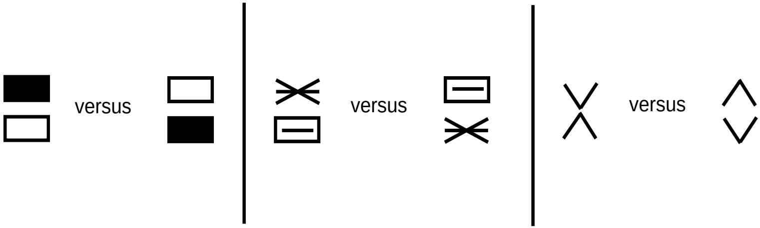
Perceptual Distance and Visual Search

Data Science - Visual Neuroscience Lecture 1

Physical distance versus perceptual distance

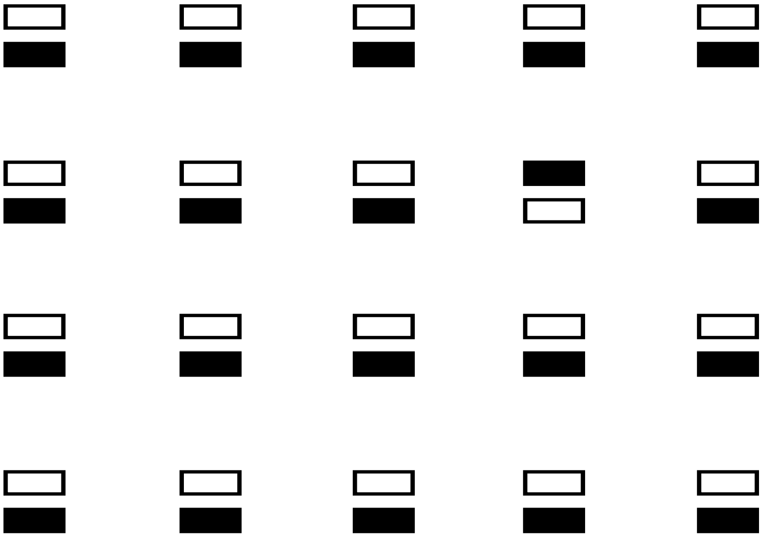
- Why are we (as yet) better at vision than machines?
- Sophisticated representation of objects.
  "Pixel distance" very different from perceptual distance
- In this module: Study experimental data that attempts to quantify perceptual distance

Measuring perceptual distance

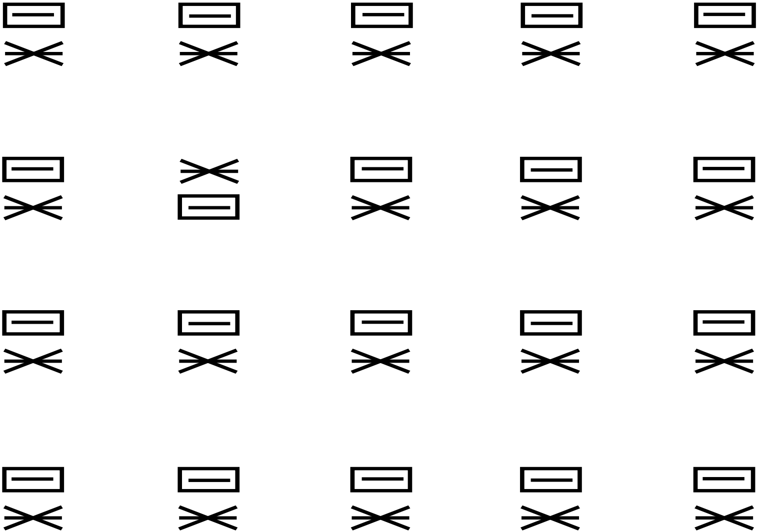


Ideas?

