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
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
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)?
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, wildly different retinal images)?

Perception & Attention
Overview of the Visual System

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

Hierarchies of specialized visual pathways, starting in retina, to LGN.
Two Streams: Ventral "what" vs. Dorsal "where"
The Retina

Retina is not a passive "camera" but a sensor for contrast enhancement that emphasizes changes over space and time.

Key principle: contrast enhancement that emphasizes changes over space and time.
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.
- Performs attentional processing.
- On- and off-center information from retina is preserved in LGN.
For coherent vision, need to detect varying degrees of all these:

- Edges differ in orientation, size (spatial frequency), and position.

Primary Visual Cortex (V1): Edge Detectors

V1 combines LGN (thalamus) inputs into oriented edge detectors:
Hubel & Wiesel Nobel Prize

Example V1 edge detector
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

Pinwheels can arise from learning and lateral connectivity: not hard-wired!
• 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

Rerouting of Visual Orientation Modules in A1
Visual Behavior After Rerouting Right Visual Field
Visual Acuity After Rerouting
So learning is powerful, but so is evolution!

Visual Acuity After Rerouting
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)
The Receptive Fields

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

The Receptive Fields
Live on mountain tops?
How many babies...
What about training on mother's faces?
Model Training on Faces
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.

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

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

Why does primary visual cortex encode oriented bars of light?
Perception and Attention

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

Correlational Learning based on natural visual scenes.

I. Why does primary visual cortex encode oriented bars of light?
The Object Recognition Problem

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

This is hard because different patterns in different locations/sizes/rotations can overlap a lot, while the same patterns in the same location cannot overlap at all!
Large amount of shape variability within and between categories.

Huge amount of view-based variability (position, orientation, size, rotation).

Object Recognition is Hard.
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")
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)
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.

Gradual Invariance Transformations (Fukushima, '80)
Convolutional Neural Networks

• very popular for "deep learning" in machine learning, Yann LeCun, Hinton, etc.
The Model: combining Fukushima with convolutional neural nets, bidirectional connectivity and learning!
Each object is presented at multiple locations, sizes, regardless of location and size. The network's job is to activate the appropriate output unit (0-19) for each object.
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.

How do we plot receptive fields for V4?
Some V4 units code for location-specific conjunctions of V1 features – this will show up as a sharp receptive field for image input.
Some V4 units code for simple features in a location invariant way.

- This will show up as smeary parallel lines in Image input.
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).
e.g., for this set of objects, corners are good.

Invariant across size transformations

Another approach to this problem is to pick features that are

Size Invariance
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
- Test on new sizes/locations
- A limited number of sizes/locations
- Then train on two new objects (using multiple sizes/locations)
- Train on these generalizations
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
(some stuff)

Input

\[ V^2 \]

In \( V^4 \) detectors
Invariant feat.
Size/location

Output

Generalization

\[ \text{19} \]
Yeah, but these objects are regularly shaped, straight lines... what about real objects?
3D Object Recognition Test
Object for one variations lighting.

Depth &
"Emer" the robot recognizing objects.

Video Demo: emer demo 1.mov
Bigger network model
Wyatte et al., 2012; O'Reilly et al., 2013
Bidirectional Dynamics

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

Need figure-ground segregation – in V2
Performance degrades significantly

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

Still missing...
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, image changes)?
Transformations: Increasingly complex featural encodings, increasing levels of spatial invariance.

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

4. Why does parietal damage cause attention problems (neglect)?
In all cases, patients with parietal/temporal lesions seem to forget about 1/2 of space, but they still see it! Self-portrait, copying, line bisection tasks:

Spatial Attention: Unilateral Neglect

Patient copying a scene:
Fixation

Valid cue

Posner Spatial Cuing Task
Posner Spatial Cueing Task

Valid cue

Cue appears

□ + □
Rosner Spatial Cuiring Task

Target location
Valid cue

Target appears,
respond with

+
Posner Spatial Cuing Task

Invalid cue + Fixation
Posner Spatial Cuing Task

• Cue appears

Invalid cue

+
Ponsner Spatial Cuing Task

Invalid cue

Target appears

Target location

respond with
Posner Spatial Cuing Task

- Invalid cue
- Valid cue
Test for "no cue" condition
Invalid cues slow down performance (relative to
"no cue" condition)
Valid cues speed up performance (relative to
Poster Spatial Cuing Task
What about smaller, unilateral parietal lesions?

- Hemisected

Subjects do not respond to targets in the neglected hemisected side of space.

- Large, unilateral parietal lesions result in neglect of the opposite (contralateral) side of space.

Effects of Parietal Lesions on Posner Task

Invalid cue

Valid cue
Effects of Parietal Lesions on Posner Task

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

Invalid cue

Valid cue
Patients perform normally in the "neutral" (no cue) condition, regardless of where the target is presented.

- Patients are hurt more than controls by invalid cues.
- Patients benefit just as much as controls from valid cues.

Effects of Parietal Lesions on Posner Task
Attention emerges from bidirectional constraint satisfaction & inhibitory competition.

Possible Models
Simple Model

Input

Object 1 (Cue)
Object 2 (Target)

Spat1
Spat2

Output

Object 1 (Cue)
Object 2 (Target)
<table>
<thead>
<tr>
<th>Patient Normal</th>
<th>Elderly Normalized (×.65)</th>
<th>Patient Normal</th>
<th>Elderly Normalized (×.65)</th>
<th>Adult Normal</th>
<th>Valid</th>
<th>Invalid</th>
</tr>
</thead>
<tbody>
<tr>
<td>418 350</td>
<td>40 350</td>
<td>760 640</td>
<td>60 540</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 030</td>
<td>350</td>
<td>60 350</td>
<td>0</td>
<td>40 390</td>
<td></td>
<td></td>
</tr>
<tr>
<td>030 418</td>
<td></td>
<td>350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68 40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Posner Task Data
The model explains the basic finding that valid cues speed target processing, while invalid cues hurt.

- No need to post "disengage" module.
- 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.
- Processed cues while invalid cues hurt.
- The model explains the finding that valid cues speed target processing, while invalid cues hurt.

Posner, Task, Sims
The model explains the basic finding that valid cues speed target processing, while invalid cues hurt processing. 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. 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.
Returning to patient with left parietal lesion...

What happens if cues are presented in contra lateral (affected)

(“Reverse Posner”)

More Posner lesion Fun
Returning to patient with left parietal lesion...

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

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

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

Returning to patient with left parietal lesion...
Inhibition of Return

• Typically, target detection is faster on invalid vs valid trials.

  However, if the cue is presented for a longer time (e.g., 500 ms),

Can explain in terms of accommodation (neural fatigue)

Performance is faster on invalid vs valid trials.

Typically, target detection is faster on trials with valid vs invalid cues.
Simple model: too simple?

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

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

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

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

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

Lesions of entire spatial pathway cause simultanagnosia: inability to concurrently recognize two objects
Spat1 has recurrent projections to encourage focus on one region of space
Spat1 has recurrent projns to encourage focus on one region of space.

But only mechanism for switching is accommodation...
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: emphasizing and collapsing across different types of distributed representations.

3. Why is visual system split into what/where pathways?
   Transformations: emphasizing levels of spatial invariance.
   Transformation: increasing wildly different retinal images.
   Distributed representations.

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

Perception and Attention
General Issues in Attention

Attention:

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

General Issues in Attention
General Issues in Attention

Attention:
• Prioritizes processing.
• Coordinates processing across different areas.
• Solves binding problems via coordination.
• Coordinates processing across different areas.

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

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

General Issues in Attention