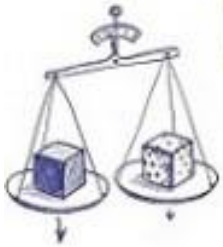


Usher Lab

Decisions

Cognition

Neural Computation

TEL AVIV UNIVERSITY אוניברסיטת תל אביב

Evidence Accumulation Modeling

Marius Usher & Moshe Glickman

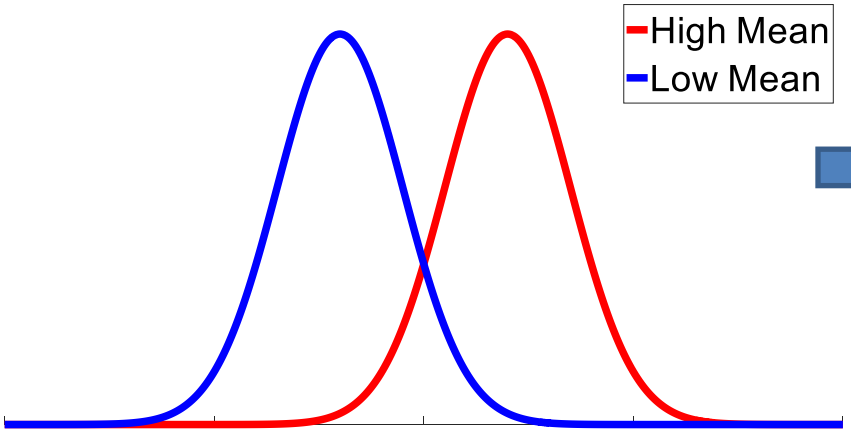
The 5th Israeli Conference on Cognition Research

February, 2018

Why do we need modelling?

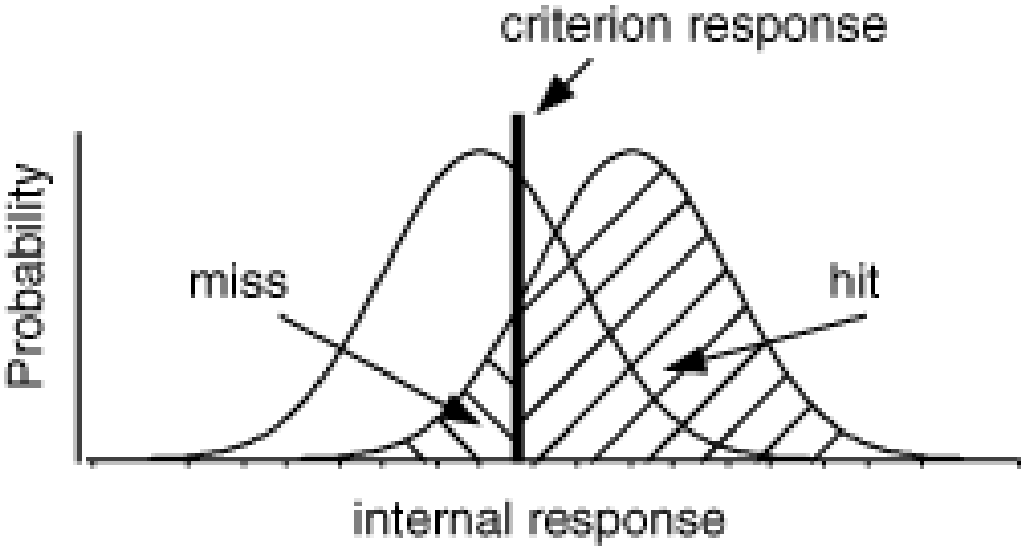
Dependent variables: choice & RT

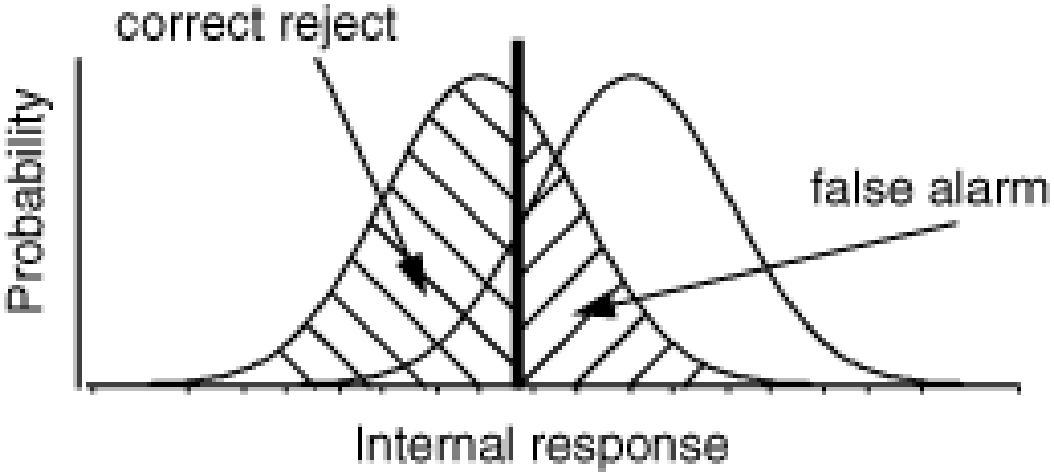
Tasks: perceptual choice, lexical-decision, VS, memory-recognition, etc.



- Extract performance level from data? (not ignore RT, correct/incorrect, and distributions); account for SAT.
- Extract latent processes: Different latent variables may show the same RT and accuracy (patients/aging/divided attention)

Signal-Detection theory





- Response bias & Sensitivity
- Better performance if taking longer to decide: speed-accuracy tradeoff (SAT); looking for more evidence takes time.
- Evidence-accumulation is a generalization of SD to multiple samples
- Accounts for SAT

Evidence integration: the problem

You are faced with noisy samples of evidence and need to decide which out of a set of perceptual hypotheses (H1,H2, H3...) gives the best match

Start with “priors” P1(0),P2(0),P3(0). Then take evidence samples, d1(t1), d(t2), d(t3) (noisy); update the posterior probabilities and compute ratio likelihood: P(H1/D)/P(H2|D)

H1, H2= hypotheses , D = observed data

Bayes Rule

P(H1 | D) = P(H1 and D)/P(D)

P(H2 | D) = P(H2 and D)/P(D)

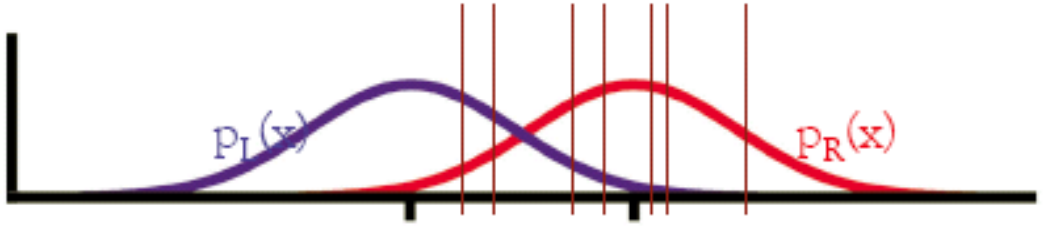
LR(t+1) = P(H1/D)/P(H2|D) =[P(D|H1)/P(D|H2)] *[P(H1)/P(H2)] [t]

- Keep going until LR(t) = accuracy criterion

An optimal decision procedure for noisy data:

the Sequential Probability Ratio Test

Mathematical idealization: During the trial, we draw noisy samples from one of two fixed distributions $p_L(x)$ or $p_R(x)$ (left or right-going dots).



The SPRT works like this: set up two thresholds $1/B$ and B and keep a running tally of the ratio of likelihood ratios:

$$R_n = \left(\frac{p_L(x_n)}{p_R(x_n)}\right) \times \ldots \times \left(\frac{p_L(x_2)}{p_R(x_2)}\right) \times \left(\frac{p_L(x_1)}{p_R(x_1)}\right)$$

When R_n first exceeds B or falls below $1/B$, declare victory for R or L.

