Perception & Attention

Perception is effortless but its underlying mechanisms are incredibly sophisticated.

- Biology of the visual system
- Representations in primary visual cortex and Hebbian learning
- Object recognition
- Attention: Interactions between systems involved in object recognition and spatial processing

Perception & Attention

Perception & Attention

Some motivating questions:

1. Why does primary visual cortex encode oriented bars of light?

Perception & Attention

Some motivating questions:

1. Why does primary visual cortex encode oriented bars of light?

2. Why is visual system split into what/where pathways?

Perception & Attention

Some motivating questions:

1. Why does primary visual cortex encode oriented bars of light?

2. Why is visual system split into what/where pathways?

3. Why does parietal damage cause attention problems (neglect)?

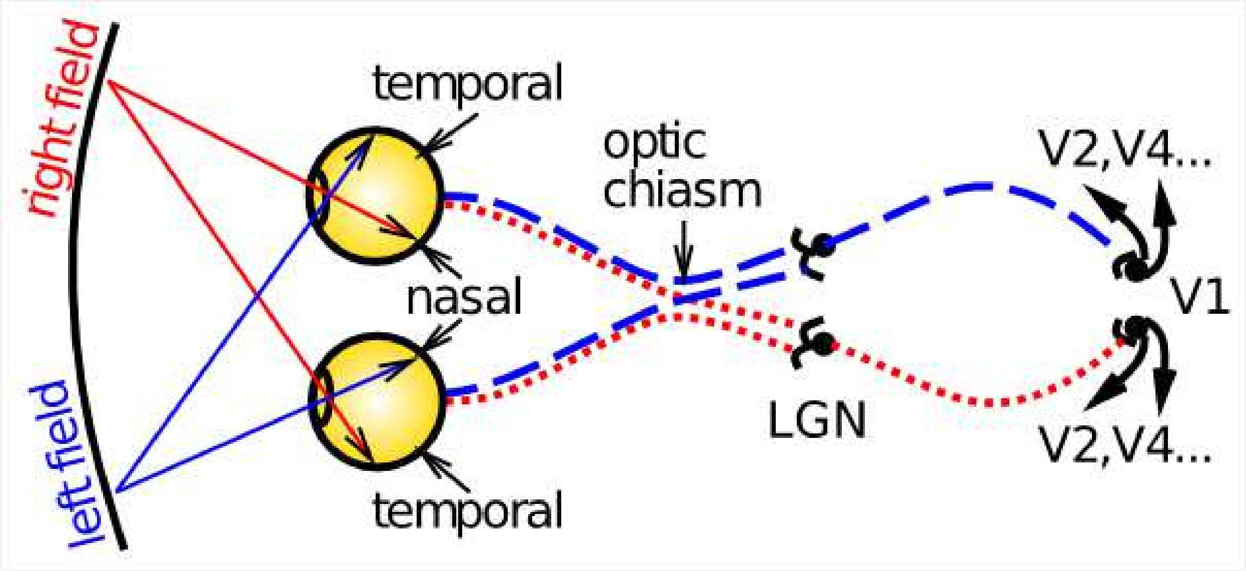
Perception & Attention

Some motivating questions:

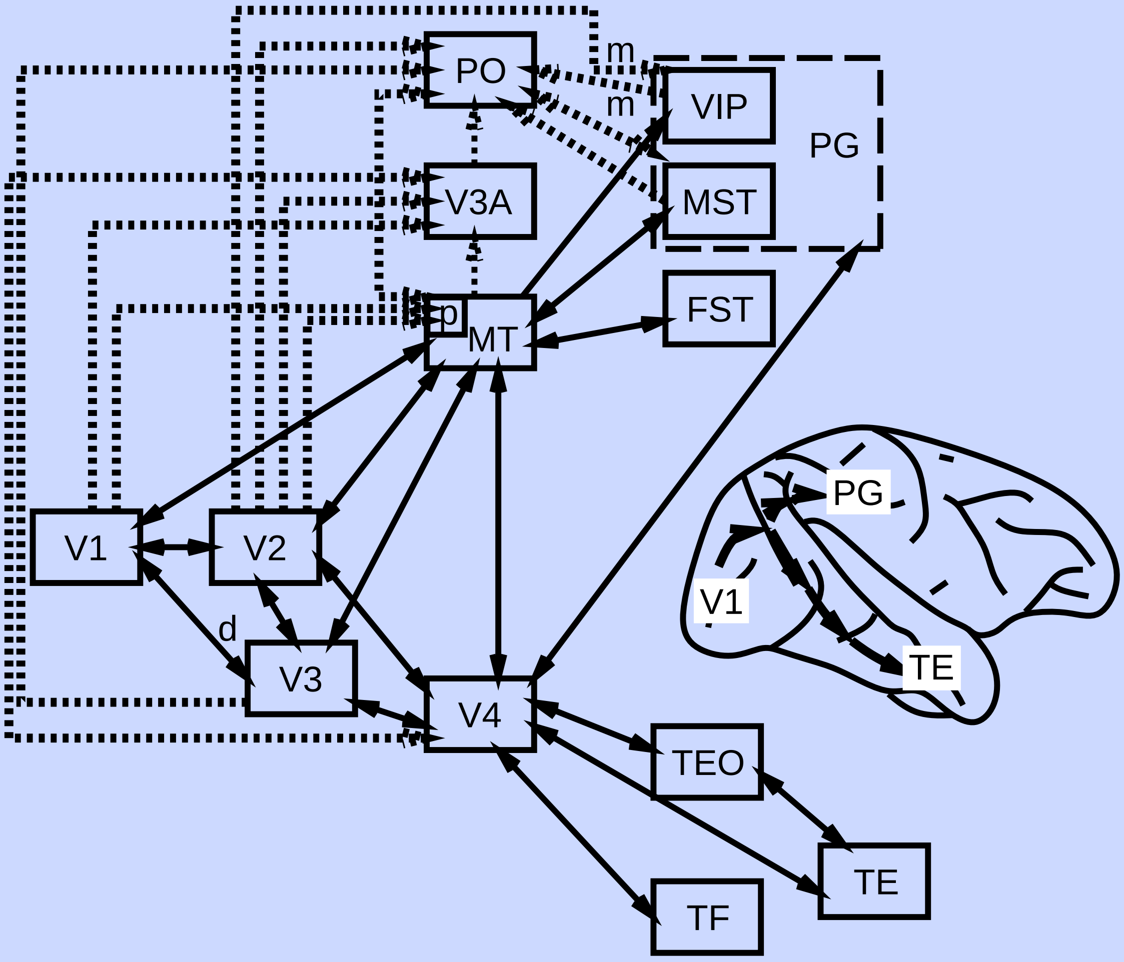
1. Why does primary visual cortex encode oriented bars of light?
2. Why is visual system split into what/where pathways?
3. Why does parietal damage cause attention problems (neglect)?
4. How do we recognize objects (across locations, sizes, rotations with wildly different retinal images)?

Overview of the Visual System

Hierarchies of specialized visual pathways, starting in retina, to LGN (thalamus), to V1 & up:



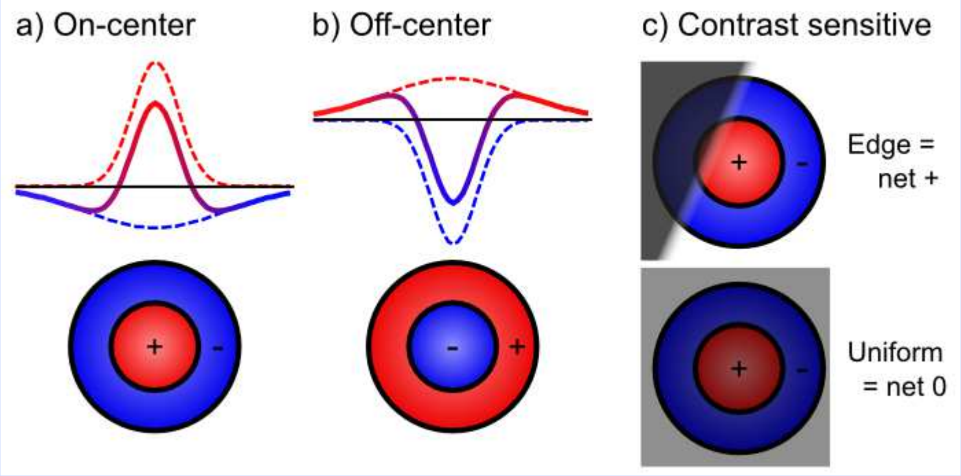
Two Streams: Ventral "what" vs. Dorsal "where"



The Retina

Retina is *not* a passive “camera”

Key principle: *contrast enhancement* that emphasizes *changes* over space & time.



retinal output ganglion cells

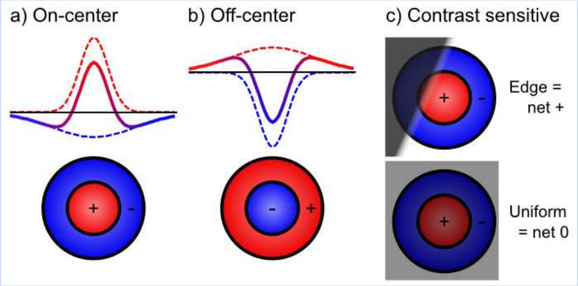
LGN of the Thalamus

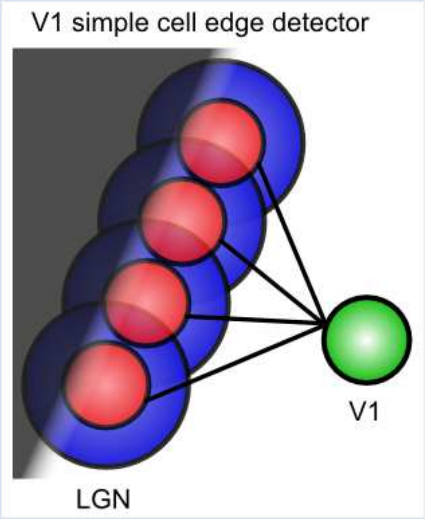
A “relay station”, but so much more.

- Organizes different types of information into different layers.
- Performs *dynamic* processing: magnocellular motion processing cells, *attentional* processing.
- On- and off-center information from retina is preserved in LGN

Primary Visual Cortex (V1): Edge Detectors

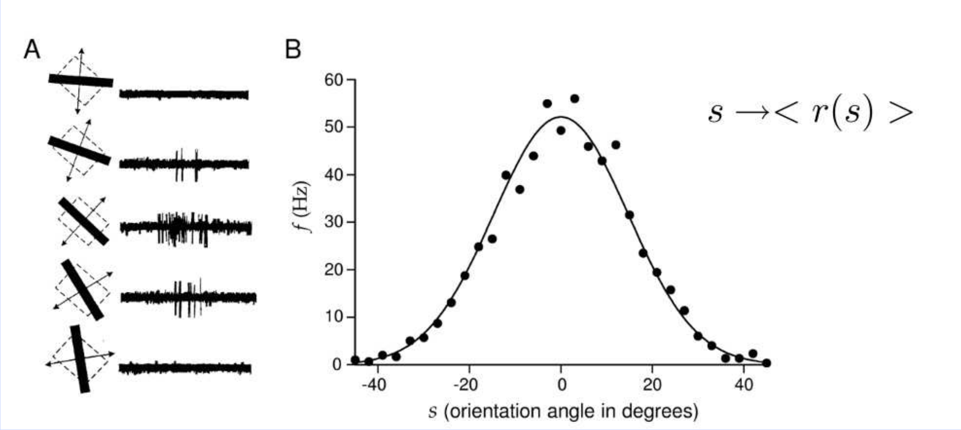
V1 combines LGN (thalamus) minputs into oriented *edge detectors*:





- Edges differ in orientation, size (spatial frequency), and position.
- For coherent vision, need to detect varying degrees of all these.

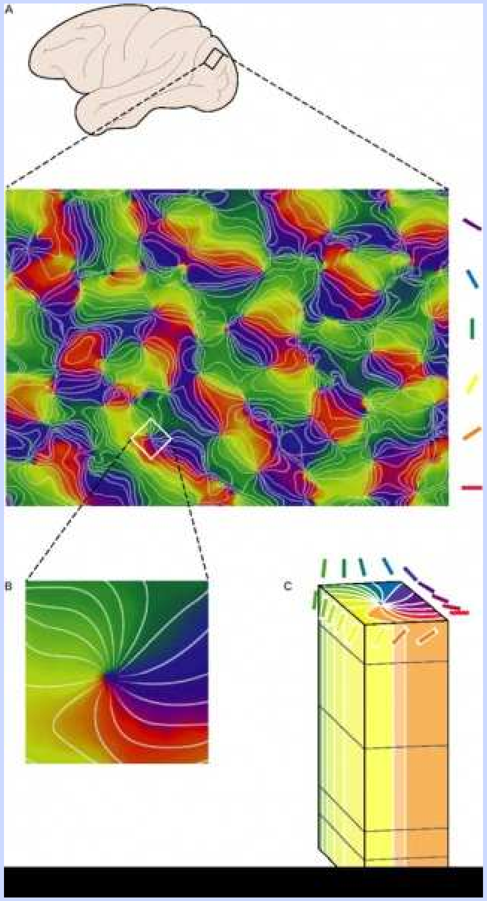
Example V1 edge detector



$s \rightarrow < r(s) >$

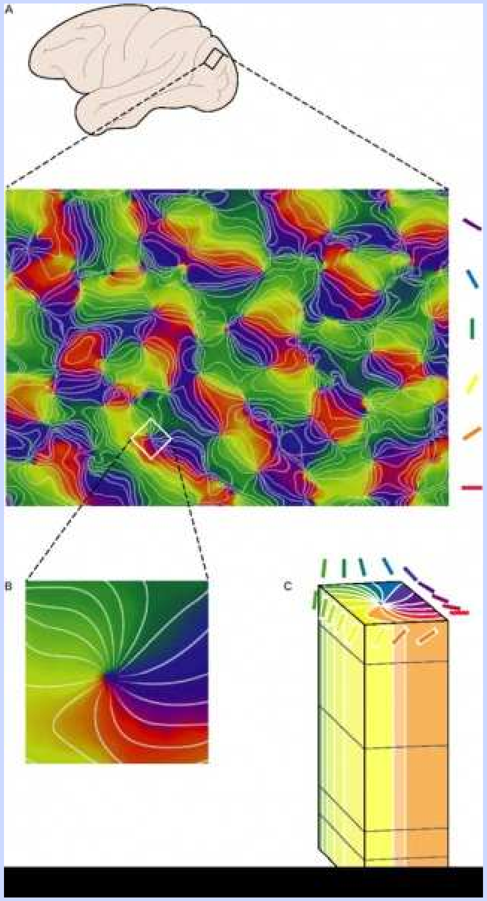
Hubel & Wiesel Nobel Prize

Primary Visual Cortex (V1): Topography



Hypercolumn: Full set of coding for each position

Primary Visual Cortex (V1): Topography

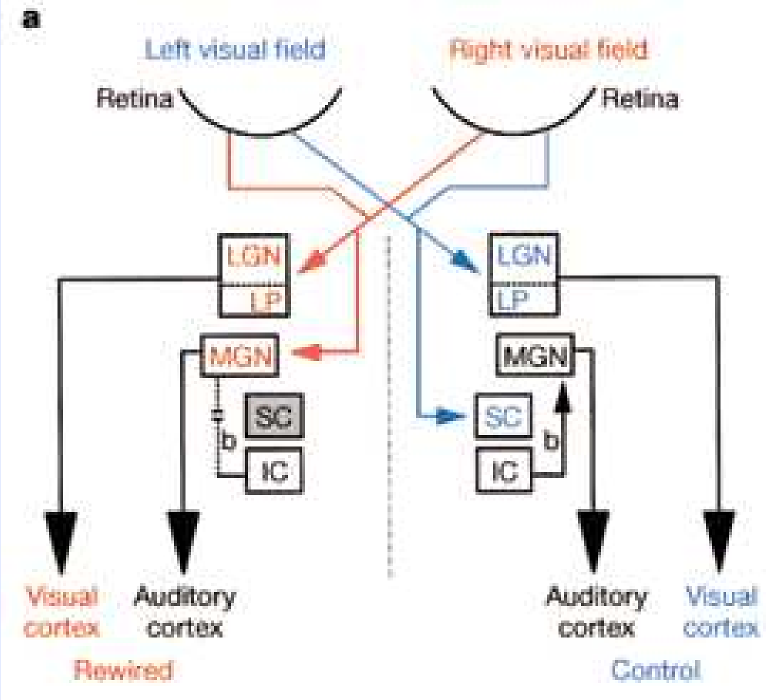


Hypercolumn: Full set of coding for each position

Pinwheel can arise from *learning* and lateral connectivity: not hard-wired!

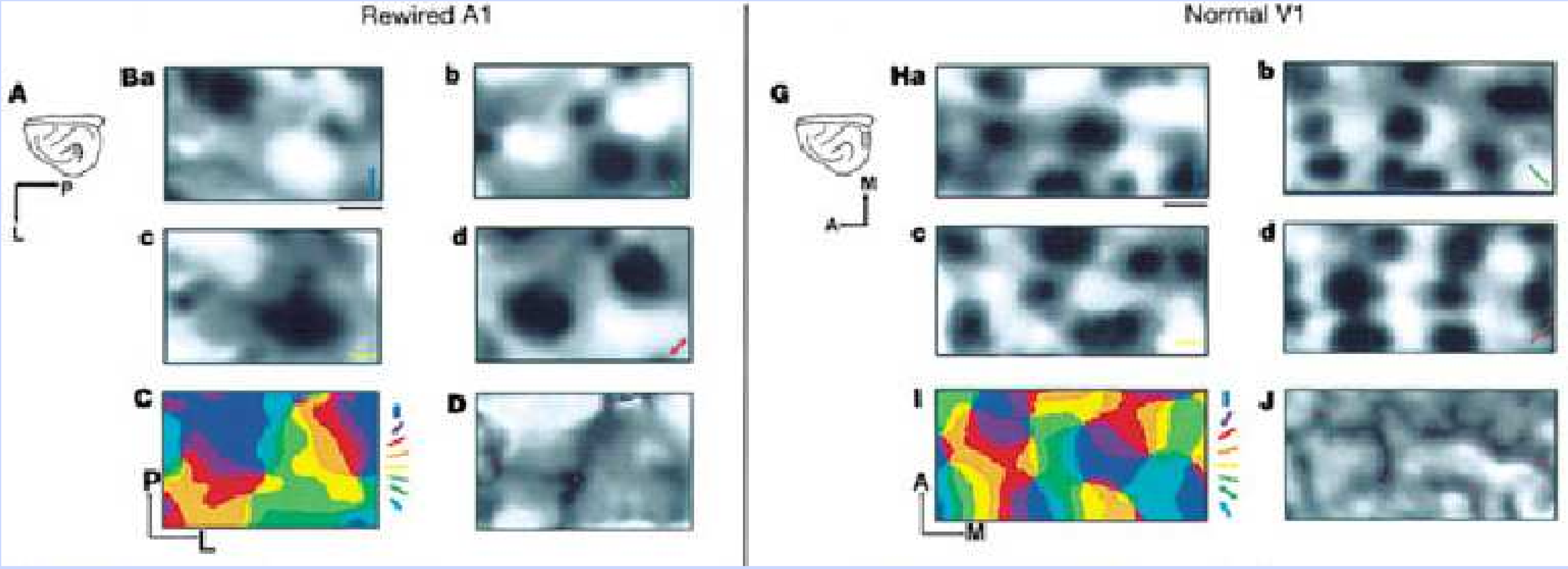
Rerouting of Visual Info to Auditory Cortex

- Sharma, Angelucci & Sur (2000), *Nature*
  Rerouted fibers from Retina→ auditory thalamus (MGN) →A1



- If visual properties are learned, they should develop in A1.

