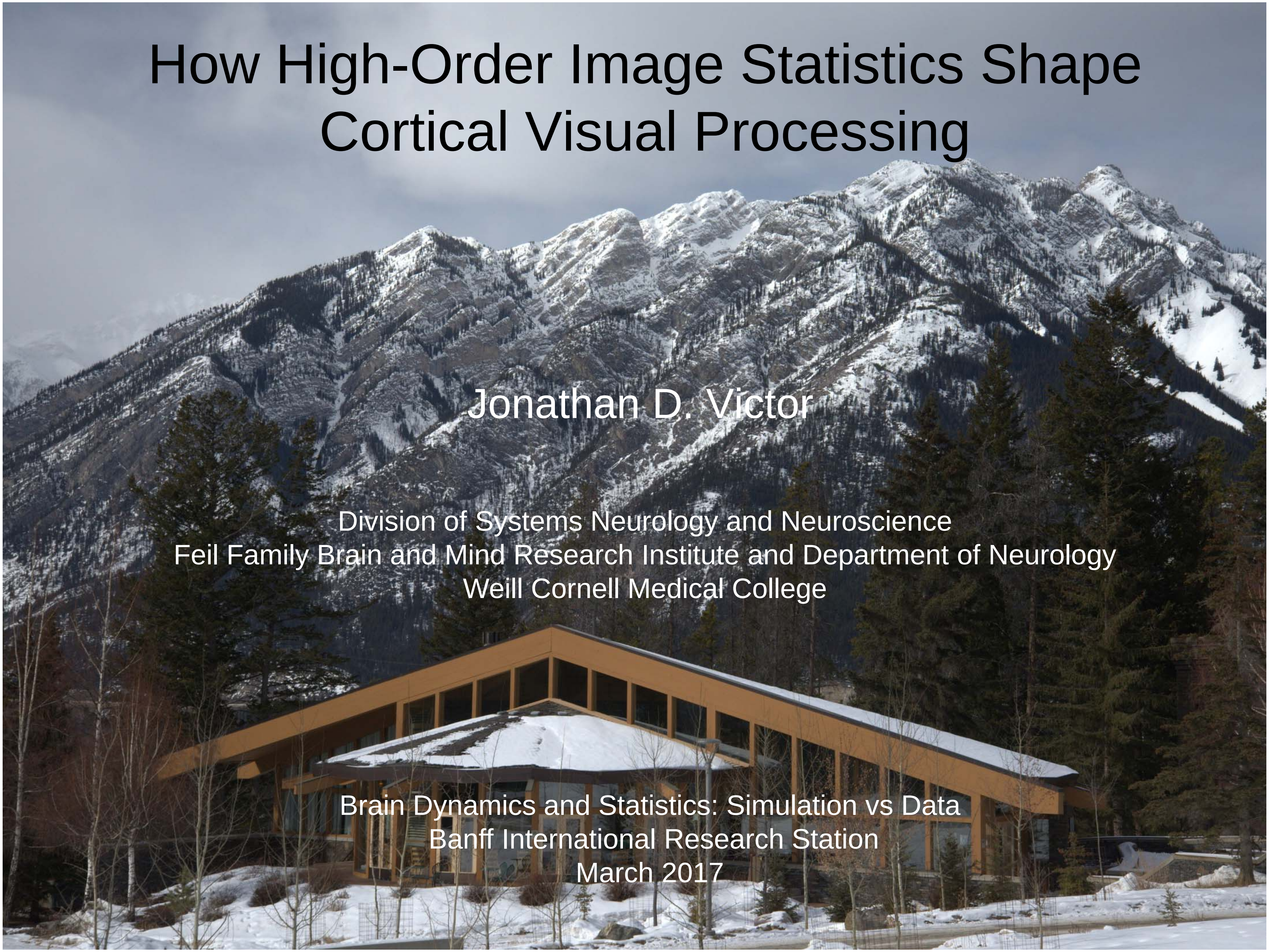


How High-Order Image Statistics Shape Cortical Visual Processing

Jonathan D. Victor

Division of Systems Neurology and Neuroscience
Feil Family Brain and Mind Research Institute and Department of Neurology
Weill Cornell Medical College

Brain Dynamics and Statistics: Simulation vs Data
Banff International Research Station
March 2017

Acknowledgments

WCMC Neurology and Neuroscience/
Brain and Mind Research Institute

Psychophysics
Mary Conte
Dan Thengone
Syed Rizvi

Physiology
Yunguo Yu
Anita Schmid
Dan Thengone
Keith Purpura
Ferenc Mechler
Ifije Ohiorhenuan
Qin Hu

University of Pennsylvania
Natural image statistics
Vijay Balasubramanian
Ann Hermundstad
Gašper Tkačik
Jason Prentice
John Briguglio

Boston University
Eye movement statistics
Michele Rucci
Xutao Kuang
Martina Poletti

Support
NEI EY7977, EY9314
WCMC Neurology and Neuroscience/BMRI

Outline

- The domain
  - understanding neural computations (why, not only how)
  - the visual system as a model
- The level of the retina
  - classic example of what we hope to achieve
  - dogma overturned
- The level of primary visual cortex
  - the emerging match between properties of the natural environment and neural computations

The Efficient Coding Hypothesis

Barlow, 1961

- Sensory systems make use of regularities in the natural world to perform efficiently

- More specifically: sensory systems devote few resources to transmitting and processing what is predictable, so they can devote more resources to what is unpredictable (and therefore, most informative)

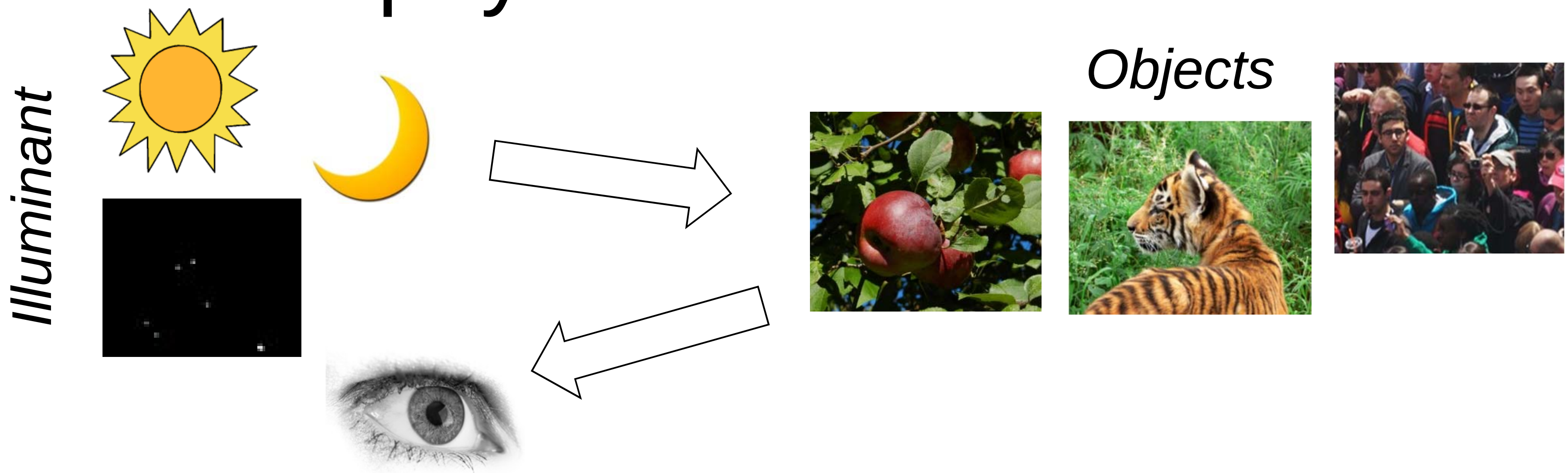
Efficient Coding is Not a Trivial Hypothesis

- It is not obvious that efficient coding is worth the costs
  – Computational costs: energy, space, time
  – Genetic and/or developmental burden
- "Efficient coding" ignores the value of specific stimuli, and the costs of different kinds of errors
- It is not obvious that the natural world even has any useful statistical regularities

Testing Barlow's ideas

- Identify statistical regularities in natural scenes
- Formulate biologically-plausible signal-processing strategies for exploiting them
- Make measurements to see whether these strategies are used

A source of statistical regularity: the physics of illumination



Light into the eye=(Illuminant intensity)*(Object reflectance)

Problem:

Illuminant:
•wide range (10^9)
•constant within a scene
•doesn’t distinguish objects

Object reflectance:
•narrow range (10^2)
•variable within a scene
•critical to distinguish objects

Solution:

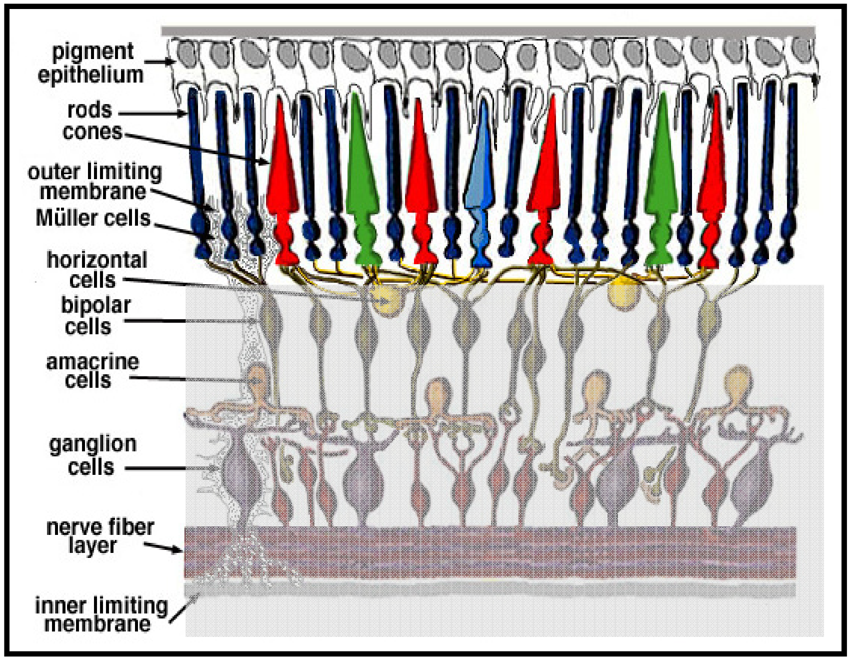
A) Compress the wide dynamic range:

log(Light into the eye)=log(Illuminant intensity)+log(Object reflectance)

B) Remove the confounding effect of the illuminant:

Subtract the average (illuminant averaged over time and space)

This accounts for the first stages of retinal processing



H. Kolb, E. Fernandez, R. Nelson: http://webvision.med.utah.edu/imageswv/

- Intensity-response curve of photoreceptors is close to a logarithmic transformation
- To a first approximation, photoreceptor adaptation subtracts the temporal average
- To a first approximation, outer plexiform circuitry subtracts the spatial average

Solution:

A) Compress the wide dynamic range:

log(Light into the eye)=log(Illuminant intensity)+log(Object reflectance)

B) Remove the confounding effect of the illuminant:

Subtract the average (illuminant averaged over time and space)

Another source of statistical regularity: the world is made of objects

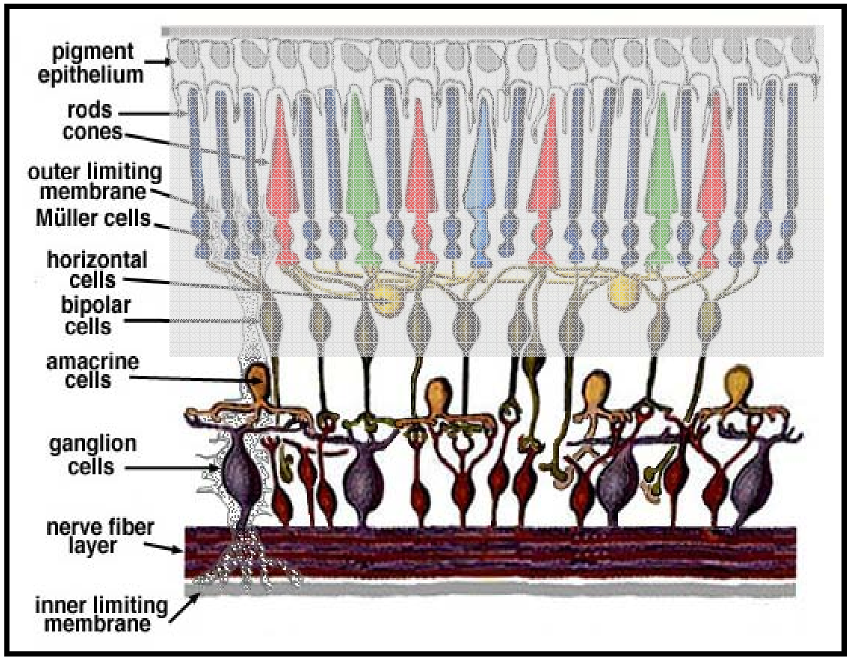
- Objects tend to be homogeneous
- So intensities at nearby points of an image tend to be similar
- Ignoring these correlations inflates the number of bits needed to signal an image

Problem: how to exploit this?