Theorem: (Wald, Barnard) Among all fixed sample or sequential tests, SPRT minimizes expected number of observations **n** for given accuracy.

For fixed *n* & *B=1* SPRT maximizes accuracy (Neyman-Pearson lemma).

Signal detection with multiple samples of evidence

Likelihood ratio with multiple evidence, e1, e2, ...

$$LR_{1,2|e_1,e_2,\ldots,e_n} = LR_{1,2|e_1} \cdot LR_{1,2|e_2} \ldots \cdot LR_{1,2|e_n}$$

Decision rule

$$LR_{1,2|e_1} \cdot LR_{1,2|e_2 \ldots} \cdot LR_{1,2|e_n} \cdot \frac{\Pr(h_1)}{\Pr(h_2)} \cdot > 1$$

Take Logs

$$\log LR_{1,2|e_1} + \log LR_{1,2|e_2 \ldots} + \log LR_{1,} + \log\left[\frac{\Pr(h_1)}{\Pr(h_2)}\right] \cdot > 0$$

Example: signal detection in the brain

Gold & Shadlen, 2001; TICS

http://www.shadlen.org/~mike/papers/mine/gold_shadlen2001c.pdf

Detect light/no-light

evidence(e) n-spikes/50 ms

450 light trials: signal

450 no-light trials: noise

Plot signal noise distribution

Decision criterion:

Ratio-likelihood:

$r = P(e|h1)/P(e|h2) > 1$

Bayes rule:

$P(e|h1)/P(e|h2) * P(h1)/P(h2) = \mathbf{P(h1|e)/P(h2|e)}$

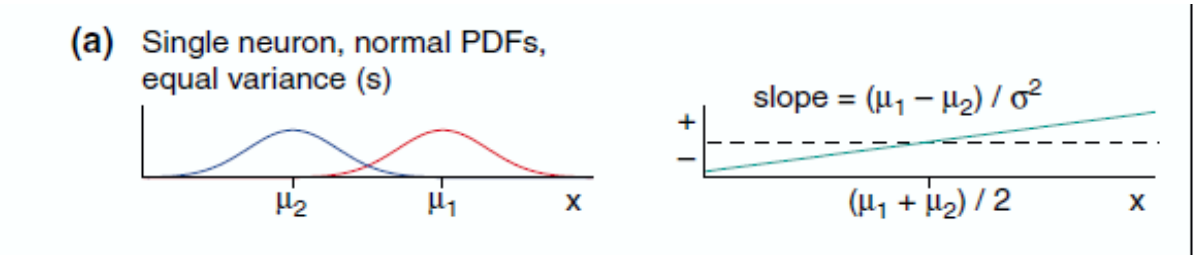
Box 1. Using the likelihood ratio

A hypothetical light detector can indicate a value (*e*) of 0 to 9 in the presence or absence of light. As indicated in Table I, the detector tends to indicate higher values in the presence of light. Columns 2 and 3 indicate the number of trials in which each value *e* was indicated in a block of 450 'light-present' trials and 450 'light-absent' trials, respectively. Columns 4 and 5 convert these counts into conditional probabilities, or likelihoods. The ratio of these likelihoods ($LR_{1,2|e}$) indicates whether it was more likely to be true that light was present or that light was absent for each given *e*. Specifically, when $LR_{1,2|e} > 1$, 'present' was more likely. Therefore, to use the detector, read the value *e* and then decide 'present' if $LR_{1,2|e} > 1$. This is equivalent to deciding 'present' if the value $e \geq 5$.

Table I. Calculating the light likelihood for a hypothetical light detector

Detector value (*e*)	No. light-present trials (h_1)	No. light-absent trials (h_2)	$Pr(e|h_1)$	$Pr(e|h_2)$	$LR_{1,2|e}$
0	0	90	0.00	0.20	0.0
1	10	80	0.02	0.17	0.1
2	20	70	0.04	0.15	0.3
3	30	60	0.06	0.13	0.5
4	40	50	0.08	0.11	0.8
5	50	40	0.11	0.08	1.3
6	60	30	0.13	0.06	2.0
7	70	20	0.15	0.04	3.5
8	80	10	0.17	0.02	8.0
9	90	0	0.20	0.00	inf

How does the brain do it?
Gold & Shadlen, 2001; TICS



$$LR_{1,2|x} = \frac{\exp\left[-\frac{1}{2\sigma^2}(x-\mu_1)^2\right]}{\exp\left[-\frac{1}{2\sigma^2}(x-\mu_2)^2\right]}$$

$$\log LR_{1,2|x} = -\frac{1}{2\sigma^2}\left[(x-\mu_1)^2-(x-\mu_2)^2\right]$$

$$= -\frac{1}{2\sigma^2}\left[2x(\mu_2-\mu_1)+\mu_1^{\,2}-\mu_2^{\,2}\right] \qquad x > (\mu_1+\mu_2)/2 \text{ (i.e. when } \log LR > 0)$$

Neural response, x, approximates Log-likelihood

How does the brain do it?
Gold & Shadlen, 2001; TICS

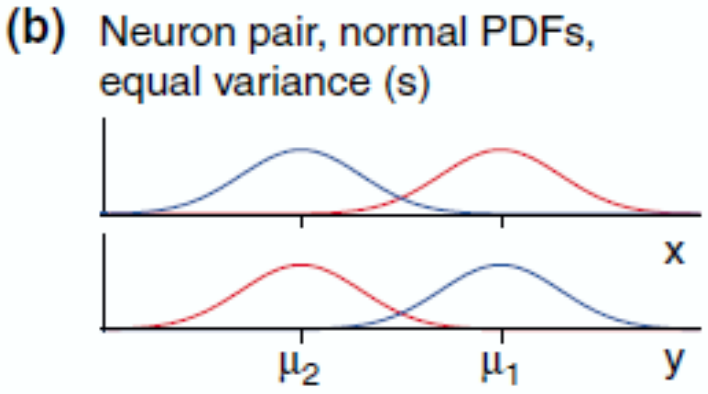
Use 2 detectors, one for h1, and one for h2

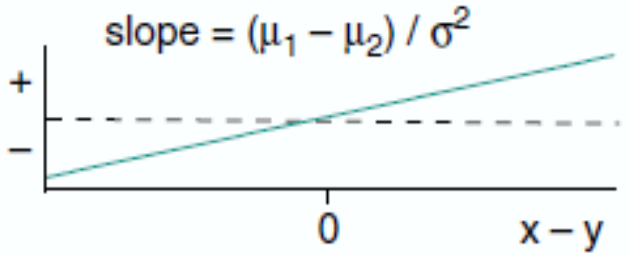
$$\log LR_{1,2|x} = -\frac{1}{2\sigma^2}[2x(\mu_2 - \mu_1) + \mu_1^{\,2} - \mu_2^{\,2}]$$

$$\log LR_{1,2|y} = -\frac{1}{2\sigma^2}[2y(\mu_1 - \mu_2) + \mu_2^2 - \mu_1^2]$$

