R V College of Engineering *Dec 24, 2019*

Search in Research: The Importance of the Theory of Probability in Real-Life

Karthik P N

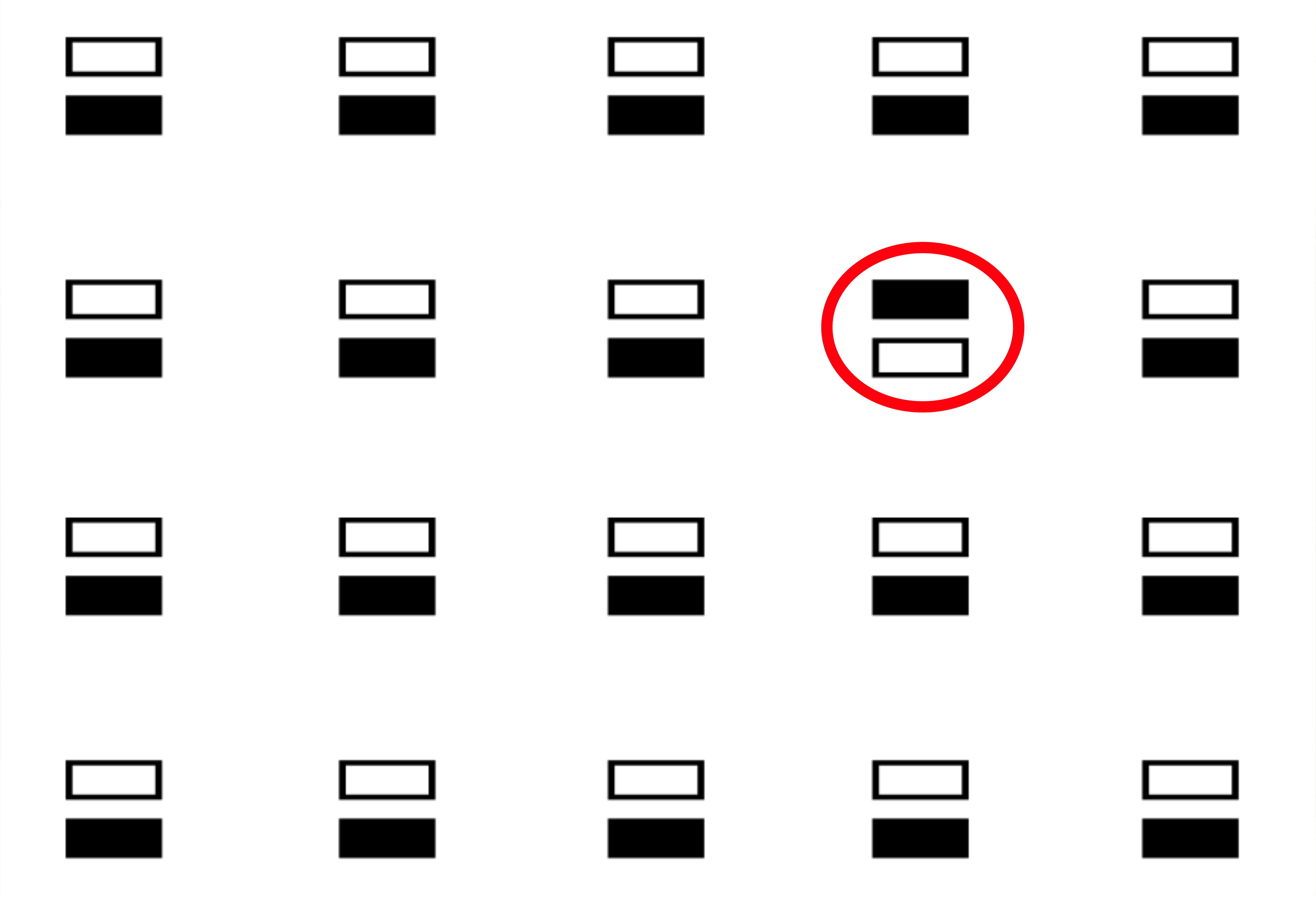
Department of ECE, IISc

Joint Work with Dr. Rajesh Sundaresan

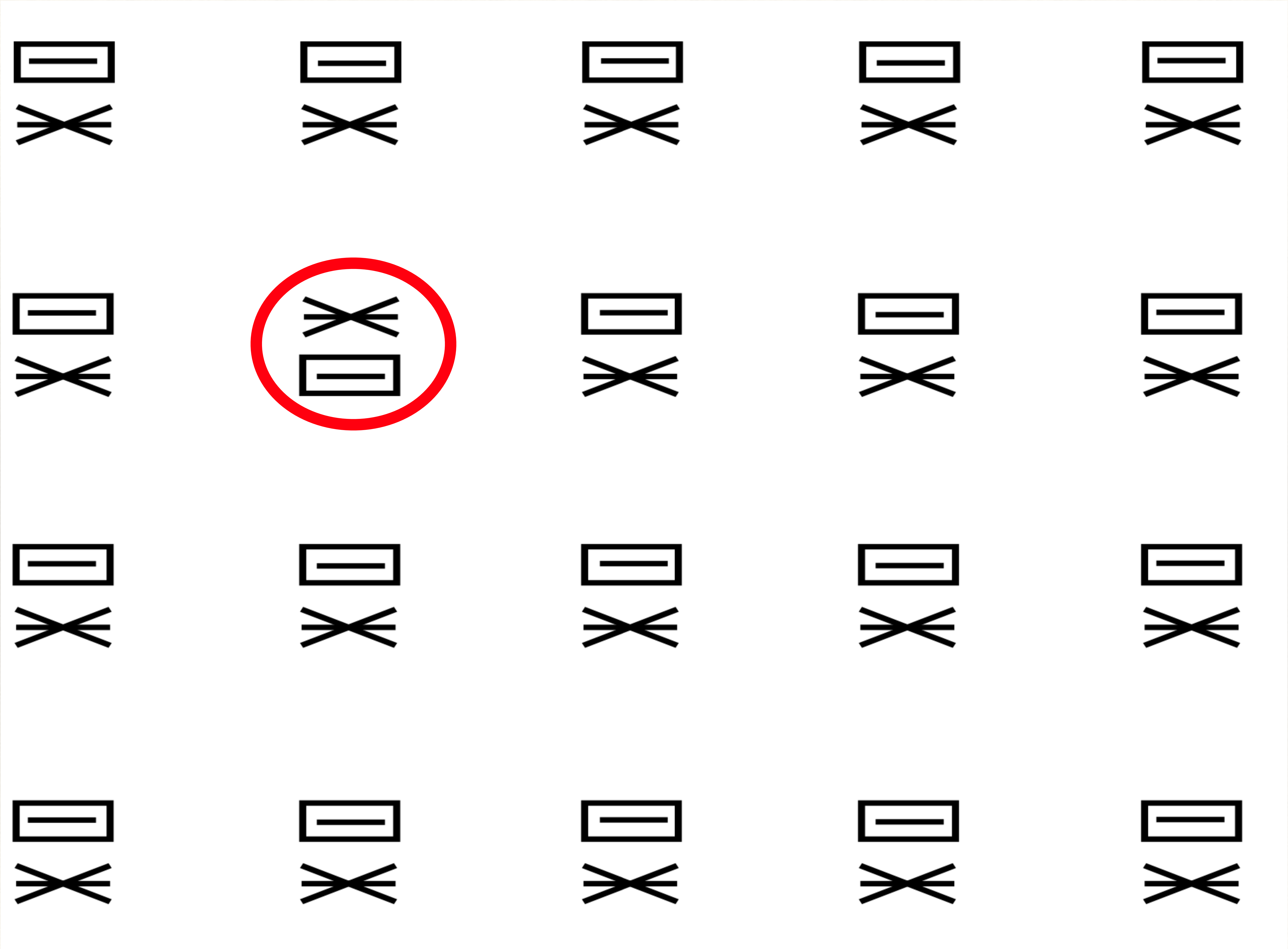
Why I am Here

- To tell you about an interesting problem
- To tell you about a genius
- To tell you what you already know! (but perhaps from how I see it)
- To take back something from you

Search in Research



Search in Research



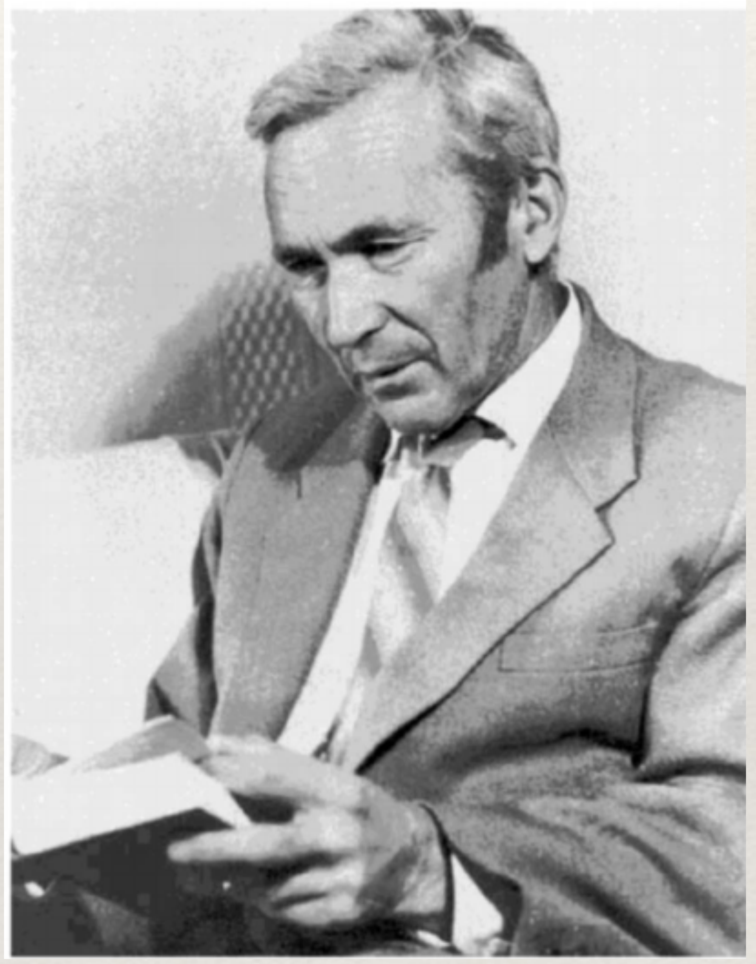
Some Points

- A learner is interested in identifying the “odd” image as quickly as possible
- The learner could make mistakes, but the chance of this happening should be low
- The learner may or may not know the images in advance
- Closeness of two images as perceived by humans is a measure of the hardness of the odd image identification problem

More on this in the coming slides

Before We Proceed, Let us Owe Our Thanks To…

A. N. Kolmogorov



The father of modern day theory of probability

Detour: Probability Theory (Through My Eyes)

A Brief History of the Evolution of Probability Theory

- Probability = chance of occurrence of an event
- The earliest known definitions of probability are from Bernoulli (1713) and de Moivre (1718)
- Classical definition of probability:

$$P(A) = \frac{\text{no. of times event } A \text{ occurs}}{\text{total no. of cases possible under the given circumstances}}$$

Drawbacks of the Classical Definition

- Requires an experiment to be carried out many times to compute the chance of an event occurring
- What if we want to compute the probability of ISRO placing a payload in Mars orbit successfully?
- Classical / frequentist approach to probability:
    - French mathematicians: happy
    - German and English mathematicians: not happy

Bertrand's Paradox

- 3 identical jewel boxes 📦 with two drawers each
