WEEK 1: FUNDAMENTAL PROBLEMS OF VISION

1) Unit formation and grouping
2) Seeing and recognizing -- form/color interactions
3) Retinal veins and blind spot
4) Perceiving surface color: Constancy, contrast, and discounting the illuminant
5) Stabilized images: Boundaries and featural color and brightness
6) Complementary processing: Unoriented and oriented filters

Addendum:
1) The noise--saturation dilemma
2) Reflectances and ratios; shunting and mass action

WHY STUDY VISION?

Among all the studies of natural causes and reasons Light chiefly delights the beholder; and among the great features of Mathematics the certainty of its demonstrations is what preeminently (tends to) elevate the mind of the investigator. Perspective, therefore must be preferred to all the discourses and systems of human learning. In this branch [of science] the beam of light is explained on those methods of demonstration which form the glory not so much of Mathematics as of Physics and are graced with the flowers of both. But its axioms being laid down at great length, I shall abridge them to a conclusive brevity, arranging them on the method both of their natural order and of mathematical demonstration; sometimes by deduction of the effects from the cause, and sometimes arguing the causes from the effects; adding also to my own conclusions some which, though not included in them, may nevertheless be inferred from them. Thus, if the Lord -- who is the light of all things -- vouchsafe to enlighten me, I will treat of Light; wherefore I will divide the present work into 3 Parts.


PREFACE

The following lecture notes for Week 1 of CN 530 contain many terms that are likely to be unfamiliar to you. In addition, many ordinary-seeming words are used in a manner that can only be fully understood in the context of material that you will not encounter until later in the course.

The purpose of this lecture is to (dis?)orient you as quickly as possible to the scope of issues addressed and methods of inquiry adopted in the course.

Please do not worry if parts, or even most, of this lecture make no sense on first presentation. If you are looking at these notes before attending the first lecture, please be aware that they may be difficult to interpret without a spoken “sound track.” Lecture notes for later weeks are more straightforward.

Please use this lecture as a “site map” for locating the contents of subsequent lectures with respect to the goals of the entire course.

THE UNITS OF VISION

The objects of perception and the space in which they seem to lie are not abstracted by a rigid metric but a far looser one than any philosopher ever proposed or any psychologist dreamed.

Lettvin (1981)
In what substrate do these functional units exist? Are they biochemical processes?? … algorithms?? … network patterns?? … aspects of consciousness??

**EMERGENT SEGMENTATION AND GROUPING**

- Proximity → Collinearity
- Gestalt “Laws”
The “Gestalt Laws” refer to perceptual grouping or unit formation include on the basis of properties such as:

- similarity
- proximity
- closure
- symmetry
- *pragnanz* (literally “pregnant”)
- “good continuation”
- “common fate”

See Koffka, 1935 and Köhler, 1947
See also: [http://www.ship.edu/~cgboeree/gestalt.html](http://www.ship.edu/~cgboeree/gestalt.html)

**COHERENT PATTERNS**

Glass patterns

*Coherence*: We select the globally most consistent groupings from among all possible local groupings.

These emergent boundary groupings are *invisible*. We continue to “see” only dots, although we are aware of and “recognize” circular groupings.
Possible groupings of oriented line segments:

- Collinear
- Perpendicular
- Oblique

Global grouping may or may not be in the same orientation as local line segments.

"Emergent features" can form via linking of local features to enable us to segment one image region from another.

We "see" the same kind of image contrast for a given element, whether or not that element is part of some larger grouping.

Q: How are these displays constructed? Why?

The observer's task in these displays is to say whether the top and bottom halves are "same" or "different."
ILLUSORY CONTOURS

Ehrenstein (1941/1987):

The form we **recognize** is not “in the image.”

We **see** the disk region as brighter than the background.

---

DOMINANCE OF GLOBAL GROUPINGS

-can distort local information
-even collinear groupings

café wall illusion


Note: For reasons that surpass understanding, the existence of the illusionworks.com web site is a stochastic process.

