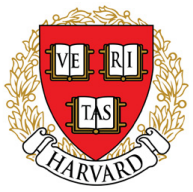
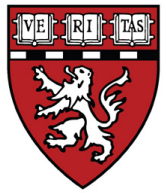
CENTER FOR
Brains
Minds+
Machines





Recurrent computations to the rescue

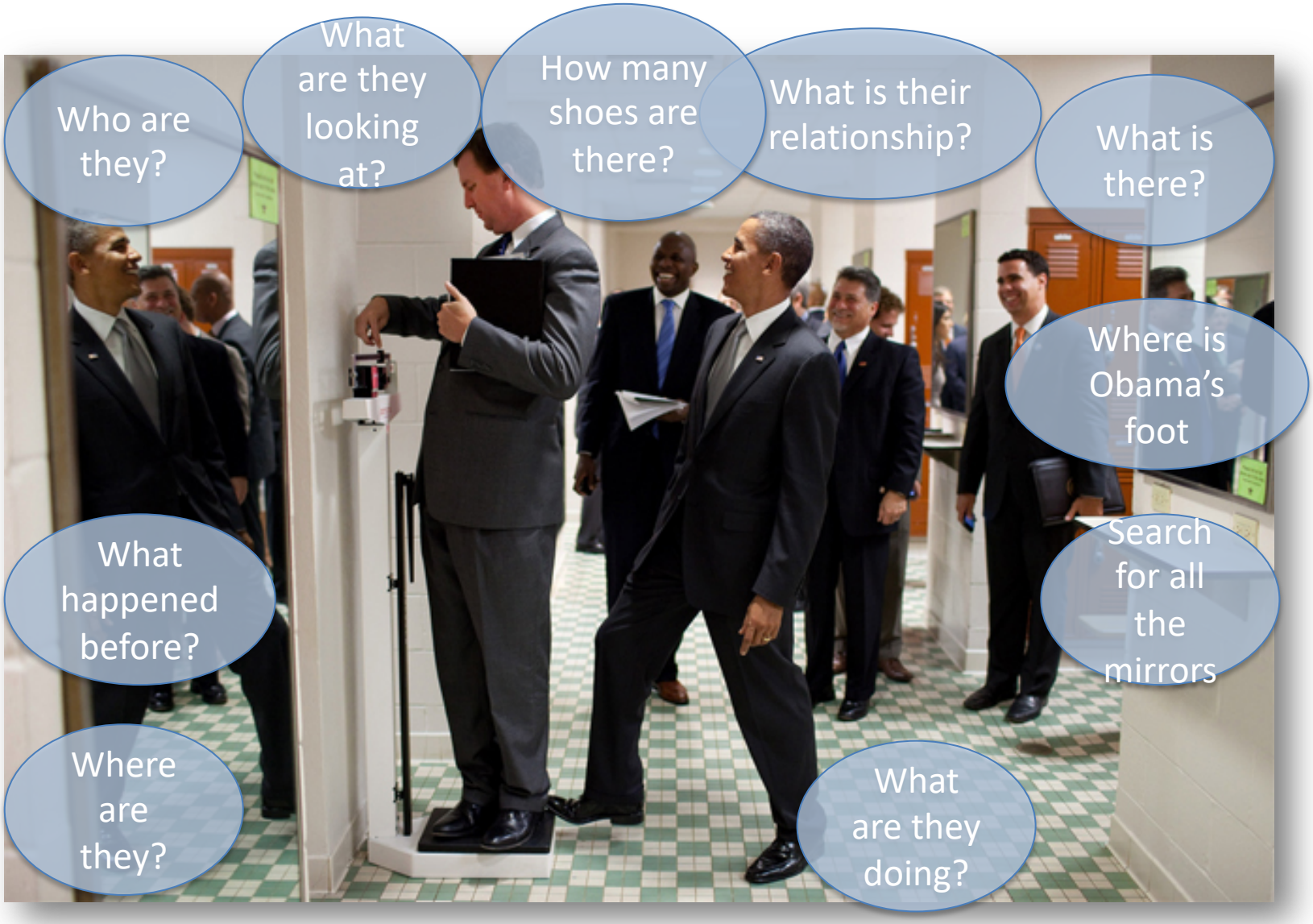
Gabriel Kreiman

gabriel.kreiman@tch.harvard.edu

http://klab.tch.harvard.edu

Scan to download
papers+data+code

An image is worth a million words



State of the Art in Image Captioning

Kreiman and Serre, 2020
Beyond the feedforward sweep

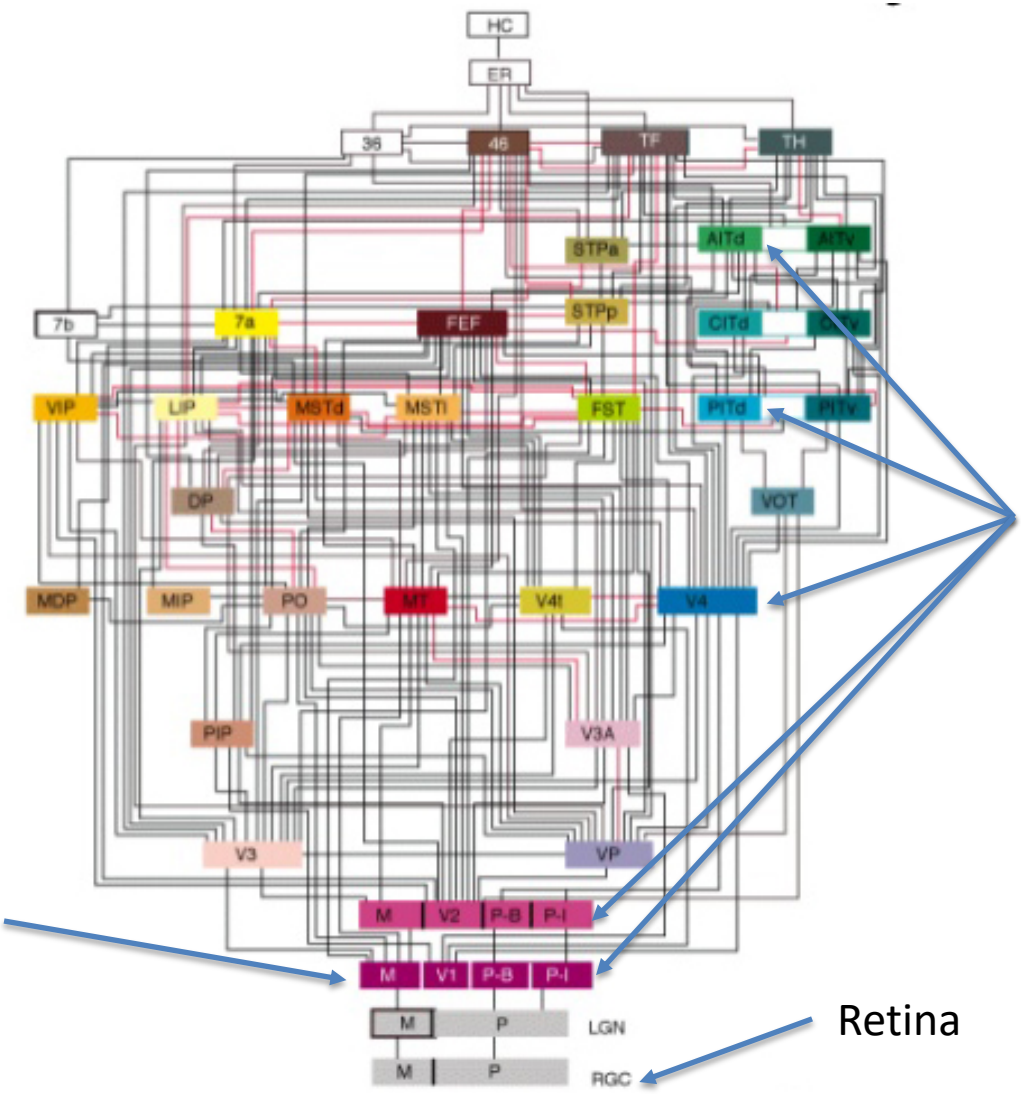
Visual cognition: a sequence of routines*

Divide et impera: break down task into simpler, reusable visual routines

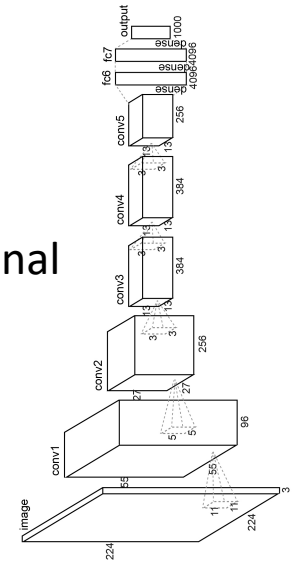
1. Extract initial sensory map → Call `VisualSampling`
2. Propose image gist → Call `FastPeriphAssess`
3. Propose foveal objects → Call `FovealRecognition`
4. Inference → Call `PatternCompletion`
5. Working memory → Call `Visual Buffer`
6. Task-dependent sampling → Call `EyeMovement`
7. Response → Call `DecisionMaking`

* Shimon Ullman. Visual Routines. 1984

Standing on the shoulders of giants: Mesoscopic connectivity of the primate visual system



