Perception & Attention

Perception is effortless but its underlying mechanisms are incredibly sophisticated. 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?
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 with wildly different retinal images)?
Hierarchies of specialized visual pathways, starting in retina, to LGN (thalamus), to V1 & up.

Overview of the Visual System
Two Streams: Ventral "what" vs. Dorsal "where"
The Retina

Retinal output ganglion cells

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

Retina is not a passive “camera”
LGN of the Thalamus

A “relay station”, but so much more.

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

Re-routing of Visual Orientation Modules in A1
Visual Behavior After Rerouting Right Visual Field

Spatial resolution (cycle per degree) at 0.15 contrast

- N3
- R6
- R7

Ferret

Left monocular field
Right monocular field

Visual Acuity After Rerouting
So learning is powerful, but so is evolution!
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 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
- FFB inhib competition for specialization (see Ch 4)
Red = on-center, Blue = off-center.

The Receptive Fields
Live on mountain tops?
How many babies
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.

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

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

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

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

Perception and Attention
The Object Recognition Problem

Problem: Recognize objects 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 can not overlap at all!
Huge amount of view-based variability

between categories

Large amount of shape variability within and

Object Recognition is Hard
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)
Gradual Invariance Transformations (Fukushima, '80)

It 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 objects 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.

Convolutional Neural Networks
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
V4 omitted here (edge detectors, hard-coded)

(important for figure-ground etc.)

Each IT unit pays attention to all of V4
Each object is presented at multiple locations, sizes.

The Objects

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

V4 Receptive Fields
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.
- Generally a given V4 unit is associated with multiple objects.

- Can also look at which Output units tend to get active for any given V4 unit.
Size Invariance

One approach to this problem is to have $V_4$ units respond to all of the $V_2$ units that represent a feature (regardless of size).
Size Invariance

- Invariant across size transformations
- Another approach to this problem is to pick features that are

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 16 objects
  – Train with 2 new objects in only some locations/sites
  – Test the net with nonstudied „views“ (sizes/locations) of new 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?
Test on new sizes/localizations: a limited number of sizes/localizations

Then train on two new objects (using multiple sizes/localizations)

Train on these localization
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.
• 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
Yeah, but these objects are regularly shaped, straight lines...

What about real objects?
3D Object Recognition Test

- Translations
- 30% planar
- Scalings
- 25% scaling
- Rotations
- 14° 2D planar
- Rotation
- 0-30° vertical depth
- Rotation + 180° flip
- +/- 20° horizontal depth
- Testing
- 2 objects left out for
- 8 objects per category
- 100 categories
- Warehouse
- Sketchup
- From Google
Depth & lighting variations for one object
Bidirectional Dynamics
A Challenge

Need figure-ground segregation – in V2

Performance degrades significantly

Cluttered Backgrounds
State of the Art
Still missing...
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 (e.g. throwing a baseball)
Neurons in area MT very sensitive to motion

Motion: motion-sensitive

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, wildly different retinal images)? Transformational invariance: 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)?
In all cases, patients with parietal/temoral lesions seem to forget about 1/2 of space, but they still see it!

Self-portrait, copying, line bisection tasks:

Patient copying a scene:
Posner Spatial Cuing Task

• Fixation

Valid cue

+
Cue appears
Valid cue
Posner Spatial Cueing Task
Posner Spatial Cuing Task

Valid cue

- Target appears, respond with target location

* +
• Fixation

Posner Spatial Cuing Task

Invalid cue
Cue appears

Invalid cue

Posner Spatial Cuing Task
Poster Spatial Cueing Task

Target location
respond with
• Target appears,

Invalid cue
Posner Spatial Cuing Task

Invalid cue

Valid cue
to "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?

Hemispheres subjects do not respond to targets in the neglected of the opposite (contralateral) side of space. Large, unilateral parietal lesions result in neglect of effect of parietal lesions on Posner Task.
Effect of Parietal Lesions on Posner Task

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

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 & inhibition competition.

Possible Models:

- Inhibit
- Engage
- Move
- Disengage
- Localize
- Interrupt
- Alert
Simple Model

Input

Object 1 (Cue)
Object 2 (Target)

V1

Spat1

Output

Object 1
Object 2

Spat2

L1a
cue

L1a
L1b

<table>
<thead>
<tr>
<th></th>
<th>Elderly Normal</th>
<th>Elderly Normal (normalized *0.65)</th>
<th>Adult Normal</th>
<th>Adult Normal (normalized *0.55)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>350</td>
<td>350</td>
<td>540</td>
<td>350</td>
</tr>
<tr>
<td>Patients normalized</td>
<td>390</td>
<td>418</td>
<td>640</td>
<td>390</td>
</tr>
<tr>
<td>Valid</td>
<td>40</td>
<td>68</td>
<td>120</td>
<td>40</td>
</tr>
<tr>
<td>Invalid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diff</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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!
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 visual field, but are disproportionately hurt by invalid cues.

No need to posit "disengage" module!

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

The model explains the basic finding that valid cues speed target processing, while invalid cues hurt.
More Posner Lesion Fun

Returning to patient with left parietal lesion...

What happens if cues are presented in contralateral (affected) hemifield? ("Reverse Posner")
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.

More Posner Lesion Fun
Typically, target detection is faster on trials with invalid cues vs valid cues.

Performance 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).
[attn simple proj]
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

and object 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

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

Conversely, simultaneously recognizing multiple objects simultaneously.

Does not address issue of spatial attention when trying to perceive and object representations (not realistic).

Has unique one-to-one mappings between low-level visual features.

Simple model: too simple?
Spat1 has recurrent projns to encourage focus on one region of space. Complex Model...
Spat1 has recurrent projections 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*
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.
• Prioritizes processing.

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

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