Solution: transmit the difference between the actual signal and the best guess based on the surround. This error signal (at neighboring point pairs) is uncorrelated.

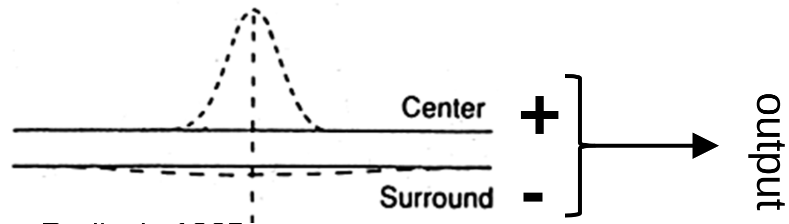
(“Predictive coding”, “redundancy removal”, “whitening”, underlies JPEG, MPEG)

This accounts for the output stage of retinal processing



H. Kolb, E. Fernandez, R. Nelson: http://webvision.med.utah.edu/imageswv/

- Ganglion cells have center-surround organization



Rodieck, 1965

- The surround computes the best guess for the center
- The difference between center and surround is an error signal, uncorrelated across ganglion cells

Atick and Redlich, 1990

Solution: transmit the difference between the actual signal and the best guess based on the surround. This error signal (at neighboring point pairs) is uncorrelated.

(“Predictive coding”, “redundancy removal”, “whitening”, underlies JPEG, MPEG)

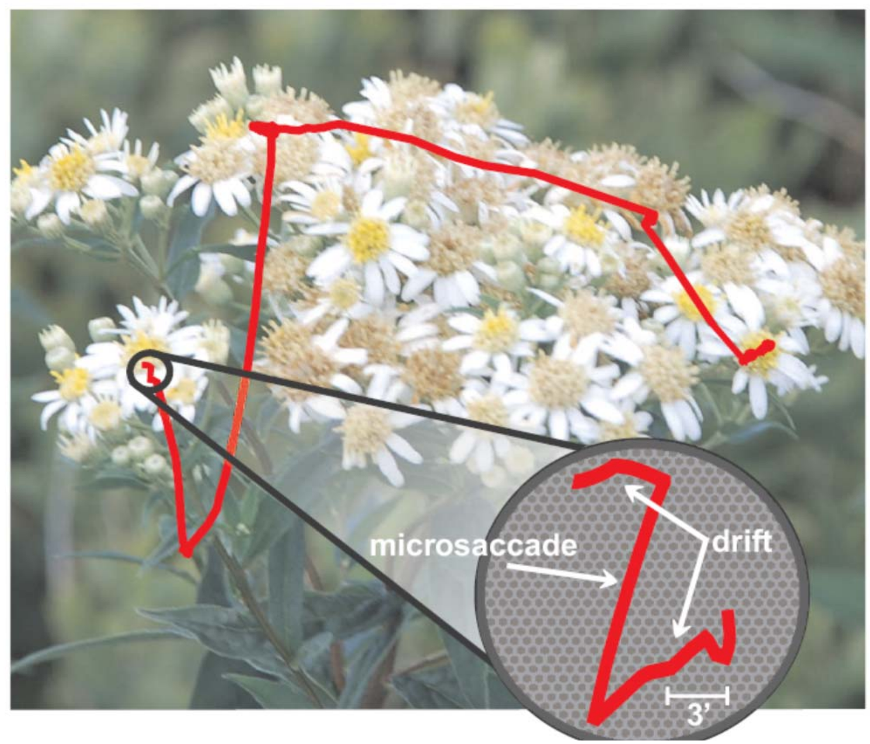
Outline

- The domain
  - understanding neural computations (why, not only how)
  - the visual system as a model

- The level of the retina
  - classic example of what we hope to achieve
  - dogma overturned

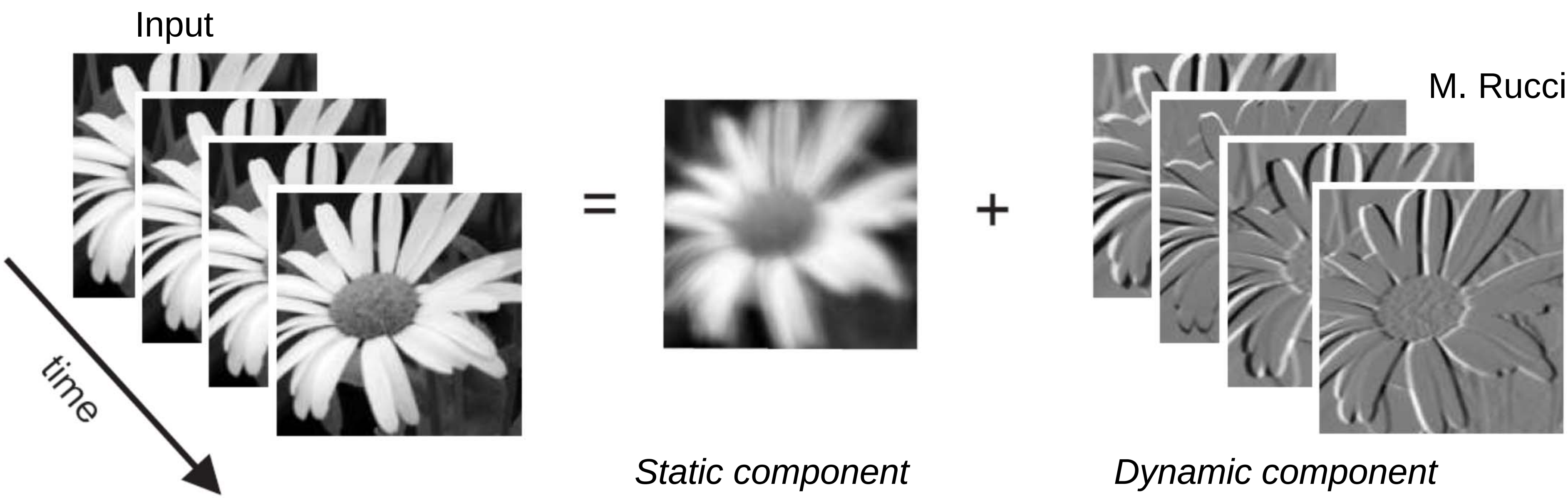
The eyes move!

- Saccades
  - Abrupt refixations, approximately 3 per sec
  - Move areas of interest into the fovea
- Smooth pursuit
  - Tracking movements
  - Stabilize the image of a moving target
- Fixational eye movements
  - Occur *between* saccades
  - Apparently random
  - They sweep across many photoreceptors
  - Function unclear



M. Rucci

What is the effect of fixational eye movements on a retinal image?



Static component
• blurred
• suppressed by temporal filtering in the retina

Dynamic component
• edge-enhanced
• accentuated by temporal filtering in the retina

It turns out (not obvious!) that the dynamic component is decorrelated at point pairs, as required to reduce redundancy

Dogma overturned:
- Decorrelation happens, but it happens via fixational eye movements prior to circuitry
- Retinal ganglion cells don't decorrelate; they begin the process of edge detection

Summary so far:

- Problem: achieving sensitivity to small luminance differences over a wide operating range
- Solution: log transform and background adaptation (photoreceptors and initial stages of neural processing)

- Problem: an output bottleneck
- Solution: reduce redundancy (fixational eye movements)

- Problem: extracting meaning from the image
- Partial solution: extract edges (center-surround processing via retinal ganglion cells)

Beyond the retina

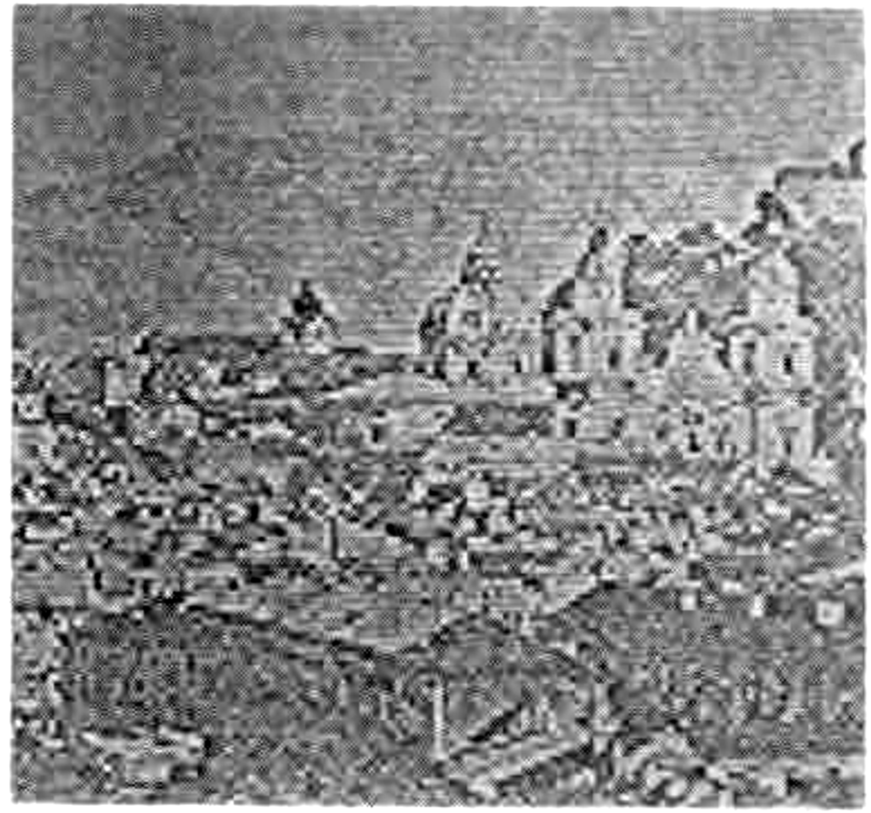
The moving retina removes 2-point correlations to achieve efficiency, but multipoint correlations (3 or more points) remain. What do we expect at later processing stages?

- Removal of these correlations to achieve further efficiency?

- Exploit these correlations to extract meaning?

