REACTION TIME DISTRIBUTIONS YIELD MORE THAN JUST PARAMETERS:

THE WEIBULL DISTRIBUTION CASE

Denis Cousineau

Indiana University

Wednesday, February 10th, 1999

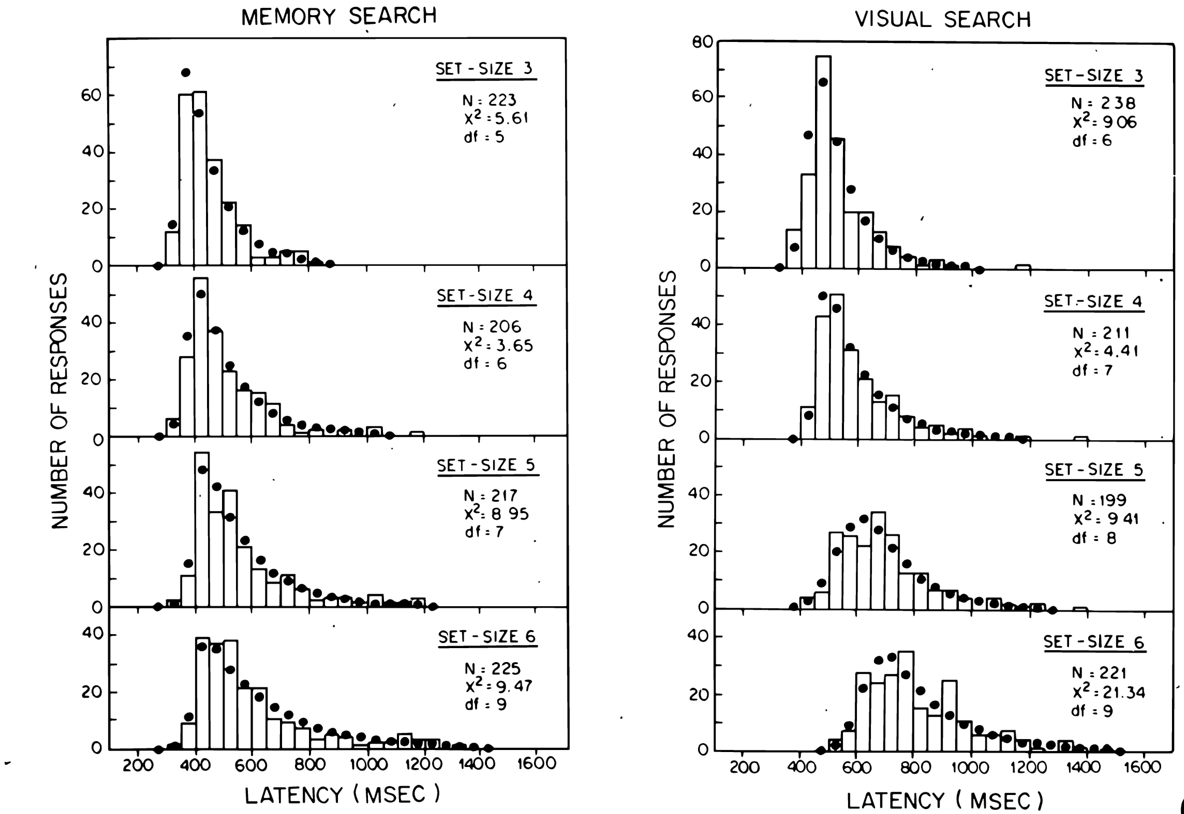
Abstract

Reaction time (RT) distributions have often been used in the past as a method to extract parameters from a RT population. These parameters are useful when the distribution is highly skewed, as is commonly the case with RT. However, one source of information that is often neglected is the distribution formula by itself. By looking at the shape of RT distribution, very broad class of models can be tested (such as competitive and multiplicative models).

At this time, the exact shape of the RT distribution is not known, but we will argue that the Weibull is the best contender. To do so, we will contrast four commonly used distributions in psychology: the ExGaussian, the LogNormal, the Double Exponential, and the Weibull distributions.

Empirical evidence will be reviewed. These results favor a Weibull distribution, suggesting that a race model could account for the results. Formally, race models are constrained by very specific assumptions. Using simulations, we violate these assumptions to see if obtained distributions still look like empirical data, so that a very broad class of race models could be compatible with the results.

Using principles of race models, a connectionist model that simulates RT will be presented. This network is better represented as a Random Walk model that learns to adjust its thresholds and drifts rates.



(Reproduced from Hockley, 1984)

- RT distribution presents the number of response lying between intervals; it can be estimated using distribution formula. This is appropriate when the distribution is not symmetrical.
- In the past: Mostly used for parameter extraction (used to replace statistics such as Mean and SD; Heathcote, Popiel, Mewhort, 1991)
- Termed as “SOFT” model (Indow, 1993), or “EMPIRICAL” model (Ratcliff, 1978): redescription of the data, not a model of the psychological processes (Kepler, 1630s).

Is the “EMPIRICAL” model adequate?

Are there psychological models that predict the “EMPIRICAL” model (Newton, 1660s).

OVERVIEW OF PART 1

1- Test four empirical models using empirical data. All these models have 3 free parameters.

- the LogNormal distribution (Ulrich and Miller, 1993)
- the ExGaussian (Ratcliff, 1978)
- DoubleExponential
- Weibull (Logan, 1992, Indow, 1993)

And discuss the notion of “Adequacy”.

2- Examine to what extent the RT distributions are compatible with a race model.

OVERVIEW OF PART 2

1- Present a model that generates “RT” distributions compatible with the observed distributions.

- All the distribution fits are computed by minimizing the log Likelihood function:

$$L = \prod f(RT_i \mid \theta)$$

Where f is the probability function of the distribution to be fit, and θ is a set of parameters.

L is also an index of fit (smaller *L* indicates better fit)

- For each session of each subject, raw data are entered into the function *L*, and an index of fit is obtained

  (no averaging, no group analysis).

- The empirical model that provides the best fit is considered the best fitting function for that session of that subject.

Experiments 1 to 4 (taken from Cousineau and Larochelle, submitted).

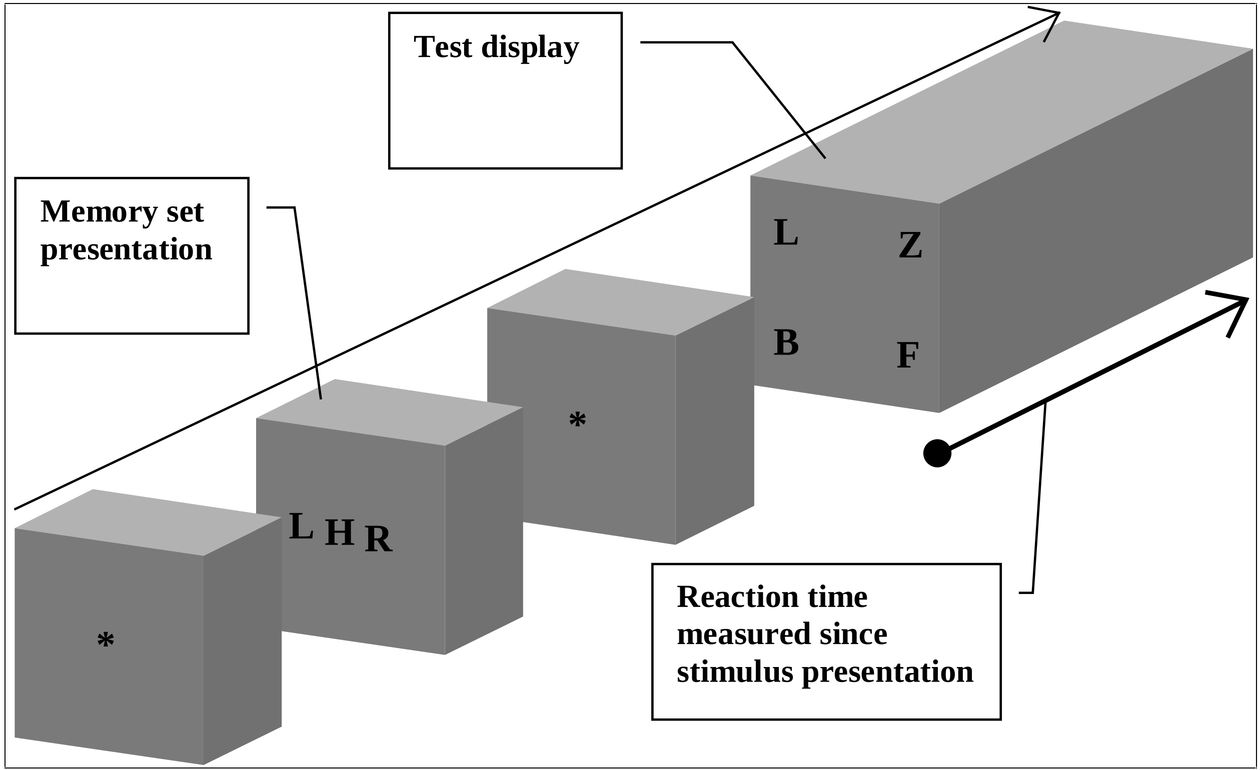
44 subjects participated in one of four memory and visual search experiments.

