Machine Learning and NeuroEngineering

机器学习与神经工程

Lecture 3 – Random-walk model for decision making

Quanying Liu （刘泉影）

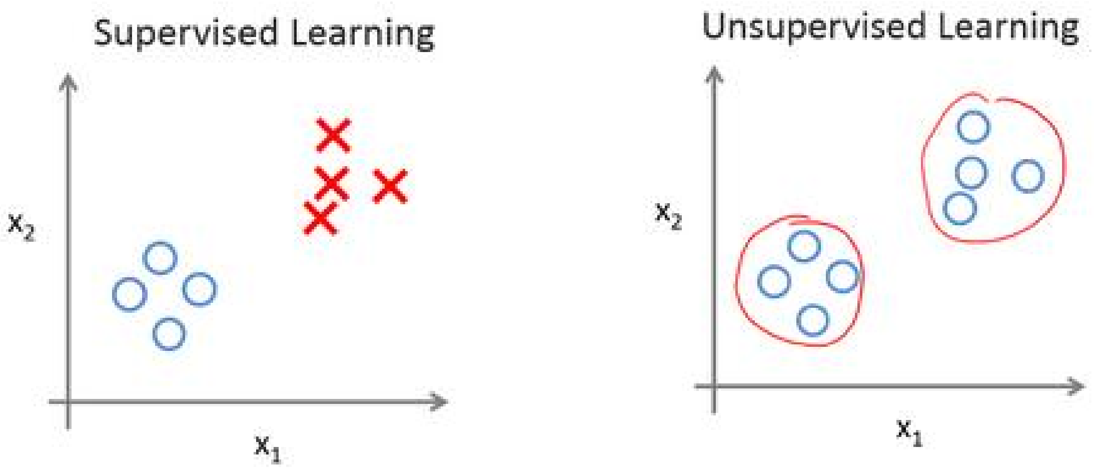
SUSTech, BME department

Email: liuqy@sustech.edu.cn

BMEB212, Spring term, 2021

Lecture 2 Recap

1. Artificial intelligence > Machine learning > Deep learning
2. Supervised & Unsupervised & Semi-supervised learning



Lecture 2 Recap

1. Artificial intelligence > Machine learning > Deep learning
2. Supervised & Unsupervised & Semi-supervised learning
3. Why do we need model? (The 4 points)

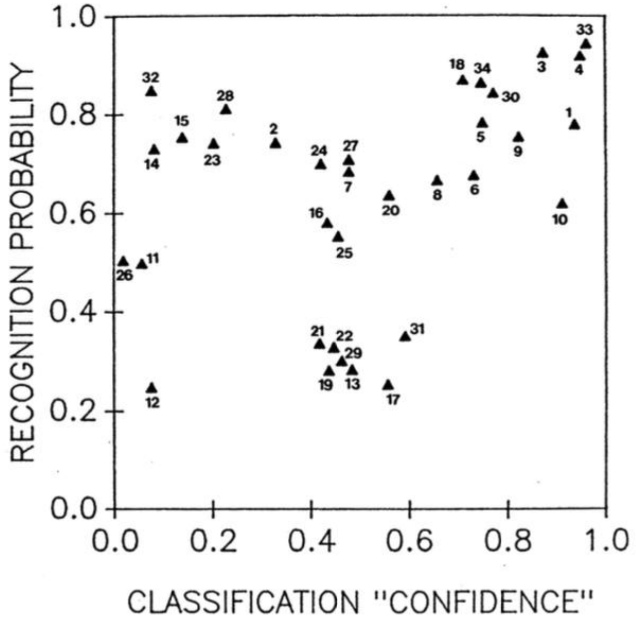
1) Data never speak for themselves. It requires a model to be understood and to be explained.
2) Verbal theorizing alone ultimately cannot substitute for quantitative analysis.
3) There are always several alternative models that vie for explanation of data and we must select among them.
4) Model selection rests on both quantitative evaluation and **intellectual and scholarly judgment**.

Lecture 2 Recap

1. Artificial intelligence > Machine learning > Deep learning
2. Supervised & Unsupervised & Semi-supervised learning
3. Why do we need model? (The 4 points)
4. Generalized Context Model

Models to study relationship:

- Correlation
- Linear regression
- Fitting a specific function
- Regression with neural network
- …



GCM: Step-by-step

$x_{i\,k}$
i: the index of face
k: the index of the feature

1. The **distance** between two faces (i&j):

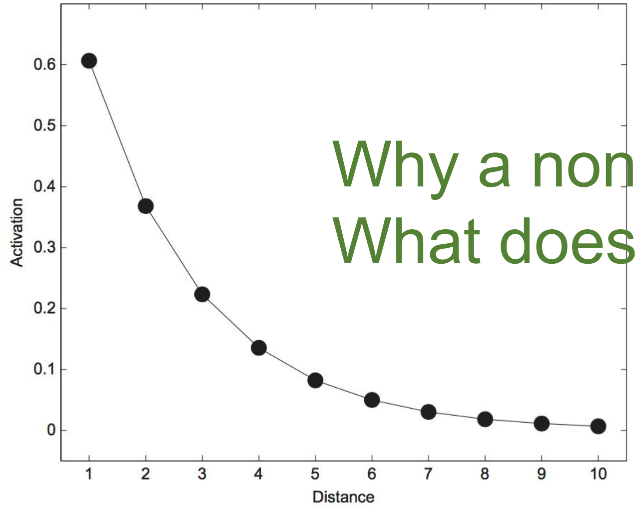
$$d_{ij} = \left(\sum_{k=1}^{K} |x_{ik} - x_{jk}|^2 \right)^{\frac{1}{2}}$$

2. The **similarity** between two faces:

$$s_{ij} = exp(-c \cdot d_{ij})$$

3. Response probabilities (to choose category A)

$$P(R_i = A|i) = \frac{\left(\sum_{j \in A} s_{ij}\right)}{\left(\sum_{j \in A} s_{ij}\right) + \left(\sum_{j \in B} s_{ij}\right)}$$

