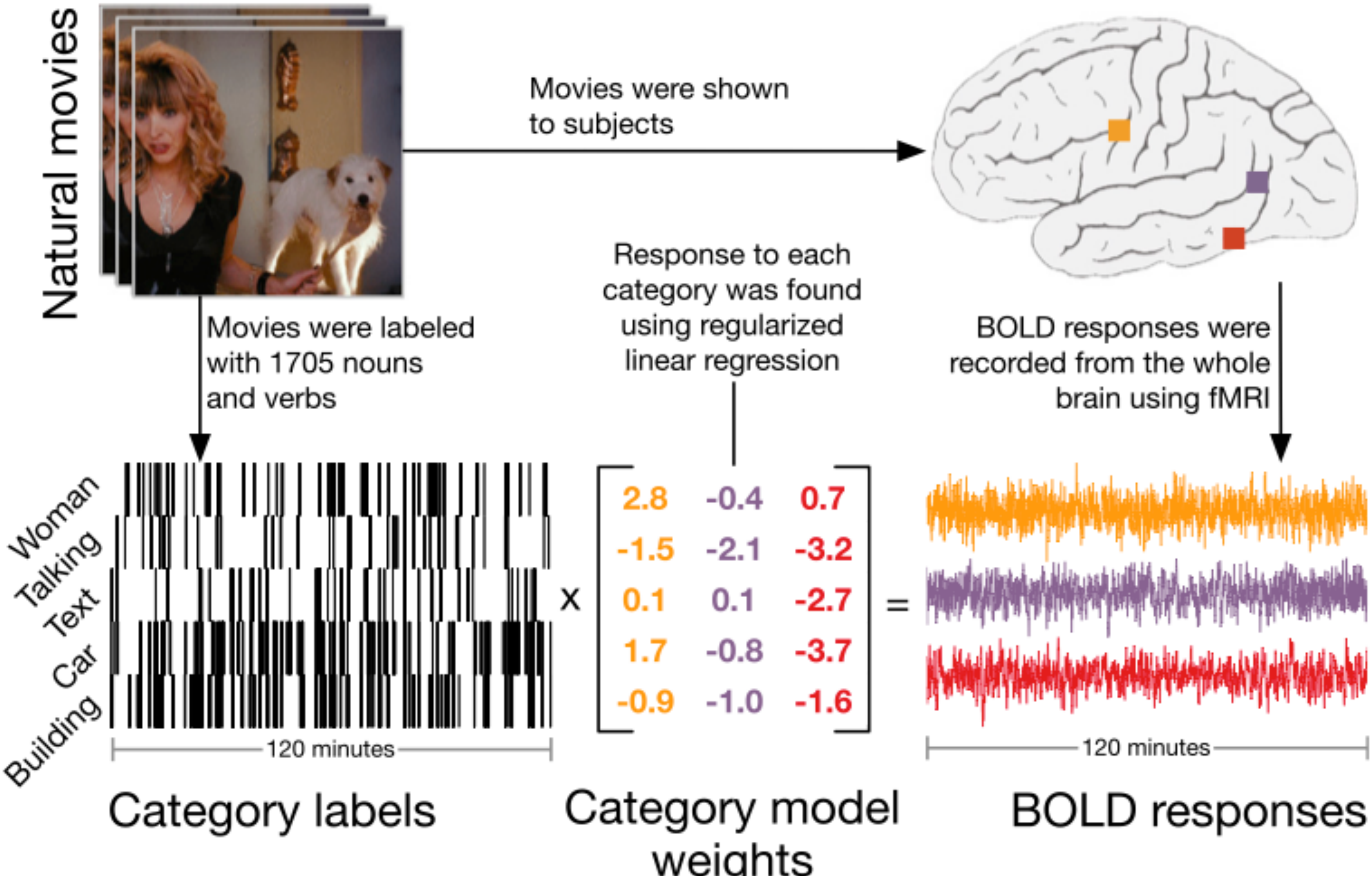
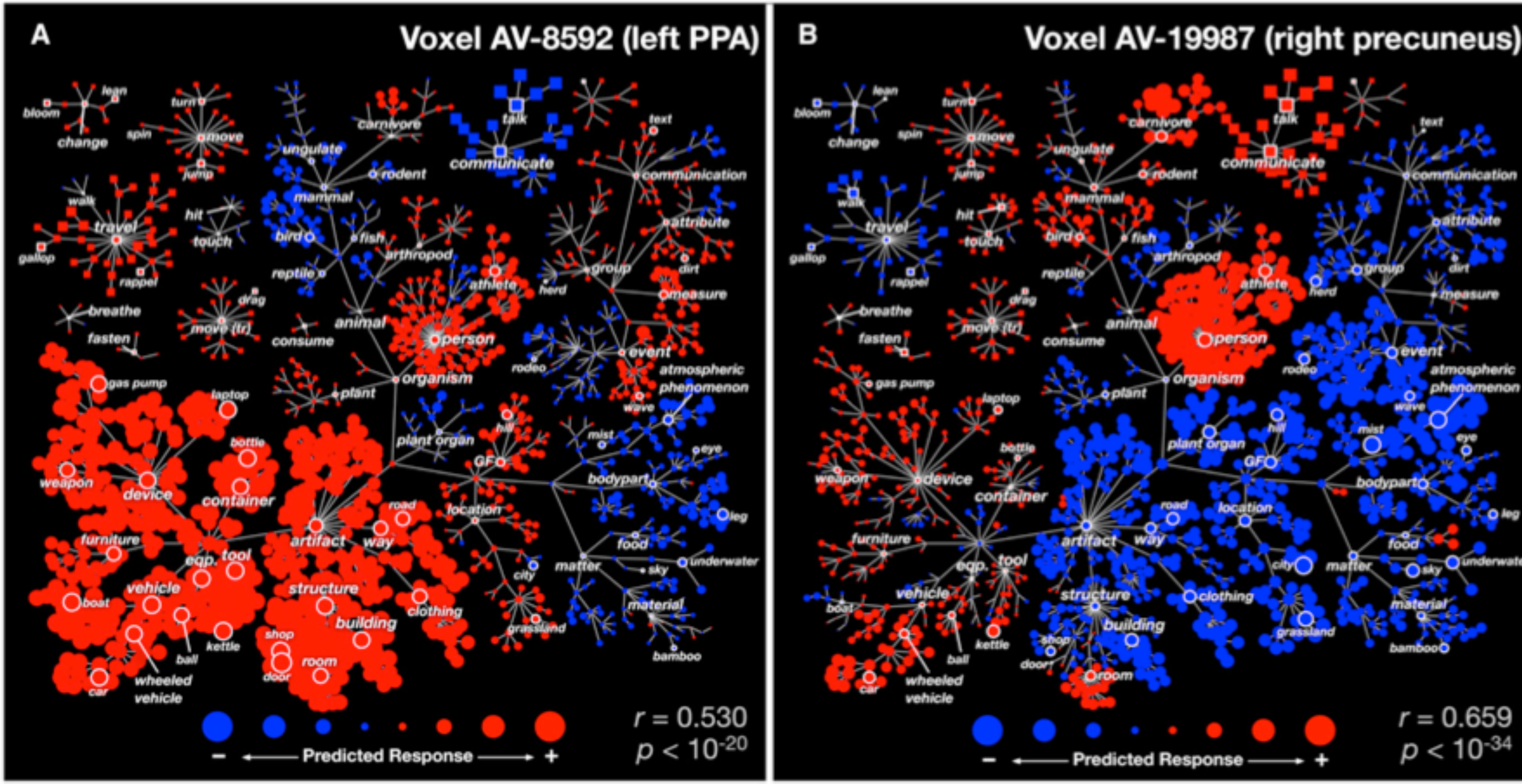
