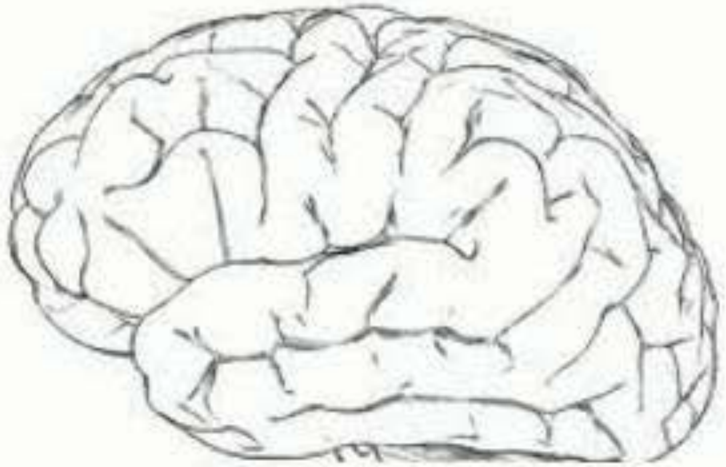
The Thousand Brains Theory of Intelligence

A framework for understanding the neocortex and building intelligent machines



Microsoft
February 21, 2019

Jeff Hawkins
Subutai Ahmad

Numenta

Numenta

Mission

1) Reverse engineer the neocortex
   - biologically accurate theories
   - test via empirical data and simulation
   - all our research is published and open

2) Apply neocortical theory to AI
   - improve current techniques
   - move toward truly intelligent systems

The Human Neocortex

75% of brain's volume

Organ of intelligence

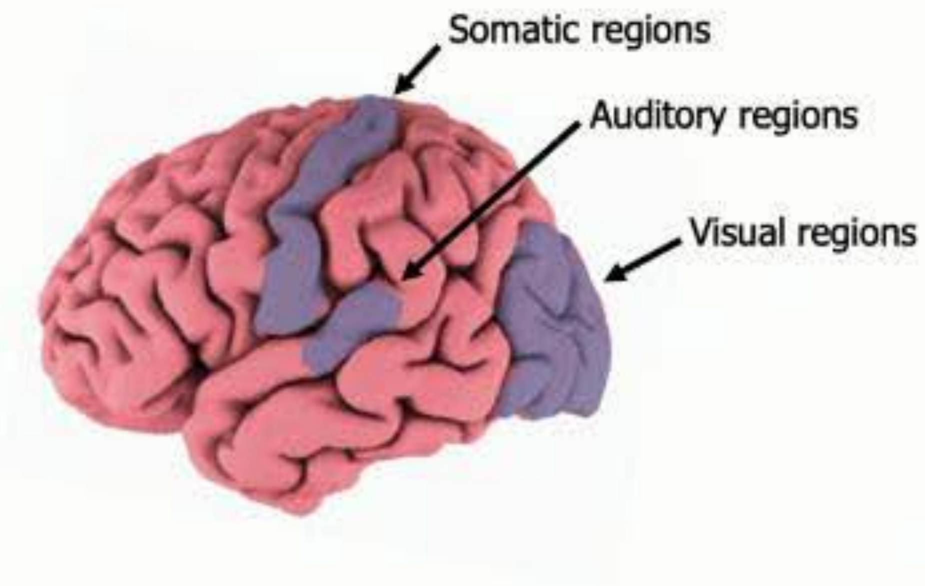
Q. What does the neocortex do?

A. The neocortex learns a model of the world

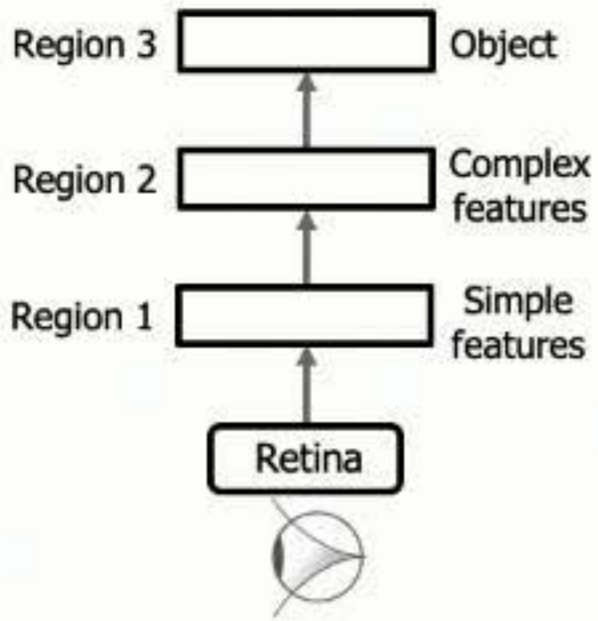
- Thousands of objects, how they look, feel, and sound
- Where objects are located
- How objects behave
- Physical and abstract objects

The model is predictive, and creates goal-oriented behaviors

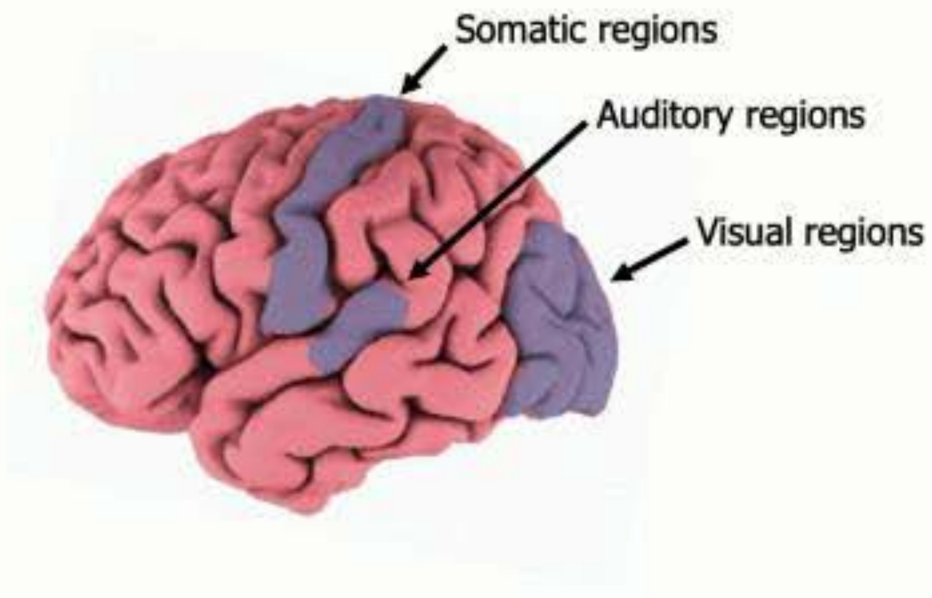
Regions and Hierarchy

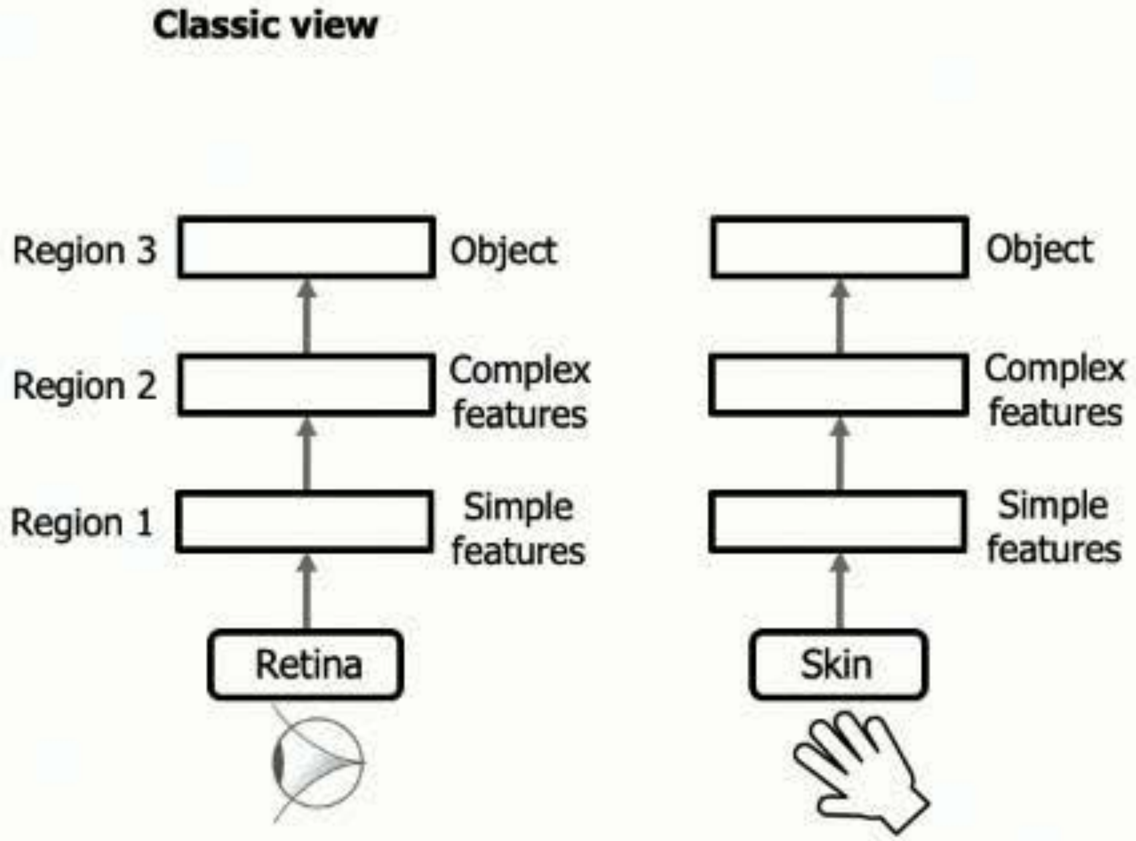


Classic view

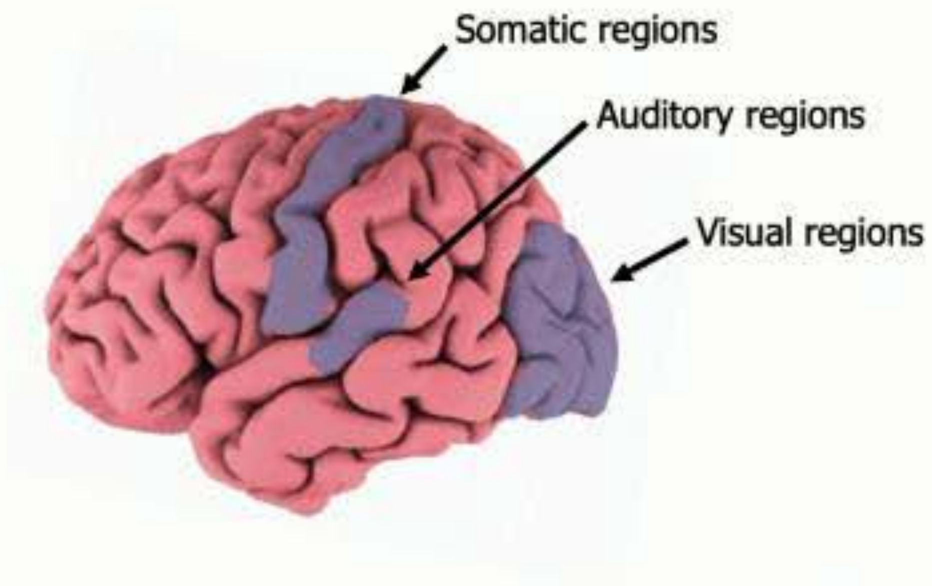


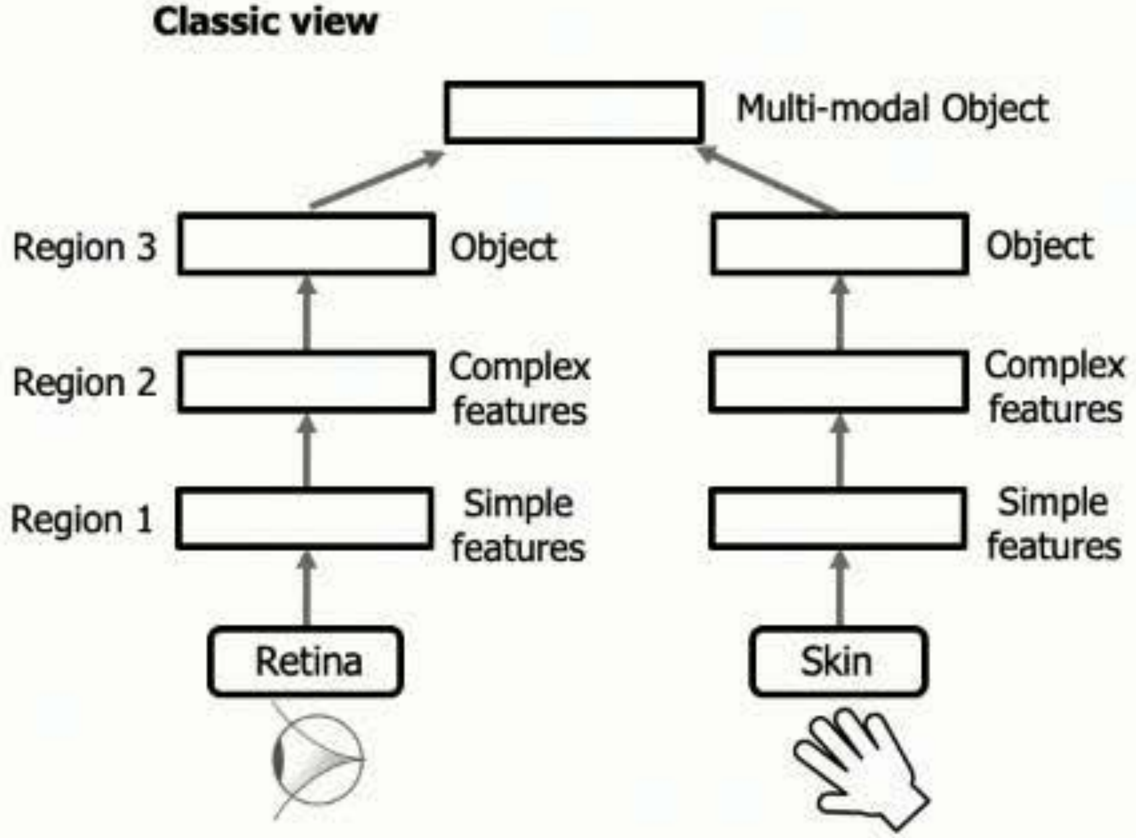
Regions and Hierarchy



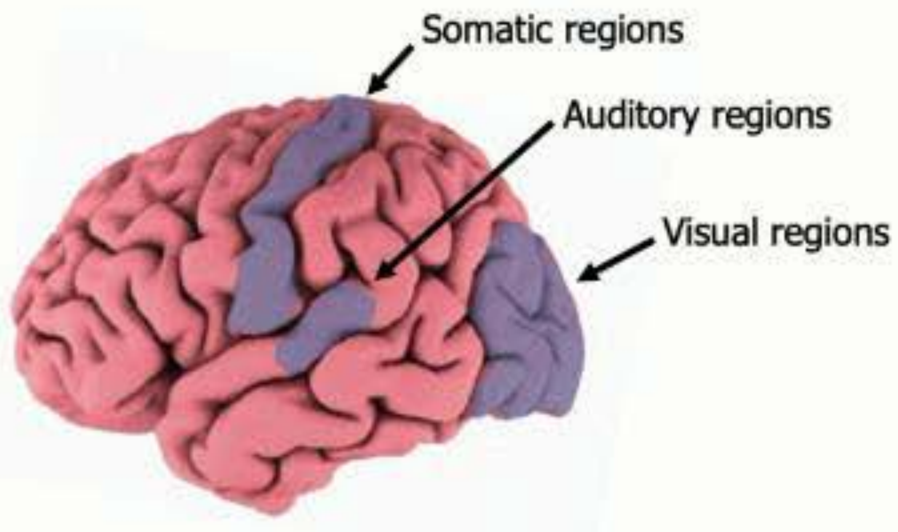


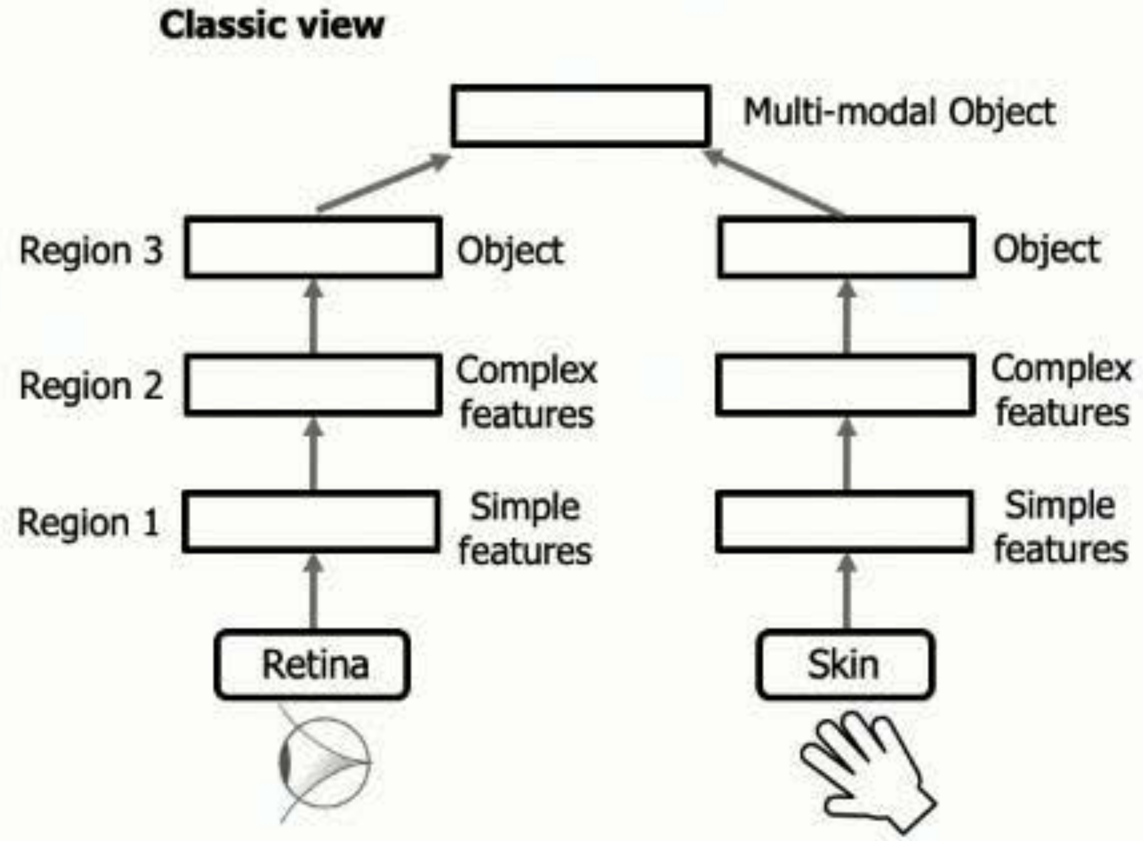
Regions and Hierarchy

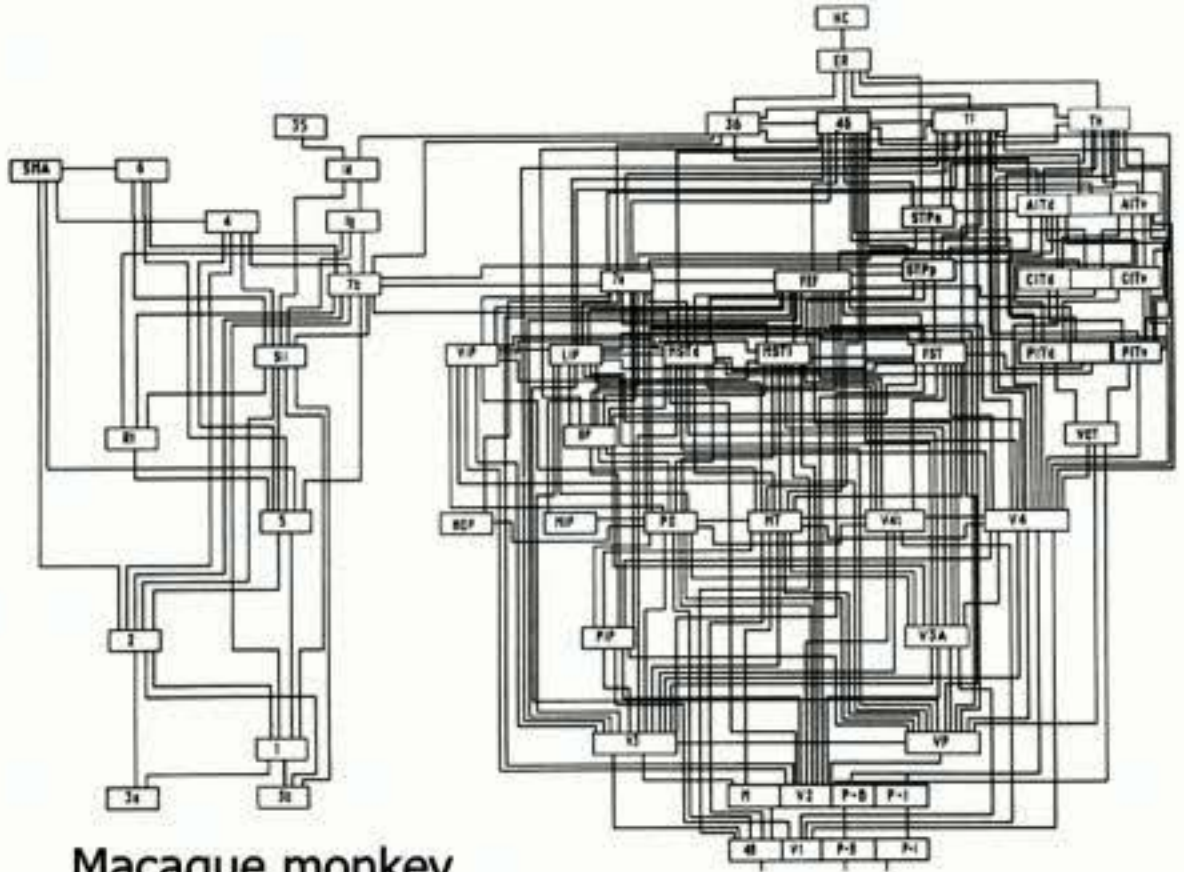




Regions and Hierarchy







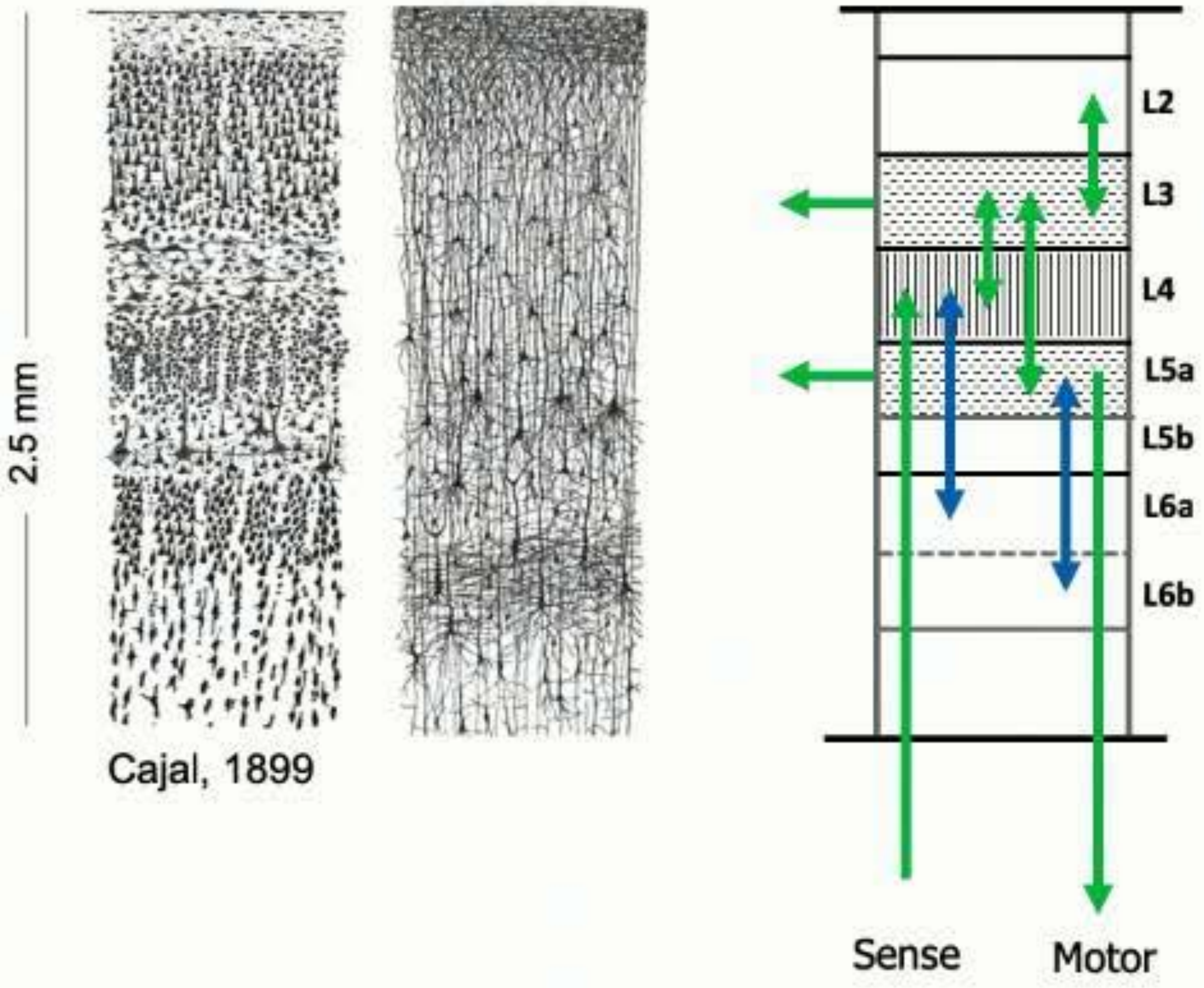
Macaque monkey

Felleman, van Essen, 1991

Most connections between regions are not hierarchical

- 40% of all possible connections exist
- Many regions get input from ten or more other regions