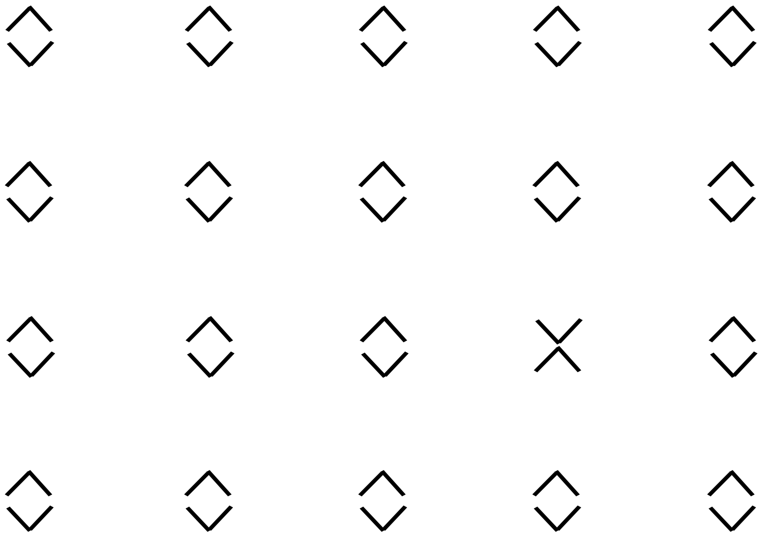
Find the odd image - 1



Find the odd image - 2



Find the odd image - 3

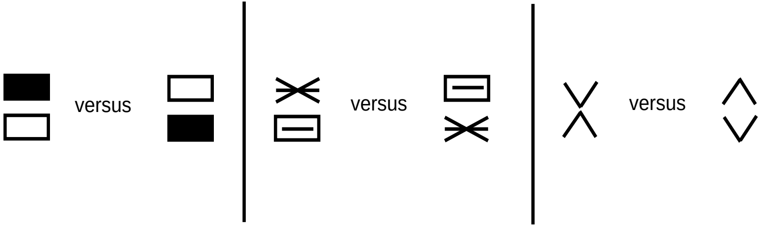


A measure of perceptual distance

Hypothesis

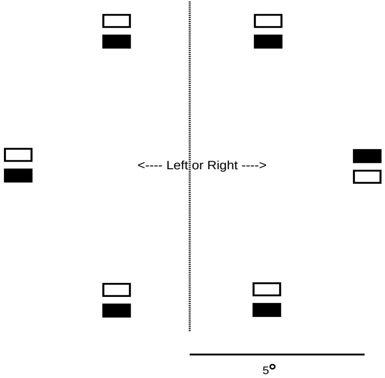
Visual search performance depends on the perceptual distance between the two images. Closer the two images in perceptual distance, the longer it takes to identify the oddball image. More specifically:

$$\text{Proposed Perceptual Distance} \propto \frac{1}{(\text{Search Time})^k}?$$



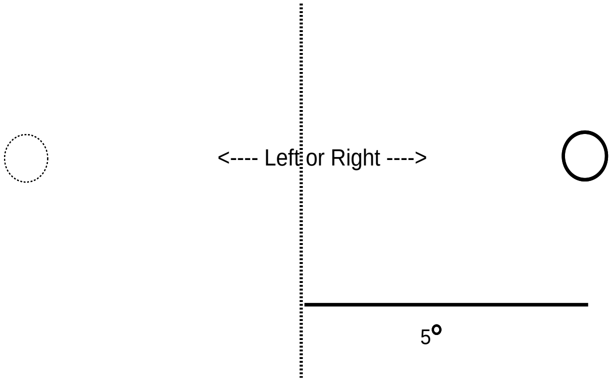
A reaction time study on humans (Arun and Olson 2010)

- Study conducted on six subjects
- Identify the location of the oddball and hit a key to tell left or right



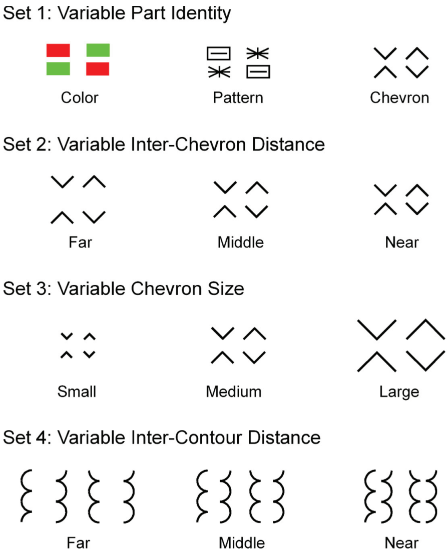
- Image displayed until reaction (which if correct, valid trial), or until 5 seconds (aborted)
  $\mathrm{RT}(i, j)$ = average reaction time
  Data averaged over both *oddball i among distracters j*

Baseline reaction time



- RT_b = baseline reaction time
- $s(i,j) = \mathrm{RT}(i,j) - \mathrm{RT}_b$
- Perceptual distance between i and j is $\propto 1/s(i,j)$
- $\mathrm{RT}_b = 328ms$.

Image pairs on which search time data was collected (Sripati and Olson 2010)



A direct view into the brain of rhesus macaques

- Try to nail the responses in the brain, and see how different they are.

Where to measure?

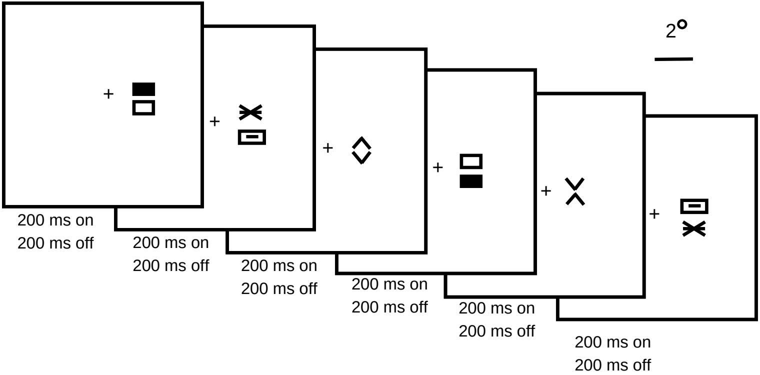
- The case for measuring in IT (Sripati and Olson 2010)
- Neurons in IT, unlike those in low-order visual areas, have receptive fields large enough to capture an entire image.
- Sensitive to global arrangement of elements within the image.
- Studies indicate that population activity in IT discriminates some images better than others. Studies also indicate that if a pair is well-discriminated by population activity in IT, then humans tend to characterise them as dissimilar.
- Perhaps then population activity in IT should predict human search efficiency.