Rerouting of Visual Orientation Modules in A1



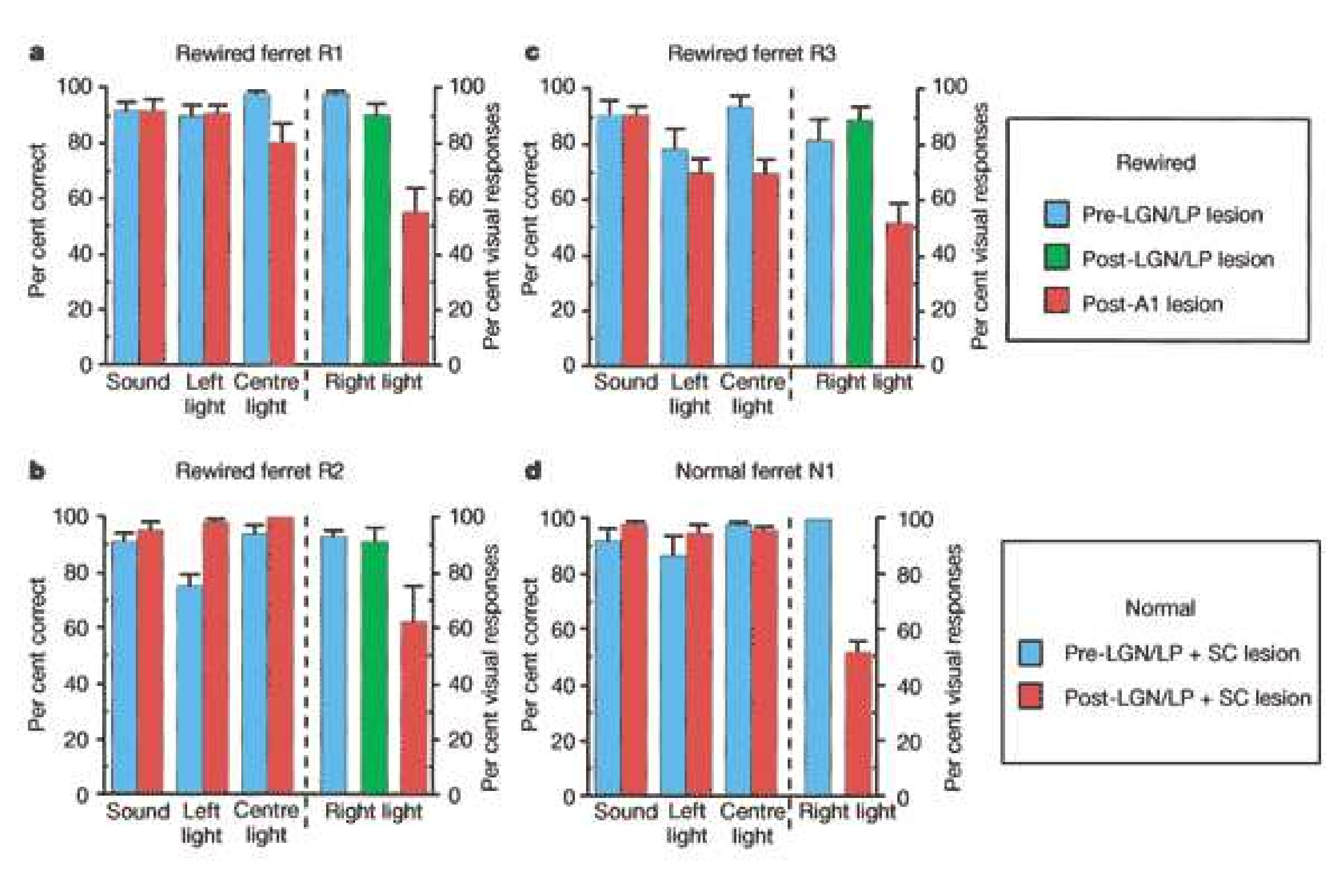
Ba-d: Orientation maps, dark - high act for given orientation (bottom right).

C: composite map of orientation preferences

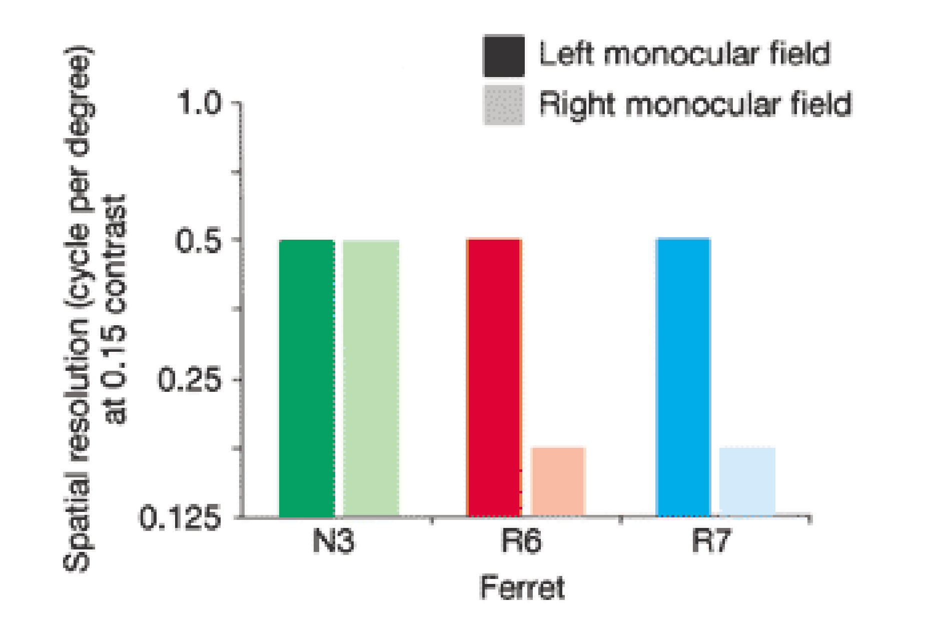
D: red dots = pinwheel centers

Visual Behavior After Rerouting Right Visual Field

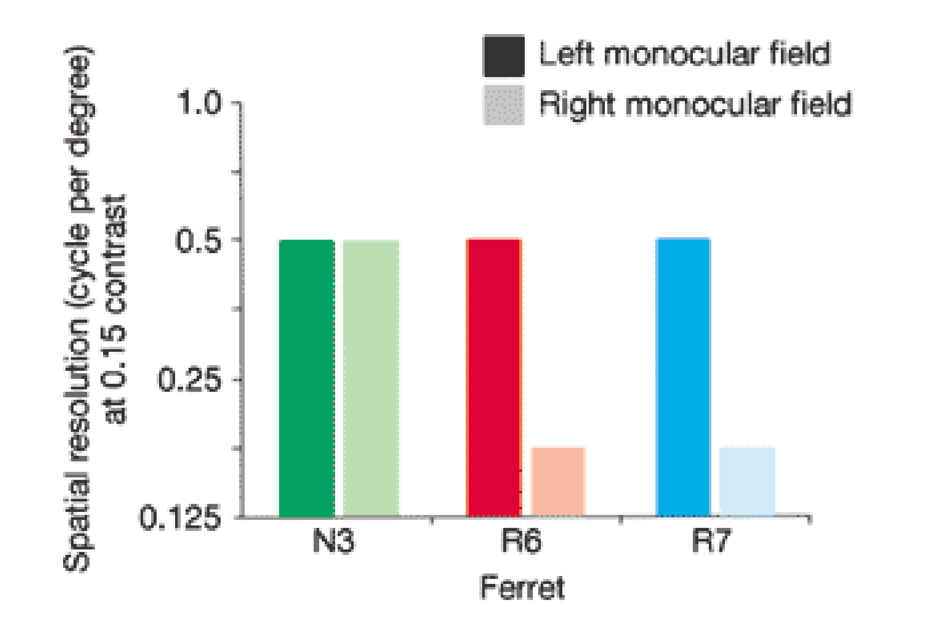
von Melchner, Pallas & Sur (2000)



Visual Acuity After Rerouting



Visual Acuity After Rerouting



→ So learning is powerful, but so is evolution!

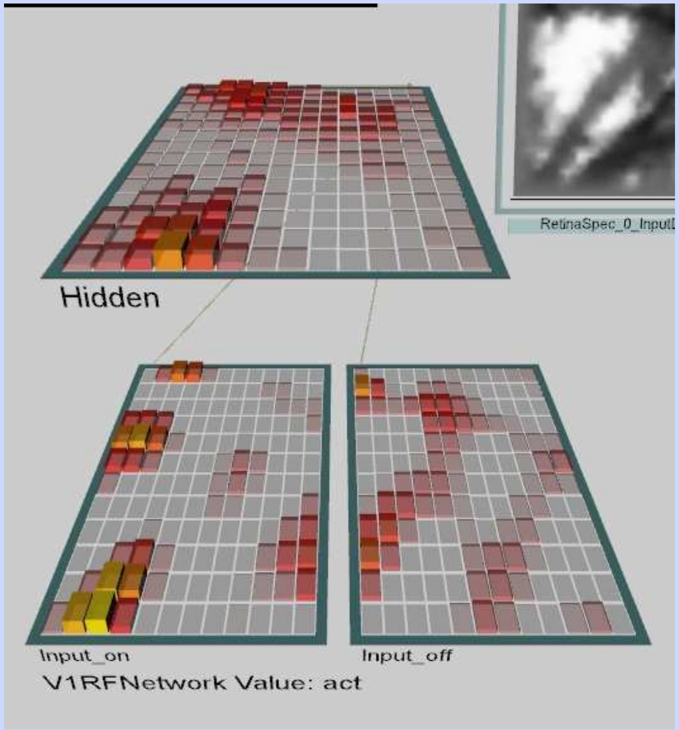
A Question

What makes visual cortex visual cortex? Why does it represent oriented bars of light?

Primary Visual Representations

Key idea: Oriented edge detectors can develop from Hebbian correlational learning based on natural visual scenes.

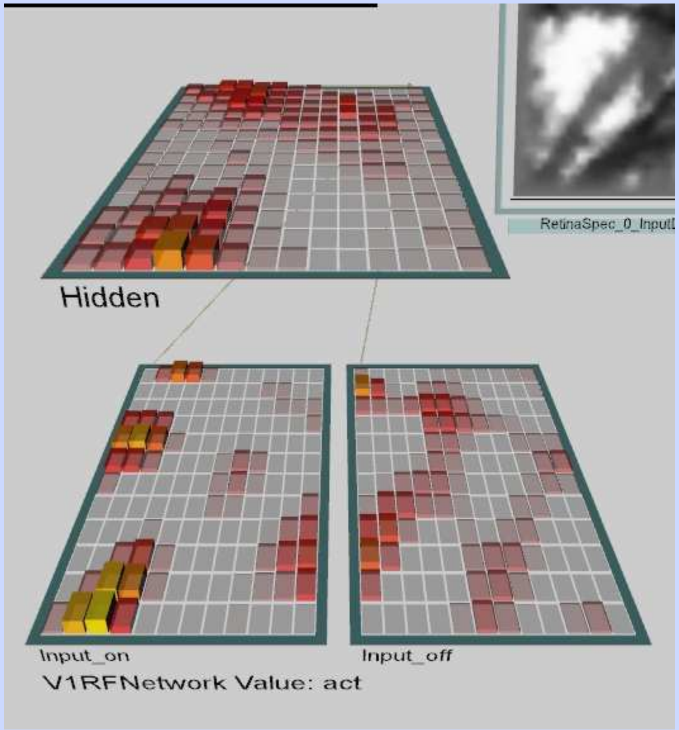
The Model: Simulating one Hypercolumn



- Natural visual scenes are preprocessed by passing them (separately) through layers of on-center and off-center inputs

- Hidden layer: edge detectors seen in layers 2/3 of V1; Layer 4 (input) just represents unoriented on/off inputs like LGN (but can be modulated by attention)

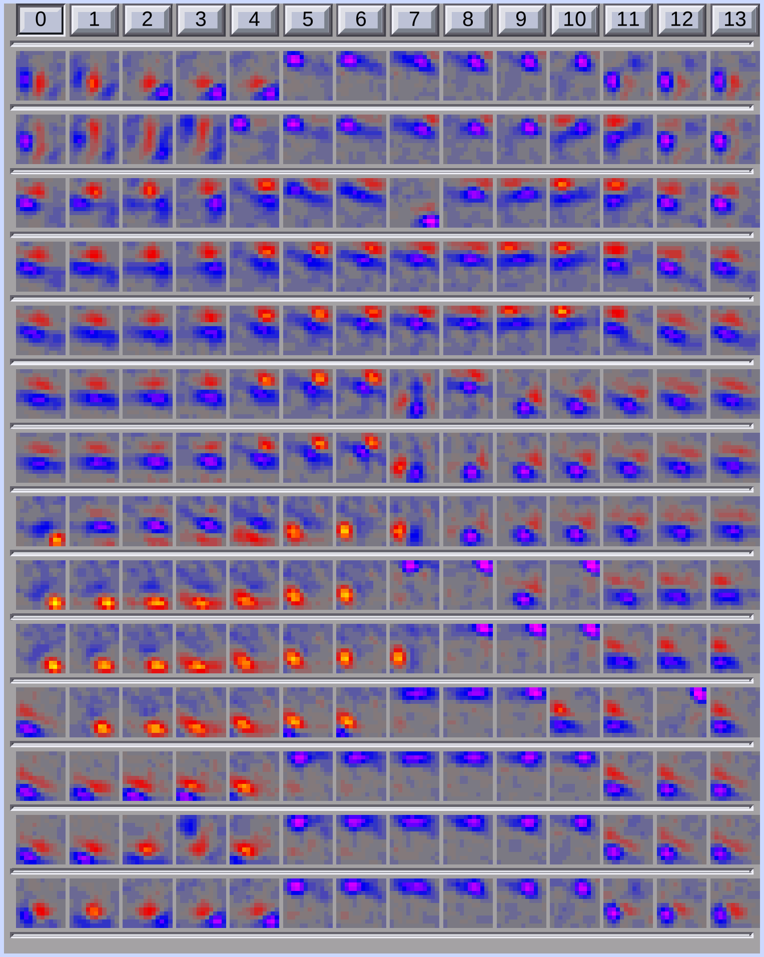
The Model: Simulating one Hypercolumn



- Hebbian learning only
- FFFB inhib competition for specialization (see Ch 4)

[v1rf.proj]

The Receptive Fields

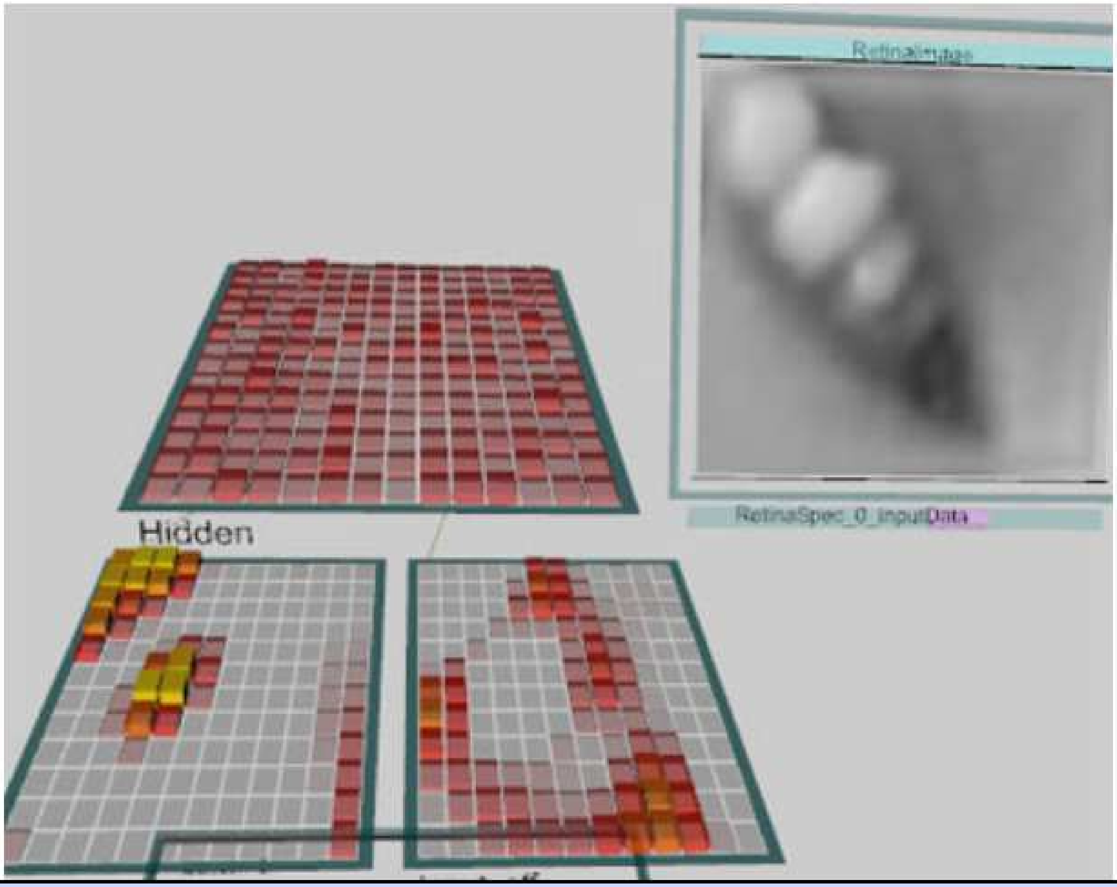


Red = on-center > off-center, Blue = off-center > on-center

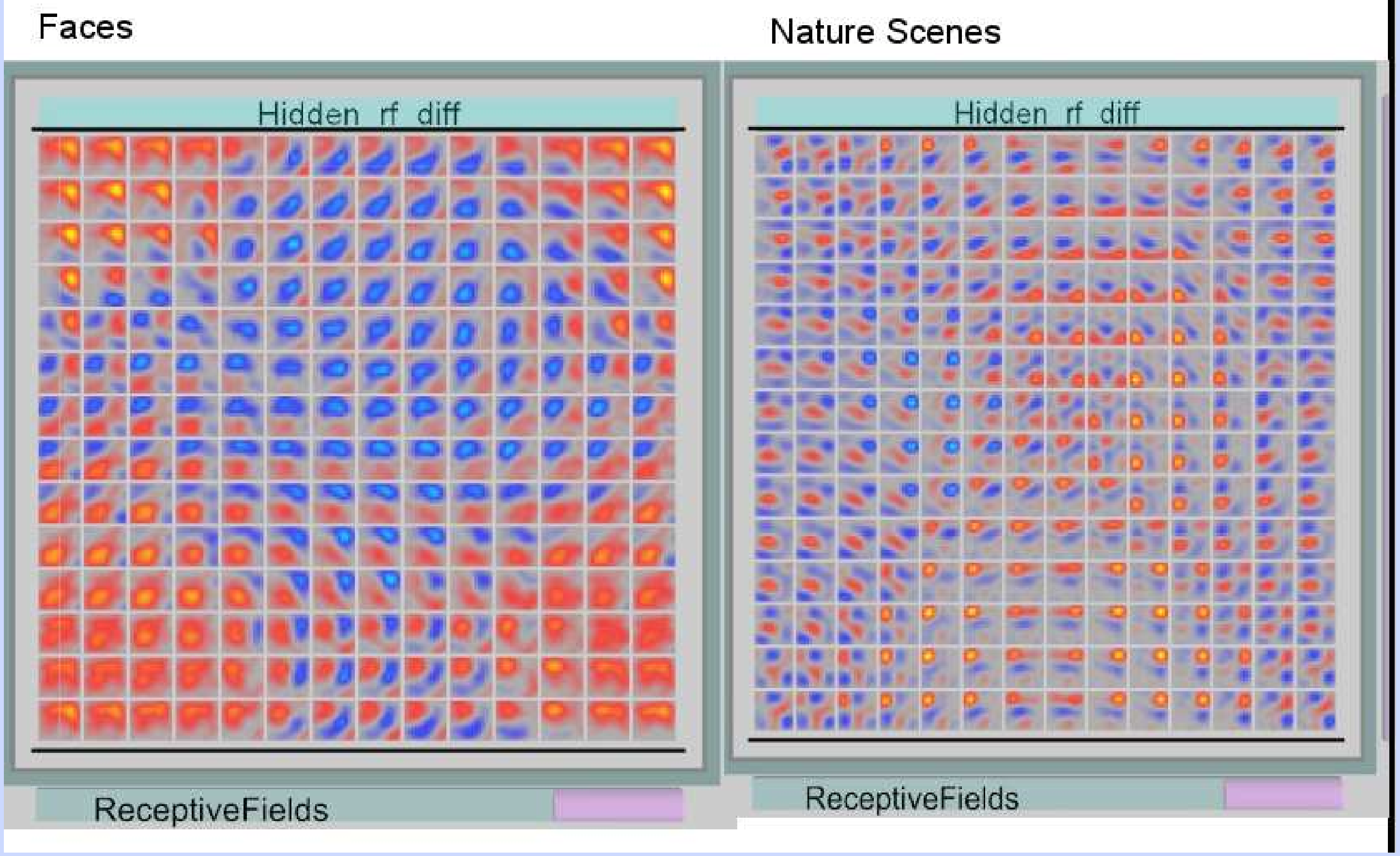
How many babies
live on mountain tops?