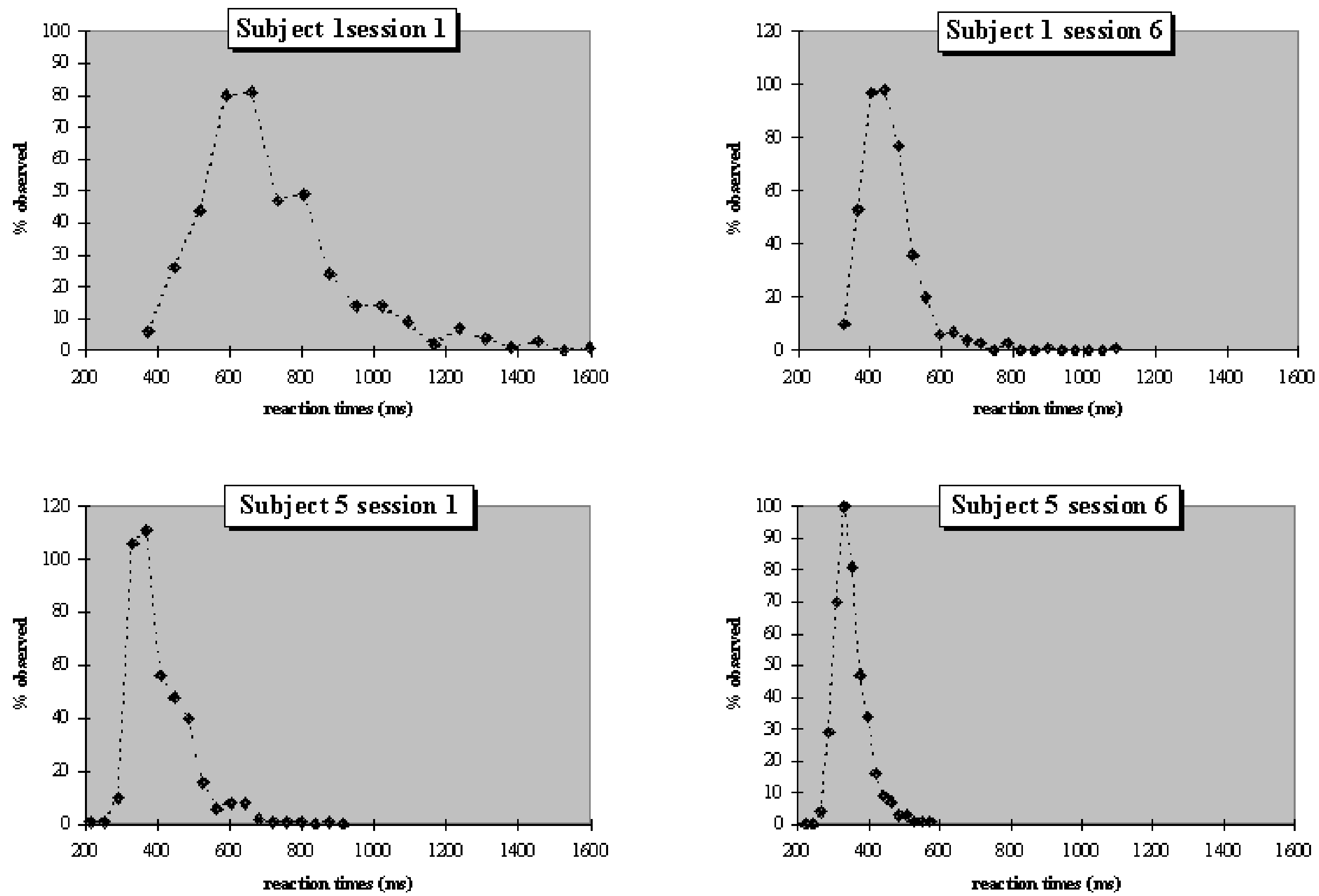
Independent variables: the size of the memory load, the size of the test display, and

the type of response (target present or target absent).

The learning lasted 6 sessions.

There were 432 – erroneous responses (~5%) observations per session.





Subject 5 distribution varied in scale only (the width);

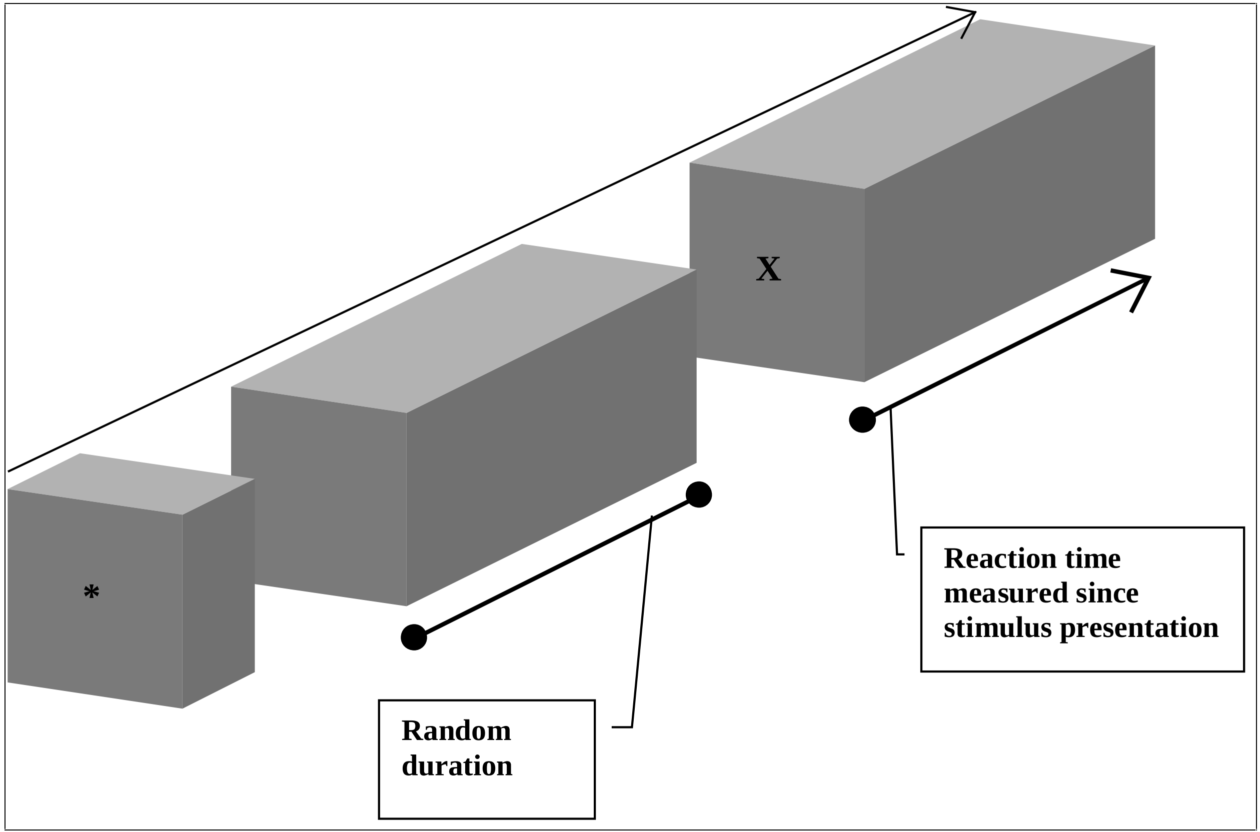
Subject 1 distribution varied both in scale and starting position.

