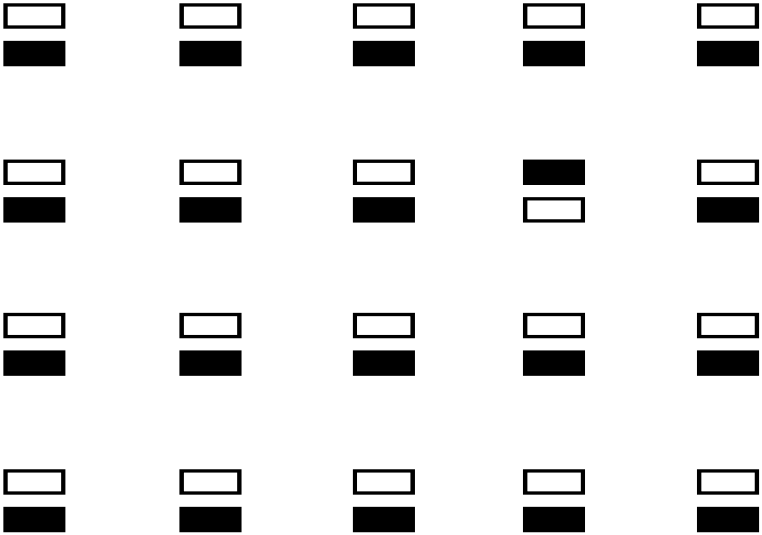
Perceptual Distance and Visual Search

Statistics 1, ISI Bangalore Centre

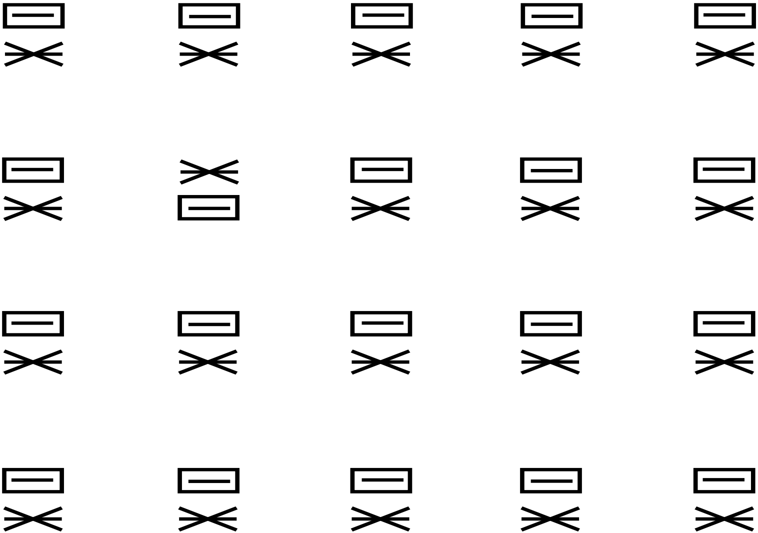
Rajesh Sundaresan
Indian Institute of Science