What about training on mother's faces??

Model Training on Faces



Difference after 100 Epochs



Some differences, but pinwheels still emerge

Perception and Attention

1. Why does primary visual cortex encode oriented bars of light?
   Correlational learning based on natural visual scenes.

Reflects reliable presence of edges in natural images, which vary in size, position, orientation and polarity.

→ model shows how documented V1 properties can result from interactions between learning, architecture (connectivity), and structure of environment.

Perception and Attention

1. Why does primary visual cortex encode oriented bars of light?
   Correlational learning based on natural visual scenes.

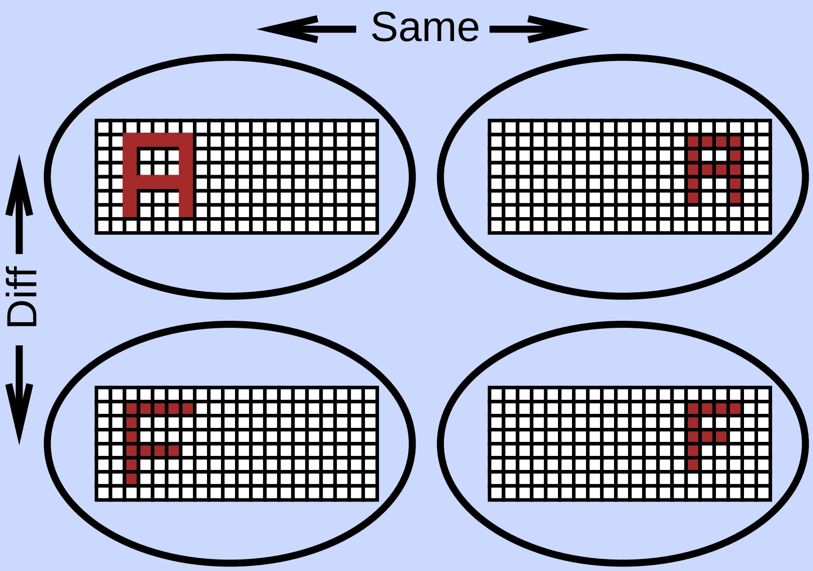
2. How do we recognize objects (across locations, sizes, rotations with wildly different retinal images)?

3. Why is visual system split into what/where pathways?

4. Why does parietal damage cause attention problems (neglect)?

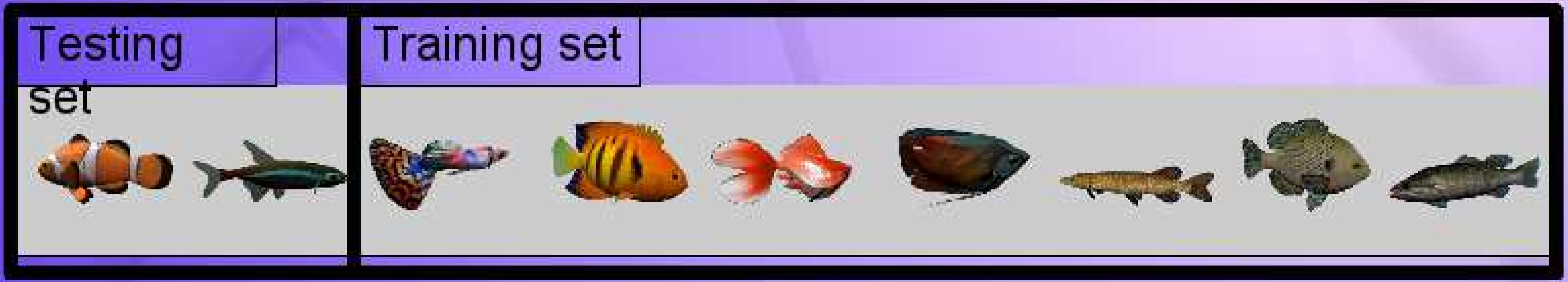
The Object Recognition Problem

Problem: Recognize object regardless of: location, size, rotation.



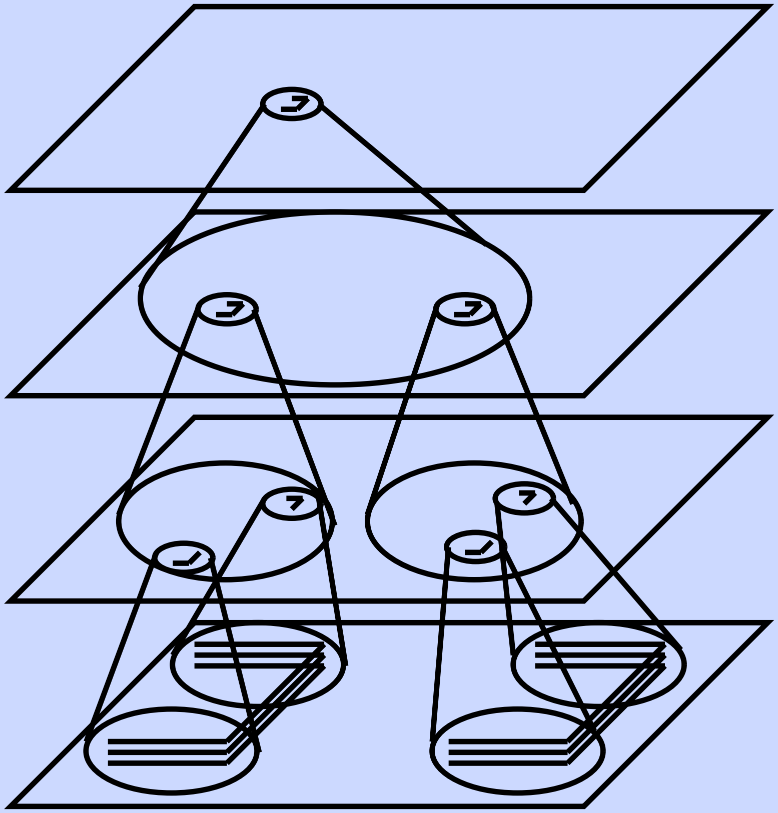
This is hard because different patterns in same location can overlap a lot, while the same patterns in different locations/sizes/rotations can not overlap at all!

Object Recognition is Hard

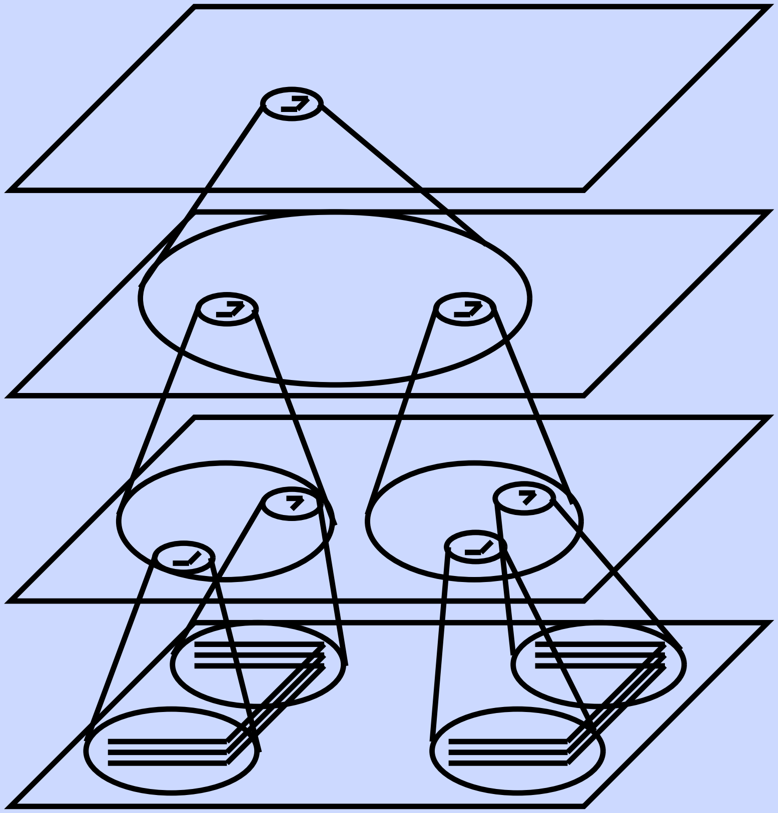


- Large amount of shape variability within and between categories
- Huge amount of view-based variability (position, orientation, size, rotation)

Gradual Invariance Transformations (Fukushima, ’80)



Gradual Invariance Transformations (Fukushima, ’80)

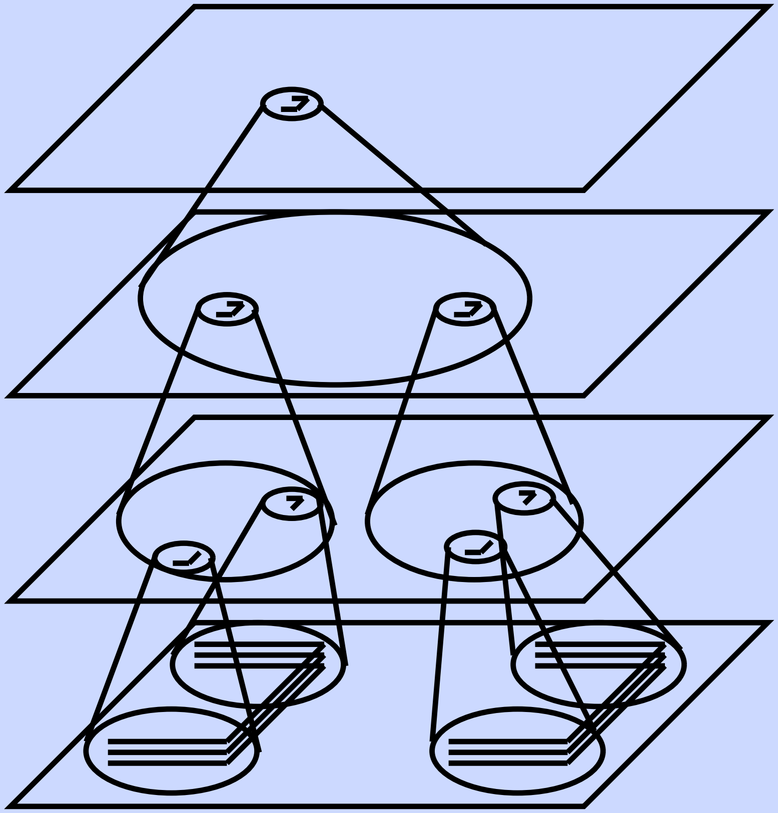


Increasing receptive field size enables:

Conjunction of features (to form more complex objects); and

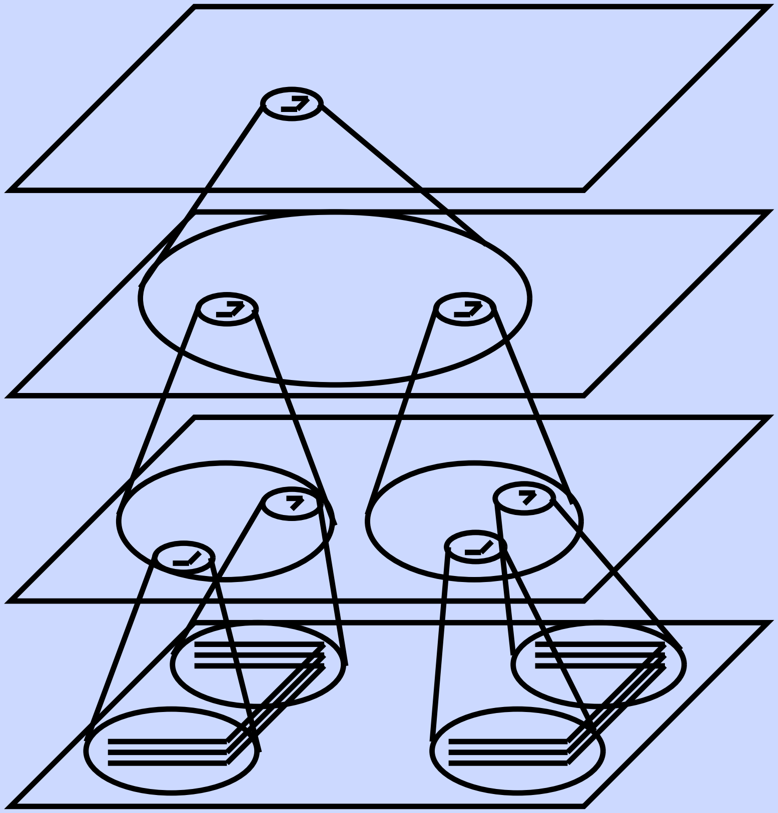
Collapsing over location information (“spatial invariance”)

Gradual Invariance Transformations (Fukushima, ’80)



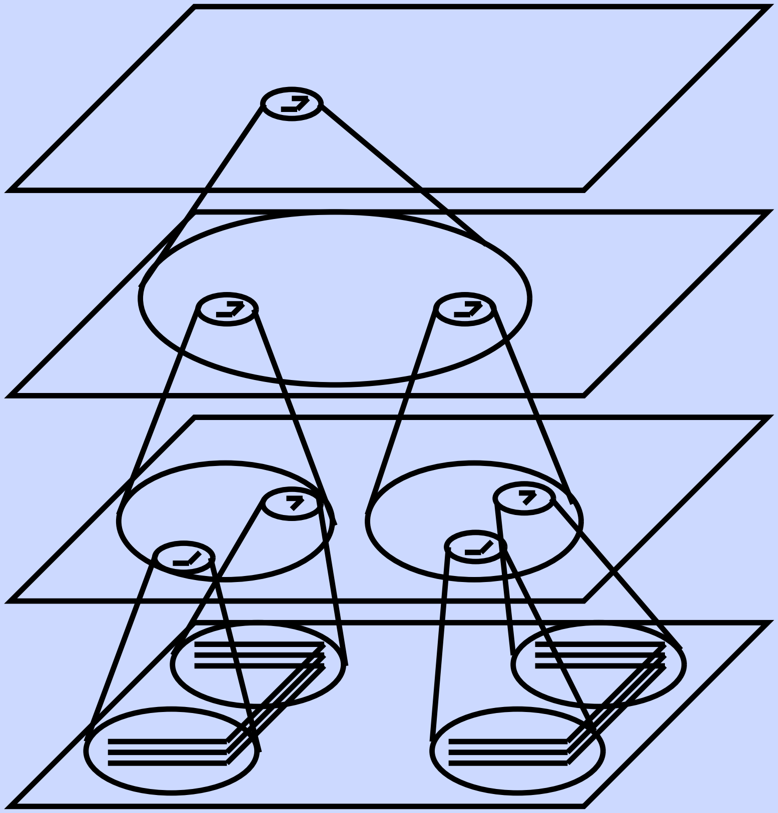
if did spatial invariance in one fell swoop: binding problem - can’t tell T from L

Gradual Invariance Transformations (Fukushima, ’80)



Goal: Units at the top of the hierarchy should represent complex object *features* in a location and size invariant fashion

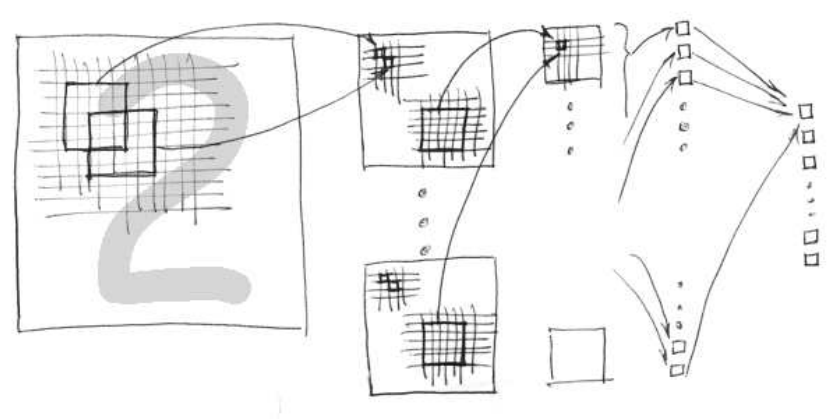
Gradual Invariance Transformations (Fukushima, ’80)

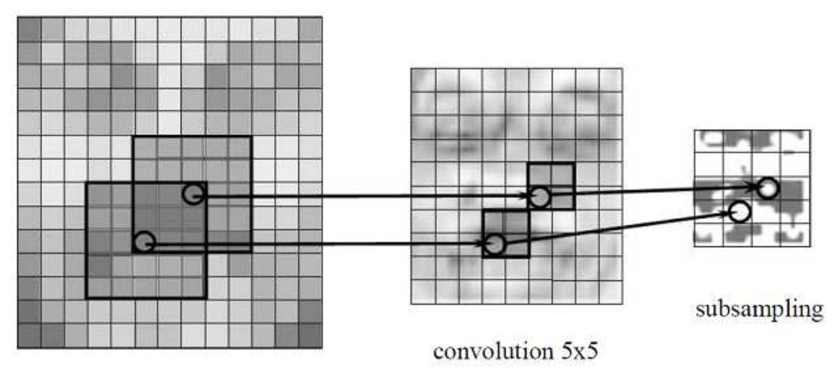


Goal: Units at the top of the hierarchy should represent complex object *features* in a location and size invariant fashion

(also want benefits of top-down amplification, pattern completion, distributed reps etc)

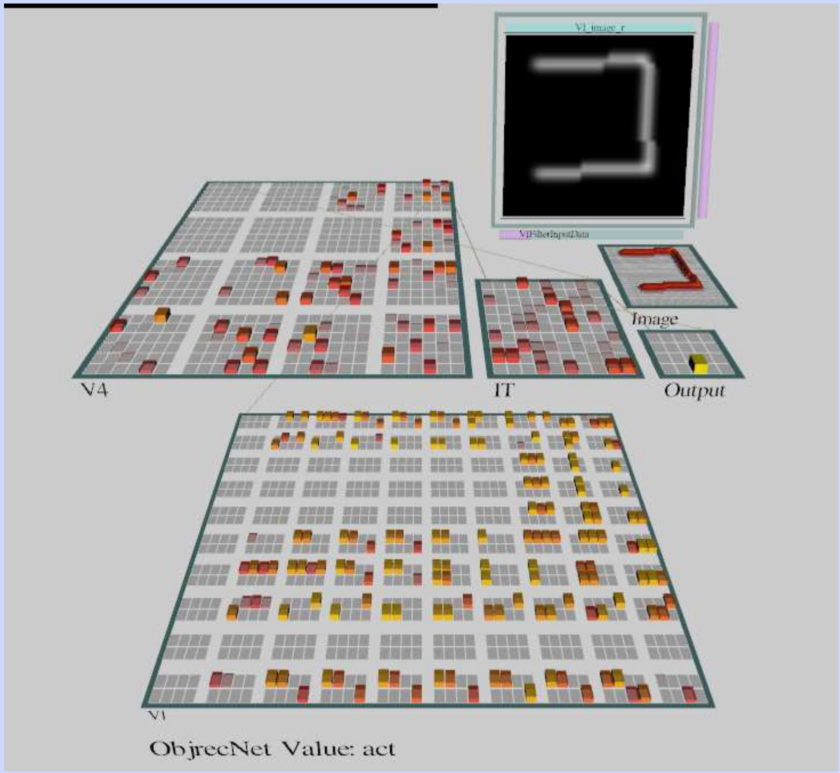
“Convolutional Neural Networks”





- very popular for “deep learning” in machine learning, Yan LeCun, Hinton etc.