~ Deep convolutional neural networks



Krizhevsky et al, NIPS 2012

Primary visual cortex

Retina

Felleman and Van Essen. *Cerebral Cortex* 1991

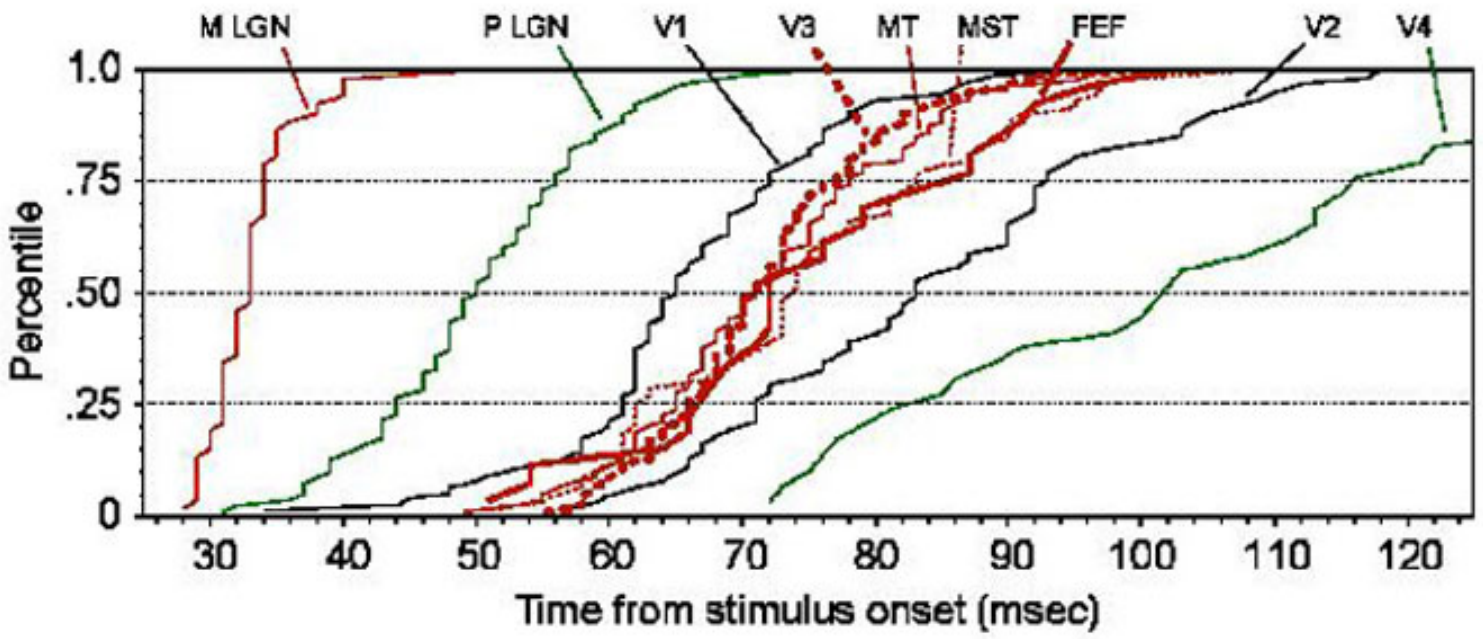
Visual cognition: a sequence of routines*

Divide et impera: break down task into simpler, reusable visual routines

1.	Extract initial sensory map	→ Call `VisualSampling`	Retina
2.	Propose image gist	→ Call `FastPeriphAssess`	V1, V2, V4
3.	Propose foveal objects	→ Call `FovealRecognition`	ITC
4.	Inference	→ Call `PatternCompletion`	Ventral Stream
5.	Working memory	→ Call `Visual Buffer`	PFC
6.	Task-dependent sampling	→ Call `EyeMovement`	Dorsal Stream
7.	Response	→ Call `DecisionMaking`	PFC

* Shimon Ullman. Visual Routines. 1984

Outline



Each additional processing step takes ~10 to 15 ms

Schmolesky et al 1998

Visual cognition: a sequence of routines*

Divide et impera: break down task into simpler, reusable visual routines

1. Extract initial sensory map	→ Call `VisualSampling`	Retina	~50 ms
2. Propose image gist	→ Call `FastPeriphAssess`	V1, V2, V4	~100 ms
3. Propose foveal objects	→ Call `FovealRecognition`	ITC	~150 ms
4. Inference	→ Call `PatternCompletion`	Ventral Stream	~200 ms
5. Working memory	→ Call `Visual Buffer`	PFC	~300 ms
6. Task-dependent sampling	→ Call `EyeMovement`	Dorsal Stream	~300 ms
7. Response	→ Call `DecisionMaking`	PFC	~300-1000 ms

* Shimon Ullman. Visual Routines. 1984

Outline

1. Pattern completion: making inferences from partial information
2. Putting vision in context
3. Active sampling: extracting information via eye movements

Outline

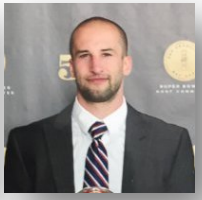
1. **Pattern completion: making inferences from partial information**
2. Putting vision in context
3. Active sampling: extracting information via eye movements

Tang et al, Neuron 2014
Tang et al, PNAS 2018

Martin Schrimpf

Hanlin Tang

Bill Lotter



Pattern completion is a hallmark of intelligence

A C E G → **I**

1, 2, 3, 5, 7, 11, → **13**

V_S_A_ R_C_G_I_I_N → **Visual Recognition**

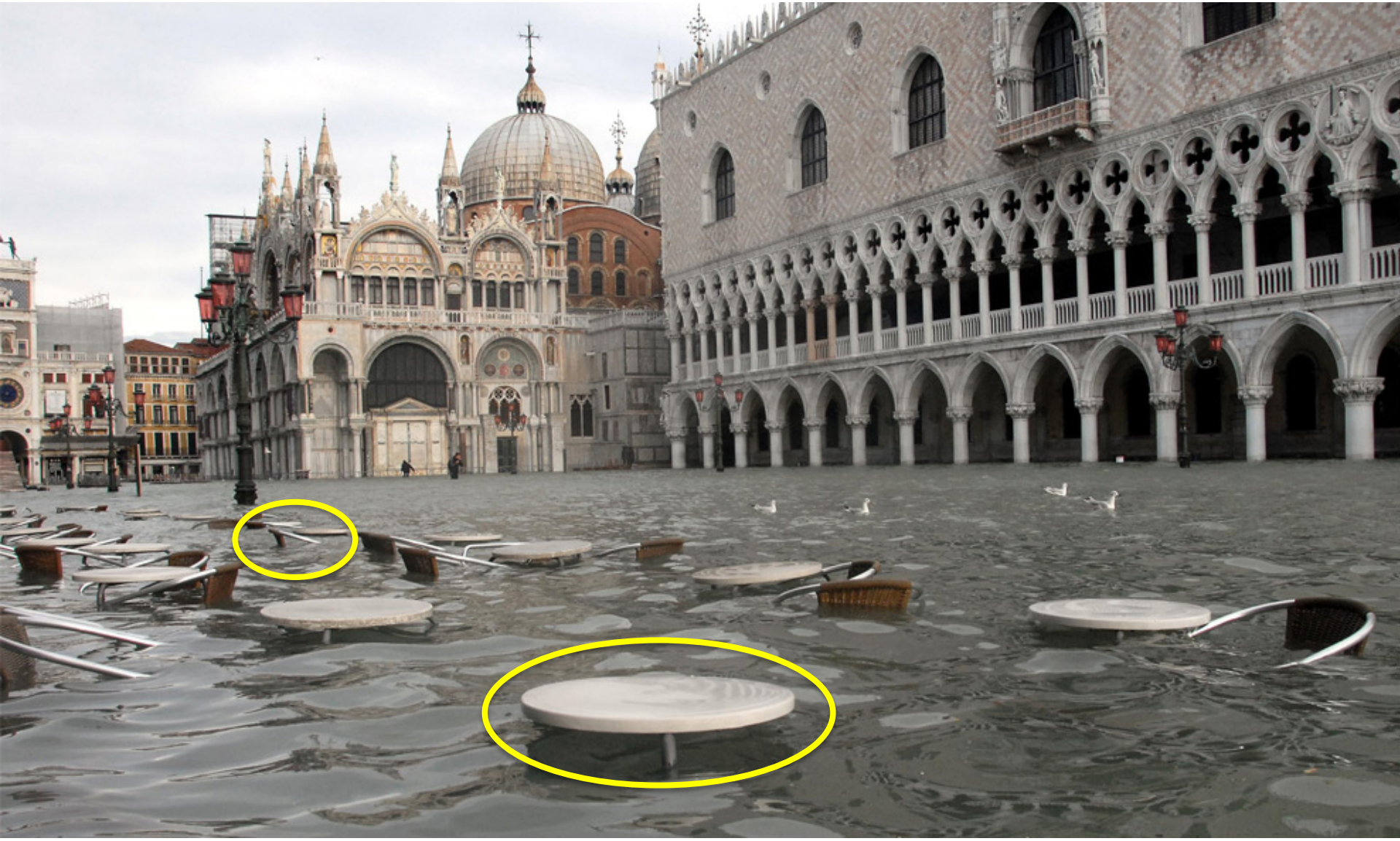
Even though it was raining heavily,
John decided to go out without an… → **Umbrella**

 → 

Also:

