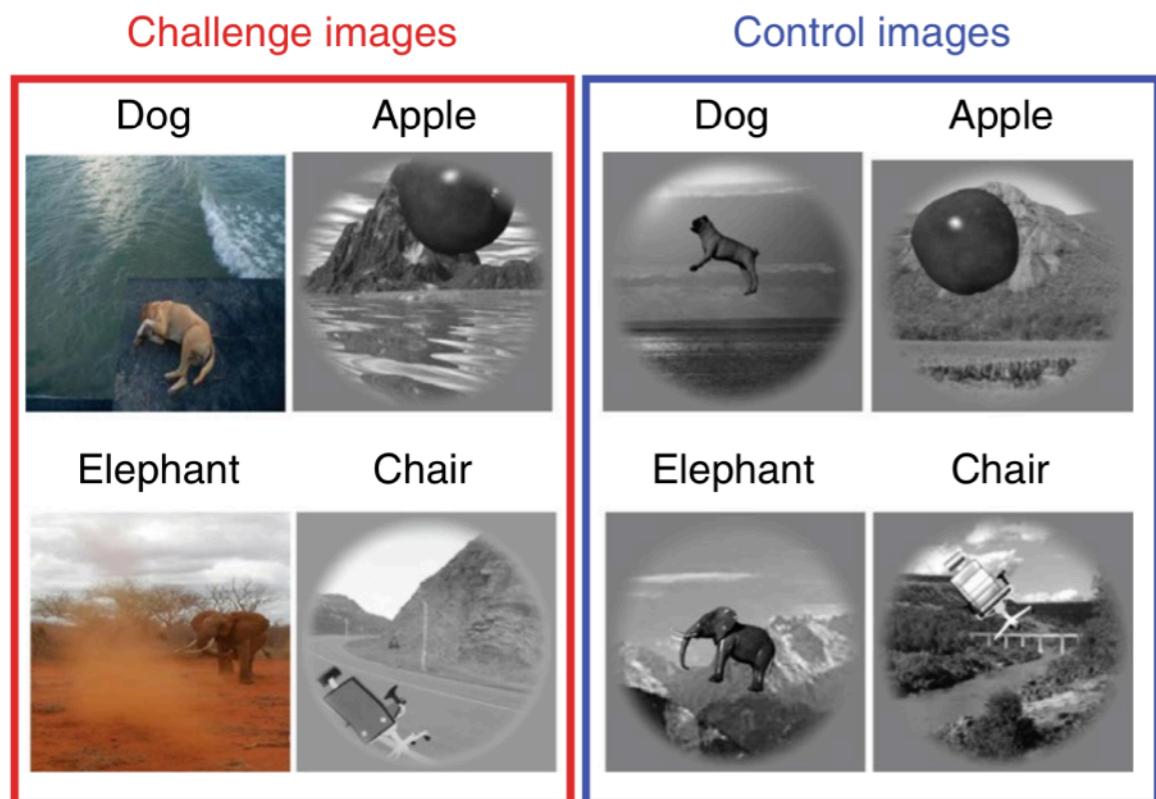
A: “ostrich”

A: “pyramid”

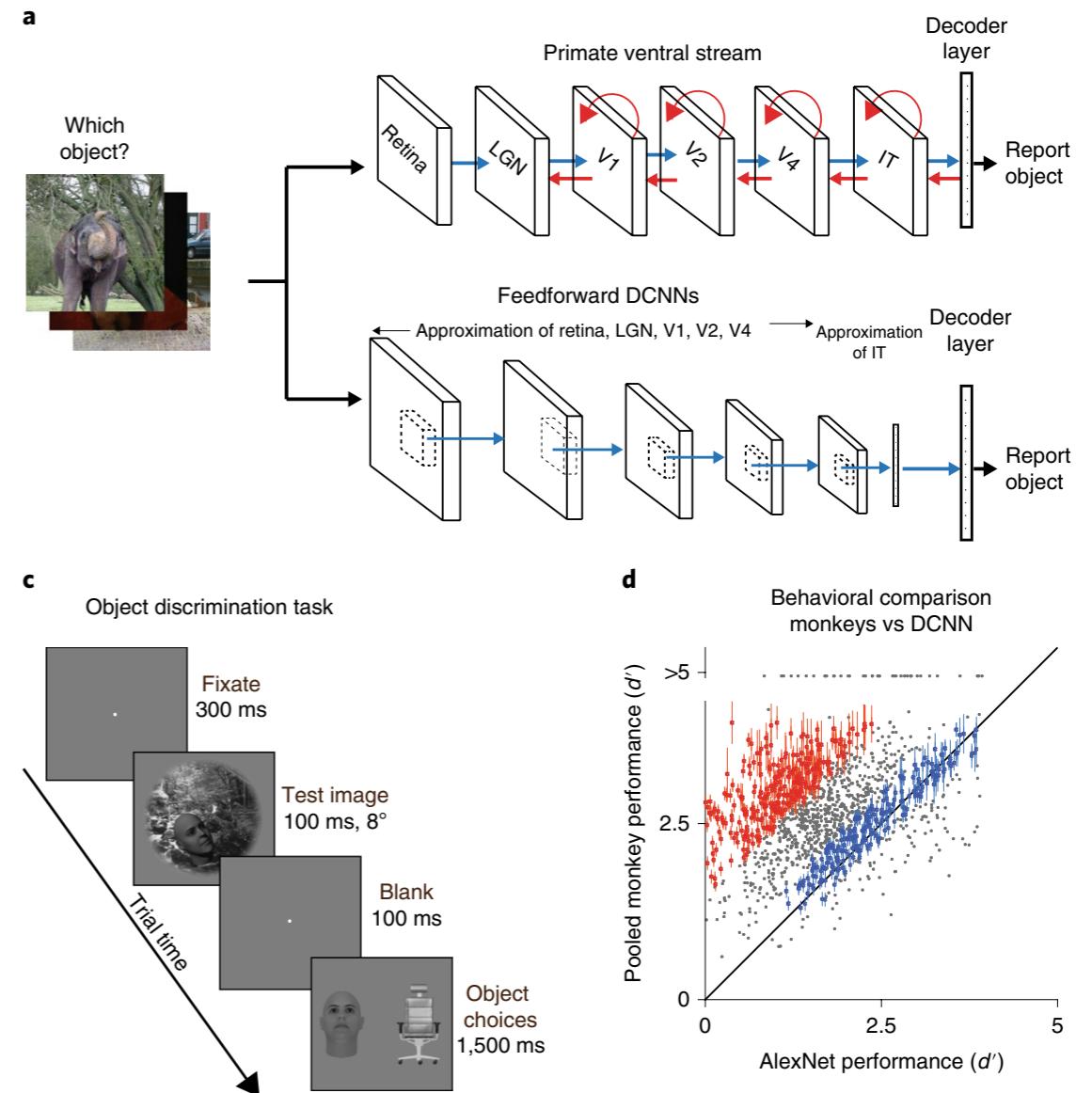