17 September 2019

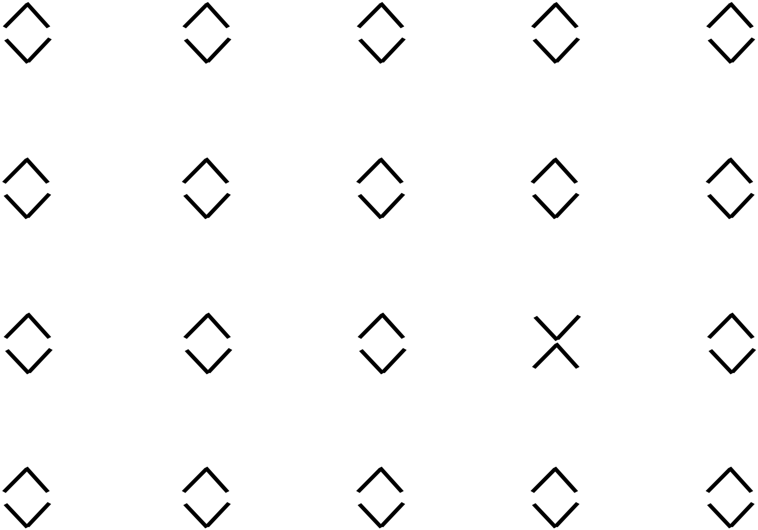
Find the odd image - 1



Find the odd image - 2



Find the odd image - 3

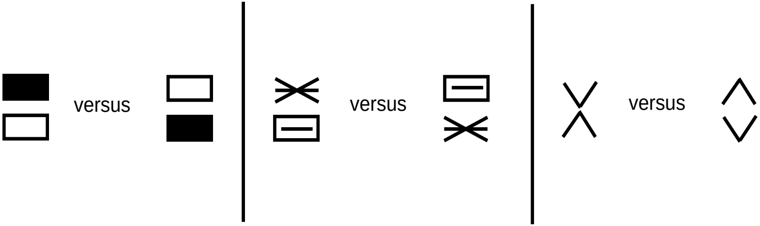


Similarity and search time

- The more similar the two images, the longer the search time
- The more the "distance" between the two images, the shorter the search time.

Neuroscientists perform such experiments

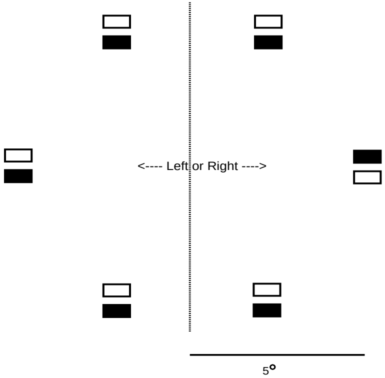
- "Perceptual distance" $= \frac{C}{(\text{search time})}$.



- Helps them interpret how objects are represented in the brain, capture important features.

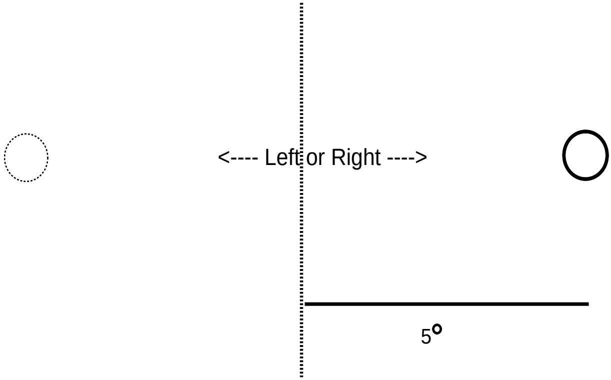
A reaction time study on humans (Arun and Olson 2010)

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



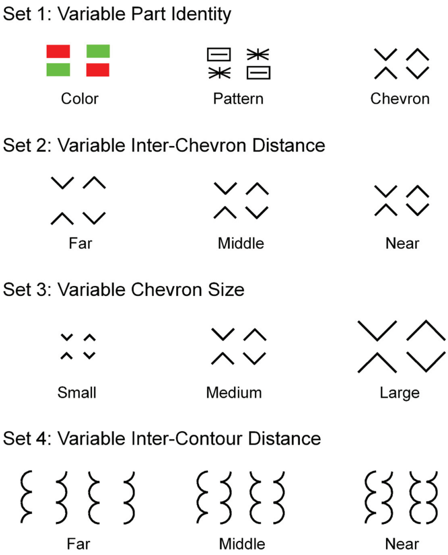
- Image displayed until reaction (which if correct, valid trial), or until 5 seconds (aborted)
  RT_{ij} = average reaction time
  Data averaged over both *oddball i among distracters j*

Baseline reaction time



- RT_b = baseline reaction time
- Search time $s_{ij} = RT_{ij} - RT_b$
- Perceptual distance between i and j is taken to be $\frac{C}{s_{ij}}$
- $RT_b = 328ms$.