---

DISCLAIMER: COGNITIVE FACTORS

From Marr 1982 -- R. C. James

... by way of Sinha & Adelson, 1997

---

PERCEPTION “RESISTS” COGNITIVE INTRUSIONS

Fig 18, Kanizsa & Vicario, 1968
CRAIK-O'BRIEN-CORN SWEET EFFECT

long-range interactions in what we see in a region.

After Todorović, 1987

The COCE is the first of many examples that we will encounter that suggests that . . .

a perceptual surface is not simply a union of local patches.

“Hold this thought” to compare with later evidence.

The kind of taxonomy of visual phenomena that we have been pursuing so far is not sufficiently constrained to help us model visual processing.

How, then, should we proceed? computational constraints? physiology?

FROM PHENOMENOLOGY TO GEOMETRY

Bar (line): not a union of dots

Surface: not a union of local patches

For these units: extension
interiors
contours separating them from surround

Not Euclidean geometry or calculus of Newton/Leibniz

“Fact: The notion of “surface area” is devoid of any meaning unless you specify the resolution at which it is to be assessed. The same goes ipso facto for the notion of “arc length.”

Koenderink Solid Shape 1990

ILLUMINATION, REFLECTANCE, AND VISION

\[ R_{x,y} = \frac{E_{x,y}}{l_{x,y}} \]
AMBIGUITY OF LOCAL INFORMATION

\[ E_{x,y} = f(I_{x,y} R_{x,y}) \]

How could you tell from purely local information whether a surface patch is planar under nonuniform illumination or smoothly curved in depth?

If you cannot make this distinction at a certain point in visual processing, what can you base subsequent processing on?

Illumination vs. Reflectance effects

Claim: We must be able to distinguish the visual effects of surface shape (orientation) and reflectance from the effects of variations in illumination* at a fairly early stage of visual processing, because virtually every other important visual competence depends on being able to make this distinction.

Note: How animals do this is still unknown for general illumination conditions! [That’s a major understatement; put another way: we’re still far away from having a general computer vision system.]

NOTE: We perceive a world of objects and events, not of light. (Gibson, 1950, 1966, 1979)

* E.g. cast shadows; variable distances from source, . . .

BRIGHTNESS CONSTANCY

REFLECTANCE: A surface property

We seem to estimate the ratio of reflectances across boundaries.
Physical reflectance is a surface property that is constant over time, specifically w/r/t variable illumination.

It would be nice if our perceptions of surface color were similarly constant.

Note: We will confront two definitions of “reflectance;” besides physical reflectance (radiance as a ratio of incident illumination), we have Grossbergian reflectance, the ratio of the magnitude of input to one node in a network relative to total input to the network.

FIRST EXPERIMENT: If the intensity of the red (long wavelength) illuminant is doubled or tripled, the colors in the Mondrian still look “much the same.” We somehow factor away the “extra” red.

Helmholtz: “Discounting the illuminant” (wavelength and intensity)

Land -- McCann “Mondrians,” I

Use different illumination gradients in different wavelengths, adjusted so as to “offset” the effects of spectral reflectance of two patches.

Different colors seen from the same spectrum

... similar to those seen in white light

GRADIENTS OF ILLUMINATION AND REFLECTANCE
“RETINEX” STRATEGY

1. Recover relative reflectances (ratios) near image edges.

\[ \frac{a}{b} \quad \frac{c}{d} \]

2. Suppress information from slowly varying region interiors.

How to go from IR_\lambda to f(R_\lambda)?

Ideally, some simple function

To be able to “discount” the illuminant (intensity) with a Retinex-like strategy, the Illumination gradient must be more gradual than the gradient of change in reflectance.

This is not true in the Gelb effect.

THE GELB EFFECT

From Kaufman, 1974.