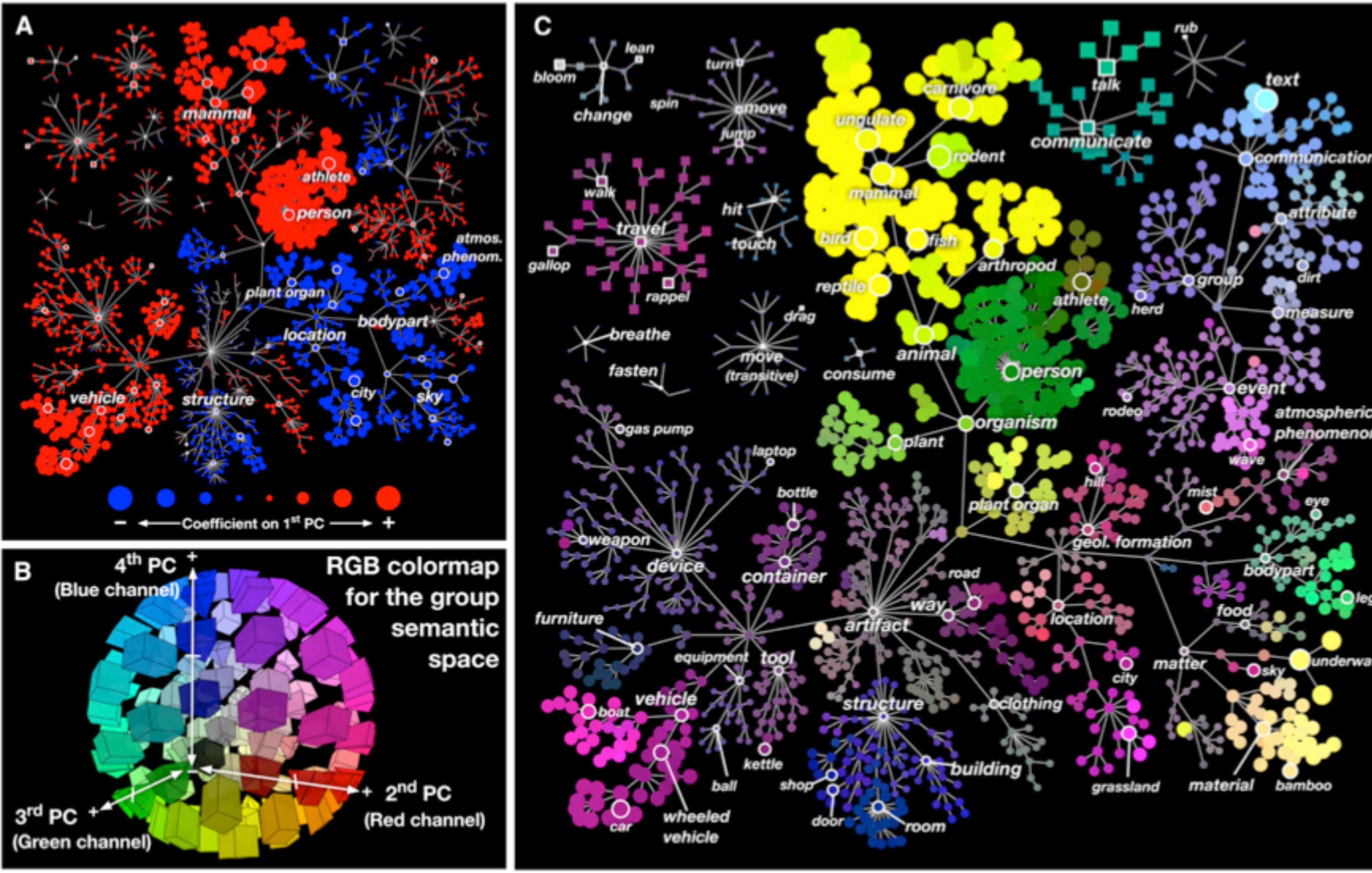
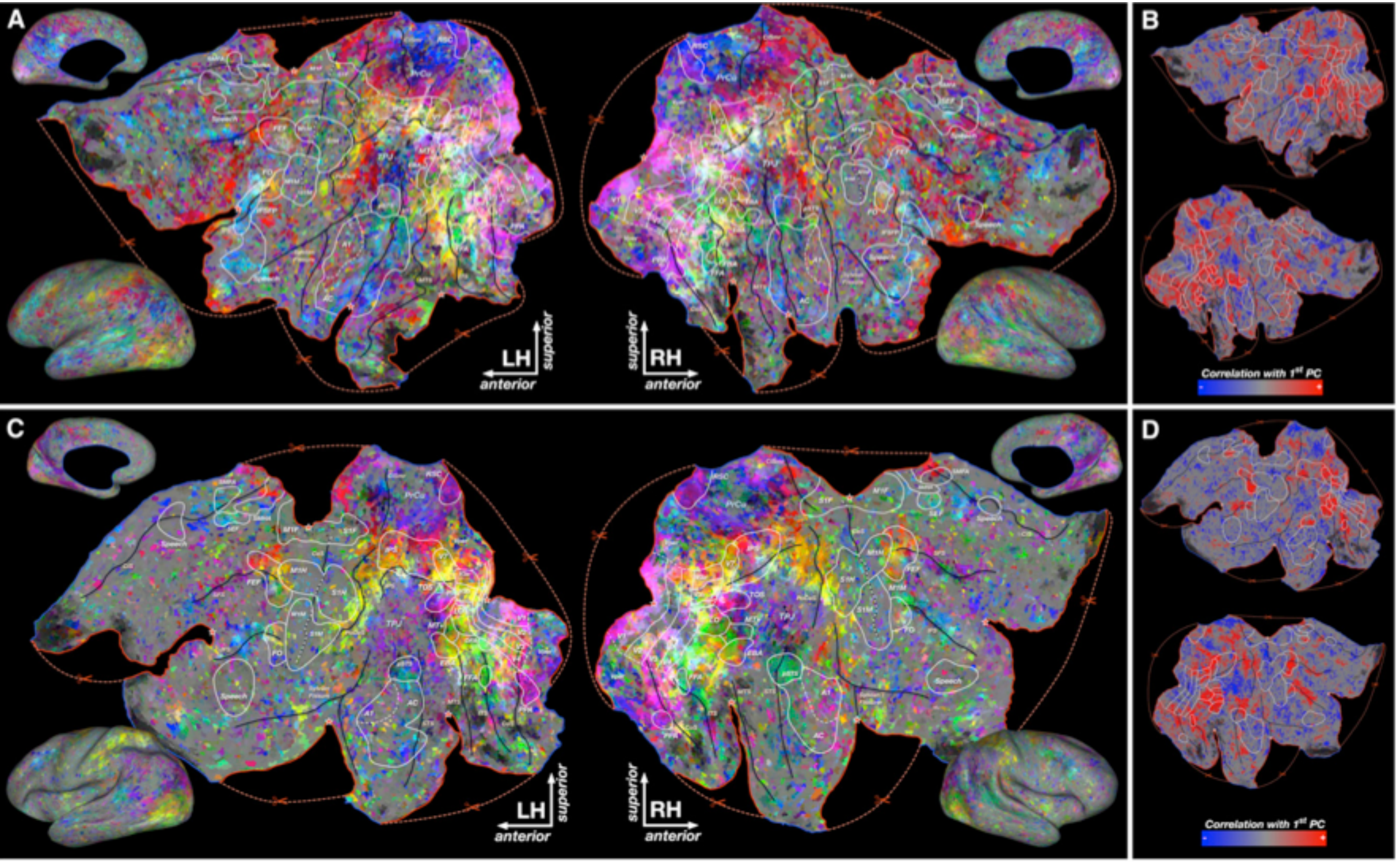