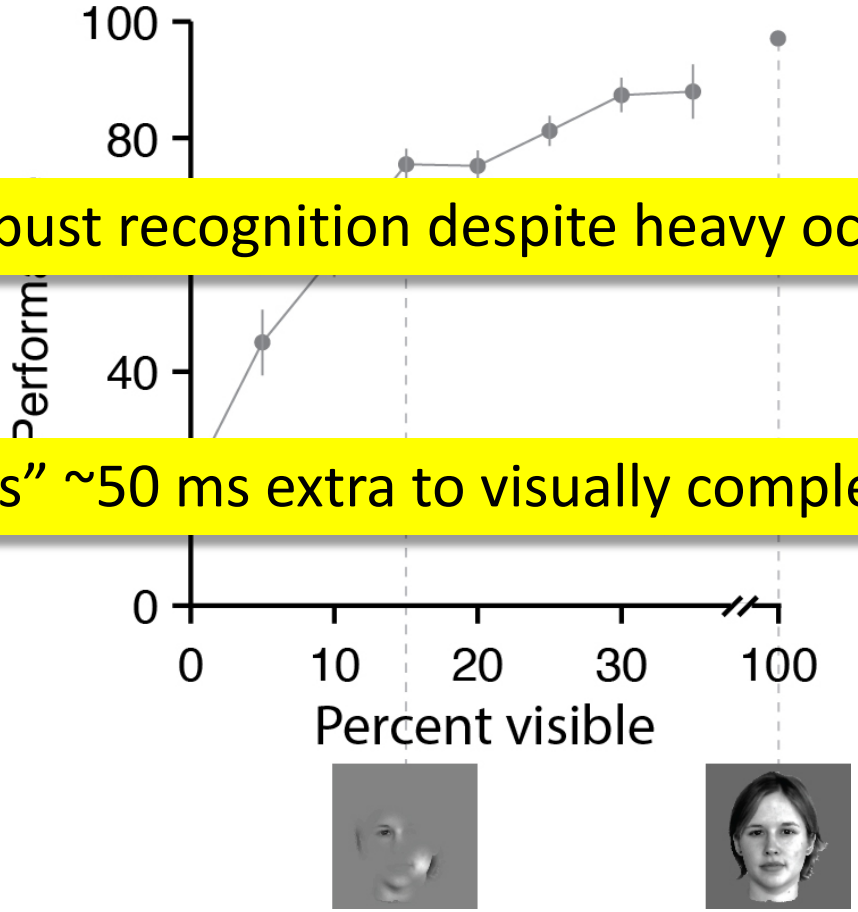
Reading a story
Social interactions

Visual recognition of occluded objects



Visual inference from partial information

Strong robustness to limited visibility

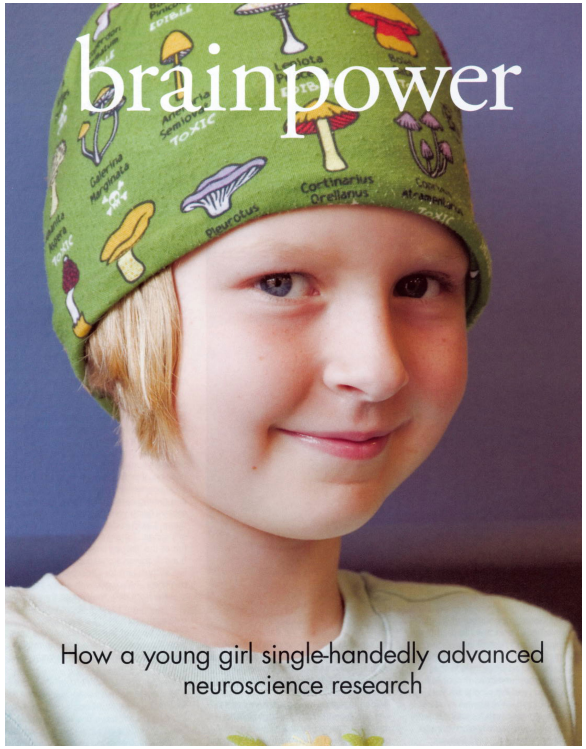


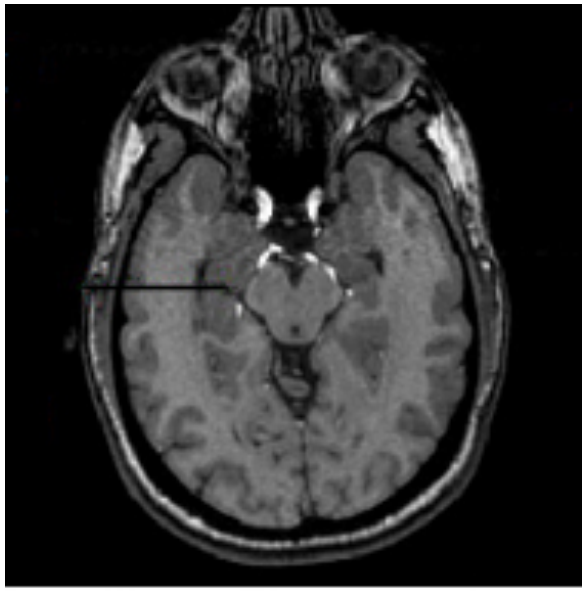
Robust recognition despite heavy occlusion

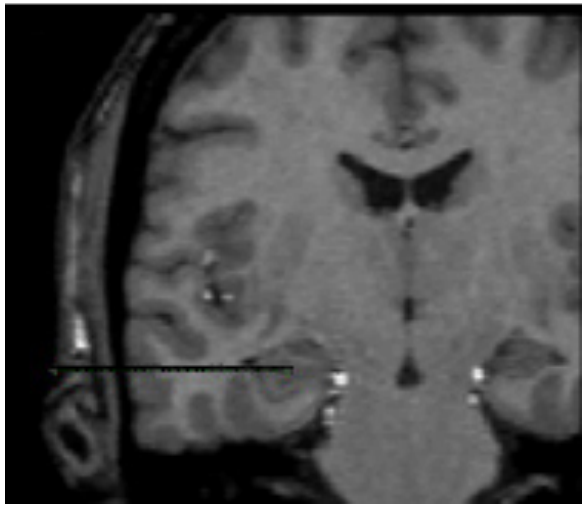
It “costs” ~50 ms extra to visually complete patterns

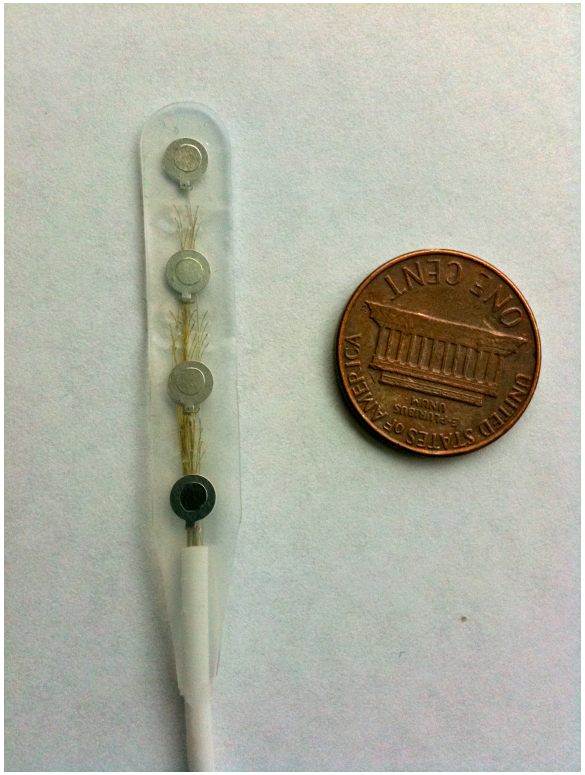
Tang et al, Neuron 2014
Tang et al, PNAS 2018