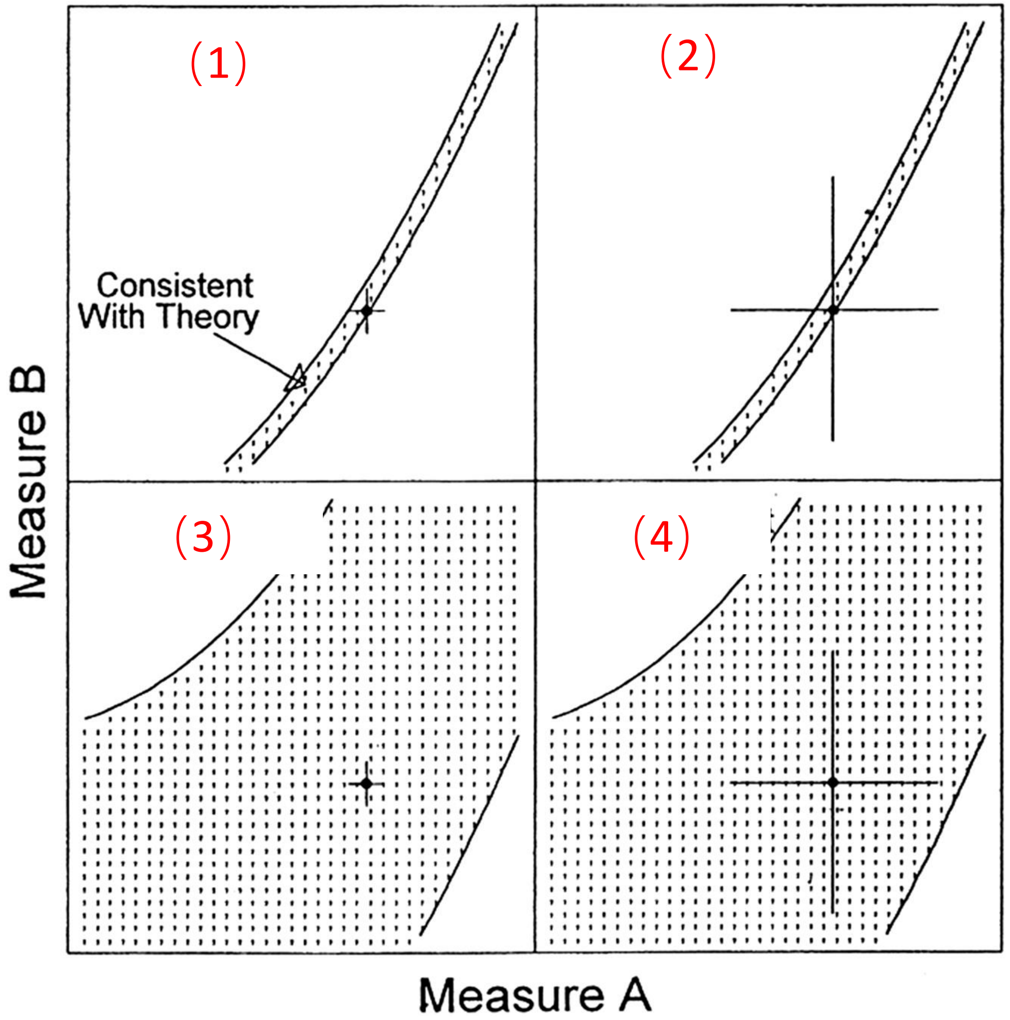


Why a nonlinear function?
What does it mean?

Lecture 2 Recap

1. Artificial intelligence > Machine learning > Deep learning
2. Supervised & Unsupervised & Semi-supervised learning
3. Why do we need model? (The 4 points)
4. Generalized Context Model
5. Scope and Falsifiability

Scope and Falsifiability



4 models:

The shadowed region is the zone of prediction from model.
The dot with error bar is the real data.

The real data supports which model the most?

(1)

Which two have better quality of data?

(1 and 3)

Lecture 3

1. A 2-choice decision making task
2. Random-walk model
3. R and RStudio
4. Implement Random-walk model with R
    - simulate data with random-walk model
    - explore the predictions of model
5. Trial-to-Trial Variability in the Random-Walk Model
6. Connecting model and data

Let’s do a cognitive task.

Please stare at the cross ‘+’ in the center of the monitor.

When the cross disappear, please Press the keyboard:

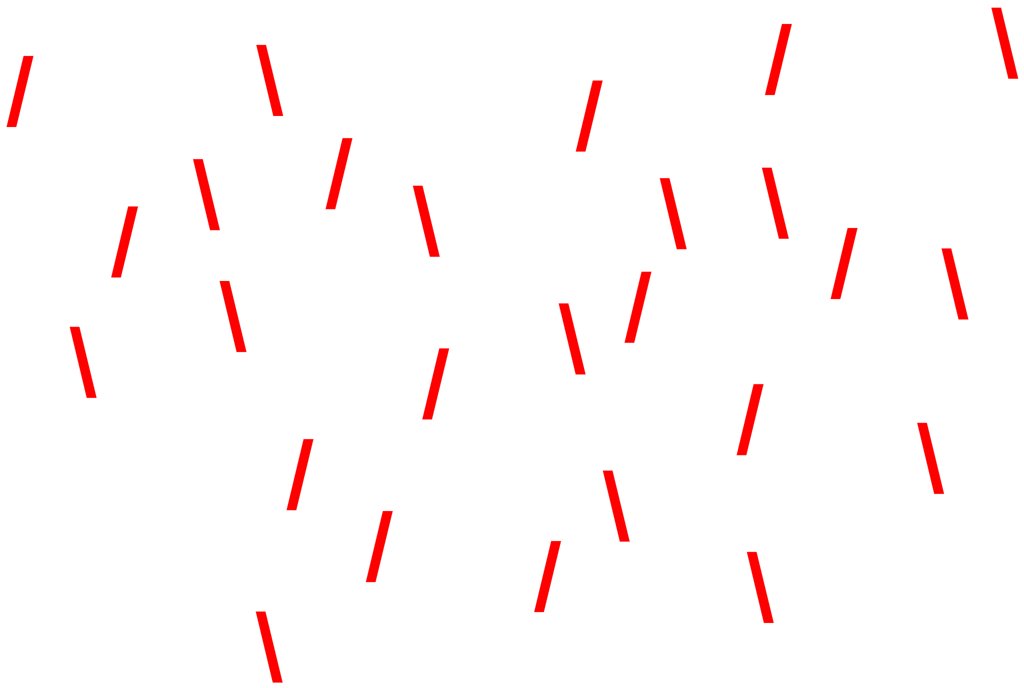
’f’(left) when you see more ’\’ ‘j’(right) when you see more ‘/’

You have to do it as soon as the ‘+’ disappears as possible, and as correct as possible.

Let’s start the task.