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

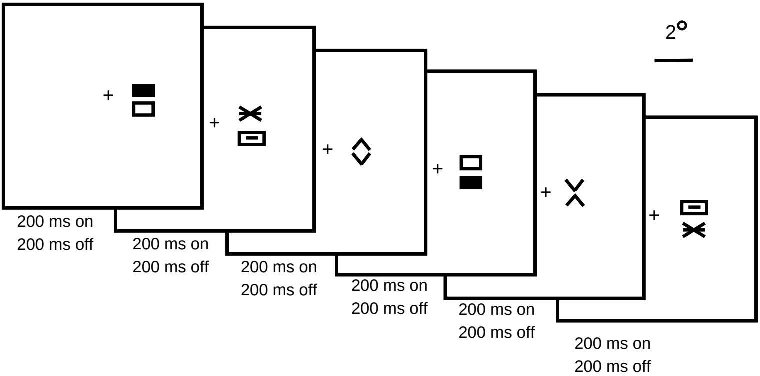


A direct view into the brain of rhesus macaques

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

A direct view into the brain of rhesus macaques

- Sripati and Olson trained two macaques to fixate on the + and made a series of stimuli appeared one after another.



- Images were randomly interleaved. Neuronal activity was recorded (in the 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 $\text{dist}(\lambda^i, \lambda^j)$ function?
- How does it relate to reaction time? Is it $\frac{C}{\text{dist}(\lambda^i, \lambda^j)}$?

First distances that come to our minds

- Try Euclidean distance, or L^1:

$$L^1_{ij} = \|\lambda^i - \lambda^j\|_1 = \sum_n \left|\lambda^i(n) - \lambda^j(n)\right|$$

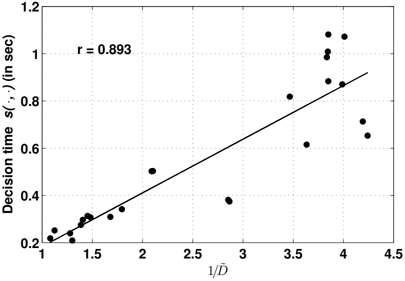
An alternative, via a model and a theory

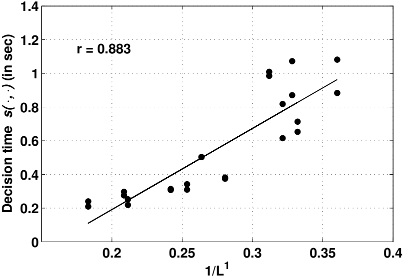
- If you look at object j, your neurons fire at rate λ^j.
- Perhaps firings are random variables $X = (X_n, 1 \leq n \leq N)$
- A good model: $X_n \sim Poisson(\lambda^j(n)T)$ over T time units.
- Let s_{ij} be the search time for identifying oddball i in a sea of j.
- An optimal search theory predicts, for a probability of error ε,

$$E[s_{ij}] \simeq \frac{\log(1/\varepsilon)}{\tilde{D}_{ij}}$$

$$\tilde{D}_{ij} = \sum_n \left[\lambda^i(n) \log \frac{\lambda^i(n)}{\lambda^j(n)} - \lambda^i(n) + \lambda^j(n)\right]$$

Compare s versus 1/dist: Which dist is better?





The ideal

- If the theory holds:

$$E[s_{ij}] = \frac{C}{dist_{ij}} \quad \iff \quad E[s_{ij} dist_{ij}] = C \quad \forall (i,j)$$

- This suggests the following hypothesis:

  Across the image pairs, the mean of the scaled search time (scaled by $dist_{ij}$) is a constant.

- Which dist ($\tilde{D}$ or L^1) comes closer to this ideal?

Data and a hypothesis test to help with comparison

- Six subjects, 12 showings of oddball i in a sea of j.
- $X_t^{(i,j)} = s_{ij,t}\tilde{D}_{ij}$, $l = 1, \ldots, 72$.