Peeking inside the human brain

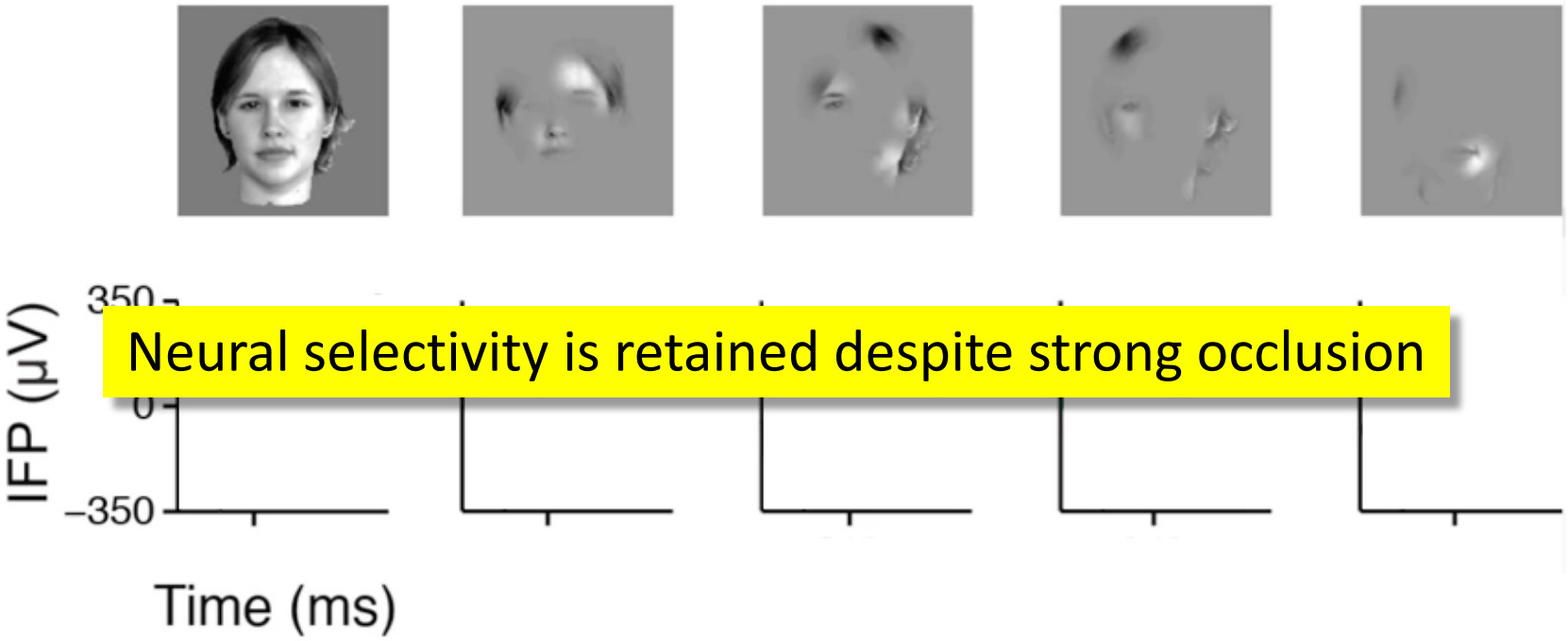








Neural mechanisms of pattern completion



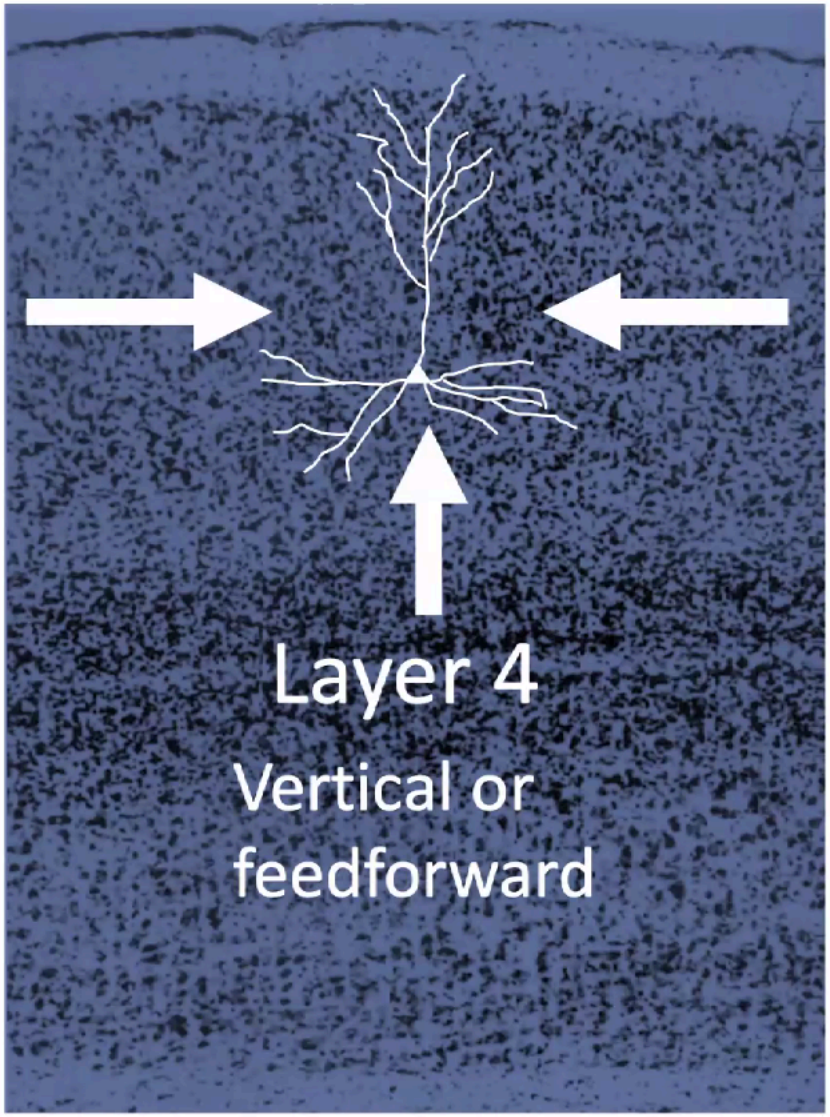
Neural selectivity is retained despite strong occlusion

Neural responses are delayed by ~50 ms

Tang et al, Neuron 2014
Tang et al, PNAS 2018

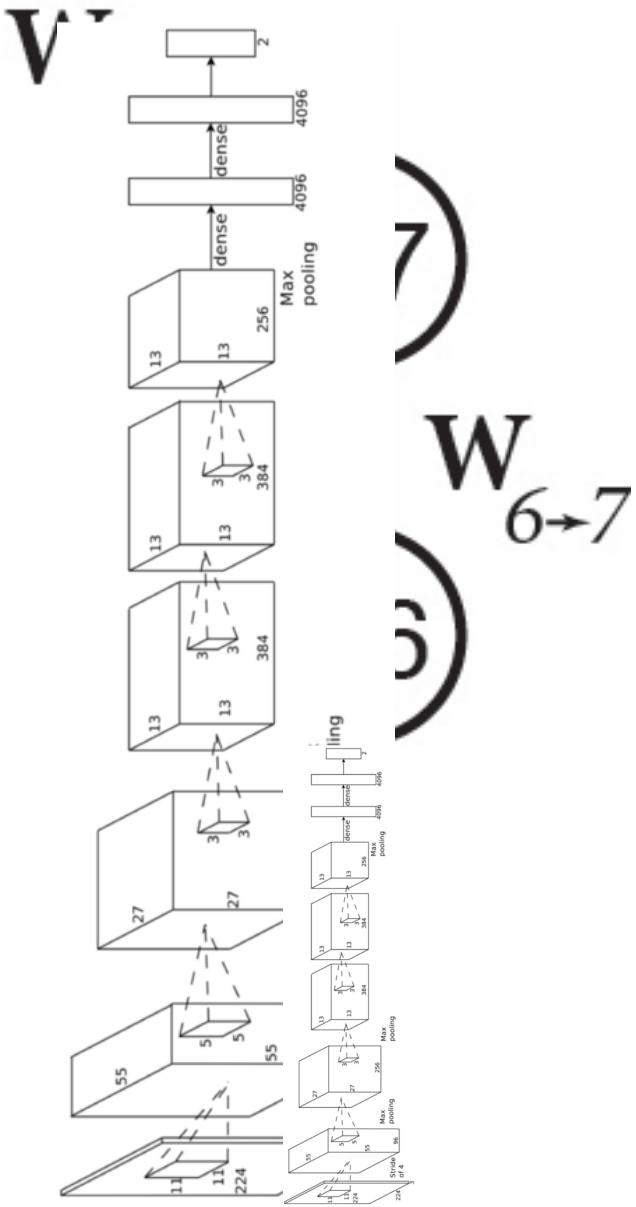
Rumination through horizontal connections

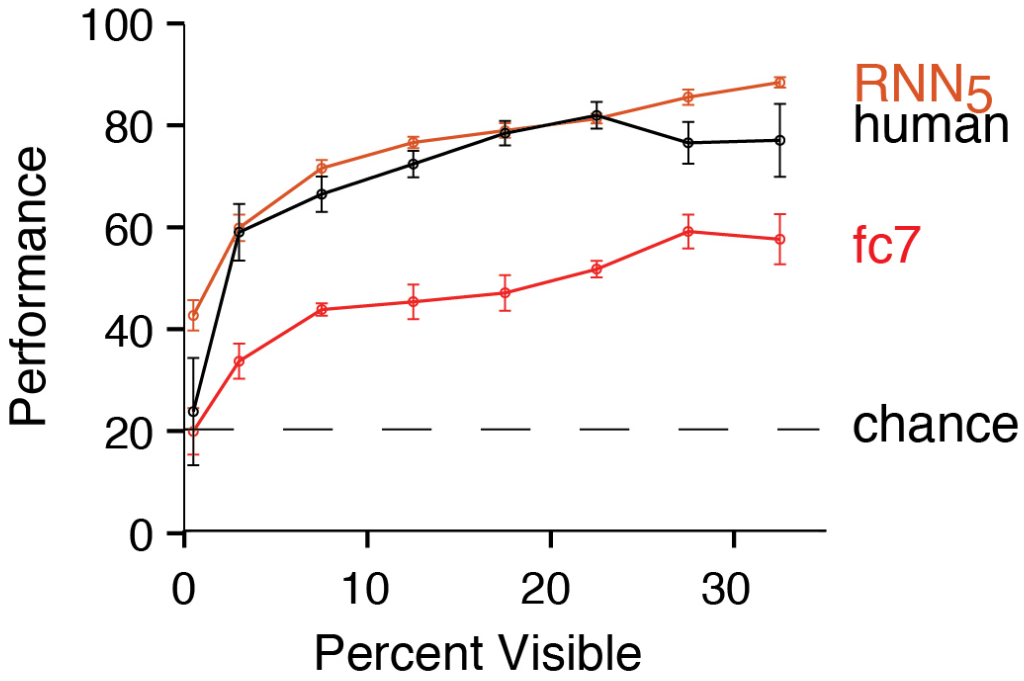
Layer 2/3
horizontal or
recurrent



1. Behavior: 50 ms cost.
2. Neurophysiology: 50 ms delay
3. Neuroanatomy: slow horizontal connection

Neurobiological recipe: add horizontal recurrent connections to deep convolutional networks





Outline

1. Pattern completion: making inferences from partial information
2. **Putting vision in context**
3. Active sampling: extracting information via eye movements