$$\log LR_{1,2|x,y} = \log LR_{1,2|x} + \log LR_{1,2|y}$$

$$\log LR_{1,2|x,y} = \frac{\mu_1 - \mu_2}{\sigma^2}(x - y)$$

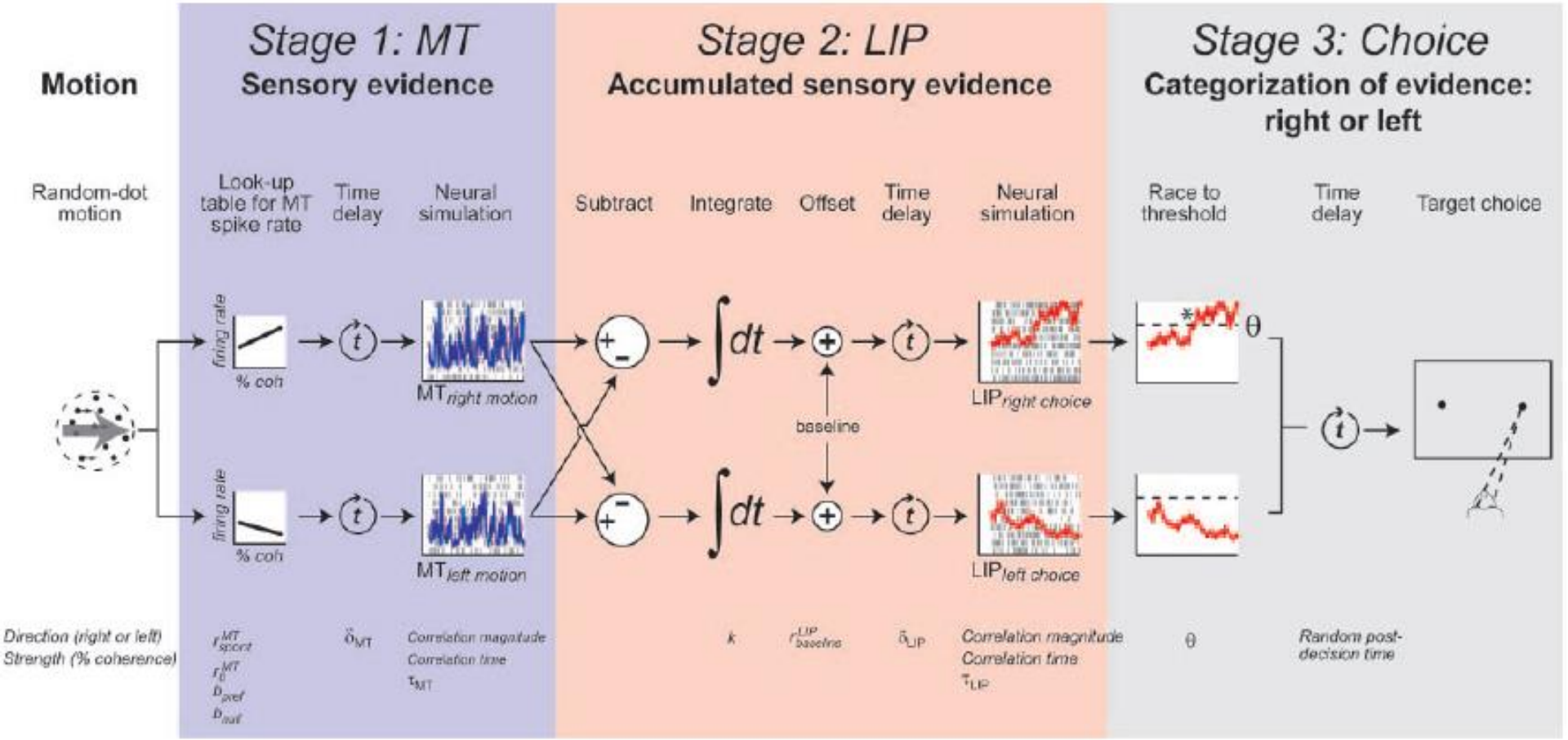




x-y, approximates Log-likelihood

Neuroscience model of perceptual decisions

(Mazurek, Roitman, Ditterich & Shadlen, 2003; *Cerebral Cortex)*

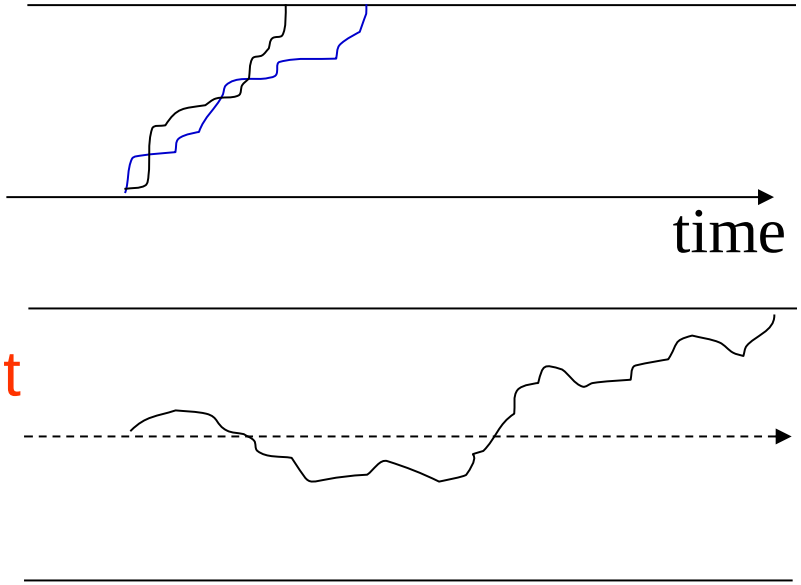


Mathematical models of choice

Task: choice on the basis of noisy *evidence* + stochastic accumulation of information.

Integrate evidence *towards response-criterion*

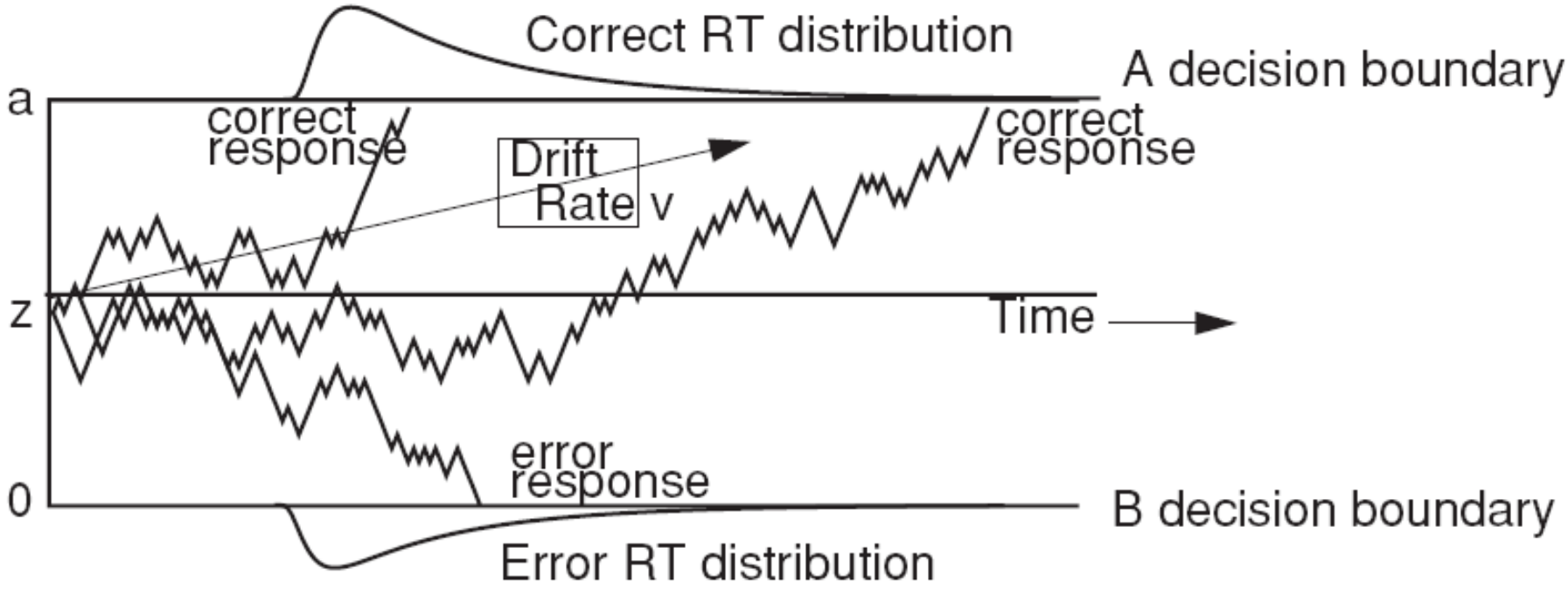
- ***Accumulators/race*** towards common response-criterion: flexible but not-efficient
- ***Drift-Diffusion*** model based on relative evidence: efficient but difficult to generalise to n-choice