- One of the boxes has gold bangles in both drawers, one has silver in both, one has gold in one and silver in another
- Probability of selecting box with one gold and one silver is?
  - Ans: 1/3
- Suppose one of the drawers is opened and found to have gold. The other drawer can have either gold or silver. So, the probability of selecting box with one gold and one silver is 1/2
- How's this possible?

Poincaré's Solution to Bertrand's Paradox

Box A

Box B

Box C

Possible cases

GG	SS	GS
GG	SS	SG
GG	GS	SS
GG	SG	SS
SS	GG	GS
SS	GG	SG
SS	GS	GG
SS	SG	GG
GS	GG	SS
GS	SS	GG
SG	GG	SS
SG	SS	GG

Poincaré's Solution to Bertrand's Paradox

Box A

Box B

Box C

Possible cases when the first drawer of the first box is G

GG	SS	GS
GG	SS	SG
GG	GS	SS
GG	SG	SS
GS	GG	SS
GS	SS	GG

$\boxed{\text{Ans is } 2/6 = 1/3}$

Bertrand's Paradox and Classical Probability

- Growing dissatisfaction among many German and English mathematicians with the classical approach
- Need for more detailed reasoning to answer Bertrand's paradoxes
- Need for an abstract theory that does not rely on conducting experiments many times

Measure Theory and The Genius of Kolmogorov

- The development of measure theory by Borel and Lebesgue (1894) was a huge contribution to mathematics
- Kolmogorov formulated axioms of probability based on Borel and Lebesgue's framework
- Kolmogorov's new theory was abstract and powerful to look beyond experimentation

Kolmogorov's Axioms of Probability

- $P(\Omega) = 1$ [here, Ω is the collection of all possible outcomes]
- If A and B are any two ***disjoint*** events, then

$$P(A \cup B) = P(A) + P(B)$$

- If $A_1, A_2, \ldots$ is any collection of pairwise disjoint events, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

"The theory of probability as mathematical discipline can and should be developed from axioms in exactly the same way as Geometry and Algebra."