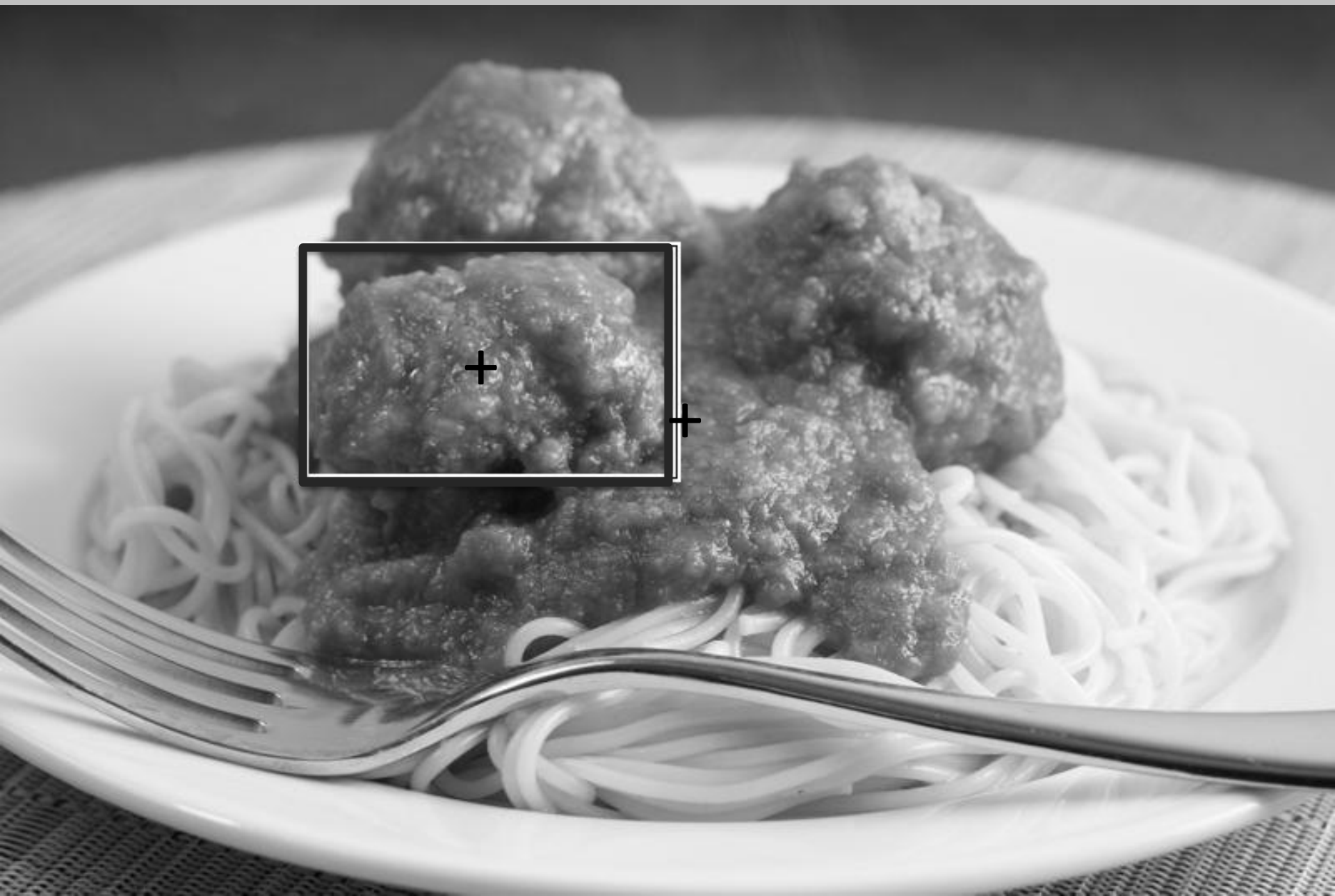
Zhang et al, CVPR 2020

Martin Schrimpf

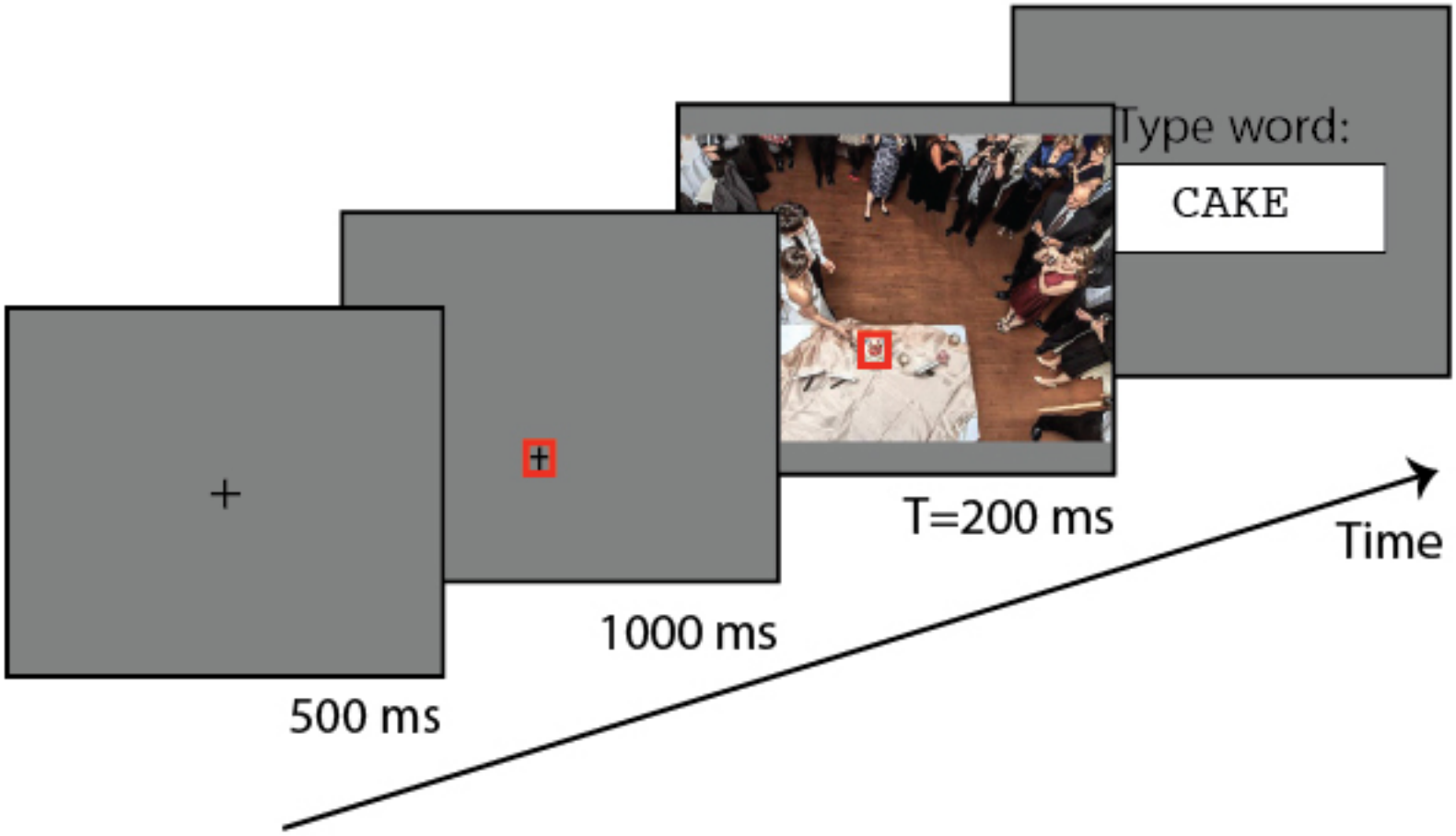
Mengmi Zhang

The role of context in vision: Example 1

The role of context in vision: Example 2

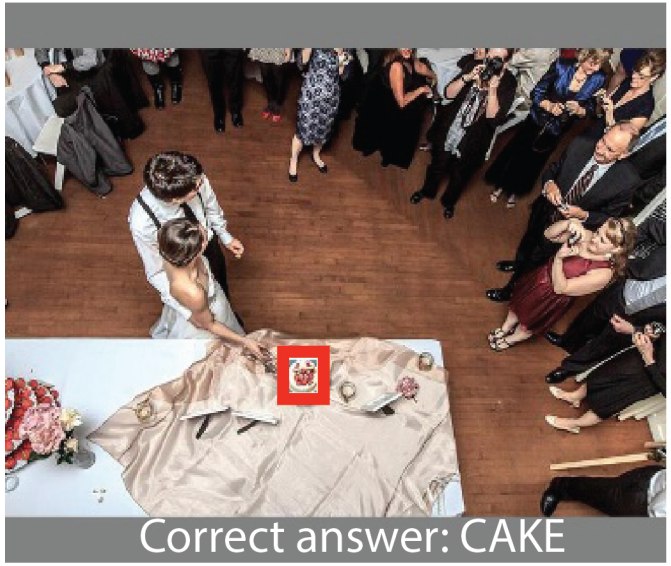


Investigating the role of context in visual recognition

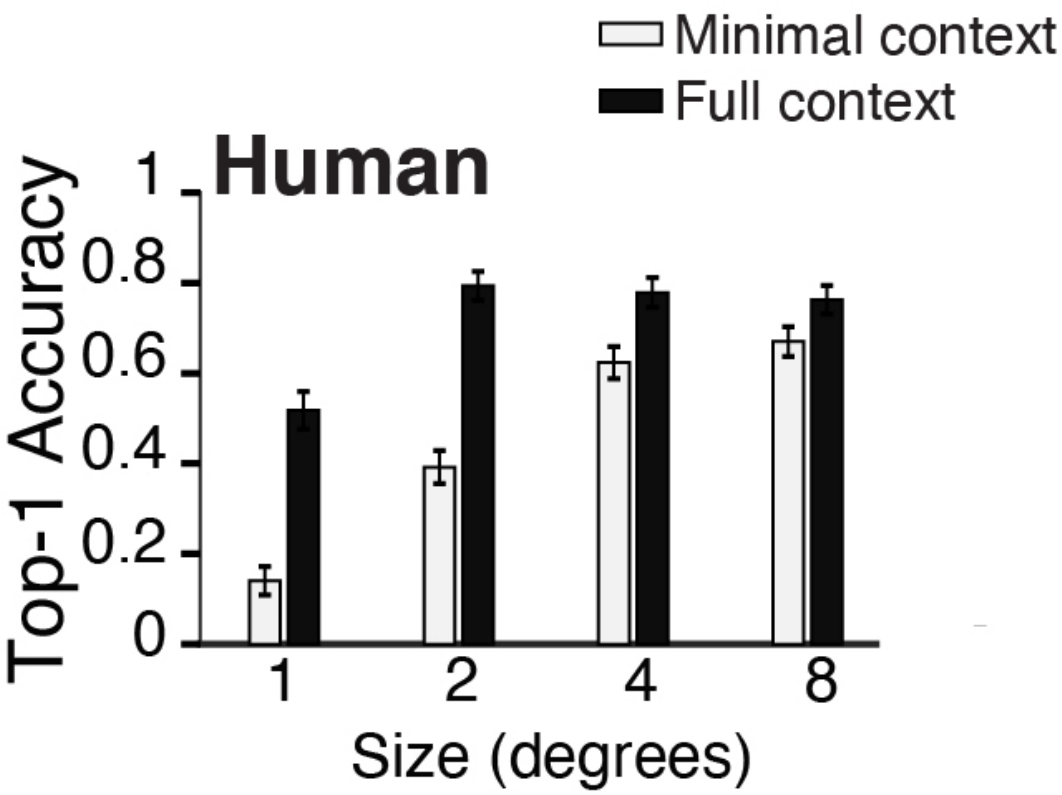


What, where, and when of contextual modulation

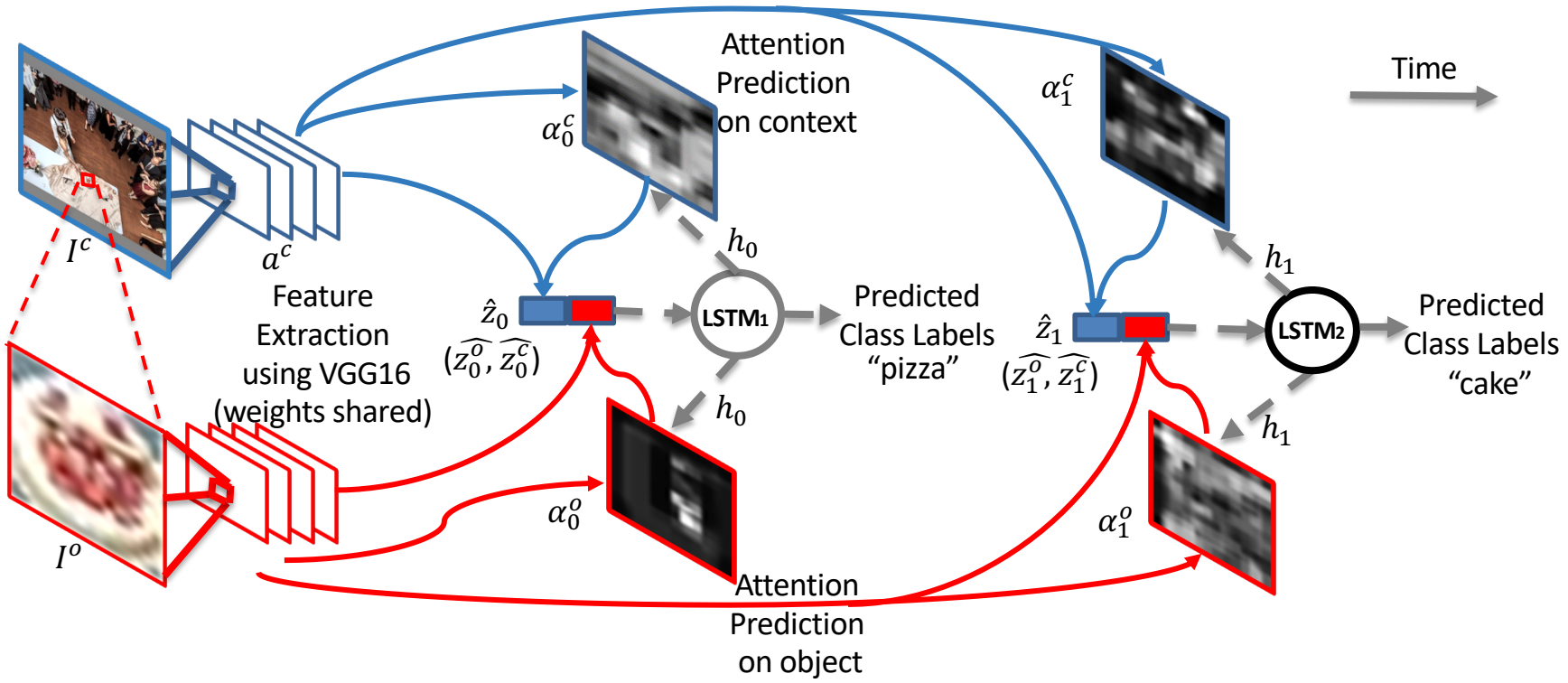
a Full context



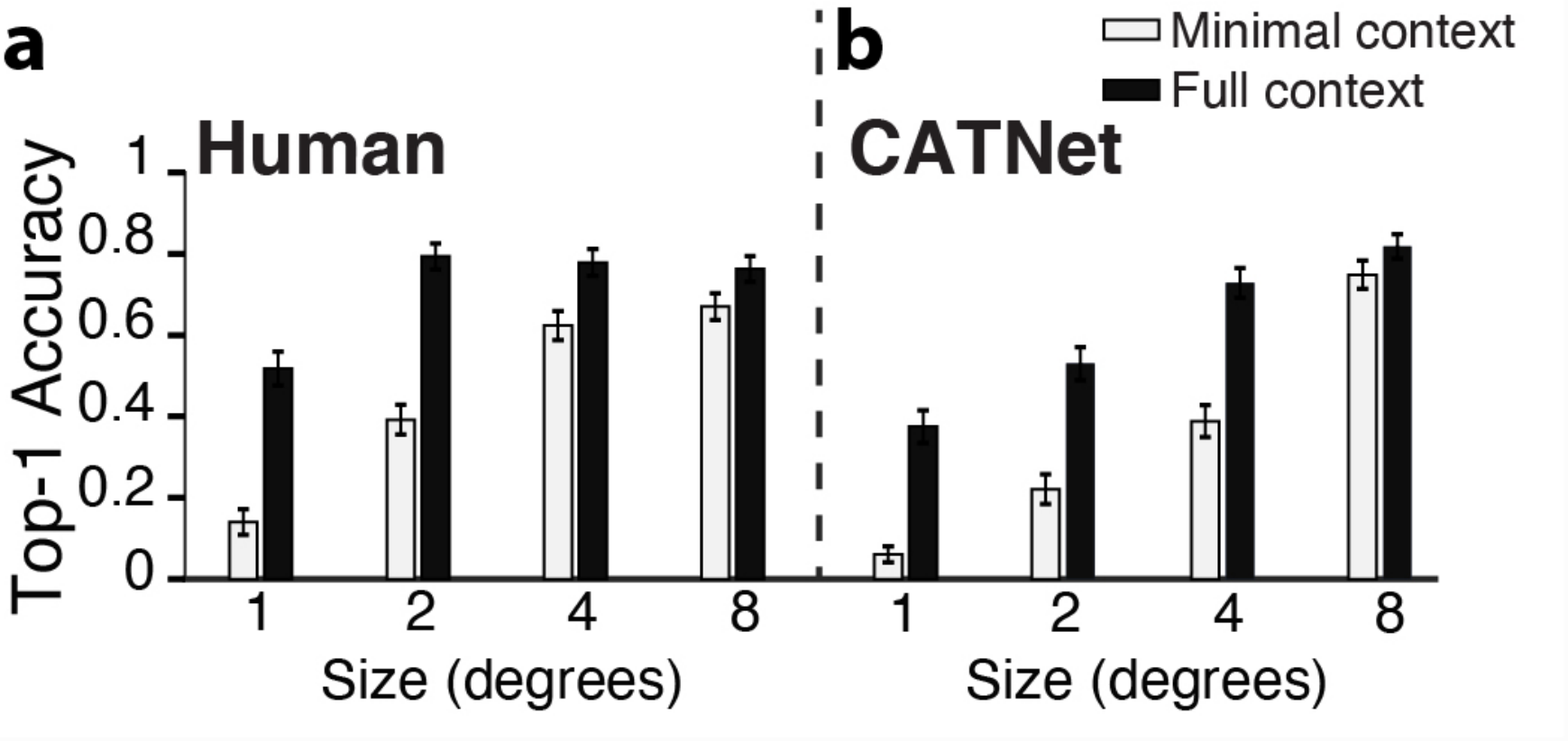