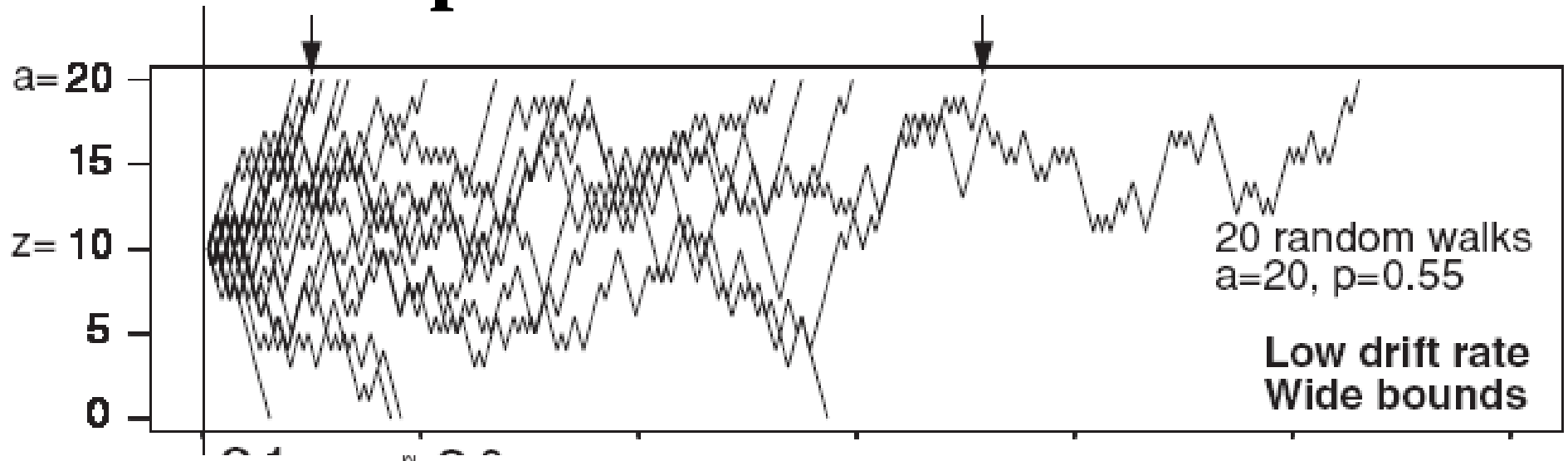
Neural competition model (Usher &McClelland, 2001)

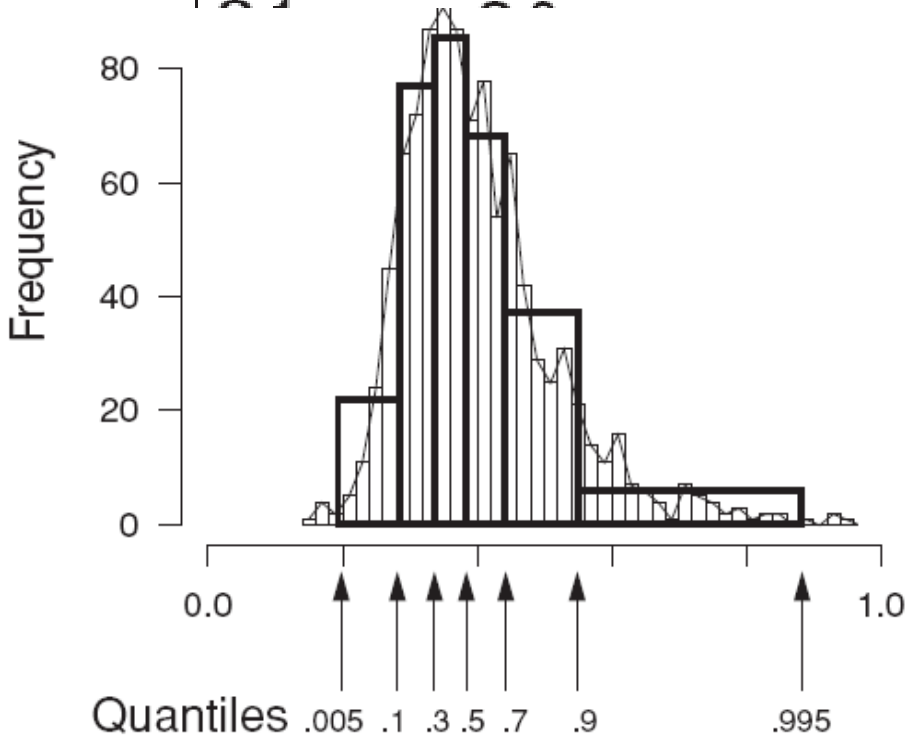
From signal-detection to 2-choice-RT random-walk/diffusion models (Ratcliff & McKoon, 2008)

- Take multiple samples x1, x2, … xn and compute $y=\Sigma(x_i - c)$
- y1=0, y2=(x1-c); y3= y2+ (x2-c); …. (y goes up or down with evidence)
- IF yn > A (respond YES); n is response time
- IF yn < -B (respond NO); A,B is response-time criteria;
- More samples takes longer but helps to average out the noise: SAT.



Response time distribution





Diffusion model:

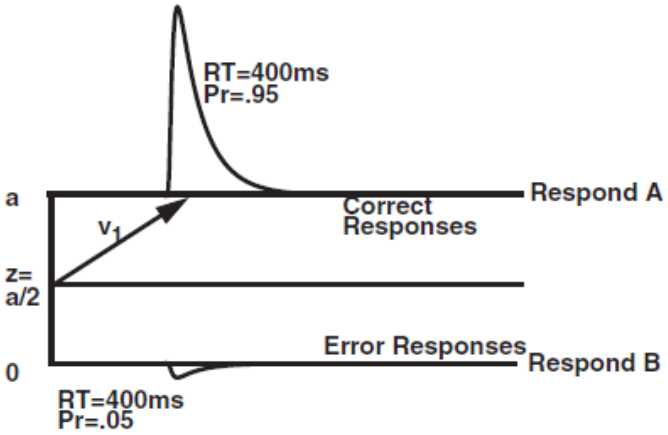
- RT_correct=RT_incorrect

To account for asymmetry add:

- Starting point variance
- Variance in drift

Diffusion model: results

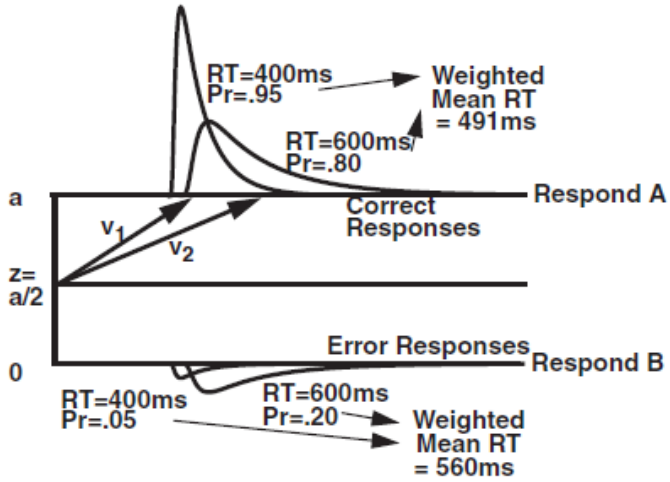
Basic model predicts same RT-distribution for correct and incorrect responses



