Week 14: Visual attention, pop-out and search

1) Facets of attention: Bottom-up and top-down
2) Feature integration theory
3) Search rate asymmetries: Pop-out and slow search
4) Guided search and iconic bottleneck
5) Surfaces and features
6) Attentional modulation of receptive fields

MEANINGS OF “ATTENTION”
(Thanks! A. Reeves)

19th Century:
- Unity of consciousness
- Span of awareness (e.g., 7 items)
- Divided attention
- Synthesis (Cf. Treisman)
- Arousal
- Vigilance
- Cathexis (fixated, unconscious attention)

20th Century:
- Effects of instructional manipulations
- Capacity or effort
- Selection (gating, filtering, gain)
- Shifting of attention
  - with capacity limitation
  - to distinguish: attention vs. memory

METABOLIC NEED FOR ATTENTION
(E. Schwartz)

TRADE-OFF: Resolution vs. field-of-view

SOLUTION: Foveal and peripheral vision --
  space-variant representation
  cortical magnification
  log-polar map

Ability to fixate (high resolution) in a limited area creates corresponding needs for “acquisition” (selection) of targets for upcoming foveations.

UP AND DOWN

“Attention” can be

- “TOP-DOWN” -- primed by internal states, OR
- “BOTTOM-UP” -- recruited by sufficiently “salient” input events

Note 1: You do not have to attend to where/what you are fixating.
Note 2: Attention can be voluntarily “moved” without a corresponding eye movement, though the two are clearly related.

Recent developments in “awake, behaving” neurophysiology make possible the asking of detailed questions about attentional effects in specific visual areas (e.g., V4, V1, LGN, ...), such as change of receptive field size, modulation of firing rate, modulation of selectivity of firing, etc.
VISUAL SEARCH

The visual search paradigm is one among the many possible topics one could study when considering visual attention.

Observer's task:

Say whether or not a particular display contains a target, as rapidly and as accurately as possible.

Does this image contain a T? Does this one?

RESEARCH ISSUES

What combinations of target and distracters make search particularly fast or particularly slow?

If search is particularly fast, and occurs in (almost) identical time with respect to number of distracter items, the target is said to “pop out.”

Note: This is a different use of “pop-out” than that employed by SG, who speaks of a “pop-out” of “figure” from “ground.”

Are there basic features underlying visual search? (“Searchons”?) If so, what are they?

TREISMAN’S FEATURE INTEGRATION THEORY

The “default framework” for the last 20 years of research … and, lately, the one most “dumped on.”

Key ideas:

- independent feature maps
- rapid (“parallel”) search for singleton activation pattern in any one map
- slow (“serial”) search when “attentional spotlight” is required to “bind” (integrate) featural data across two maps

EXAMPLE

adapted from Treisman, 1988
ENGINEERING CONSENSUS

Note: In many publications of recent years, the hypothesis that “preattentive” processing is rapid and parallel, while “attentive” processing is slow and serial is treated as a fact, without benefit of actual proof.

Lately, a mini-boom of challenges to that hypotheses has occurred, and many alternative accounts have emerged. See especially Ahissar & Hochstein, 2002

See also Eriksen, (1990) re: “spotlight” and “zoom lens” metaphors for attention.

In any case... Treisman and colleagues have studied . . .

SEARCH RATE ASYMMETRIES

When does switching the identity of target and distracter items result in a qualitatively easier (or harder) search?

Key hypothesis:

It is easier to search a feature map -- which, by definition, codes for an important, basic feature -- for a single instance (presence) of something, than to check many locations in a feature map for the absence of something that is found in several places.

I.e. in communicating with object recognition stages, the feature maps can more readily “shout out” the existence of some property in the visual field, than they can transmit a dense, retinotopic map of where that property is or is not found.

SEARCH RATE ASYMMETRY

Which is the “feature” -- gap or closure?

WISHFUL THINKING

Search rate asymmetries are a fact of many data sets.

The interpretation of search-rate asymmetries with respect to “feature integration theory” suggests that they can be used to discover the identity of basic visual feature dimensions.

I.e. the dimensions that support “parallel” visual search in one-half of search asymmetry may be the elementary ones of all early visual processing!

So, consider...
Entries on a table of experiments listed by Treisman and Gormican, (1988):

line length,
gray [level],
curved/straight,
line orientation,
color,
circles vs. ellipses,
intersection,
juncture,
convergence,
closure and terminators
containment (convex, concave).

Sound familiar?

Note: The skepticism of the preceding slides concerns interpretation, not the validity of the phenomenon of search rate asymmetry, which is in need of explanation.

In a recent personal communication Bart Anderson -- the latest in a distinguished line of perceptual psychologists to not get tenure at MIT -- described an appealing approach to thinking about SRAs, w/r/t entropy:

Which is more unlikely . . . finding a little chunk of order in a sea of disorder, or finding a little chunk of disorder, in a sea of order?

It is not surprising that we are more sensitive to the more unlikely occurrence.

Theoretical prediction of “ideal” search experiment by feature integration theory:

Slopes of data lines (msec/item) should be either 0 (zero) or “very steep” (e.g. 40 msec/item) and uniform across experiments, because a search is either parallel or serial.

In fact, a continuum of search slopes has been found, and one can, for certain stimulus dimensions, “tune” the difficulty of a search to get a particular desired slope!