	Distribution			
Conditions	Weibull	LogNormal	ExGaussian	D.Exponential
Over experiments 1 to 4 (n = 44 subjects x 6 sessions = 264 distributions)				
Target absent trials				
Best fitted	195	69	0	0
Mean fit	2667.5	2681.8	3028.8	2923.7
% Best fitted	73.8%	26.1%	0.0%	0.0%
Target present trials				
Best fitted	215	47	1	1
Mean fit	2587.9	2611.1	2892.9	2816.0
% Best fitted	81.44%	17.8	0.3%	0.3%
Experiment 1 per display size x memory size conditions (n = 24 subjects x 6 sessions x 9 conditions = 1296 distributions)				
Target absent trials				
Best fitted	1122	164	1	9
Mean fit	279.8	282.7	322.3	300.9
% Best fitted	86.6%	12.7%	0.0%	0.7%
Target present trials				
Best fitted	1187	100	0	9
Mean fit	269.6	270.0	311.6	291.8
% Best fitted	91.6%	7.7%	0.0%	0.7%

Experiment 5 (unpublished)

59 subjects in a simple RT experiment.

Independent variable: the SOA before the target (an X) appeared. SOA was uniform, between 1 and 10 seconds. There was only one session of 60 observations.



	Distribution			
Conditions	Weibull	LogNormal	ExGaussian	D.Exponential
Experiment 5 (n = 59 subjects x 1 sessions)				
Best fitted	54	2	0	3
Mean fit	44.5	49.5	56.1	51.9
% Best fitted	96.4%	3.6%	0.0%	5.4%

Discussion of the empirical results

Overall, the Weibull provided the best fit in over 80% of the RT distributions analyzed.

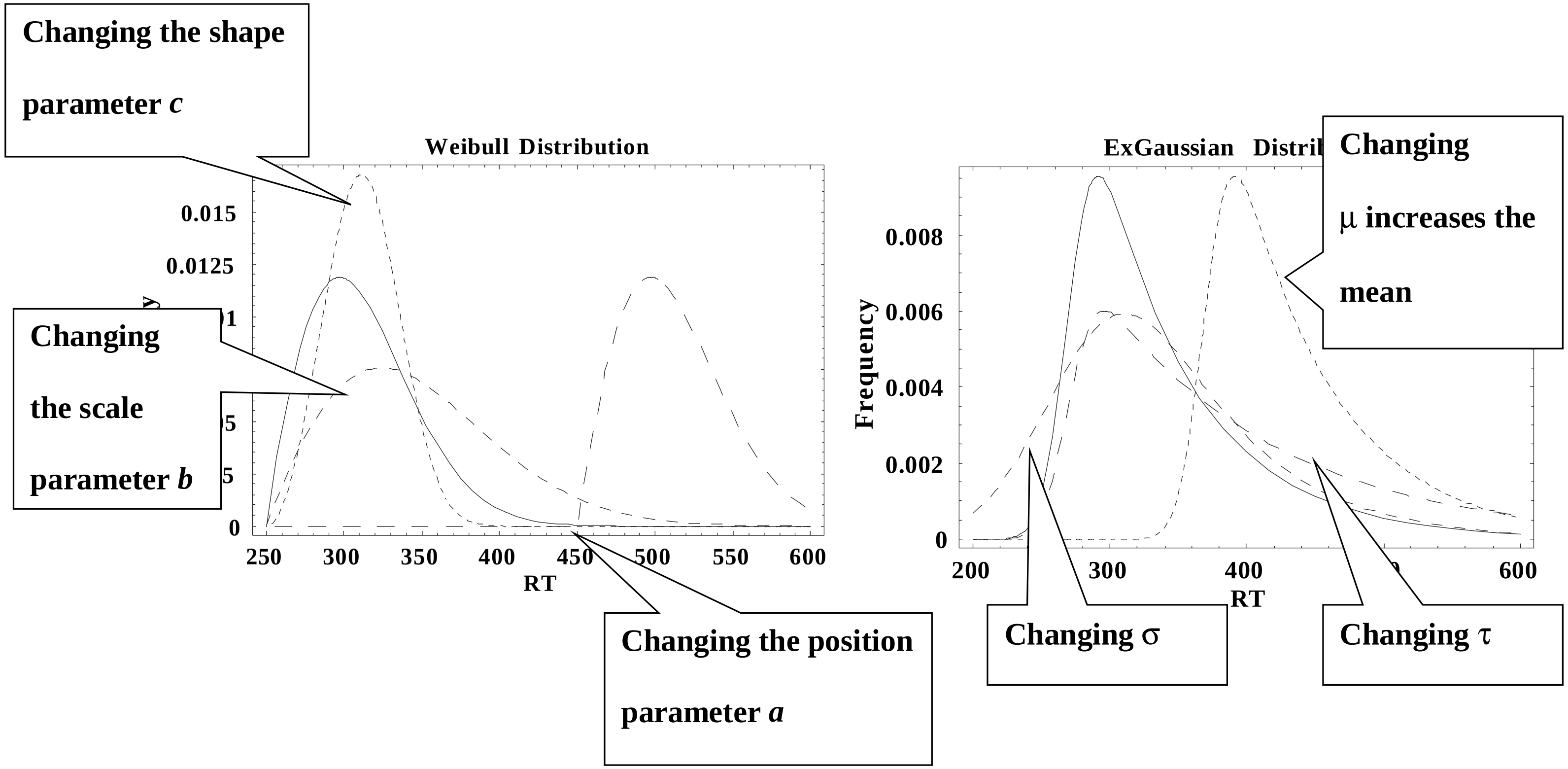
However, the LogNormal also provided a good fit (as measured by the LogLikelihood index of fit).

The ExGaussian is the worst distribution, with respect to LogLikelihood and % best fitted.

How can we measure the “adequacy” of an empirical model?

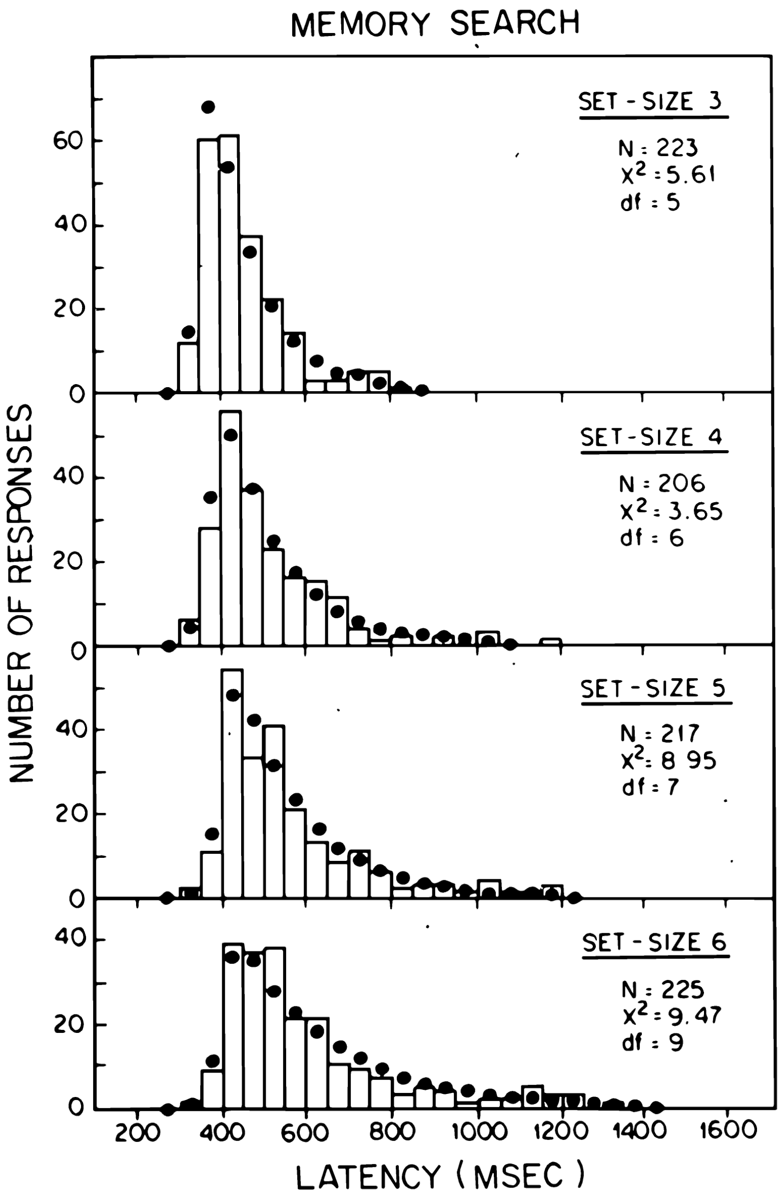
a) By the quality of the fit. But! Have to demonstrate that this is not the result of an “artifact”, i.e. an ability of the Weibull to overfit, by virtue of its equation. Requires Response Surface Analyses (Myung and al., 1998)