Local Circuits



Cajal, 1899

Dozens of neuron types

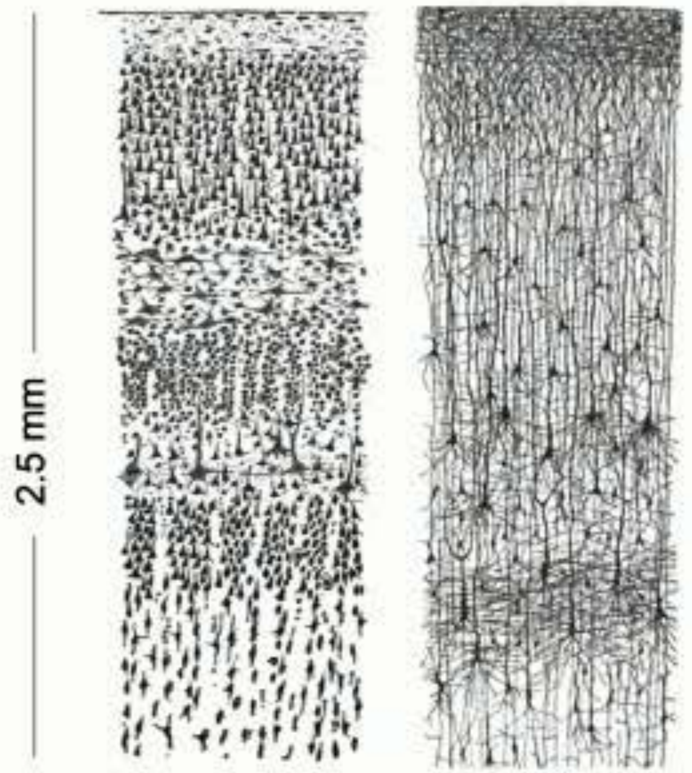
Organized in layers

Local projections cross all layers

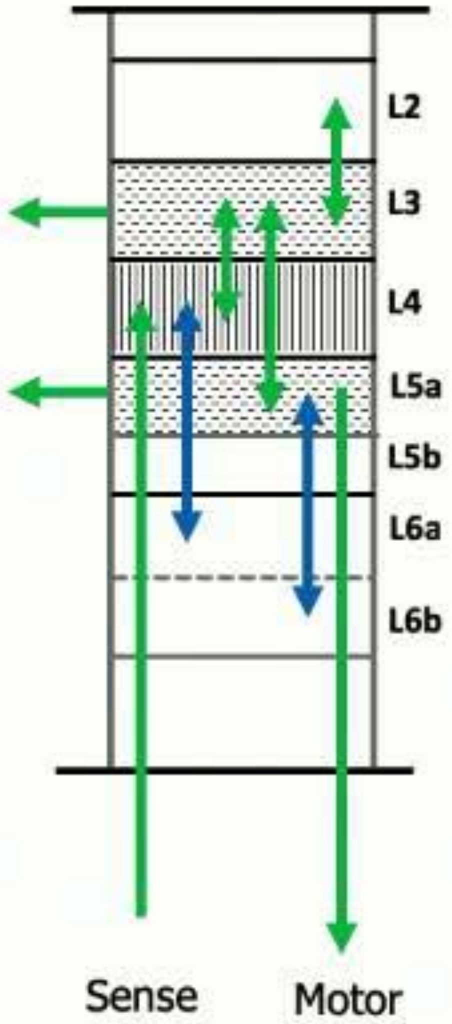
Long-distance horizontal projections in some layers

All regions have a motor output

Local Circuits



Cajal, 1899



Dozens of neuron types

Organized in layers

Local projections cross all layers

Long-distance horizontal projections in some layers

All regions have a motor output

Remarkably the same in every region

Complex circuit → complex function

Vernon Mountcastle's Big Idea

1) All areas of the neocortex look the same because they perform the same intrinsic function.

2) What makes one region visual and another auditory is what it is connected to.

3) A "cortical column" ($1mm^2$) is the unit of replication.

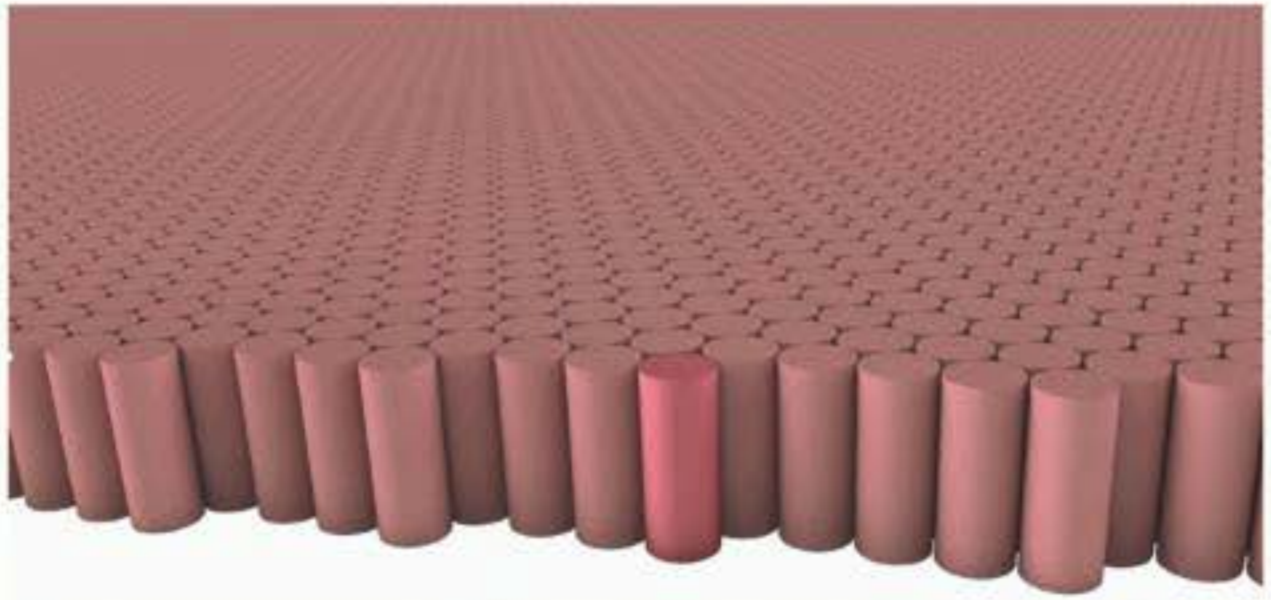
Mountcastle, 1978

Vernon Mountcastle's Big Idea

1) All areas of the neocortex look the same because they perform the same intrinsic function.

2) What makes one region visual and another auditory is what it is connected to.

3) A "cortical column" ($1mm^2$) is the unit of replication.



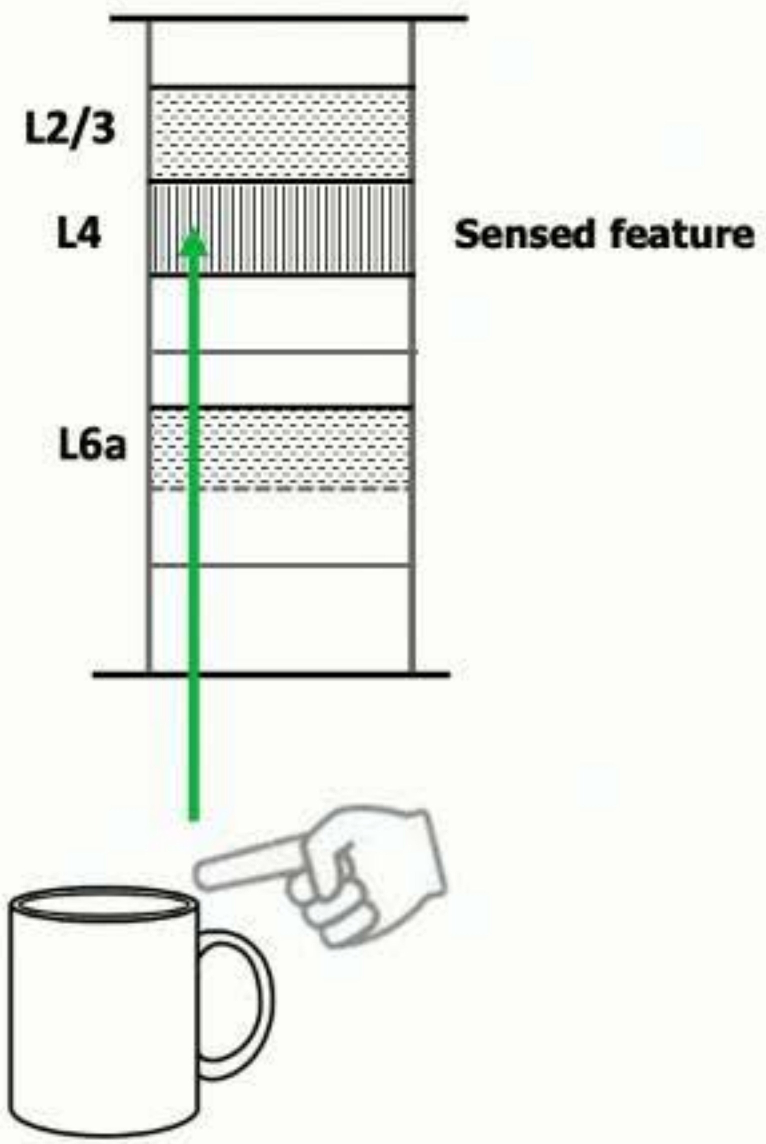
Corollary:

Every column must perform the same functions as the entire neocortex.

Mountcastle, 1978

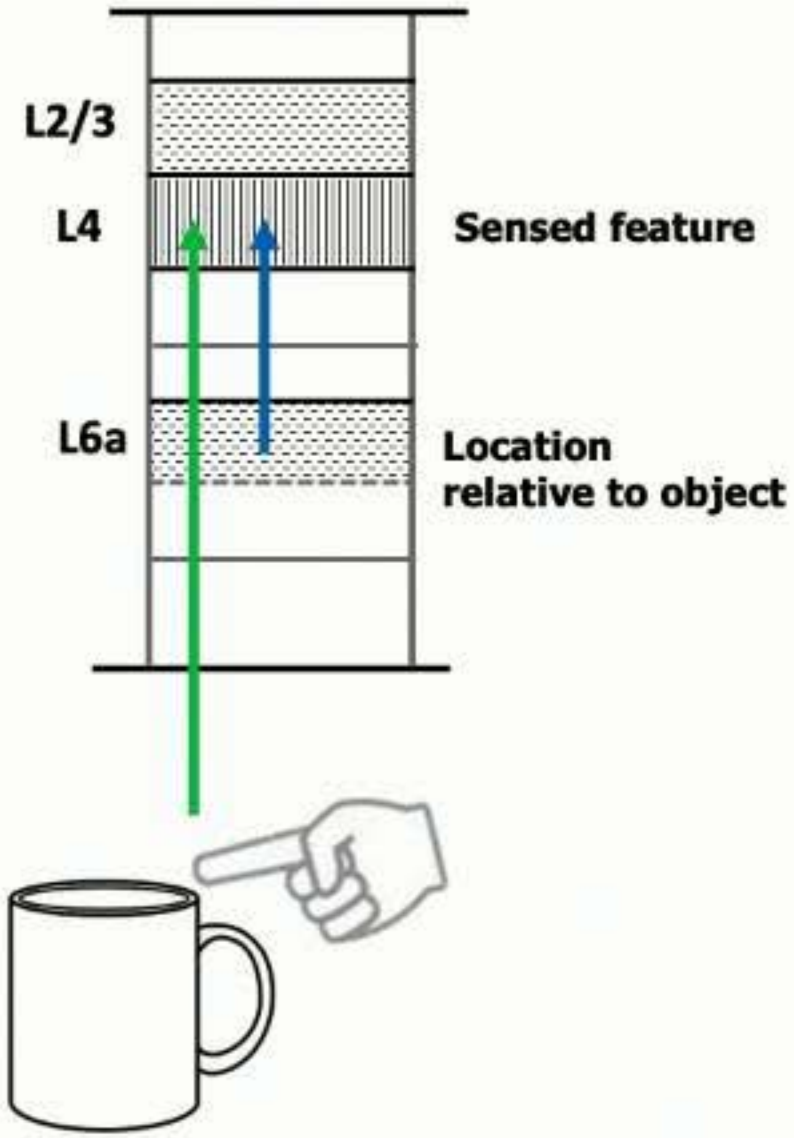
Thought Experiment

"A Theory of How Columns in the Neocortex Enable Learning the Structure of the World" (Hawkins, et. al., 2017)



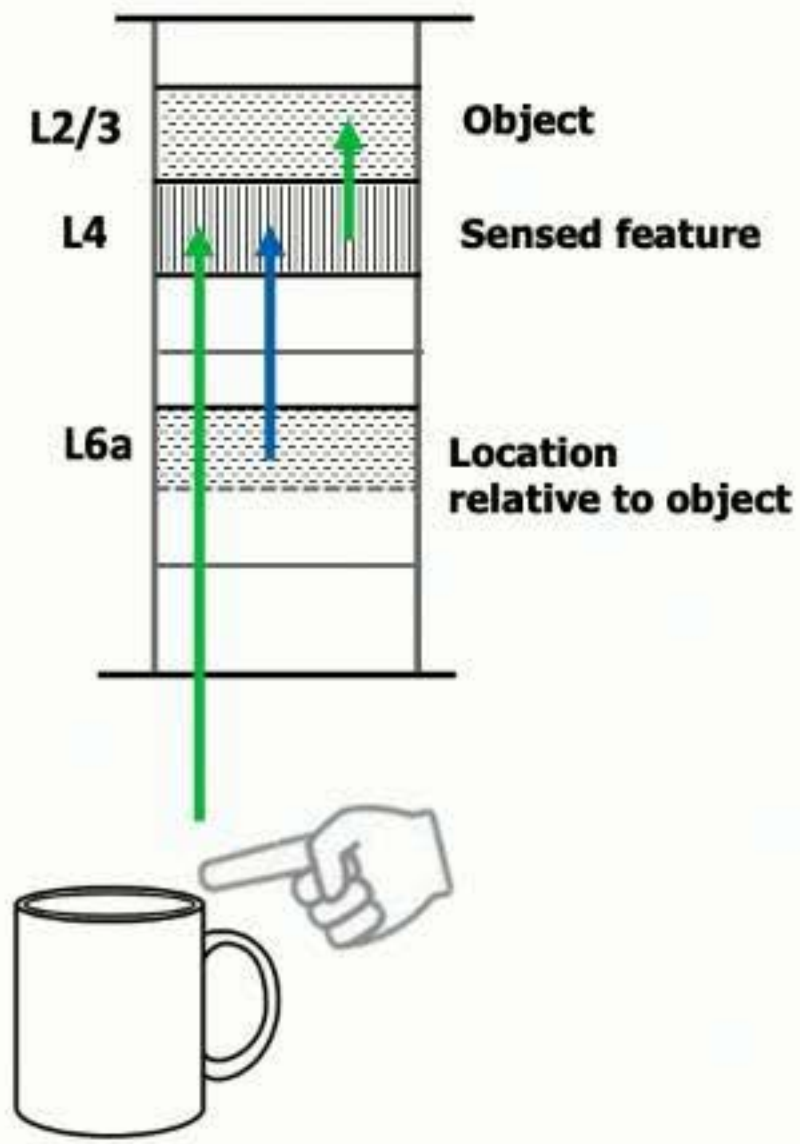
A single column learns completes models of objects by integrating features and locations over time.

"A Theory of How Columns in the Neocortex Enable Learning the Structure of the World" (Hawkins, et. al., 2017)



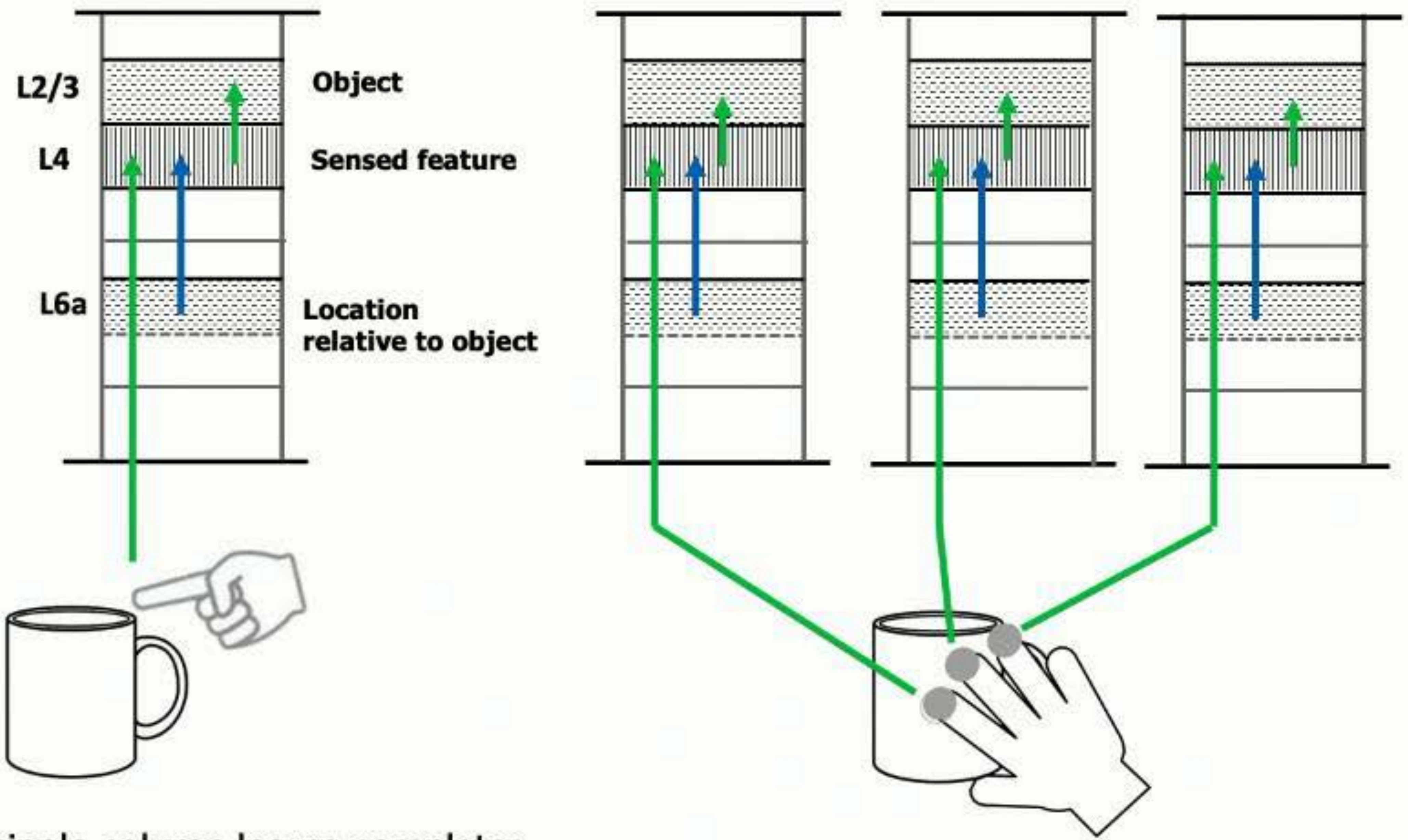
A single column learns completes models of objects by integrating features and locations over time.

"A Theory of How Columns in the Neocortex Enable Learning the Structure of the World" (Hawkins, et. al., 2017)



A single column learns completes models of objects by integrating features and locations over time.