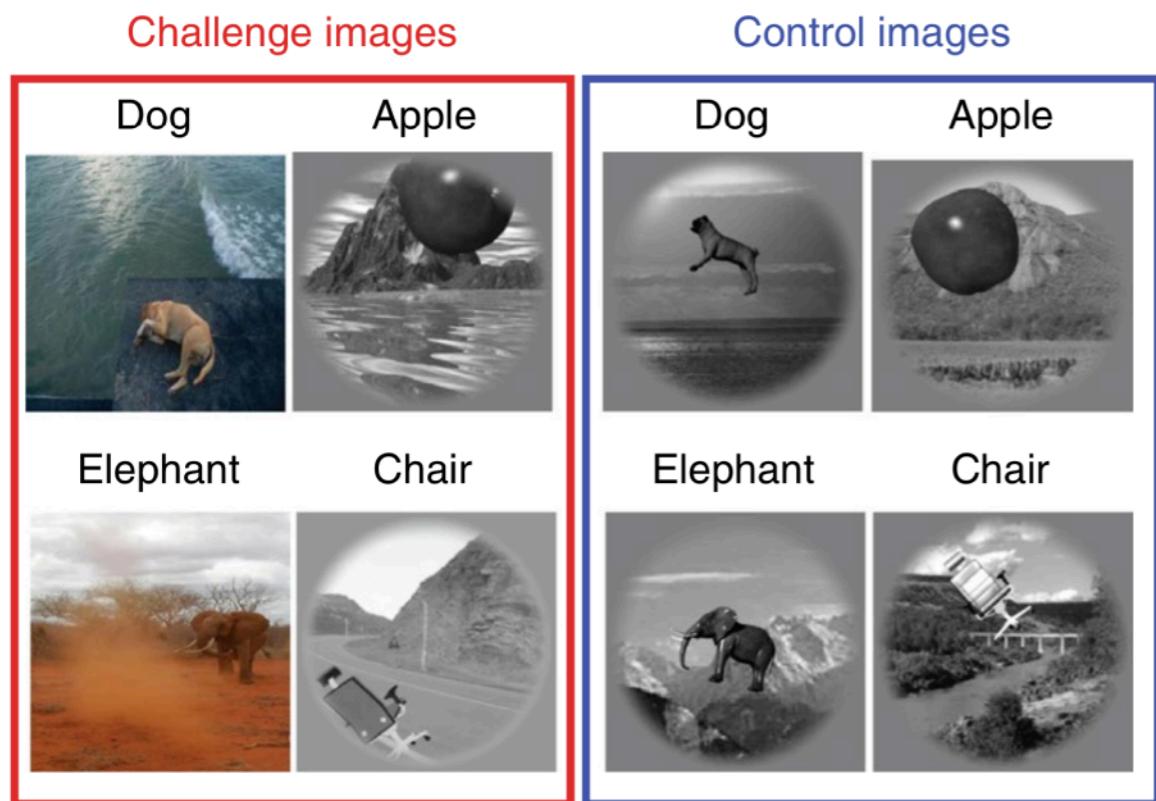
Adversarial testing of the limit of discriminative model