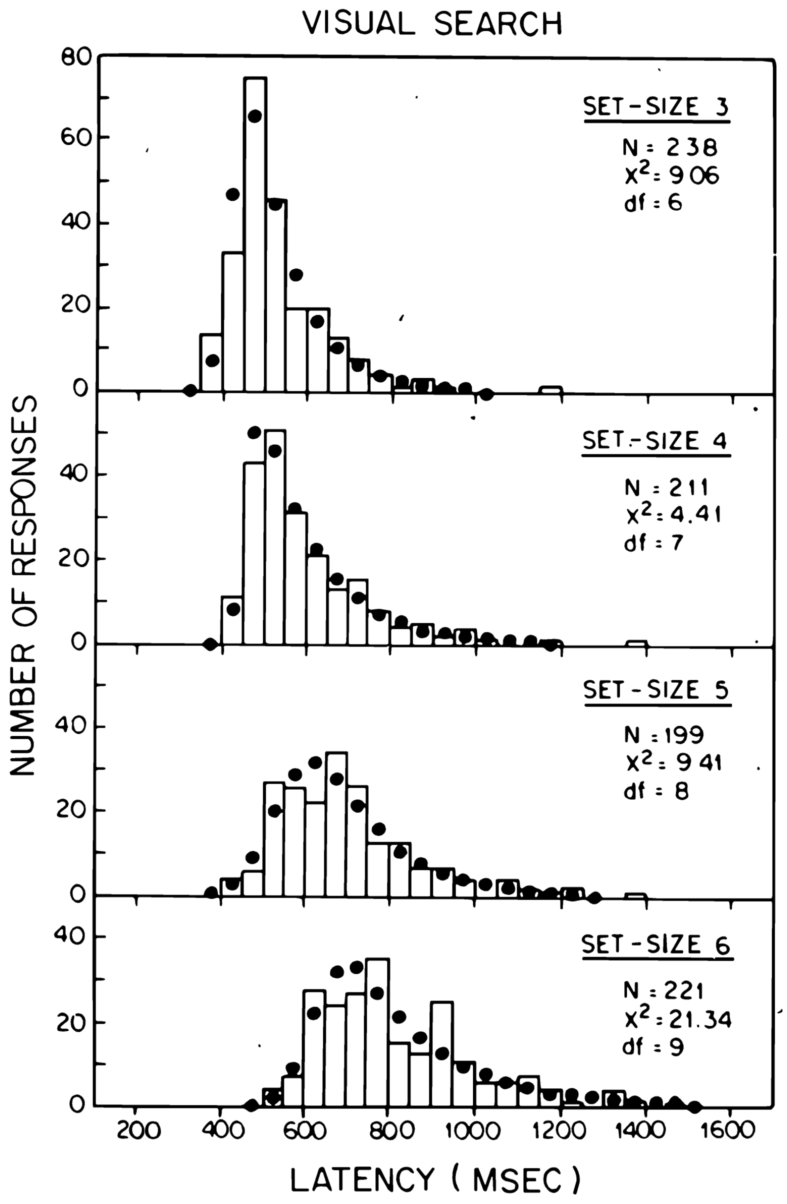
b) By the usefulness of the parameters extracted with the distribution.



(See Hockley, 1984)

c) By its possible relation to psychological theories.


MEMORY SEARCH
SET - SIZE 3
N = 223
X² = 5.61
df = 5
SET - SIZE 4
N = 206
X² = 3.65
df = 6
SET - SIZE 5
N = 217
X² = 8 95
df = 7
SET - SIZE 6
N = 225
X² = 9.47
df = 9
NUMBER OF RESPONSES
LATENCY (MSEC)


VISUAL SEARCH
SET - SIZE 3
N = 238
X² = 9 06
df = 6
SET - SIZE 4
N = 211
X² = 4.41
df = 7
SET - SIZE 5
N = 199
X² = 9 41
df = 8
SET - SIZE 6
N = 221
X² = 21.34
df = 9
NUMBER OF RESPONSES
LATENCY (MSEC)

Let's consider a class of model:

- Competitive models where one or a small number of units can trigger the response.

  $RT = \text{Min}(Y_i) \Rightarrow RT \sim \text{Weibull}$ (Asymptotic theory of extreme, Galambos, 1978).

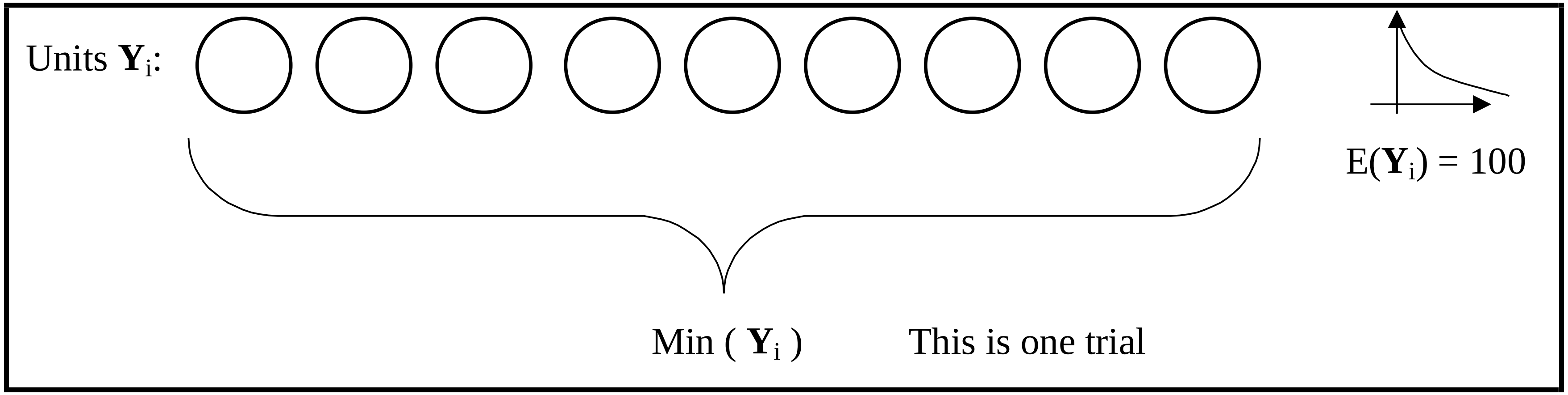
  Ex: Race models (Marley, 1989)

Assumptions that come with the formal definition of a competitive model (Gumbell, 1958):

- The number of units Y_i must be large ($n \sim 400$)
- Each Y_i must be lower bounded and continuously distributed
- Each Y_i is not contaminated by other source of variance (such as noise)
- All Y_i are independent and identically distributed (i. i. d.)

These assumptions are too restrictive to be of any use in psychological models. Can we relax some of them?