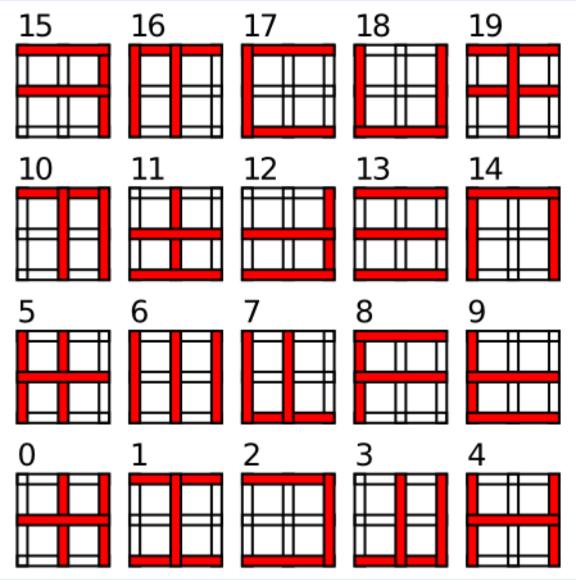
The Model: combining Fukushima with convolutional neural nets, bidirectional connectivity and learning!



V1 = oriented line (edge) detectors, hard-coded
V4 units encode conjunctions of V1 edges across a subset of space
Each IT unit pays attention to all of V4

(V2 omitted here, important for figure-ground etc)

The Objects



Each object is presented at multiple locations, sizes

Network's job is to activate the appropriate Output unit (0-19) for each object, regardless of location and size

[objrec.proj]

Activation-Based Receptive Fields

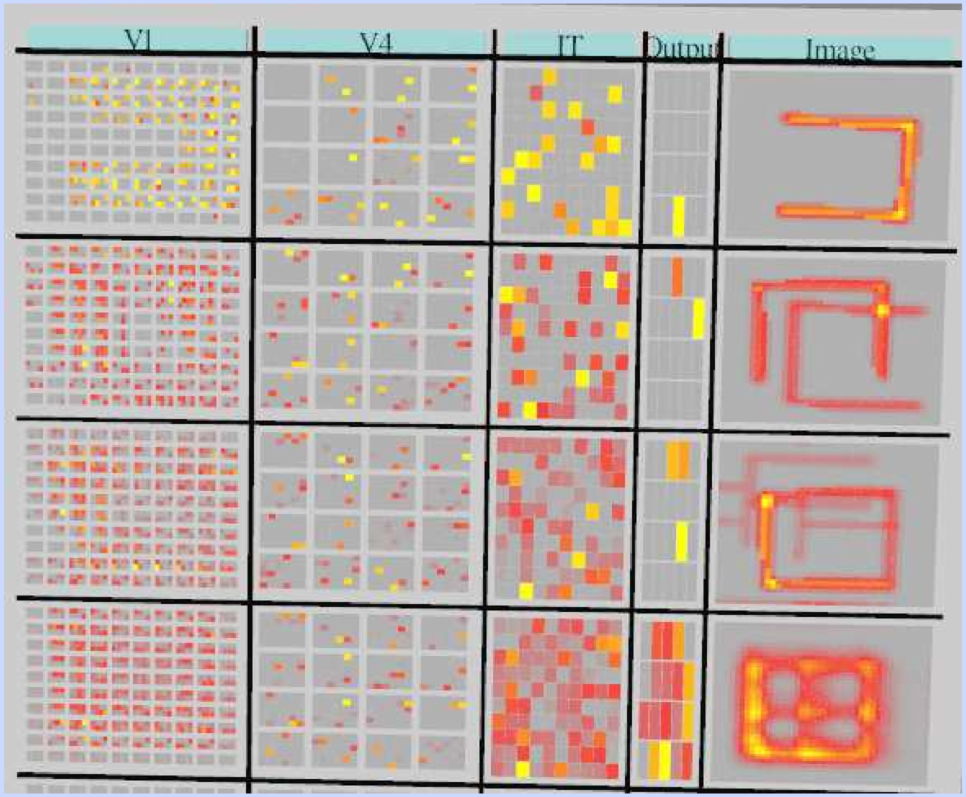
- How do we plot receptive fields for V4?

- Receiving weights show which V1 units a V4 unit responds to, but they don’t show what *thing in the world* the unit responds to

- Solution: Show the network lots of input patterns.

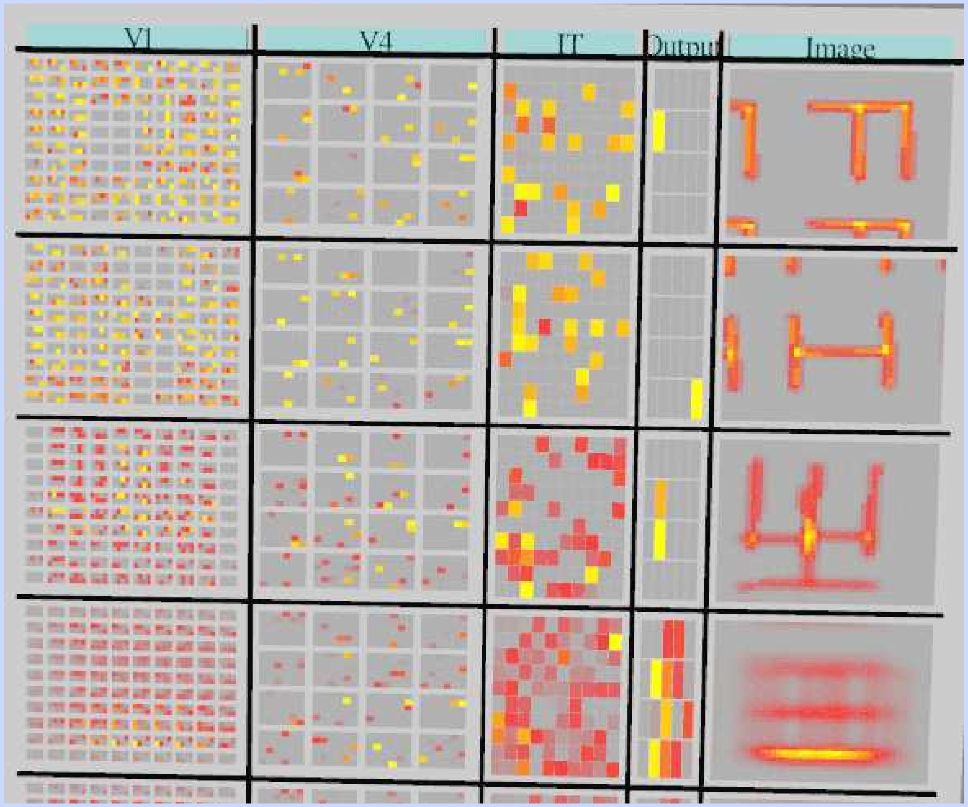
- Then, display a *composite* of all the input patterns that activate the unit.

V4 Receptive Fields



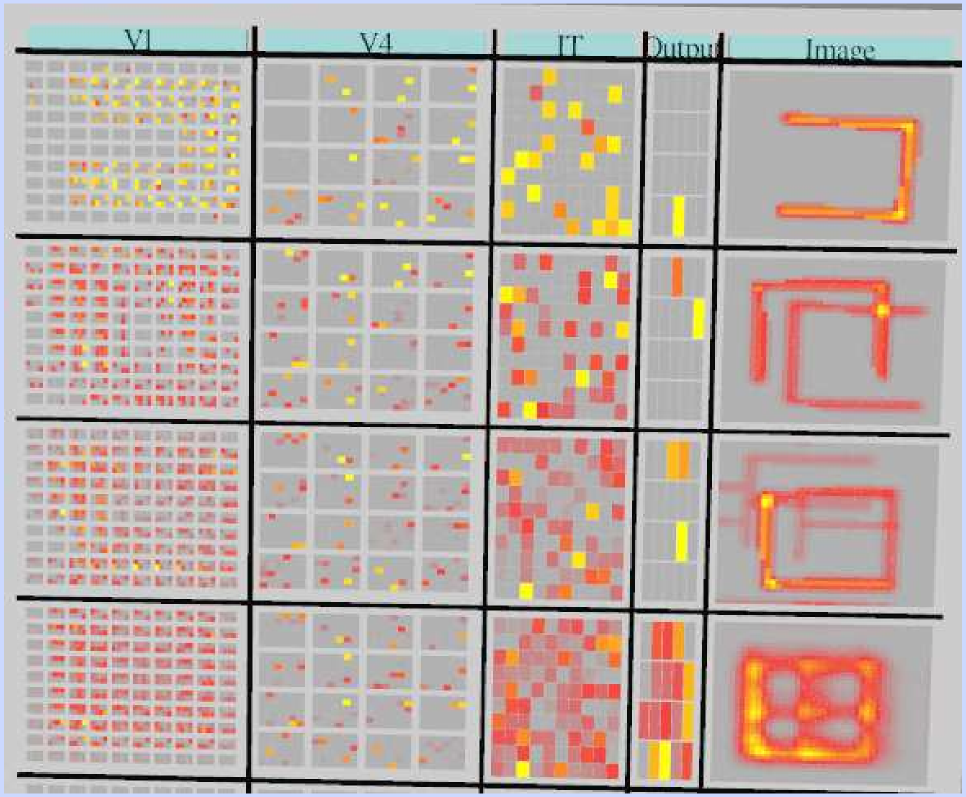
- Some V4 units code for location-specific conjunctions of V1 features
  - This will show up as a sharp receptive field for Image input

V4 Receptive Fields



- Some V4 units code for simple features in a location invariant way
  - This will show up as smeary parallel lines in Image input

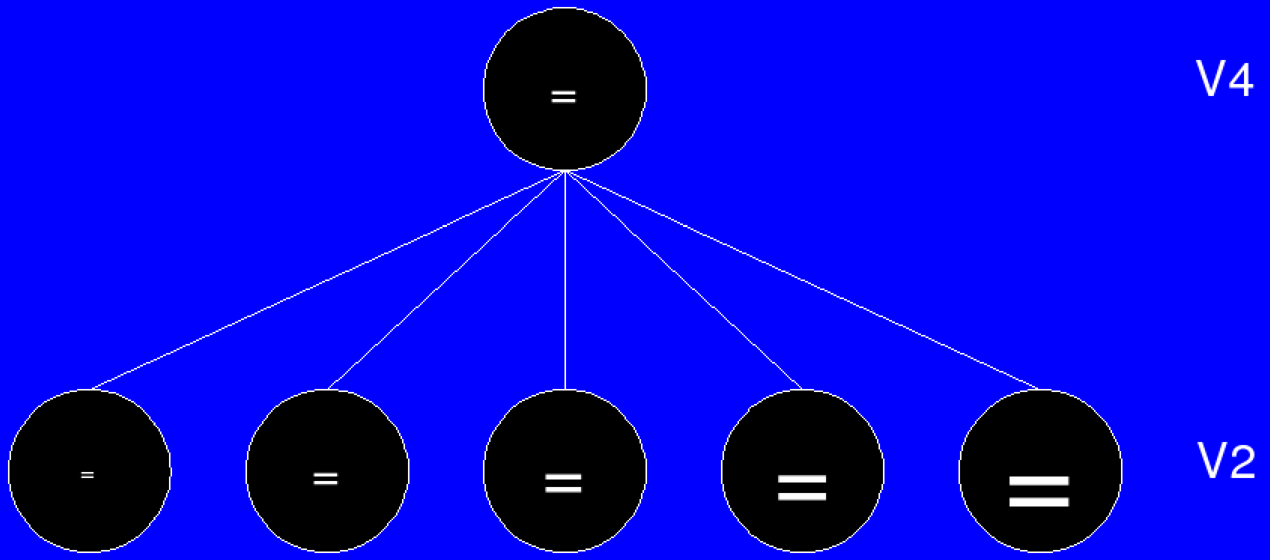
V4 Receptive Fields for Output



- Can also look at which Output units tend to get active for any given V4 unit
  - Generally a given V4 unit is associated with multiple objects

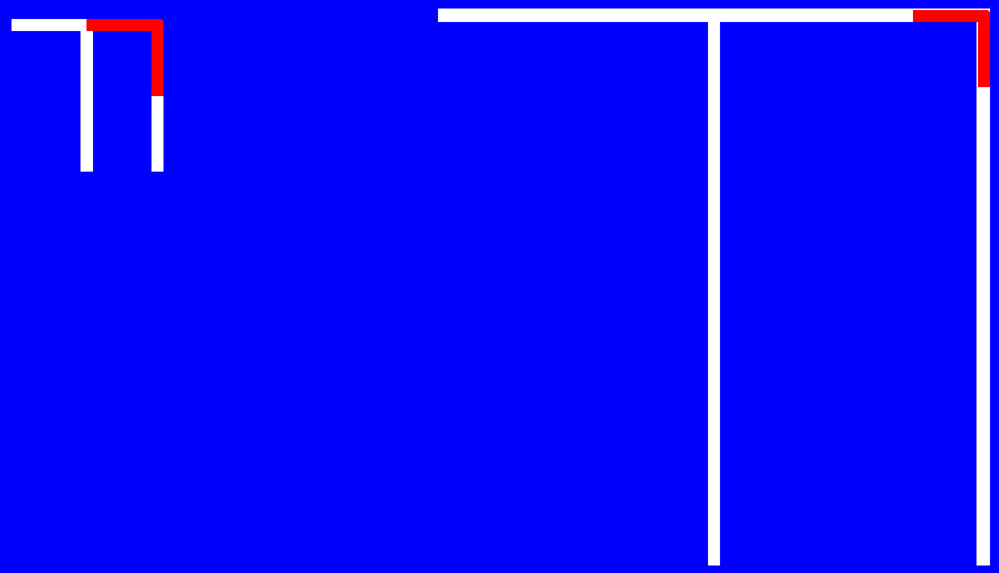
Size Invariance

- One approach to this problem is to have V4 units respond to all of the V2 units that represent a feature (regardless of size)



Size Invariance

- Another approach to this problem is to **pick features that are invariant across size transformations**
- e.g., for this set of objects, corners are good!

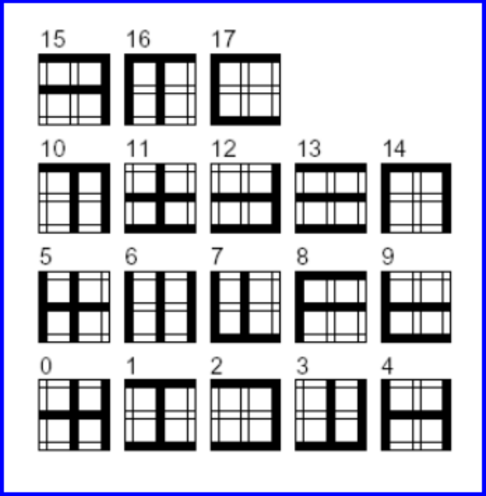


Generalization

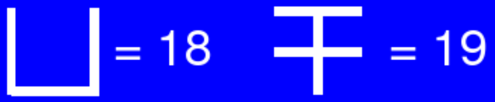
- Can the network generalize to unseen views of studied objects?
- In other words: Does training the net to recognize a set of objects in a size/location invariant fashion help it recognize new objects in a size/location invariant fashion?
- Procedure:
  - Take a net trained on 18 objects
  - Train with 2 new objects in only some locations/sizes
  - Test the net with nonstudied “views” (sizes/locations) of new objects

Generalization

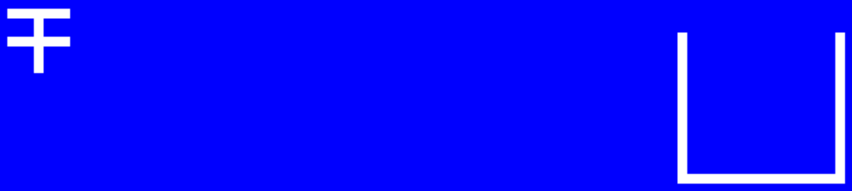
- Train on these using multiple sizes/locations



- Then train on two new objects (using a limited number of sizes/locations)



- Test on new sizes/locations:



Generalization

- Can the network generalize to unseen views of studied objects? *yes*

- Approx. 90% correct on novel views following training on just 6% of possible sizes/locations

Generalization

- Can the network generalize to unseen views of studied objects? *yes*

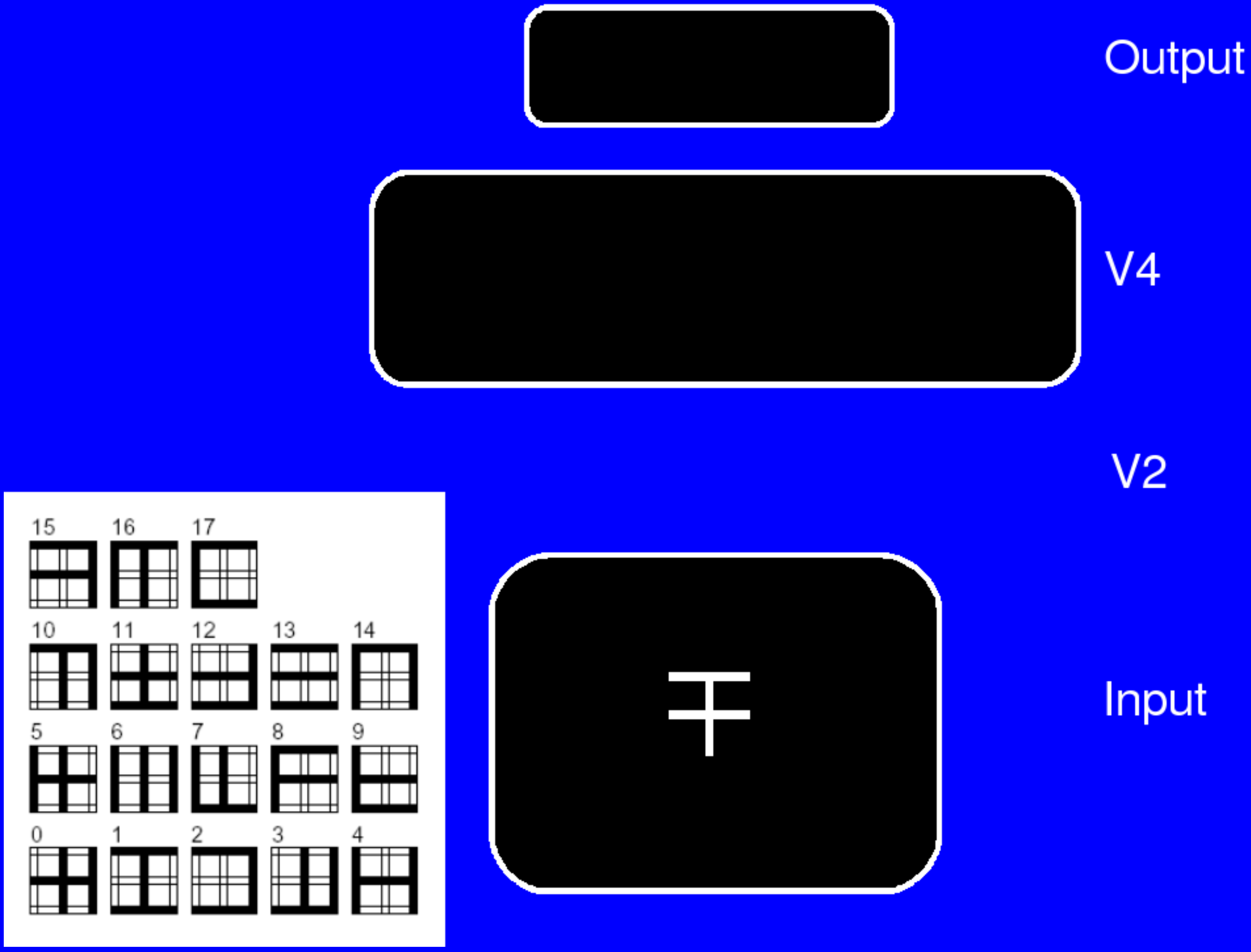
- Approx. 90% correct on novel views following training on just 6% of possible sizes/locations

  Explanation: Distributed representations and Hebb learning!

