Accumulator and Race Models: An Overview

Denis Cousineau,
Denis.Cousineau@UMontreal.CA

Abstract

In cognitive psychology, and following the computer revolution of the 80s, neural network models have taken a large part of all the simulations performed. However, there is also in cognitive psychology a long history of models that stem from the Signal Detection Theory model and resulted in the so-called Sampling models. I will describe some of these models (the random walk model, the Poisson race model) and show their relation to SDT. Further, I will end up with a new accumulator model that has all the features of a neural network (learning capabilities, resistance to partial destruction, etc.).

Objectives and overview of the talk

- *Present a brief overview of all the accumulator-race-sampling-signal detection-etc. models, all grouped under the general term: sampling models.*
- *Understand their differences, and the specific details*
- *Provide methods to fit these models*

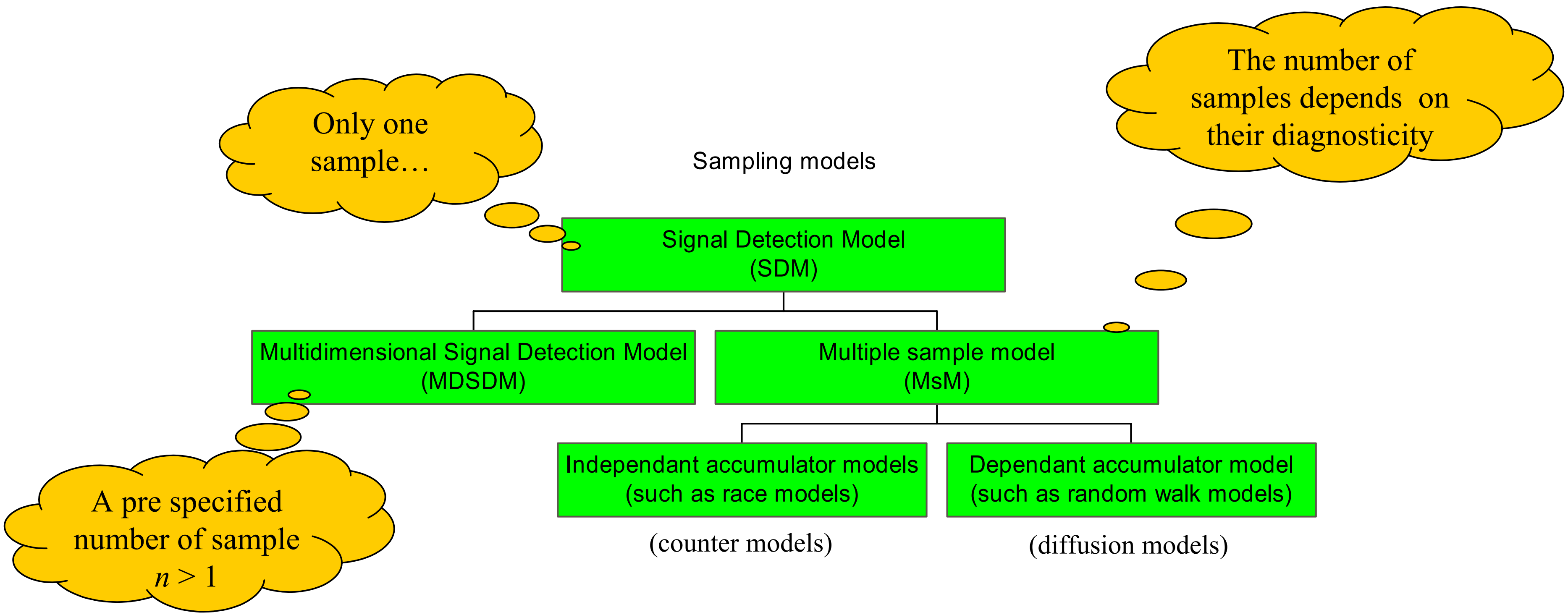
- *The possible observations in psychology*
  - choice, preference, error pattern
  - response times to manifest a choice
  - ➔ A complete model should make predictions on both

- *Historically,*
  - the sampling models were studied first (1960..)
  - the neural networks dominate the scene now. Why?