Natural movies
Movies were shown
to subjects
Movies were labeled
with 1705 nouns
and verbs
Response to each
category was found
using regularized
linear regression
BOLD responses were
recorded from the whole
brain using fMRI
Woman
Talking
Text
Car
Building
120 minutes
x
2.8 -0.4 0.7
-1.5 -2.1 -3.2
0.1 0.1 -2.7
1.7 -0.8 -3.7
-0.9 -1.0 -1.6
=
120 minutes
Category labels
Category model
weights
BOLD responses


A
Voxel AV-8592 (left PPA)
B
Voxel AV-19987 (right precuneus)
$r = 0.530$
$p < 10^{-20}$
$r = 0.659$
$p < 10^{-34}$
– Predicted Response +
– Predicted Response +


A
mammal
athlete
person
atmos. phenom.
plant organ
bodypart
location
vehicle
structure
city
sky
Coefficient on 1st PC
B
4th PC
(Blue channel)
RGB colormap for the group semantic space
3rd PC
(Green channel)
2nd PC
(Red channel)
C
bloom
lean
change
turn
spin
move
jump
walk
travel
gallop
rappel
hit
touch
breathe
fasten
drag
move (transitive)
consume
carnivore
ungulate
rodent
mammal
bird
fish
arthropod
reptile
animal
athlete
person
organism
plant
plant organ
talk
communicate
rub
text
communication
attribute
dirt
group
herd
measure
event
rodeo
atmospheric phenomenon
wave
gas pump
laptop
bottle
weapon
device
container
furniture
equipment
tool
road
way
artifact
hill
mist
geol. formation
eye
bodypart
leg
location
food
matter
sky
underwater
city
clothing
structure
boat
vehicle
ball
kettle
shop
building
door
room
car
wheeled vehicle
grassland
material
bamboo


A
B
C
D
LH
RH
superior
anterior
superior
anterior
Correlation with 1st PC
LH
RH
superior
anterior
superior
anterior
Correlation with 1st PC
PrCu
RSC
FEF
TPJ
TOS
Speech
AC
FFA
EBA
PPA

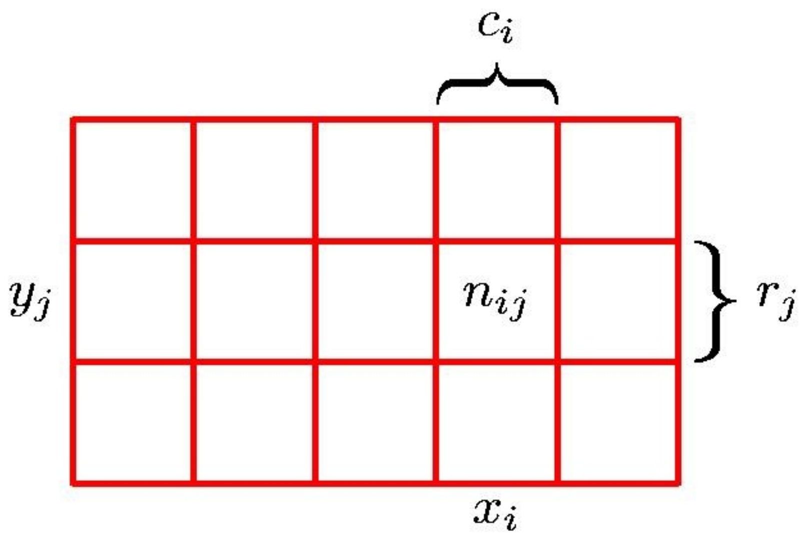
MACHINE LEARNING

PROBABILITY THEORY

LECTURE 4: BASICS, BAYES, DISTRIBUTIONS

Probability Theory

$$p(X, Y)$$



Marginal Probability

$$p(X = x_i) = \frac{c_i}{N}.$$

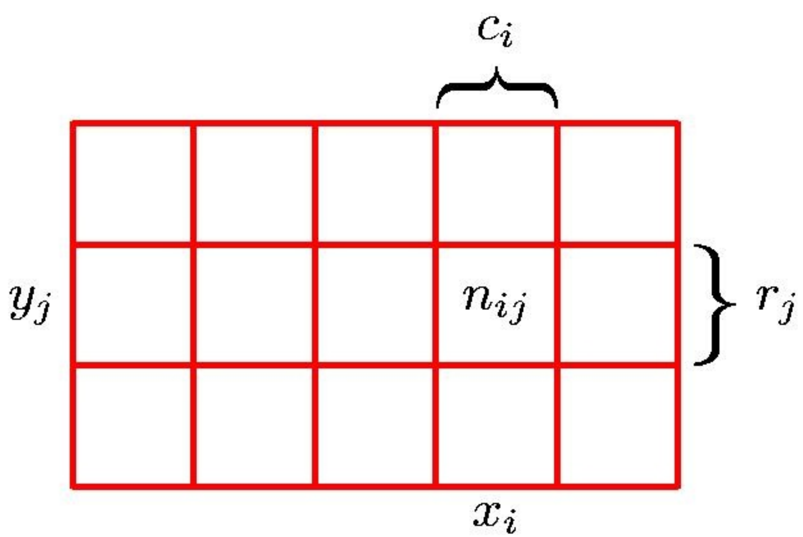
Joint Probability

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

Conditional Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

Probability Theory



Sum Rule

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N} \sum_{j=1}^{L} n_{ij}$$

$$= \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

Product Rule

$$\begin{aligned} p(X = x_i, Y = y_j) &= \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} \\ &= p(Y = y_j | X = x_i) p(X = x_i) \end{aligned}$$

The Rules of Probability

Sum Rule

$$p(X) = \sum_Y p(X, Y)$$

Product Rule

$$p(X, Y) = p(Y|X)p(X)$$

Bayes' Theorem

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

$$p(X) = \sum_Y p(X|Y)p(Y)$$

posterior $\propto$ likelihood × prior

Example Problem 1

1. (Kohler) Bad news at the doctor's

   After your yearly checkup the doctor has bad news. You tested positive for a serious disease and the test is 99% accurate. That is, the probability that the test result is positive given that you have the disease is 0.99. The test is very accurate: the probability that you do not have the disease given the test resuly is negative is also 0.99.

   The only good news is that the disease is rare, striking only one in 10,000 people. What are your chances given this information?

Solution

Let's let T^+ stand for the test being positive and T^- stand for the test being negative.

Similarly, let's let D^+ stand for the having the disease and D^- stand for not having the disease.

What we want to know is $P(D^+|T^+)$, the probability that we actually have the disease given that the test was positive.

Solution

Let's let T^+ stand for the test being positive and T^- stand for the test being negative.
Similarly, let's let D^+ stand for the having the disease and D^- stand for not having the disease.
What we want to know is $P(D^+|T^+)$, the probability that we actually have the disease given that the test was positive. Using Bayes rule:

$$P(D^+|T^+) = \frac{P(T^+|D^+)(P(D^+)}{P(T^+)},$$

$$= \frac{P(T^+|D^+)(P(D^+)}{P(T^+, D^+) + P(T^+, D^-)},$$

$$\frac{P(T^+|D^+)(P(D^+)}{P(T^+|D^+)P(D^+) + P(T^+|D^-)P(D^-)},$$

$$= \frac{1}{1 + \frac{P(T^+|D^-)P(D^-)}{P(T^+|D^+)P(D^+)}},$$

$$= \frac{1}{1 + \frac{(0.01)(0.9999)}{(0.99)(10^{-4})}},$$

Example Problem 2

2. (MacCay) Monte Hall On a game show a contestant is told the rules as follows:

   There are three doors labeled 1, 2 and 3. A single prize has been hidden behind one of them. You get to select one door. Initially your chosen door will *not* be opened. Instead the gameshow host will open one of the two remaining doors and will do so in a way so as to not reveal the prize.

   At this point you will be given a fresh choice of door. You can either stick to your choice or choose the remaining unopened door.

   Should the contestant stick with the originally chosen door or switch? Or does it matter?