A. N. Kolmogorov

From Outcomes to Numbers

- The notion of outcome of an experiment is too primitive to work with
- Numbers are more appealing than mere outcomes.
  E.g.: gambling, wireless communications

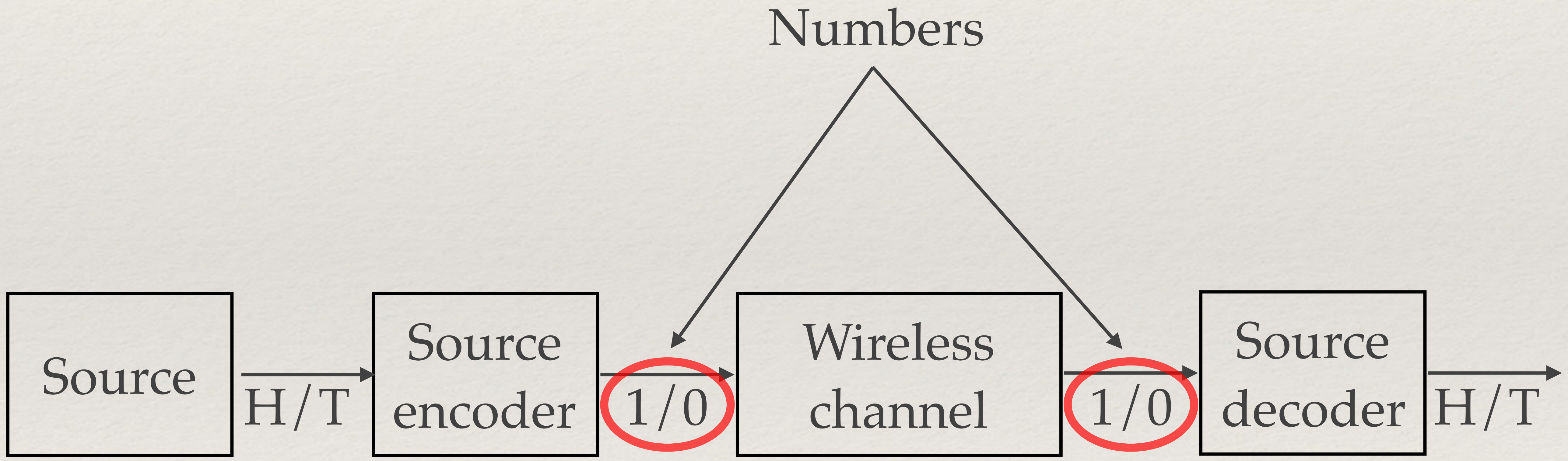


Fig: Transmitter wants to convey heads/tails information to receiver

Random Variables

- A random variable is a **function**
- Domain of the function: Ω
- Codomain of the function: $\mathbb{R}$

$$X\colon \Omega \to \mathbb{R}$$

Discrete and Continuous Random Variables

Discrete RVs

- Domain: Ω
- Range: a set which is either finite or countably infinite

Continuous RVs

- Domain: Ω
- Range: an uncountably infinite set

$$\{0,1\}, \quad \{1,2,3\}$$

$$\mathbb{N}, \quad \mathbb{Q}$$

$$\mathbb{R}$$

Discrete and Continuous Random Variables

Discrete RVs

- Domain: Ω
- Range: a set which is either finite or countably infinite

Continuous RVs

- Domain: Ω
- Range: an uncountably infinite set

The particular values that a random variable takes are not so important.
What is important is the **probability** with which the random variable takes those values.

Random Variables and Probability

$$\Omega \to \mathbb{R} \underset{F_X}{\overset{p_X}{\longrightarrow}} [0,1]$$

Cumulative Distribution Function (CDF)

$$F_X : \mathbb{R} \to [0,1]$$

$$F_X(x) = P(X \leq x)$$

Defined for all random variables

Probability Mass Function (PMF)

$$p_X : \mathbb{R} \to [0,1]$$

$$p_X(x) = P(X = x)$$

Defined for discrete random variables

Probability Density Function (PDF)

$$f_X : \mathbb{R} \to [0,\infty)$$

$$f_X(x) = \frac{d}{dx} F(x)$$

Defined for continuous RVs

Most Popular Discrete Distributions

$$X : \Omega \to \{0,1\}$$

- **Bernoulli** distribution:
  - Defined by a parameter $p \in [0,1]$
  - The PMF is given by

$$P(X = k) = p^k (1 - p)^{1-k}, \quad k \in \{0,1\}$$

  - A Bernoulli random variable is an indicator of success or failure

Most Popular Discrete Distributions

$$X : \Omega \to \{0, \ldots, n\}$$

- **Binomial** distribution:
  - Defined by a two parameters $n \in \{0,1,\ldots\}$ and $p \in [0,1]$
  - The PMF is given by

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{1-k}, \quad k \in \{0,\ldots,n\}$$

  - A Binomial random variable counts how many successes have occurred in a sequence of trials each of whose outcomes is either a success or a failure

Most Popular Discrete Distributions

$$X : \Omega \to \{0,1,2,\ldots\}$$

- **Poisson** distribution:
  - Defined by a parameter $\lambda \in (0,\infty)$
  - The PMF is given by

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \{0,1,2,\ldots\}$$

  - A Poisson random variable can be used for <u>counting</u> the number of customers who arrive at a bank, number of neurons which fire upon seeing an image, etc

Most Popular Discrete Distributions

$$X : \Omega \to \{1,2,\ldots\}$$

- **Geometric** distribution:
  - Defined by a parameter $p \in [0.1]$
  - The PMF is given by

$$P(X = k) = p(1 - p)^{k-1}, \quad k \in \{1,2,\ldots\}$$

  - A Geometric random variable counts the number of trials until the first success is seen in a sequence of trials each of whose outcomes is either a success or a failure