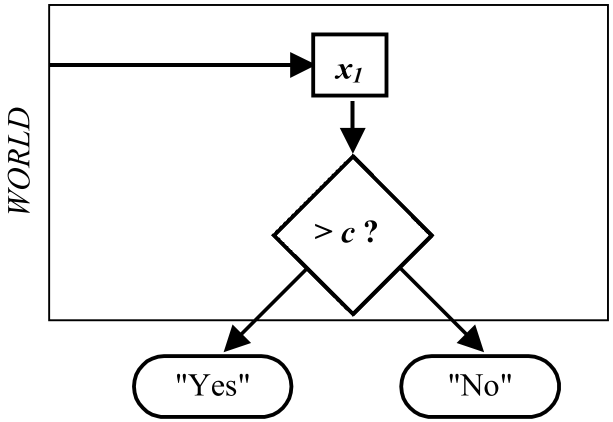
Genealogy of the sampling models

The foundational idea of the sampling models

- *The world is sampled, i.e. we only get noisy activations from it.*
- *Therefore, decision is difficult and a decision criterion is required.*
- *Each sample can be called an "evidence"*

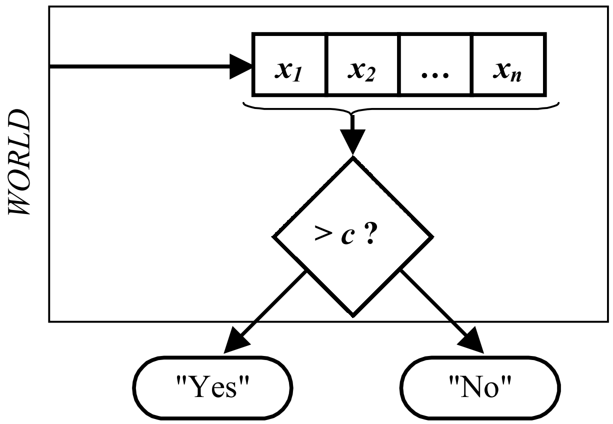


General architecture of the models: The SDM (1/5)



- **SDM:**
  - *You "grab" something from the real world (a sample) and based on this, you make a decision*
  - *There are four types of response:*
    - Pr(Hit), Pr(FA), Pr(CR), Pr(Miss)
  - *There is no way to predict response times (RT)*

General architecture of the models: The MSDSM (2/5)

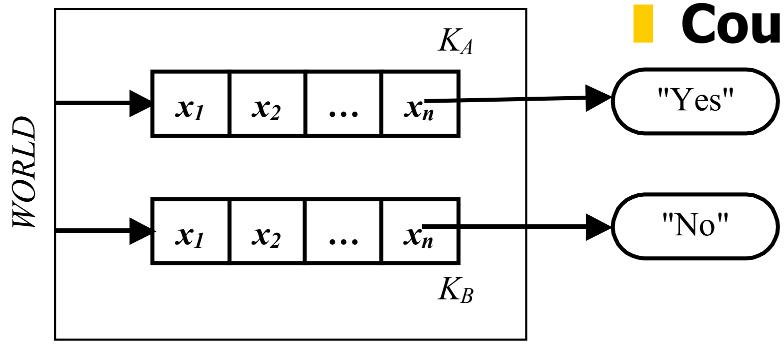


- **MDSDM:**
  - *You take exactly n samples, combine them in some way, and decide*
  - *Combining (x_i) could be: "Take the largest" (presumably the most reliable).*

General architecture of the models: The MSDSM (3/5)

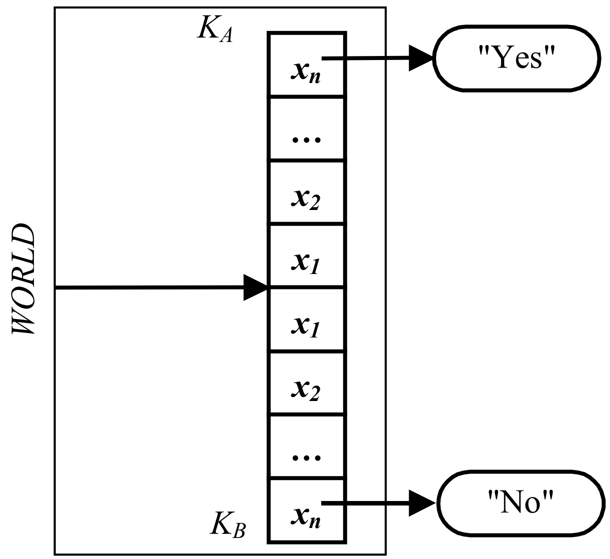
- **Both SDM and MDSDM can't predict response times unless subsidiary assumptions are added:**
  - *Gurnsay's work on the SDM*
  - *Palmer's work on the MSDSM*

General architecture of the models: The MSM (4/5)



- **Counter or Race models**
  - *There is a special trigger: an evidence falling into the last available slot automatically and immediately triggers a response.*
  - *Incoming evidences arrive independently for each possible response.*
  - *At worst, it will take $K_A + K_B - 1$ samples.*