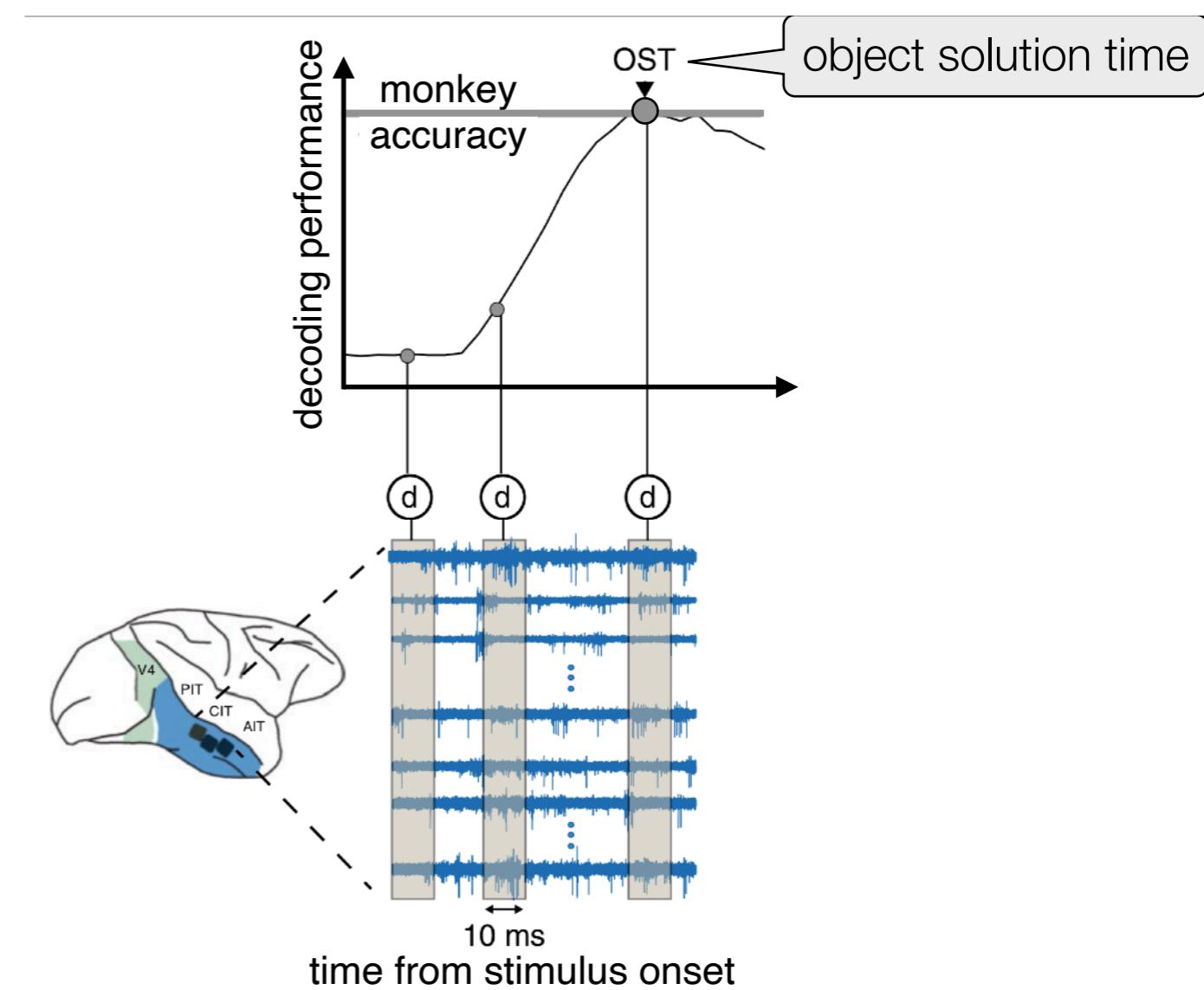
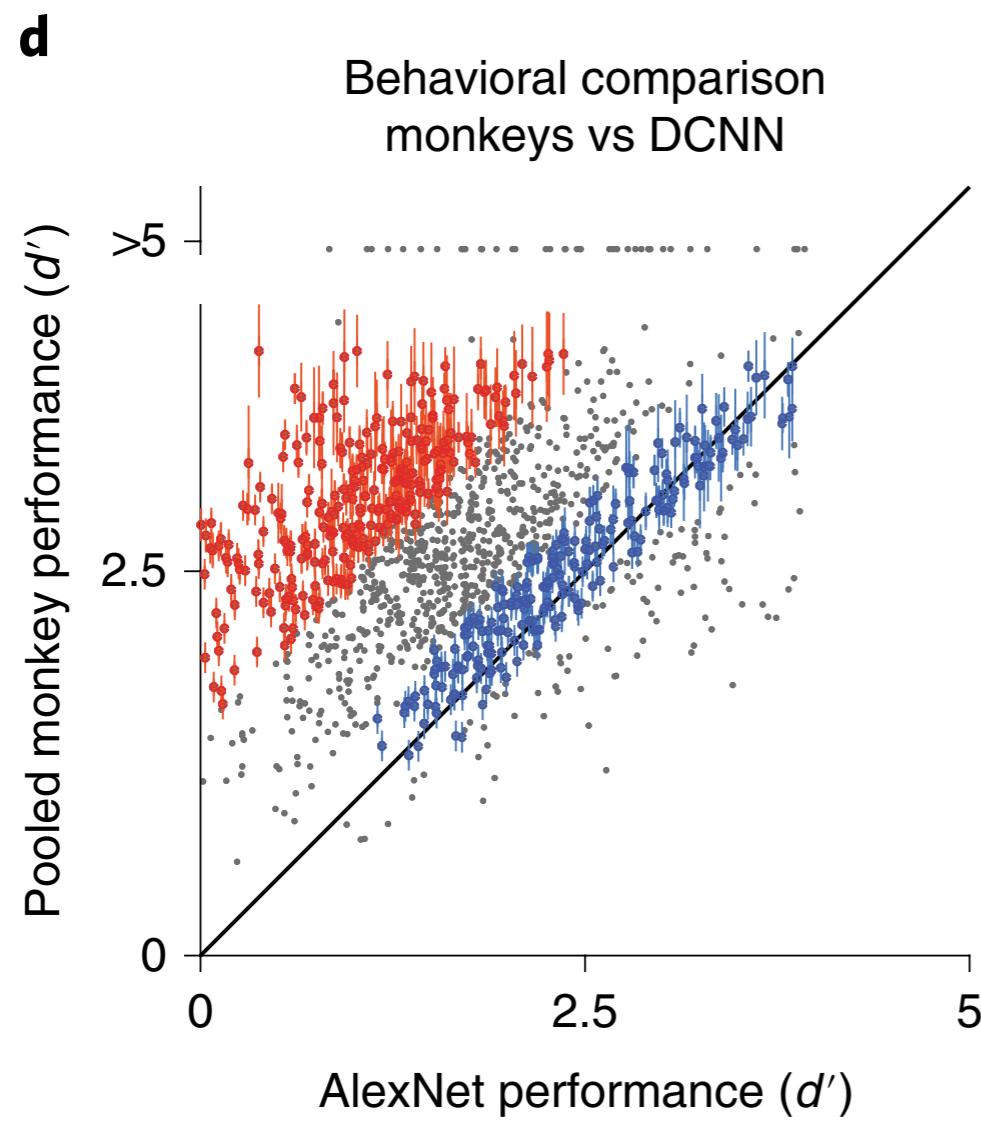