Numerical simulations:



We simulated units Y_i then find the “Simulated Reaction time” as Min(Y_i);

One “session” consists of 100 trials (replications);

100 sessions were included in the analyses.

a) Varying the number of units Y_i competing against each other.

Number of	then distribution of Min(Y_i) is			
Units Y_i is	Weibull	LogNormal	ExGaussian	D.Exponential
25	100	0	0	0
50	100	0	0	0
100	100	0	0	0
200	99	1	0	0
400	98	1	0	1
800	98	1	0	1
1600	100	0	0	0

The number of competing units does not seem to be a critical factor in race models.

b) Varying the distribution of each single unit

Units Y_i are	then distribution of Min(Y_i) is				
Distributed as	Weibull	LogNormal	ExGaussian	D.Exponential	
Exponential	100	0	0	0	Lower-bounded
Uniform	100	0	0	0	“
Poisson	79	3	0	18	Discrete
Normal	0	0	0	100	no lower-bound

The lower-bounded, continuous assumption IS critical in race models.

c) Contaminating the units: Adding normally distributed random noise with zero mean to each unit

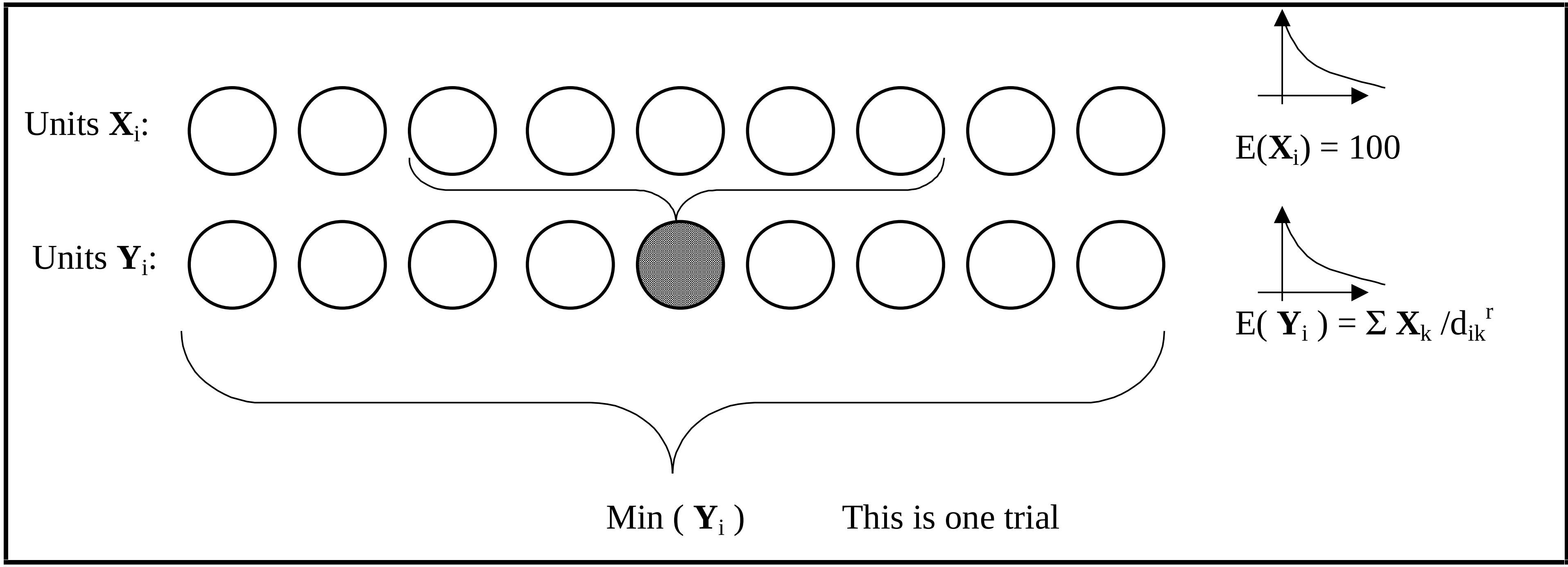
[This creates a convolution between an exponential main component and normal noise]

Ratio of	then distribution of Min(Y_i) is			
Noise/Mean$_{Signal}$	Weibull	LogNormal	ExGaussian	D.Exponential
0.25%	100	0	0	0
0.50%	98	2	0	0
1.00%	72	28	0	0
2.00%	58	42	0	0
4.00%	86	13	0	1
8.00%	57	18	0	25

A small amount of contamination can rapidly violate the lower-bounded assumption. [Apparently the Weibull is not overfitting]

With 1% of noise, results are similar (with respect to % Best fitted) to empirical data.

d) Simulating dependencies between nodes.



1- Generate a layer $\mathbf{X}_i$

2- Generate a layer $\mathbf{Y}_i$ such that $E(\mathbf{Y}_i) = \Sigma\, \mathbf{X}_k / d_{ik}^{\,r}$

Where d_{ik} is the distance between the unit $\mathbf{X}_k$ and $\mathbf{Y}_i$, and r is a decay parameter.

3- Choose random values for $\mathbf{Y}_i$

4- Find the minima = 1 trial

⇒ no effect on %Best fitted for r ranging from ½ to 3.

Discussion on the race model assumptions.

a) Race models in general predict a Weibull distribution

b) Adding a small amount of noise to each units increases the likelihood of observing a LogNormal distribution

c) Inhibition between units close to each other does not change the results: The scale of the distribution is increased, but not the proportion best fitted by Weibull

d) Competition among Means or Sums will never yield Weibull (Central limit theorem)
⇒ Min and Means are incompatible when trying to account for empirical RT distributions

Theory		Observable	example
Race model	→	Weibull	Instance-based (Logan, 1992), Horse Race model (Marley, 1989) Random Walk models (later section)
?	←	Weibull	**?**
?	→	Distribution similar to a Weibull	**?**

General conclusion of part 1:

a) The Weibull seems to be the best “EMPIRICAL” model known so far;

b) The LogNormal distribution is very close to the Weibull in shape;

c) The LogNormal is more frequently seen if there is noise inside of a competitive system;

d) The use of the ExGaussian should be avoided.

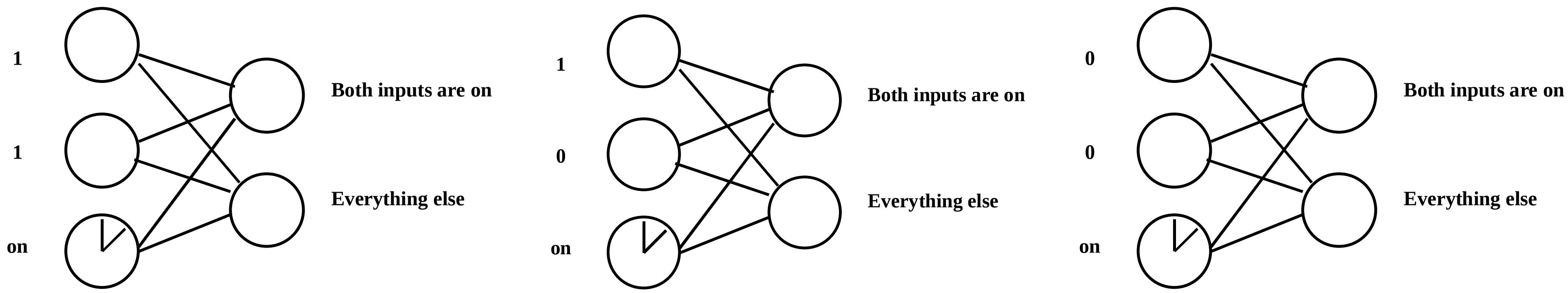
PART 2

- Objective: Build a distributed network that
  – learns to categorize simple stimuli
  – produces a Weibull distribution
- The architecture of this model is similar to two-layer connectionist models, but here ends the similarity.
- This network will use a Min-based decision rule using delays.
- Inputs are
  0: no input visible
  1: very clear input

  A time-out is always part of the input

[Temporary name of this model: n-Min network w/ time-out or Temporal conjunction detector w/ time-out]