Multipoint correlations carry visual form

original image



2-point correlations removed
(only multipoint correlations)

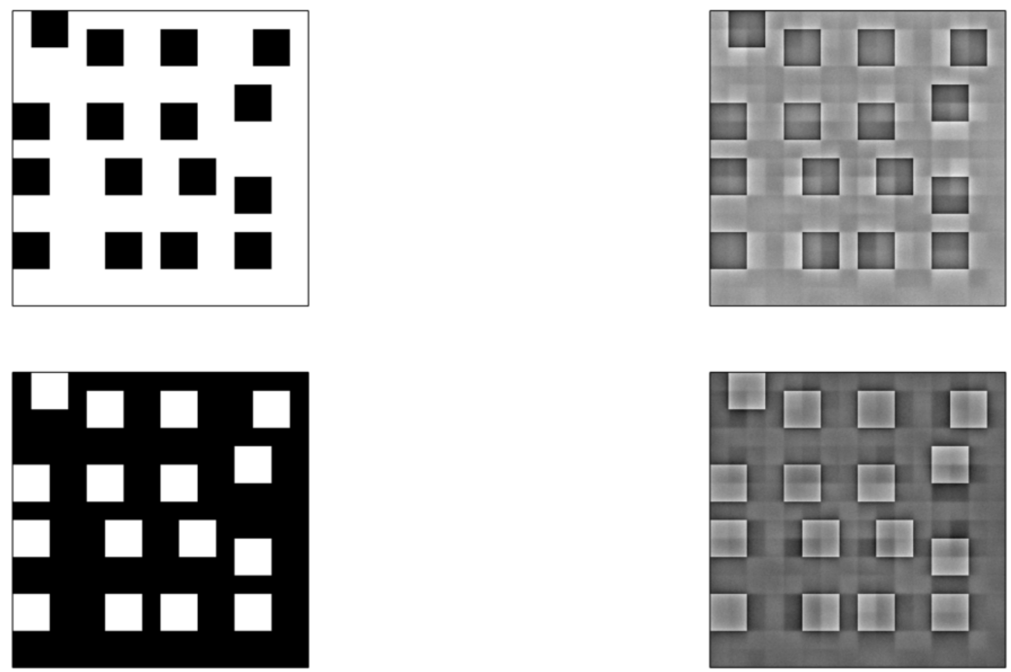
Oppenheim and Lim, 1981

Multipoint correlations carry illusory contours

original image

only multipoint correlations

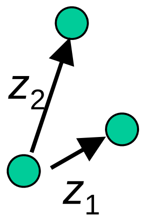
and figure-ground



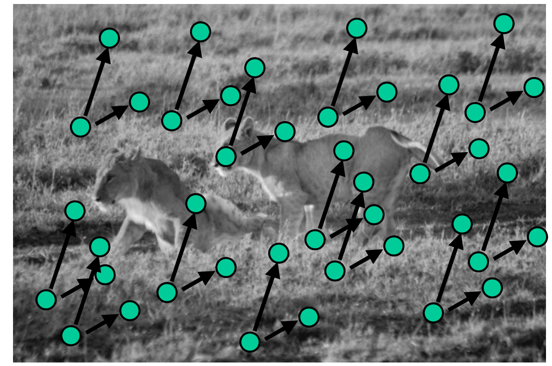
Y.Yu, 2014

Computing multipoint correlations takes a lot of resources

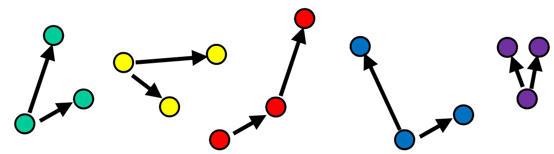
A single three-point correlation:



The three-point correlation for this template is the average of the product of image intensities $I(x)I(x+z_1)I(x+z_2)$ across the image (all values of x).



Do this tabulation separately, for all three-point templates.



…

Then do the tabulation for all four-point templates.

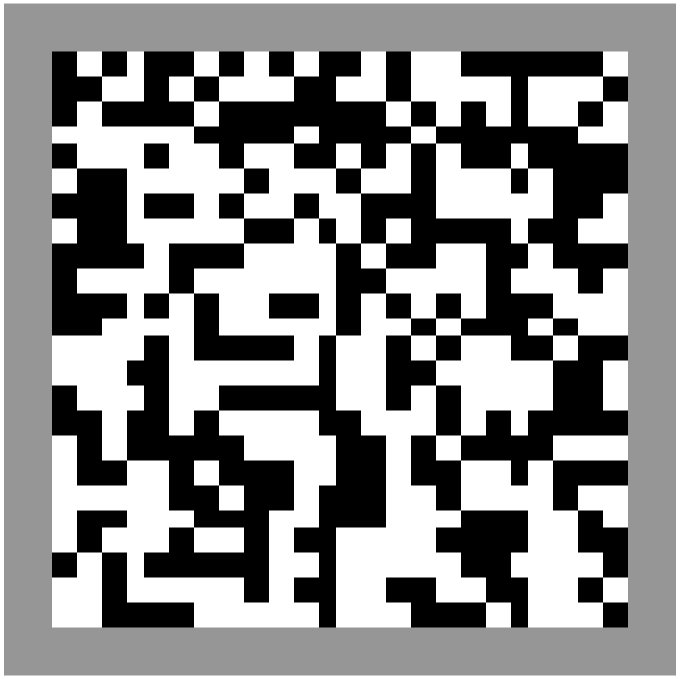
⋮

So what does the visual system do?

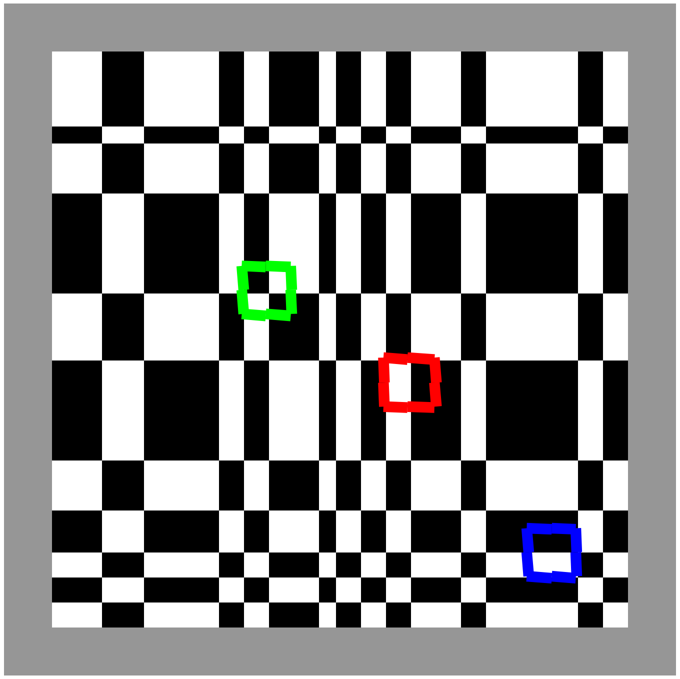
- Compute all of them?
- Compute the ones that are easy?
- Compute the ones that are helpful?

To answer this, we need to dissect the statistical properties of natural scenes

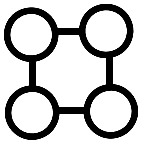
We can create images that contain only one kind of multipoint correlation



random

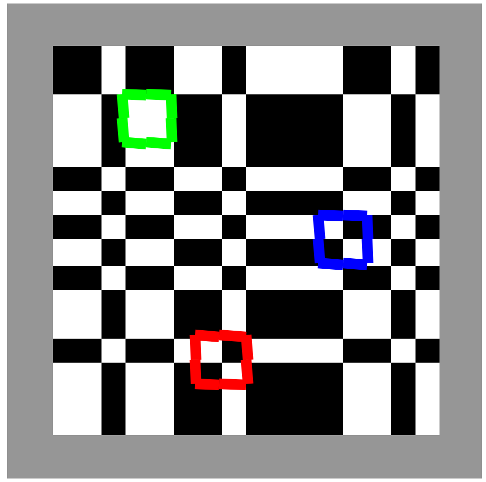


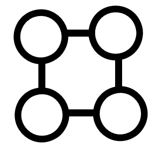
one kind of four-point correlation



with white=+1,black=-1, product is always +1

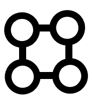
This generalizes...



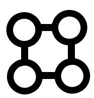


with white=+1,black=-1,
product is always +1

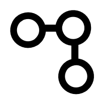
other template shapes and parities



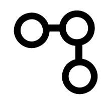
product is always +1



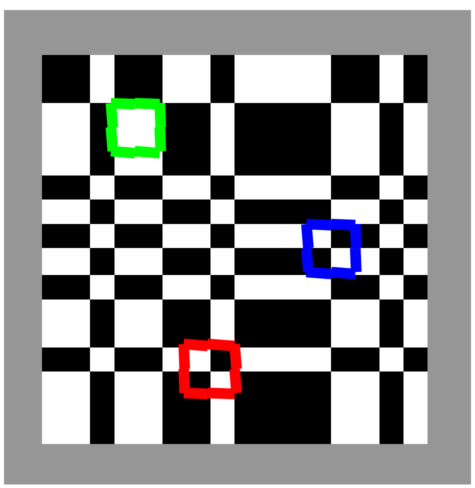
product is always -1



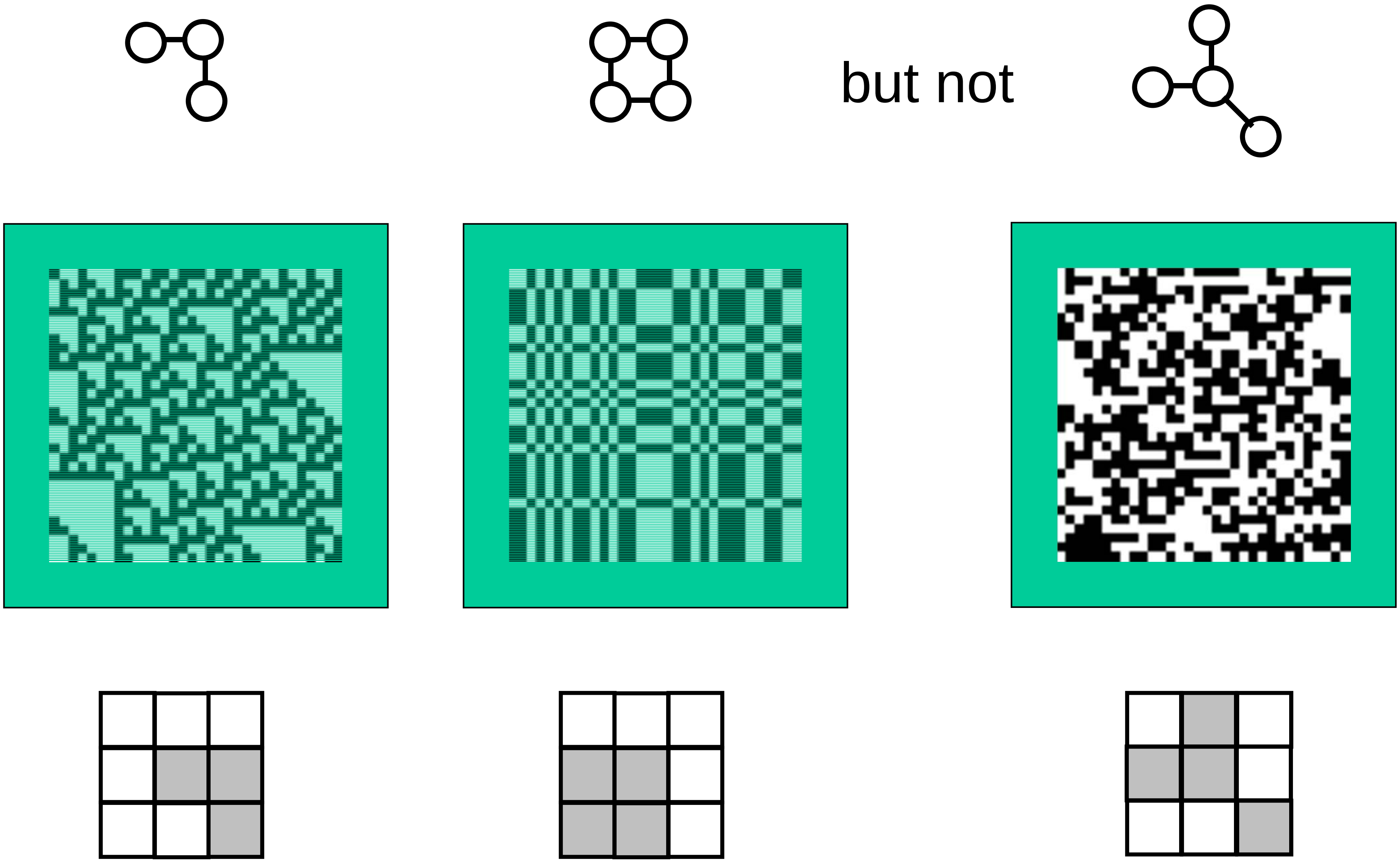
product is always +1



product is always -1



But only some multipoint correlations are visually salient



These simple demos show:

- Natural images contain multipoint correlations
- They carry visual features (edges, corners)
- Multipoint correlations can be visually salient, even in highly un-natural images
- Salience is selective: some multipoint correlations are salient, others are less so

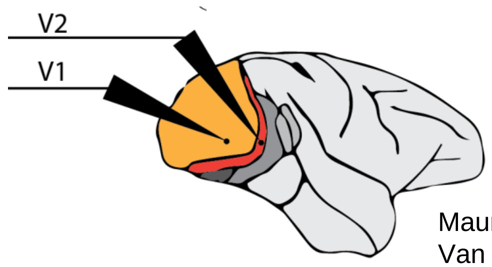
Where are these correlations extracted?

Does selective visual salience of multipoint correlations correspond to the properties of natural images?

Where do the computations take place?

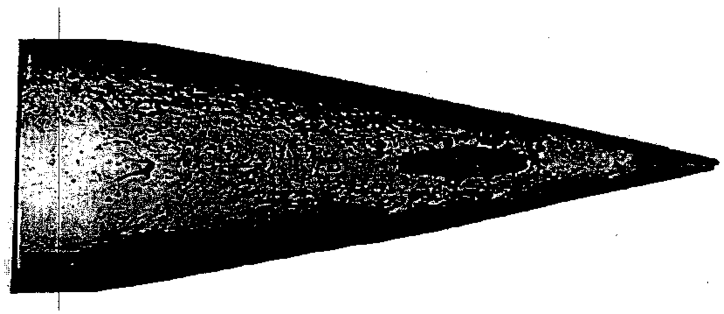
Y. Yu, A. Schmid

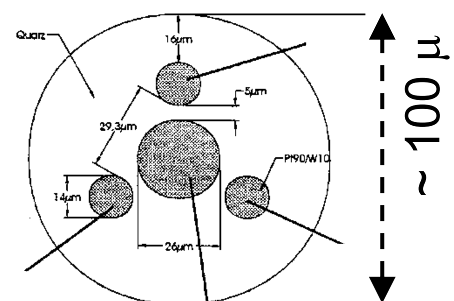
- Recordings from macaque primary visual cortex (V1) and secondary visual cortex (V2)



Maunsell and
Van Essen 1983

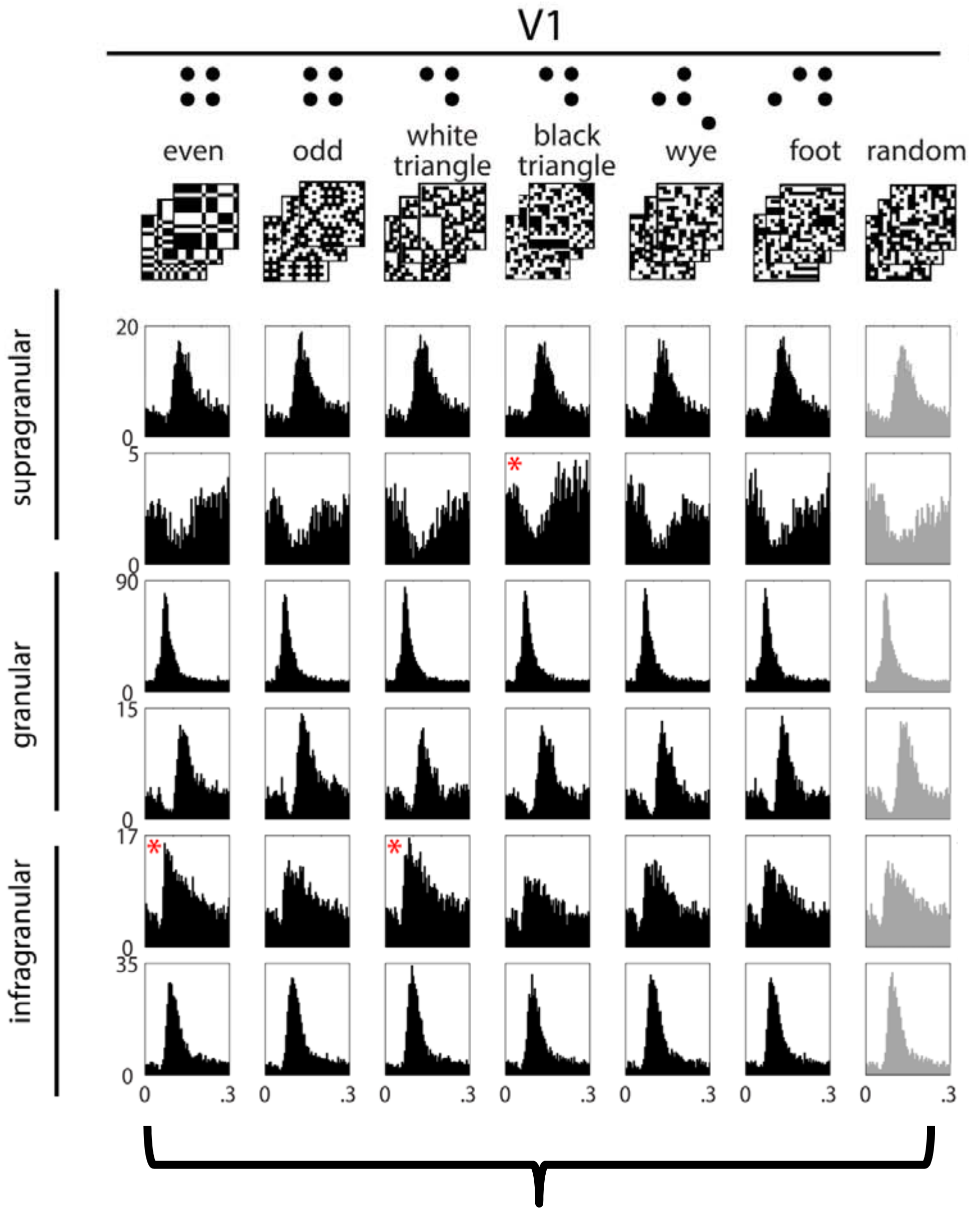
  - anesthetized (propofol, sufentanil), paralyzed
  - 6 tetrodes (24 contacts), with spike sorting





  - lesions and histology for laminar localization

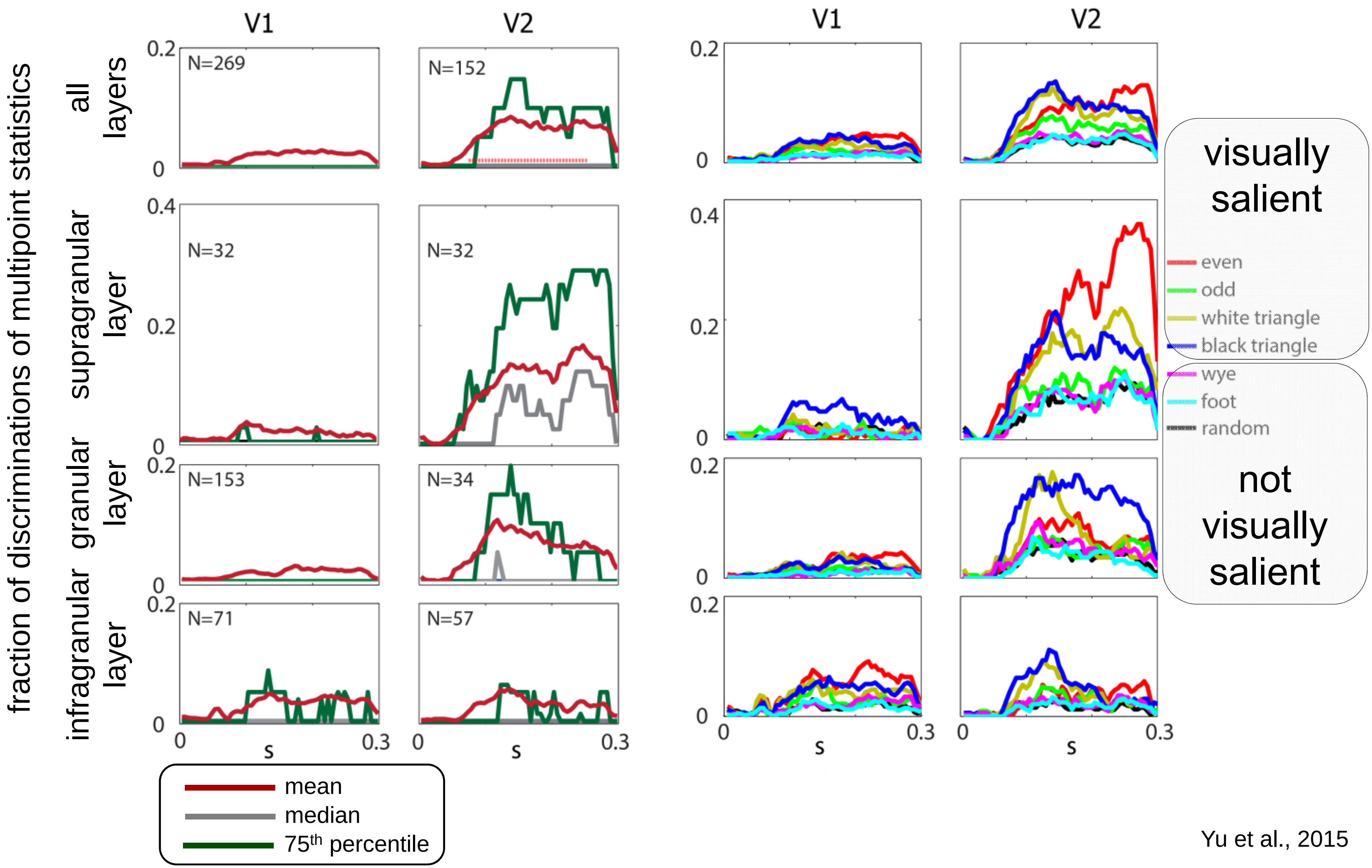
Single-neuron responses in V1 and V2



V1: responses usually independent of multipoint statistics

Yu et al., 2015

V2 neurons are sensitive to visually salient multipoint correlations



Yu et al., 2015

So far:

- Natural images contain multipoint correlations
- They carry visual features (edges, corners)
- Multipoint correlations can be visually salient, even in highly un-natural images
- Salience is selective: some multipoint correlations are salient, others are less so

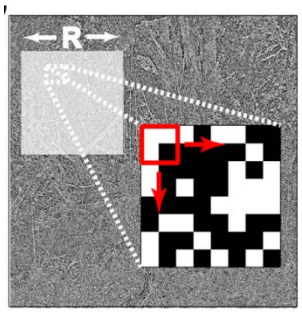
Where are multipoint correlations extracted?