Adversarial testing of the limit of discriminative model



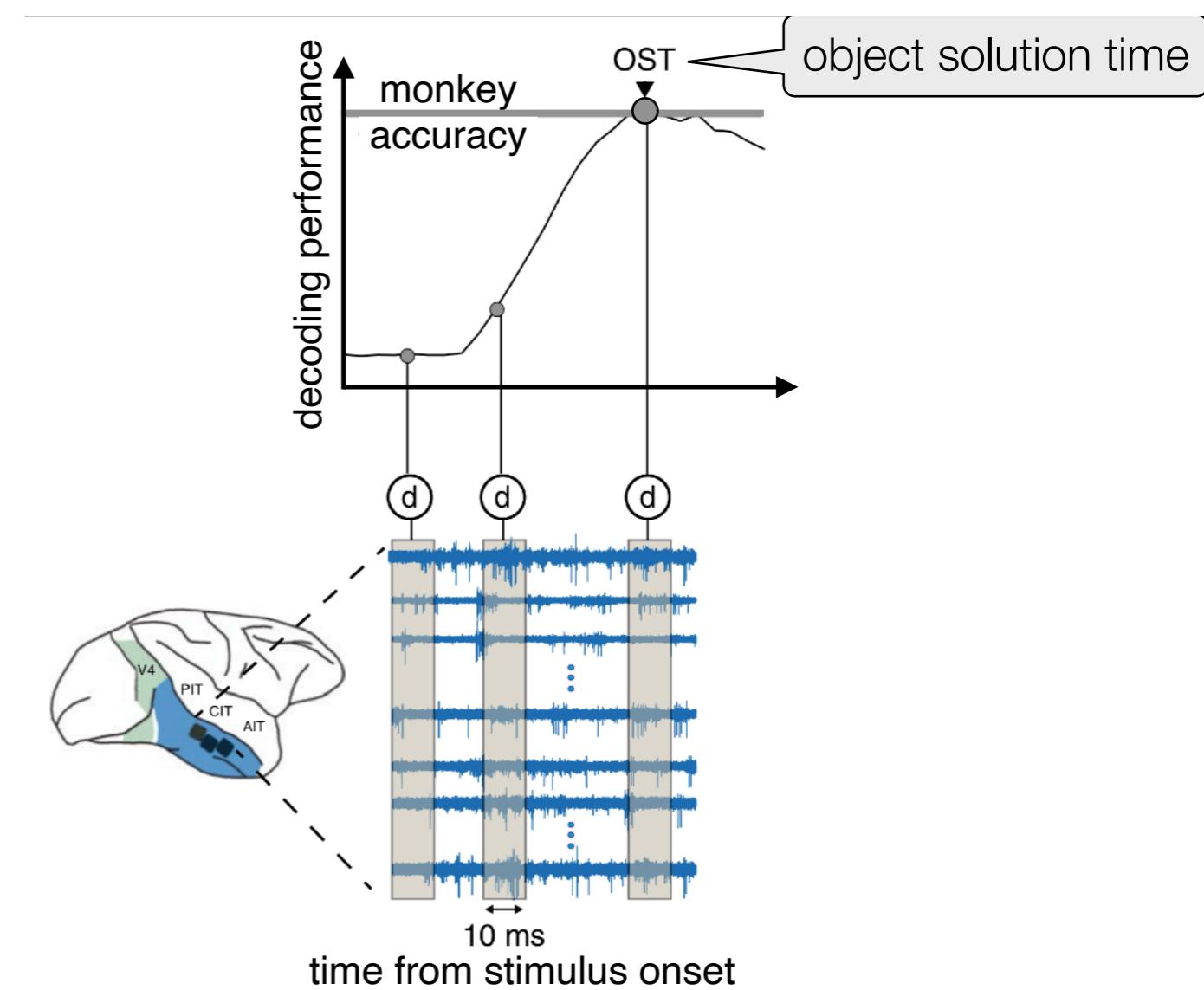