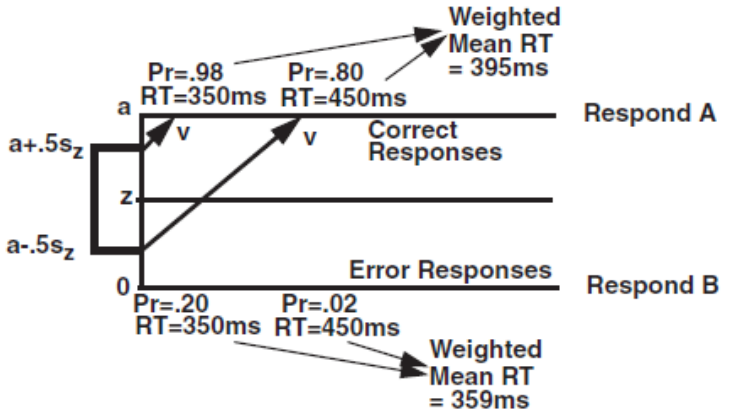
Experimental-RT are not always the same for correct vs incorrect

Modification of the diffusion model can account for differences:

- Starting point variability: RT(errors) < RT(corrects)



- Variability in drift from trial to trial): RT(errors) > RT(corrects)

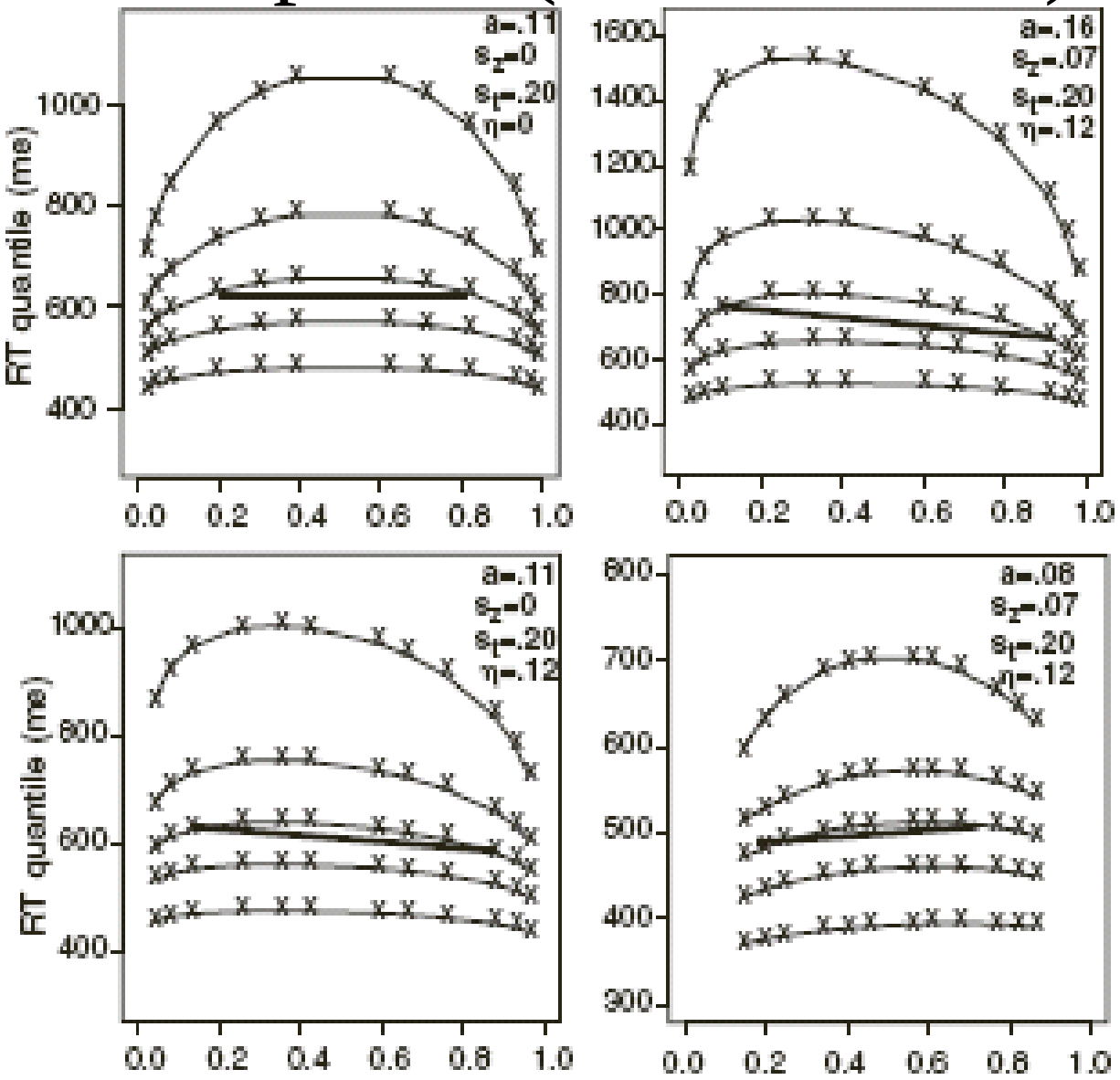


Diffusion model: latency as a function of signal/noise and response (correct/error)

Ratcliff & McKoon, 2008 (*Neural Comp.*) Quantile-RTs:

Latency-probability functions

- Stimulus difficulty: ***drift rate***



Summary on diffusion/race

Accumulators: no stochastic accumulation.

- Variability due to variance in starting point.
- Increasing the support for weak option → faster RT
- Non-optimal decision

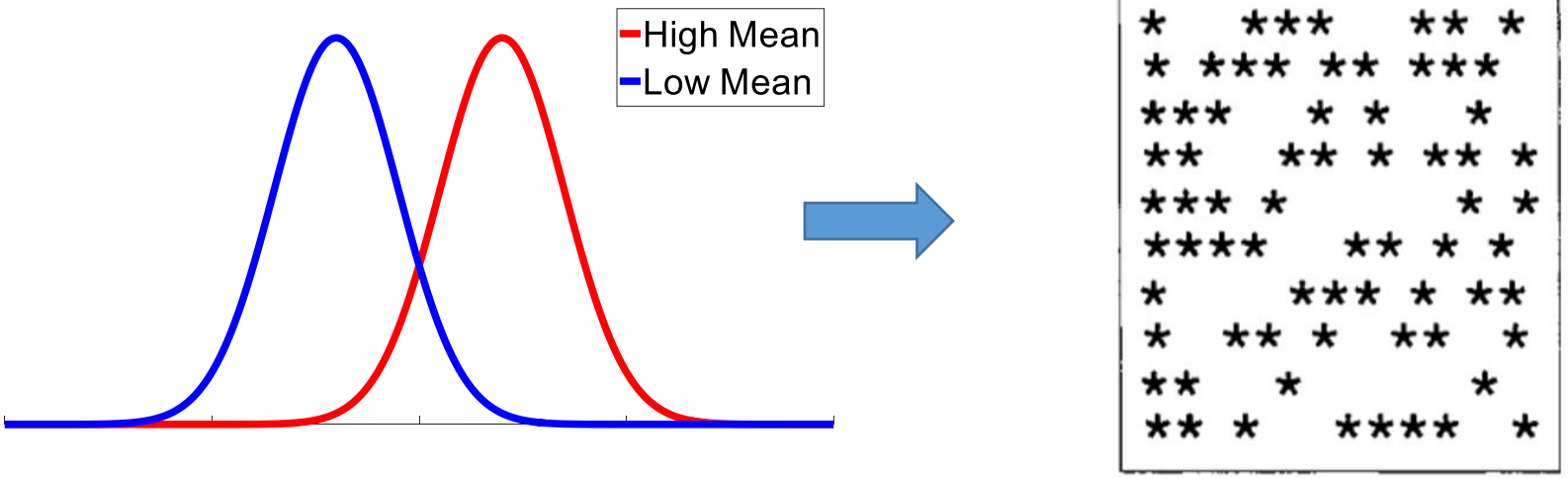
Diffusion (random-walk)

- Predicts equal RT for correct and incorrect RT.
- To account for unequal correct/errors, assumes variance in starting point and in drift across trials.
- Optimal decision.