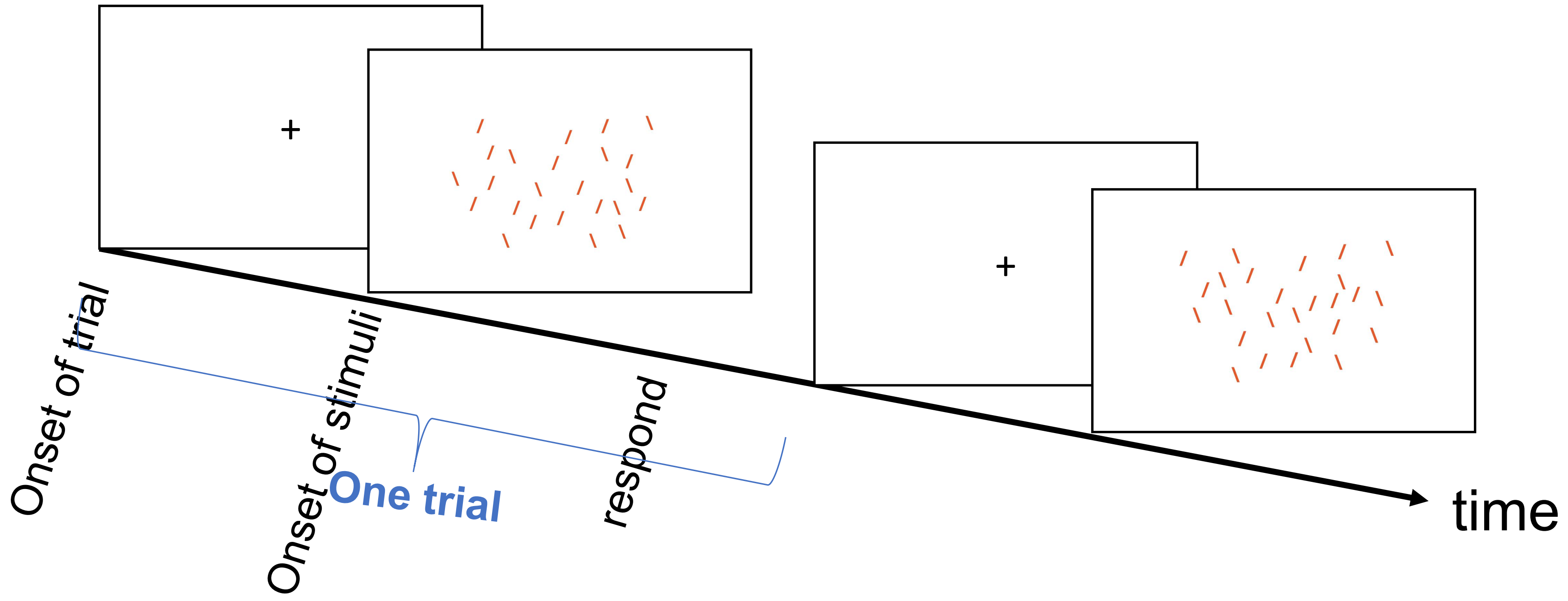
+

+



+

OK. To be short we finish the task now. But in the real cognitive experiment, we would do it hundreds of times.



We will record:

response time (onset of stimuli → you response to the stimuli)

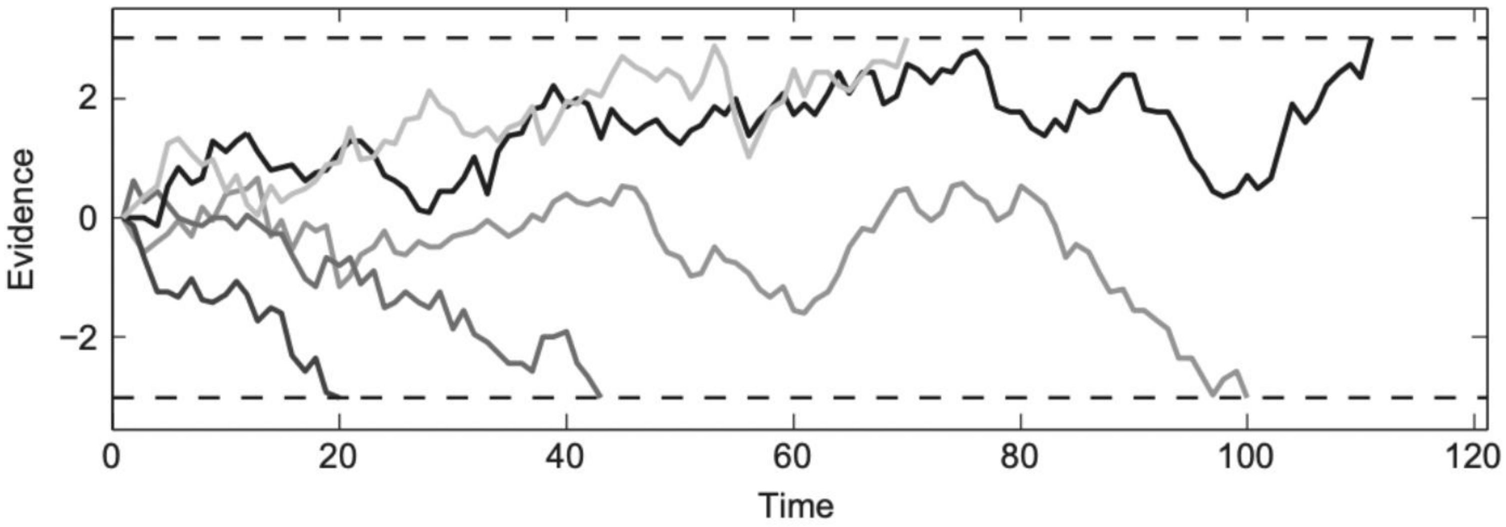
accuracy (whether your response is correctly)

Random-walk model

Assumption:

When a stimulus is presented, **not** all information is available to the decision maker instantaneously.

Instead, people gradually build up the evidence required to make a decision.