Wolfe and colleagues therefore proposed that the distinction between serial and parallel could be usefully modified to allow for an “essentially” serial process to be “guided” by partial activations from a process that sums (normalized?) activation distributions across feature maps.
GUIDED SEARCH 2.0

Basic Components of Guided Search

Note: 2.0 is not the latest release!

ICONIC BOTTLENECK

Nakayama (1990) proposed that the interface between our vast store of visual memory and the dense retinotopic maps of visual inputs may have a capacity on the order of 100 pixels!

ICONIC ISSUES

According to this proposal: the “iconic bottleneck” is the rate-limiting factor on what can be processed “deeply,” with the aid of attention.

A certain savings is afforded by the presumed pyramid of redundant, multi-scale information in feature maps, but note:

This imposes an additional burden on attention — to “find” the right level of the pyramid to access for a particular transfer of information through the iconic bottleneck!

Also, EITHER attention must “find” the correct spatial location within a level of the pyramid to solve the registration problem — whereby the information in the icon must be sufficiently scaled, rotated, etc. to be recognized — OR — a sufficiently redundant “multi-view" recognition system must be postulated, to handle the variability of input data for a particular category of object or event.

SPATIAL AND OBJECT SEARCH (SOS) MODEL

Grossberg, Mingolla and Ross (1994) proposed a model of visual search that provides an alternative to the “standard model” of preattentive -- parallel attentive -- serial processing.

Key idea: Candidate regions -- “proto-objects” that are formed rapidly and asynchronously

A single candidate region may act as a unit for comparison with stored templates of target attributes.

A candidate region may be
1) the desired target
2) entirely lacking in target features ==> quickly rejected
3) ambiguous ==> processed more deeply, possibly by being segmented into smaller regions
MACROCIRCUIT OF SOS MODEL

STEP 1: retinotopic feature coding

STEP 2: spatial grouping by featural qualities

STEP 3: candidate region selection

MATCH? if no, go to step 2

STEP 4: object recognition

SIMPLIFY! SIMPLIFY!

A “quick and dirty” implementation of the SOS model requires only two equations involving high-school algebra! It can simulate data patterns from certain visual search experiments:

\[ RT = R + (N + 1)(S + M) \]
-- (target present)

\[ RT = R + 2N(S + M) \]
-- (target absent), where

RT is reaction time

R is the duration necessary for retinotopic feature coding (Step 1)

N is the number of candidate regions (mean for a class of scenes -- computed by super-simple “image processing”)

S is the duration necessary to segment and select a candidate region (Steps 2 and 3)

M is the duration necessary to match a candidate region with the target representation (Step 4)

SEARCH EXAMPLE

Cohen & Ivry (1991) search when items are “clumped” is more difficult than for when items are spread out.

SOS ON CLUMPING

Analysis: Sparse scenes have fewer candidate regions

Cf. long-range completion and spatial impenetrability
Enns & Rensink (1990) showed that search can be rapid for combinations of shaded polygons that yield an “immediate” impression of 3-D structure, but slow for corresponding combinations of polygons that look like 2-D objects.

Similarly, He & Nakayama (1994): visual search stereograms -- Disparities for each element (target or distracter) could be reversed. Search was easy for normal and reversed L shapes, but hard when those same L-shaped regions were perceived as visible parts of an occluded (amodally completed) square region.

**PHYSIOLOGY**


Abstract. Single cells were recorded in the visual cortex of monkeys trained to attend to stimuli at one location in the visual field and ignore stimuli at another. When both locations were within the receptive field of a cell in the prestriate area V4 or the inferior temporal cortex, the response to the unattended stimulus was dramatically reduced. Cells in the striate cortex were unaffected by attention. The filtering of irrelevant information from the receptive fields of extrastriate neurons may underlie the ability to identify and remember the properties of a particular object out of the many that may be represented on the retina.

Read this article!
BOTTOM LINE

1) Attention can be operationally manipulated.
   • Single-unit electrophysiology on an awake, behaving monkey
   • Reinforcement paradigms used to “force” the monkey
to attend in order to perform a task at all

2) Attention can be shown to drastically reorganize “default”
receptive field characteristics, including size, in V4 and IT.

Researchers are now asking “how low” can the effects of
attention be measured (e.g. V1, LGN), and in what ways does
attention affect cell functions.

Cf: fMRI studies

WHAT NEXT?

Functional MRI reveals spatially specific attentional modulation
in human primary visual cortex
D. C. SOMERS, A. M. DALE, A. E. SEIFFERT, & R. B. H. TOOTELL‡

(Some of) WHAT (I hope) YOU LEARNED IN CN530

(Bad*) modeling is easy; understanding biological vision --
whether by empirical measurement or by modeling -- is hard.

The separation of computational theory, algorithm and
representation, and (biological) implementation can be overdone.

Recurrent competitive networks may underlie a metric of vision
“... far looser one than any philosopher ever proposed or any
psychologist dreamed.” (Lettvin, 1981)

The aspects of visual stimulation that are informative (useful)
to animals are (perhaps) not what you thought. (J. J. Gibson.)
(Consider early stereo algorithms and da Vinci stereopsis.)

* writing equations that describe the way one thinks vision is or
“ought to be”, unconstrained by any influences of data.

AND THE UNITS OF VISION ARE . . .

The units of vision are emergent wholes (gestalts),
not a fixed set of elementary features.

Certain gestalts are recur consistently in characteristic
situations, but “if the doors of perception were cleansed . . .”

Many important units of vision are scientifically describable,
(but they may seem strange; e.g. “candidate regions.”)