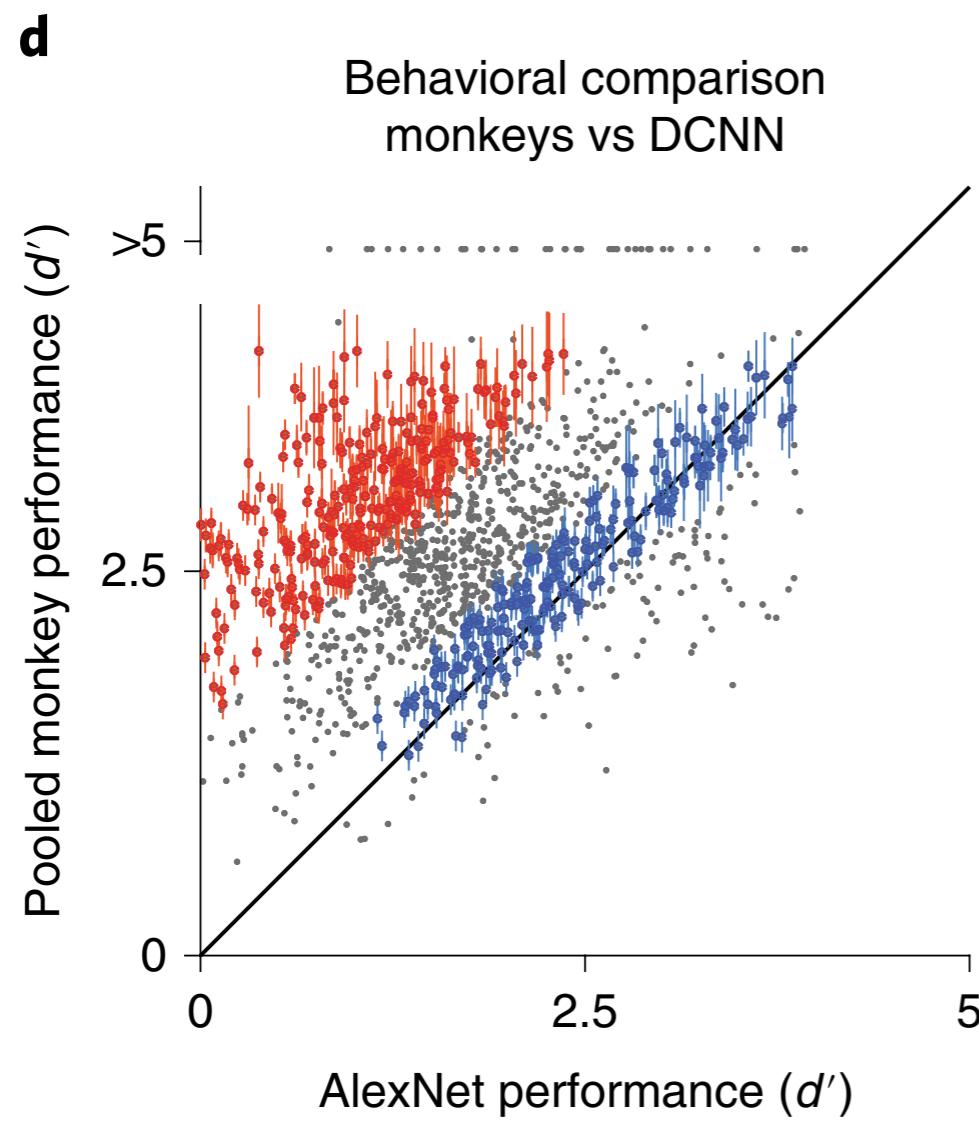
Adversarial testing of the limit of discriminative model