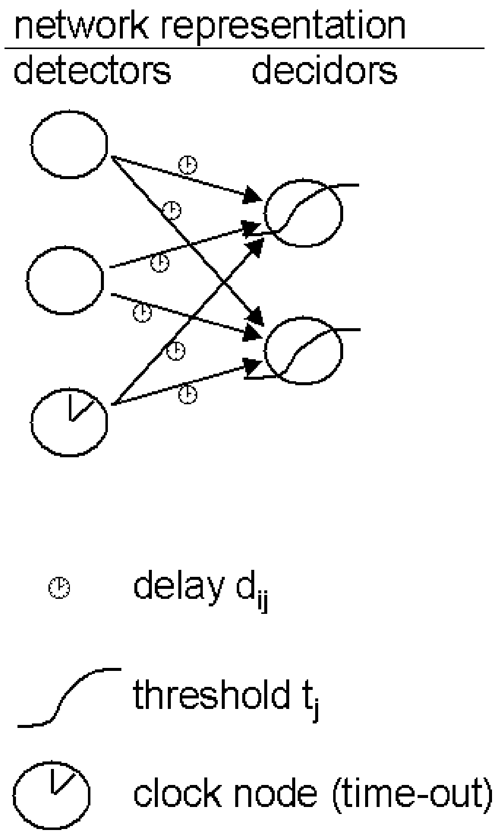
Introductory example: The AND (conjunction of two inputs)

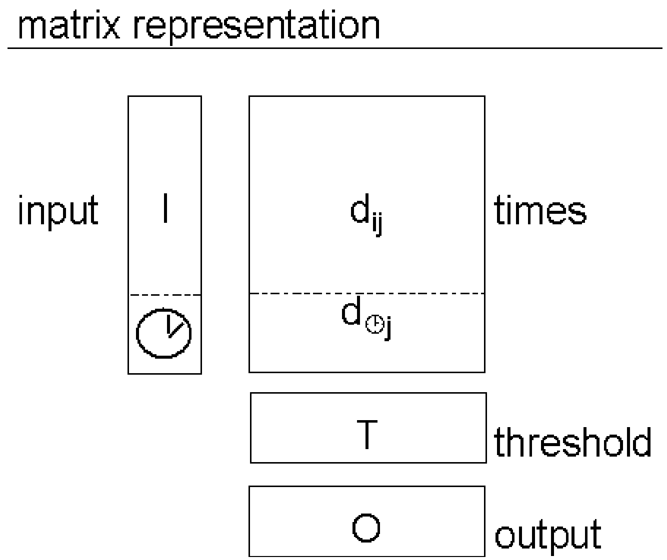


Rule: Answer "Both" <u>as soon as</u> two evidences are detected

or else, answer "Everything else" if one evidence is detected

or else, answer "Everything else" after a while.





Decision:

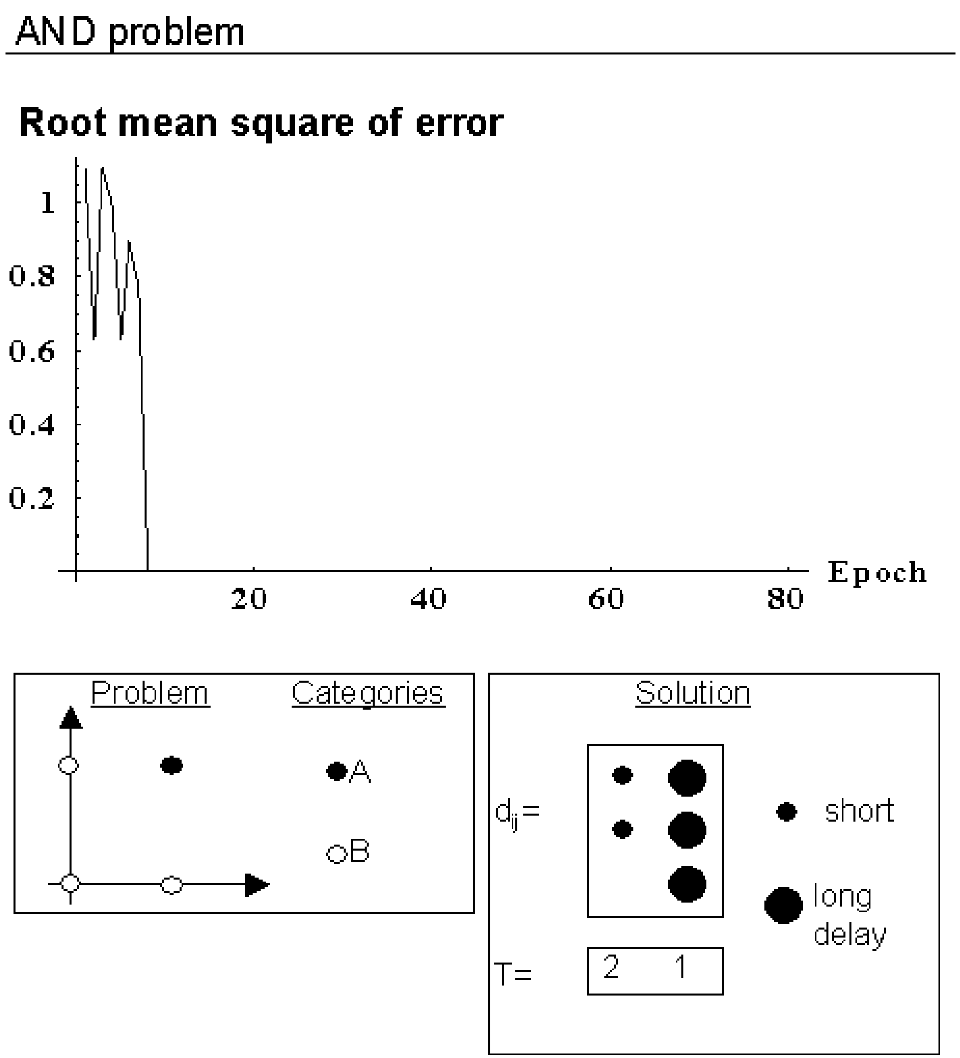
n^{-}Min(S_I / I | n = T) v Min($S_{\oplus}$)