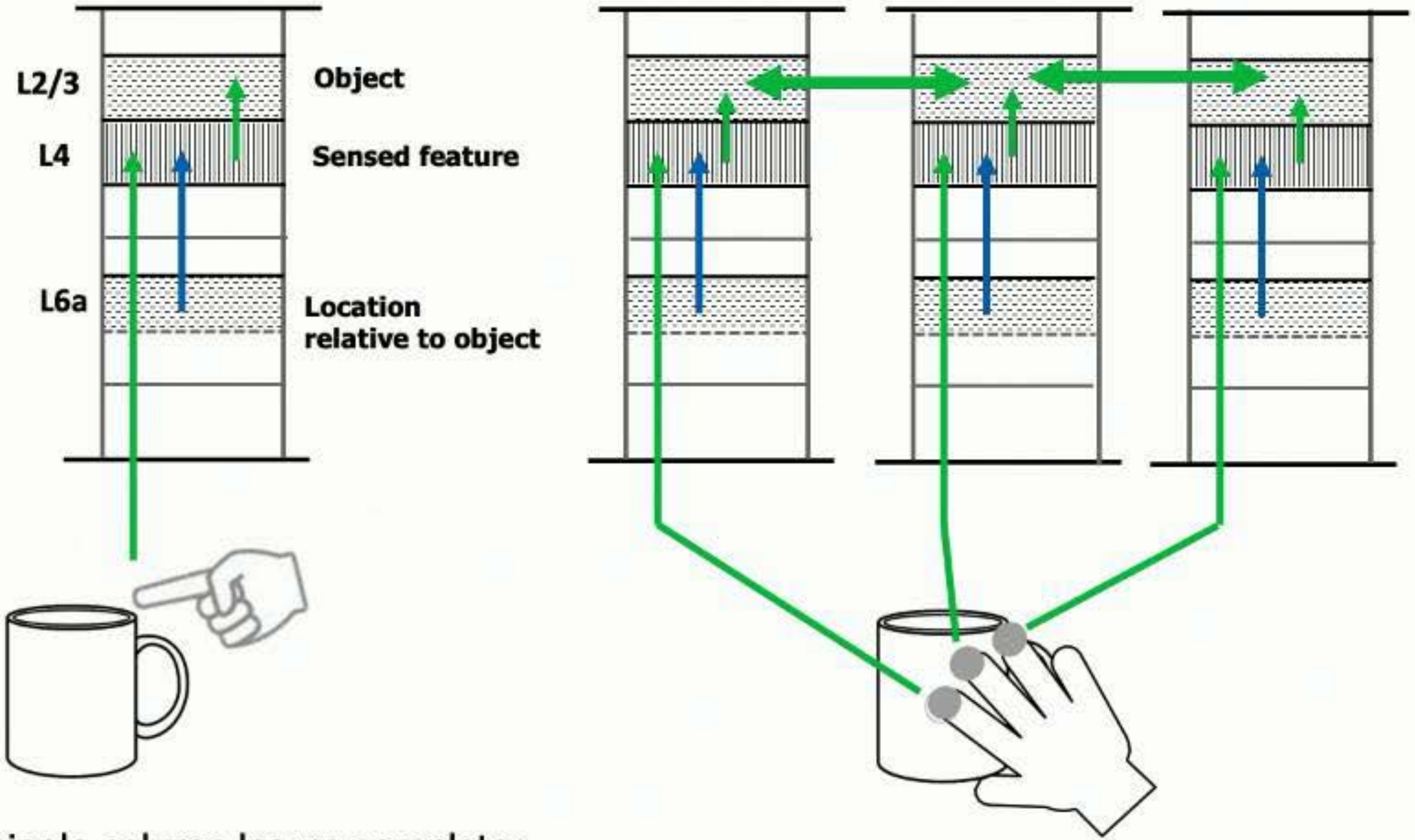
“A Theory of How Columns in the Neocortex Enable Learning the Structure of the World” (Hawkins, et. al., 2017)



A single column learns completes models of objects by integrating features and locations over time.

Multiple columns can infer objects in a single sensation by “voting” on object identity.

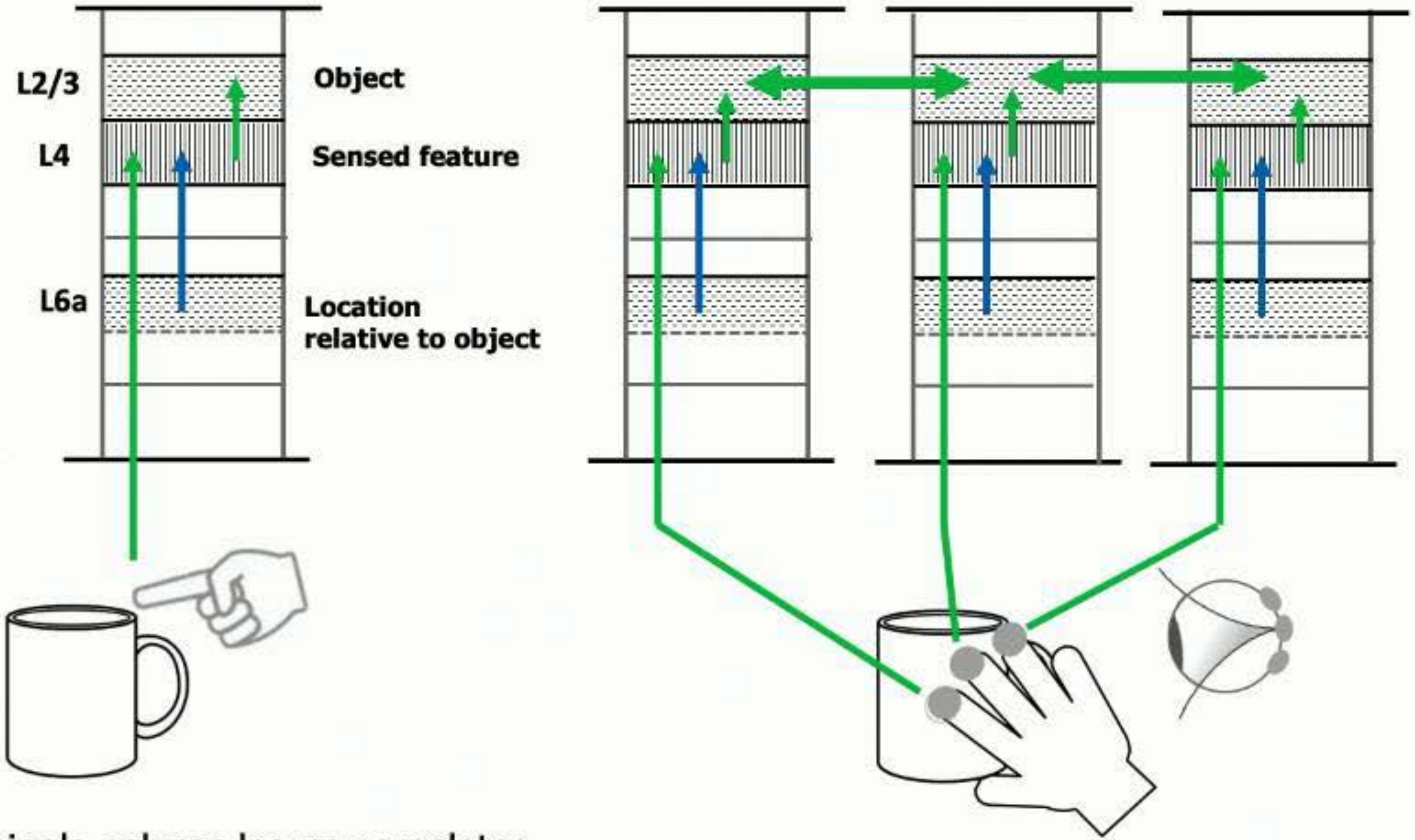
“A Theory of How Columns in the Neocortex Enable Learning the Structure of the World” (Hawkins, et. al., 2017)



A single column learns completes models of objects by integrating features and locations over time.

Multiple columns can infer objects in a single sensation by “voting” on object identity.

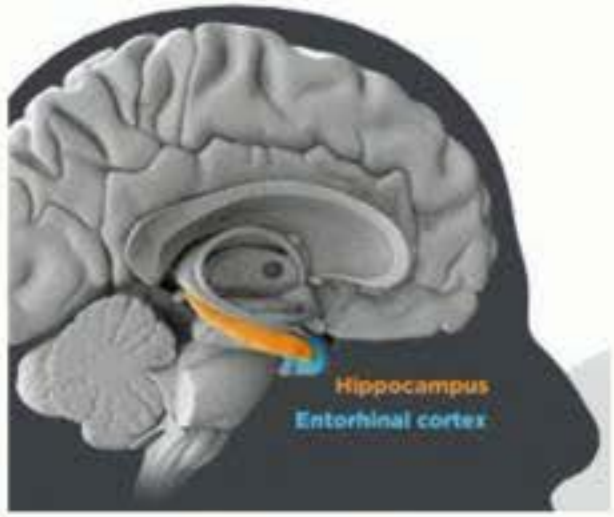
"A Theory of How Columns in the Neocortex Enable Learning the Structure of the World" (Hawkins, et. al., 2017)



A single column learns completes models of objects by integrating features and locations over time.

Multiple columns can infer objects in a single sensation by "voting" on object identity.

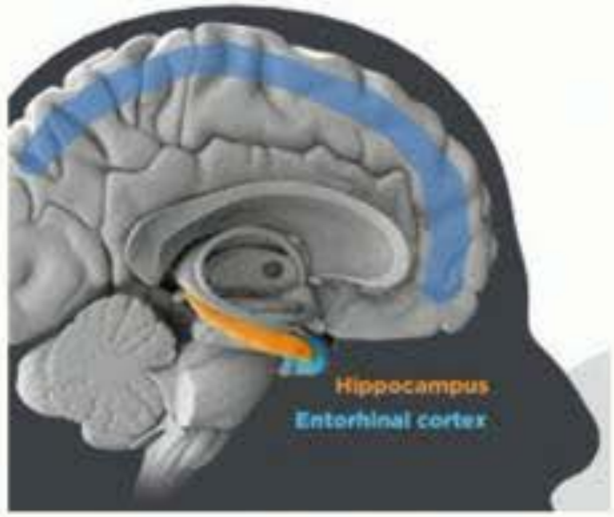
Reference Frames in the Brain



"Grid cells" in entorhinal cortex
- Create reference frames for environments
- Represent location of body
- Needed for mapping environments and moving body

Moser, 2005

Reference Frames in the Brain



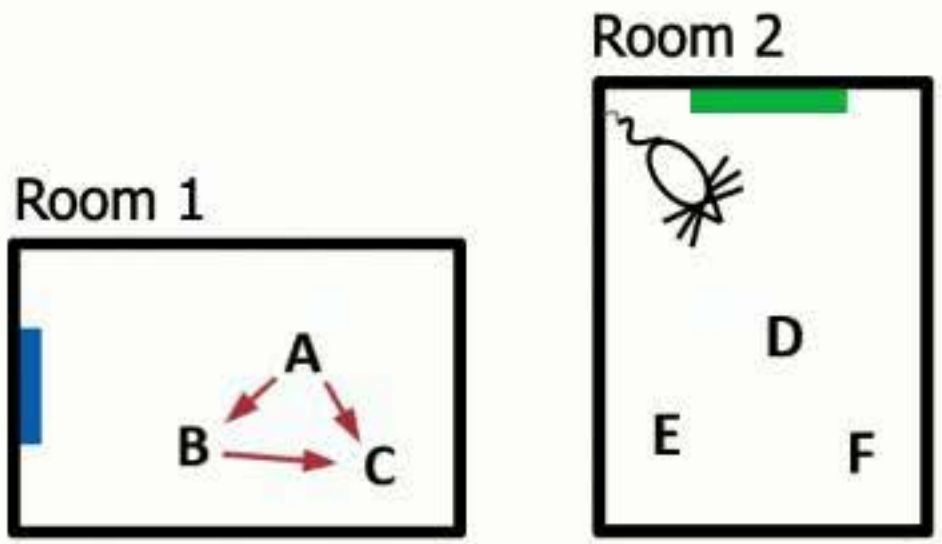
"Grid cells" in entorhinal cortex
- **Create reference frames for environments**
- **Represent location of body**
- **Needed for mapping environments and moving body**

Moser, 2005

Grid cells exist in every cortical column (hypothesis)
- **Create reference frames for objects**
- **Represent location of column's input**
- **Needed for learning the structure objects and moving limbs**

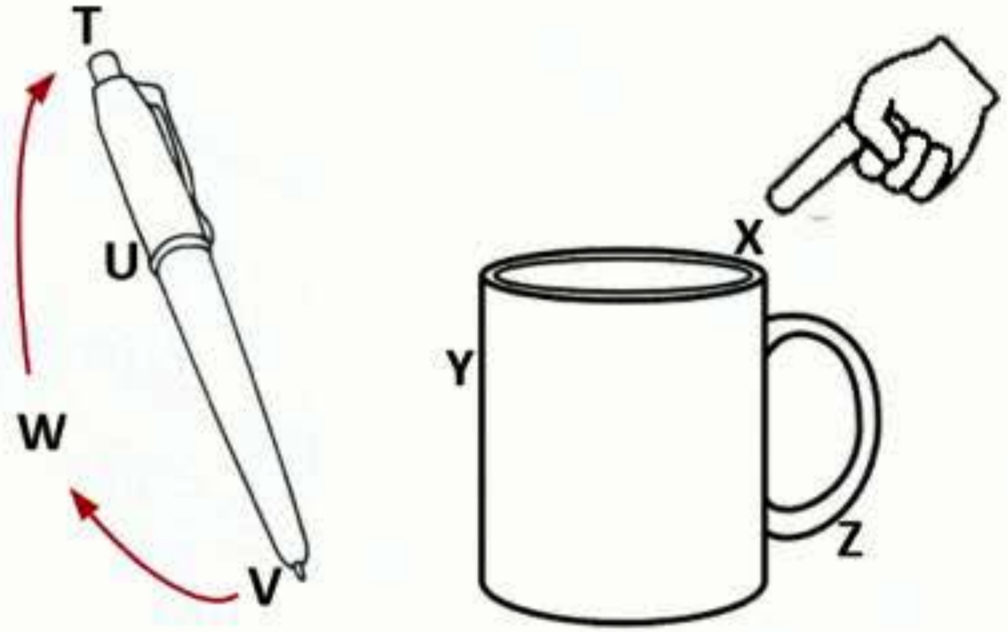
Hawkins et. al., 2017
Hawkins et. al., 2018
Lewis et. al., 2018

Entorhinal Cortex

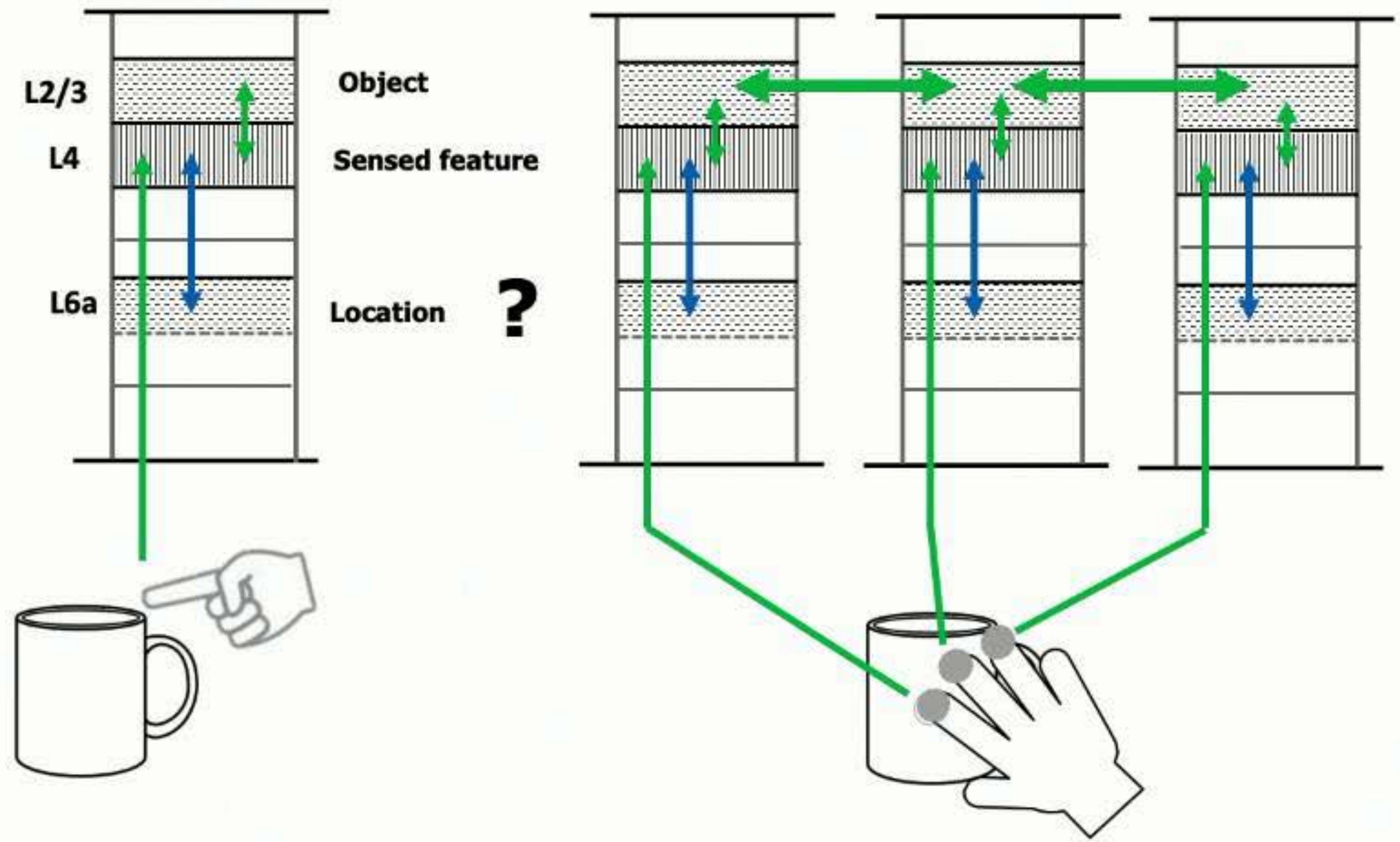


Grid cells represent location of body relative to room.

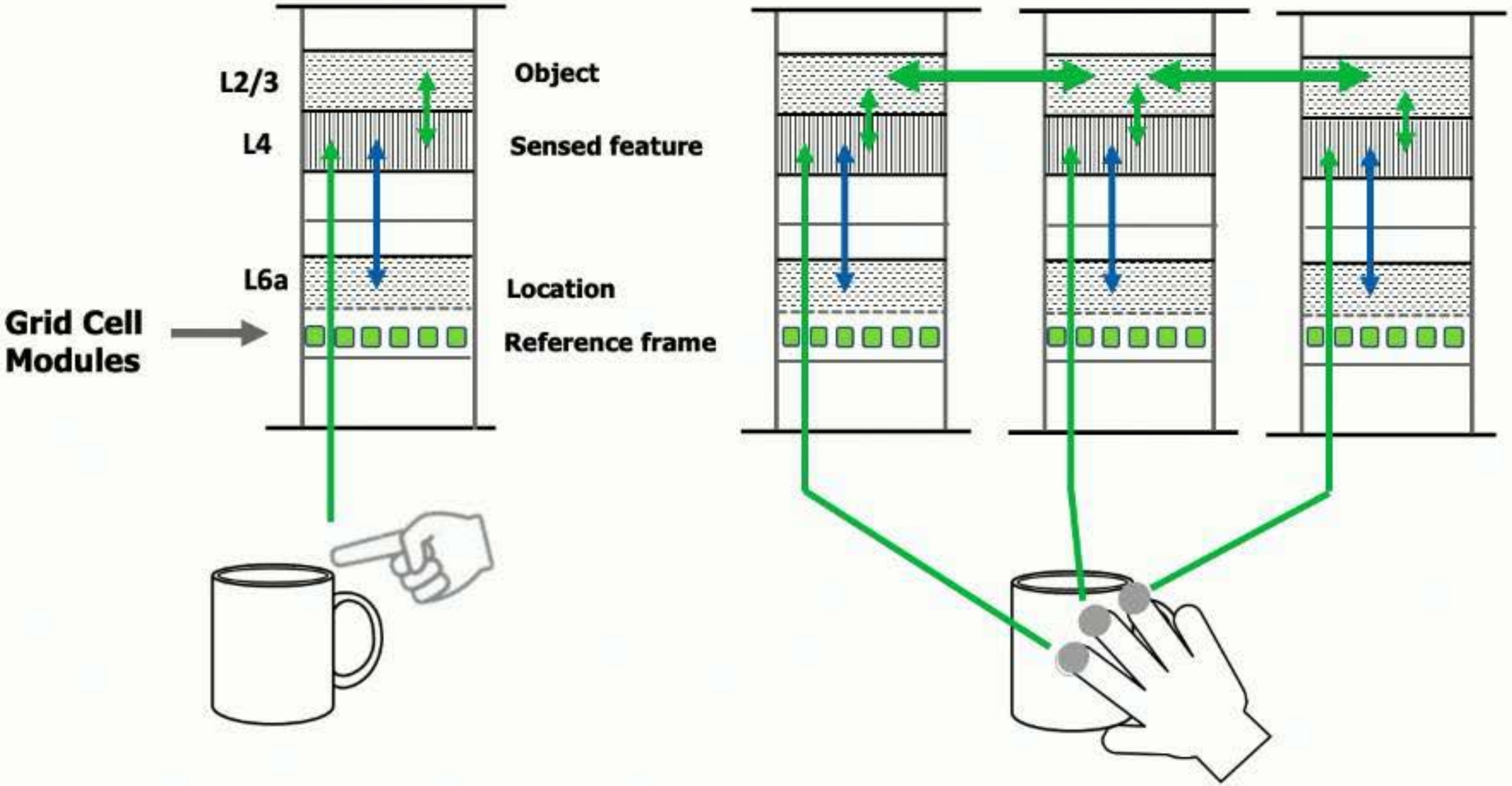
Neocortex



Grid cells represent location of sensor relative to object.



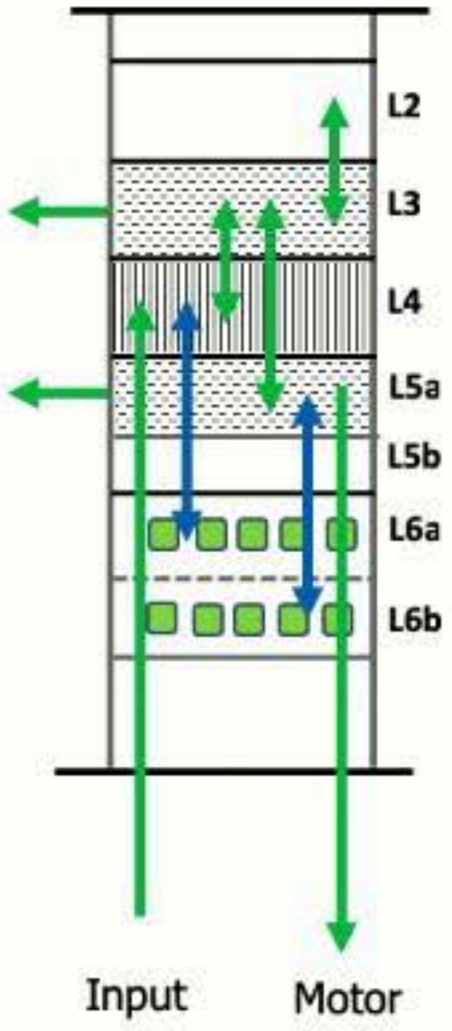
L2/3
L4
L6a
Object
Sensed feature
Location
?