How to extract latent variables from data

- Data fitting illustration
- The Fast-DM program
- Do it yourself - Maximum-Likelihood estimation

Data Fitting Illustration

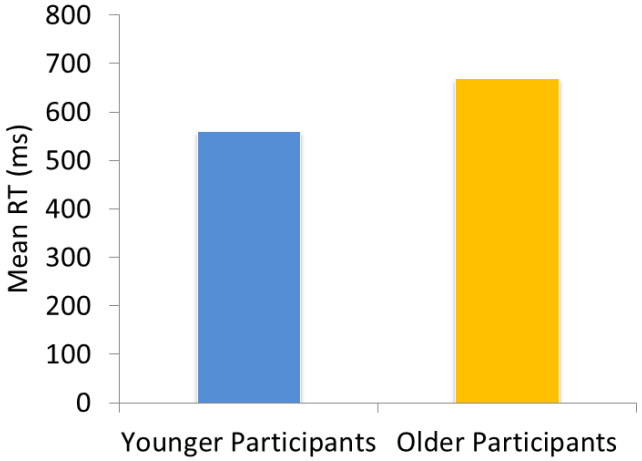


- **Task**
  - Decide whether the display originated from a **high** or **low** mean distribution
- **2 Groups**
  - Younger participants (college age)
  - Older Participants (60 – 74 years old)

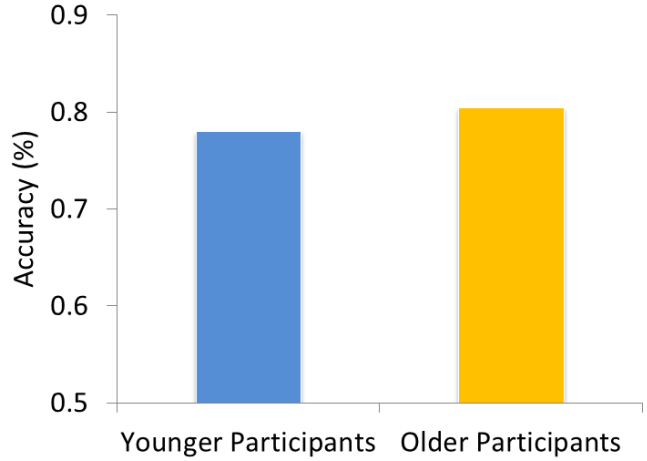
Ratcliff et al., 2006

RT & Accuracy Results

The older participants were slower as compared to the younger ones

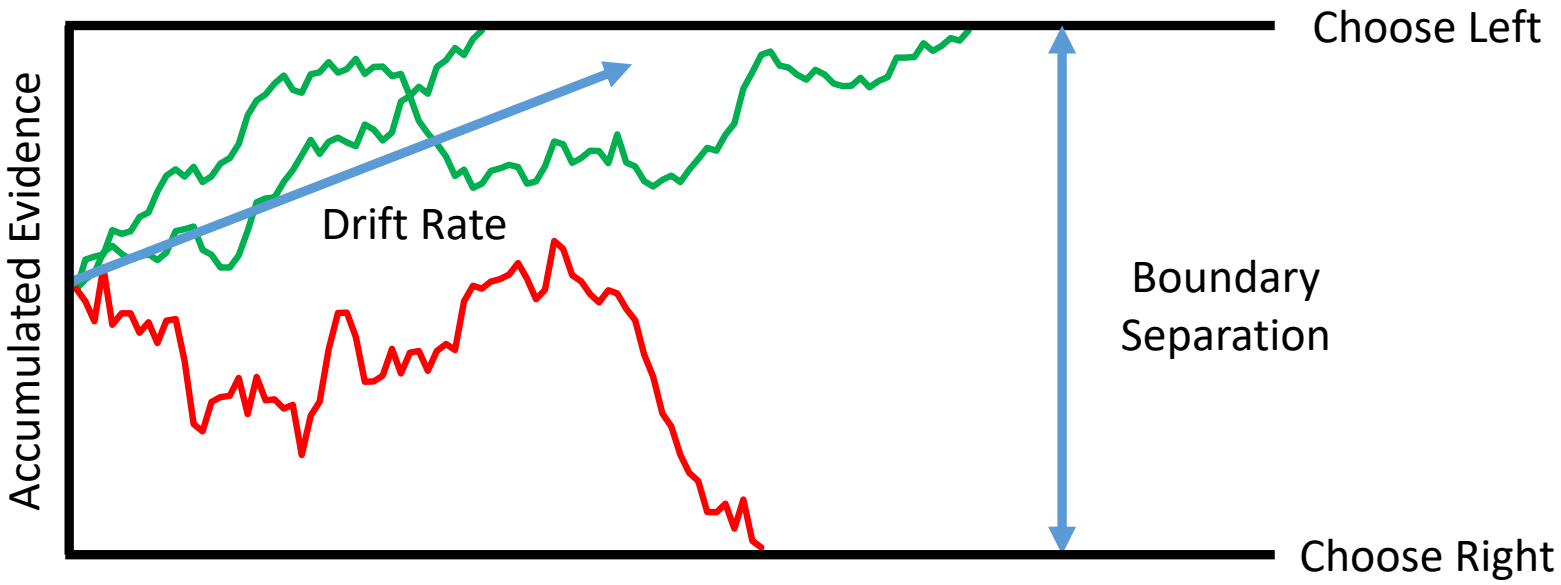


The older participants were more accurate than the younger ones



How can we quantify the trade-off between speed and accuracy?

Drift–Diffusion Model (DDM)



Drift (v)
Strength of the evidence

Boundary Separation (a)
Response caution

Non-decisional time (to)
Encoding and response execution

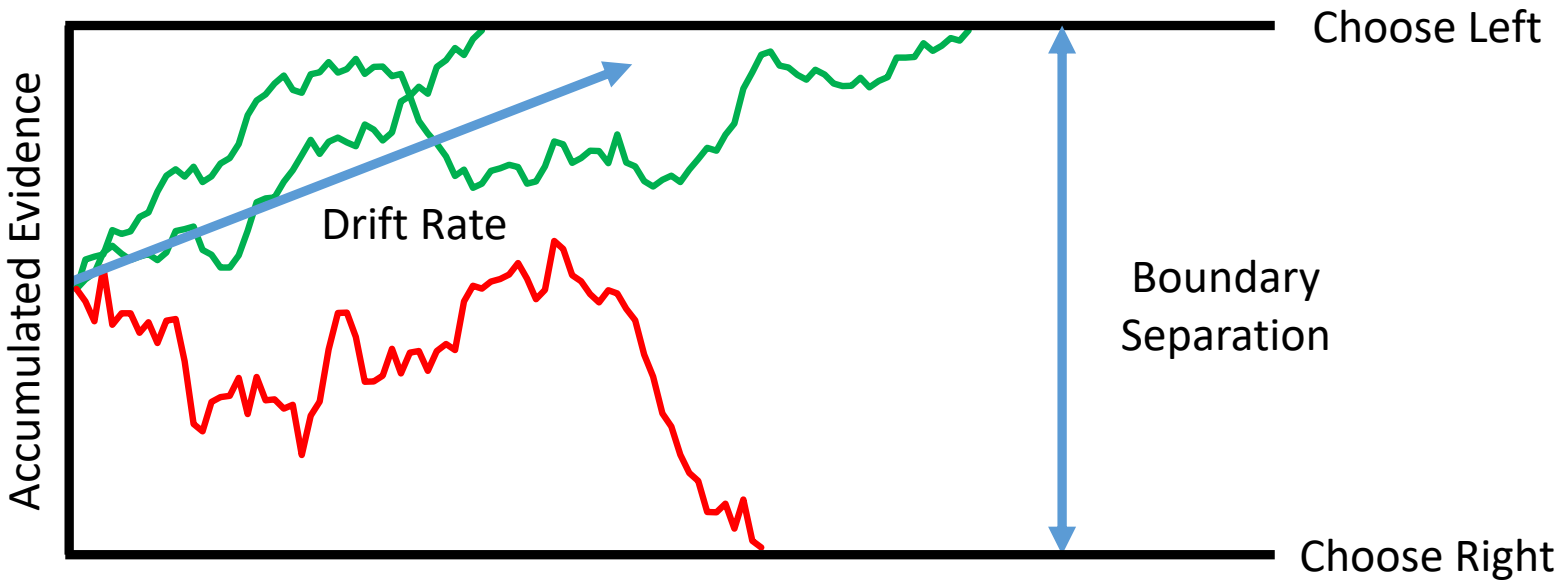
Starting Point Bias (zr)