Most Popular Continuous Distributions

$$X : \Omega \to \mathbb{R}$$

- **Normal (or Gaussian)** distribution:
  - Defined by a two parameters $\mu \in \mathbb{R}$ and $\sigma \in (0,\infty)$
  - The PDF is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

  - A Gaussian random variable represents the amount of noise in an additive white Gaussian noise (AWGN) channel

Most Popular Continuous Distributions

$$X : \Omega \to (0,\infty)$$

- **Exponential** distribution:
  - Defined by a parameter $\eta > 0$
  - The PDF is given by

$$f_X(x) = \eta\, e^{-\eta x}, \quad x \in (0,\infty)$$

  - An Exponential random variable represents the lifetime of a component such as an electric bulb

Expectation

- Represents one single value that the random variable takes "typically"
- Denoted by the operator $\mathbb{E}[\,\cdot\,]$ and defined as follows:
  - Discrete random variables:

$$\mathbb{E}[X] = \sum_x x\,P(X = x) = \sum_x x\,p_X(x)$$

    provided the RHS is not of the form $\infty - \infty$

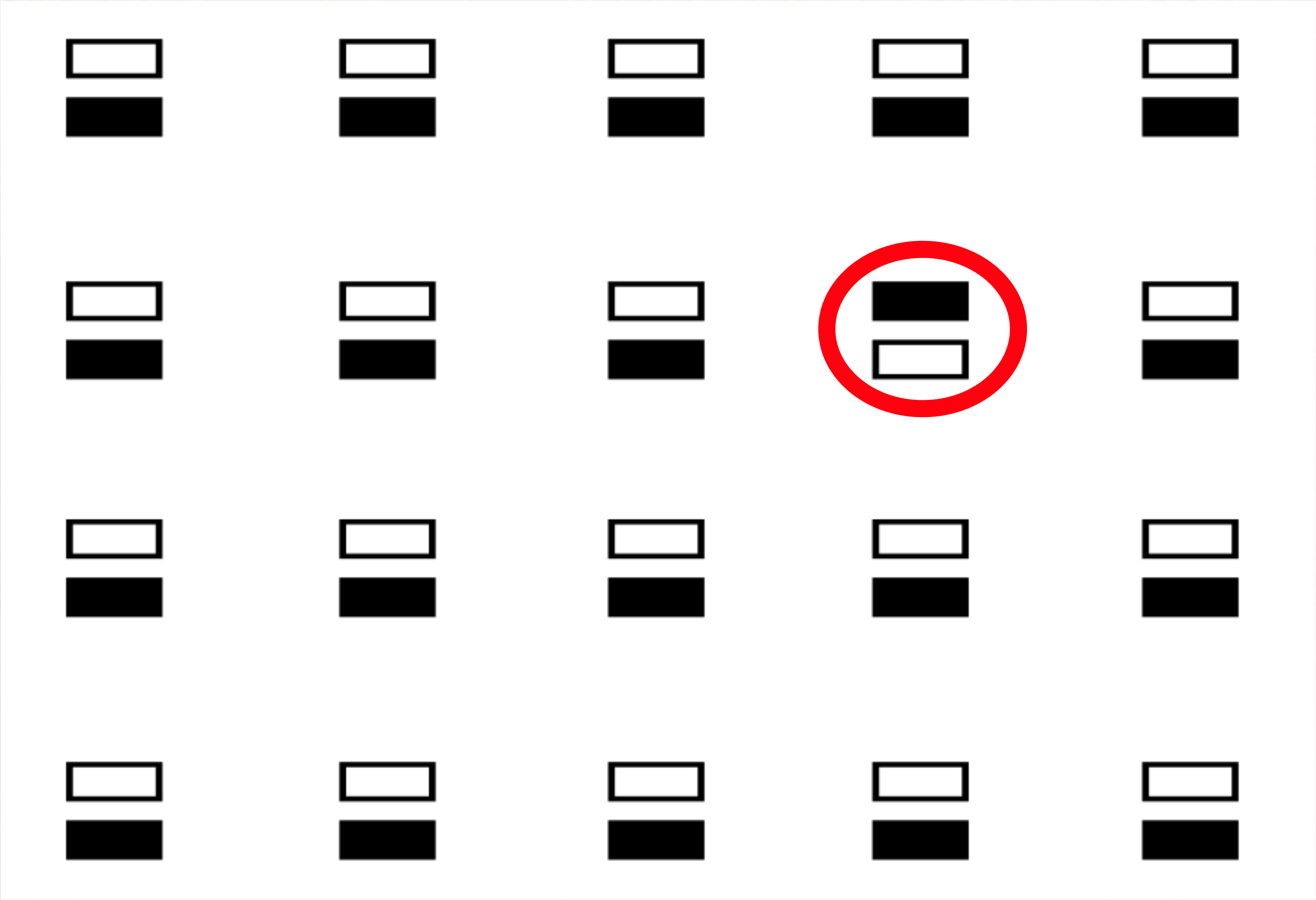
  - Continuous random variables:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\,f_X(x)\,dx$$