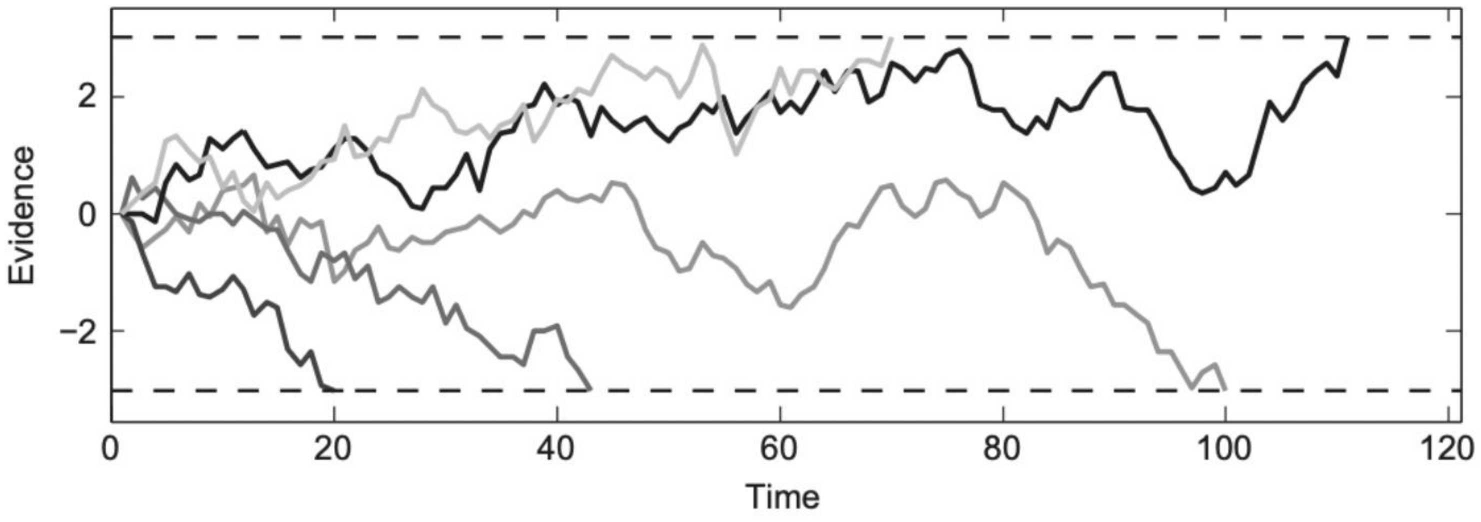


When the number of / and \ is equivalent.

Evidence is accumulated with time (sample steps).

Drift: the evidence from one sample step.

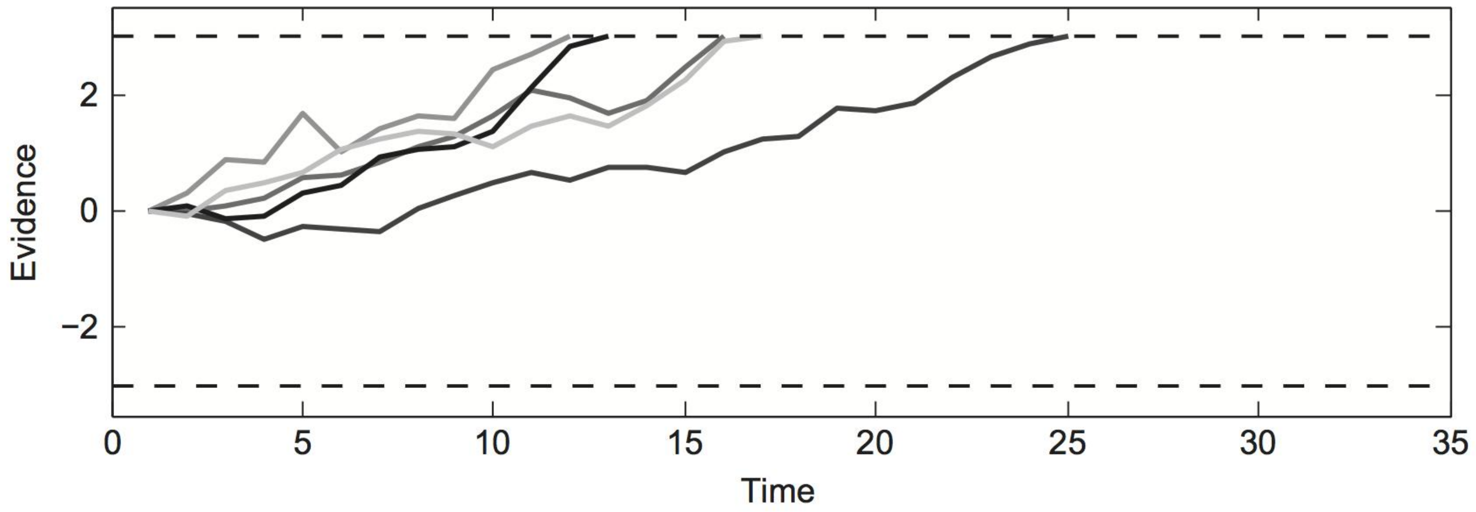
When the stimulus are non-informative, the probability to left and right is equal.



drift rate = 0

Do you take similar time to make decisions for left and right?
1(yes) or 2(no)

When the stimulus are informative, it favors one direction.



drift rate = 0.2

Does it take less time for the up boundary, compared to the bottom boundary? 1(yes) or 2(no)

R and RStudio

R is 'GNU S', a freely available language and environment for statistical computing and graphics.

It provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc.

It is cross-platform, suitable for Windows, Mac, and Linux.

Download and install R: http://cran.r-project.org/

RStudio (a free interface for R programming):

Download RStudio: www.rstudio.com/

Implement random-walk model with R

```
# random-walk model
nreps <- 10000      # number of trials
nsamples <- 2000  # samples to collect evidence for each decision

drift <- 0.0          # noninformative stimulus
sdrw <- 0.3          # s.t.d in the evidence
criterion <- 3       # distance between boundary and baseline

latencies <- rep (0, nreps)  # a vector to store the simulated response latencies
responses <- rep (0, nreps)  # a vector to store the simulated responses
evidence <- matrix (0, nreps, nsamples+1)  # a matrix to store the evidence for each response

for (i in c(1:nreps)) {
  evidence[i ,] <- cumsum ( c ( 0 , rnorm(nsamples , drift , sdrw) )  )
  p <- which ( abs (evidence[i, ] )>criterion) [1]  # find the first evidence bigger than criterion
  responses[i] <- sign (evidence[i,p])    # responses: positive (left) or negative (right)
  latencies[i] <- p }                      # response latency
```

Draw samples from a normal distribution

Plot the simulated data

```
## plot up to 5 random-walk paths
tbpn <- min(nreps, 5)
plot(1:max(latencies[1:tbpn])+10,type='n',las = 1,
     ylim = c(-criterion, criterion),
     ylab = 'Evidence', xlab='Decision time')
for (i in c(1:tbpn)){
  lines(evidence[i, 1:(latencies[i])])
}
abline(h=c(criterion, -criterion), lty='dashed')
```

```
## plot histograms of latencies
par(mfrow=c(2 , 1))
toprt <- latencies[responses>0]
topprop <- length(toprt)/nreps
hist(toprt, col='gray',
     xlab='Decision time', xlim=c(0,max(latencies)),
     main=paste('Top responses (',
as.numeric(topprop),')m=',
as.character(signif(mean(toprt),4)),
                sep=''),las=1)
botrt <- latencies[responses<0]
botprop <- length(botrt)/nreps
hist(botrt, col='gray',
     xlab='Decision time', xlim = c(0, max(latencies)),
     main=paste (‘Bottom responses (’, as.numeric
(botprop), ‘) m=’, as.character(signif(mean(botrt),4)) ))
```

Trial-to-trial variability

Trial-to-trial variability refers to changes in the values of parameters between different simulated trials.