Lewis et. al., 2018

Cortical Columns Are Complete Sensory-motor Modeling Systems

Hawkins et. al., 2018
Klukas et. al., 2019

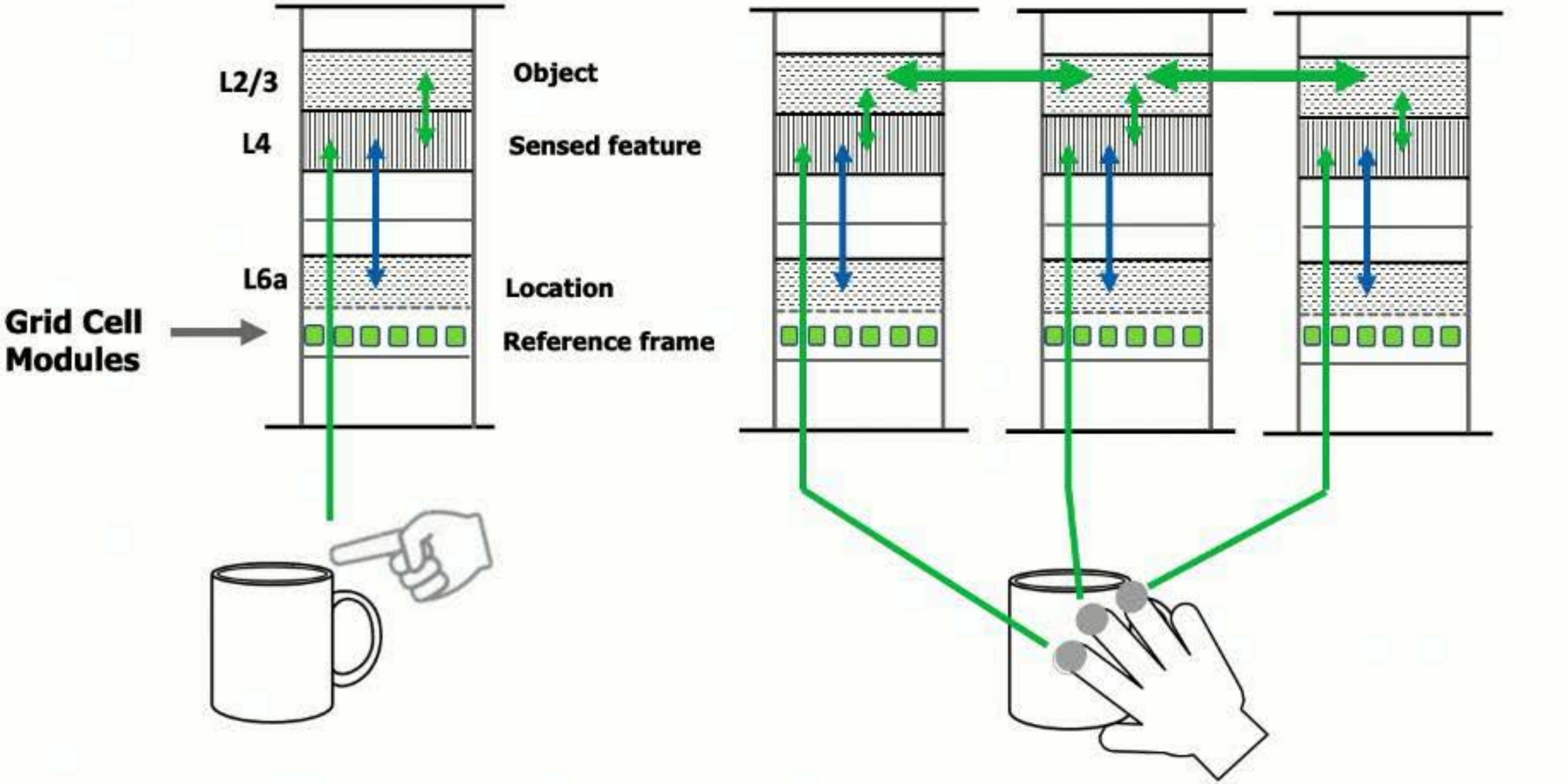


Two reference frames

Learns:
- **Dimensionality of object**
- **Morphology**
- **Changes in morphology** (how objects behave)
- **Compositional and Recursive structure**

Generates motor behaviors

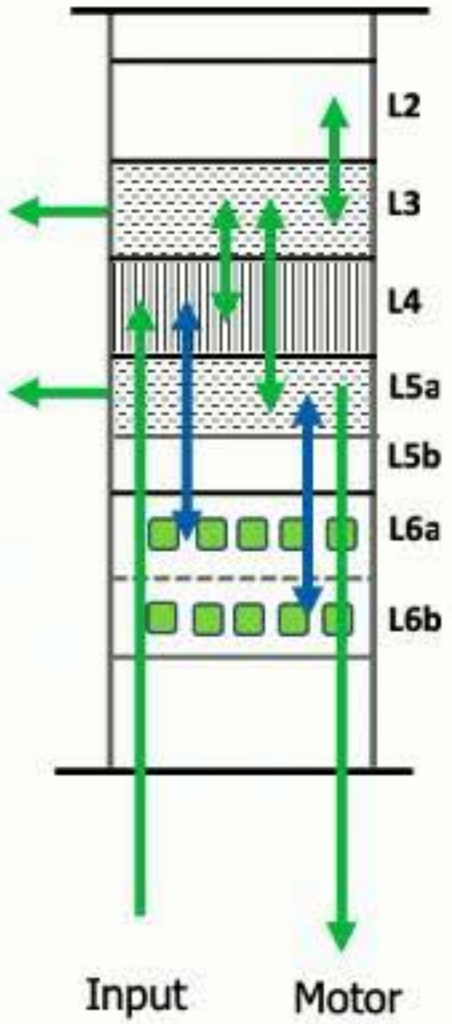
Applies to:
- **Physical objects**
- **Abstract objects**



Lewis et. al., 2018

Cortical Columns Are Complete Sensory-motor Modeling Systems

Hawkins et. al., 2018
Klukas et. al., 2019



Two reference frames

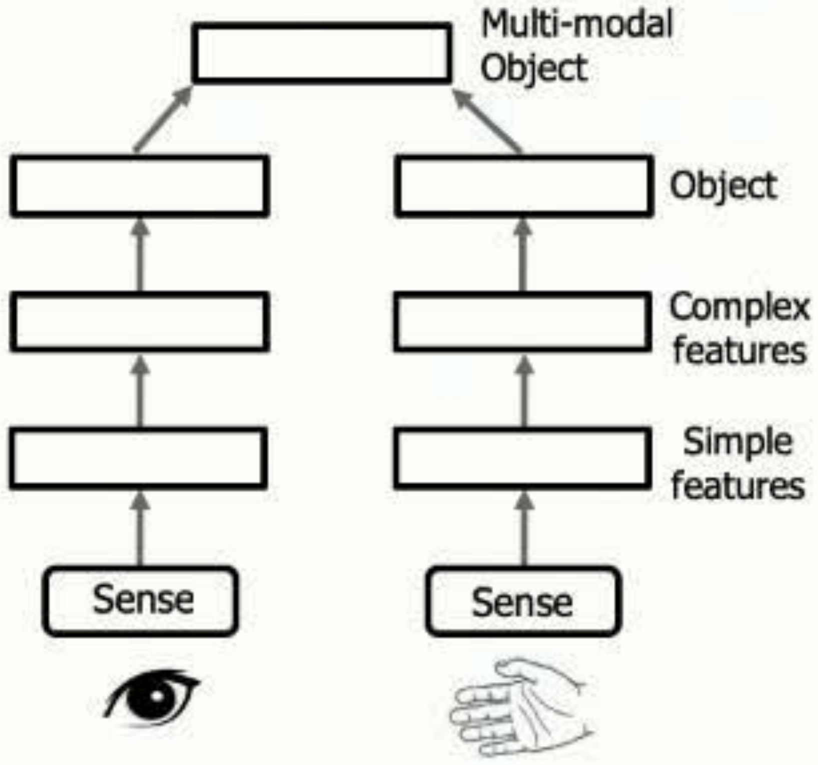
Learns:
- **Dimensionality of object**
- **Morphology**
- **Changes in morphology** (how objects behave)
- **Compositional and Recursive structure**

Generates motor behaviors

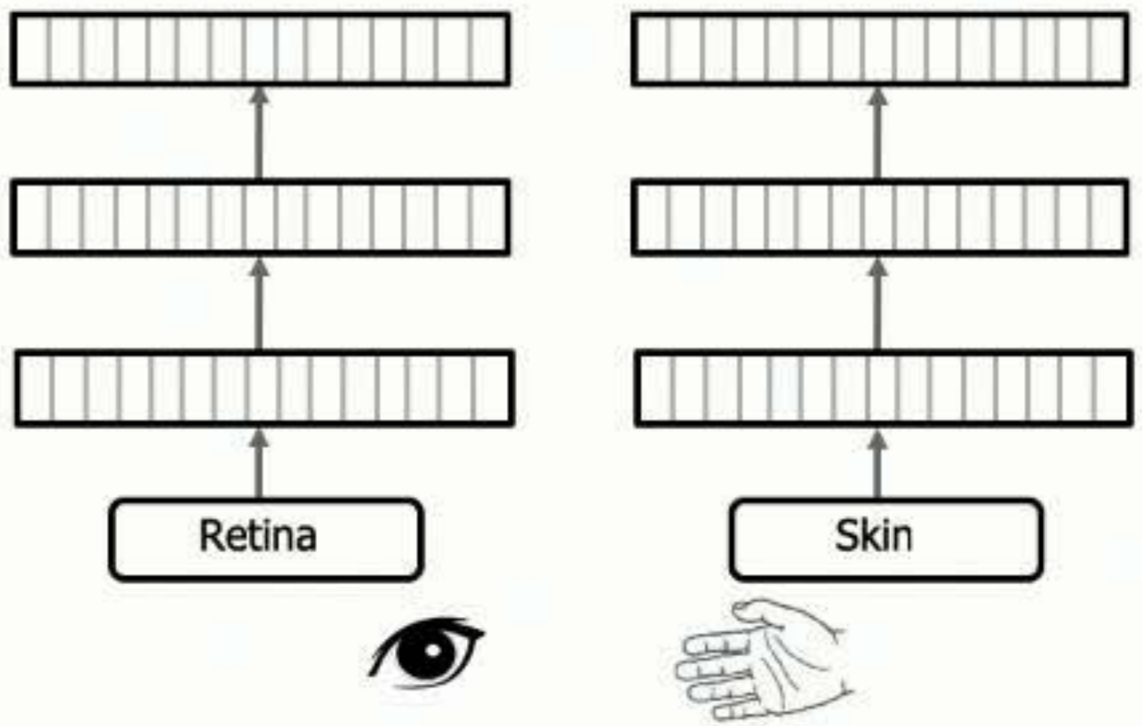
Applies to:
- **Physical objects**
- **Abstract objects**

The Thousand Brains Theory of Intelligence

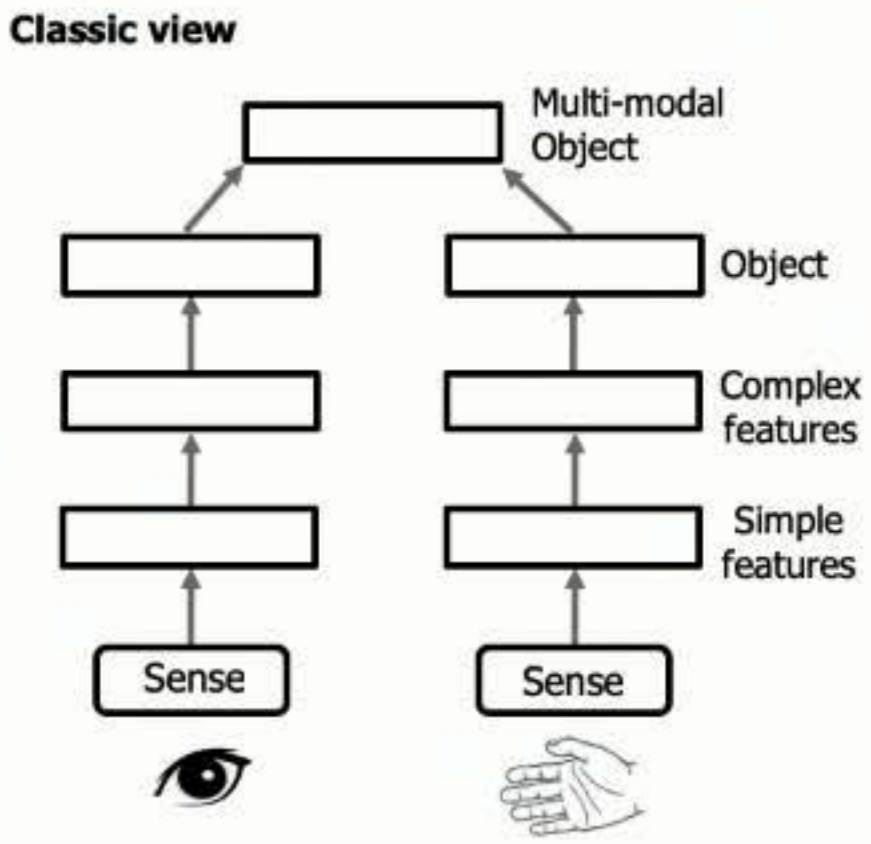
Classic view

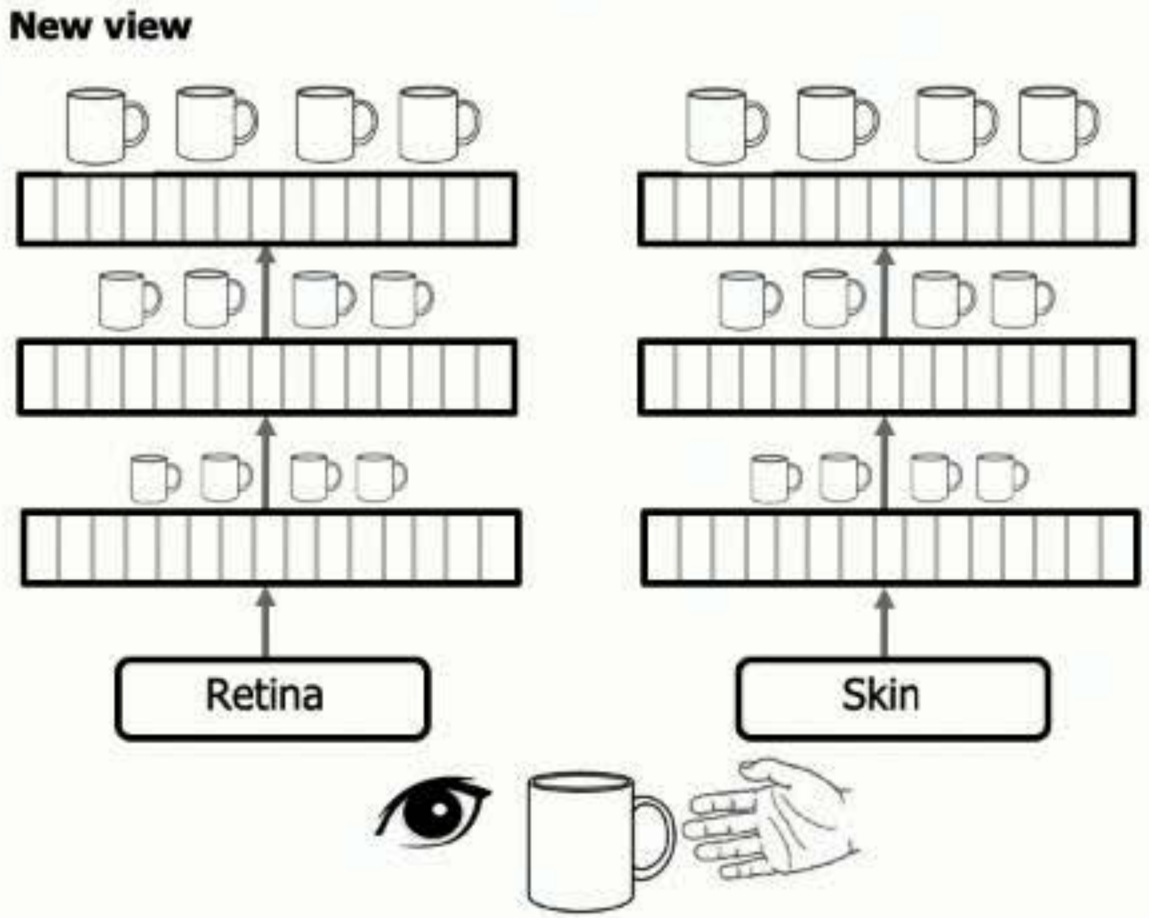


New view



The Thousand Brains Theory of Intelligence



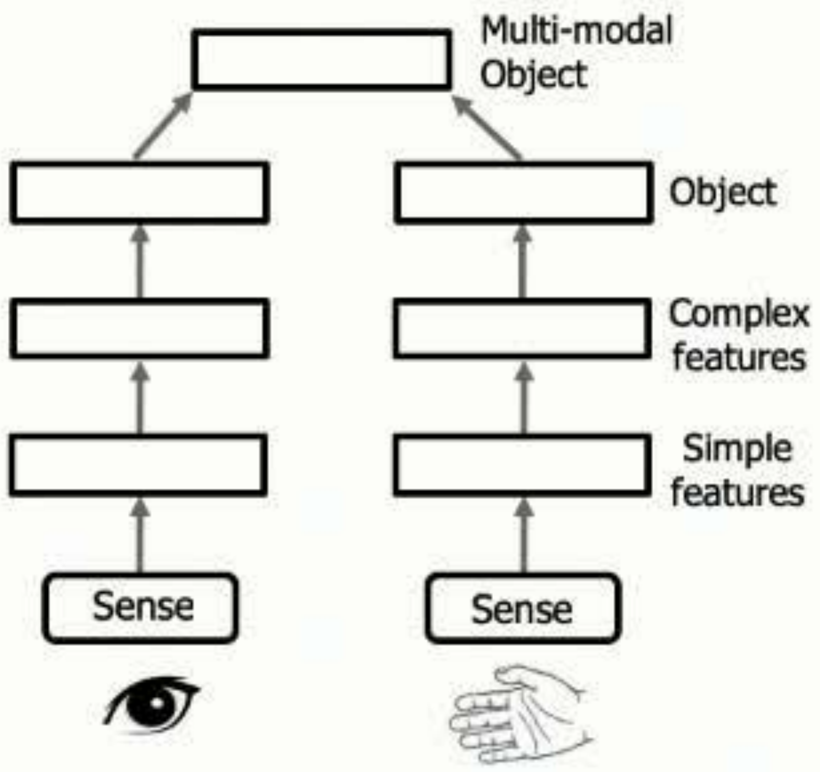


Many models of every object

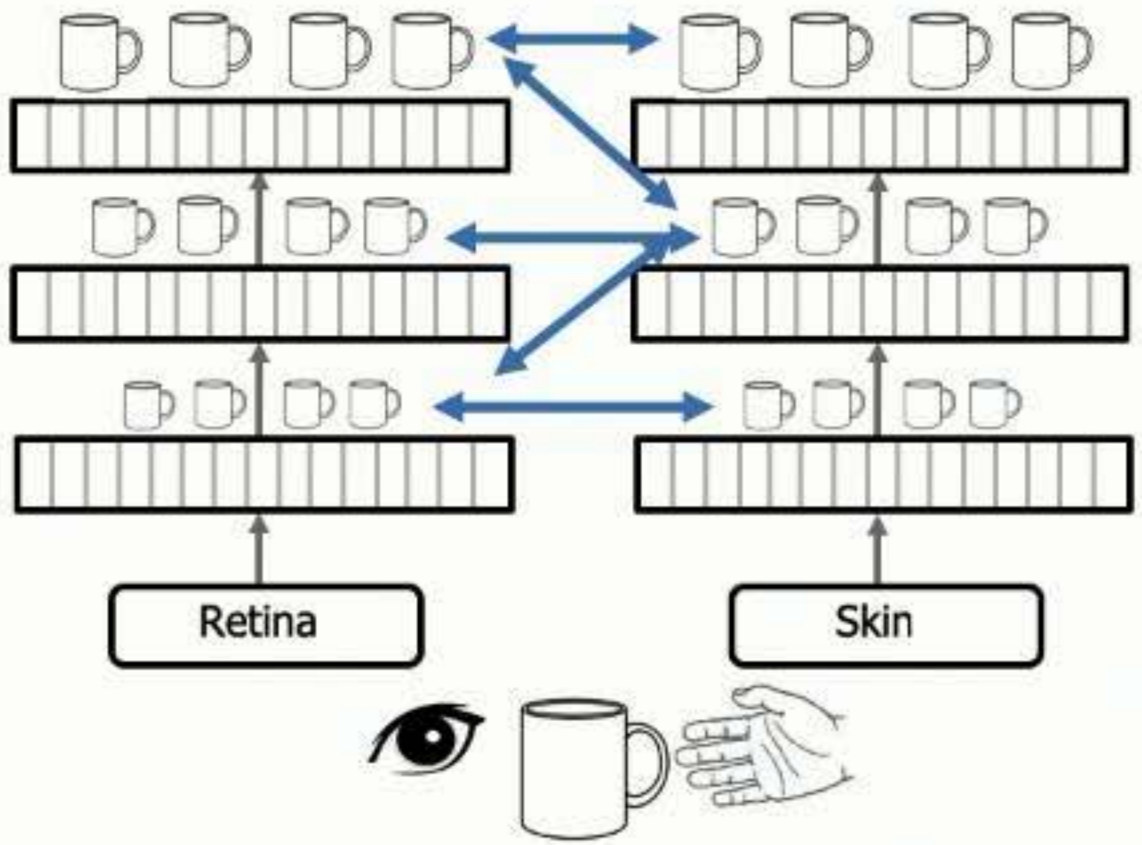
Models differ based on input

The Thousand Brains Theory of Intelligence

Classic view



New view



Many models of every object

Models differ based on input

Long-range connections
- **resolve ambiguity**
- **form a singular percept** ("sensor fusion")

Will Neocortical Principles Will Be Essential for AI?