    provided the RHS is not of the form $\infty - \infty$

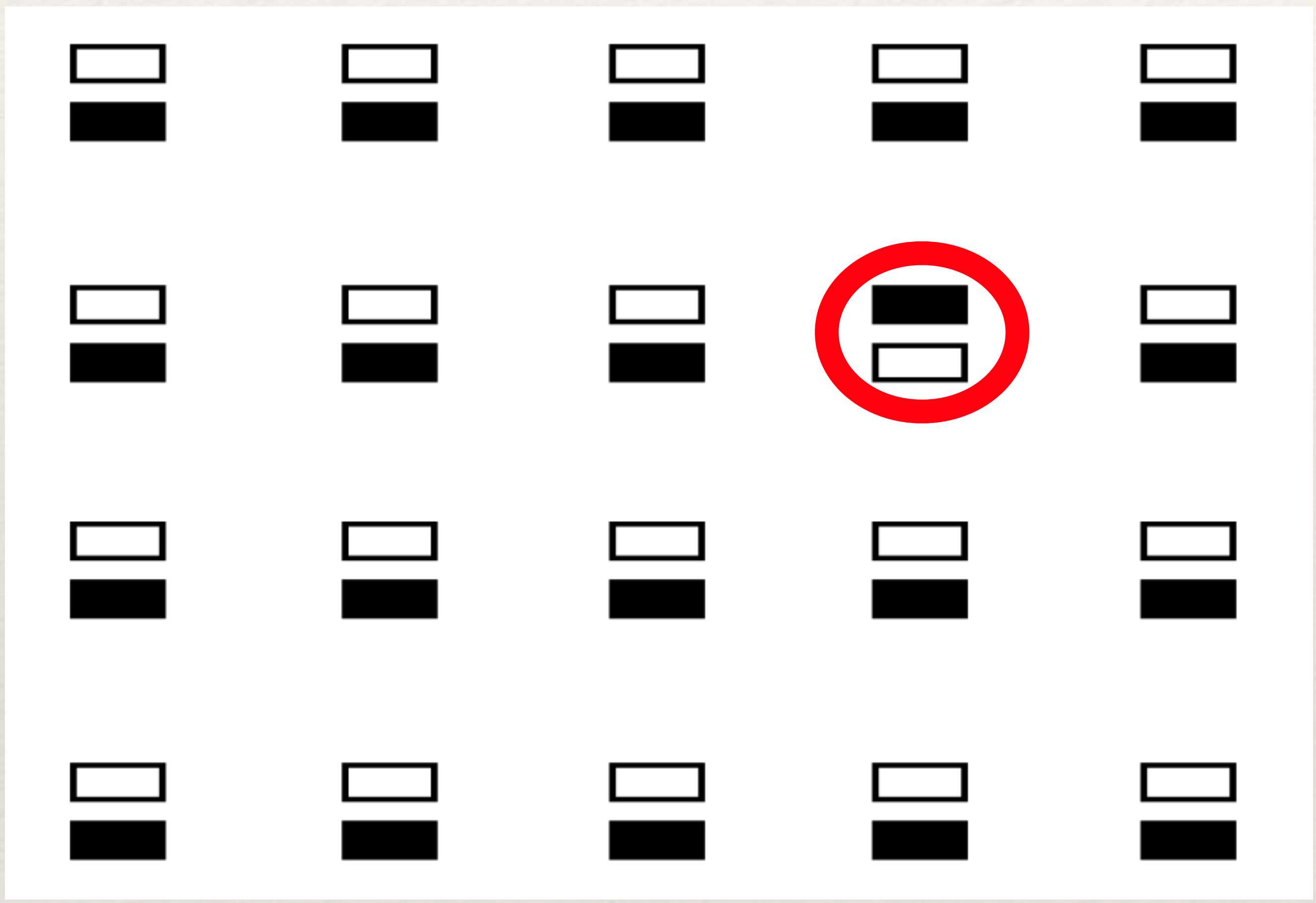
Odd Image Search:
Applying Probability to Solve the Problem

Recall the Setup

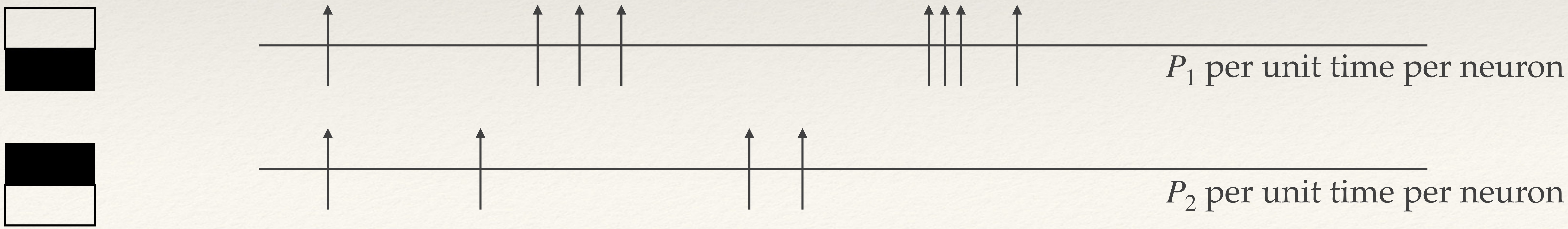


- A certain number of images displayed on screen
- One of the images is different from the rest
- The goal is to identify the odd image in the shortest possible time, while also being correct with high probability

Inside the Brain



- Suppose there are K locations
- Suppose there are two images I_1 and I_2. You know these images beforehand
- One of the images (say I_2) is present at $K - 1$ locations, and the other is present at the remaining location (which is not known to you)
- You are the learner, and you wish to identify the location where image I_1 is present
- You can look at only one location at any given time
- When you look at the image at any location, the neurons in your brain fire at some rate
  - Firing rate when you look at image I_1: P_1
  - Firing rate when you look at image I_2: P_2



The Poisson Model

- Neuroscientists[1] conducted an experiment in which they allowed the human subjects to look at each location for a fixed duration of 5 seconds
- They discovered that the number of neuronal firings follows a Poisson distribution
- Thus, we have two Poisson distributions to work with: Poisson(P_1) and Poisson(P_2)

[1] A. P. Sripati and C. R. Olson, "Global image dissimilarity in macaque inferotemporal cortex predicts human visual search efficiency," J. Neurosci., vol. 30, no. 4, pp. 1258–1269, Jan. 2010.