Exercise: Look around you and point to the surface of an object whose appearance would be closest to that of a moon rock, if that moon rock were in the room with you now.
RATIOS FROM EARLY BIOLOGICAL PROCESSING

If early representation looks like this we need “filling-in” for perceived surface interiors; otherwise, we would “see” a world of line drawings.

RATIOS NOT ENOUGH

BRIGHTNESS CONTRAST

Two small disk patches of equal luminance. One annulus is of high luminance. The other is of low luminance.

Contrast effect:
Perceived brightness of inner disk varies in direction opposite to luminance of annuli.

Claim: This occurs because some representation of the “sum of ratios” of inputs for the two scenes is approximately constant for each view.

That is, normalization occurs, whereby – for some functional spatial domain – the quantity of total “energy” in a representation (or output of some network) is conserved, as the sum of inputs varies over some range. (Cf. Simulation Assignment 1.)

Normalization within functional spatial domains -- total “energy” in a representation is conserved, as the sum of inputs varies.

Not just “lateral inhibition” -- but “anchoring".

Gilchrist e amici
The retina is the “interface” between a mammal and its visual environment.

Starting from the interface, research on visual perception could be motivated by:

1) a detailed analysis of the visual system’s mechanisms or
2) a detailed analysis of the visual environment.

Ultimately, these two lines of inquiry would converge, because the visual environment has shaped the evolution of our visual systems.

Pattern formed on retina by a dark line

Completion needed for “real” contours

Note: this statement is still controversial.
**Completion Over the Blind Spot**

- **DEMO:** Close right eye and fixate upper cross with left eye.
- Hold page at about 1 ft from the eye, and move it back and forth in depth slightly until the disk on the left disappears.
- Similarly, when fixating the lower cross, the gap in the black line can be made to fall on the blind spot, and the line is seen as continuous.


---

**Eye Micromovements**

- The eye jiggles constantly in its orbit. ("tremor," approximately 40 Hertz*)
- The shadows of retinal veins do not move relative to the photoreceptor mosaic.
- They therefore form "stabilized images."

**But stabilized images fade.**
- (Time scale: seconds)


**Note:** While the stabilization of veins accounts for our not "seeing" them, the line that we do see still makes an **incomplete** pattern on the retina.

* Local expert: Prof. Rucci

---

**Compensation for Retinal Gaps**

- **Emergent boundary formation (completion)**
  - *Which boundaries to connect?*

- **Featural filling-in**
  - What color and brightness do we **SEE**?

* Not the same as "recognition"

---

Examples of emergent segmentation:
- 1) Collinear (Beck, Bozzi)
- 2) Perpendicular to line ends (Beck, Ehrenstein, Varin neon)
- 3) Diagonal, oblique (Beck, Kennedy)
- 4) "Other" (Beck's T's and L's, dalmation)

The contrasts and hues that we "see" are not the only relevant perceptual structures.

We segment and group images on the basis of "emergent boundaries," that are based on image contrasts, but not isomorphic to the pattern of contrasts.

* Why?
How to stabilize an image... the hard way!

Caps needed to be of low mass and low moment of inertia; why?

Fig 24 from Yarbus (1967)

When indicated boundaries are stabilized...

...white and black are no longer visible, but the effects of contrast remain!

STABILIZED IMAGE EXPERIMENTS

YARBUS EXPERIMENT (1967):

A red dot moves back and forth.

NOTE: time scale of seconds

HOMMADE STABILIZATION

NOTE: time scale of seconds

BOUNDARIES AND FILLING-IN

An image like this:

When stabilized on the retina,

stabilize

red

black

red (same)

darker red

lighter red

stabilize
**INFERENCES FROM YARBUS EXPERIMENT**

*Boundaries restrict featural filling-in* (color and brightness).
When boundaries fade, colors flow.
Boundary fading does not imply color fading.

Two subsystems in early vision:
- **boundary system**
- **feature system** (brightness, color)

**BOUNDARIES AND VISIBILITY**

Q: What happens if there are *no boundaries* in a visual field?