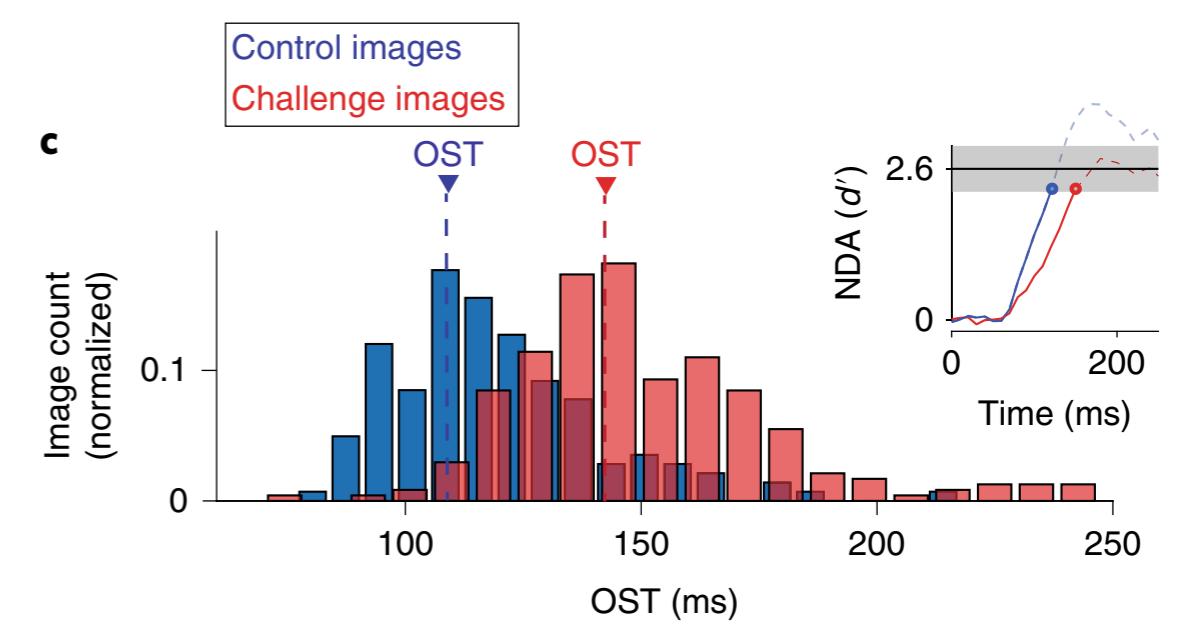
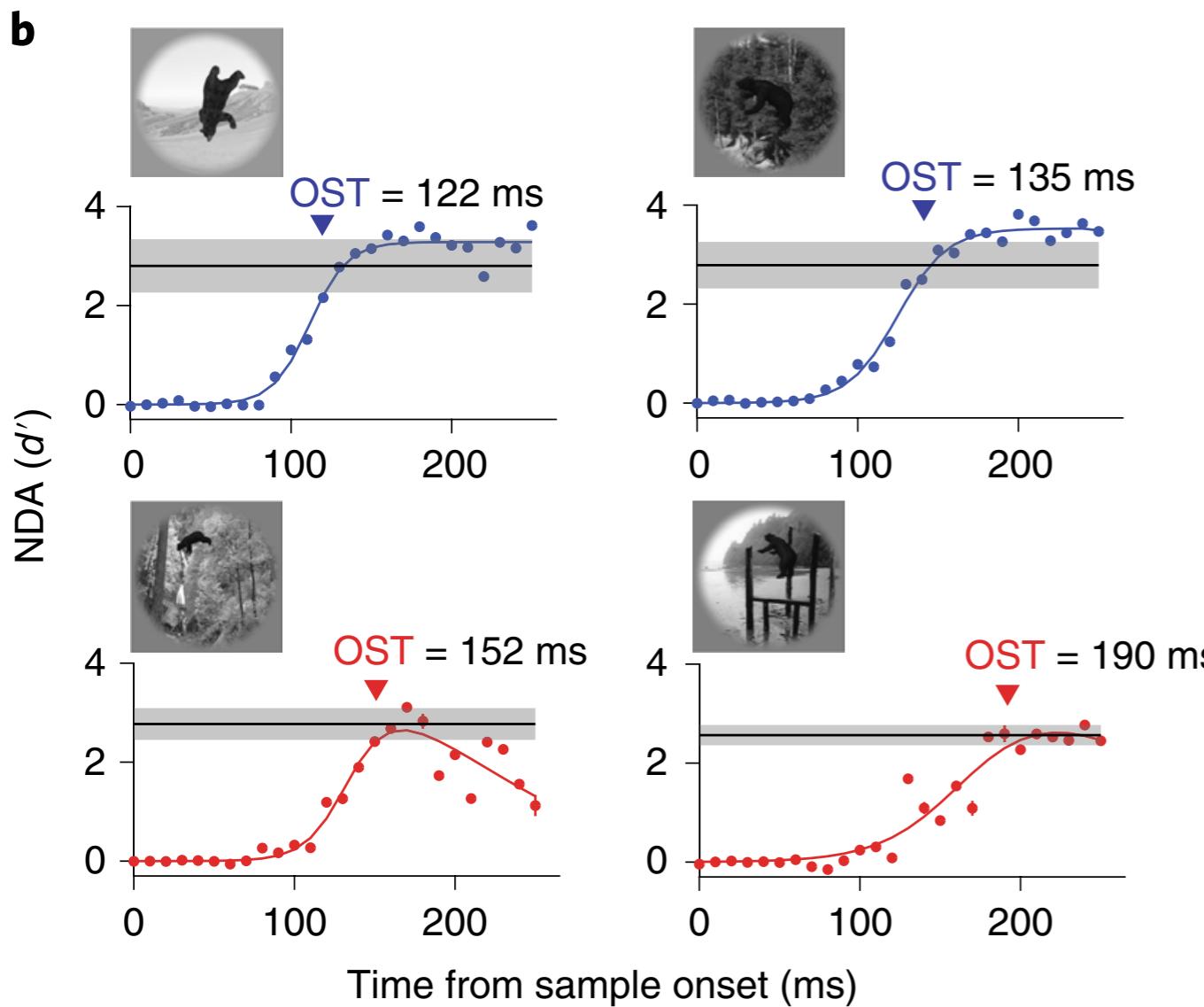
Adversarial testing of the limit of discriminative model



Adversarial testing of the limit of discriminative model



Object solution time is higher for challenging tasks



Top-down feedback and recurrent network