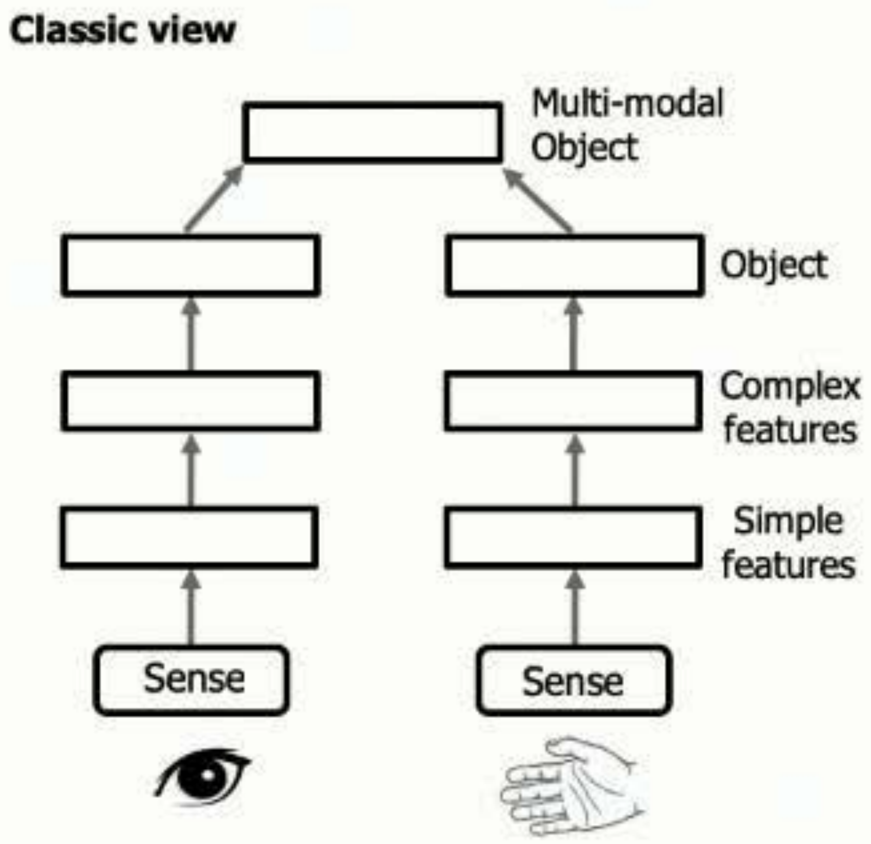
Medium and Long Term

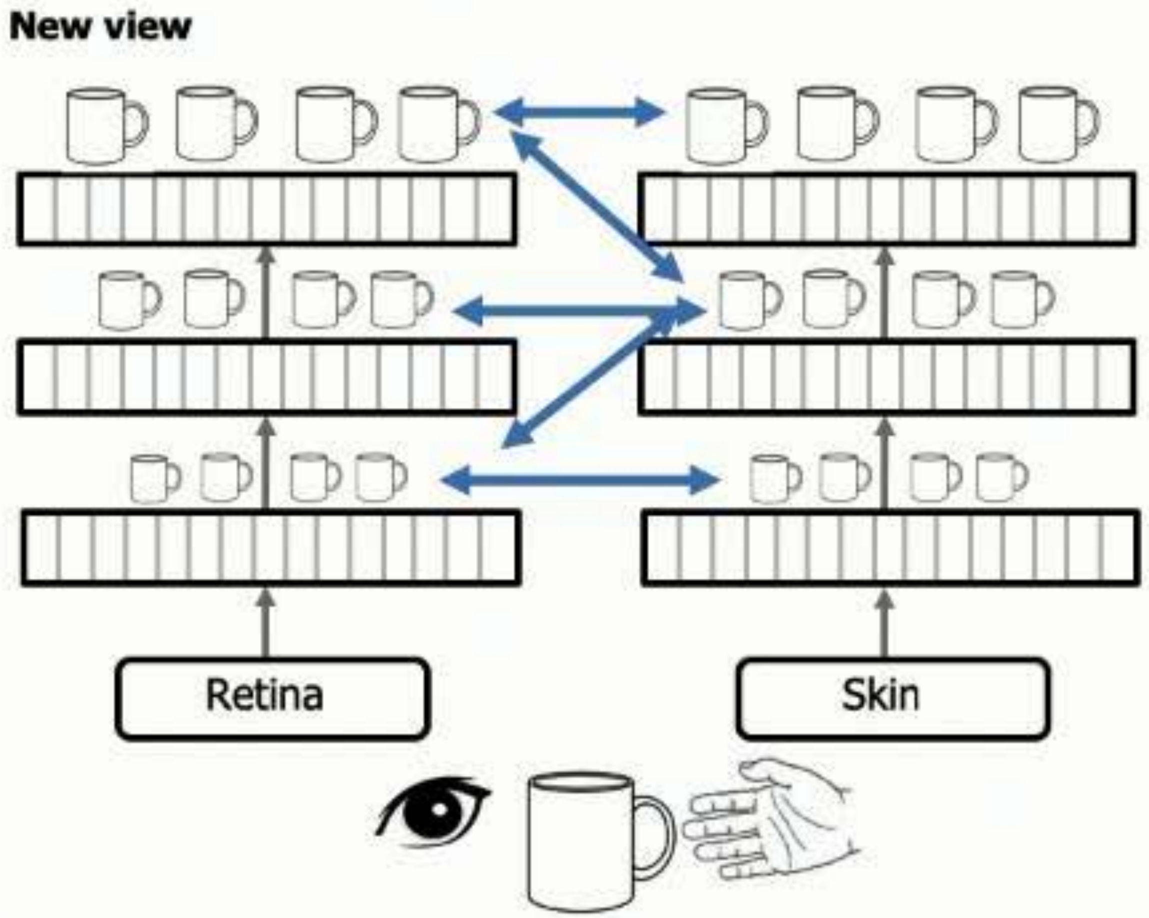
- Sensory-motor learning and inference
  (AI and Robotics are not separable)
- Models based on object-centric reference frames
- Many small models with voting

Near Term

- Sparse representations : robustness
- Predictive neuron model : continuous on-line learning

The Thousand Brains Theory of Intelligence



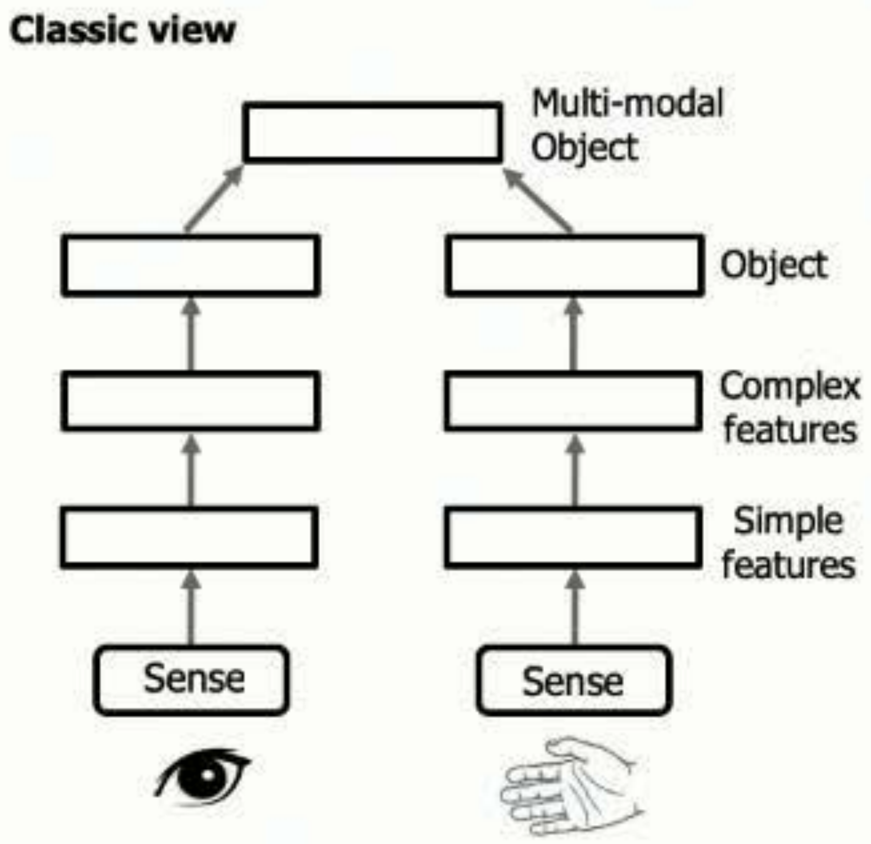


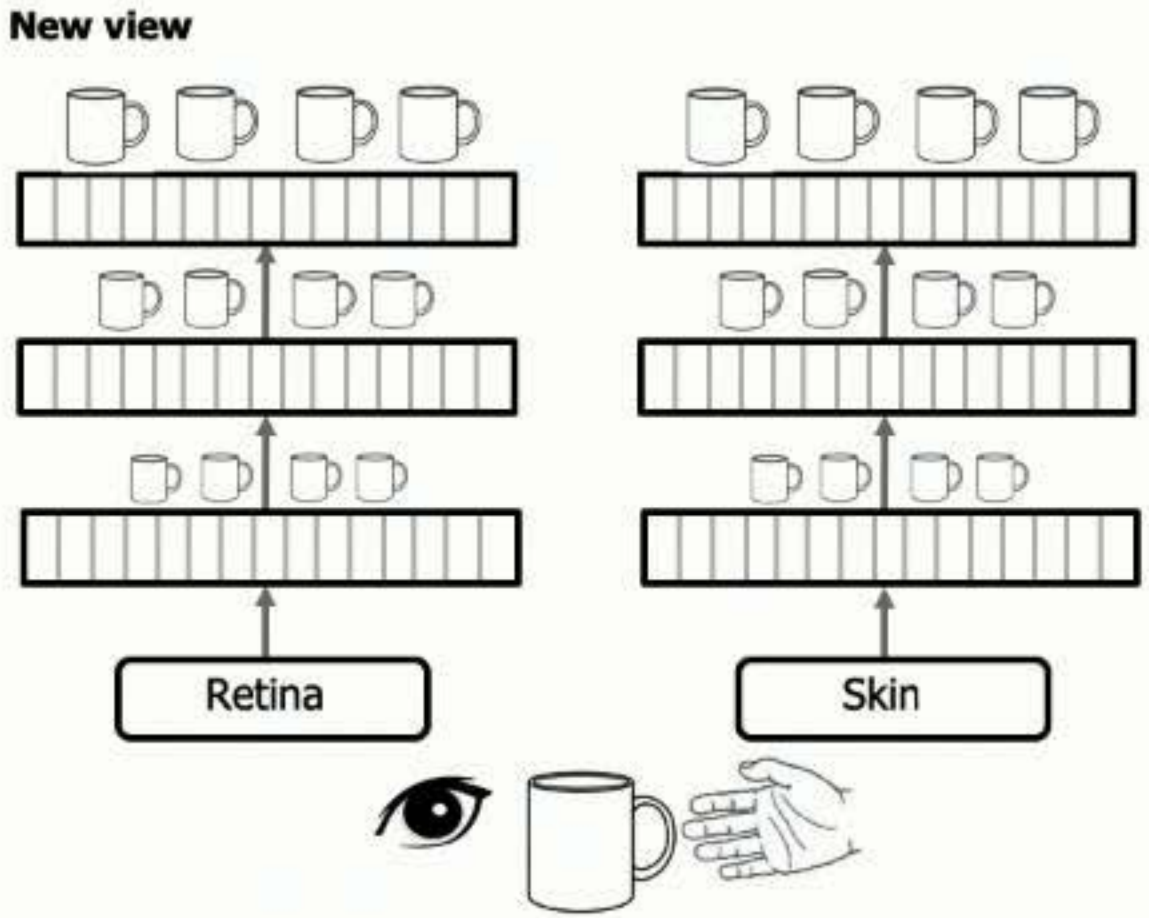
Many models of every object

Models differ based on input

Long-range connections
- **resolve ambiguity**
- **form a singular percept** ("sensor fusion")

The Thousand Brains Theory of Intelligence



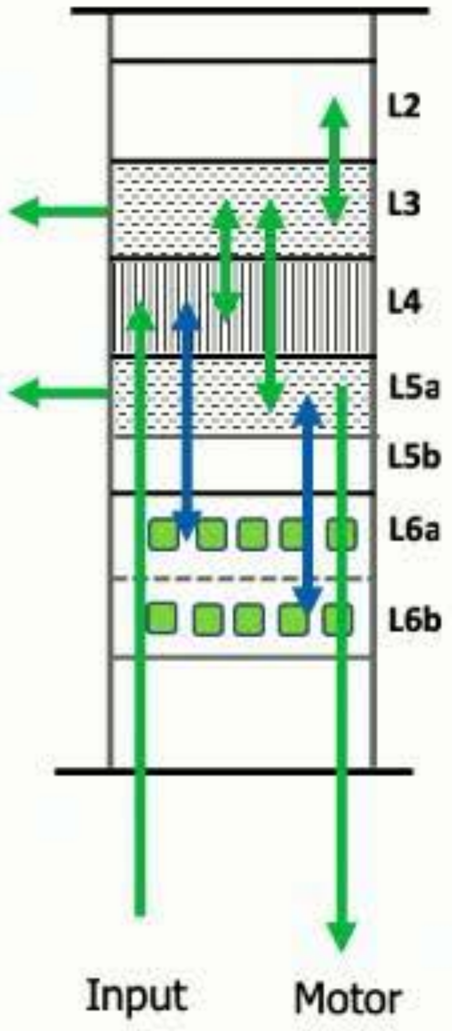


Many models of every object

Models differ based on input

Cortical Columns Are Complete Sensory-motor Modeling Systems

Hawkins et. al., 2018
Klukas et. al., 2019



Two reference frames

Learns:
- **Dimensionality of object**
- **Morphology**
- **Changes in morphology** (how objects behave)
- **Compositional and Recursive structure**

Generates motor behaviors

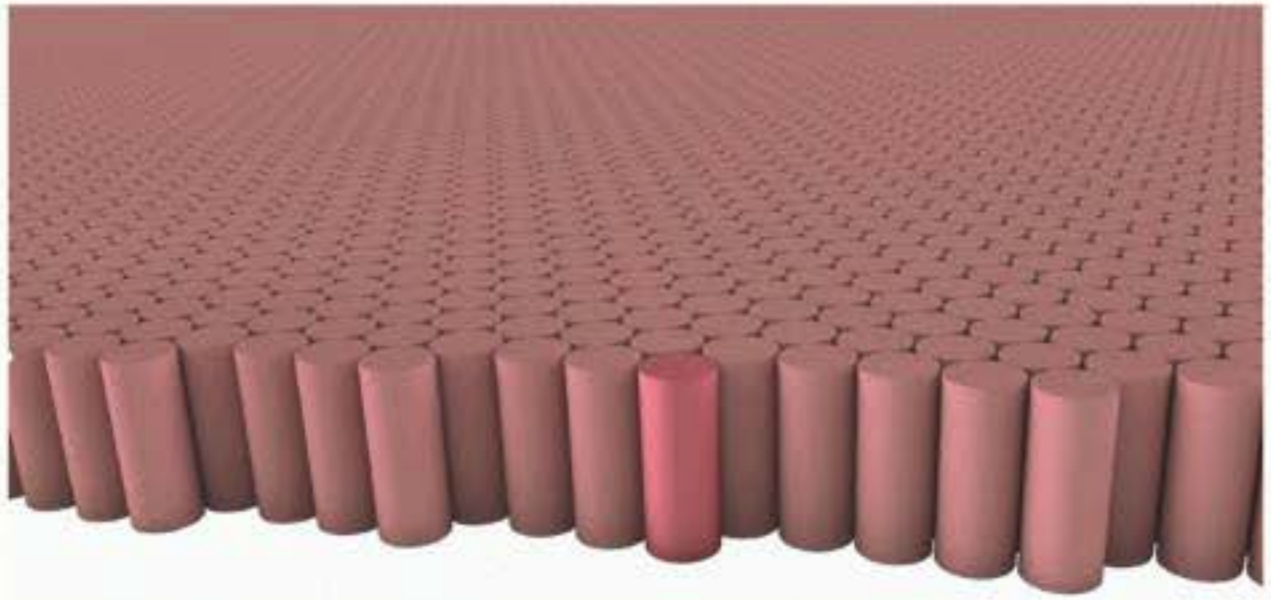
Applies to:
- **Physical objects**
- **Abstract objects**

Vernon Mountcastle's Big Idea

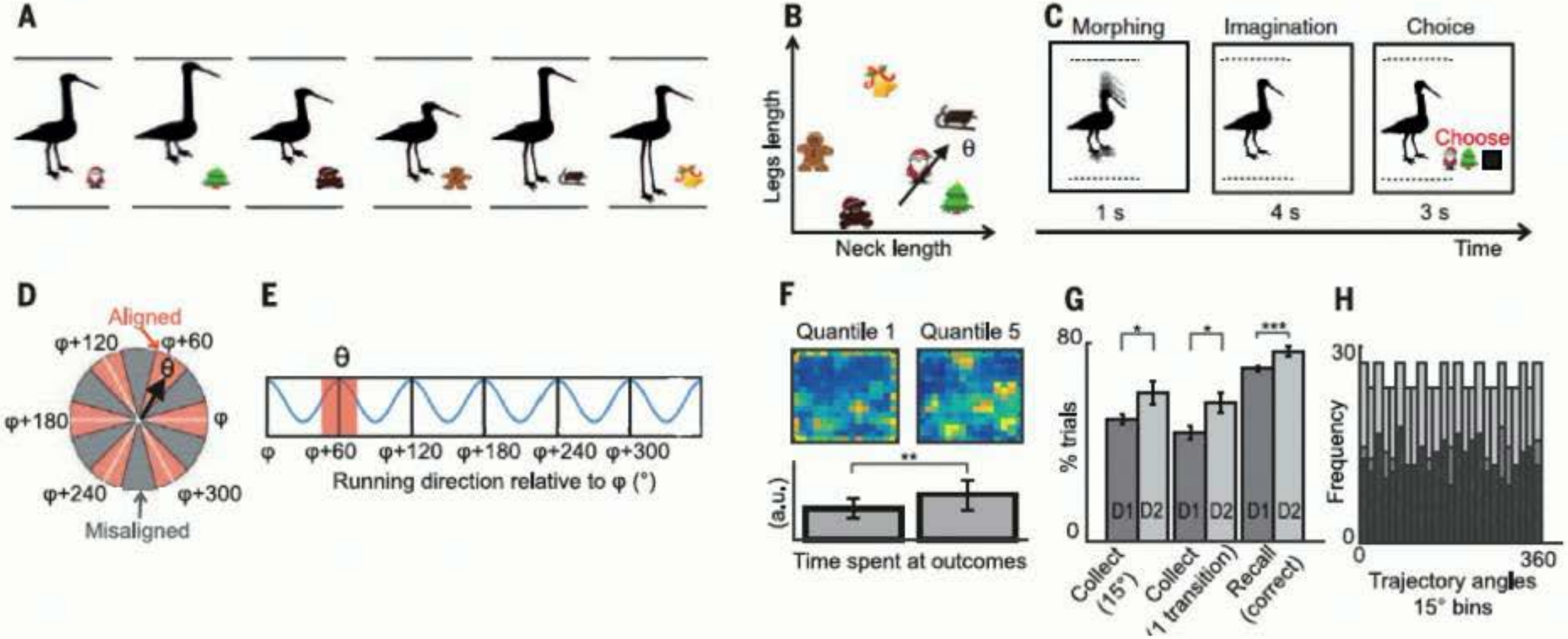
1) All areas of the neocortex look the same because they perform the same intrinsic function.

2) What makes one region visual and another auditory is what it is connected to.

3) A "cortical column" ($1mm^2$) is the unit of replication.



Mountcastle, 1978



Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. Nature

Numenta

Mission

1) Reverse engineer the neocortex
   - biologically accurate theories
   - test via empirical data and simulation
   - all our research is published and open

2) Apply neocortical theory to AI
   - improve current techniques
   - move toward truly intelligent systems

Outline

1) Robustness

Sparse representations in the brain

Incorporating sparsity into deep learning networks

2) Continuous learning

3) Unsupervised learning

Outline

1) Robustness

Sparse representations in the brain

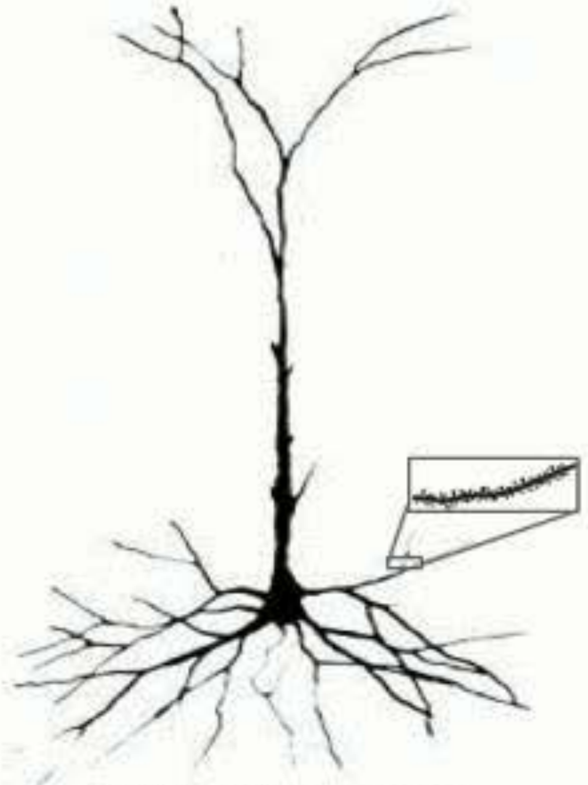
Incorporating sparsity into deep learning networks

2) Continuous learning / unsupervised learning

Biological neurons

Neurons continuously make predictions and learn from errors

Neurons Operate On Highly Sparse Representations



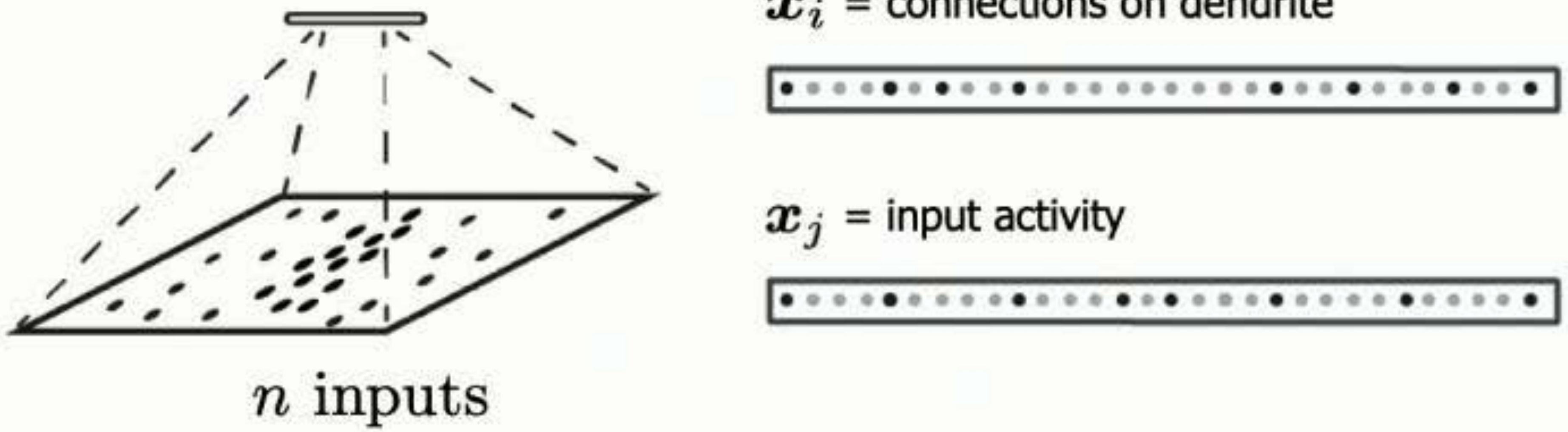
Pyramidal neuron
3K to 10K synapses

On a single neuron, 8-20 synapses on tiny segments of dendrites can recognize patterns.

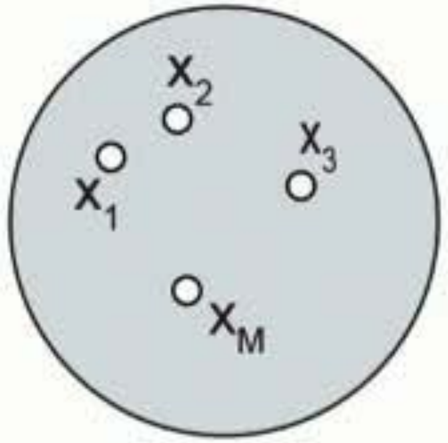
Thousands of other neurons send input to it.

How can neurons recognize patterns robustly using a tiny fraction of available connections?