It is based on the physical and psychological circumstances in an experiment do not remain invariant.

- Stimuli are encoded better or worse by our brain.
- We pay more or less attention.
- **Mistakes**: start the decision process before the stimulus is presented…

Variability in the starting value of the random walk

Variability in the drift rate

…

Trial-to-Trial Variability in the drift rate

```
#random walk model with unequal latencies between responses classes
nreps <- 1000
nsamples <- 2000

drift <- 0.03    # 0 = noninformative stimulus; >0 = informative
sdrw <- 0.3
criterion <- 3
t2tsd  <- c(0.0, 0.025)   # trial-to-trial s.d.

latencies <- rep(0,nreps)
responses <- rep(0,nreps)
evidence <- matrix(0, nreps, nsamples+1)

for (i in c(1:nreps)) {
  sp <- rnorm(1, 0.5, t2tsd[1])     # sampling starting point: mean = 0,  std = 0
  dr <- rnorm(1, drift, t2tsd[2])   # sampling drift rate: mean = drift,  std = 0.025
  evidence[i,] <- cumsum(c(sp, rnorm(nsamples, dr, sdrw)))
  p <-  which(abs(evidence[i,])>criterion)[1]
  responses[i] <- sign(evidence[i,p])
  latencies[i]  <- p }
```

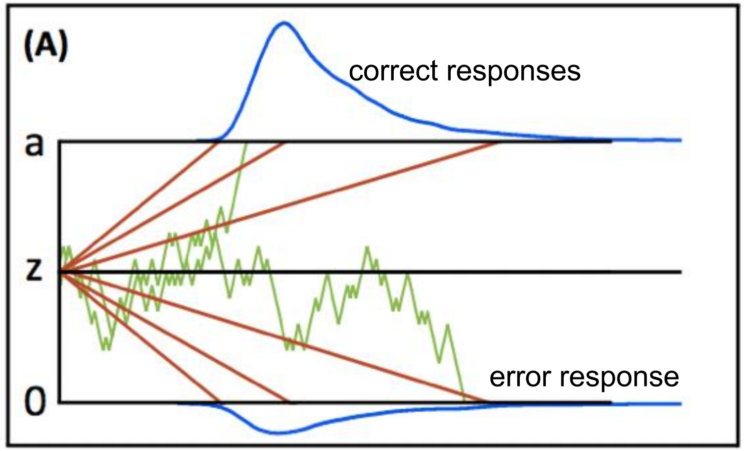
1 longer
2 shorter

Assuming both the starting point and drift rate are Gaussian distribution.

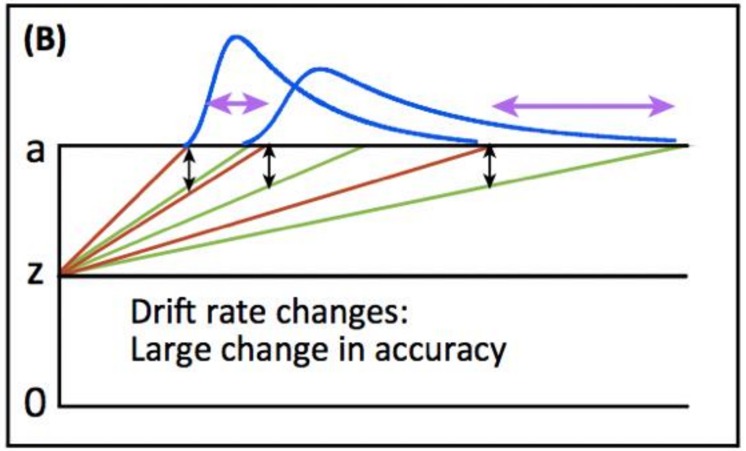
Drift-diffusion models

The parameters of the model are boundary separation (a), starting point (z), drift rate (v, one of each condition), nondecision time (Ter),

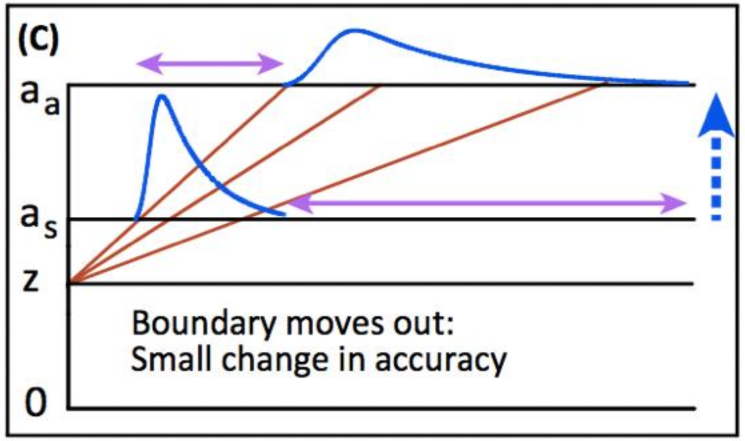
Two simulated paths in the diffusion model



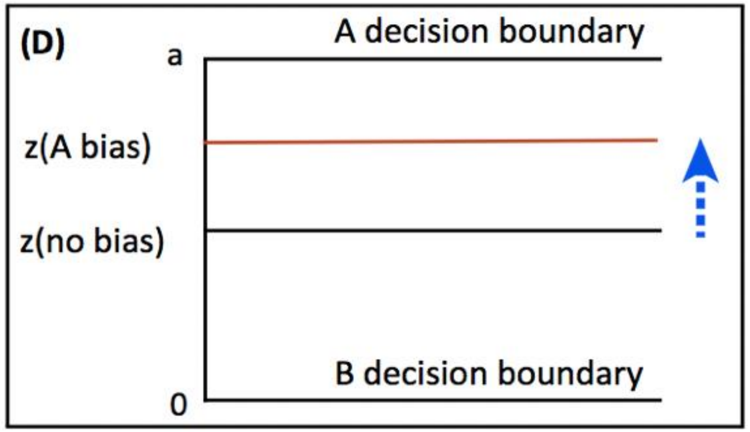
Effects of lowering drift rate by a fixed amount