Let H_i represent the hypothesis that the prize is behind door i.

And without loss of generality let's assume the contestant chooses door 1 and the gameshow host opens door 3.

A priori we must assume that

$$P(H_i) = \frac{1}{3}. i = 1, 2, 3$$

The key observation is that if the contestant chose a door with the prize, then the host can choose either of the remaining doors, but if the ocntestant *didn't* choose the prize then the host is restricted in choosing the door to open. Where D is the door the host is thinking of picking, these options can be expressed as;

$$P(D=2|H_1) = \frac{1}{2},$$
$$P(D=3|H_1) = \frac{1}{2},$$

$$P(D=2|H_2) = 0,$$
$$P(D=3|H_2) = 1,$$

$$P(D=2|H_3) = 1,$$
$$P(D=3|H_3) = 0$$

Now for Bayes rule:

$$P(H_i)|D=3) = \frac{P(D=3|H_i)p(H_i)}{P(D=3)}$$

So, enumerating the options,

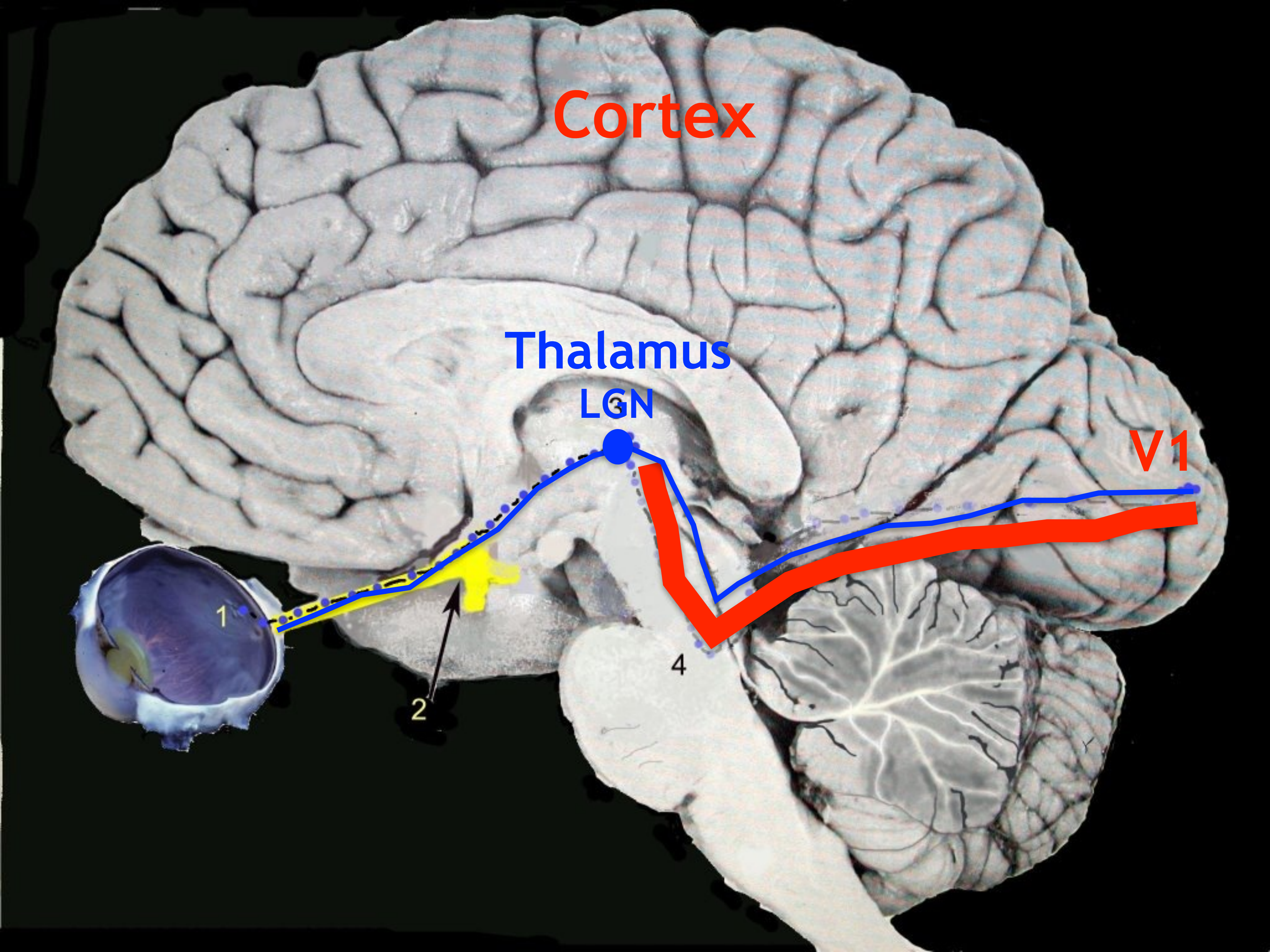
$$P(H_1)|D=3) = \frac{\frac{1}{2} \times \frac{1}{3}}{P(D=3)}$$

$$P(H_2)|D=3) = \frac{1 \times \frac{1}{3}}{P(D=3)},$$

$$P(H_3)|D=3) = \frac{0 \times \frac{1}{3}}{P(D=3)}$$

By inspection, after door 3 has been opened, H_2 is twice as likely as H_1, so the contestant should switch.

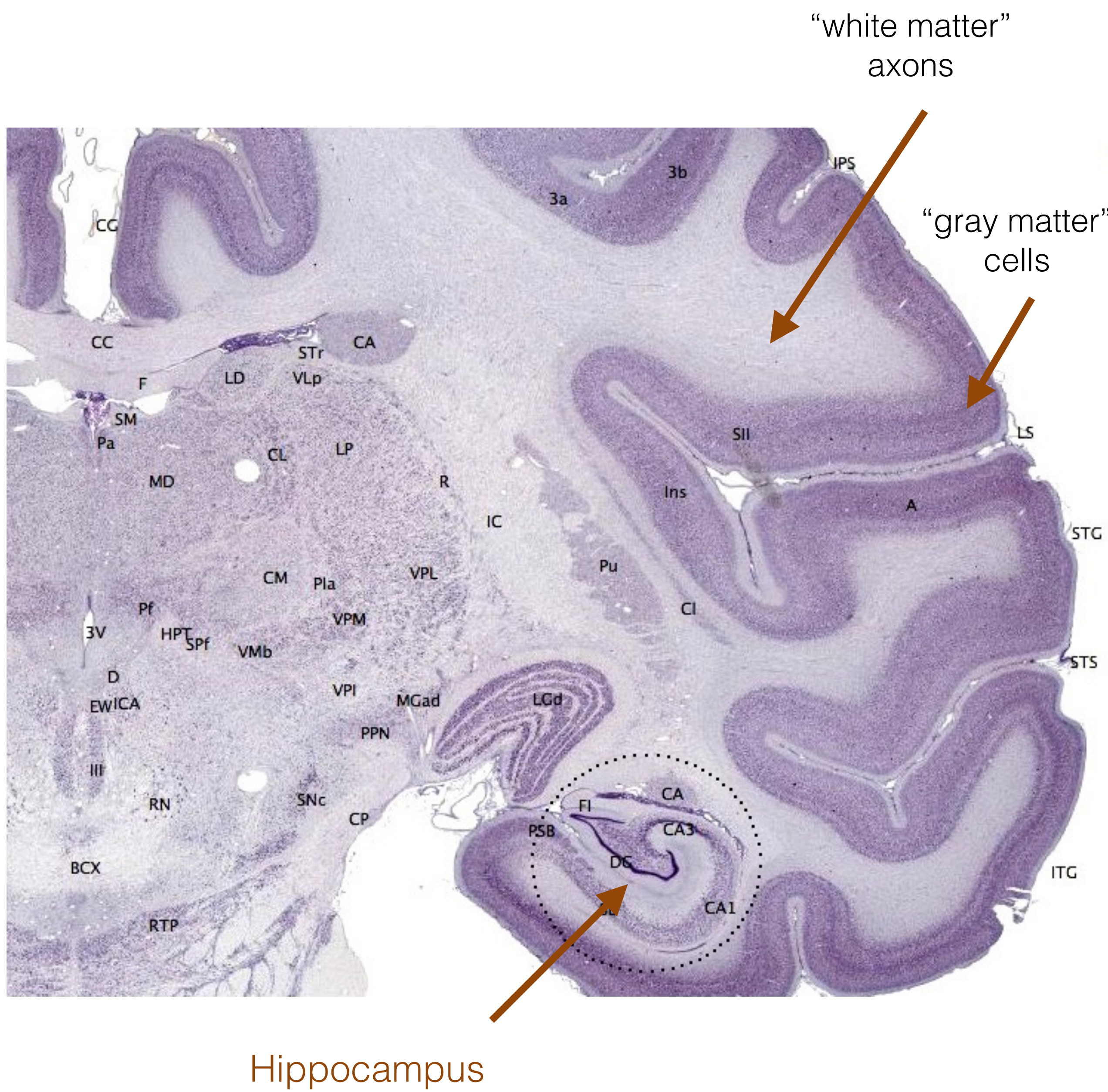
What is P(D=3)?