Binary Sparse Vector Matching

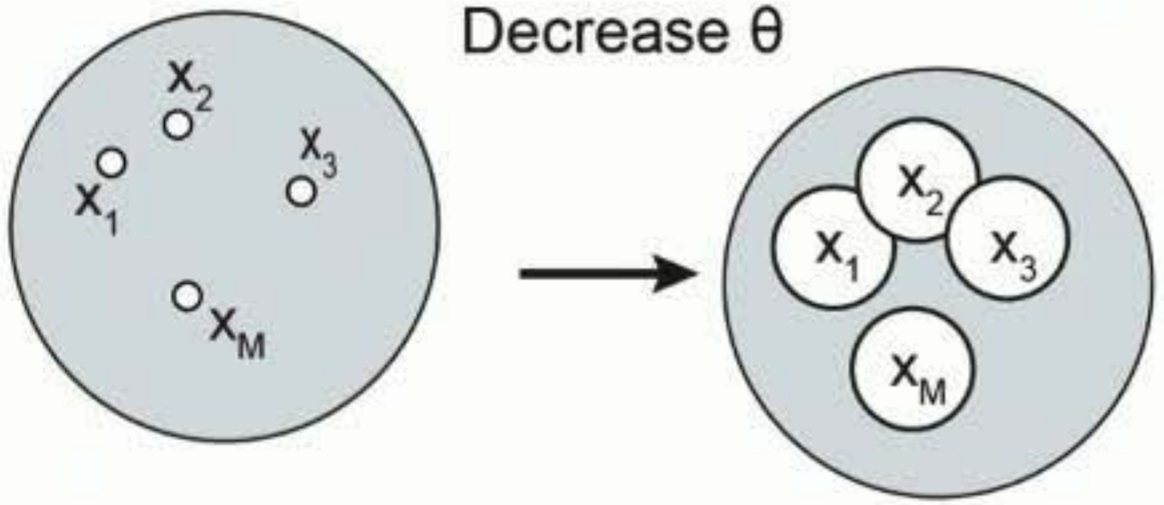


Combinatorics of Sparse Vector Matching



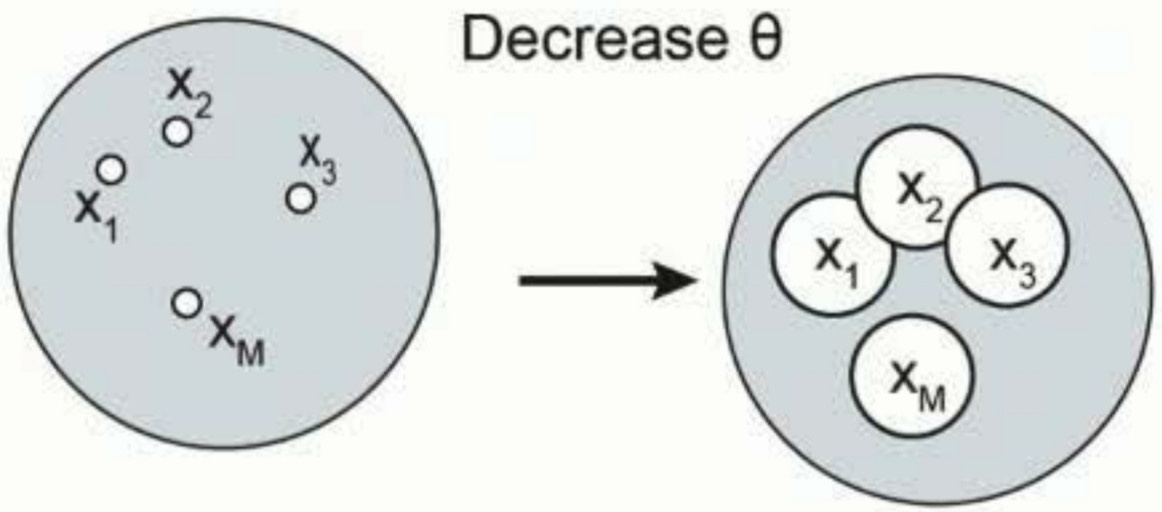
Combinatorics of Sparse Vector Matching

Decrease θ



We can get excellent noise robustness by reducing θ. What we care about are the false positives.

Combinatorics of Sparse Vector Matching



We can get excellent noise robustness by reducing θ. What we care about are the false positives.

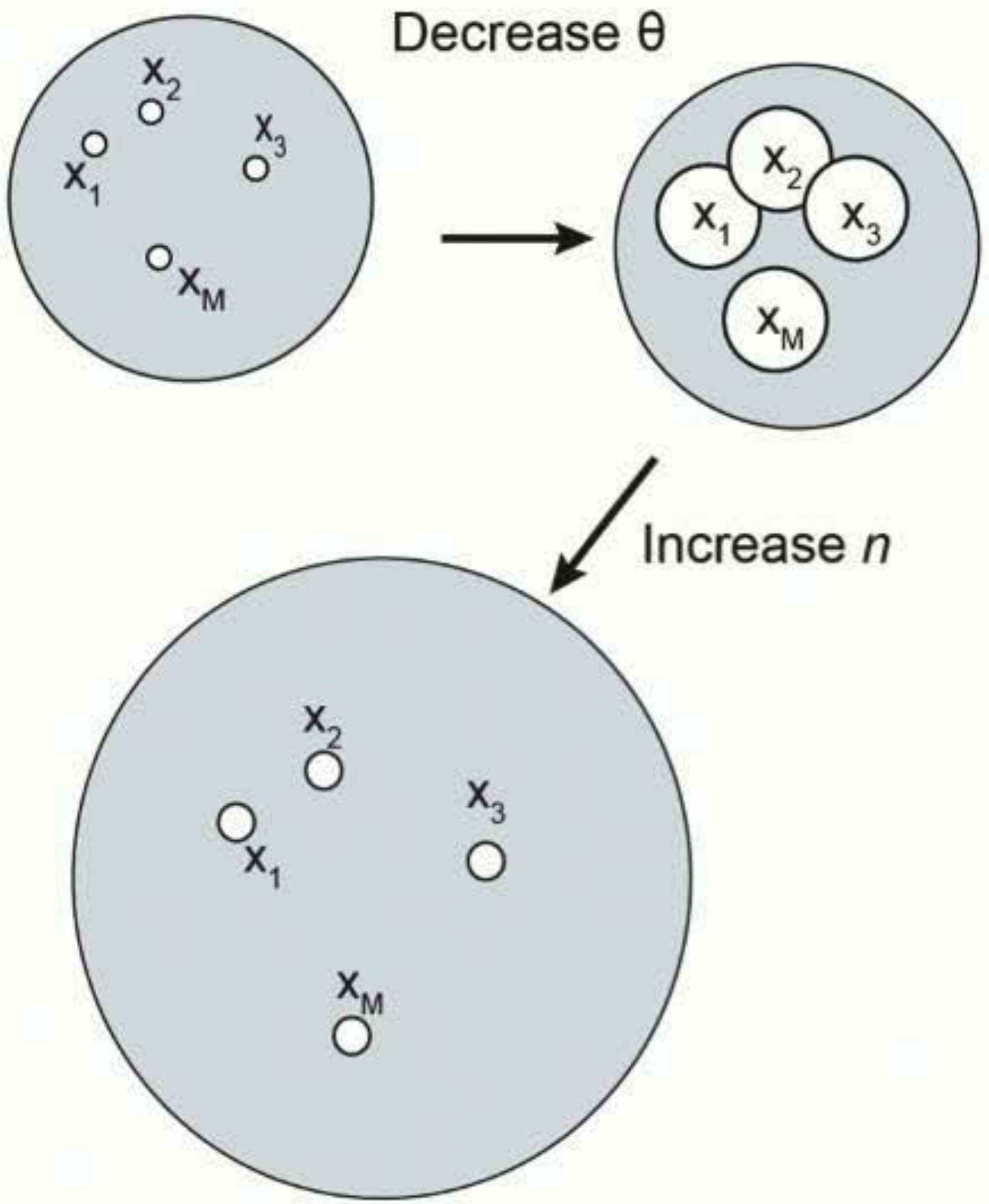
Can compute the probability of a random vector $\boldsymbol{x}_j$ matching a given $\boldsymbol{x}_i$:

$$P(\boldsymbol{x}_i \cdot \boldsymbol{x}_j \geq \theta) = \frac{\sum_{b=\theta}^{|\boldsymbol{x}_i|} \mid \Omega^n(\boldsymbol{x}_i, b, |\boldsymbol{x}_j|) \mid}{\binom{n}{|\boldsymbol{x}_j|}}$$

Numerator: volume around point (white)
Denominator: full volume of space (grey)

$$|\Omega^n(\boldsymbol{x}_i, b, k)| = \binom{|\boldsymbol{x}_i|}{b}\binom{n - |\boldsymbol{x}_i|}{k - b}$$

Combinatorics of Sparse Vector Matching



We can get excellent noise robustness by reducing θ. What we care about are the false positives.

Can compute the probability of a random vector $\boldsymbol{x}_j$ matching a given $\boldsymbol{x}_i$:

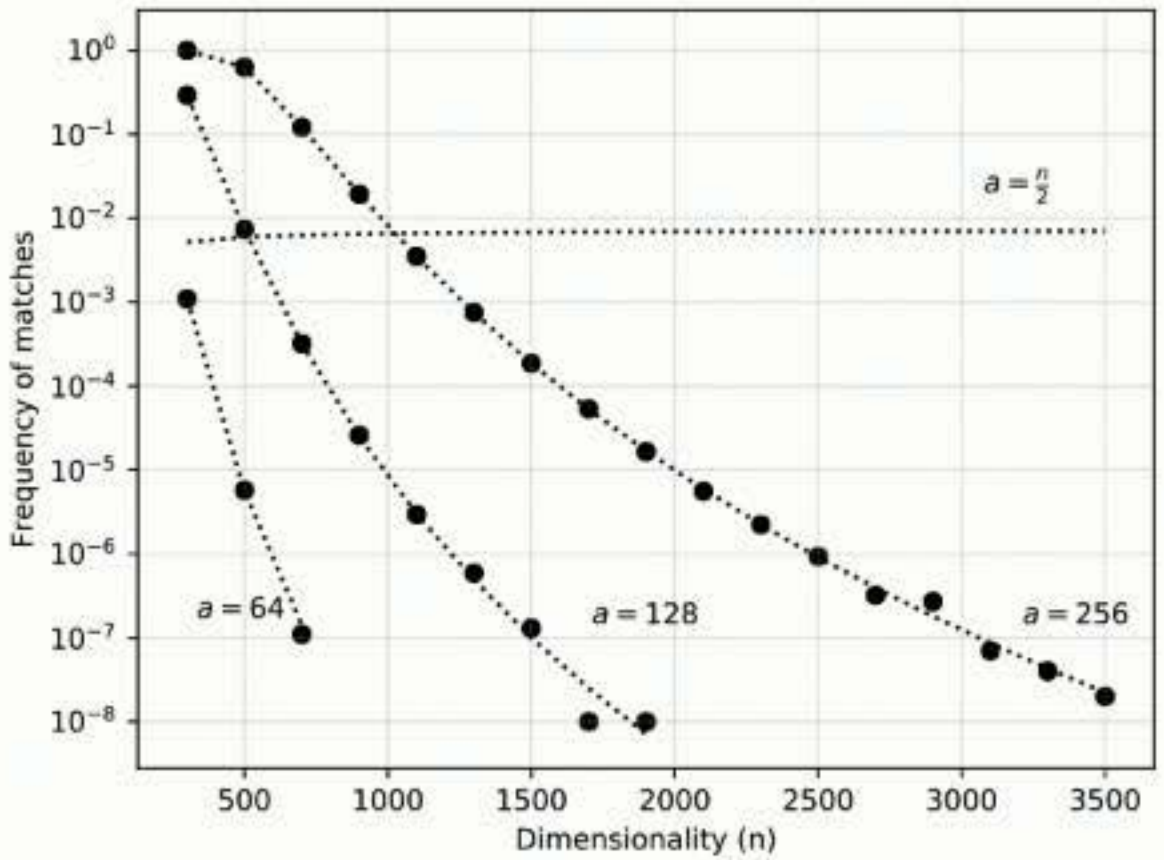
$$P(\boldsymbol{x}_i \cdot \boldsymbol{x}_j \geq \theta) = \frac{\sum_{b=\theta}^{|\boldsymbol{x}_i|} |\ \Omega^n(\boldsymbol{x}_i, b, |\boldsymbol{x}_j|)\ |}{\binom{n}{|\boldsymbol{x}_j|}}$$

Numerator: volume around point (white)
Denominator: full volume of space (grey)

$$|\Omega^n(\boldsymbol{x}_i, b, k)| = \binom{|\boldsymbol{x}_i|}{b}\binom{n - |\boldsymbol{x}_i|}{k - b}$$

Sparse High Dimensional Representations Are Highly Robust

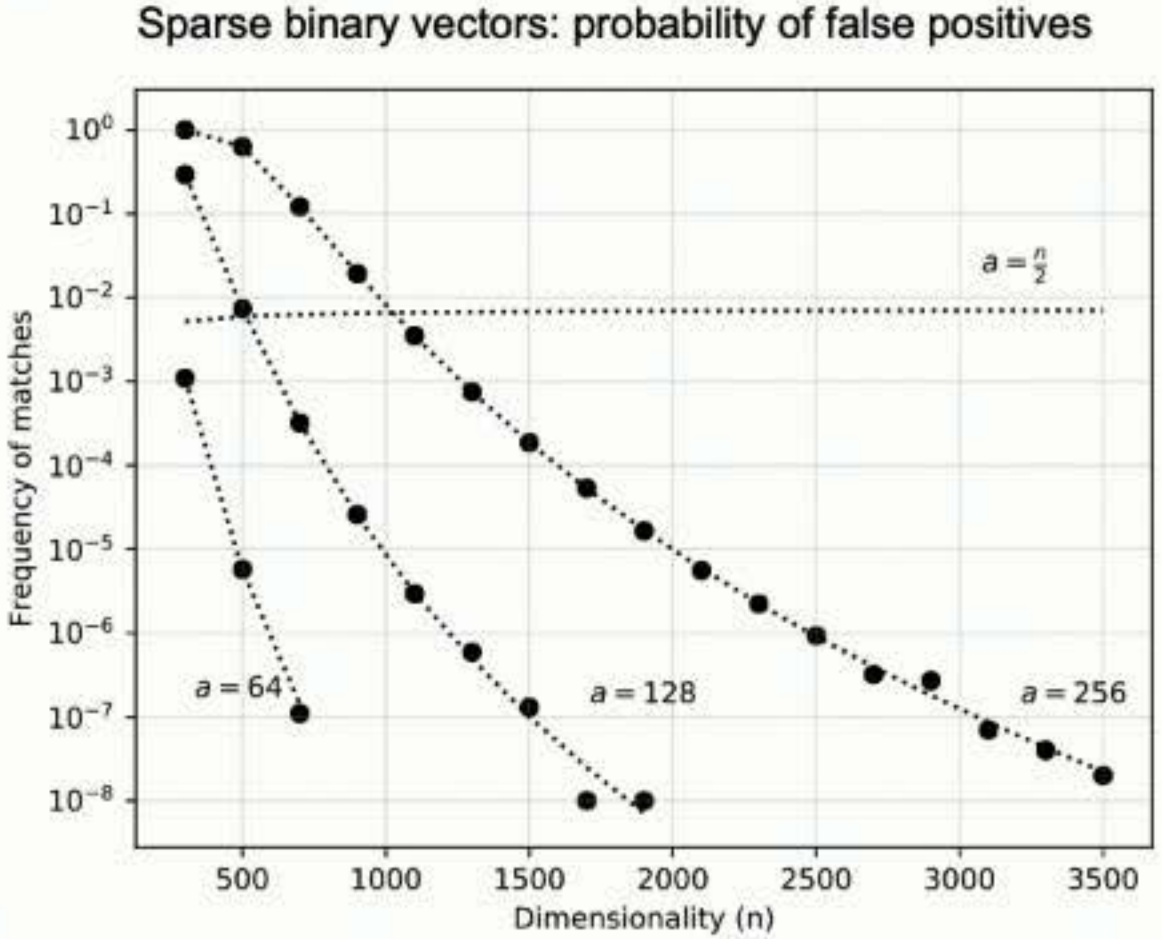
Sparse binary vectors: probability of false positives



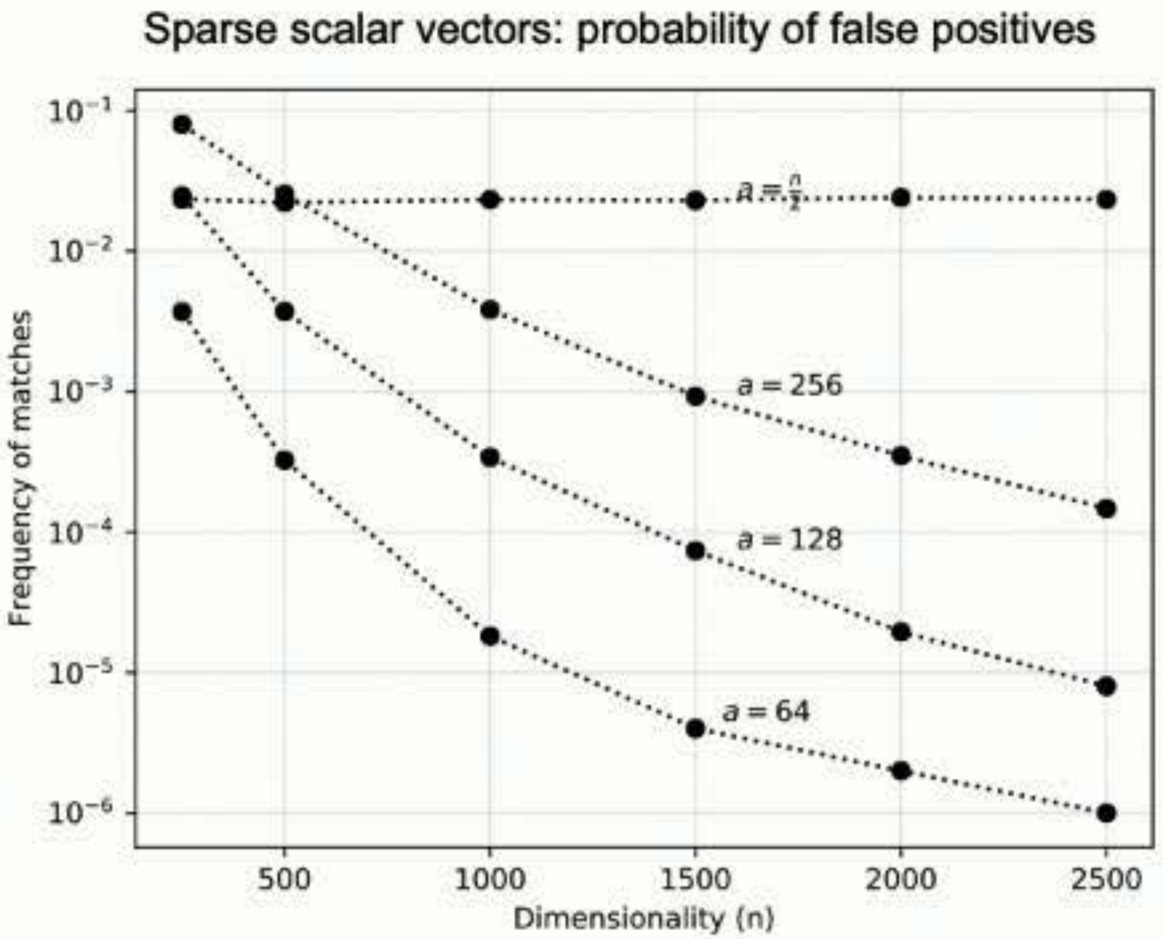
$$|\boldsymbol{x}_i| = 24, \theta = 12, a = |\boldsymbol{x}_j|$$

1) False positive error decreases exponentially with dimensionality with sparsity.
2) The number of connections can be quite small, even with threshold at 50%.
3) Error rates do not decrease when activity is dense (a=n/2).
4) Assume uniform random distribution of vectors.

Sparse High Dimensional Representations Are Highly Robust

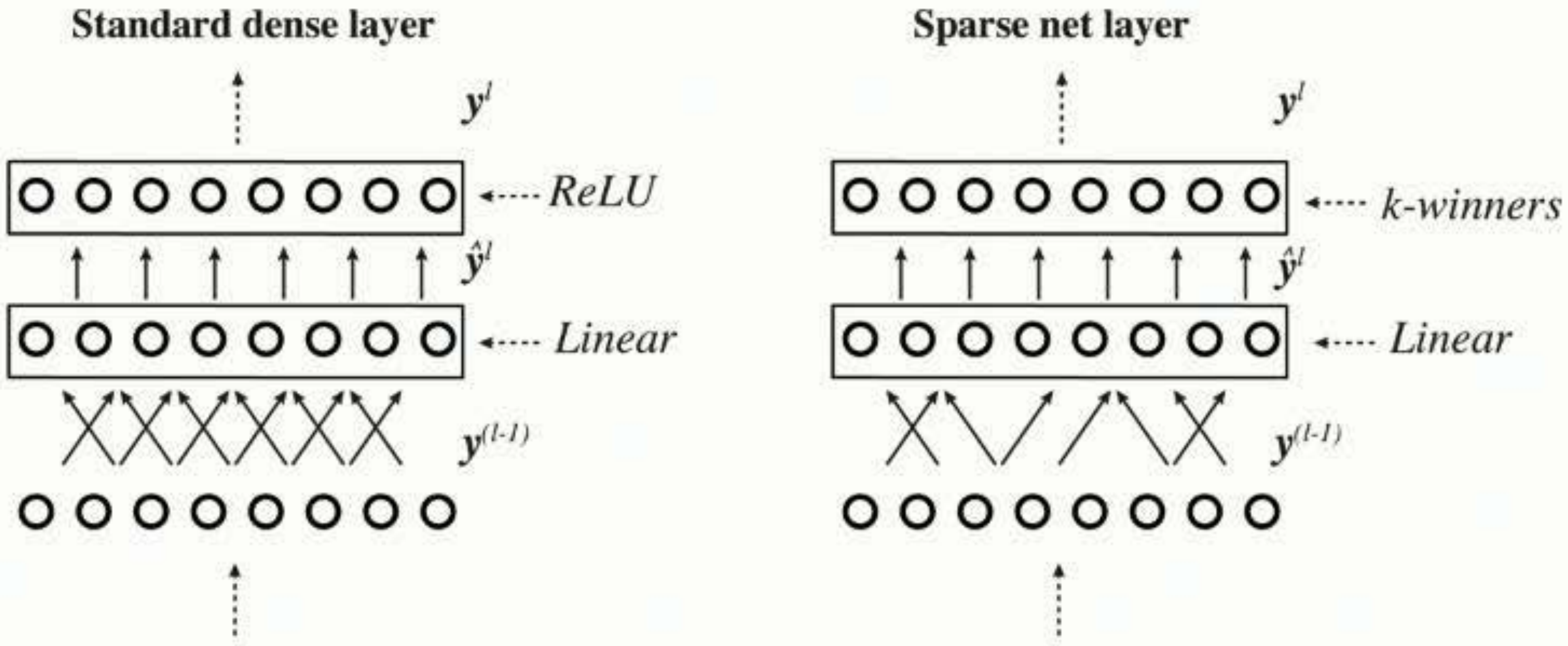


$$|\boldsymbol{x}_i| = 24, \theta = 12, a = |\boldsymbol{x}_j|$$



1) False positive error decreases exponentially with dimensionality with sparsity.
2) The number of connections can be quite small, even with threshold at 50%.
3) Error rates do not decrease when activity is dense (a=n/2).
4) Assume uniform random distribution of vectors.