_Ganzfeld:_ Completely homogeneous visual field -- can be approximated by wearing goggles made from halves of a ping-pong ball.

A: *When viewing a ganzfeld, brightness fades.* (time scale?)

Note: Homogeneous regions of any image are _de facto_ stabilized.
e.g. “clear blue sky”

Homogeneous regions carry no information.

**BOUNDARY PROCESSING**

Boundaries: how to
- _detect_
- _sharpen_ (sometimes)
- and _complete_?

Unoriented and _oriented receptive fields_ (masks, filters, kernels)

**SIMPLE CELLS: ORIENTED LOCAL CONTRAST FILTERS**

Sensitive to:
1) orientation
2) amount of contrast
3) direction-of-contrast (i.e. contrast polarity)
4) spatial scale
5) position

_Not_ (just) “edge detectors.”

Hubel & Wiesel, 1962 . . . Nobel Prize
Local oriented-contrast filtering alone is *not* enough:

Again, boundary ≠ brightness.

Cf: Shapley & Gordon, 1985

Evidently, while sensitive to:
1) orientation,
2) amount of contrast,
3) spatial scale, and
4) position,
*human* boundary processing is

*insensitive* to *direction of contrast*
(i.e., *contrast polarity* of “edges”),

*when completing over gaps*  
(*i.e.,* over a sufficiently large spatial scale.)

**COMPLEMENTARITY**

 Boundaries: *Completion*  
- oriented
- inward
- insensitive to direction-of-contrast

 Surfaces: *Filling-in*  
- unoriented
- outward
- sensitive to direction-of-contrast

**UNITS OF VISUAL REPRESENTATIONS**

**UNCERTAINTY:** Some kinds of information are *incompatible*; they cannot be represented simultaneously by a single unit.

**QUANTIZATION:** How to maximize resolution, given that vision cannot “do calculus” with infinitesimals?
UNCERTAINTY AND QUANTIZATION

Thin perceptual line: Oriented filter:
NOT a union of points NOT an edge detector
Fuzzy/statistical object Uncertainty
Boundary and feature Localization of “edgel” AND plausible grouping directions

OBJECT RECOGNITION SYSTEM
Learns and remembers perceptual codes

Boundary Contour System
- grouping
- completion
- sharpening

Feature Contour System
- filling-in
- of brightness, color

PREPROCESSING
- ratios at edges

INPUT

NOTE:
This diagram includes “cognitive expectations”

WHERE’S THE “INTELLIGENCE” IN VISION?

Helmholtz: “as if” unconscious inference
Gestaltists: field theory
Gibson: direct perception; “pickup” of environmental information
Marr: computational theory
Grossberg: network architecture & resonance
Schwartz: maps
You: _______

WHAT ARE THE UNITS OF VISION?

This course is about the geometry of vision.

(Recall Lettvin quote.)

In the visual universe, we have, at times:
coherence among disconnected elements, and
segmentations through homogeneous regions!
What kind of geometry does this?
What could possibly be its functional units?

Consider: The “preattentive” visual system at times makes use of direction-of-contrast information and at other times ignores it, thereby rivaling the performance of the most efficient cognitive engine yet invented: the thermos bottle!
Addendum

1) The noise--saturation dilemma
2) Reflectances and ratios; shunting and mass action

THE PLAN (including next week)

Feedforward shunting competitive networks

Distance-dependent interactions

Functional view of network properties:
- *motivated* by physiology
- *described* abstractly.

Try to understand certain *molar* perceptual
(or "computational") properties
with reference to simplified *micro*-interactions.

COMPETITION

A maximum and a minimum number of
sites (processes) can turn *on* or *off*;
each site's process is essentially
"all or none."

Key
- **o off** (unexcited site)
- **o on** (excited site)

Competition is ubiquitous in the biological world.