Starting Point Variability (szr)

Drift Variability (sv)

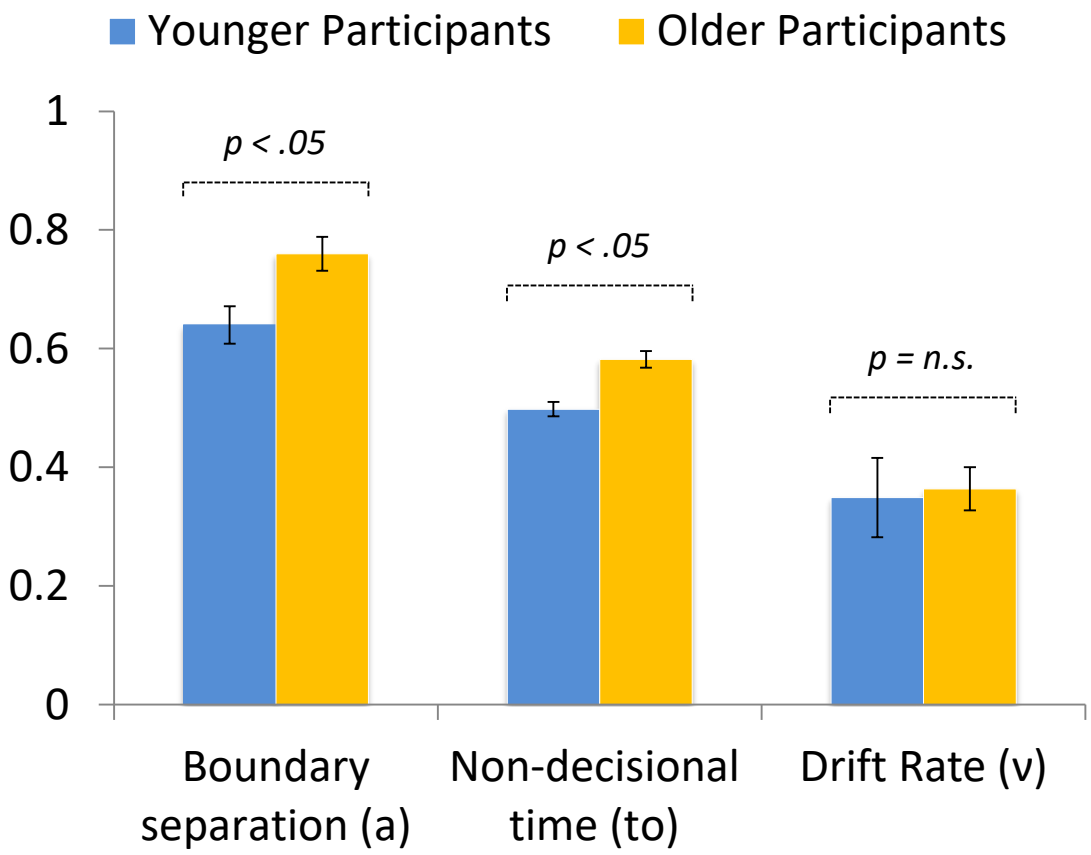
Non-decisional time Variability (sto)

Drift–Diffusion Model (DDM)



Parameter	RT	Accuracy
Higher Drift (higher strength of evidence)	Shorter RTs	Higher accuracy
Higher boundary separation (more cautious)	Longer RTs	Higher accuracy
Higher non-decisional time (longer peripheral processes)	Longer RTs	Same accuracy

DDM Data Fitting Results



- Older participants: higher boundary separation and non-decisional time
- No differences in the drift rate (the quality of the evidence was intact)

How to extract latent variables from data

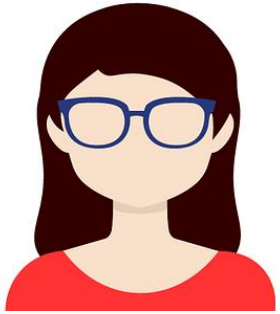
- Data fitting illustration
- The Fast-DM program
- Do it yourself - Maximum-Likelihood estimation

5 Conditions

Subject A
Base-line

Subject B
Speed condition

Subject C
Accuracy condition



Subject D
Difficult condition

Subject E
Beer condition

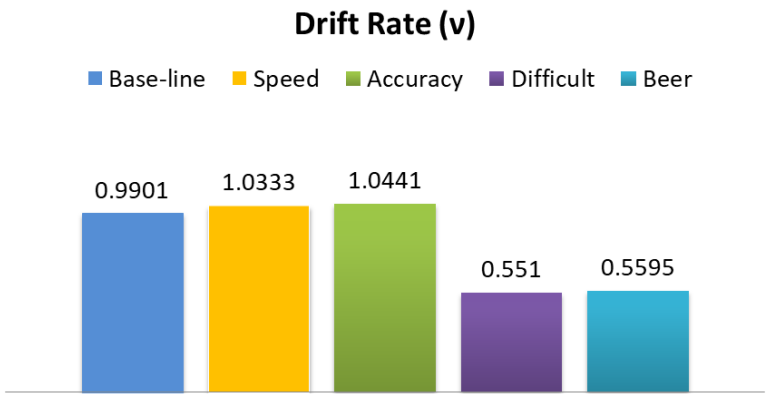
5 Conditions

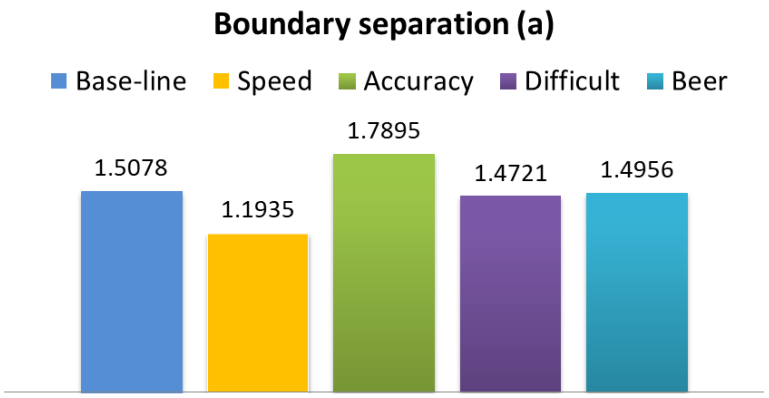
Subject	Condition	Mean-RT	Accuracy
	Base-Line	745 ms	0.83
	Speed	618 ms	0.77
	Accuracy	897 ms	0.85
	Difficult	828 ms	0.71
	Beer	1031 ms	0.67