Cortex
Thalamus
LGN
V1
1
2
4

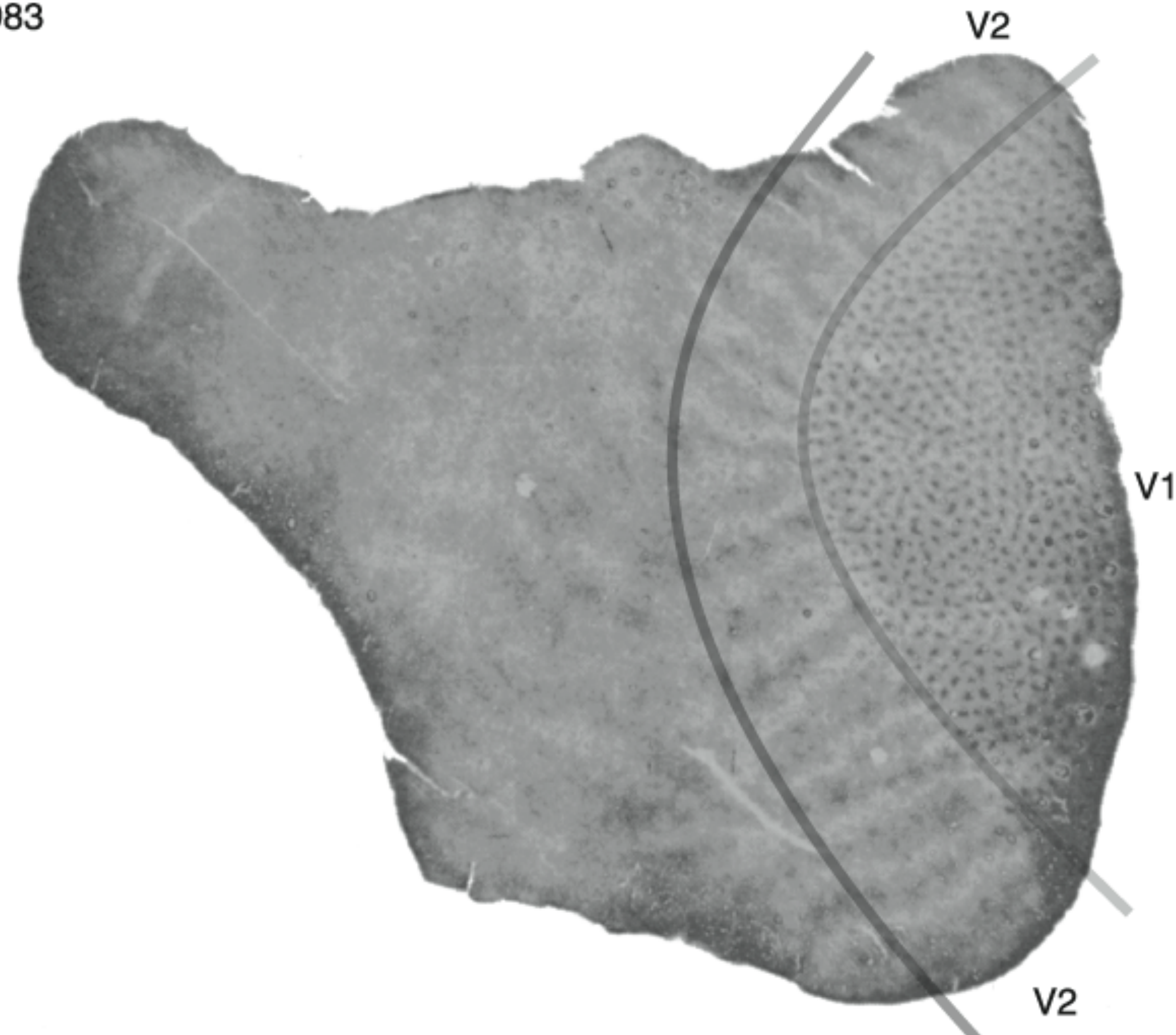
A slice through the macaque brain cortex

The cortex has been likened to a “Pizza” in that the cells are located in a thin outer layer 6 credit cards thick. This slice shows the cortical layers terminating in the Hippocampus, which has a critical role in the creation of new memories.