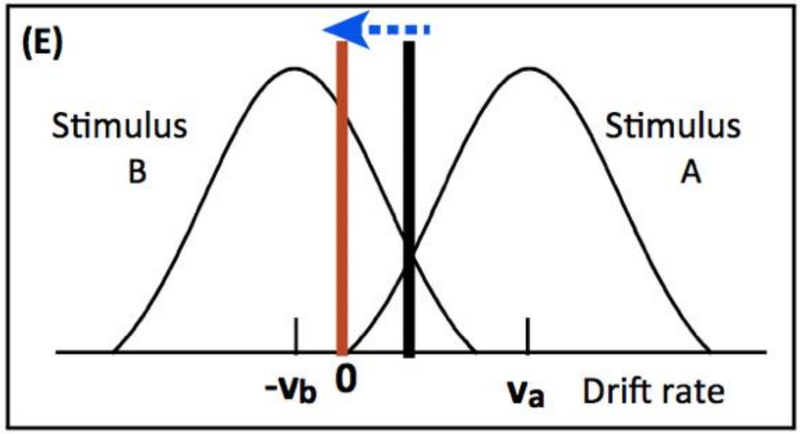
Effects of moving a boundary away



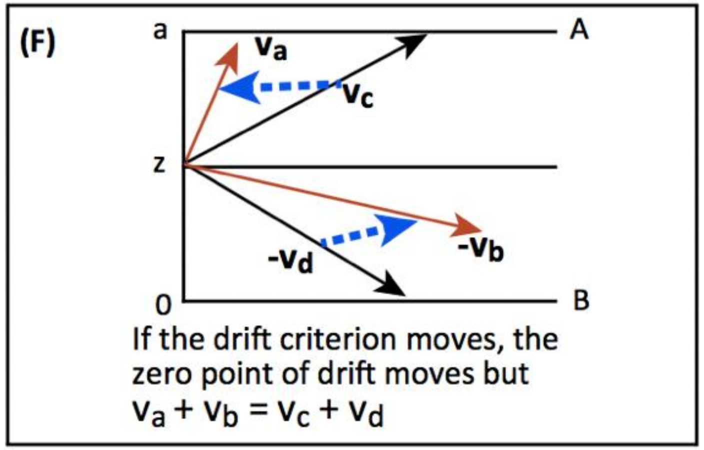
Change in the starting point (RT effect as C)



Bias towards A by changing the zero point of drift rate

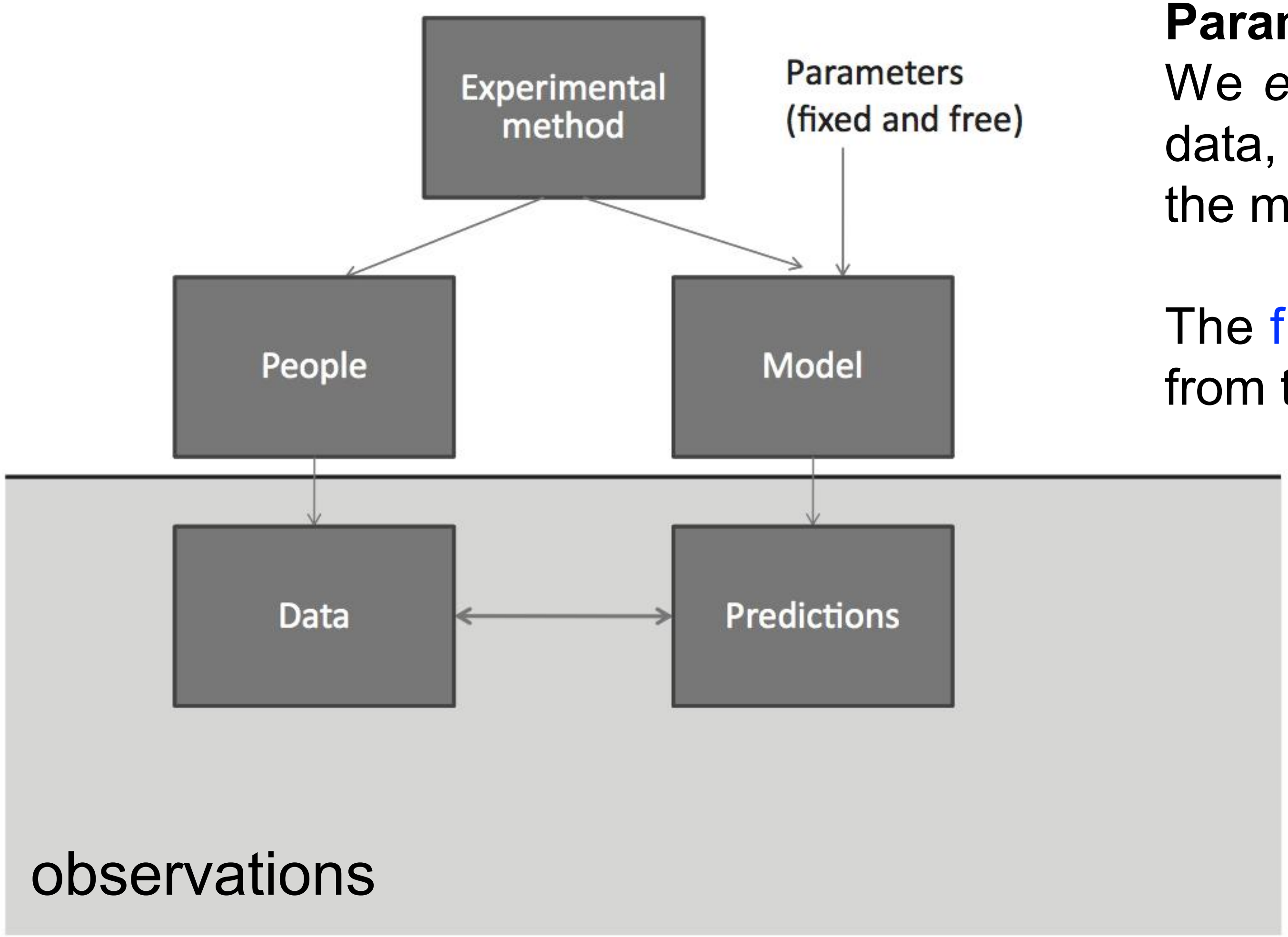


Effects of (E)