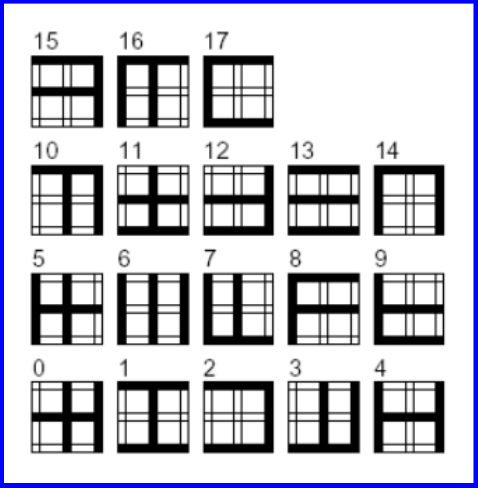
- V4 represents object **features** in a location/size invariant way

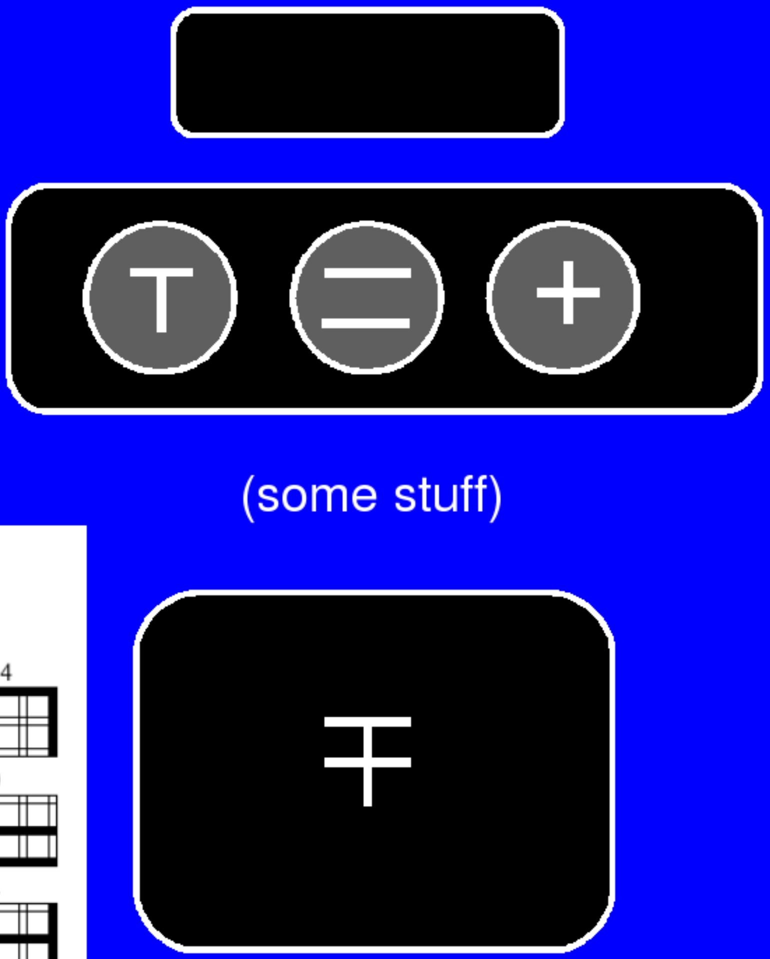
- Each object activates a distributed pattern of these invariant feature detectors

Generalization



Generalization





Output

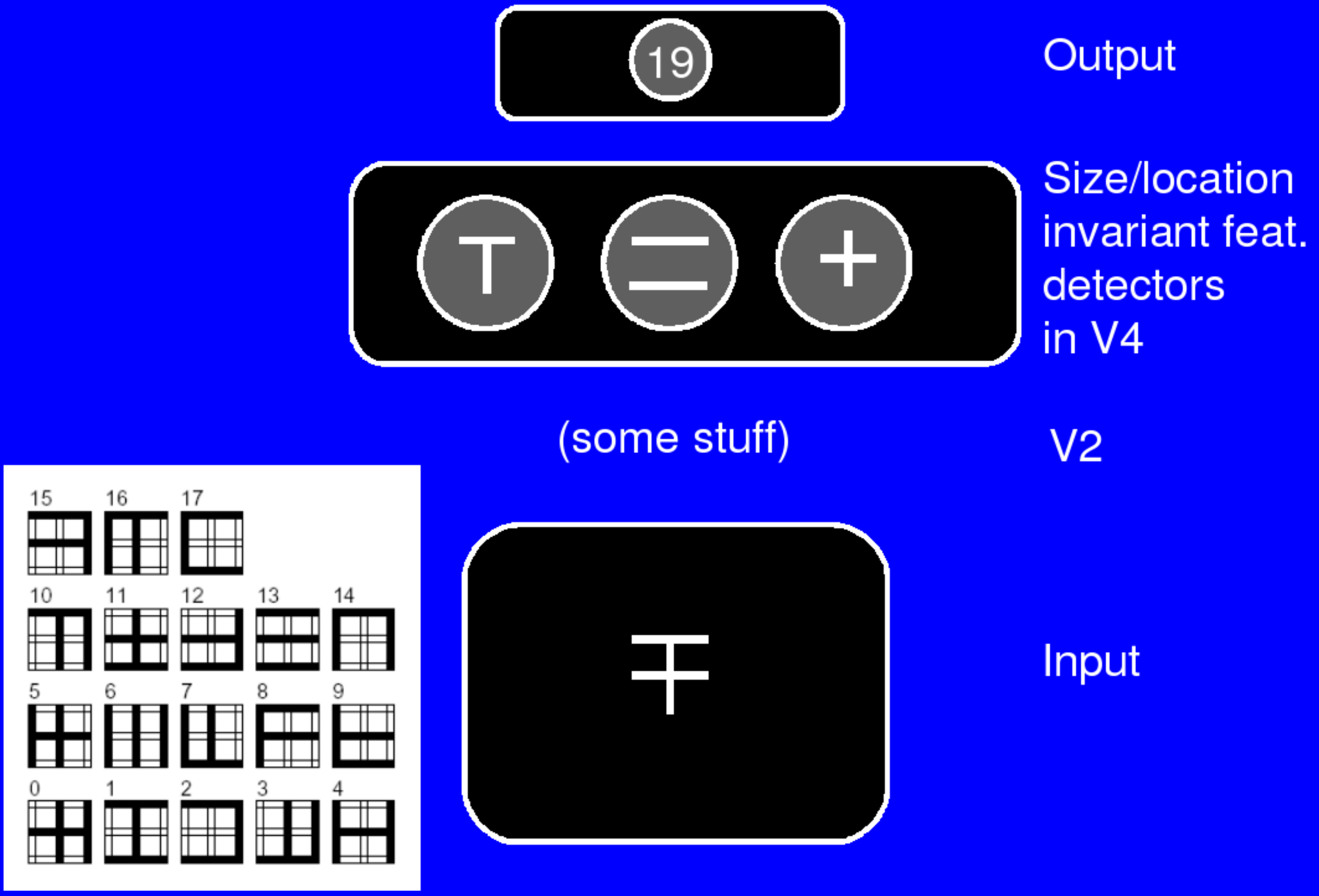
Size/location
invariant feat.
detectors
in V4

(some stuff)

V2

Input

Generalization



Output

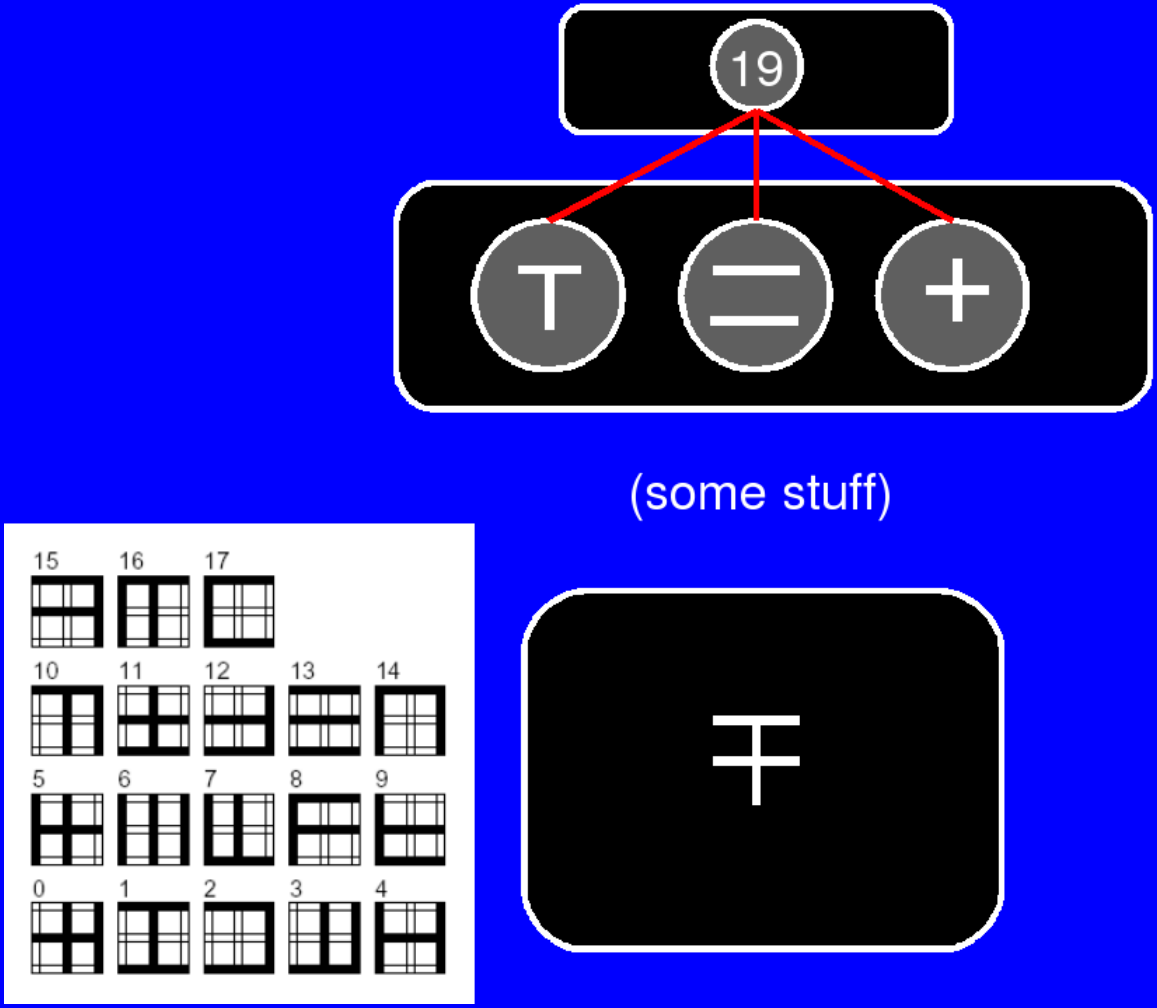
Size/location invariant feat. detectors in V4

(some stuff)

V2

Input

Generalization



Output

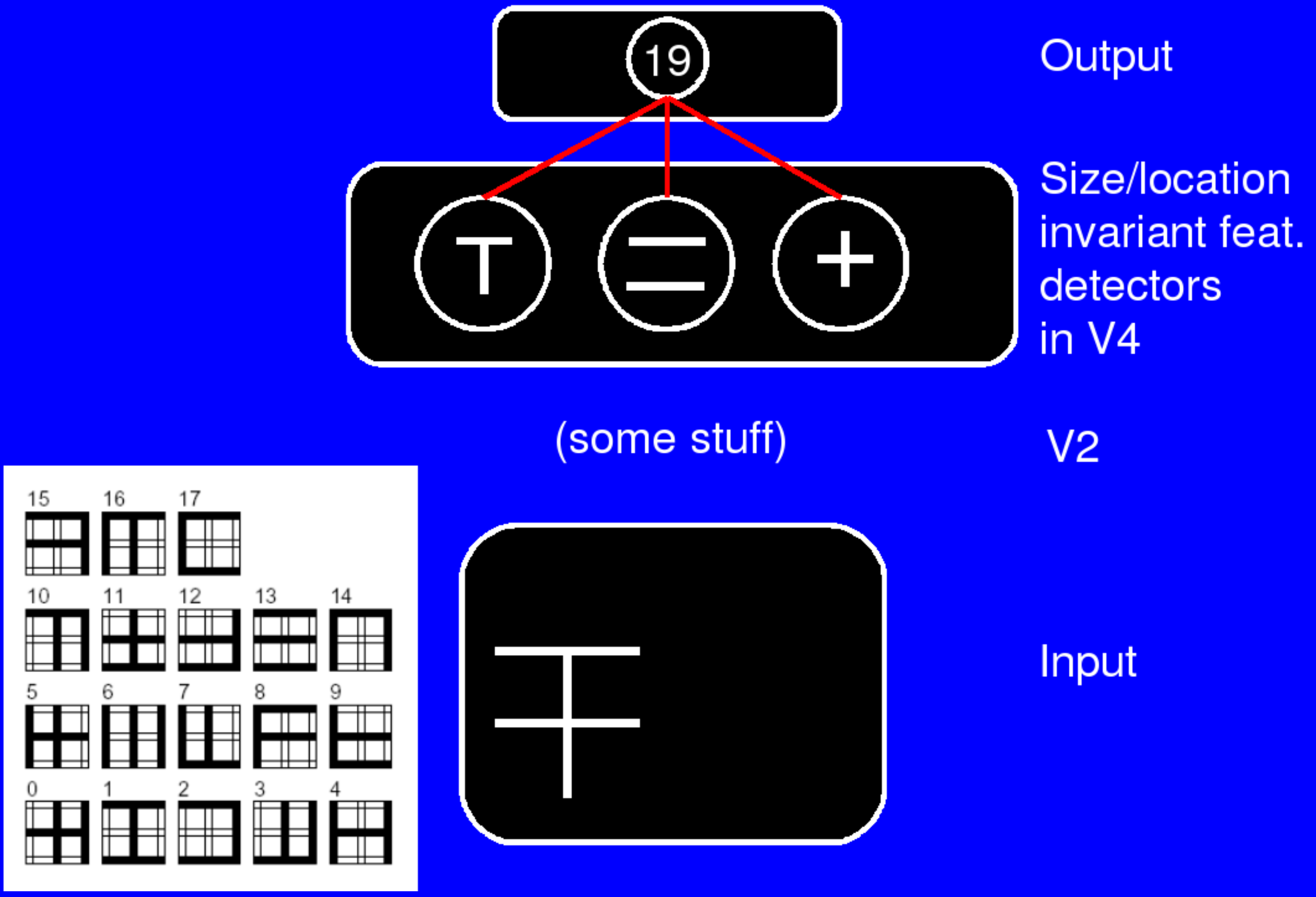
Size/location invariant feat. detectors in V4

(some stuff)