Context enhances visual recognition



CATNet:
Context-aware Attention Two-Stream Network

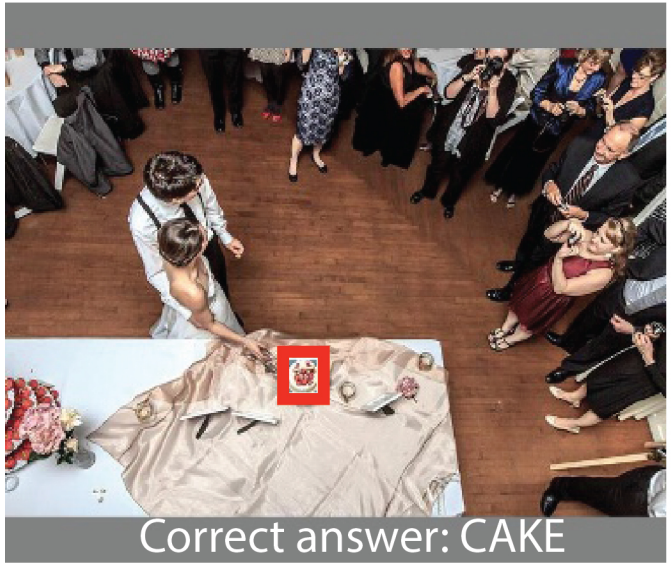


Context enhances visual recognition in the model

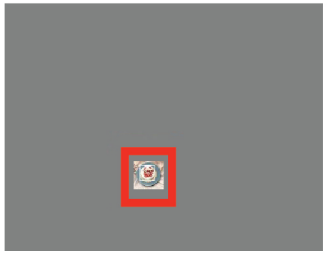


What, where, and when of contextual modulation

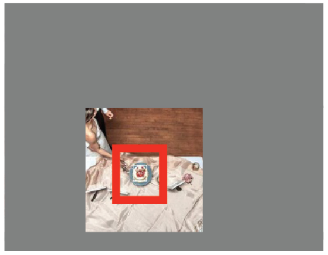
a Full context



b Minimal context



c Context area



Outline

1. Pattern completion: making inferences from partial information
2. Putting vision in context
3. **Active sampling: extracting information via eye movements**