Ratcliff et al. (2016) Trends in Cognitive Science

Connecting Model and Data

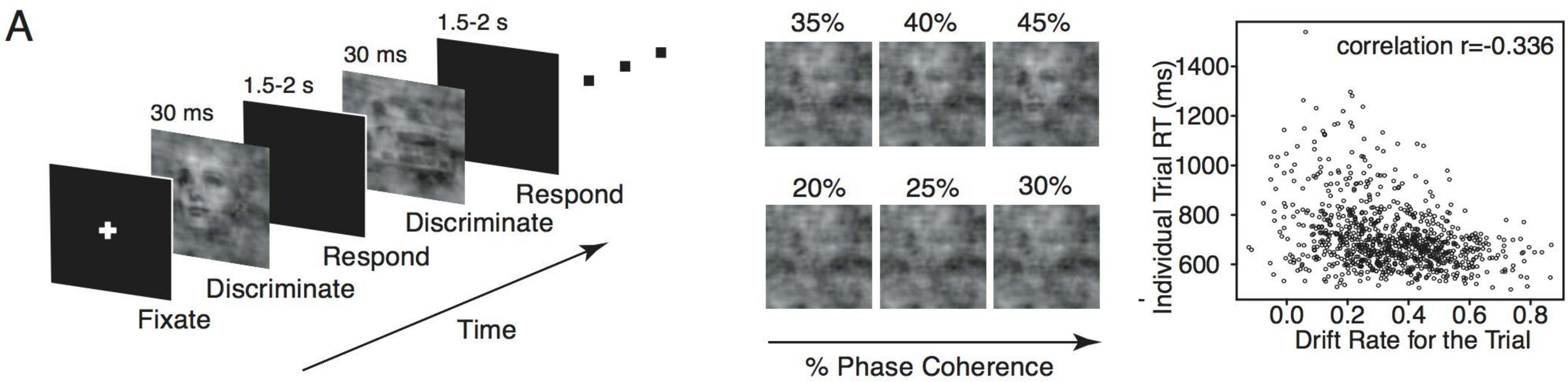


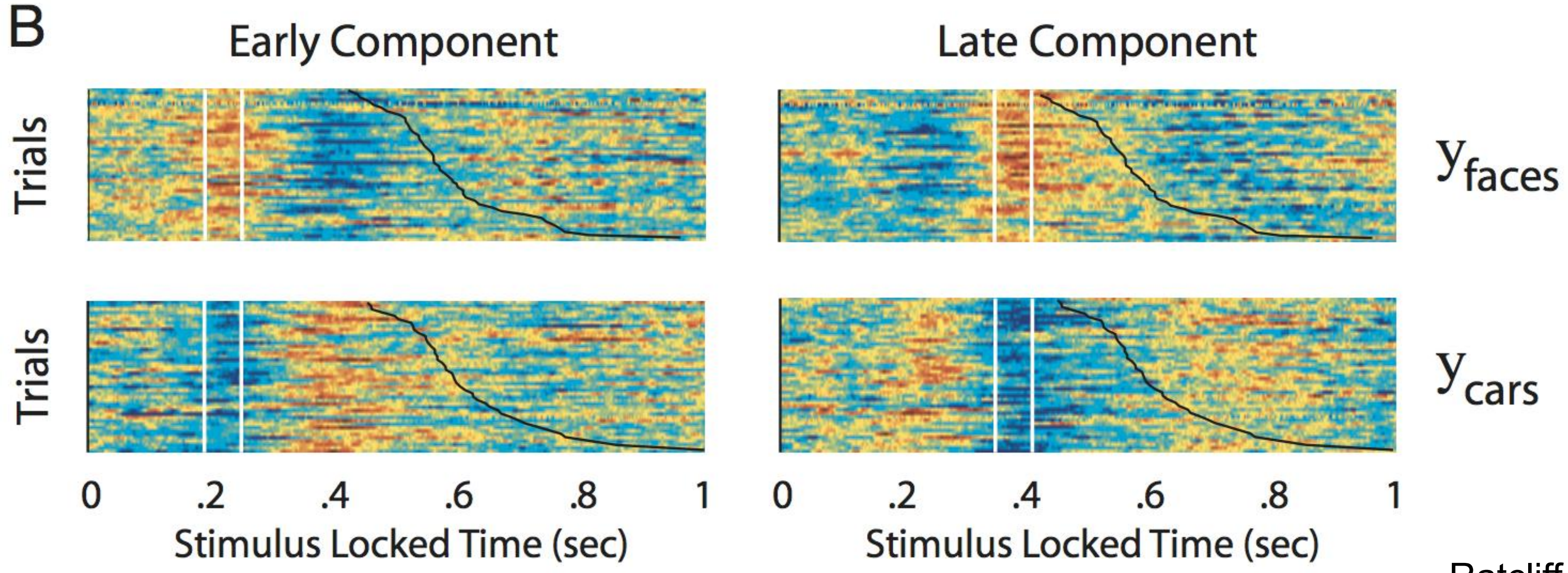
Parameters: free & fixed
We *estimate* the free parameters from the data, by finding those values that maximally fit the model's predictions with the data.

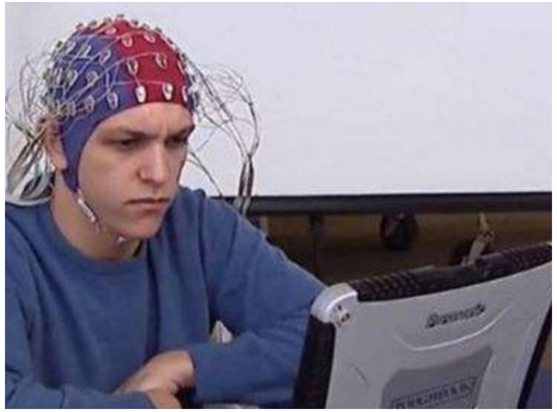
The fixed parameters, that are not estimated from the data, are invariant across datasets.

How to estimate free parameters?

We will introduce **the principles of parameter estimation** next lecture.



