Human Visual Attention

Lecture #2
CS510, Spring 2018
Review: Human Vision System

D. Milner & M. Goodale,
*The Visual Brain in Action*, p. 22
What is attention?

• “The selective aspect of processing” – Kosslyn

• “processes that enable an observer to recruit resources for processing selected aspects of the retinal image more fully than nonselected aspects” – Palmer
Overt vs. Covert Attention

• Overt attention: observable movements of eyes, head & body to orient eyes
  – Foveas: 90% of receptors, $\pm 2^\circ$
  – Allocation to 3D point in space
    • Vergence & focus
  – Average dwell time: ~300ms
• Saccadic movement
  • Very fast: ~30ms, up to 900$^\circ$/sec
  • Suppression: no input during saccade
  – World appears as sequence of displaced, small, high resolution, stereo images with low resolution peripheries
Overt vs. Covert Attention (II)

- You don’t process all the data in your foveal image
- Covert attention: selection of retinal data to process (“inner eye”)
  - Cannot be observed directly
  - Its existence is not in dispute
  - Its form is a matter of intense debate
  - Assumption: insufficient resources necessitate covert attention.
- Covert attention is the subject of this talk
3 Models of Covert Attention

1. Feature Integration Theory\(^5\)
   - “Pre-attentive” low-level features computed in parallel across the image
     - E.g. color, edge orientations, motion
   - In visual search, attention can jump to locations based on pre-attentive features (“pop-out”)
   - Conjunctions of features or complex features require sequential search
   - Implicitly assumes attention is like a spotlight
Feature Integration Theory (II)

Find the red target

Find the round target

Find the red ‘X’ target
2. Integrated Competition Hypothesis

- “Pop-out” effect depends on:
  - Homogeneity of distractors
  - Homogeneity of targets (seq. pres.)
- Primary role of attention is *segmentation* (or grouping)
- Low-level features important as the basis of segmentation
### Integrated Competition Hypothesis (II)

Find the upright T

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Find the upright T

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Find the upright T

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

*Colorado State University*
Theory #3

Task: Which Line is Longer?
3 Models (III)

• Inattentinal Blindness Theory\(^7\)
  – When concentrating on the task, most subjects will not see additional objects
    • Depends on semantics of additional object
  – Additional objects are interpreted
    • Cause priming effects
  – Hypothesis: all objects in visual field are interpreted
    • Attention is a late effect, caused by attentional bottleneck
Gorilla

• Category Example
• Memorable
• Worth knowing
That's why autistic savants can make perspective drawings without being taught how. They're drawing what they see, which is all the little changes in size and texture that tell you one object is closer up and another object is farther away. Normal people can't see all those little changes without a lot of training and effort, because their brains process them unconsciously. So normal people are drawing what they "see," which is the finished object, after their brains have put it all together. Normal people don't draw a dog, they draw a concept of a dog. Autistic people draw the dog.

It's ironic that we always say autistic children are in their own little worlds, because if Dr. Snyder is right it's normal people who are living inside their heads. Autistic people are experiencing the actual world much more directly and accurately than normal people, with all their inattentional blindness and their change blindness and their every-other-kind-of-blindness.
Whose right?

• Most likely explanation: everyone!
• Vision is a multi-stage process
  – Every stage is coarsely parallel
  – Every stage is resource limited
  – Every stage attends based on data, task
• There is evidence for attention at the level of:
  – Windows (spatial)\(^8\)
  – Features\(^9, 11\)
  – Objects\(^{10, 20}\)
  – Awareness\(^7\)
Capacity Theory

• Vision is a resource-limited pipeline (or conveyer belt)

• Data and task demands determine where and how attention occurs
  – Psychological support for shifting from spatial to feature-based attention
  – ERP support for spatial attention preceding feature-based attention
Spatial vs. Feature Attention

This was a small sample. Statistically, subjects make more errors when the distracter resembles the non-target class.

Determine if target is ‘X’ or ‘O’; ignore peripheral distracters.
Statistically, subjects do better on this task than the previous task. Why? Because spatial attention eliminates the distracter before feature-based attention can be confused.

Task: Same as before, only now there are additional distracters in a cross around the target.
Relation to Human Object Recognition

- Attention occurs at every processing stage
- Human object recognition can be modeled as a four stage pipeline:
  1. Early vision (spatial attention)
  2. Feature extraction (feature attention)
  3. Categorization (object category att.)
  4. Expertise (object attention)

- What do we know about these types of visual attention?
Early Spatial Attention

• Appears as early as LGN\textsuperscript{8}
• Manifests as anticipatory rise in baseline activity\textsuperscript{8}
  – Both points suggest strong top-down component
• Very fast, even without cueing
  – 80-130 ms post stimulus\textsuperscript{15}
  – Suggests low-level features (e.g. edges)
  – Above random correlation to NVT\textsuperscript{21}
• Probably coarsely parallel
  – Up to 5 objects independently tracked\textsuperscript{16}
  – Evidence for at least 2 parallel fixations without motion\textsuperscript{17}
• Probably selects scales as well as positions\textsuperscript{18}
• Minimum scale for spatial attention\textsuperscript{19}
  – Larger than minimum scale for visual resolution
Feature-based Attention

- Slower than spatial attention
  - 140-180 ms post stimulus\textsuperscript{15}
  - Feedback to striate cortex 235 ms ps\textsuperscript{11}
- May itself be several mechanisms
  - Color, shape processed separately\textsuperscript{22}
- Feature conjunction tasks require attention; single features do not\textsuperscript{23}
Object-based Attention

- Object Category Attention
  - Attentional blink is category specific$^{24}$
- Object Instance Attention
  - Target objects draw attention; distracters do not.$^{20}$
  - Not as good at it as we think.$^{25}$