Zhang et al, Nature Communications 2018

Mengmi Zhang

High-resolution fovea, low-resolution periphery

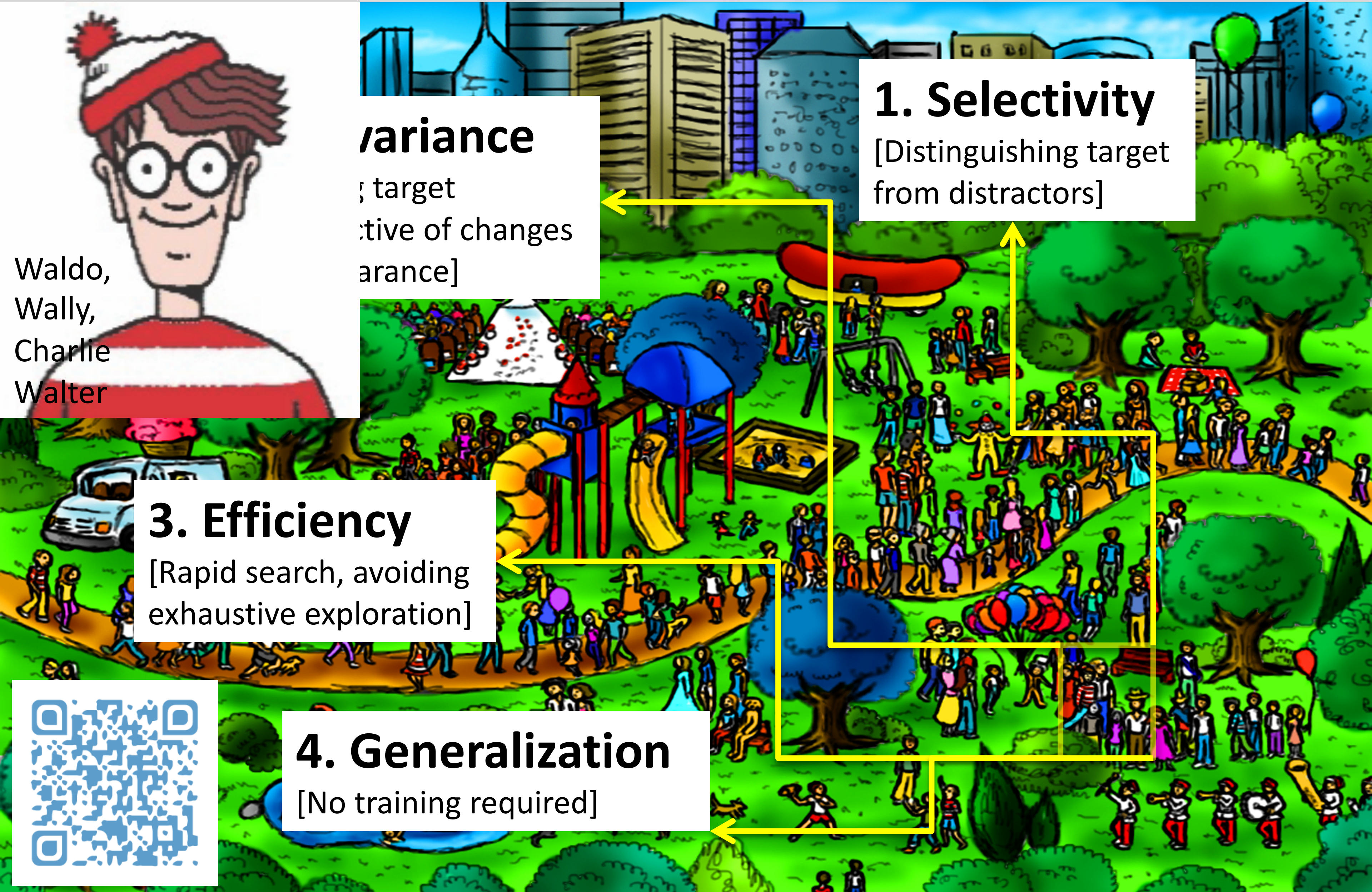
We are legally blind outside the fovea

Reading depends strictly on foveal resolution. Try to fixate on the letter "R" shown here in large bold font. Make sure that you do not move your eyes away from the R. If you do, then your high resolution area rapidly shifts to whatever location you are fixating on. Once you are fixating, try to read a word that is four lines below the letter "R". This task is basically impossible for us because the resolution drops sharply outside of the fovea. The notion that we can capture the entire visual scene at high resolution is merely an illusion created by our rapid eye movements and the fact that whenever we land on a particular location, it appears in high resolution!

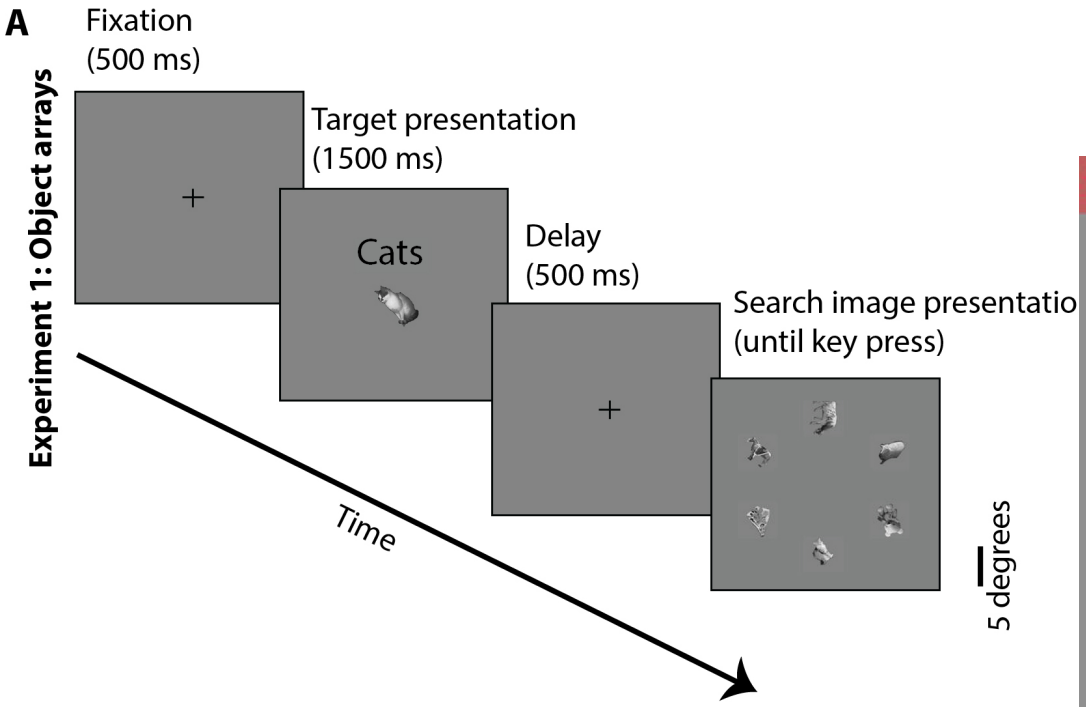
DO NOT MOVE YOUR EYES!

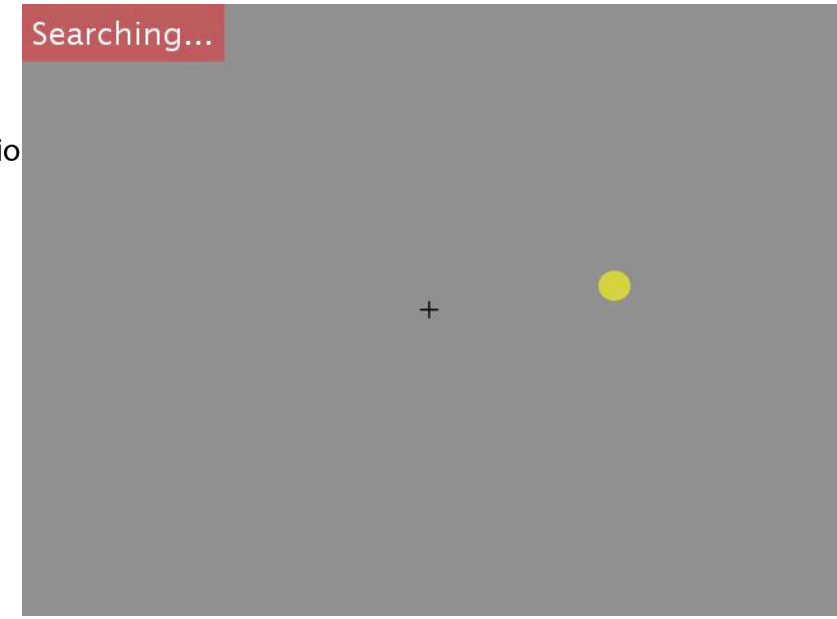
Eye movements are critical for scene understanding