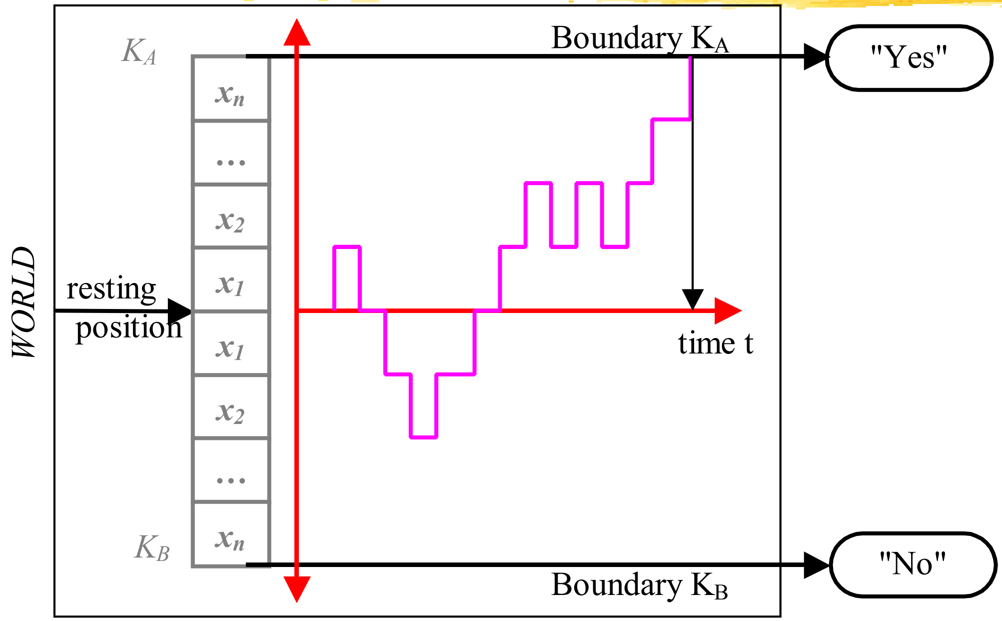
General architecture of the models: The MSM (5/5)



- **Diffusion/Random walk models**
  - *The same trigger is used*
  - *An evidence for one interpretation is necessarily an evidence against the other interpretation (they are in opposition).*
  - *At worst, it may take an infinite number of samples*

Generally speaking, the race model and the diffusion model accumulate evidence: I call them *Accumulator models*

Illustrating the random walk model (1/2)



Will I buy a new car this year?

- *Mental deliberations can be modeled with a random walk model (Busemeyer et al.).*
- *Being biased toward one choice means having a different resting position (Ratcliff et al.).*
- *You can check the model by asking at various times "If you had to decide now, what would you tend to decide?"*

Example is statistics: Wald decision process

- *Suppose a situation where each sample is very costly: Take only what is needed to reach a certain confidence level.*

Illustrating the random walk model (2/2)

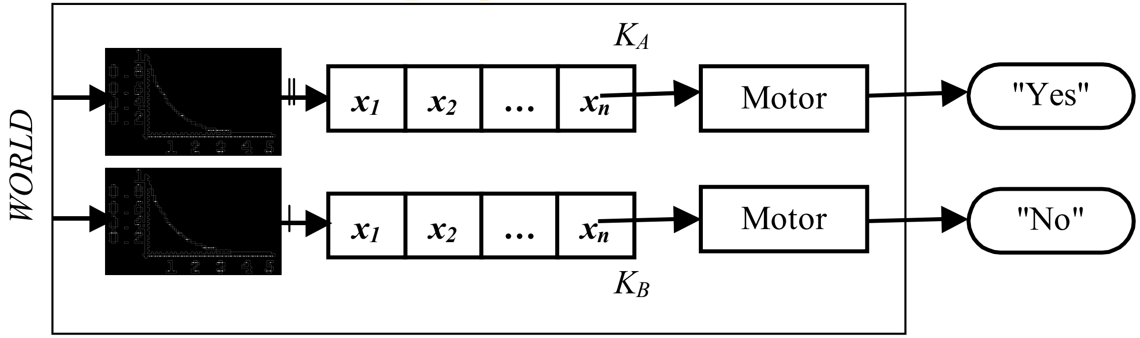
Things to consider:

- *Is the time between two evidence discrete (as shown) or continuous?*
- *Are the evidence unitary (as shown) or continuous? In the first case, it is called a Random Walk model, in the second, a Diffusion model.*