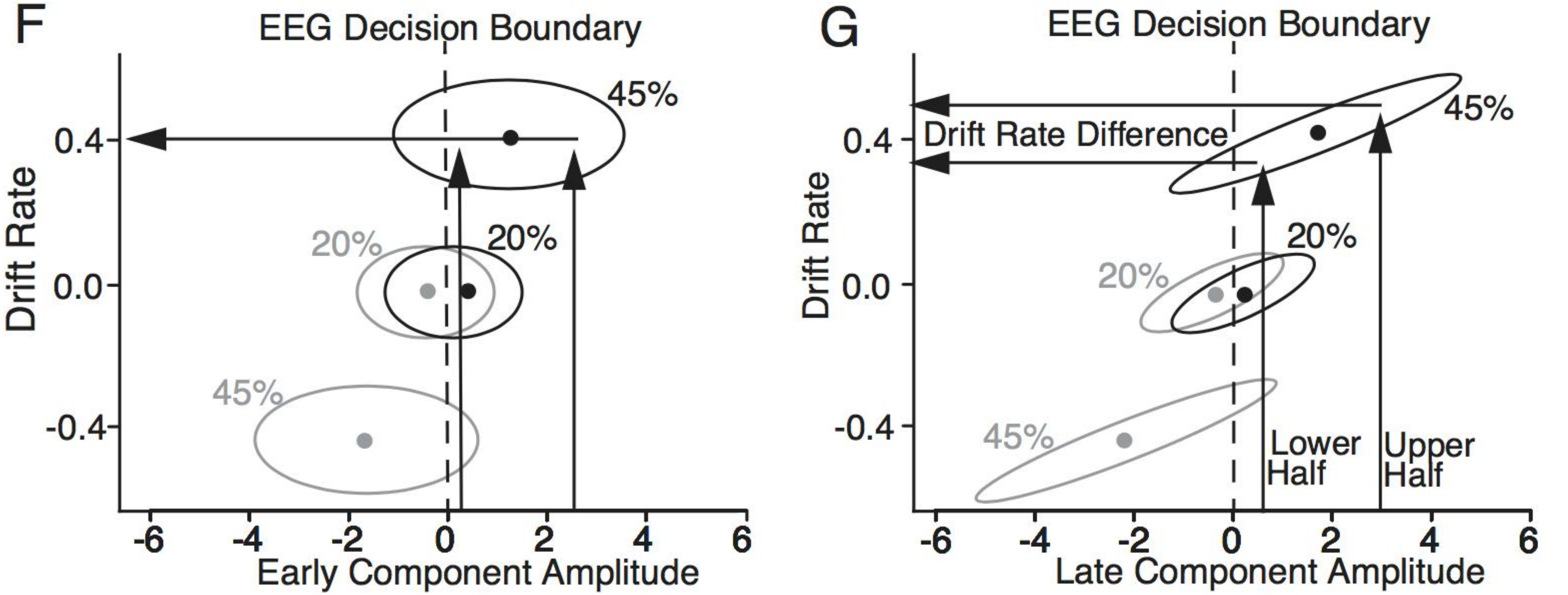




Ratcliff et al. (2009) PNAS

Fit the <u>EEG amplitude</u> and <u>drift rate</u> with **Bivariate Gaussian distributions.**

$$f(x,y) = \left(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right)^{-1} exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right]$$



Which one is associated with the drift rate?

1 (the early component of EEG) or 2 (late component of EEG)

Summary of Lecture 3

1. A 2-choice decision making task
2. Random-walk model
3. R and RStudio
4. Implement Random-walk model with R
    - simulate data with random-walk model
    - explore the predictions of model
5. Trial-to-Trial Variability in the Random-Walk Model
6. Connecting model and data

Recommended materials

Textbook

- Computational Modeling of Cognition and Behavior, Chapter 2

Must read.

Research Paper

- Roger Ratcliff, Marios G. Philiastides, and Paul Sajd. (2009), Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG, PNAS

Must read. Good for your to understand how to use models.

- Roger Ratcliff, Philip L. Smith, Scott D. Brown, and Gail McKoon (2016), Diffusion Decision Model: Current Issues and History, Trends in Cognitive Science

Not obliged. For fun.

Homework 1

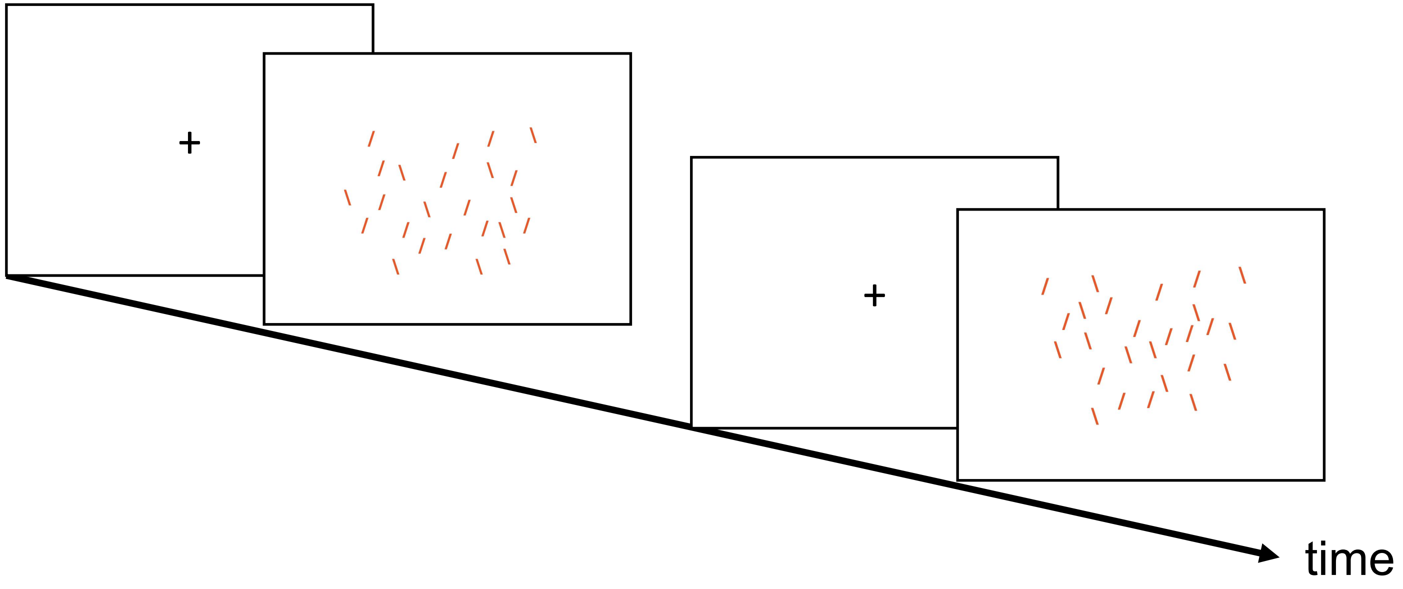
Tip: An example on YouTube
https://www.youtube.com/watch?v=i6xMBig-pP4

Some code is available at github

Implement the task with Python （可以用PyGame）

DDL: March 8, 2021

(上课前，把code发邮件给曲由之quyz@mail.sustech.edu.cn，课上抽查)



Requirements

1. **60 trials in total**
2. The **number** of \ and / is sampled from a uniform distribution [20, 40] and cannot be equal.
3. Randomize the **locations** of \ and /
4. **Record the following information**

1) Subject ID	5) The number of \ in a trial
2) Trial ID	6) The number of / in a trial
3) Time: onset of stimulus	7) The response: left or right
4) Time: response is made	8) Response time: time in 4) – time in 3)

输出到一个csv文件
一行就是一个trial的信息
每一行有8列

Quiz

1 or 2

When we are under time pressure,
the error responses are **faster** than correct responses?

1 (yes) or 2 (no)

1

The error responses are always **faster** than correct responses?

1 (yes) **or** **2 (no)**

2

when the task is difficult and no time pressure

When we made **fast** errors, the drift rates of error responses are **bigger** than the drift rates of correct responses?

1 (yes) **or** **2 (no)**

2

lower starting point

Learning will increase the drift rate, shift the starting point and the boundary.

1 (yes) or 2 (no)

I am not sure. We need study it.