Four key properties of visual search



Active sampling during visual search

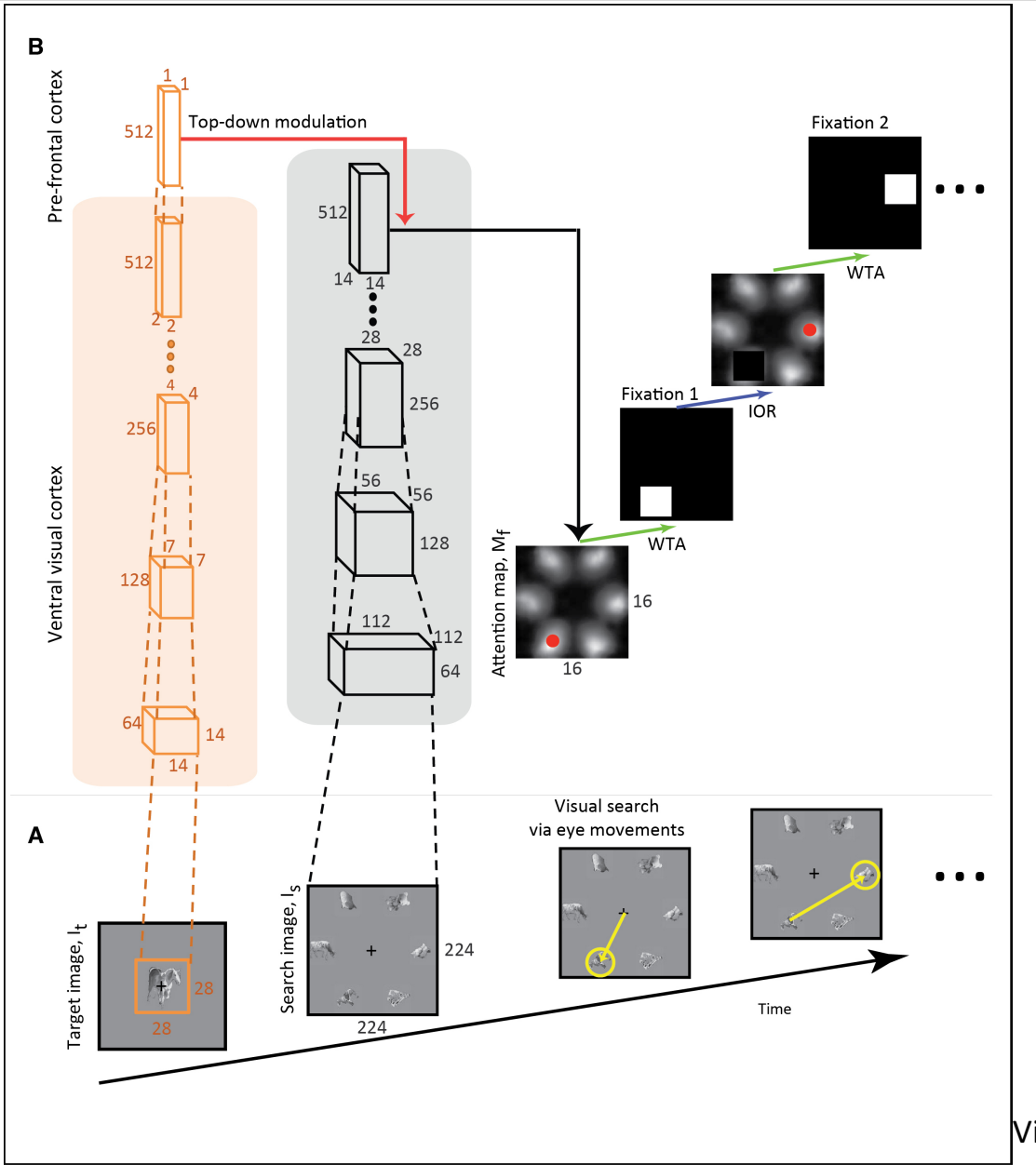




Zhang et al, 2018

Mengmi Zhang

Invariant Visual Search Network (IVSN)

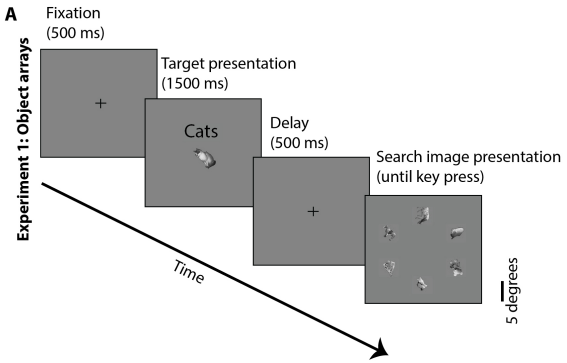


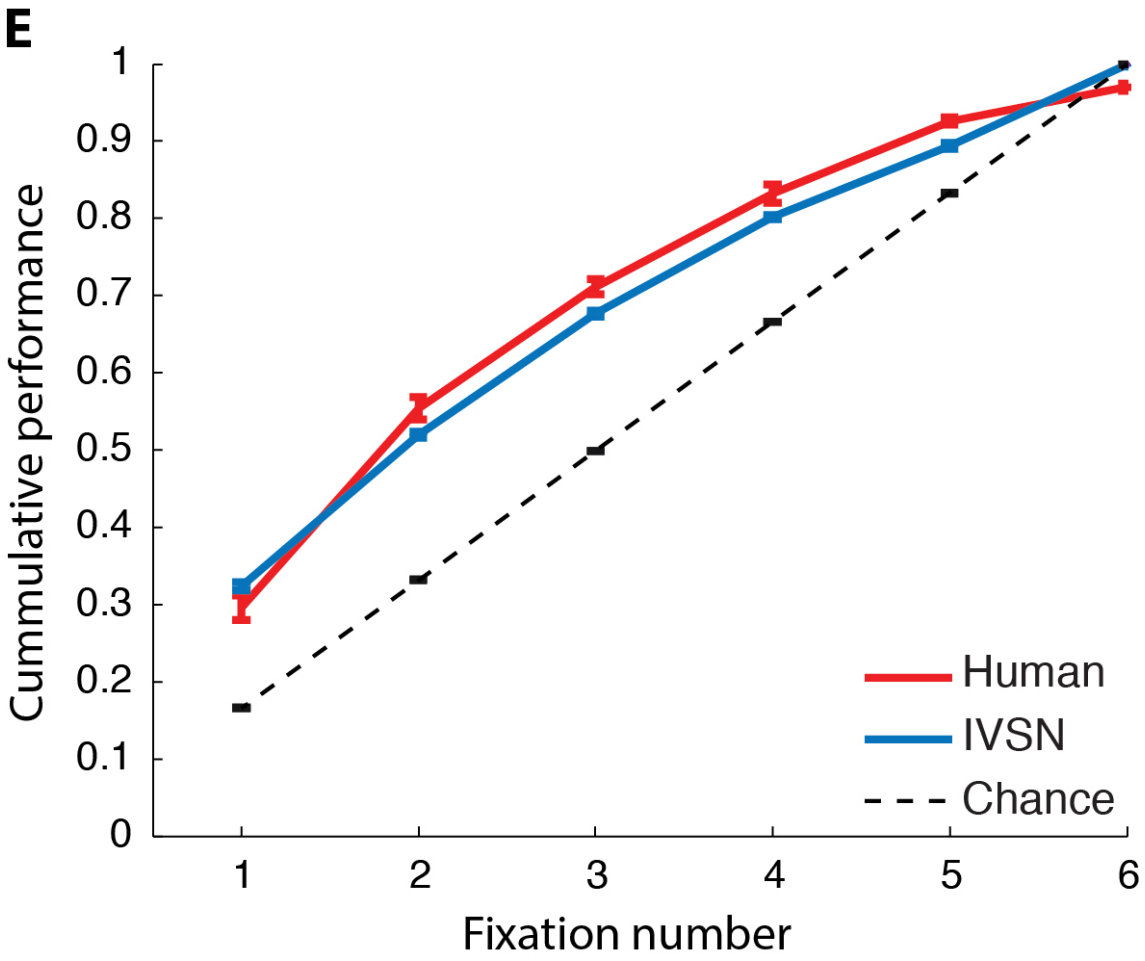
Zhang et al. Nature Communications 2018

Scan to download
paper+code+data

VGG16 (Simonyan et al 2014)
Visual search circuitry: e.g. Bichot et al (2015)

Neurally inspired model captures sampling behavior





Finding any Waldo:
zero-shot invariant and efficient visual search.
Zhang et al, 2018

Summary

Attractor-based recurrent neural networks can perform pattern completion

Recurrent connections and eccentricity dependence help incorporate context

Top-down signals can guide eye movements during visual search

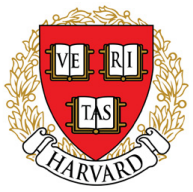
Divide and Conquer Strategy: A sequence of flexible, reusable, interchangeable **visual routines**

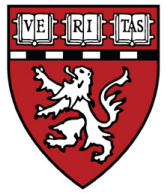
Algorithm discovery: neuroscience + behavior + computational models

http://klab.tch.harvard.edu
gabriel.kreiman@tch.harvard.edu

Scan to download
papers+data+code

CENTER FOR
Brains
Minds+
Machines





Recurrent computations to the rescue

Gabriel Kreiman

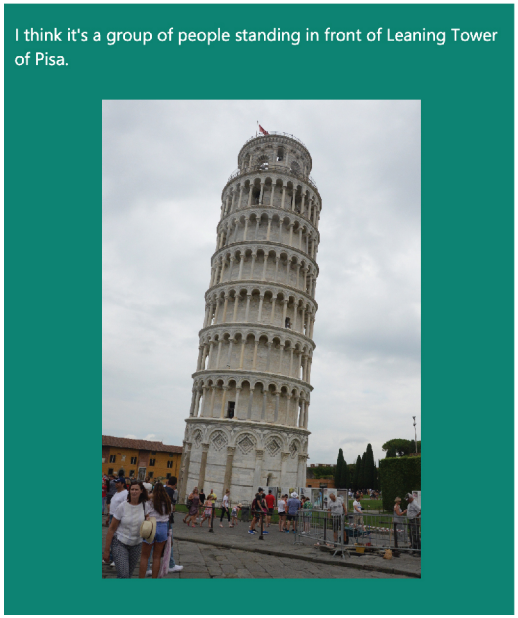
gabriel.kreiman@tch.harvard.edu

http://klab.tch.harvard.edu

Scan to download
papers+data+code

A long and exciting way forward

A



Microsoft Captioning Bot
https://www.captionbot.ai/

Kreiman and Serre, 2020
Beyond the feedforward sweep

The what, where, and when of contextual modulation

a Full context