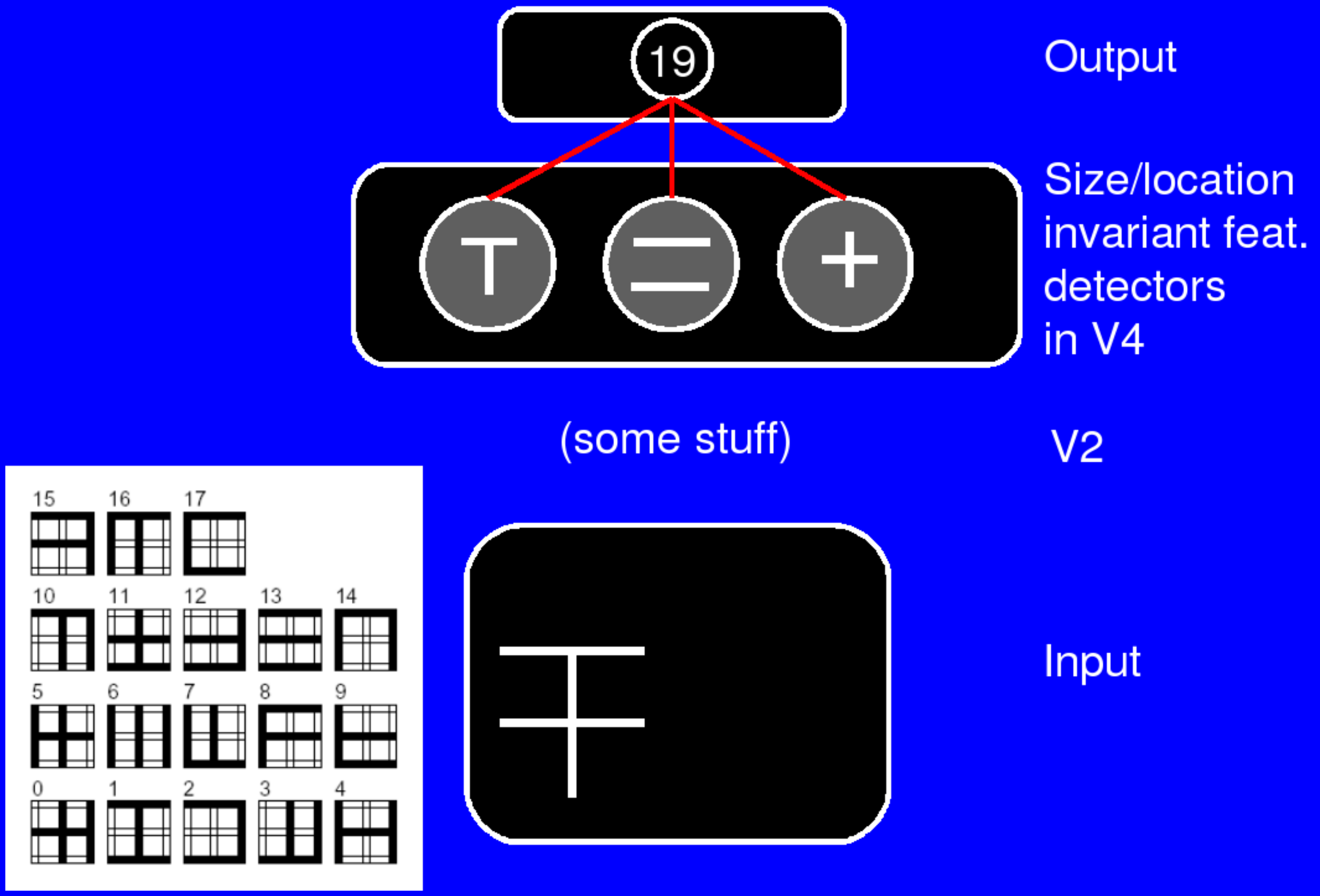
V2

Input

Generalization



Generalization



Output

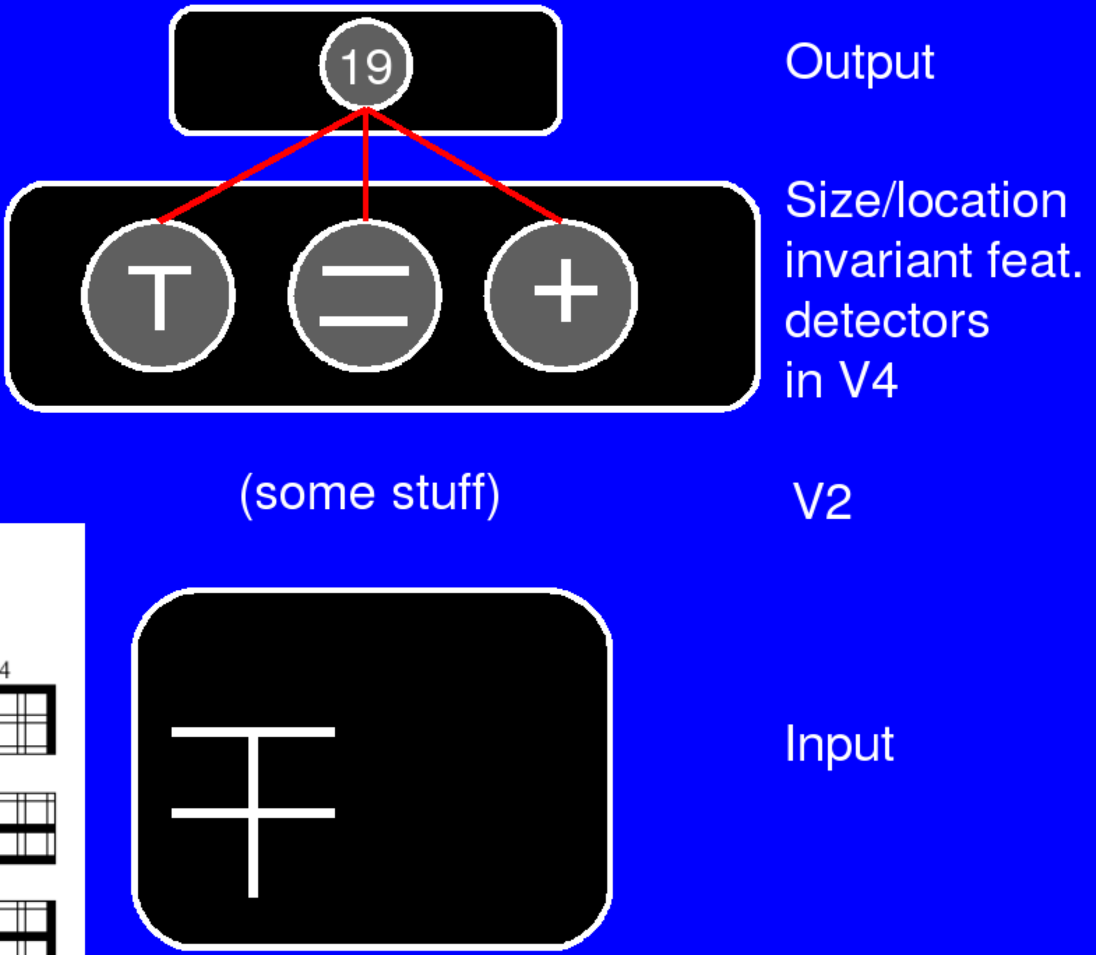
Size/location invariant feat. detectors in V4

(some stuff)

V2

Input

Generalization



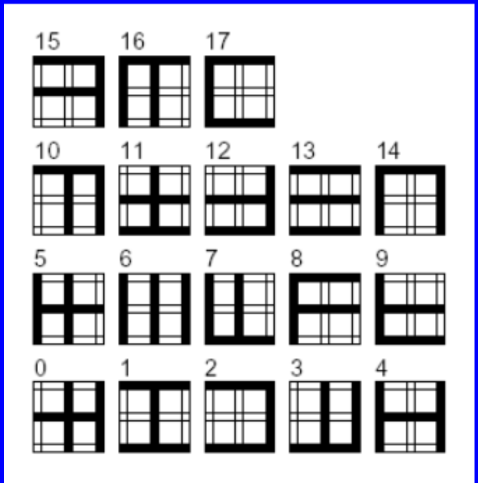
Output

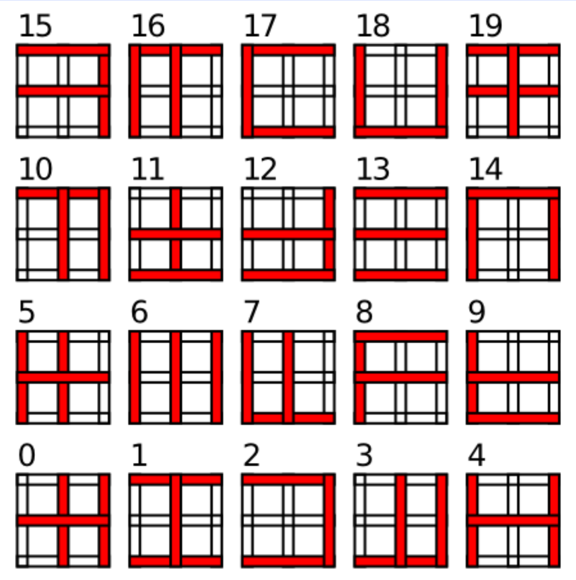
Size/location
invariant feat.
detectors
in V4

(some stuff)

V2

Input

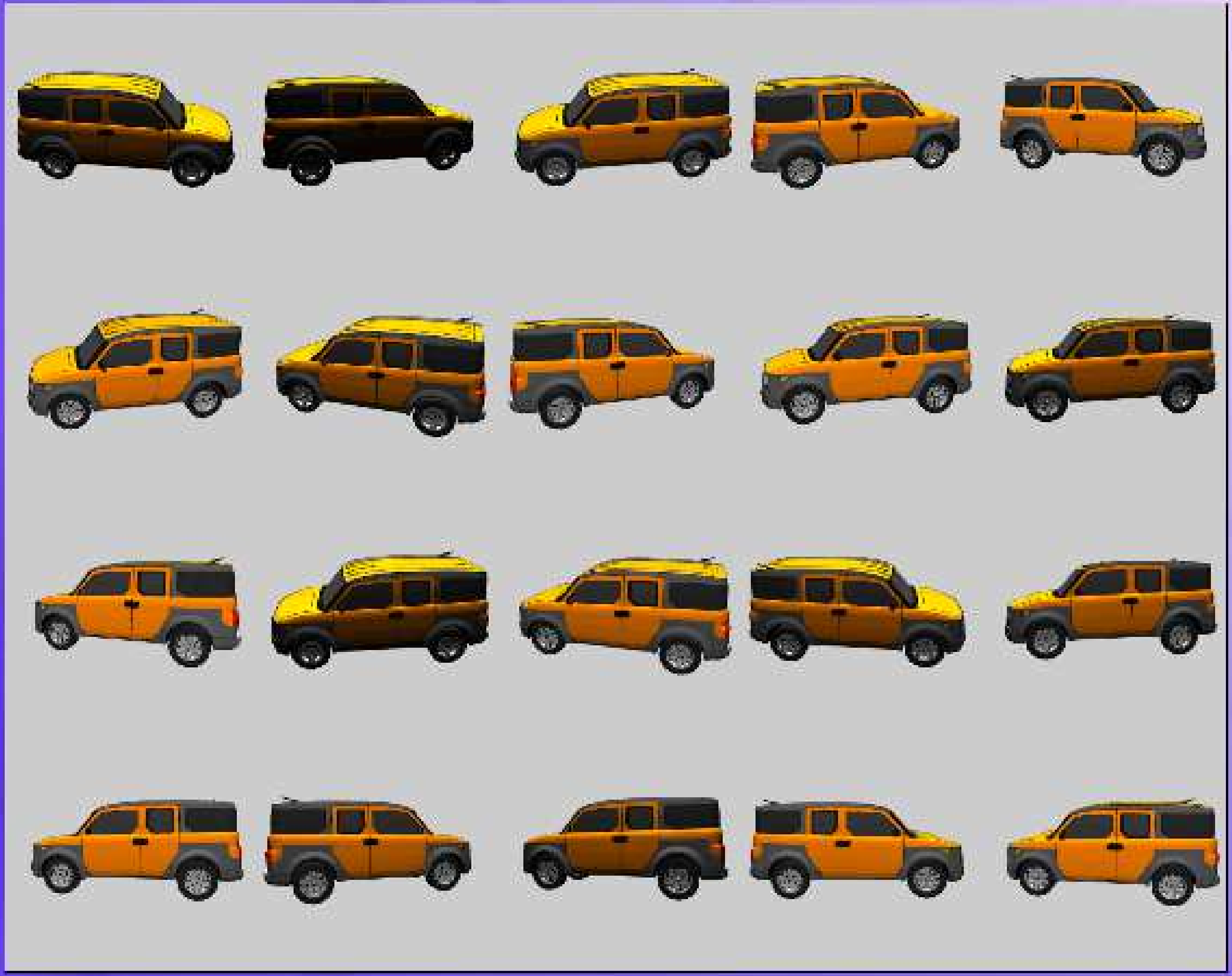




Yeah, but these objects are regularly shaped, straight lines...
what about real objects?

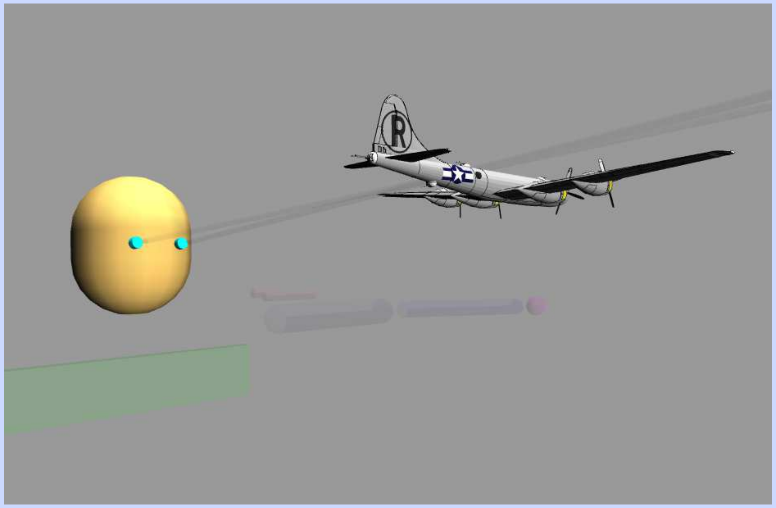
3D Object Recognition Test

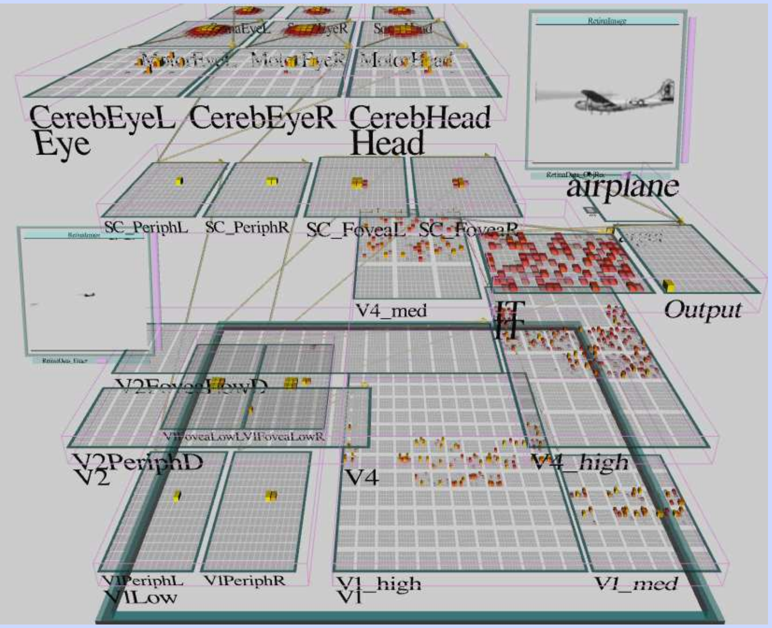
- From Google SketchUp Warehouse
- 100 categories
- 8+ objects per categ
- 2 objects left out for testing
- +/- 20° horiz depth rotation + 180° flip
- 0-30° vertical depth rotation
- 14° 2D planar rotations
- 25% scaling
- 30% planar translations



- Depth & lighting variations for one object

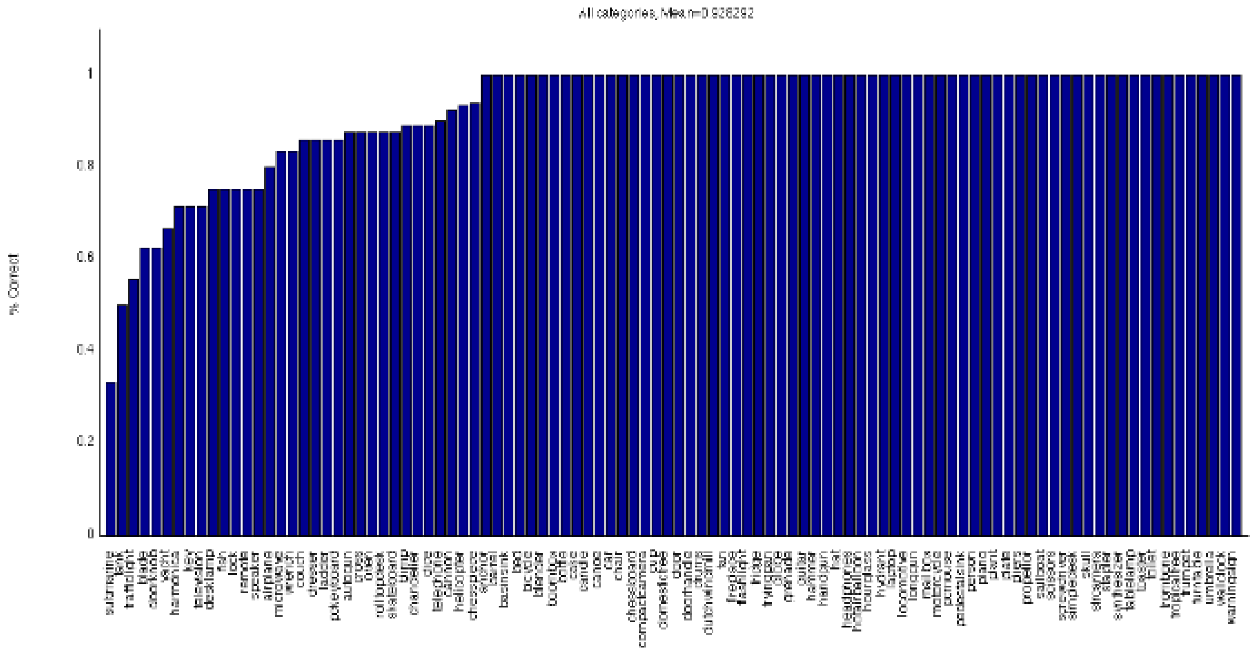
“Emer” the robot recognizing objects..



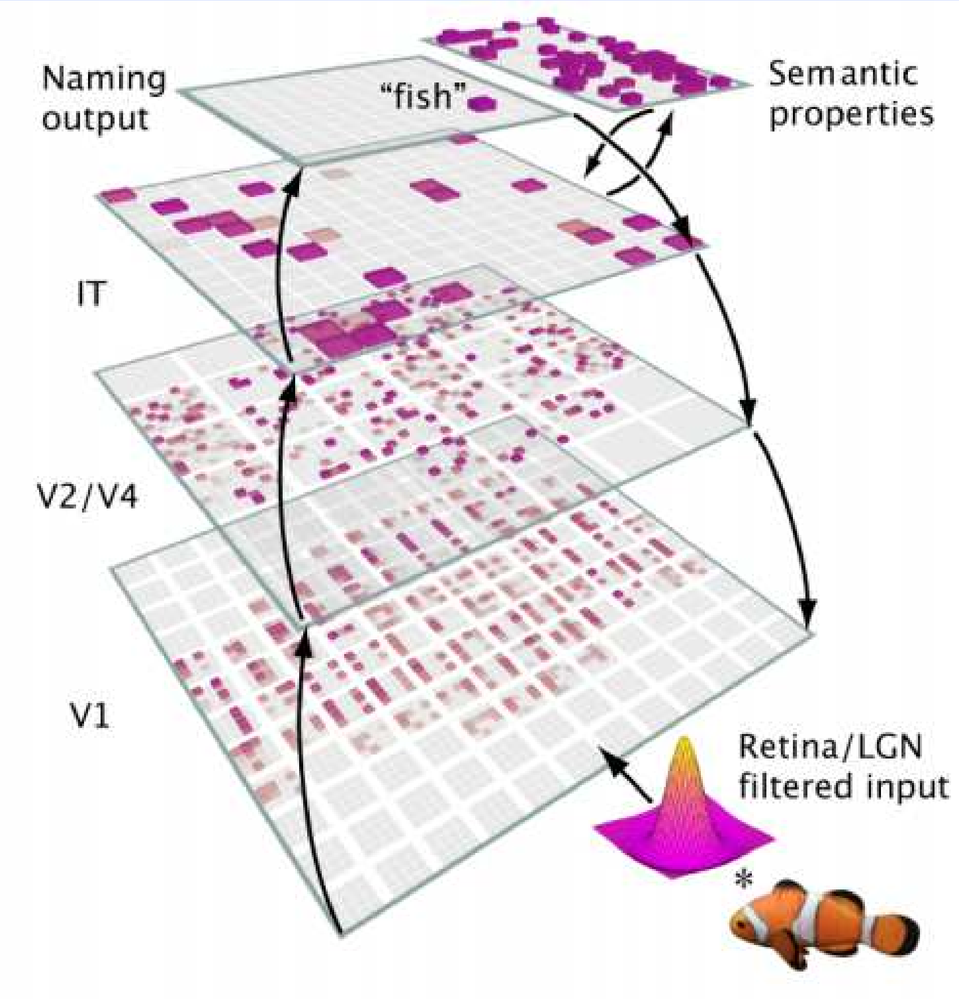


Video Demo: emer_demo_1.mov

Generalization Results: 92.8%

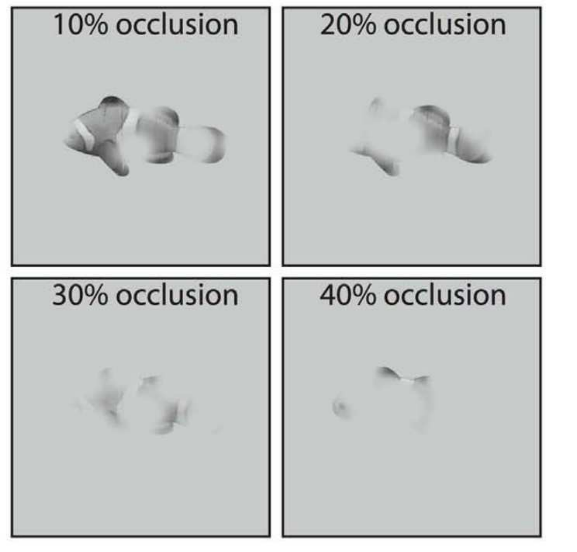


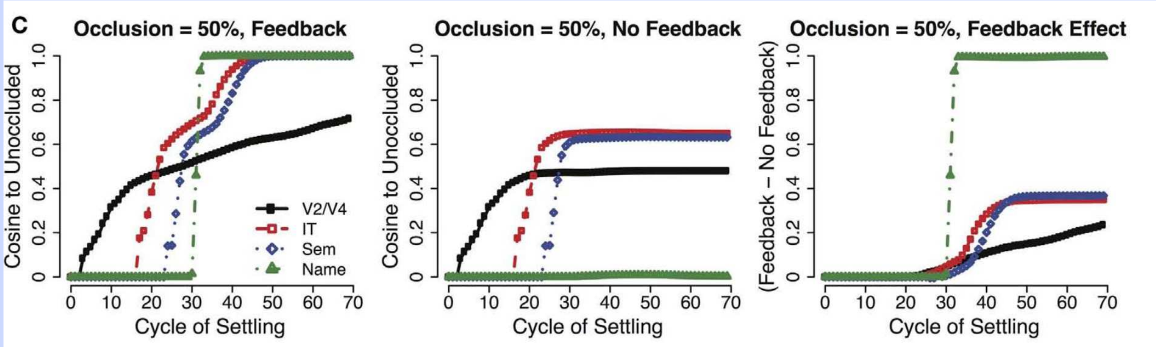
Bigger network model, bidirectional dynamics