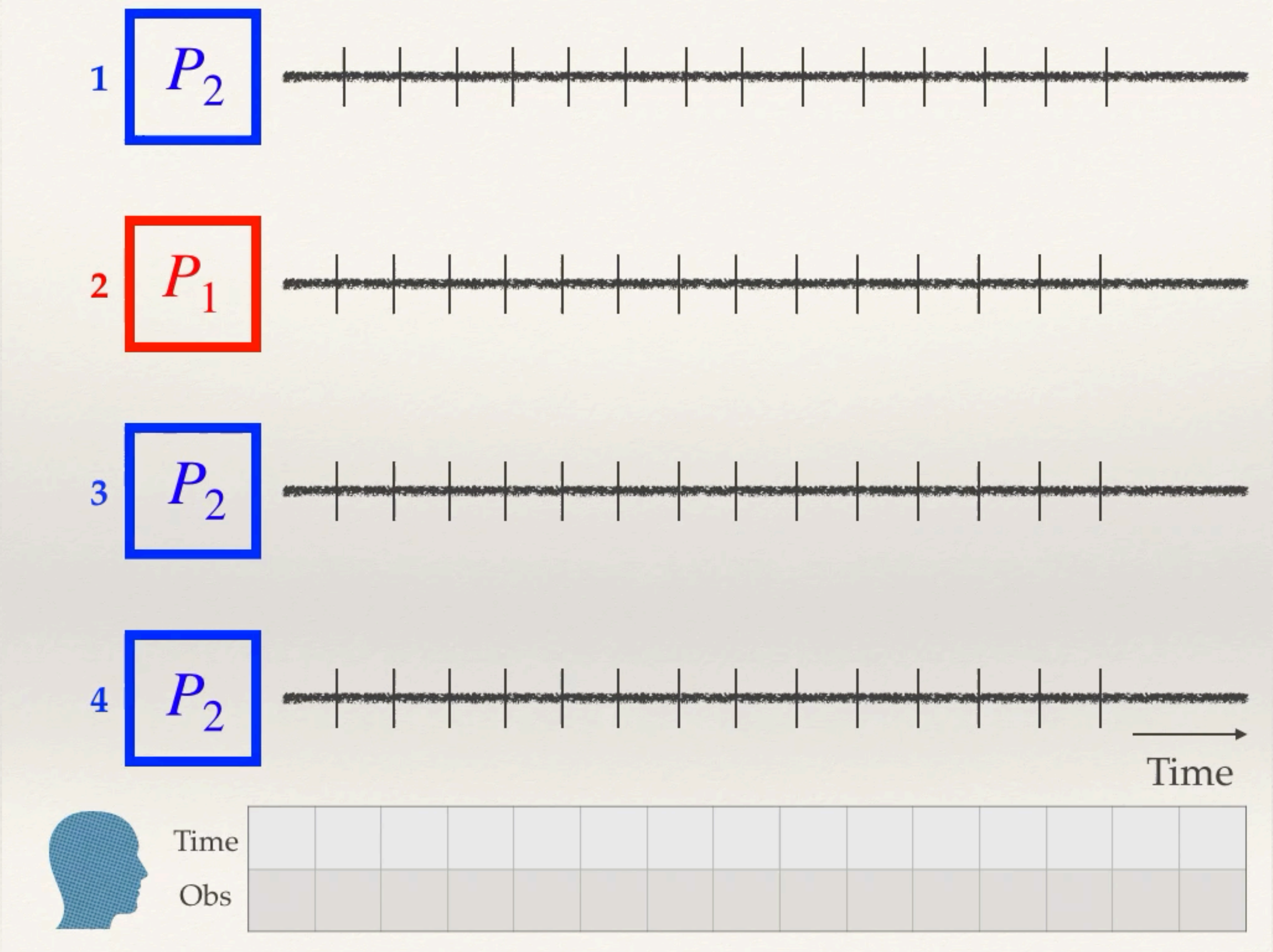
Hypothesising the Possible Answers

- Let us hypothesise all the possible answers to our problem
- There are K possible answers in which the location of the odd image is one of the possible K locations
- Thus, we have the following hypotheses:

$$\begin{aligned}
\mathscr{H}_1 : &\quad \text{odd image location} = 1 \\
\mathscr{H}_2 : &\quad \text{odd image location} = 2 \\
&\vdots \\
\mathscr{H}_K : &\quad \text{odd image location} = K
\end{aligned}$$

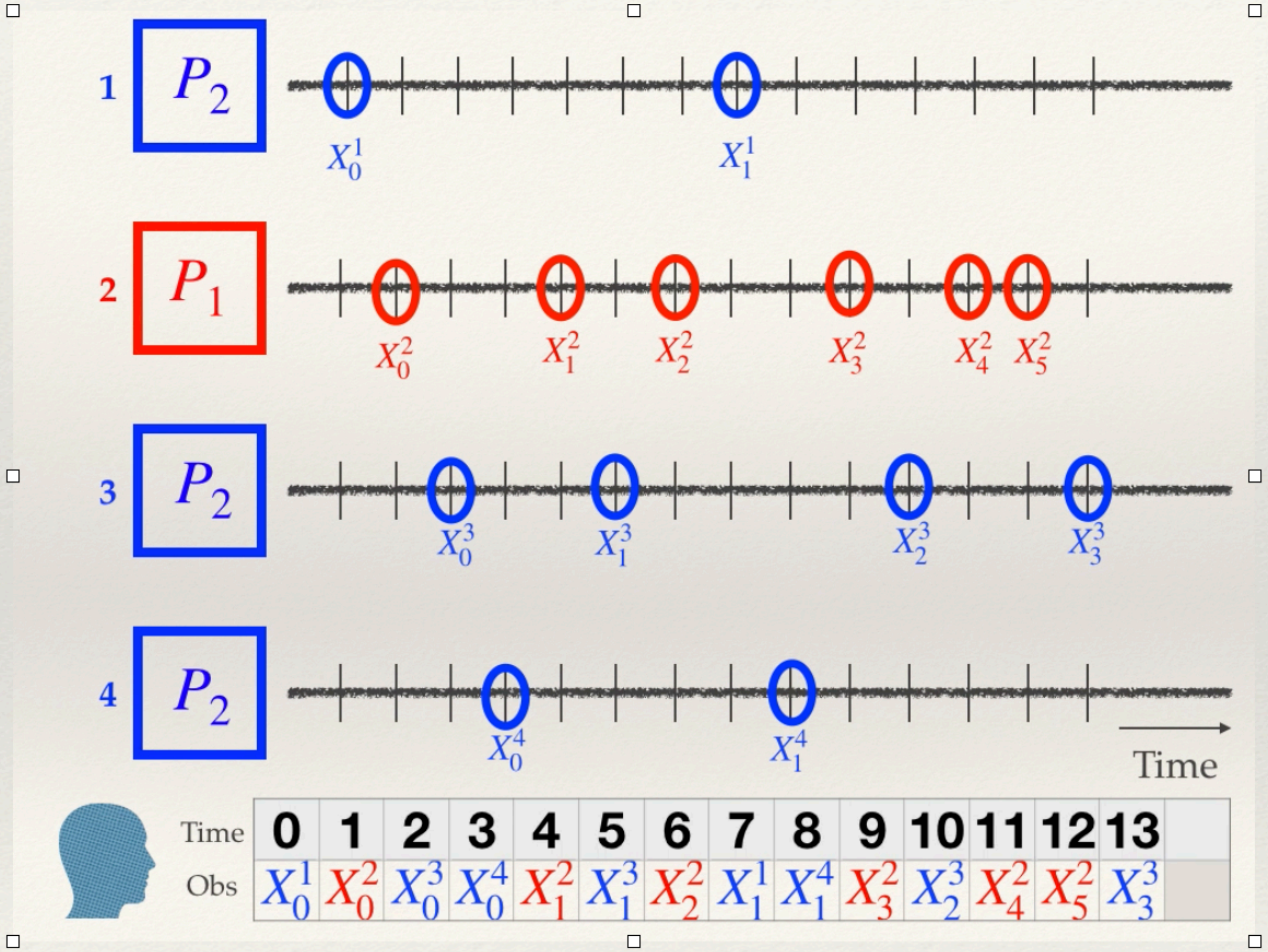
- Goal is to identify as quickly as possible which hypothesis is the true hypothesis