Why competition on the *cellular* level?

*What is a cell*, computationally speaking?

*Infinity* does not exist in biology;
so our neural models should not assume
that it does -- not even *implicitly.*
NOISE-SATURATION DILEMMA
Pattern registration by cellular systems

Given the constraints of a finite number of sites with all-or-none processes, how does nature prevent the following?

Let: 
\[ I_n = x^n \]
\[ I \] stands for intensity, or energy of input

<table>
<thead>
<tr>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( x_3 )</th>
<th>( x_4 )</th>
<th>( x_5 )</th>
<th>( x_6 )</th>
<th>( x_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I_1 )</td>
<td>( I_2 )</td>
<td>( I_3 )</td>
<td>( I_4 )</td>
<td>( I_5 )</td>
<td>( I_6 )</td>
<td>( I_7 )</td>
</tr>
</tbody>
</table>

range in which input signals can not be distinguished from endogenous noise

effective operating range

saturation range

NOISE-SATURATION DILEMMA, AGAIN

NEED: automatic gain control (retuning of sensitivity)

HARDWARE AND WETWARE
The conventional engineering solution for an automatic gain control . . .

CONSTRUCTION OF SIMPLE SHUNTING NETWORK, PART 1

In a neural network:
1) Let each input source fluctuate in amplitude;
2) Let the total number of convergent sources fluctuate during pattern processing.

But: Each node has a fixed operating range. (Outputs stay within upper and lower bounds.)

\[ \begin{array}{c|c|c|c|c|c|c|c} \hline \text{output range} & x_1 & x_2 & x_3 & \quad \quad & x_{n-1} & x_n \ \hline \end{array} \]

Grossberg, 1973:
If \( x_i \)'s are sensitive to small inputs, why don't they \textit{saturate} in response to large inputs?
If \( x_i \)'s are sensitive to large inputs, why don't small inputs get lost in \textit{endogenous noise}?
The previous panel’s point can be put like this:

Normalization is easy; knowing when and where to normalize and when and where to not normalize is hard.

A current leading theory of lightness perception (Gilchrist et al.) is based on interactions among “frameworks” -- i.e. regions of the scene within and between which ratios of luminance are to be compared.

Good news: The formula’s for computing the ratios and related quantities are pretty well worked out.

Bad news: There is nothing like an effective procedure for determining what the frameworks are for even modestly complex “scenes.”

Classical definition:
the ratio of reflected ($R$) to incident ($I$) light for a Lambertian (i.e., ideal, perfectly diffusing) surface.

$$0 < R/I < 1 \text{ for each surface}$$

COMPARE:
Grossberg’s definition of reflectance of network inputs:

Input to a node: $I_i$ or $I_i(t)$ for $i = 1, \ldots n$

Total input: $I = \sum_j I_j$

“Reflectance” of an input: $\theta_i = \frac{I_i}{I}$

What about Grossberg’s phrase “cell or population”?

**Physical Reflectance**

**Ratio Sensitivity in Shunting Networks**

**Notation:**

- $v_i$ name of the $i$th node
- $x_i$ potential or activity level of the $i$th node
- $f(x_i)$ output signal function of the $i$th node (often involving, e.g., a rectification, thresholding, or sigmoidal transform)
CONSTRUCTION OF SIMPLE SHUNTING NETWORK, PART 2

Assume: (1) Inputs perturb cell sites, and (2) activity at each site is “all or none.”

Graphical convention: Each little circle stands for a site.

Activity $x_1(t)$ $x_2(t)$ $x_3(t)$ $x_n(t)$ $V_1$ $V_2$ $V_3$ $V_n$

$I_1(t)$ $I_2(t)$ $I_3(t)$ $I_n(t)$

Activity

Assume: $B$ total number of sites at each cell

Inhibitory inputs affect only these: $x_i(t)$ excited sites at time $t$

Excitatory inputs affect only these: $B - x_i(t)$ unexcited sites at time $t$

WHY (GO FORTH AND) MULTIPLY?