Wyatte et al., 2012; O'Reilly et al., 2013

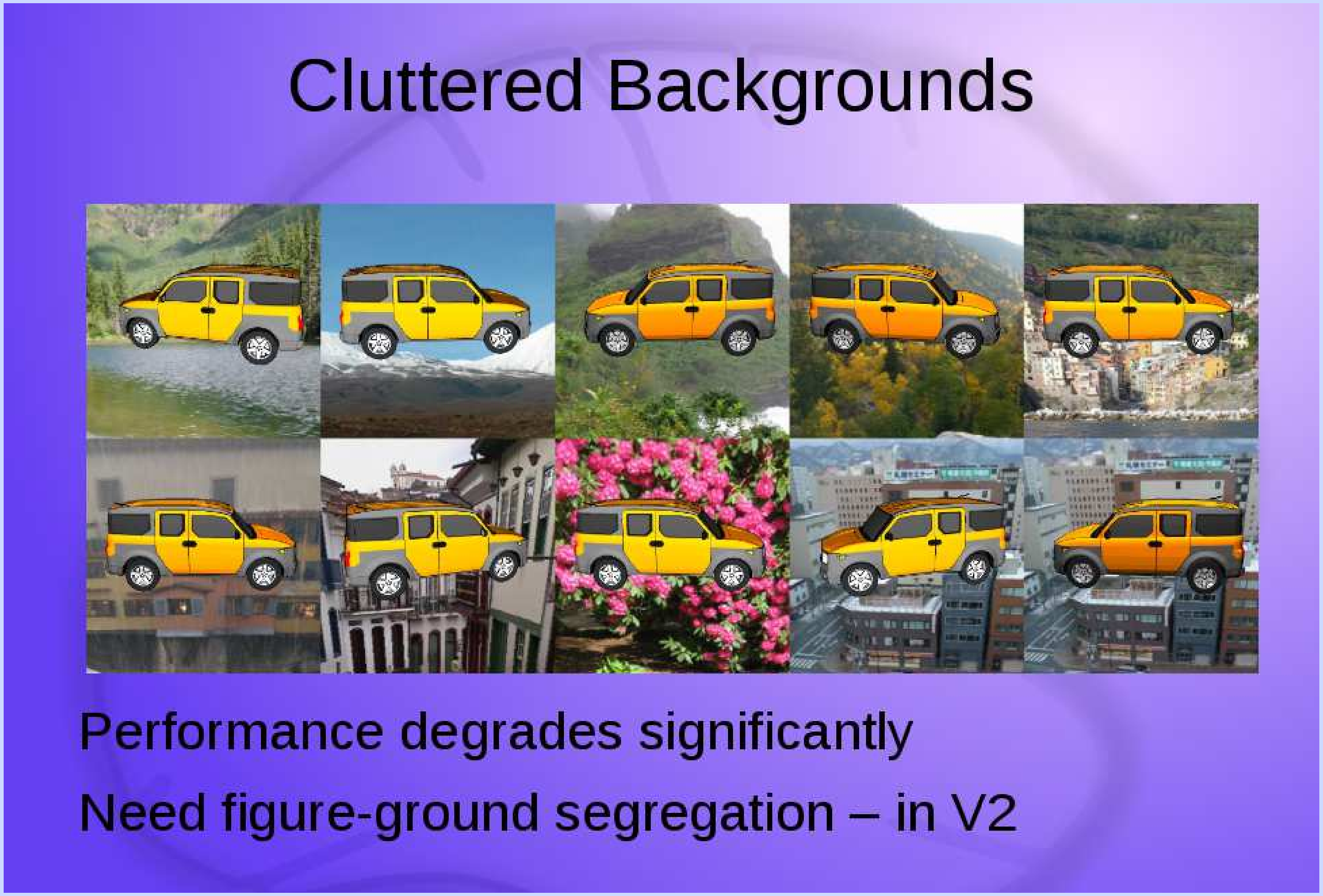
Bidirectional Dynamics





O'Reilly et al., 2013; Wyatte et al., 2012

A Challenge



Cluttered Backgrounds

Performance degrades significantly

Need figure-ground segregation – in V2

Deep Predictive Learning: What-Where-Integration

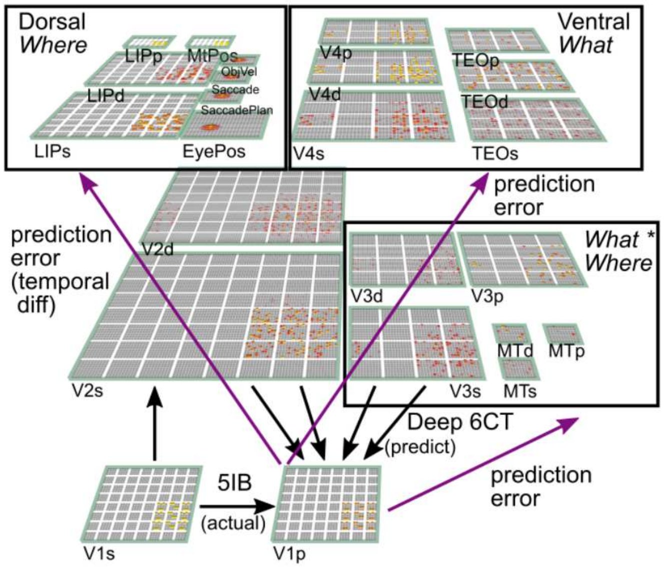


Figure 2: The three-visual-stream deep predictive learning model (What-Where-Integration or WWI model). The dorsal *Where* pathway learns first, using abstracted *spatial blob* representations, to predict where an object will move next, based on prior motion history, visual motion, and saccade efferent copy signals. It then provides strong top-down inputs to lower areas to drive accurate spatial predictions, leaving the residual error to be more about *What* and *What* * *Where* integration information. The V3 and MT areas constitute the *What* * *Where* integration pathway, sitting on top of V2 and learning to integrate visual features plus spatial information to accurately drive fully detailed predictions over the V1 pulvinar (V1p) "projection screen" layer (i.e., the cells distributed throughout the pulvinar that receive strong 5IB driver inputs). V4 and TEO are the *What* pathway, and learn abstracted object feature representations, which uniquely generalize to novel objects, and, after some initial learning, drive strong top-down inputs to lower areas. Most of the learning throughout the network is driven by a common predictive error signal encoded via a temporal difference over the pulvinar (V1p and other *p* layers), reflecting the difference between prediction (minus phase) and actual outcome (plus phase). *s* suffix = superficial layer, *d* = deep layer.

- uses predictive learning via recurrent feedback but no supervised target labels!

O'Reilly et al, 2017 *arXiv*

Still missing...

Motion: motion_noise.mp4

Still missing...

Motion: motion_noise.mp4

- Neurons in *area MT* very sensitive to motion
- Lots of work on how downstream areas integrate motion signals across time to detect *coherence* (e.g. Shadlen, Newsome, etc)
- Thomas Serre has shown that motion signals very reliable for discriminating between particular *actions* (eg throwing a baseball)

Still missing...

Motion: motion_noise.mp4

- Neurons in *area MT* very sensitive to motion

- Lots of work on how downstream areas integrate motion signals across time to detect *coherence* (e.g. Shadlen, Newsome, etc)

- Thomas Serre has shown that motion signals very reliable for discriminating between particular *actions* (eg throwing a baseball)

- Should be able to solve problem via bidirectional influence of motion integration signals, object recognition, and spatial attention (next)....

Perception and Attention

1. Why does primary visual cortex encode oriented bars of light? *Correlational learning based on natural visual scenes.*

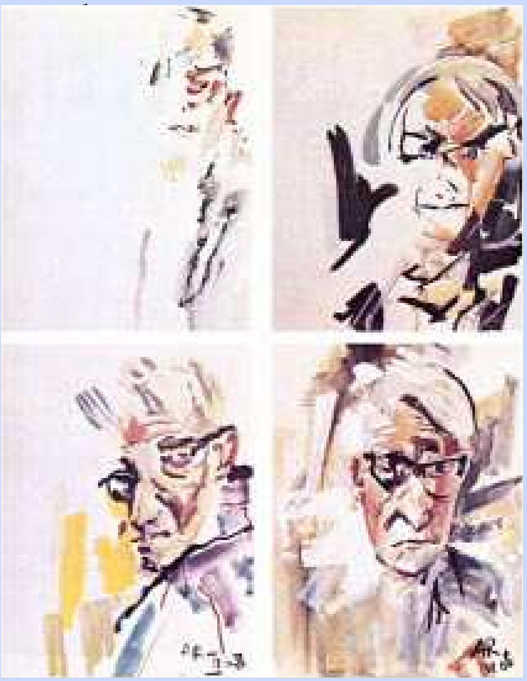
2. How do we recognize objects (across locations, sizes, rotations with wildly different retinal images)? *Transformations: increasingly complex featural encodings, increasing levels of spatial invariance; Distributed representations.*

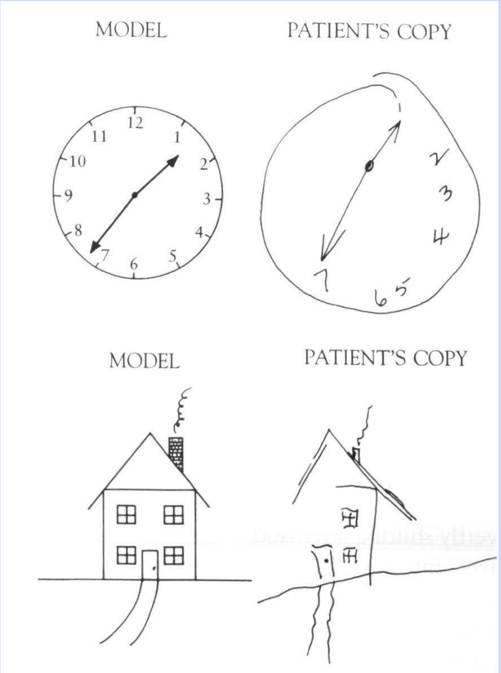
3. Why is visual system split into what/where pathways?

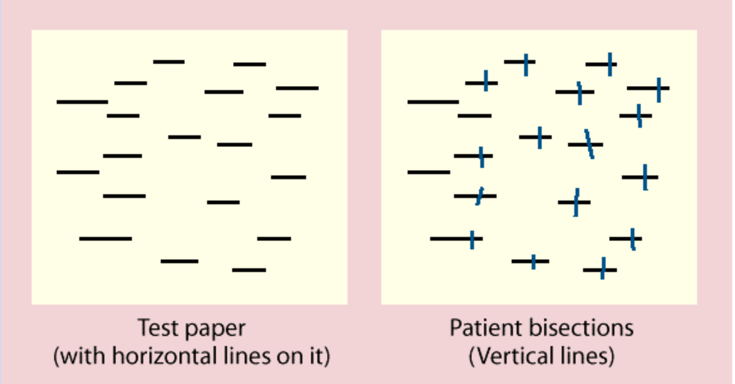
4. Why does parietal damage cause attention problems (neglect)?

Spatial Attention: Unilateral Neglect

Patient copying a scene:





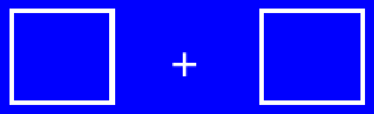


Self portrait, copying, line bisection tasks:
In all cases, patients with parietal/temporal lesions seem to forget about 1/2 of space! *but they still see it!*

Posner Spatial Cuing Task

Valid cue

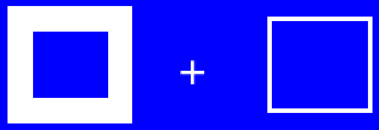
- Fixation



Posner Spatial Cuing Task

Valid cue

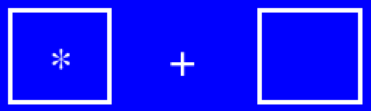
- Cue appears



Posner Spatial Cuing Task

Valid cue

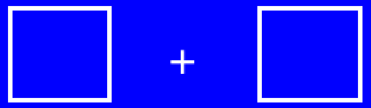
- Target appears, respond with target location



Posner Spatial Cuing Task

Invalid cue

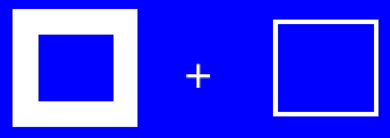
- Fixation



Posner Spatial Cuing Task

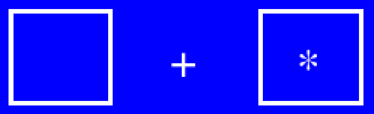
Invalid cue

- Cue appears



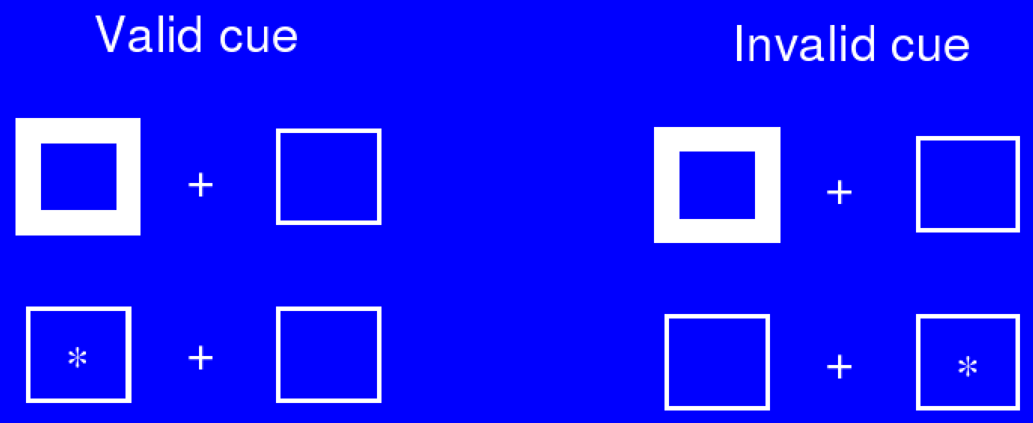
Posner Spatial Cuing Task

Invalid cue



- Target appears, respond with target location

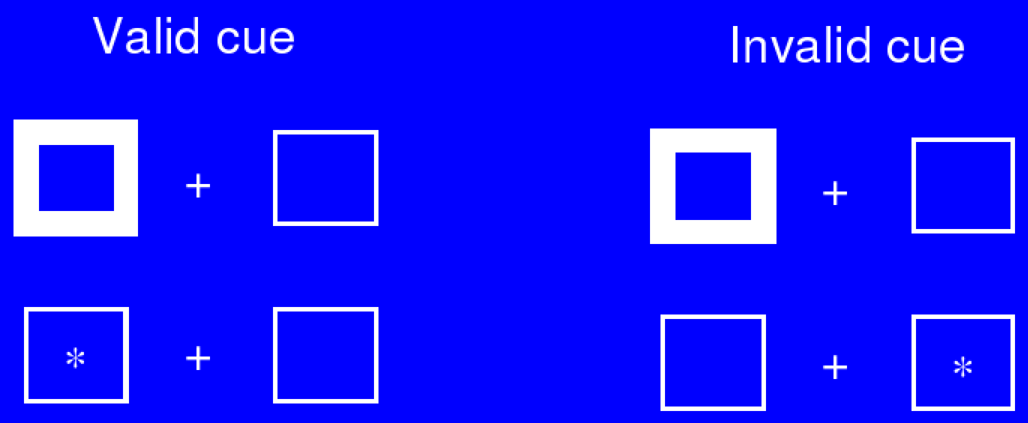
Posner Spatial Cuing Task



Valid cue

Invalid cue

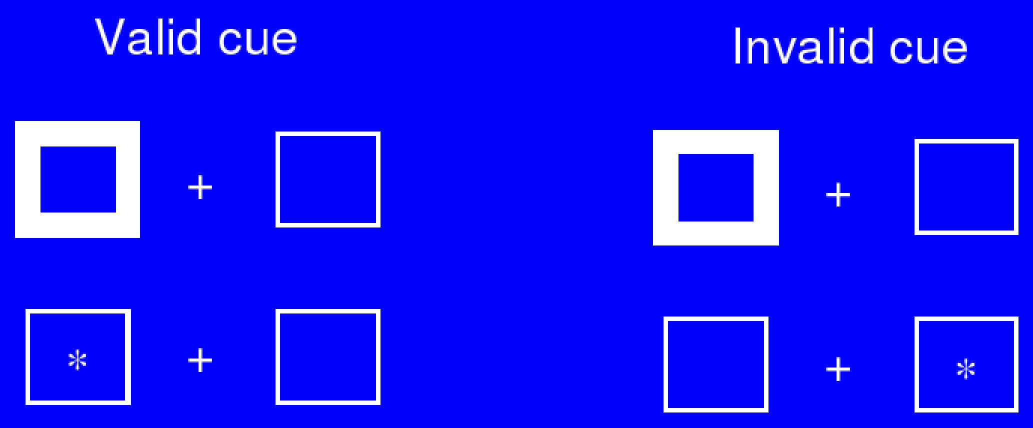
Posner Spatial Cuing Task


Valid cue

Invalid cue

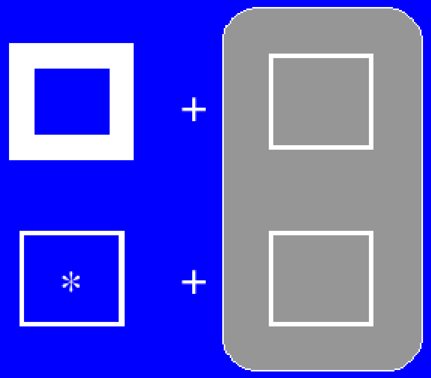
- Valid cues speed up performance (relative to “no cue” condition)
- Invalid cues slow down performance (relative to “no cue” condition)

Effects of Parietal Lesions on Posner Task

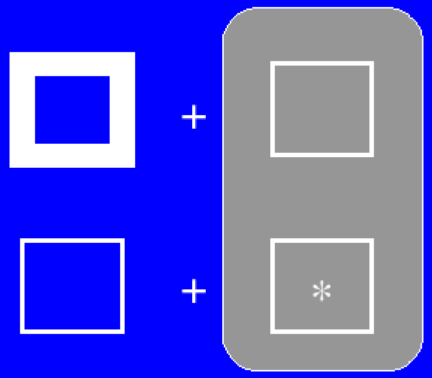


- Large, unilateral parietal lesions result in **neglect** of the opposite (contralateral) side of space
- Subjects do not respond to targets in the neglected hemifield
- What about smaller, unilateral parietal lesions?