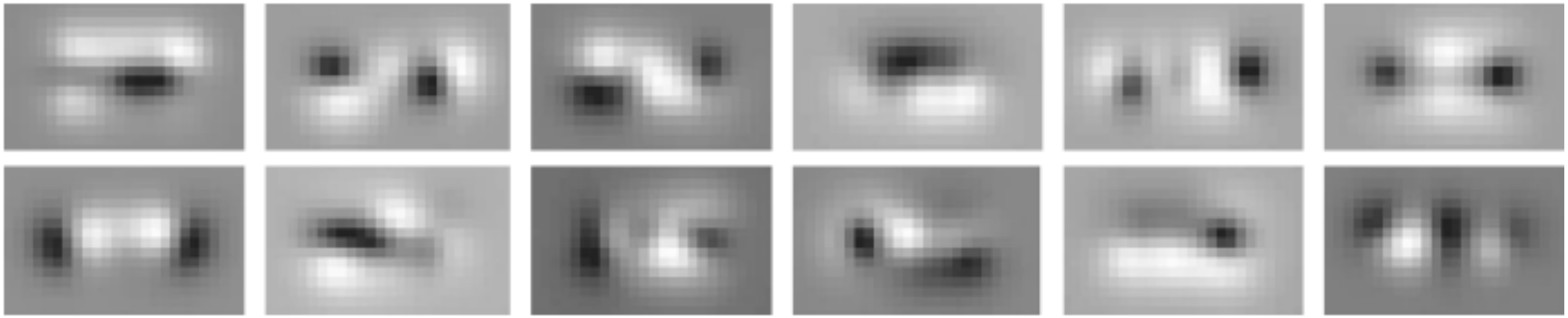


Cytochrome Oxidase in Monkey Visual Cortex

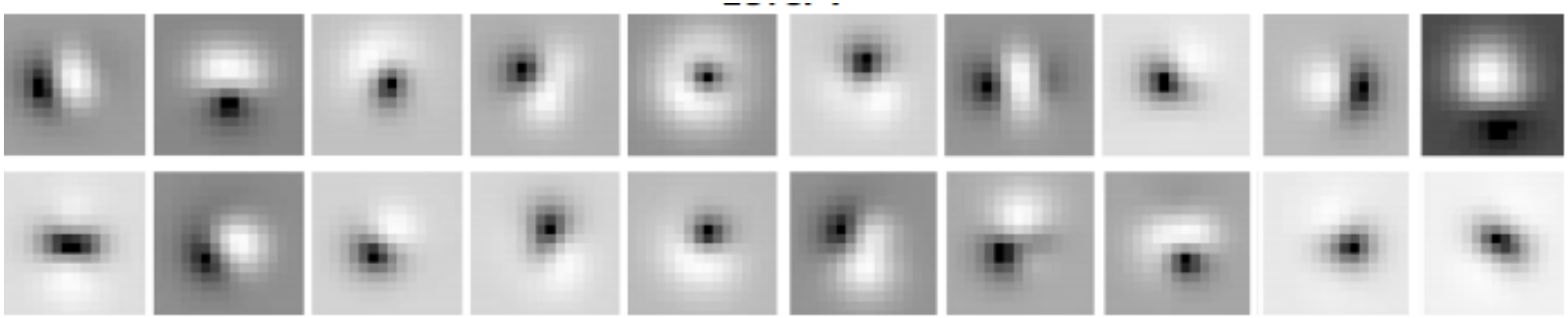
Tootell et al 1983



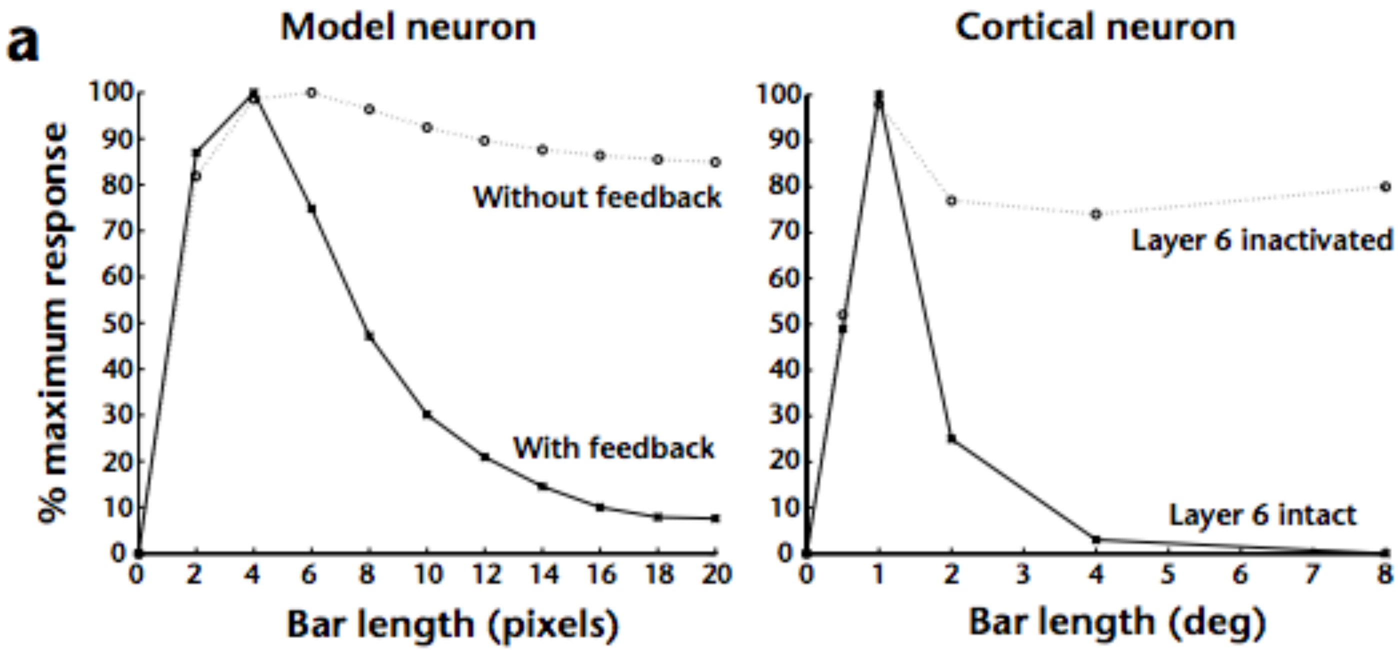
V2



V1



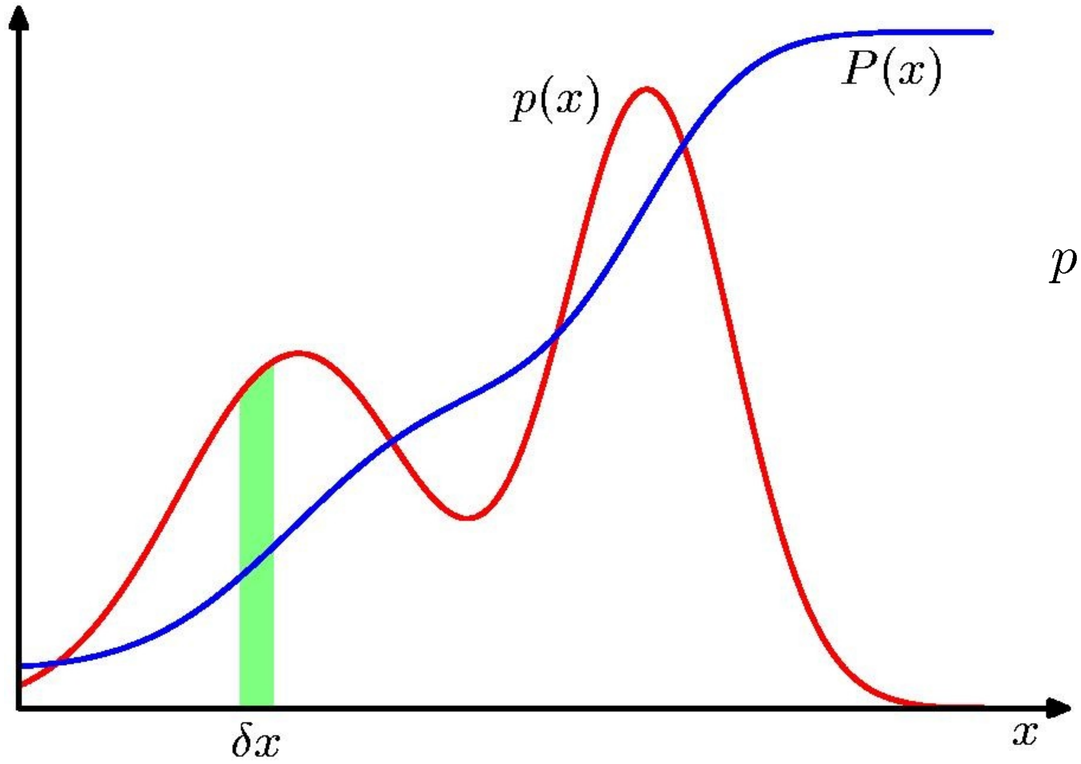
$$P(V1, V2) = P(V1|V2)P(V2)$$



Rao, R., Ballard, D. Predictive coding in the visual cortex:
a functional interpretation of some extra-classical receptive-field effects.
Nat Neurosci **2,** 79–87 (1999)

> 3,000 citations

Probability Densities



$$p(x \in (a,b)) = \int_a^b p(x)\,\mathrm{d}x$$

$$P(z) = \int_{-\infty}^{z} p(x)\,\mathrm{d}x$$

$$p(x) \geqslant 0 \qquad \int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1$$

Expectations

$$\mathbb{E}[f] = \sum_x p(x) f(x)$$

$$\mathbb{E}[f] = \int p(x) f(x) \, \mathrm{d}x$$

$$\mathbb{E}_x[f|y] = \sum_x p(x|y) f(x)$$

Conditional Expectation (discrete)

$$\mathbb{E}[f] \simeq \frac{1}{N} \sum_{n=1}^{N} f(x_n)$$

Approximate Expectation (discrete and continuous)