Experimental procedure on rhesus macaques and recording (done by Arun and Olson)

- Cleared by CMU institutional animal care and use committee
- Two macaques were surgically fitted with:
  - a cranial implant for neuronal activity recording;
  - a scleral search coil for recording eye movements.
- Data was collected over several days. Before each day's experiment, an electrode was inserted so that the tip was 1 cm above the inferotemporal cortex.
- The electrodes were pushed, reproducably, along tracks forming a square grid with 1 mm spacing.
- Neuronal activity was recorded. Individual neurons' action potentials then isolated using a commercially available tool (Plexon).

A direct view into the brain of rhesus macaques

- Two macaques were trained to fixate on the + while a series of stimuli appeared one after another.



- Images were randomly interleaved. Neuronal activity recorded (inferotemporal cortex) over several 2 second rounds.

The neuronal data

- Inferotemporal cortex - gross object features emerge here
- Firing rates of $N = 174$ neurons in response to these six images
- Data collected in a similar manner for a total of 24 images
- For each image i, the neuronal response is summarized by the firing rate vector $(\lambda^i(n), 1 \leq n \leq N)$.

$$\text{Image } i \mapsto \lambda^i = \begin{pmatrix} \lambda^i(1) \\ \lambda^i(2) \\ \vdots \\ \lambda^i(N) \end{pmatrix}$$

The main question

- For the pair (i, j), perceptual distance ought to be a function of how "different" λ^i and λ^j are.
- What function?
- How does it relate to reaction time?

A model grounded in a theory

- What would the prefrontal cortex do if it got observations from the human analogue of the inferotemporal cortex and could control the eye?

Aspects of search

- Find in the shortest possible time. Cost = delay.
- Local focus. You could choose where you wanted to look next.
- Two types of pictures. But you didn't "know" either. Learnt which is which on the fly.
- But you learnt just enough to tell a picture in location 1 was same as or different from the picture in location 2.
- When you changed focus, you often chose a location nearer to the current location.
- You waited until you were sure about the oddball location.

A model for search - sequential hypothesis testing

- Hypothesis $h = (\ell, i, j)$: The oddball location is ℓ and its type i among distracters j. Ground truth.
- Divide time into slots.
- Control: Given observations and decisions in all previous slots (history),
  - decide to stop and declare the oddball, or
  - decide to continue, and direct the eye to focus on location b, one of the six locations.
- Observation: If the object in location b is k, then N Poisson point processes with rates $(\lambda^k(n), 1 \leq n \leq N)$.
- Policy π: For each time slot, given history, a prescription for action.
  To stop or not to stop?
  If continue, where to look?
  If stop, what to decide?

Performance

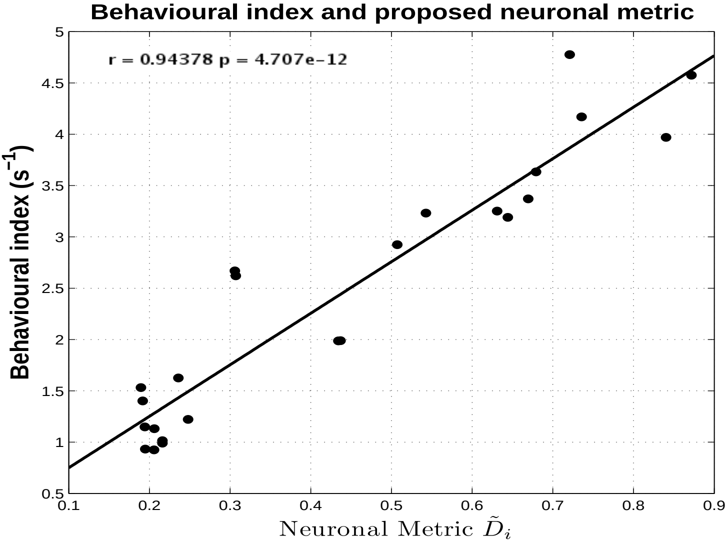
- For each ground truth h, your policy shall make an error with probability at most ε.

- What is the expected time to stop for a fixed positive ε?

- The average search delay is the average over all hypotheses h with i as oddball and j as distracter.

- What function of λ^i and λ^j?
  Difficult to evaluate. We will do some asymptotics as $\varepsilon \to 0$ to get the following.

We will process data to get this correlation plot



What we will learn in this module

- Hypothesis testing
- Hypothesis testing with a stopping criterion
- Data processing inequality, and relative entropy
- A brief view into asymptotic analysis
- Testing for a distribution - Kolmogorov-Smirnoff test
- ANOVA and variants