A Visualisation of the Problem



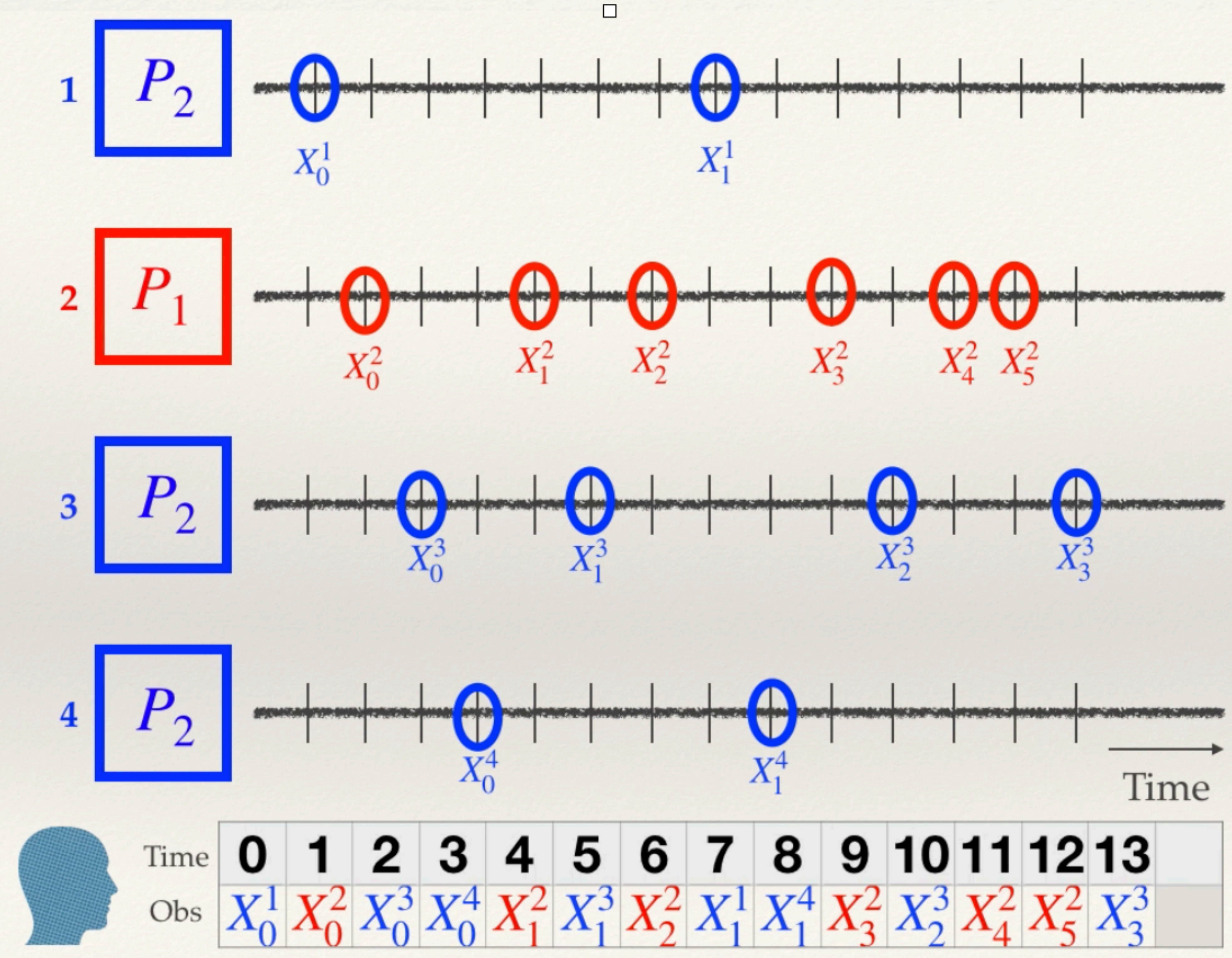
- Divide time into discrete slots of 5 seconds each
- In each time slot, you pick a location and observe the image at that location
- Suppose L_t is the location you pick to observe in time slot t
- Let $\bar{X}_t$ is the number of neuronal firings observed in time slot t
- At the beginning of the next time slot, you have to decide whether to
    - Stop and declare the location of the odd image, or
    - Continue to pick the next location

A Visualisation of the Problem



- The time at which you will stop depends on your policy of how you pick locations to observe, and is therefore random
- There can be many possible policies that you may choose to use. E.g.: round robin, uniform sampling, etc
- Given a policy π, let $\tau(\pi)$ be the time taken to stop under policy π
- We want to minimise $\mathbb{E}[\tau(\pi)]$

Important Questions



- How do we capture the behaviour of $\mathbb{E}[\tau(\pi)]$ as a function of
  - Probability of error
  - "Closeness" of the two images shown to you
- What policy should we use in order to identify the location of the odd image as quickly as possible while meeting the probability of error constraint?

A Measure of Closeness Between Two Images

- A metric to capture the notion of how humans perceive two images as different
- The metric should be asymmetric since the time taken to identify 0 in a sea of Os is different from that taken to identify O in a sea of 0s
- The most widely employed asymmetric measure of closeness is Kullback-Leibler (KL) Divergence
- KL divergence between two discrete distributions P and Q is given by

$$D(P\|Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$$

A Game Between You and Nature: Capturing the Hardness of the Problem

- You want to identify the location of the odd image as quickly as possible
- Nature does not want you to do so. It tries to confuse you as much as possible so that you take more time
- The only thing you can control is which location to pick next based on all the locations picked and observations obtained in the past
- Let's say location 1 is where the odd image is
- Consider another location $\ell \neq 1$
- If you decide to pick the next location to observe using a probability distribution λ in each time slot, then the average distance between hypotheses $\mathscr{H}_1$ and $\mathscr{H}_\ell$ is

$$\sum_{a=1}^{K} \lambda(a)\, D(P_1^a \,||\, P_\ell^a)$$

A Game Between You and Nature: Capturing the Hardness of the Problem

- Therefore, the distance to the nearest alternative hypothesis is

$$\min_{\ell \neq 1} \sum_{a=1}^{K} \lambda(a)\, D(P_1^a \,||\, P_\ell^a)$$

A Game Between You and Nature: Capturing the Hardness of the Problem

- We have to pick λ so that the distance to the nearest alternative is as large as possible (thereby beating nature to the best possible extent)