Effects of Parietal Lesions on Posner Task



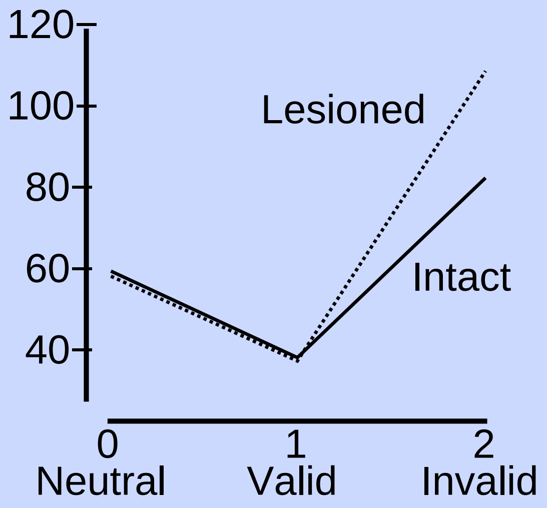
Valid cue



Invalid cue

- Say that you have a small, left parietal lesion, so the right side is affected
- Run the Posner task with cues in the ipsilateral (left) side of space

Effects of Parietal Lesions on Posner Task

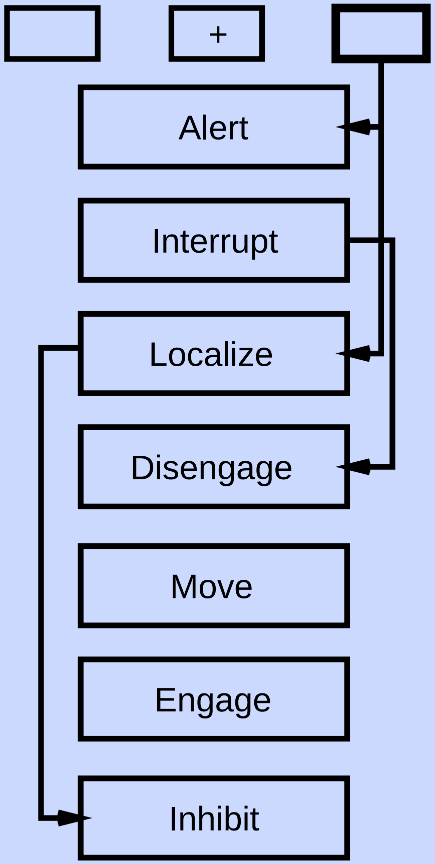


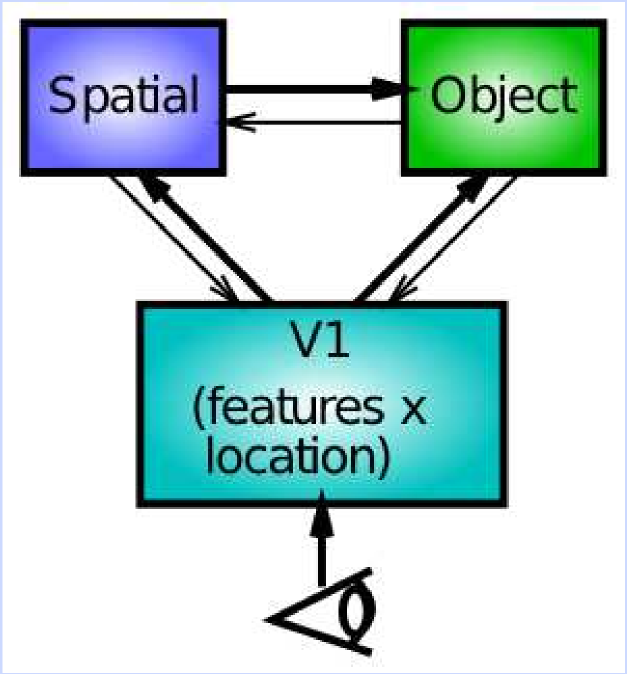
- Patients perform normally in the “neutral” (no cue) condition, *regardless* of where the target is presented

- Patients benefit just as much as controls from valid cues

- Patients are hurt more than controls by invalid cues

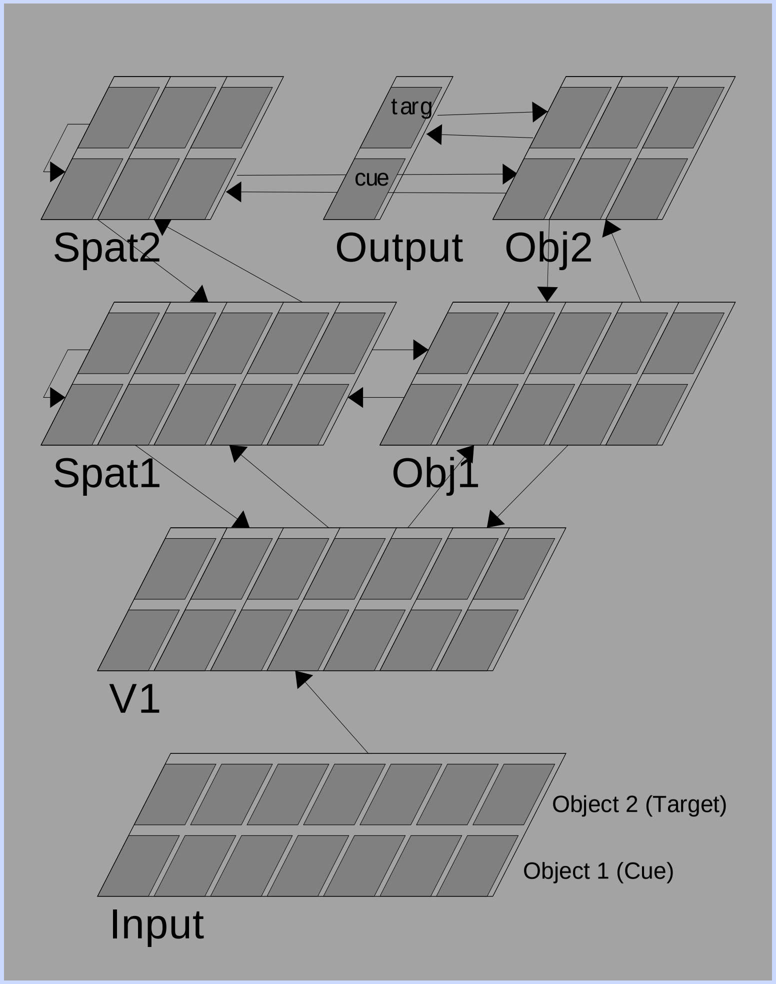
Possible Models





Attention emerges from bidirectional constraint satisfaction & inhibitory competition.

Simple Model



[attn_simple.proj]

Posner Task Data

	Valid	Invalid	Diff
Adult Normal	350	390	40
Elderly Normal	540	600	60
Patients	640	760	120
Elderly normalized (*.65)	350	390	40
Patients normalized (*.55)	350	418	68

Posner Task Sims

- The model explains the basic finding that valid cues speed target processing, while invalid cues hurt

- Also explains finding that patients with small unilateral parietal lesions benefit normally from valid cues in ipsilateral field but are disproportionately hurt by invalid cues.

- No need to posit “disengage” module!

Posner Task Sims

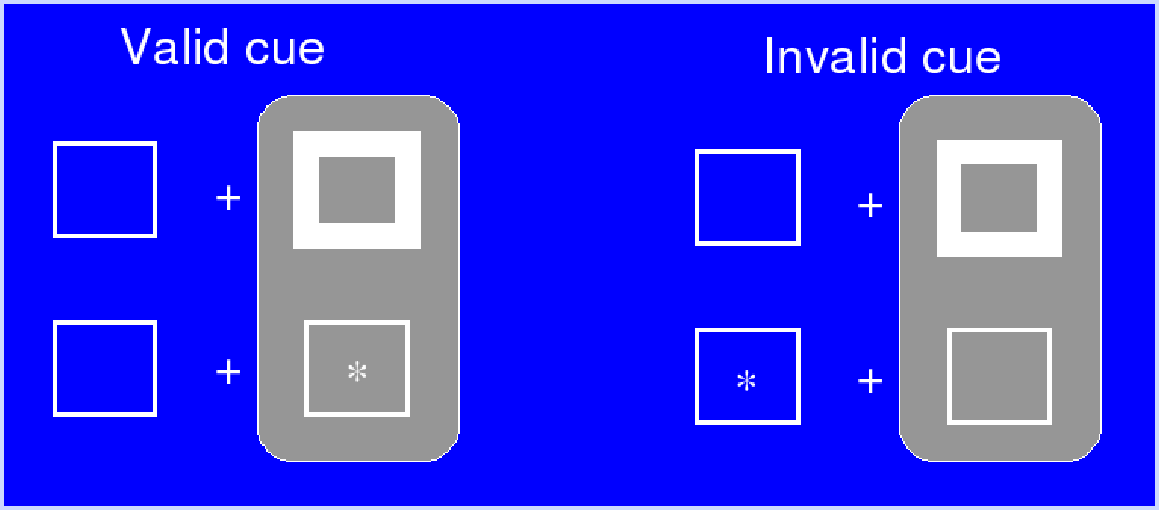
- The model explains the basic finding that valid cues speed target processing, while invalid cues hurt

- Also explains finding that patients with small unilateral parietal lesions benefit normally from valid cues in ipsilateral field but are disproportionately hurt by invalid cues.

- No need to posit "disengage" module!

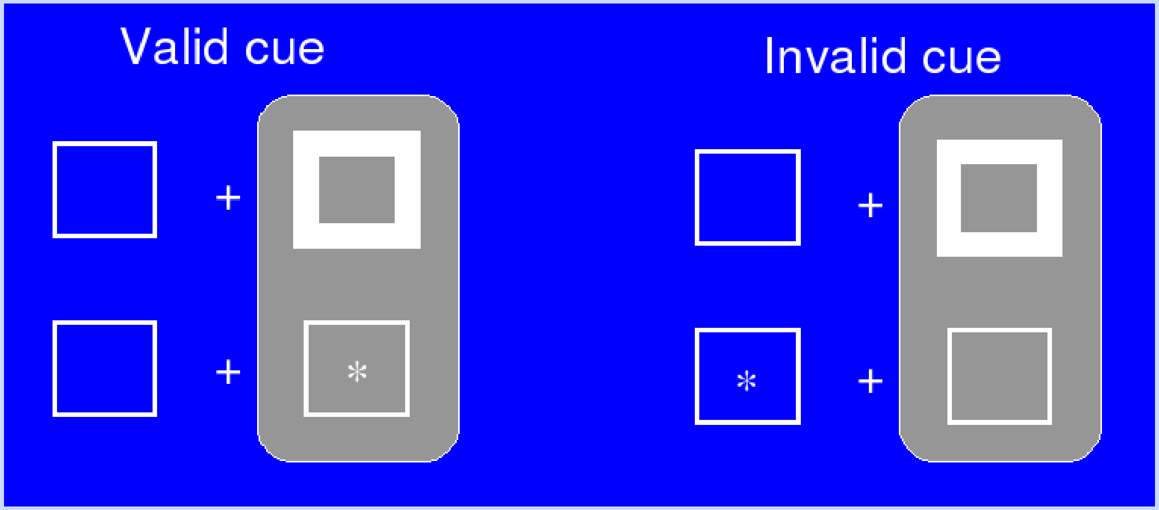
- Also explains finding of **neglect** of contralateral visual field after large, unilateral parietal lesions when some stimulus is present in ipsilateral field ("extinction")

More Posner Lesion Fun



- Returning to patient with left parietal lesion...
- What happens if cues are presented in **contralateral** (affected) hemifield? (“Reverse Posner”)

More Posner Lesion Fun



Returning to patient with left parietal lesion...

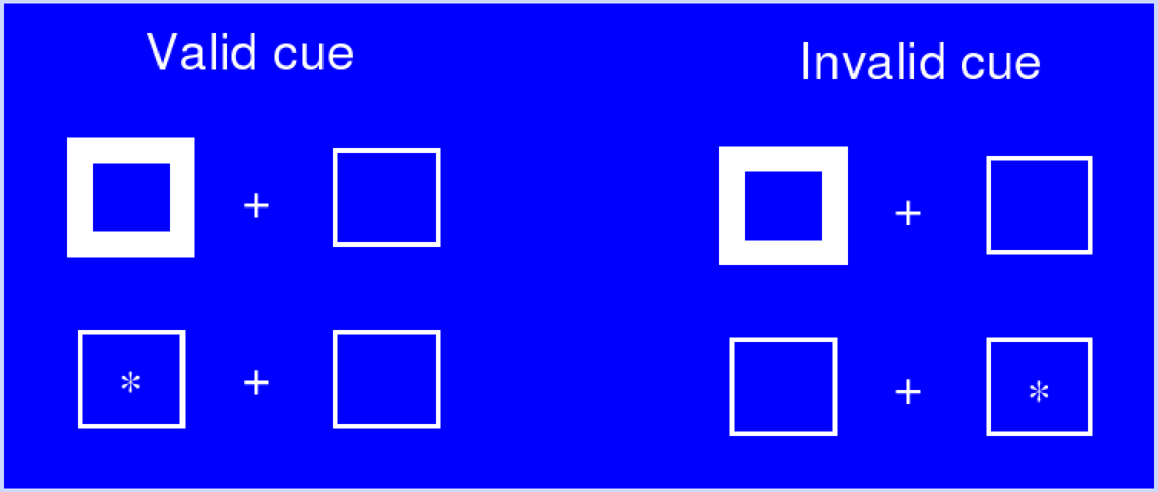
- What happens if cues are presented in **contralateral** (affected) hemifield?

  Predictions:

- Smaller benefit for valid cues
- Patients should be hurt less than controls by invalid cues.

[attn_simple.proj]

Inhibition of Return



- Typically, target detection is faster on trials with valid vs invalid cues

- **However**, if the cue is presented for a longer time (eg. 500 ms), performance is faster on ***invalid*** vs valid trials

- Can explain in terms of **accommodation** (neural fatigue)

[attn_simple.proj]

Simple model: too simple?

- Has unique one-to-one mappings between low-level visual features and object representations (not realistic)
- Does not address issue of spatial attention when trying to perceive multiple objects simultaneously

Simple model: too simple?

- Has unique one-to-one mappings between low-level visual features and object representations (not realistic)

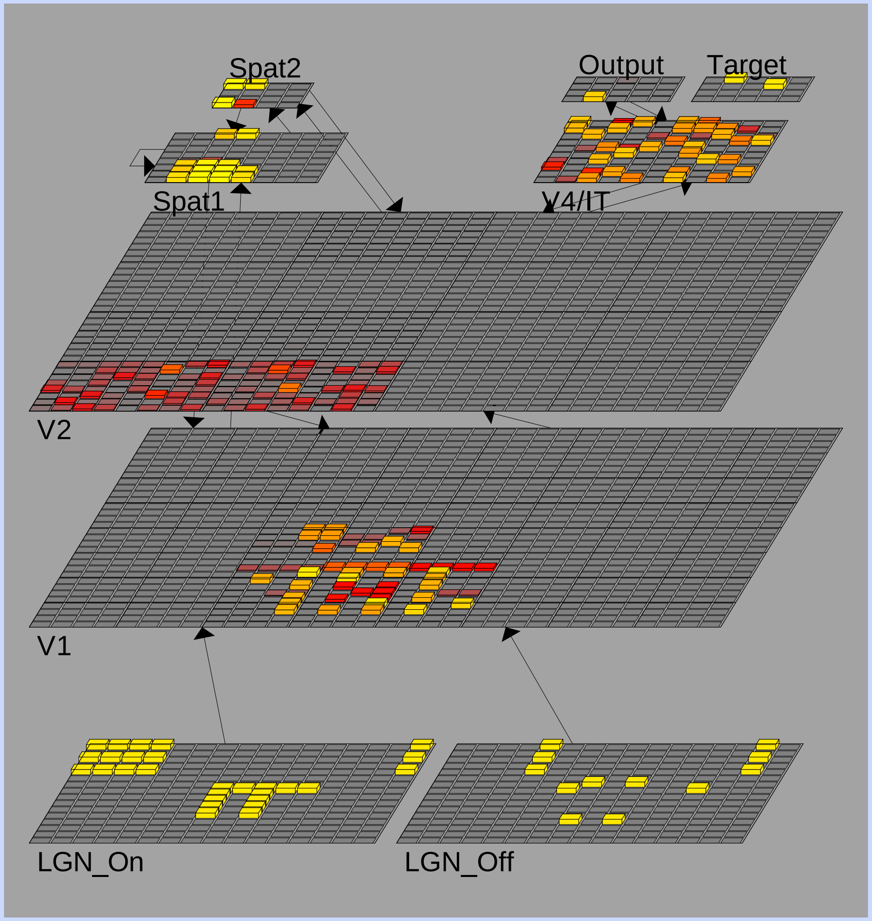
- Does not address issue of spatial attention when trying to perceive multiple objects simultaneously

- “Complex” model combines more realistic model of object recognition (starting from LGN) with simple attention model

  → Can use spatial attention to restrict object processing pathway to one object at a time, enabling it to sequentially process multiple objects.

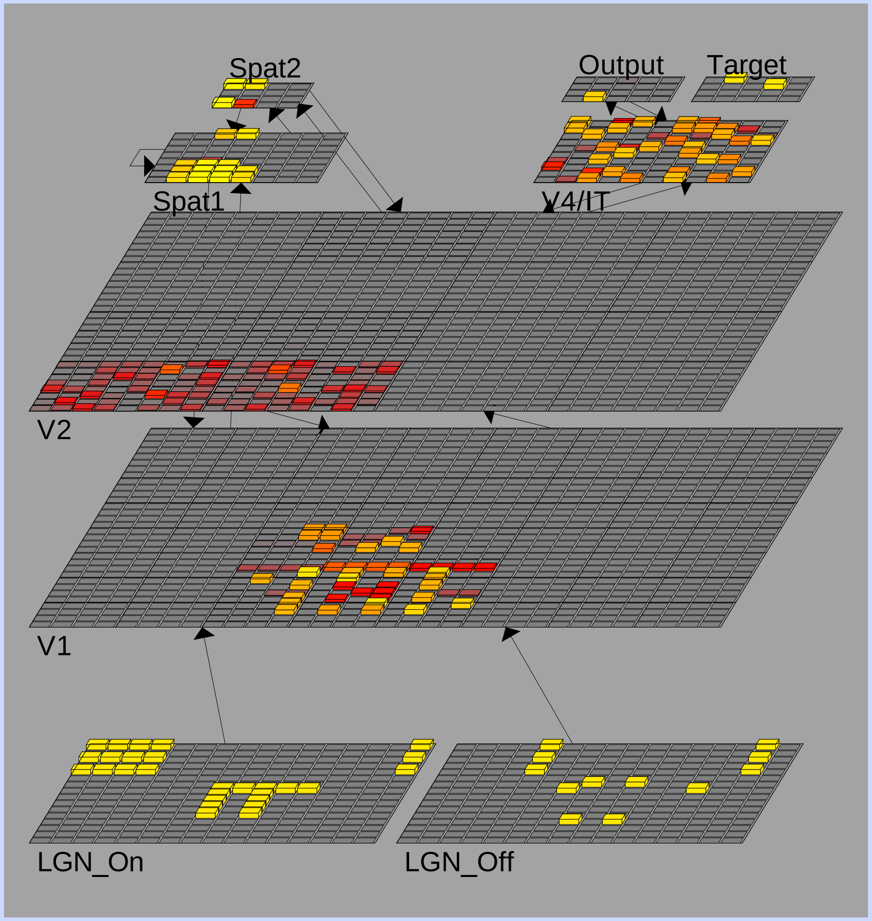
- Lesions of entire spatial pathway cause *simultanagnosia*: inability to concurrently recognize two objects

Complex Model



Spat1 has recurrent projns to encourage focus on one region of space

Complex Model



Spat1 has recurrent projns to encourage focus on one region of space

But only mechanism for switching is accommodation...

Perception and Attention

1. Why does primary visual cortex encode oriented bars of light?
   Correlational learning based on natural visual scenes.

2. How do we recognize objects (across locations, sizes, rotations with wildly different retinal images)? *Transformations: increasingly complex featural encodings, increasing levels of spatial invariance; Distributed representations.*

3. Why is visual system split into what/where pathways?
   Transformations: emphasizing and collapsing across different types of relevant distinctions; attention

4. Why does parietal damage cause attention problems (neglect)?
   Attention as an emergent property of competition

General Issues in Attention

Attention:

- Prioritizes processing.
- Coordinates processing across different areas.
- Solves binding problems via coordination.

General Issues in Attention

Attention:

- Prioritizes processing.
- Coordinates processing across different areas.
- Solves binding problems via coordination.

But attention should be much more flexible than just spatial bias!

Later: how to incorporate goals, reinforcement probability, into attentional allocation