Assume:

Activity of excited sites spontaneously decays to equilibrium (first term.)

Activity of unexcited sites is increased by inputs, in proportion to size of input and number of unexcited sites (mass action*, connoted by multiplication in second term).

* implies statistical independence, therefore multiplication:

$$\frac{dx_i}{dt} = -Ax_i + (B - x_i)I_i$$

Expected value: $E(X \text{ and } Y) = E(X)E(Y)$

MASS ACTION

Reaction rate for particles of two types in a medium?

Analogy: Consider a chemical reaction occurring among particles of two types that are sparsely distributed in a fluid container (e.g. yeast and sugar, for fermentation).

Reaction occurs only upon collision of particles of different types.

Random process: Rate of reaction is proportional to product of the concentrations of the two types of particles.

In network equation, multiply input size by measure of available (inactive) sites,...

because there are many sites, and a single input is thought of as arriving at particular sites randomly along similarly numerous, “all or none” channels.
**NEED FOR INHIBITION**

Consider:  
\[
\frac{dx_i}{dt} = -Ax_i + (B - x_i)I_i
\]

To compute equilibrium response, set:  
\[
\frac{dx_i}{dt} = 0
\]

Solution:  
\[
x_i = \frac{B\theta_i I_i}{A + \theta_i I}
\]

Remember:  
\[
\theta_i = \frac{I_i}{\sum_k I_k}
\]

If \( I \) is small relative to \( A \), inputs are lost in noise,* because \( A \) of denominator dominates.

If \( I \) is large relative to \( A \), all nodes saturate at the value \( B \).

*CHEATING ALERT: There’s no “noise” in the equation.

**COOPERATIVE AND COMPETITIVE NETWORKS**

Network constructed so far “didn’t have a chance.”

To be sensitive to Grossbergian reflectances, each node’s output must be influenced by all of the inputs, by definition:

\[
\theta_i = \frac{I_i}{\sum_k I_k} = \frac{I_i}{I_i + \sum_k I_k}
\]

**INPUT**

**OUTPUT**

<table>
<thead>
<tr>
<th>Increasing ( I_i )</th>
<th>tends to increase ( \theta_i )</th>
<th>COOPERATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increasing ( I_j ) (( j \neq i ))</td>
<td>tends to decrease ( \theta_i )</td>
<td>COMPETITIVE</td>
</tr>
</tbody>
</table>

G: “on-center, off-surround”

network anatomy (Kuffler, 1953)

CAREFUL: There’s no “distance” between nodes in this network!

**SHUNTING INHIBITION**

Next, suppose that excited sites are inhibited by mass action:

\[
\frac{dx_i}{dt} = -Ax_i + (B - x_i)I_i - x_i \sum_{k \neq i} I_k
\]

Here, each node receives “equal weight” inhibitory input from all input sources except the one with the same index value.

NOTE: The use of indices does not here imply a functional spatial ordering (“neighbors”) among nodes.

**EFFECTS OF SHUNTING INHIBITION**

Now, at equilibrium:

\[
x_i \rightarrow \theta_i \text{ as } I \rightarrow \infty
\]

*The word “pattern” here means a normalized vector quantity; elsewhere, the word may be used to mean any vector quantity.

QUIZ: Sensitivity of node potential to (Grossbergian) reflectances is said to “model” brightness constancy. Why?
NORMALIZATION: CONSERVATION OF TOTAL ACTIVITY

As total input increases, what happens to total activity in the network?

\[ x = \sum_k x_k = \frac{BI}{A+I}, \quad \text{since } \sum_k \theta_k = 1 \]

Result, total network activity asymptotically approaches a constant value, \( B \).

This fact is said to "model" (or "explain") brightness contrast.

Why?
Note: There are aspects of the phenomenon of brightness contrast that can not be well-modeled by this equation.