Intracortically, mostly within the circuitry of V2

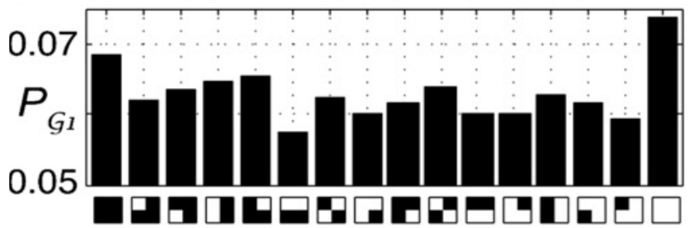
Does selective visual salience of multipoint correlations correspond to the properties of natural images?

Determining informativeness of multi-point correlations in natural images

natural image

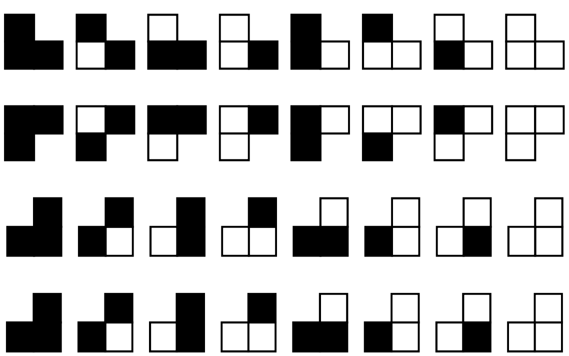


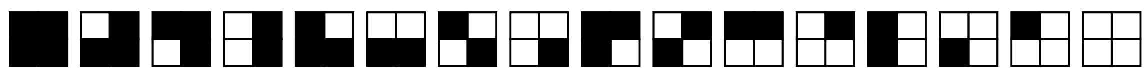
determine probability
of each 4-check block



Can this distribution be guessed from a simpler measurement?

If you know the 3-check probabilities, *can you guess the 4-check probabilities?*





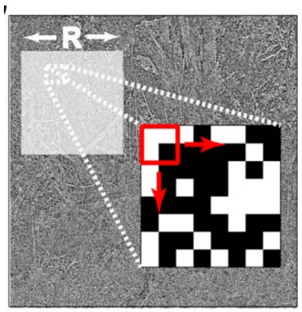
To quantify this:

1) Guess 4-check distribution that is are *as random as possible*, given the 3-check probabilities
2) If the 4-check distribution differs, *how much information is necessary to specify it?*

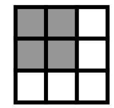
Tkačik, Prentice, Victor, and Balasubramanian, 2010

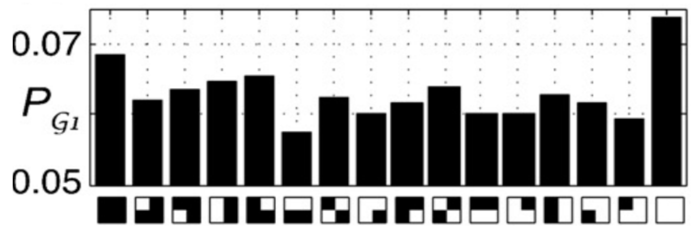
What kinds of four-point correlations are informative about natural images?

decorrelated natural image



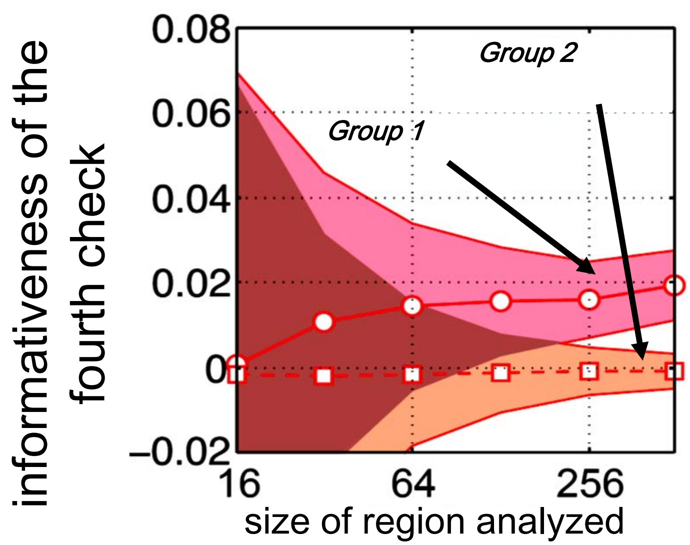
determine probability of each 4-check block





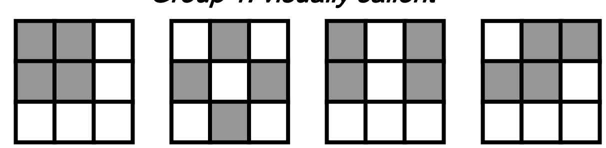
determine informativeness of the fourth check

repeat for each template



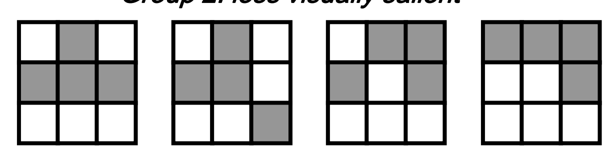
repeat for each region size

Group 1: visually salient



informative

Group 2: less visually salient



uninformative

So far:

- Natural images contain multipoint correlations
- They carry visual features (edges, corners)
- Multipoint correlations can be visually salient, even in highly un-natural images
- Salience is selective: some multipoint correlations are salient, others are less so

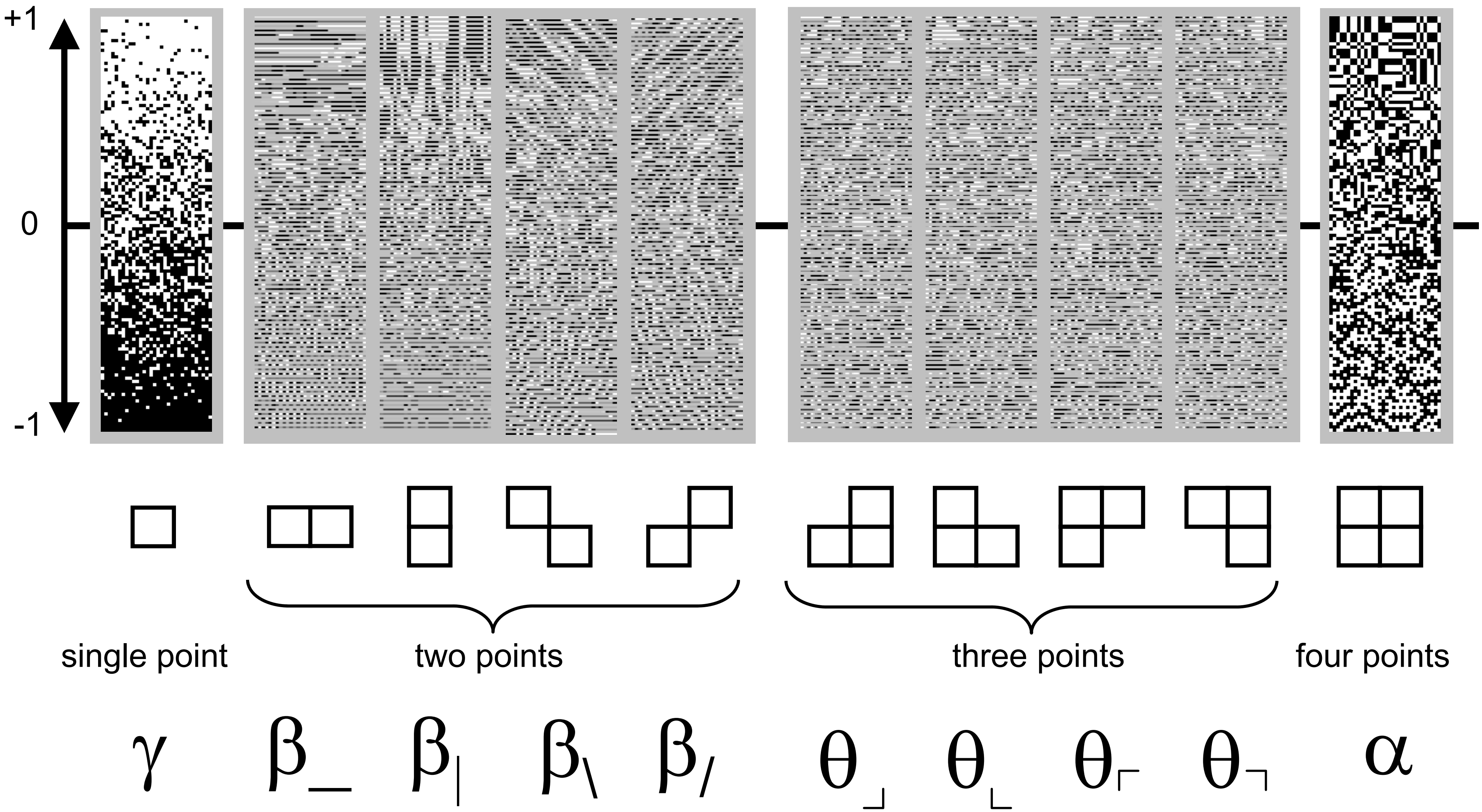
Where are multipoint correlations extracted?

Intracortically, mostly within the circuitry of V2

Does selective visual salience of multipoint correlations correspond to the properties of natural images?

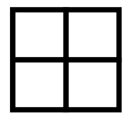
Yes. And we can focus on correlations within a 2x2 block.