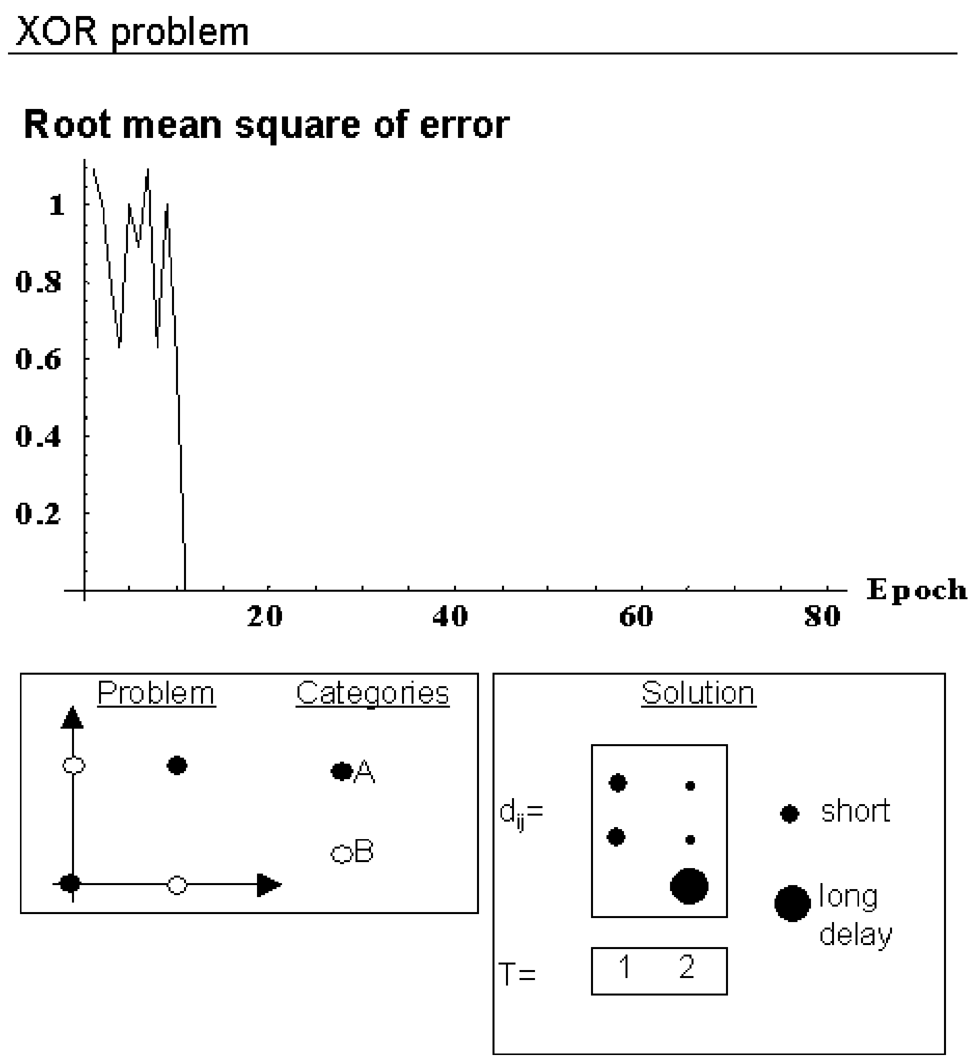
Learning on delays

Slow down by a constant the connection d_{ij} that is responsible of a false alarm

Learning on activation function (threshold)

Modify by ± a constant the threshold t_j of the decidor that miss the decision



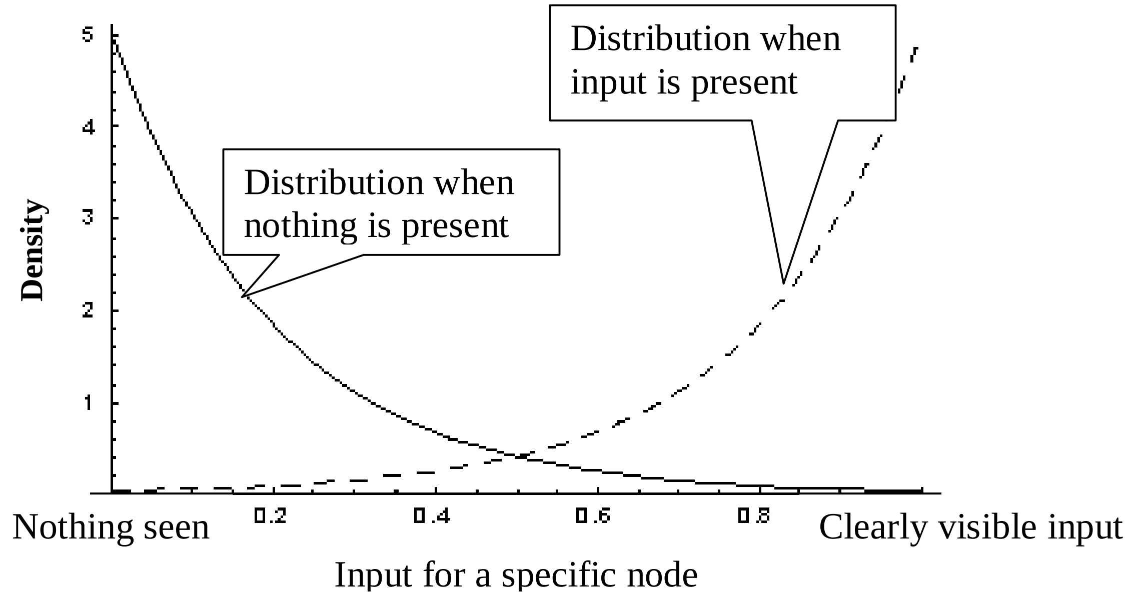


Properties: a) It is faster for detecting conjunction than disjunction

b) It can learn as easily non-linearly separable problems with only two layers.

Adding variability to the network, in order to assess the distribution of simulated reaction times.

Variability is introduced through noise in the input:



The actual delay before a decidor receive an activation is

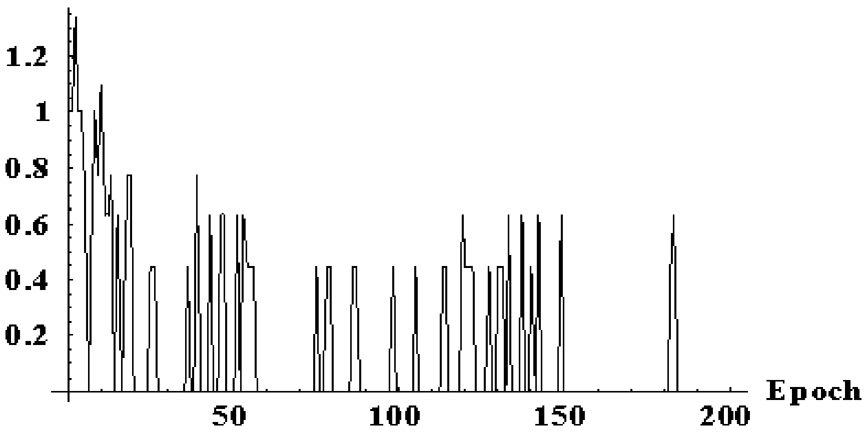
proportional to d_{ij} and

inversely proportional to input's clarity: d_{ij} / I_I

Therefore, if no input is seen (input of zero), the connection will never fire ($d_{ij} / 0 = \infty$)