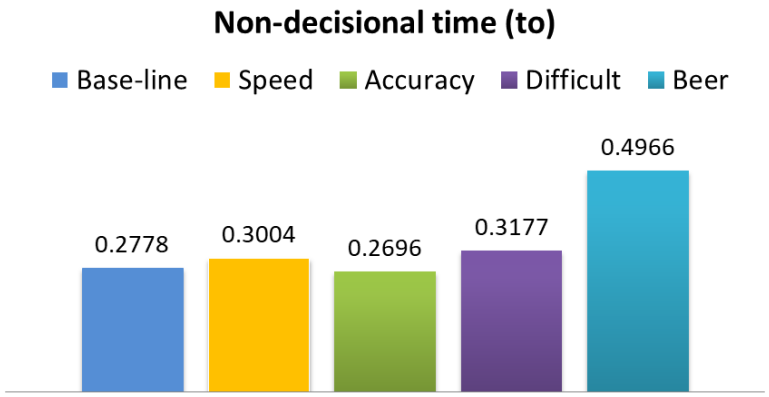
Data Fitting Program - Fast-DM

1. Google - Voss & Voss Fast-DM
2. Download the Fast-DM windows binaries
3. Save your data in a csv file (txt file)
4. Create a control file with a text editor (experiment.ctl)
5. Run the Fast-DM.exe file
6. Read results into your favorite statistics software for further analysis

Comparison of the Fitted Parameters







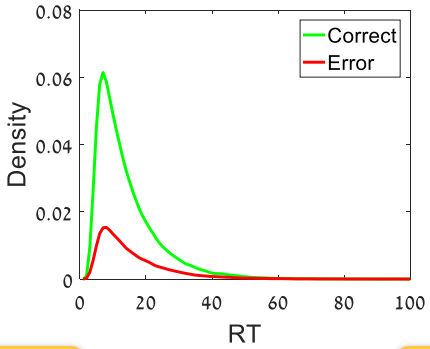
- The participants in the base-line, speed and accuracy conditions have the same drifts and non-decision times, however their boundary separation varies
- The participant in the difficult condition has lower drift
- The participant in the beer condition has lower drift and higher non-decisional time

How to extract latent variables from data

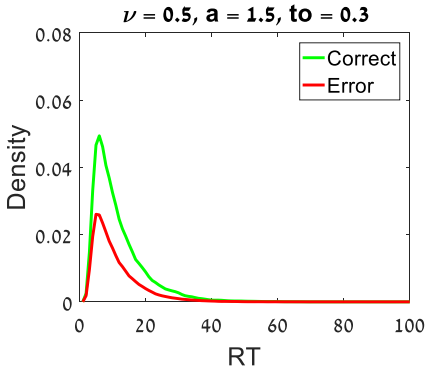
- Data fitting illustration
- The Fast-DM program
- Do it yourself - Maximum-Likelihood estimation

Maximum Likelihood Estimation

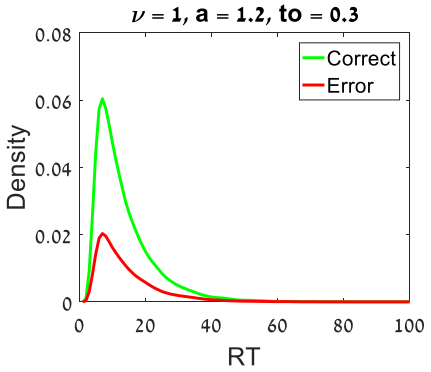
Empirical Distribution



Simulated Distribution 1



Simulated Distribution 2



Which simulated distribution is more similar to the empirical data?

It is more likely that the empirical distribution was generated using the 2nd set of parameters, rather than the 1st set of parameters

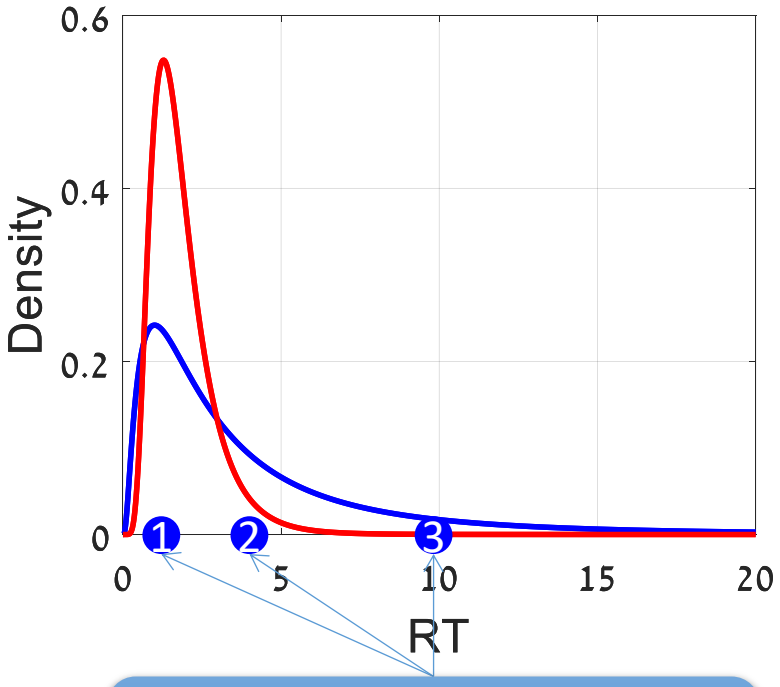
Maximum Likelihood Estimation

1. How can we quantify the relations between the empirical and simulated distributions?

2. How can we search for the best fitting set of parameters?

Quantifying the Similarity

2 Simulated distributions (Blue & Red)



3 Empirical Data Points

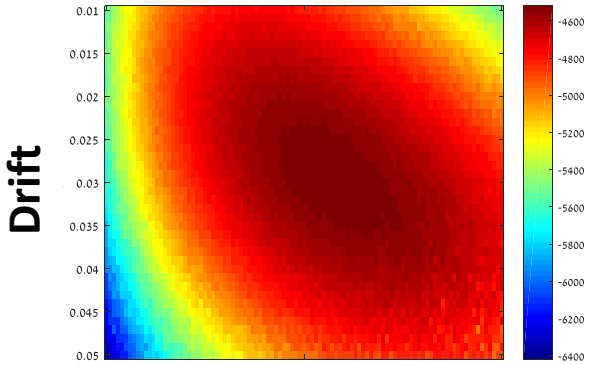
- Likelihood function:

$$L(\theta|X) = \prod_{i=1}^{N} f(x_i|\theta) = f(x_1|\theta) \cdot f(x_2|\theta) \cdot f(x_3|\theta)$$

- Red distribution → lower likelihood (data points 2 & 3)
- Blue distribution → higher likelihood
- Likelihood → Log-Likelihood

Search Strategies

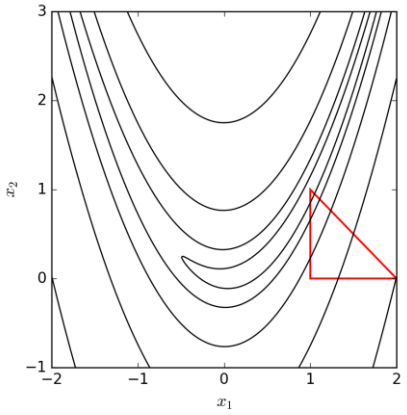
Grid Search



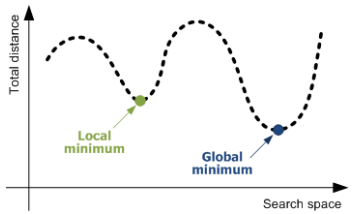
Disadvantages: not efficient

In case there is a large number of parameters, the grid search method can be very time consuming

Search Algorithms



Disadvantages: local minima



Mixture strategy:

- Use coarse grid search to find 10 – 20 best fitting sets of parameters
- Use these sets as starting points to the search algorithms (e.g., simplex)

More readings

•Ratcliff & McKoon (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation.*

•Mazurek, Roitman, Ditterich & Shadlen, 2003; *Cerebral Cortex*

•Voss, A., & Voss, J. (2007). Fast-dm: A Free Program for Efficient Diffusion Model Analysis. Behavioral Research Methods, 39, 767-775.

•Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: current issues and history. Trends in cognitive sciences, 20(4), 260-281.