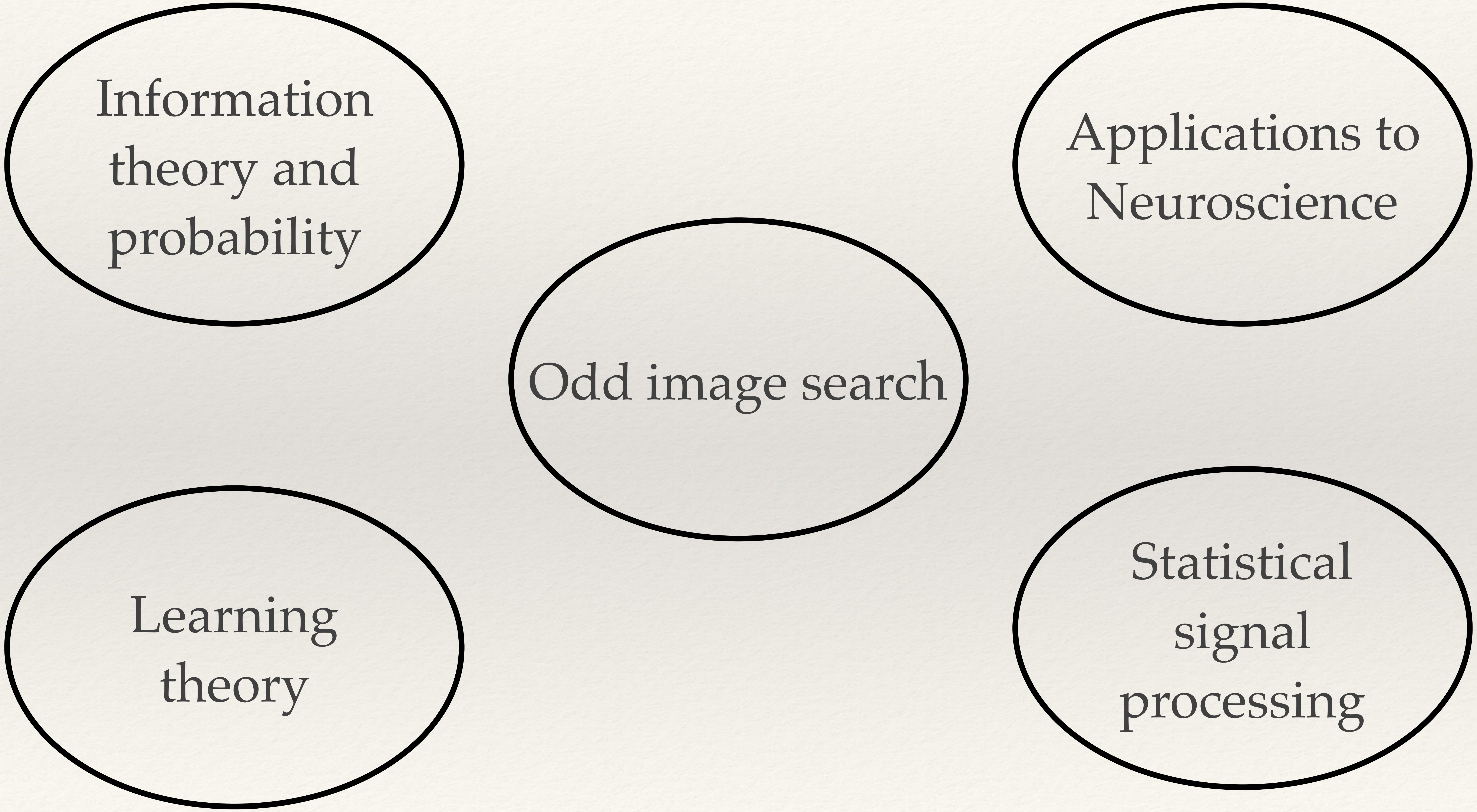
Hardness of the problem: amount of effort required to battle against nature

$$\max_{\lambda} \min_{\ell \neq 1} \sum_{a=1}^{K} \lambda(a)\, D(P_1^a \,||\, P_\ell^a)$$

Results

- A limiting lower bound on $E[\tau(\pi)]$ in terms of probability of error and the hardness of the problem
- A policy that meets the lower bound as probability of error vanishes

Intersection of Many Disciplines



Looking Beyond

- Generalisations to distributions other than Poisson[2]
- Dependence between successive observations from each location[3]
- Finite sample regime (fixed and positive probability of error)

[2] G. R. Prabhu, S. Bhashyam, A. Gopalan, and R. Sundaresan, "Learning to detect an oddball target with observations from an exponential family," 2017.

[3] P. N. Karthik and R. Sundaresan, "Learning to detect an Odd Markov Arm," presented at IEEE International Symposium on Information Theory (ISIT), Paris, France. A detailed manuscript is available at https://arxiv.org/pdf/1904.11361.pdf

Opportunities

Rajesh Sundaresan
IISc

Srikrishna Bhashyam
IIT Madras

Aditya Gopalan
IISc

Venugopal Veeravalli
UIUC

Jonathan Scarlett
NUS

Emilie Kaufmann
INRIA

Aurélien Garivier
ENS

Sandeep Juneja
TIFR Mumbai

References on Probability

- A first course in probability, Sheldon Ross
- Probability and Random Processes, Geoffrey Grimmett and David Stirzaker
- Probability and Random Processes with Applications to Signal Processing and Communications, Miller and Childers
- Introduction to Probability, Dmitiri Bertsekas (MIT lecture notes)
- Probability Foundations for Electrical Engineers, NPTEL lectures by Krishna Jagannathan, IIT Madras

Lectures and Seminars on Probability

- Mathematics Training and Talent Search (MTTS) Programme
  - Rigorous mathematical training to students and faculty with emphasis on foundations
- Bangalore Probability Seminar (BPS)
  - Conducted on Mondays at IISc and ISI Bangalore
- Lectures on Probability and Stochastic Processes (LPS)
  - Conducted at ISI Kolkata, ISI Delhi, ISI Bangalore and IISc every year, during Nov-Dec

Thank You!