Noise = 0.025

Root mean square of error



d_{ij} =

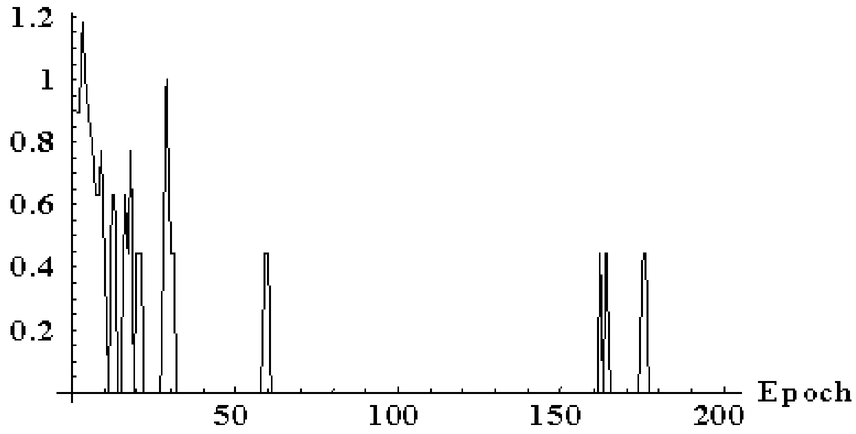
116.0	101.5
113.4	102.1
129.6	129.2

T =

1	2

Noise = 0.010

Root mean square of error



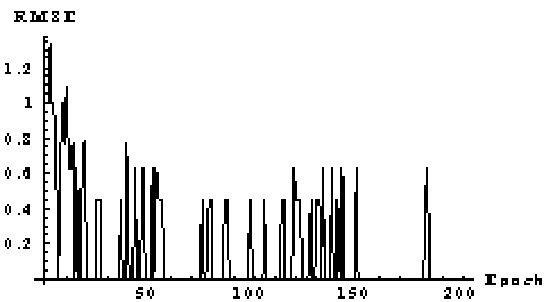
d_{ij} =

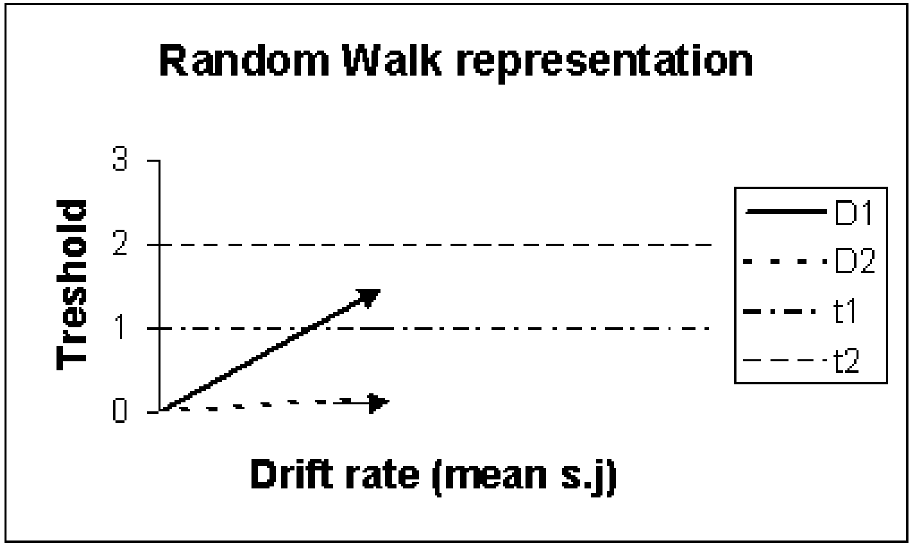
101.1	95.8
103.1	95.6
110.7	110.4

T =

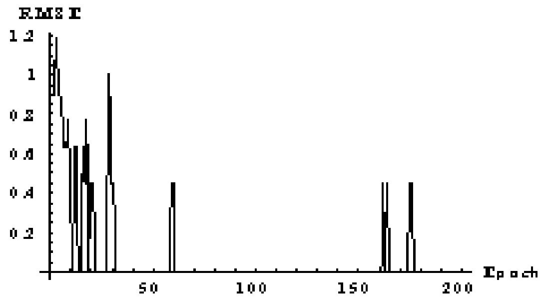
1	2

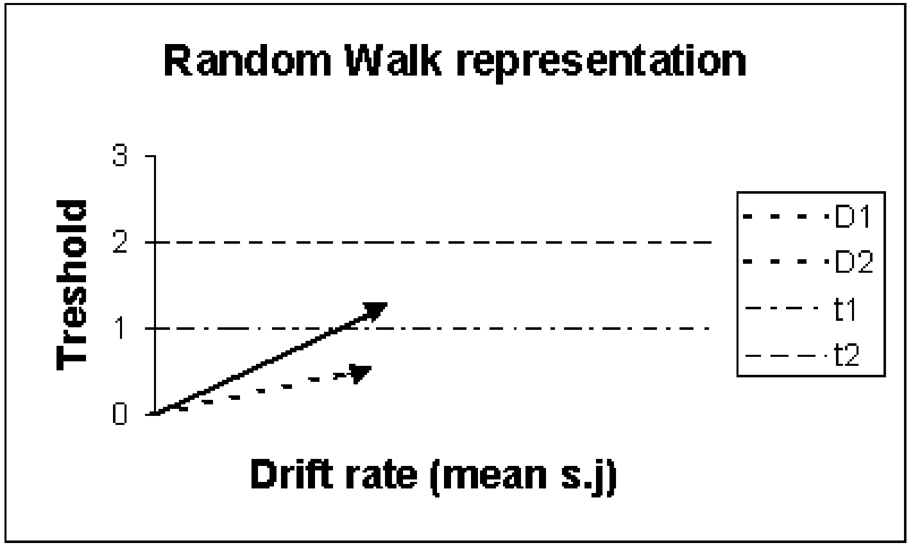
Noise = 0.025





Noise = 0.010





Parameters			Panel a	Panel b
Free				
F	Distribution function of the noise		Exponential	Exponential
η	Variability in the noise		0.025	0.010
Learned				
$d_{.i}$ (D_i)	Drift rates	D_1 - D_2	12.88505	6.4297
t_i	Thresholds	t_1 - t_2	1	1

Further results:

a) Thresholds are not affected by noise,

   but they are affected by redundancy inputs

b) Difference between drift rates is increased when noise increase

   and is reduced when there are redundant inputs.

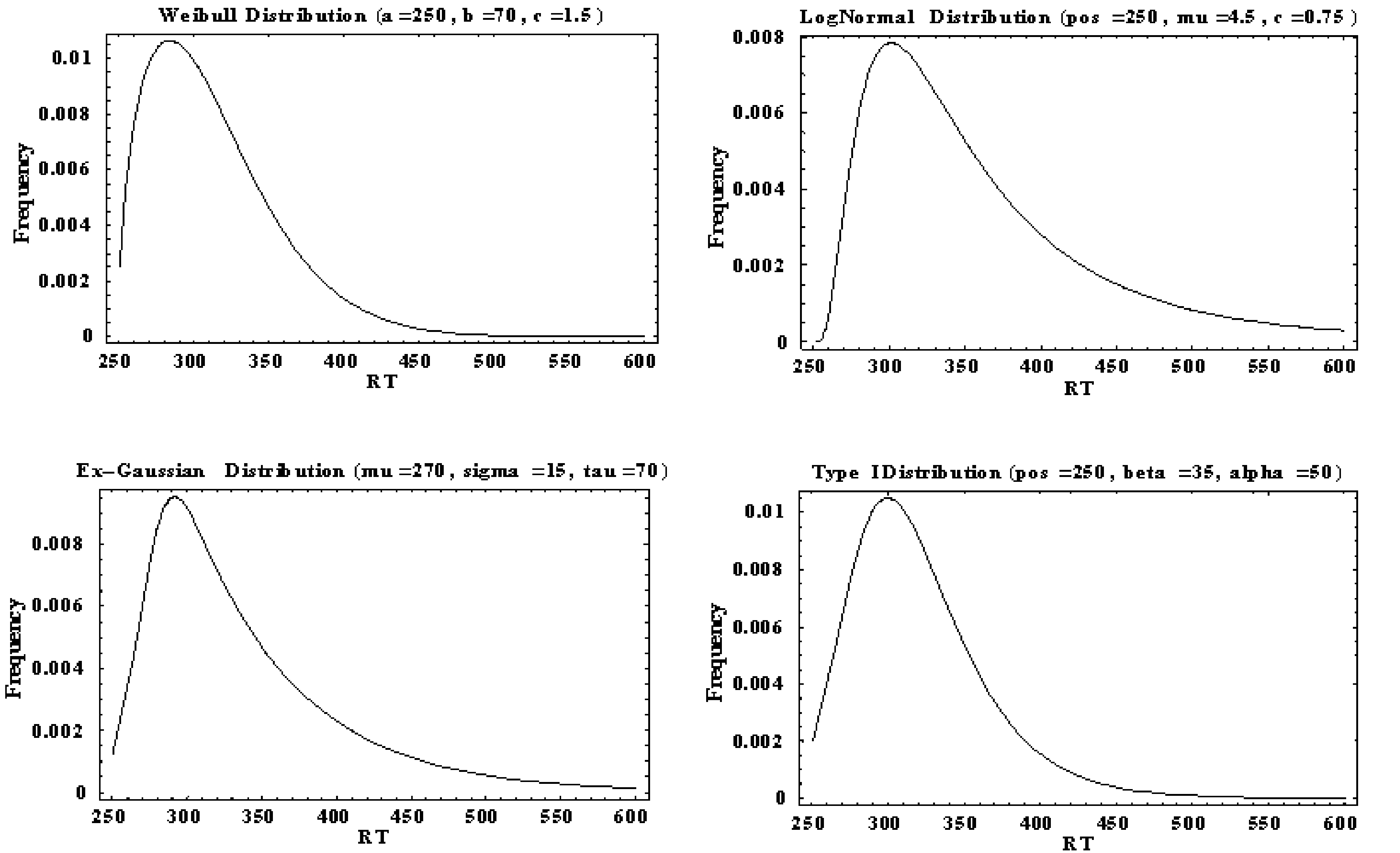
c) Even with a small network, the simulated RT distribution is best fitted by a Weibull

The knowledge that the Weibull distribution best fit the RT distribution is valuable:

a) Weibull parameters seems useful

b) Bring some support in favor of race models

c) Competitive model principles (the use of Minima) has proven useful to generate an extension of the Random Walk model that learns, has only two free parameters, and replicate empirical RT distributions.



Note: Although not seen in the figure, the Type I distribution is symetrical. Also, both the Type I and the ExGaussian distribution have no lower bound.