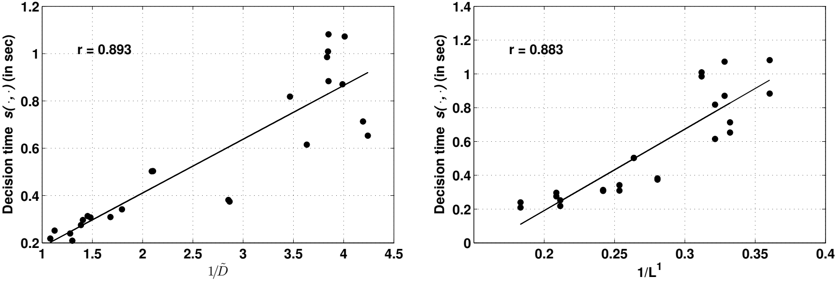
$g_1 = (i_1, j_1)$	$g_2 = (i_2, j_2)$	$\ldots$	$g_{24} = (i_{24}, j_{24})$
$X_1^{(g_1)}$	$X_1^{(g_2)}$	$\ldots$	$X_1^{(g_{24})}$
$\vdots$	$\vdots$		$\vdots$
$X_{72}^{(g_1)}$	$X_{72}^{(g_2)}$	$\ldots$	$X_{72}^{(g_{24})}$

- A hypothesis test:

  Null hypothesis: All groups have the same mean

  Alternative hypothesis: Not so

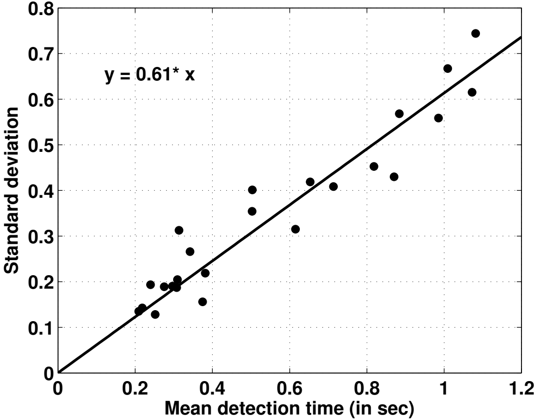
Understanding the data, which parametric family?



- Variance increases with mean
- Gamma distribution: $p_S(s) = \frac{\beta^\alpha}{\Gamma(\alpha)} s^{\alpha-1} e^{-\beta s}, \quad s > 0$
  Shape $\alpha > 0$, rate $\beta > 0$
- $E[S] = \frac{\alpha}{\beta}, \quad \sigma(S) = \frac{\sqrt{\alpha}}{\beta} = \frac{1}{\sqrt{\alpha}} E[S].$

Do the groups have a common shape?

Is $\sigma(S^{(g)})$ linear in $E[S^{(g)}]$ across $g = 1, \ldots, 24$?



$$\frac{1}{\sqrt{\alpha}} = 0.61 \quad \Longrightarrow \quad \alpha = \left(\frac{1}{0.61}\right)^2 = 2.7$$

Equality of means test, look at scaled search times

- $X_1^{(g)}, \ldots, X_{72}^{(g)}$ iid Gamma(α, β_g), $\quad g = 1, \ldots, G$.
- $g = 1, \ldots G$ stands for the different pairs of images.
  $G = 24$ pairs of images in all.
- Null hypothesis: $\beta_1 = \beta_2 = \ldots = \beta_G$
- Alternative hypothesis: Not all equal.

Equality of means test, look at scaled search times

- The test when α is known: Reject null hypothesis if:

$$\alpha \log \frac{\text{AM across groups}}{\text{GM across groups}} > \tau$$

  for a suitable threshold τ.

- Here, don't use it for accepting/rejecting null hypothesis. Instead use it for comparison.

- Compute AM/GM for dist = $\tilde{D}$ and dist = L^1.

diff	α log(AM/GM)
$\tilde{D}$	0.0600
L^1	0.2061

- Conclusion: $\tilde{D}$ provides a better 'clustering' than L^1.

Summary

- We looked at a visual search problem.
- Neuroscientists are interested in this problem because they want to capture a notion of ‘perceptual distance’ between objects.
- Perceptual Distance $= \frac{C}{(\text{search time})}$
- Another measure of distance – how different are the responses to the images?
- What function of the firing rate vectors is an appropriate distance?
  - Search times are indeed C divided by some dist.
  - A (relative entropic) distance $\tilde{D}$ explains observed search times better than L^1.
  - Conclusion obtained by verifying that $s_{ij}\tilde{D}_{ij}$ is more clustered than $s_{ij}L^1_{ij}$.
- Data set link: `https://ece.iisc.ac.in/~rajeshs/E0259/02_data_visual_neuroscience.htm`