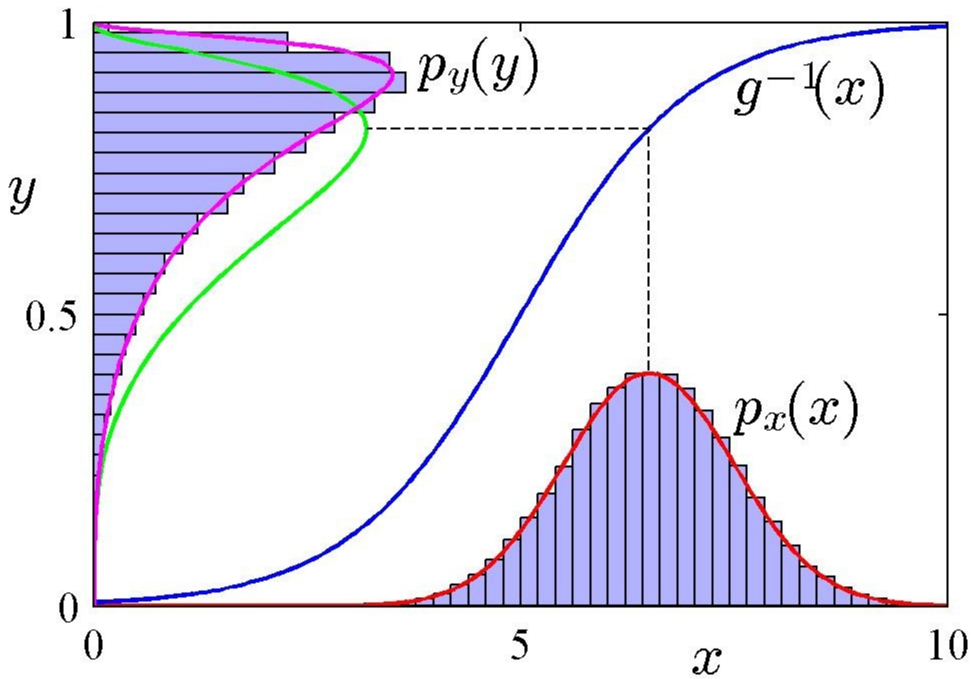
Variances and Covariances

$$\mathrm{var}[f] = \mathbb{E}\left[\left(f(x) - \mathbb{E}[f(x)]\right)^2\right] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

$$\begin{aligned}\mathrm{cov}[x, y] &= \mathbb{E}_{x,y}\left[\{x - \mathbb{E}[x]\}\{y - \mathbb{E}[y]\}\right] \\ &= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y]\end{aligned}$$

$$\begin{aligned}\mathrm{cov}[\mathbf{x}, \mathbf{y}] &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\}\{\mathbf{y}^{\mathrm{T}} - \mathbb{E}[\mathbf{y}^{\mathrm{T}}]\}\right] \\ &= \mathbb{E}_{\mathbf{x},\mathbf{y}}[\mathbf{x}\mathbf{y}^{\mathrm{T}}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}^{\mathrm{T}}]\end{aligned}$$

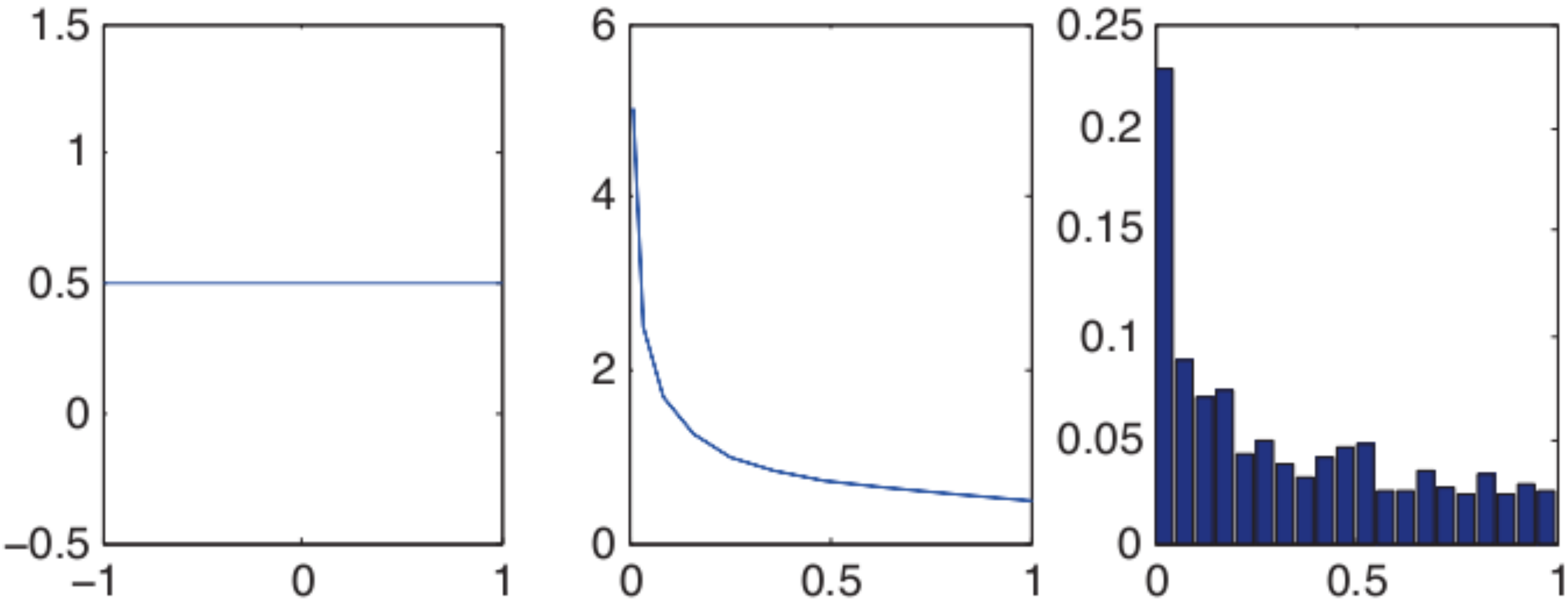
Transformed Densities



$$
\begin{aligned}
p_y(y) &= p_x(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| \\
&= p_x(g(y))\,|g'(y)|
\end{aligned}
$$