Differentiable Sparse Layer



1) Weight matrix is sparse
   Most of the weights are zero, and maintained as zero throughout

2) Outputs of top-k units are maintained, the rest are set to 0
   (analogous to ReLU: gradient is 1 for top k units, 0 elsewhere)

3) An exponential boosting term favors units with low activation frequency
   This helps maximize the overall entropy of the layer.

4) Convolutional layer is identical except we didn't use sparse weights

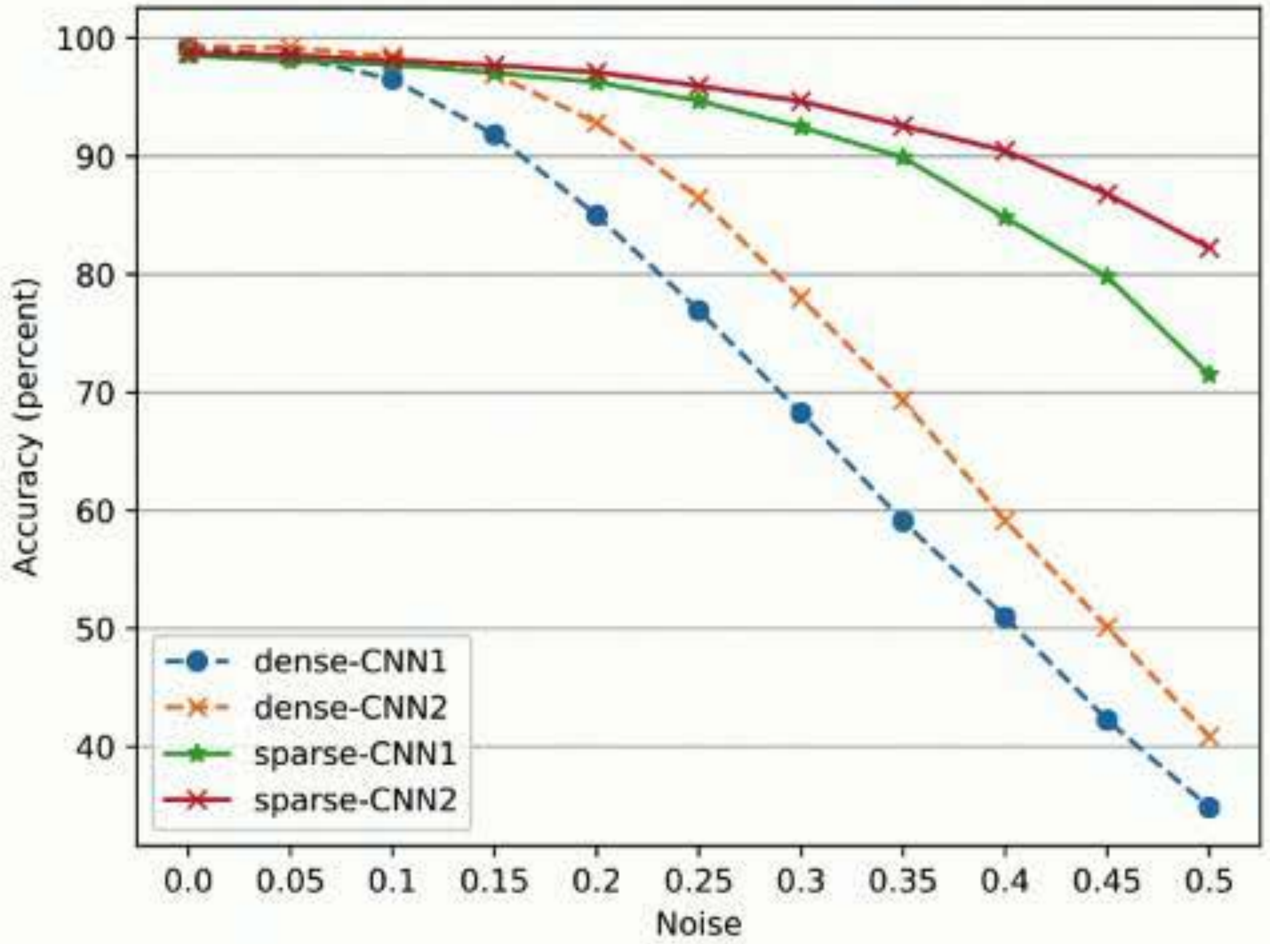
(Hawkins et al., 2011)
(Makhzani & Frey, 2015)
(Ahmad & Scheinkman, 2019)

MNIST With Sparse Networks

MNIST

Network	Test Score
Dense CNN-1	99.23± 0.04
Dense CNN-2	99.38± 0.10
Sparse CNN-1	98.85± 0.09
Sparse CNN-2	98.89± 0.12

MNIST: Accuracy vs noise



1) Networks used CNN layers + one (sparse) linear layer + one softmax output layer.
2) State of the art test set accuracy is between 98.3% and 99.4% (without data augmentation)

Sparse Networks Are Significantly Better On Noisy Data

Dense CNN

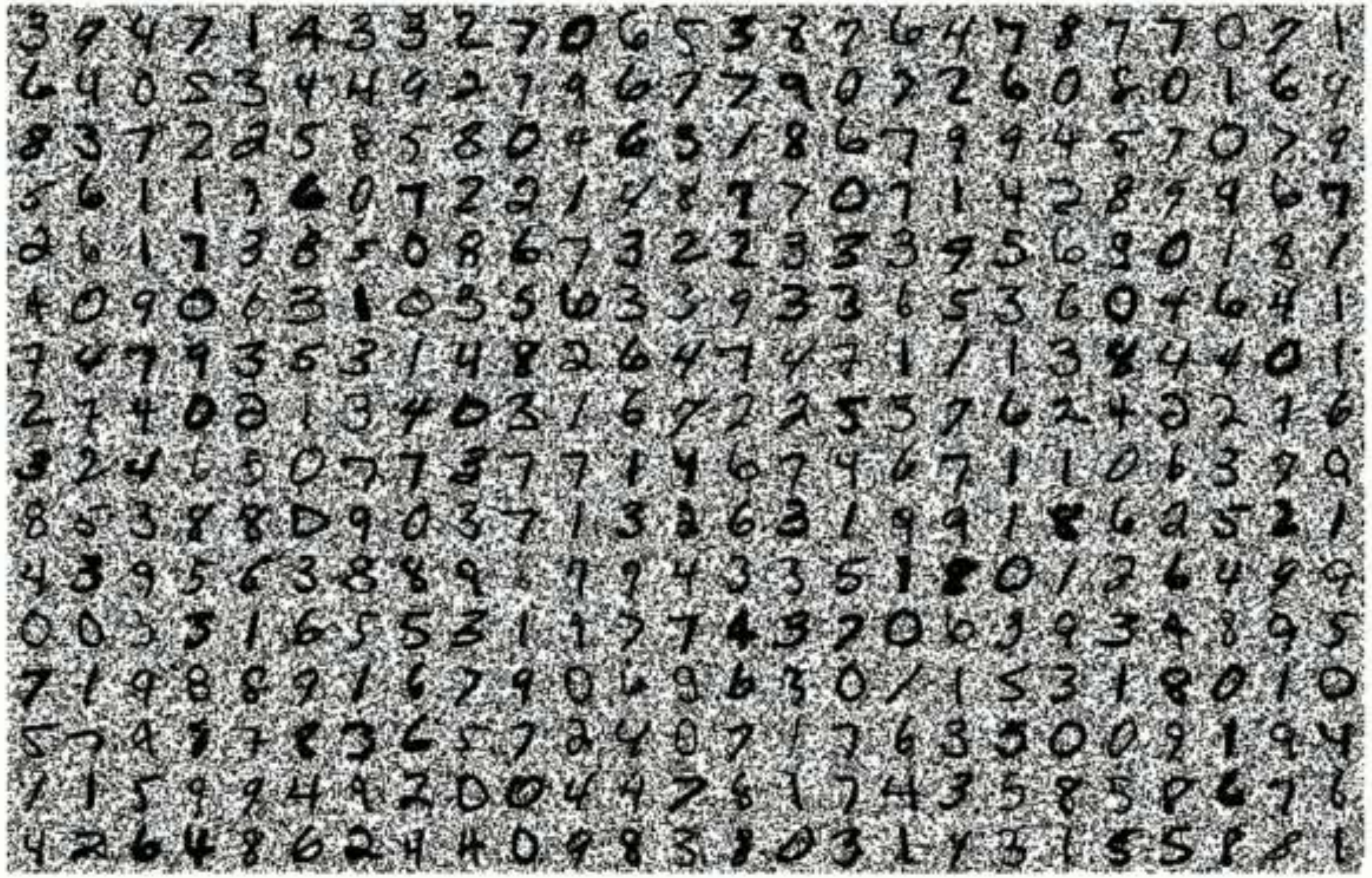
97 %

SparseNet

98 %

10% Noise

Sparse Networks Are Significantly Better On Noisy Data



30% Noise

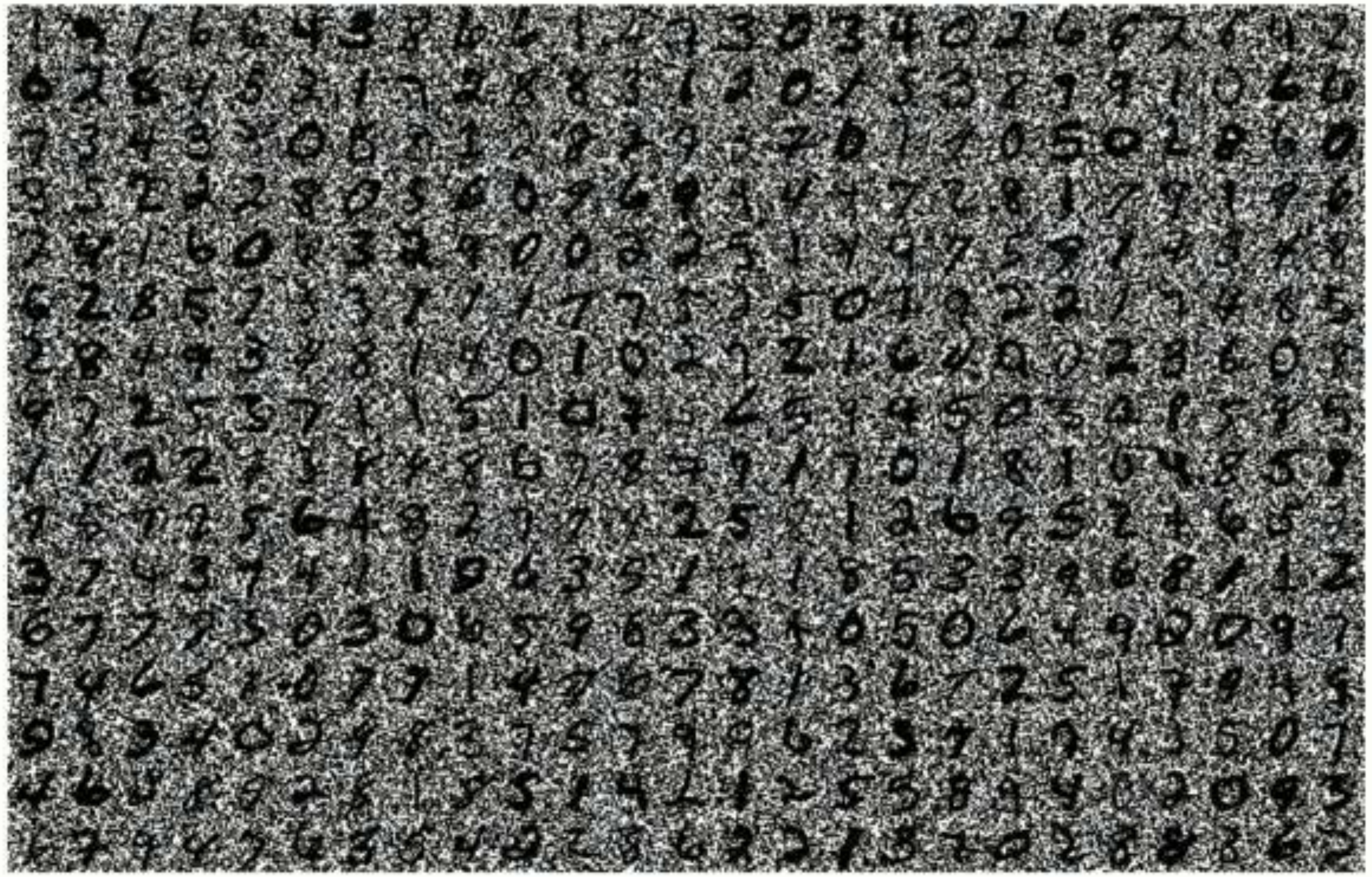
Dense CNN

64 %

SparseNet

92 %

Sparse Networks Are Significantly Better On Noisy Data



Dense CNN

34 %

SparseNet

72 %

50% Noise

Google Speech Commands Dataset

<u>**Dataset of spoken one word commands**</u>

- Released by Google in 2017
- 65,000 utterances, thousands of individuals
- Harder than MNIST
- State of the art is around 95 - 97.5% for 10 categories
- Tested accuracy with noisy sounds

Network	Test Score	Noise Score
Dense CNN-2 (Dr=0.0)	96.37 ± 0.37	$8{,}730 \pm 471$
Dense CNN-2 (Dr=0.5)	95.69 ± 0.48	$7{,}681 \pm 368$
Sparse CNN-2	96.65 ± 0.21	$11{,}233 \pm 1013$
Super-sparse CNN-2	96.57 ± 0.16	$10{,}752 \pm 942$

1) Networks used two CNN layers + one sparse linear layer + one softmax output layer.
2) Batchnorm used for all hidden layers
3) Audio files were converted to 32-MFCC coefficients, with data augmentation during training.
4) Super-sparse net had a very sparse linear layer: 6.7% sparsity and 10% of weights as non-zero

Outline

1) Robustness

Sparse representations in the brain

Incorporating sparsity into deep learning networks

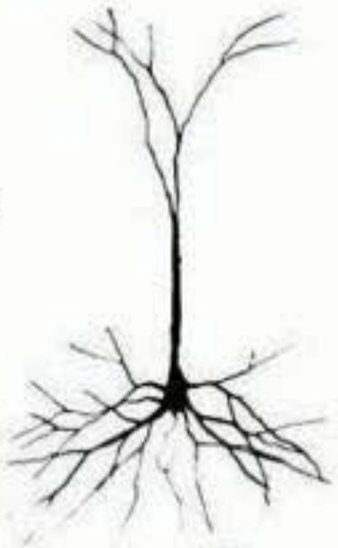
2) Continuous learning / unsupervised learning

Biological neurons

Neurons continuously make predictions and learn from errors

Biological Neurons Are Complex

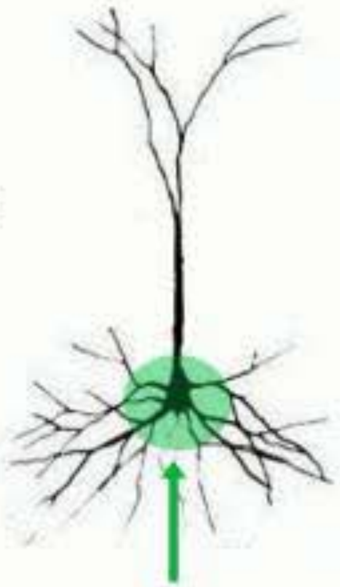
Pyramidal neuron
3K to 10K synapses



(Poirazi et al., 2003)
(Hawkins & Ahmad, 2016)

Biological Neurons Are Complex

Pyramidal neuron
3K to 10K synapses

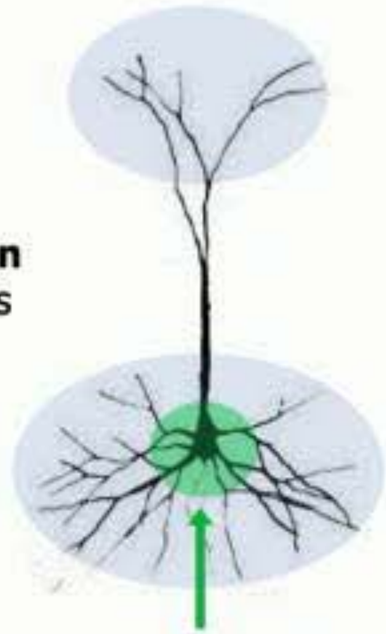


Feedforward
Weighted sum + non-linearity
Drive the cell, classic point neuron
10% of synapses

(Poirazi et al., 2003)
(Hawkins & Ahmad, 2016)

Biological Neurons Are Complex

Pyramidal neuron
3K to 10K synapses



Distal dendrites
- 8-20 clustered synapses generate dendritic spikes
- Does not cause cell to fire
- Primes the cell to fire strongly in the near future
- Can detect hundreds of independent sparse patterns

Feedforward
Weighted sum + non-linearity
Drive the cell, classic point neuron
10% of synapses

(Poirazi et al., 2003)
(Hawkins & Ahmad, 2016)

Predictions And Continuous Learning In Neurons

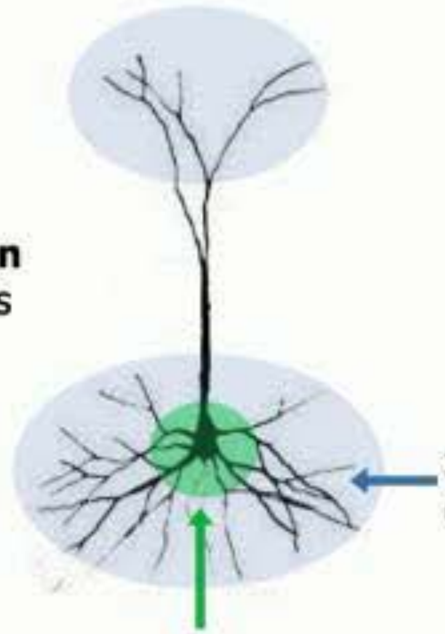
Pyramidal neuron
3K to 10K synapses

Feedforward patterns

(Poirazi et al., 2003)
(Hawkins & Ahmad, 2016)

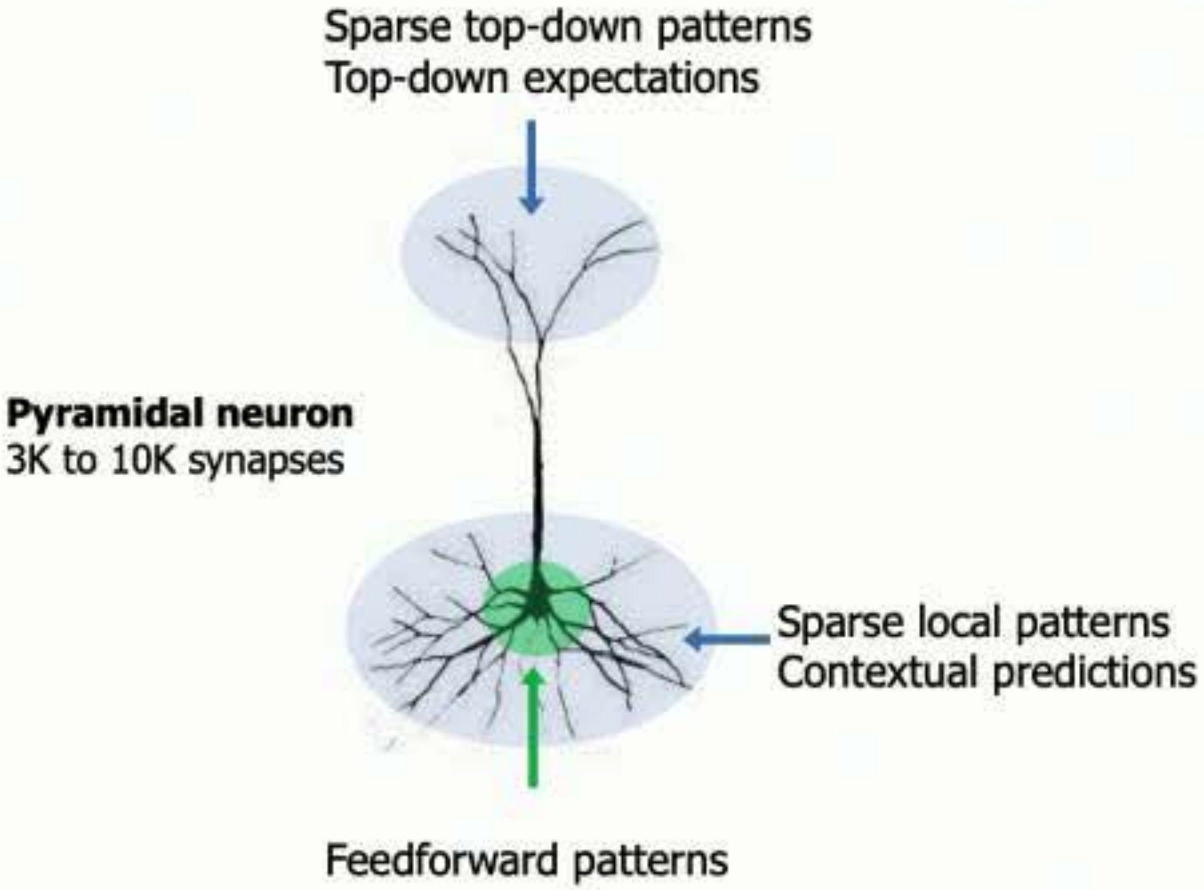
Predictions And Continuous Learning In Neurons

Pyramidal neuron
3K to 10K synapses



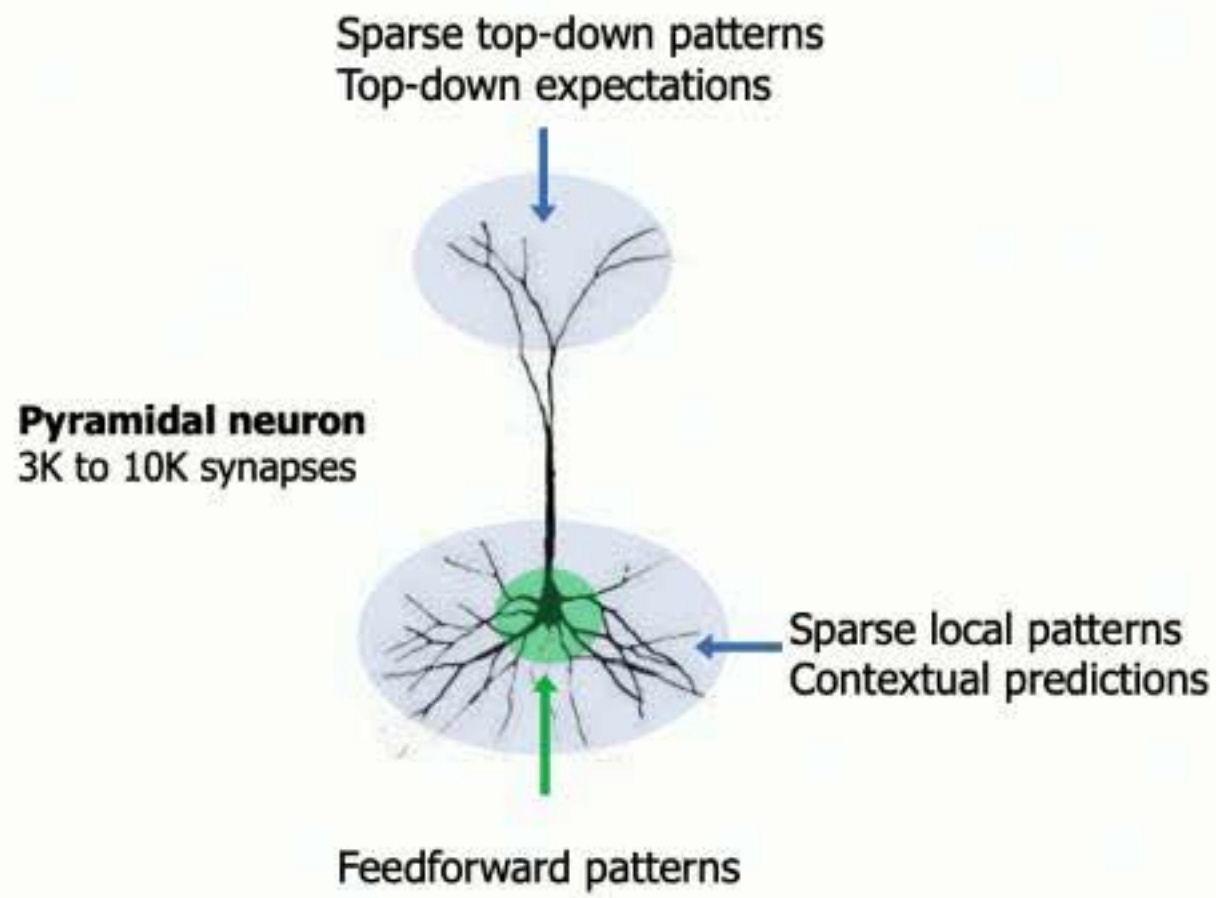
(Poirazi et al., 2003)
(Hawkins & Ahmad, 2016)

Predictions And Continuous Learning In Neurons



(Poirazi et al., 2003)
(Hawkins & Ahmad, 2016)

Predictions And Continuous Learning In Neurons



Simple learning rules

If cell becomes active:

1) If there was a prediction, reinforce that segment
2) If there was no prediction, grow connections by subsampling cells active in the past

If cell is not active:

1) If there was a prediction, weaken than segments

- Learning consists of growing new connections, i.e., highly sparse vectors.