With artificial images, we can separately manipulate *all* correlations within a 2x2 block



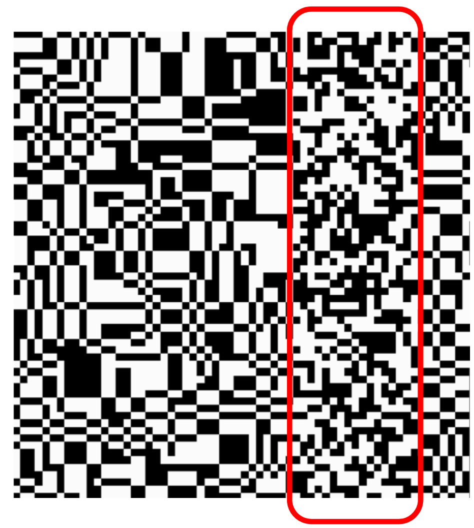
Making perceptual measurements

target: structured
background: random



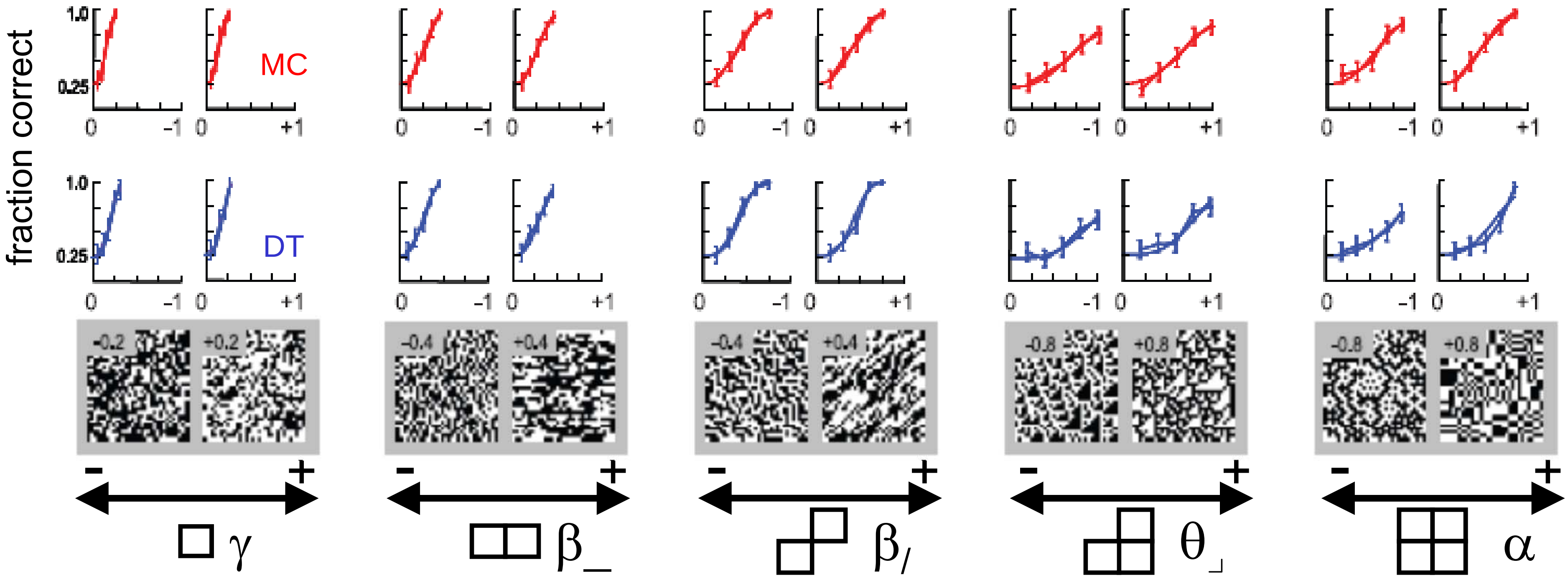
α=0.85

target: random
background: structured



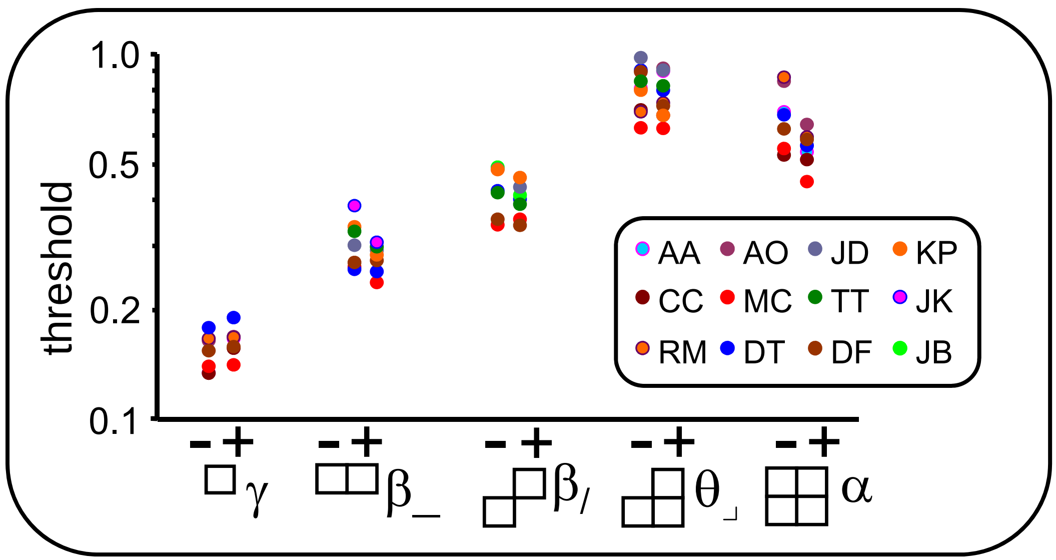
Where is the distinguished segment?
(120 ms viewing time)

Sensitivity is selective, and similar across observers

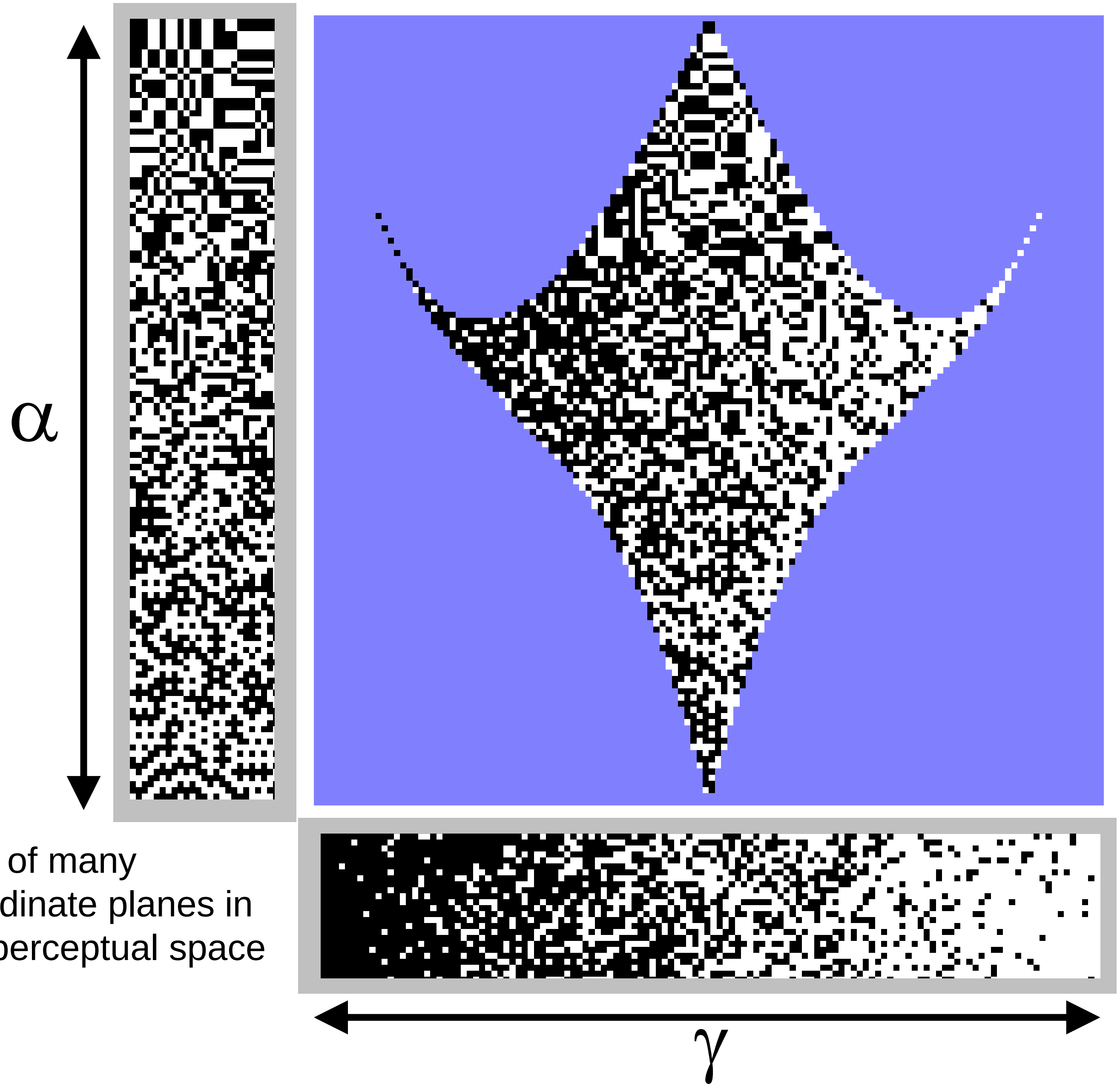


MC: ~10^4 hours of viewing experience

DT: ~10^1 hours of viewing experience

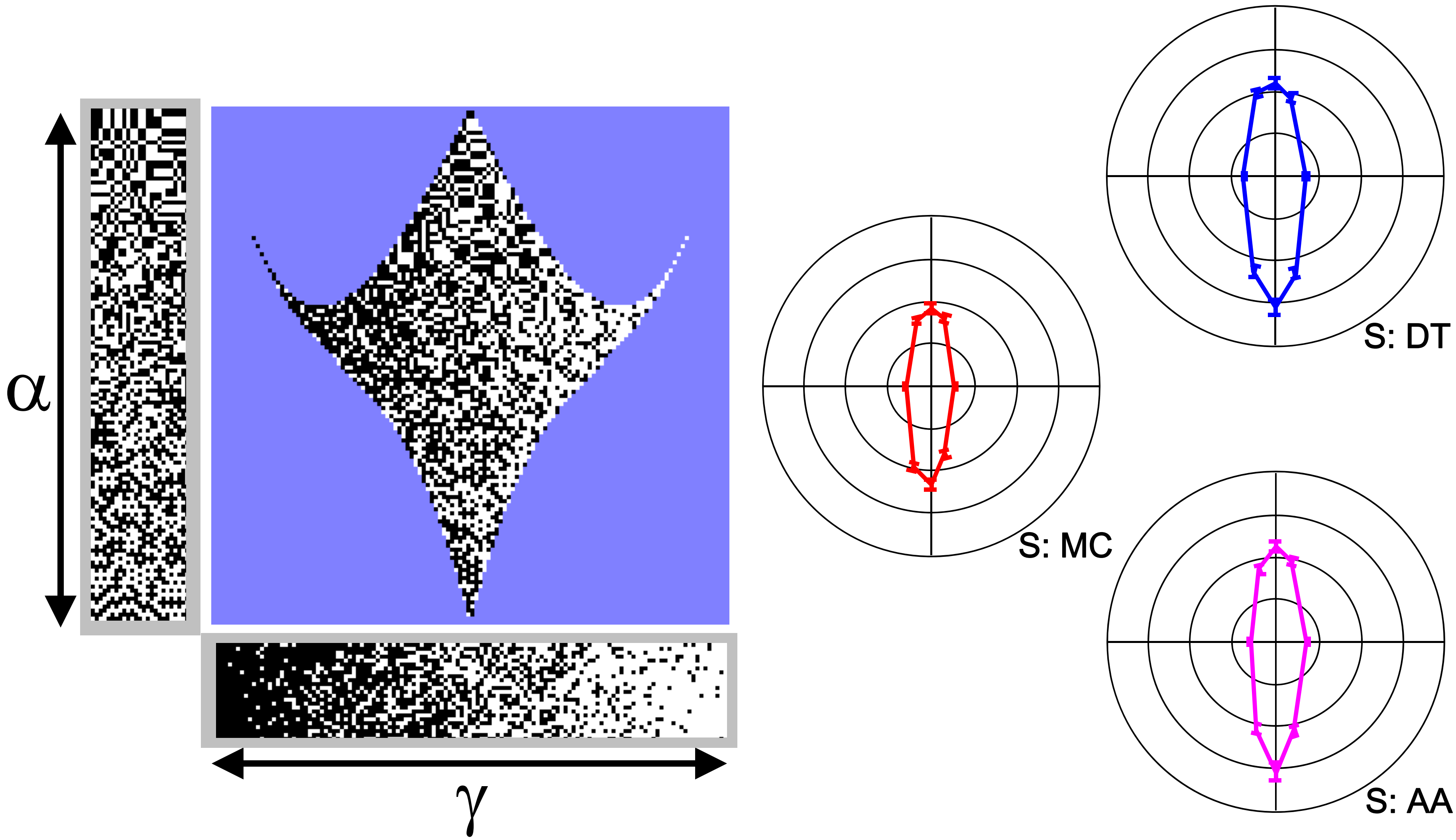


How do the coordinates interact perceptually?