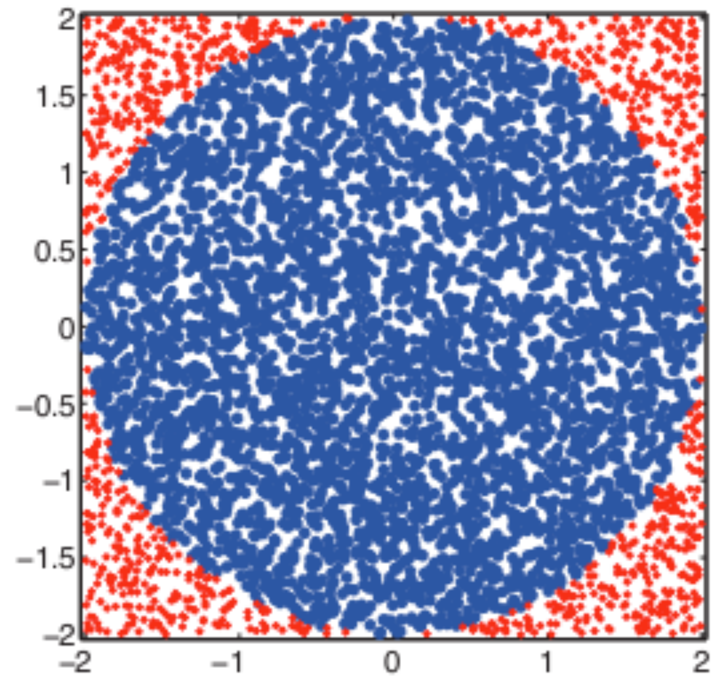
Transformed densities by sampling

e.g. p(y) given $y=x^2$ and p(x) is uniform

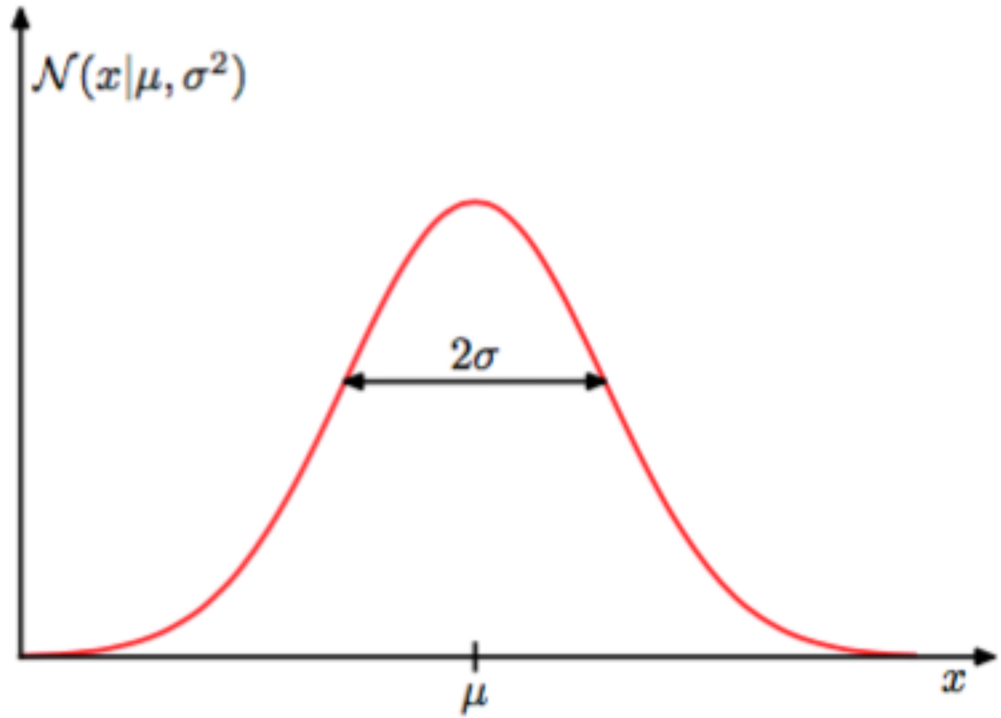


Monte Carlo sampling for π

$$
\begin{aligned}
I &= (2r)(2r)\int\int f(x,y)p(x)p(y)dxdy \\
&= 4r^2\int\int f(x,y)p(x)p(y)dxdy \\
&\approx 4r^2\frac{1}{S}\sum_{s=1}^{S} f(x_s, y_s)
\end{aligned}
$$

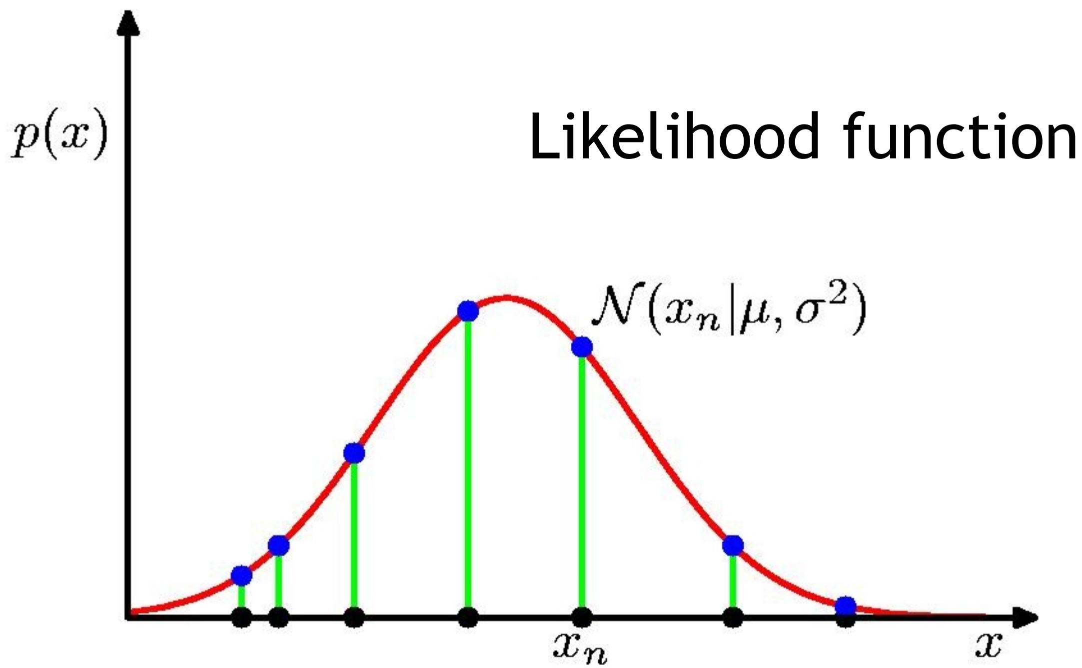


Gaussian (Normal) distribution



$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

Gaussian Parameter Estimation



$$p(\mathbf{x}|\mu, \sigma^2) = \prod_{n=1}^{N} \mathcal{N}\left(x_n|\mu, \sigma^2\right)$$

Gaussian parameters by Maximum (Log) Likelihood

$$\ln p\left(\mathbf{x}|\mu,\sigma^2\right) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi)$$

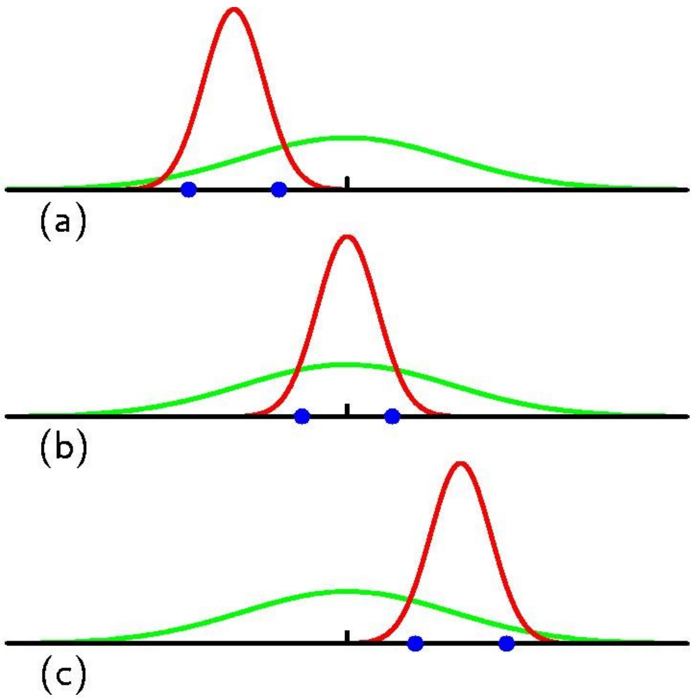
$$\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n \qquad \sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{\mathrm{ML}})^2$$

Properties of μ_{ML} and σ^2_{ML}

$$\mathbb{E}[\mu_{\mathrm{ML}}] = \mu$$

$$\mathbb{E}[\sigma^2_{\mathrm{ML}}] = \left(\frac{N-1}{N}\right)\sigma^2$$

$$\begin{aligned} \widetilde{\sigma}^2 &= \frac{N}{N-1}\sigma^2_{\mathrm{ML}} \\ &= \frac{1}{N-1}\sum_{n=1}^{N}(x_n - \mu_{\mathrm{ML}})^2 \end{aligned}$$



End of Lec 4: Basic Probabilities