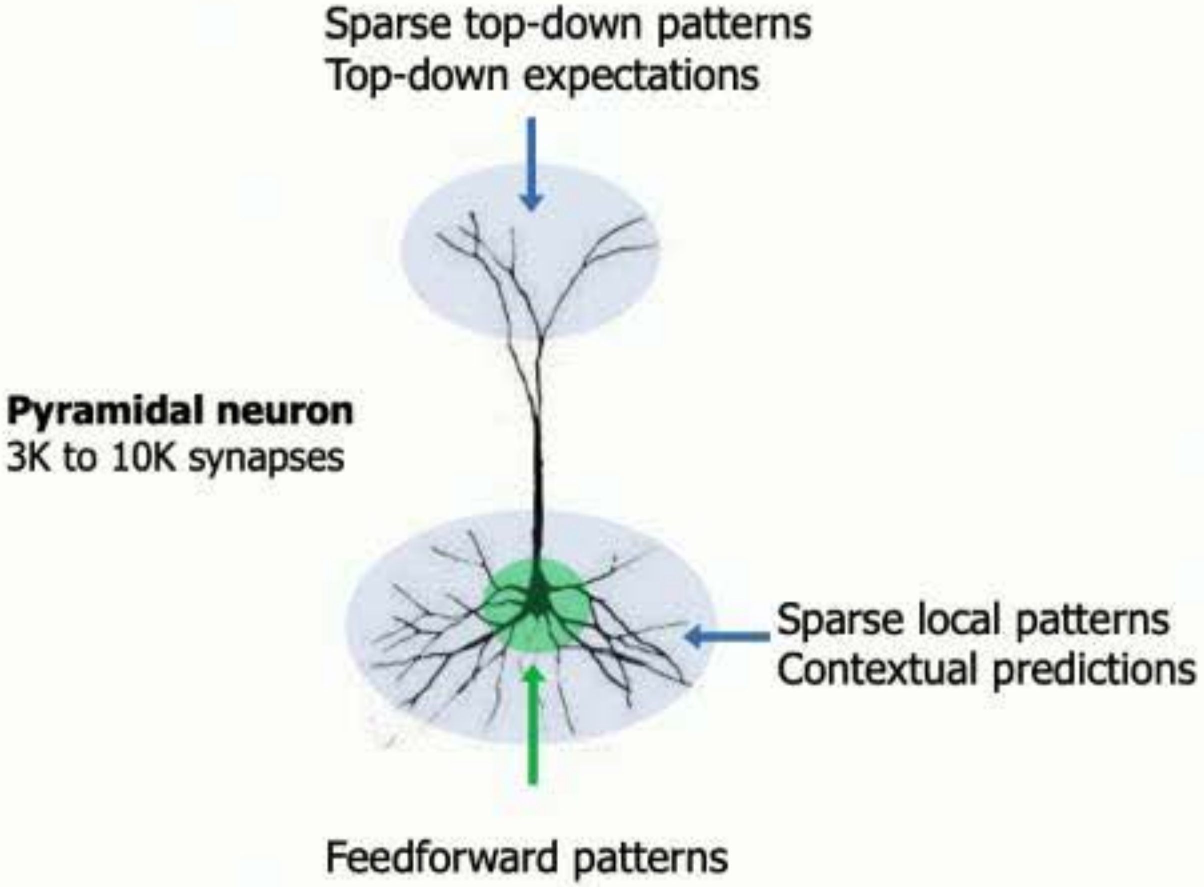
- Each neuron can be associated with hundreds of such sparse contextual patterns spread throughout dendrites.

- Each neuron is constantly trying to make predictions and learn from its mistakes.

- Everything is continuously learning but because vectors are sparse, patterns don't interfere with each other.

(Poirazi et al., 2003)
(Hawkins & Ahmad, 2016)

Predictions And Continuous Learning In Neurons



Simple learning rules

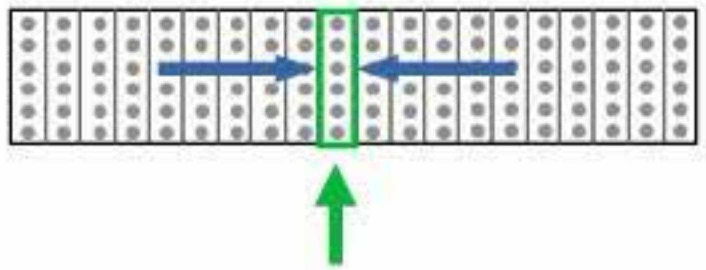
<u>If cell becomes active:</u>

1) If there was a prediction, reinforce that segment
2) If there was no prediction, grow connections by subsampling cells active in the past

<u>If cell is not active:</u>

1) If there was a prediction, weaken than segments

Network of pyramidal neurons can form a powerful predictive learning algorithm

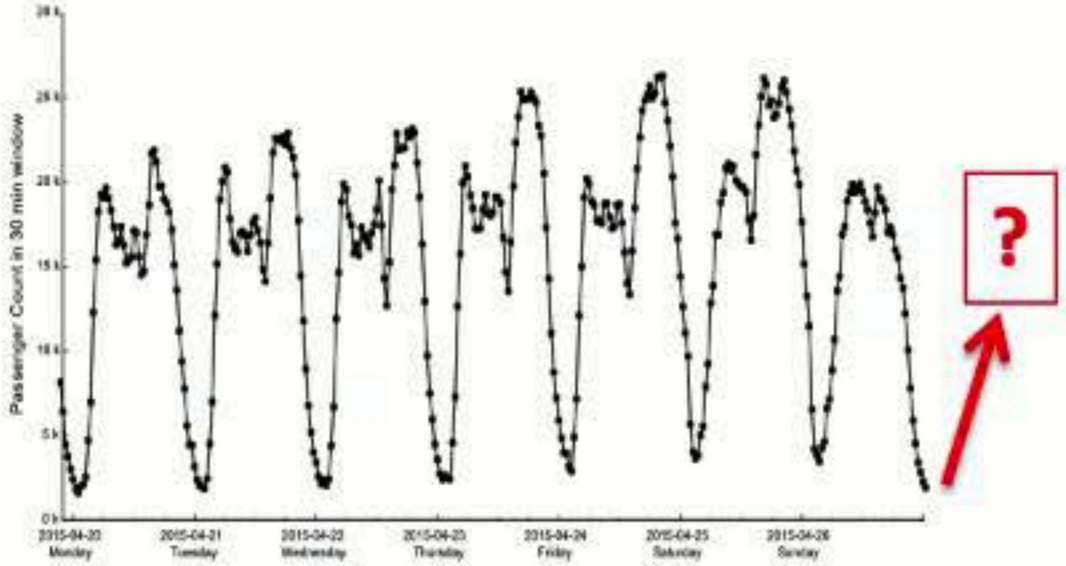


1) Associates past activity as context for current activity
2) Learns continuously without forgetting past patterns
3) Can learn complex high-Markov order sequences
4) Sparse representations lead to fault tolerance

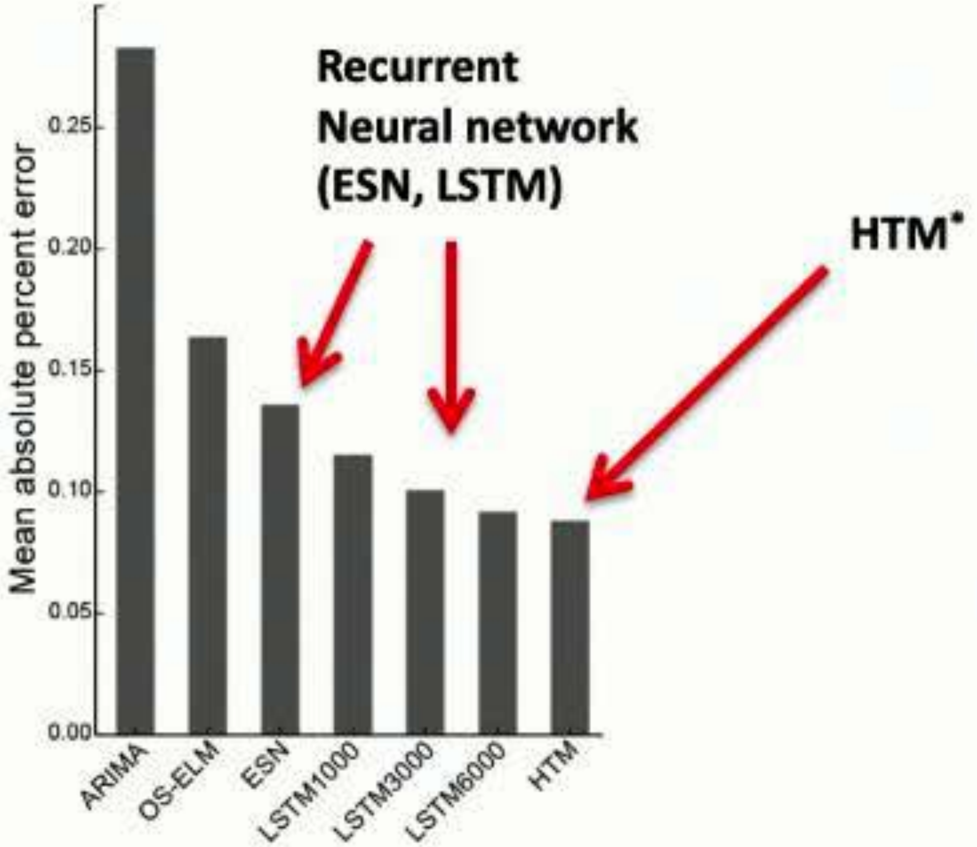
(Poirazi et al., 2003)
(Hawkins & Ahmad, 2016)

Continuous Learning With Streaming Data Sources

NYC Taxi demand datastream



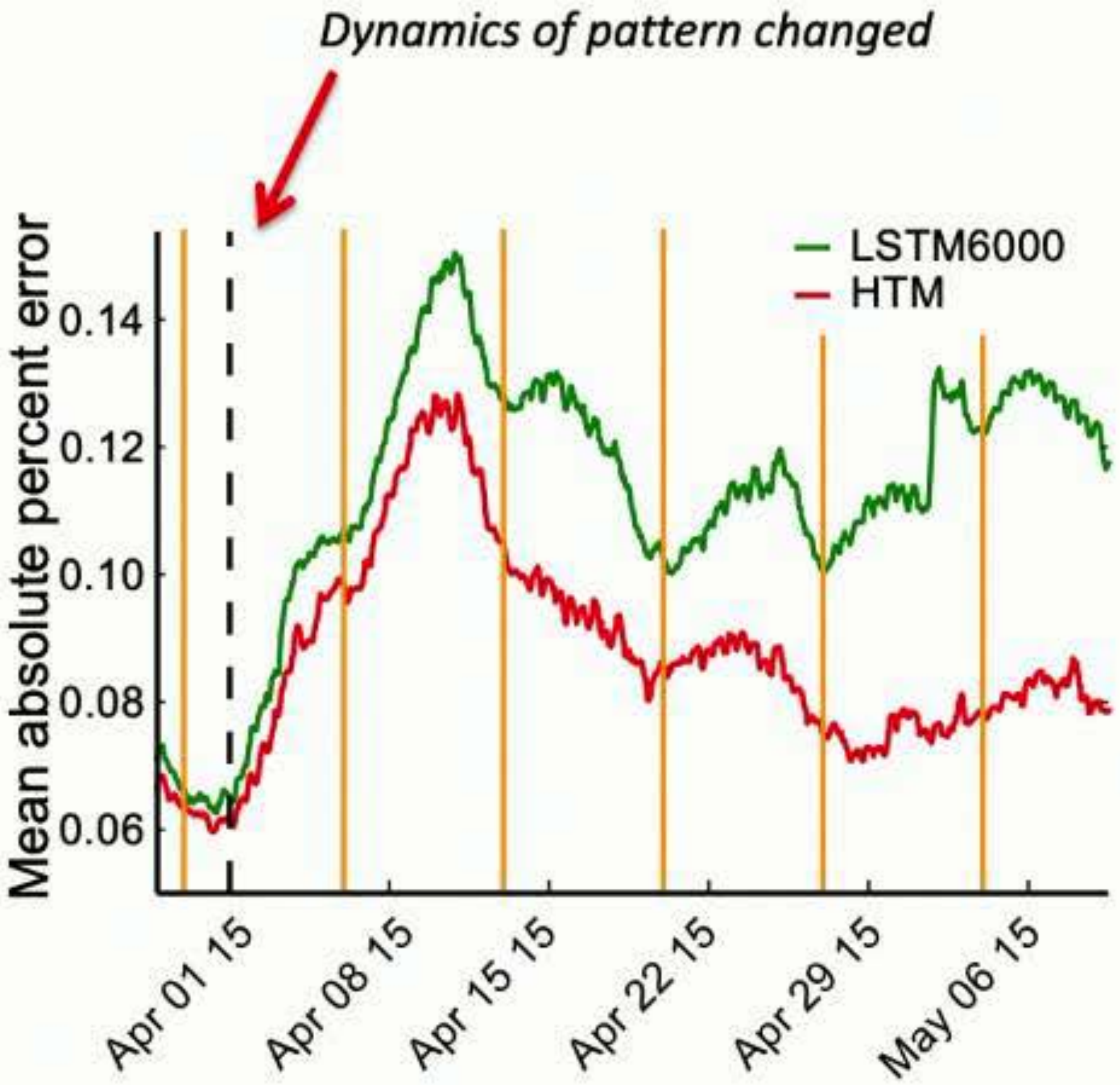
Source: http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml



HTM = Hierarchical Temporal Memory

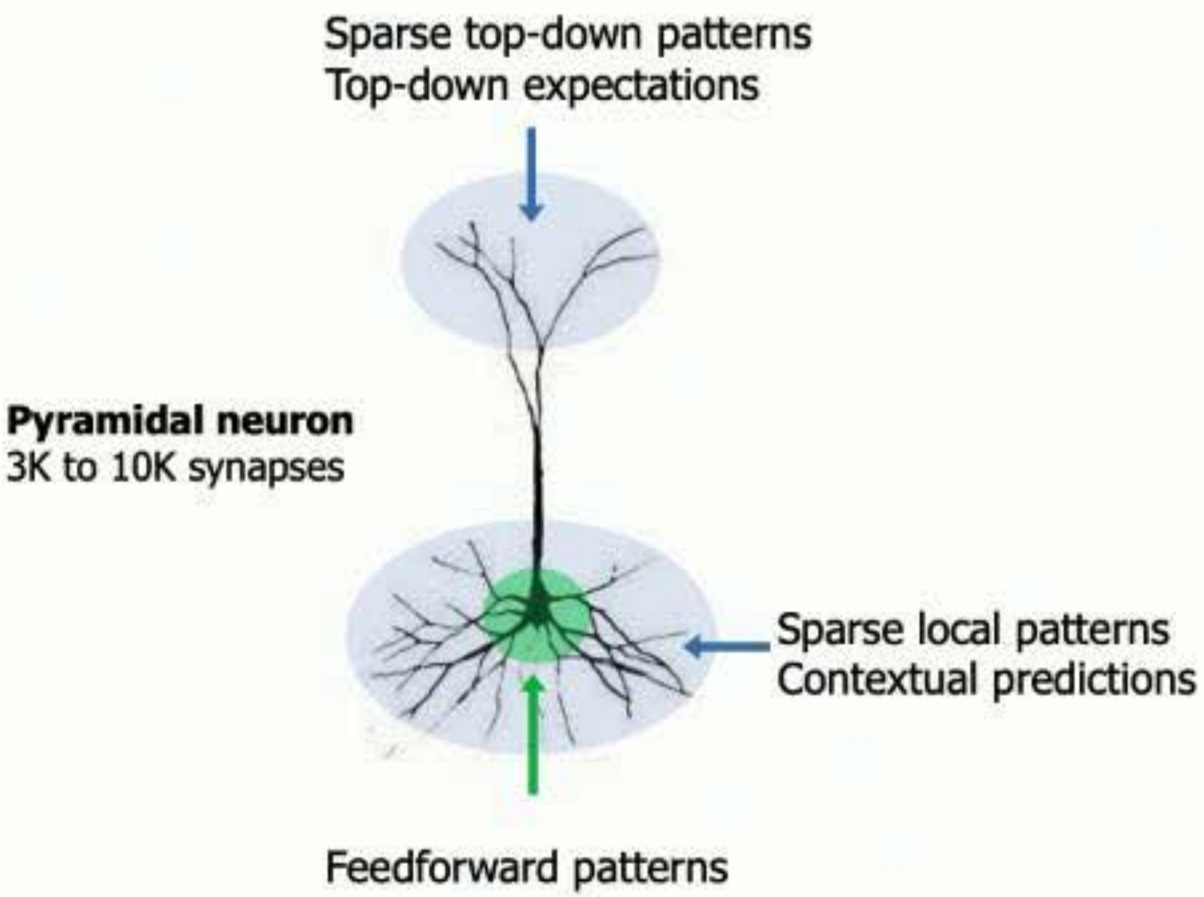
(Cui et al, Neural Computation, Nov 2016)

Adapts Quickly To Changing Statistics



(Cui et al, Neural Computation, 2016)

Predictions And Continuous Learning In Neurons



Simple learning rules

<u>If cell becomes active:</u>

1) If there was a prediction, reinforce that segment
2) If there was no prediction, grow connections by subsampling cells active in the past

<u>If cell is not active:</u>

1) If there was a prediction, weaken than segments

- Learning consists of growing new connections, i.e., highly sparse vectors.

- Each neuron can be associated with hundreds of such sparse contextual patterns spread throughout dendrites.

- Each neuron is constantly trying to make predictions and learn from its mistakes.

- Everything is continuously learning but because vectors are sparse, patterns don't interfere with each other.

(Poirazi et al., 2003)
(Hawkins & Ahmad, 2016)

Research Roadmap

1) Robustness

    Sparse representations in the brain
    Incorporate sparsity into deep learning networks

    Scale to larger problems
    Test with adversarial systems

2) Continuous learning / unsupervised learning

    Understand biological neurons
    Continuously make predictions and learn from errors

    Integrate neuron model into deep learning systems
    Implement predictive learning rules

3) "1000 Brains Theory"

    Distributed voting
    Many small models, across sensory modalities
    Object-centric reference frames

Opportunities For Collaboration

1) Applications
   Test robustness in adversarial and security scenarios.
   Test with different domains, such as robotics, NLP, and IoT
   Test with different DL architectures and paradigms, such as RNNs, and RL.

1) Scaling
   Attack much larger problems.
   Acceleration and power efficiency (e.g. FPGA implementations).

Numenta

Jeff Hawkins

Subutai Ahmad

Marcus Lewis

Mirko Klukas

Luiz Scheinkman

Contact: `sahmad@numenta.com` `jhawkins@numenta.com`
@SubutaiAhmad

Research Roadmap

1) Robustness

Sparse representations in the brain
Incorporate sparsity into deep learning networks

Scale to larger problems
Test with adversarial systems

2) Continuous learning / unsupervised learning

Understand biological neurons
Continuously make predictions and learn from errors

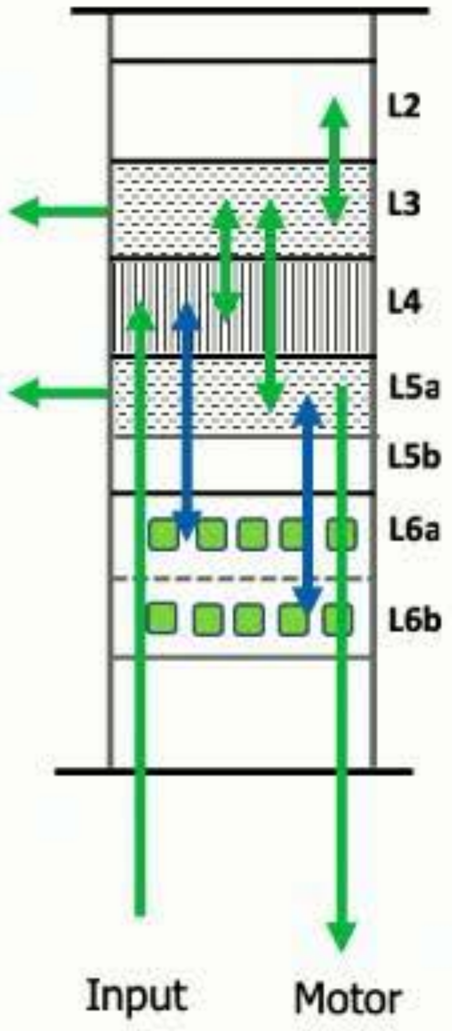
Integrate neuron model into deep learning systems
Implement predictive learning rules

3) "1000 Brains Theory"

Distributed voting
Many small models, across sensory modalities
Object-centric reference frames

Cortical Columns Are Complete Sensory-motor Modeling Systems

Hawkins et. al., 2018
Klukas et. al., 2019



Two reference frames

Learns:
- **Dimensionality of object**
- **Morphology**
- **Changes in morphology** (how objects behave)
- **Compositional and Recursive structure**

Generates motor behaviors

Applies to:
- **Physical objects**
- **Abstract objects**