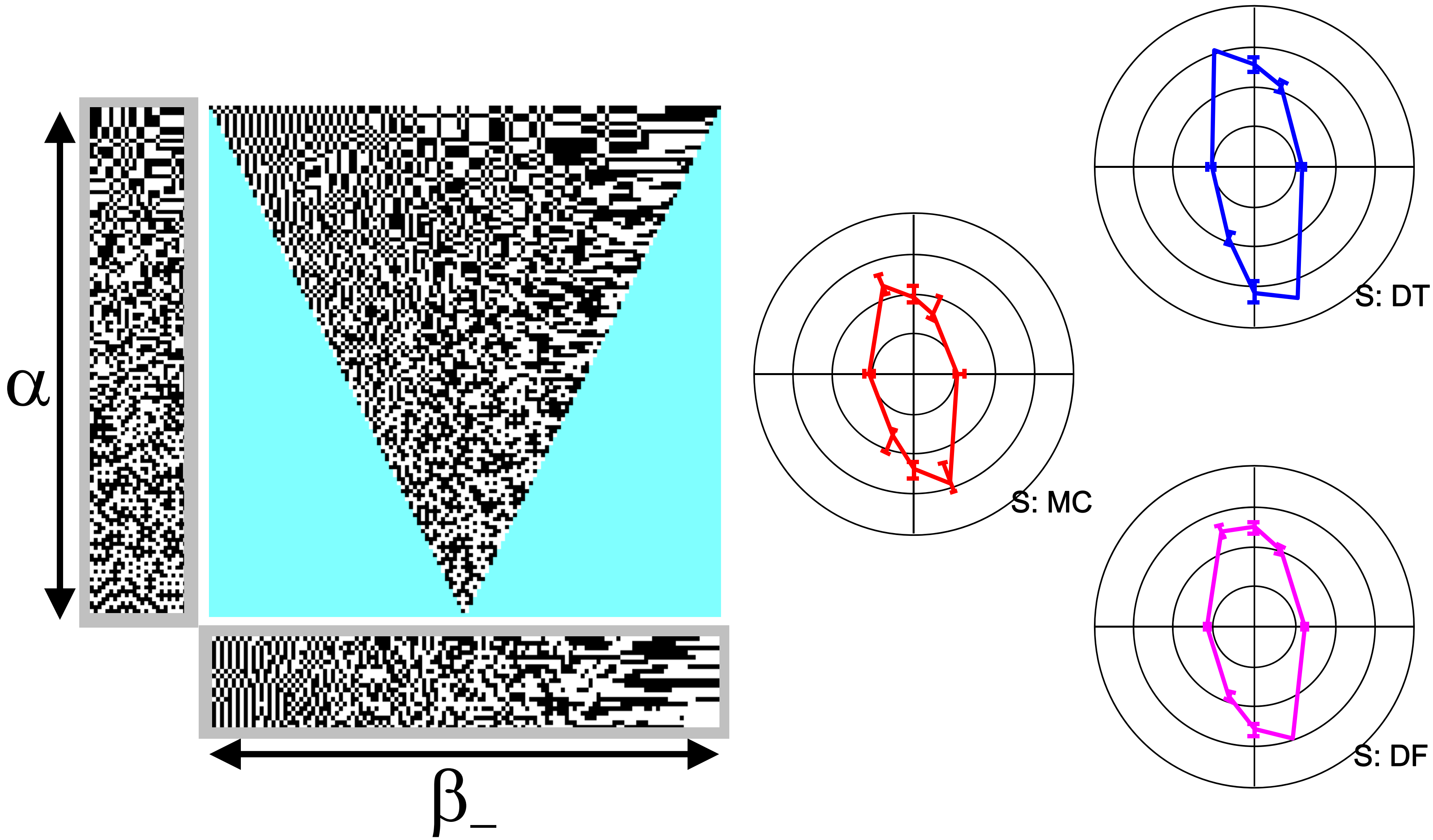


One of many coordinate planes in the perceptual space

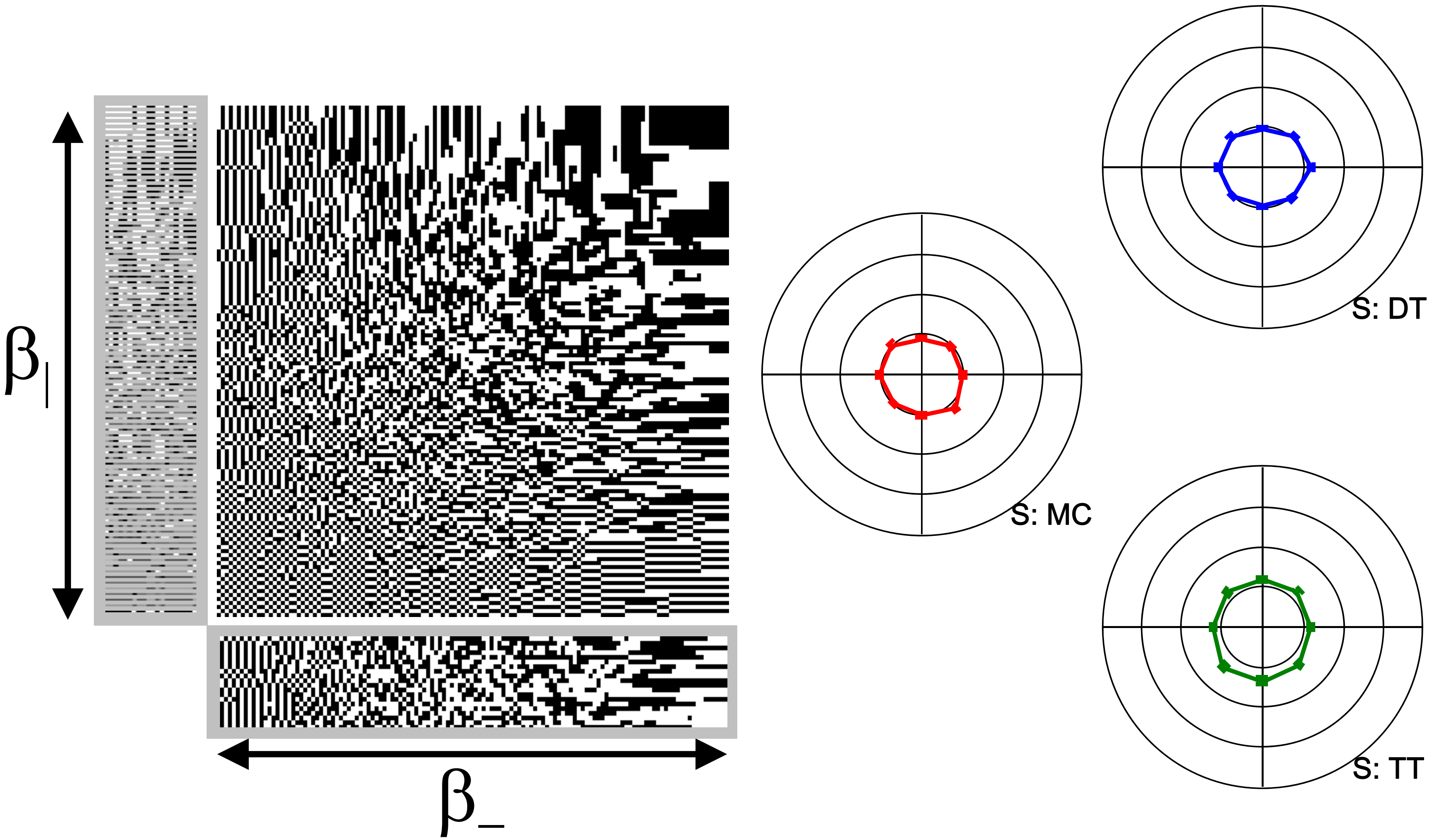
Pairwise interactions



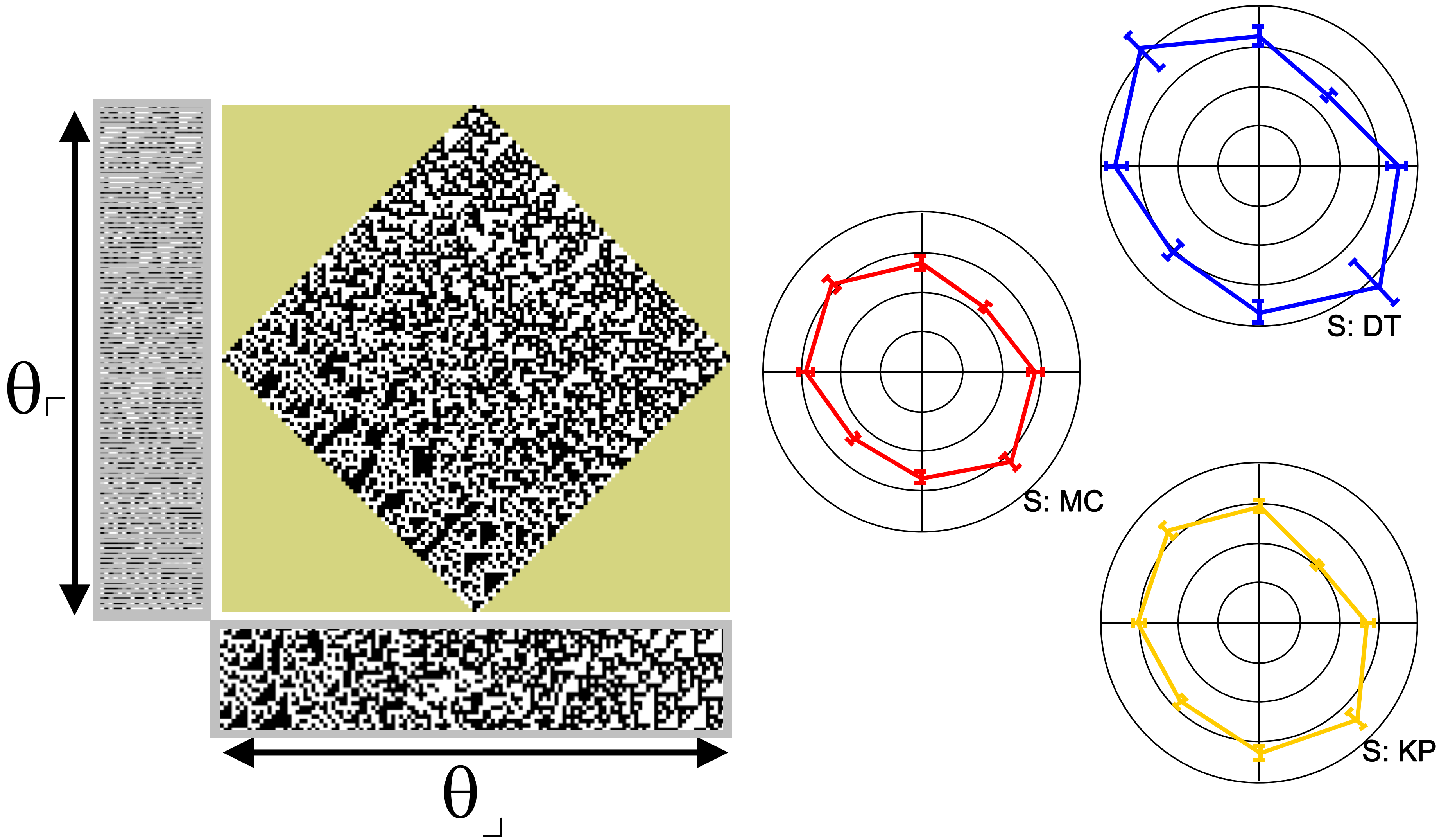
Pairwise interactions



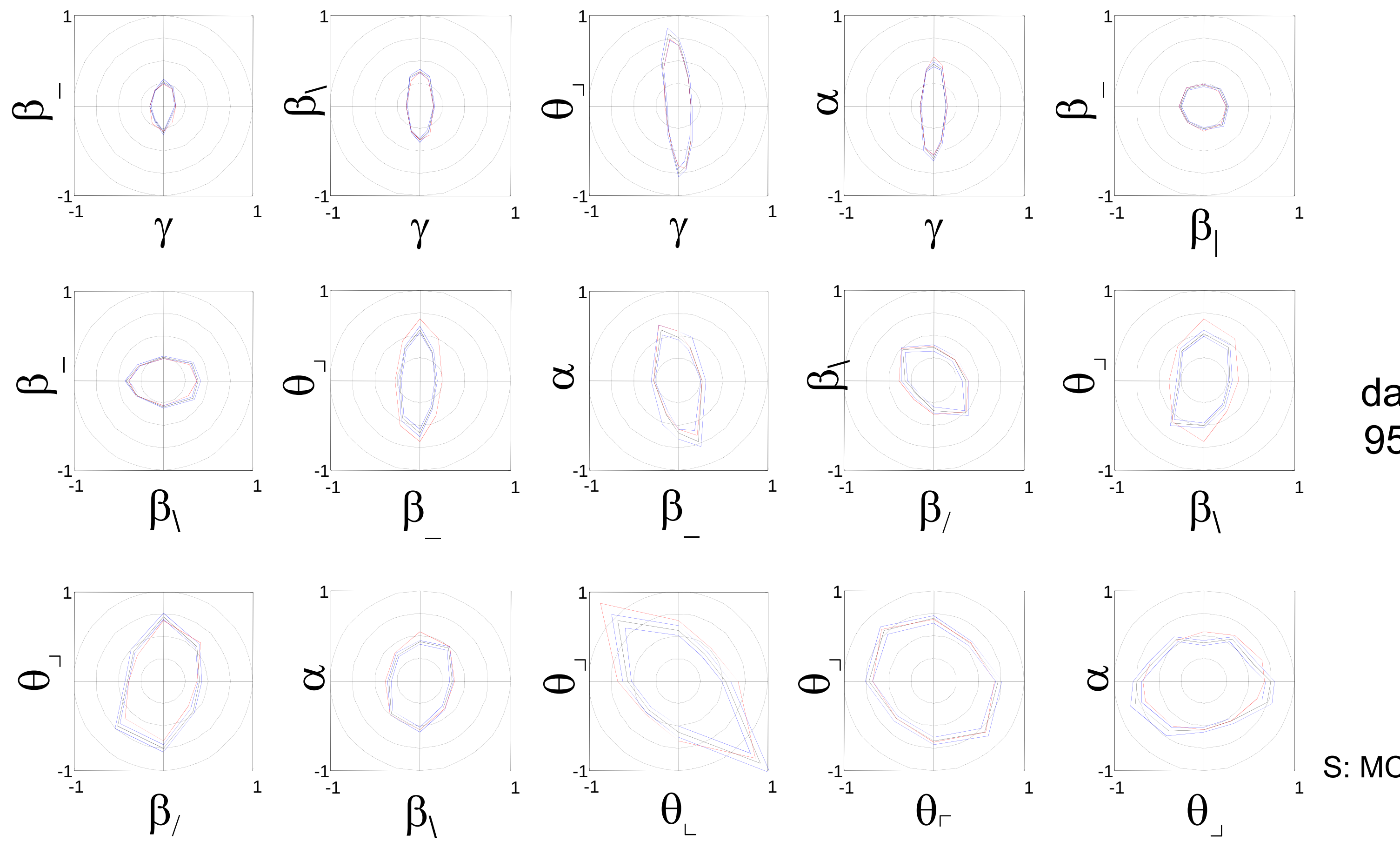
Pairwise interactions

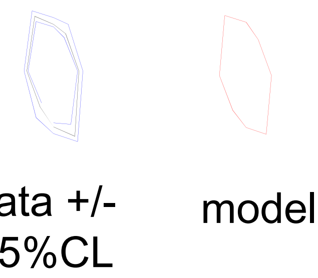


Pairwise interactions



A simple model accounts for perceptual thresholds





>95% of variance explained

In each plane, isodiscrimination contours are approximately elliptical.

$$\text{Perceptual distance} = \sqrt{\sum_{i,j} Q_{i,j} c_i c_j}$$

c_i: the coordinates
$Q_{i,j}$: the metric

Summarizing:

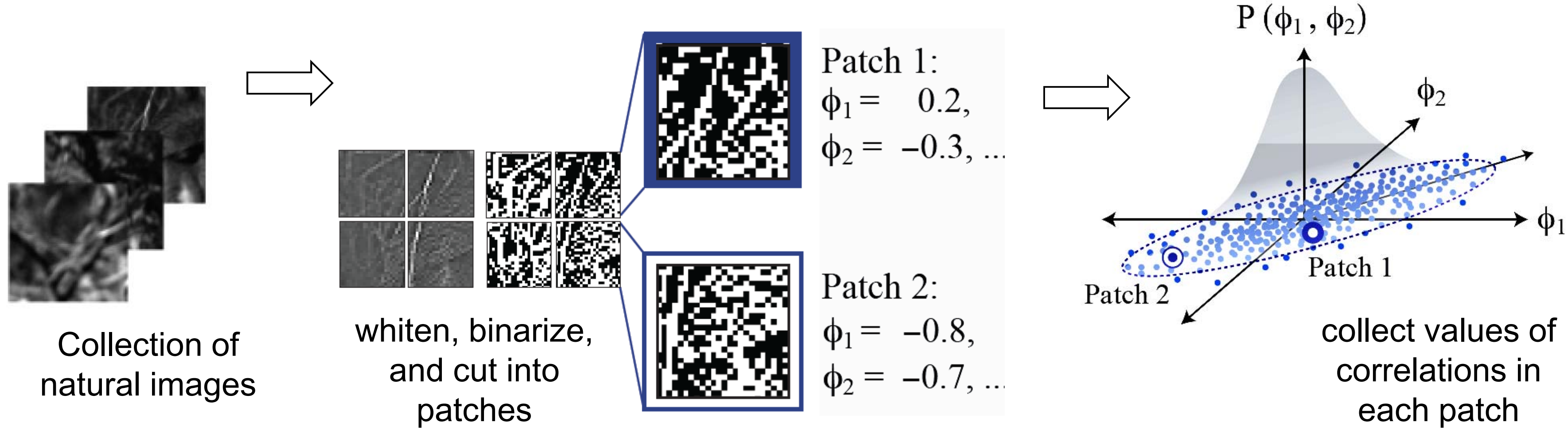
- There is a stereotyped pattern of selective sensitivity to informative correlations, and uninformative correlations are ignored

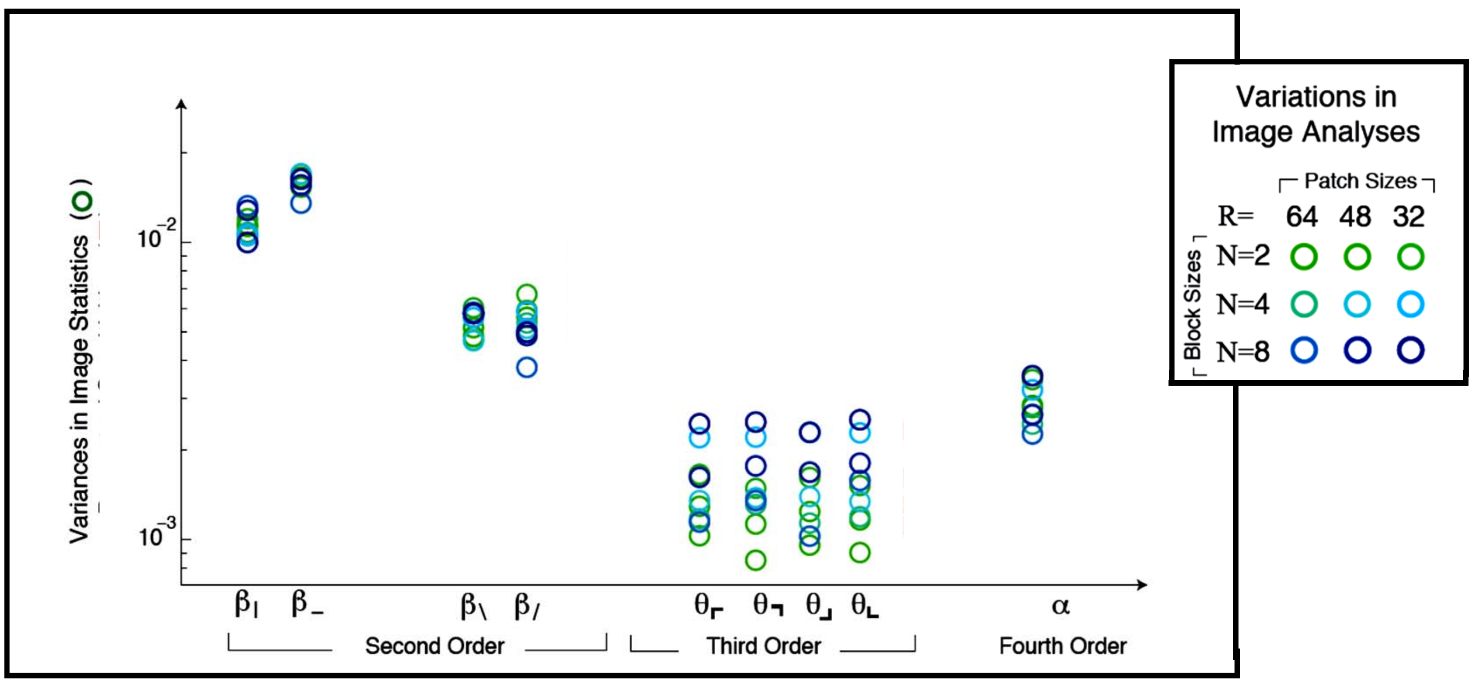
- These correlations are extracted in visual cortex, mostly in V2

- A simple phenomenological model accounts for perceptual sensitivities

- Why this particular pattern of sensitivities?

Sensitivity to each coordinate is matched to its range of variation in natural images





(Hermundstad, Briguglio, Conte, Victor, Balasubramanian, Tkačik, 2014)

Conclusions: the efficient coding principle, revisited

- In the sensory periphery, redundancy is removed. This means that fewer resources are devoted to signal components with predictably greater range.

- In cortex, sensitivity to image features covaries with their occurrence in natural images. Thus, the opposite occurs: more resources are devoted to signal components with greater range.

- Why this difference? Different goals, different coding regimes:
  - Peripherally: efficiency in the face of a transmission bottleneck
  - Centrally: making inferences in the face of sampling noise


559
Thank you!