- **Ratcliff summarized his diffusion model with the following parameters:**
  - δ *: average drift rate toward one boundary, as a function of the condition*
  - η*: variability around* δ
  - K_A*,* K_B*,* z*: Boundaries and resting position*
  - T_0*: residual time (motor and perceptual).*

One example: The Poisson race model

Description



- **It assumes that:**
  - *the time between two evidence is continuous and random*
  - *Since it represents spikes of activity (in this model), they are exponentially distributed (Poisson distribution).*
  - *Presumably:*
    - If the stimuli eliciting a "Yes" response is present, the rate for the first channel τ_A is greater than the rate τ_B for the second channel.
    - Otherwise the rate for the first channel τ_A is smaller than the rate τ_B for the second.
  - *There might be a residual time for motor response, call it T_0.*
  - *We thus have quit many parameters:*
    - $\tau_{A|A}$, $\tau_{B|A}$, $\tau_{A|B}$, $\tau_{B|B}$, K_A, K_B, T_0

How to fit - 1

- **b.1) Full simulation method**
  - *write a program that returns mean RT(hit), RT(miss), RT(CR), RT(FA), P(hit), P(miss), P(CR), P(FA) given a set of parameter:* $\tau_{A|A}$, $\tau_{B|A}$, $\tau_{A|B}$, $\tau_{B|B}$, K_A, K_B, T_0 ➔
  - *Compute the SSE between observed and simulated statistics:*

$$\sum_{i=1}^{8} \frac{\left(Obs_i - Exp_i\right)^2}{Exp_i}$$

  - *Minimize the SSE through multiple call to the simulation with different parameters*
  - *The program to the right is pretty inefficient (~8 sec. per simulation)…*

- **One simulation**
  - *Repeat 1000 times (for stability)*
    - Assume stimulus A is presented
    - Repeat K_A times
      - get t_A ~ Exponential ($\tau_{A/A}$)
    - $RT_A = \Sigma\ t_A + T_0$
    - idem for the other accumulator RT_B
    - If $RT_A > RT_B$ then Hit else miss.
    - idem when the other stimulus is being presented
  - *Compute:*
    - *RT(hit), RT(miss), RT(CR), RT(FA)*
    - *P(hit), P(miss). P(CR), P(FA).*

Interlude: Shortcuts

- **Most of the previous program is occupied with generating individual "spikes":**
  - Repeat K_A times
    - get $t_A \sim$ Exponential ($\tau_{A|A}$)
- **There is a shortcut, since we know what is the distribution of total times it takes to receive K_A evidence separated by exponentially random times:**
  - Let Pr(T = t) be the probability that the total time is t.
  - Assume (for simplicity) that K_A is 2.
  - The time T is equal to the time of the first and the second spike, $T_1 + T_2$
  - Both are unknown. Assume that the first is equal to z, $T_1 = z$, $0 < z < t$
  - Then, $\Pr(T = t) = Pr(T_1 = z)\ Pr(T_2 = t - z)$ for all possible z.

$$= \int_0^t Pr(T_1 = z) * Pr(T_2 < t - z)\, dz$$

- This equation has a solution, the Gamma distribution with parameter $\tau_{A|A}$, K_A

How to fit - 2

- **b.2) Simpler simulation method**
  - this program runs many times faster: No loop, no summation (near 2 sec. per simulation) ➔
  - There is still a need for a minimization routine, but many are available (they are built-in in Mathematica, Matlab).

- **One simulation**
  - Repeat 1000 times (for stability)
    - Assume stimulus A is presented
    - get $RT_A \sim \text{Gamma}(\tau_{A|A}, K_A) + T_0$
    - idem for the other accumulator RT_B
    - If $RT_A > RT_B$ then Hit else miss.
    - idem when the other stimulus is being presented
  - Compute
    - RT(hit), RT(miss), RT(CR), RT(FA)
    - P(hit), P(miss). P(CR), P(FA).

Interlude:
Some knowledge of probability theory

- **Since the two accumulators are independent, we can write:**
  - For the percent correct:

$$\Pr(\text{Hit}) = \Pr(RT_A = t) \times \Pr(RT_B > t) \text{ for all } t$$
$$= \int_0^\infty \Pr(RT_A = t) \times (1 - PR(RT_B < t))\, dt$$

  - For the mean RT:

$$\overline{RT}(\text{Hit}) = t \times \Pr(RT_A = t) \times \Pr(RT_B > t) \text{ for all } t$$
$$= \int_0^\infty t \times \Pr(RT_A = t) \times (1 - PR(RT_B < t))\, dt$$

- **However, these equations generally have no solution…**
- **Is it the end?**

- **No.**

- **Many computer programs can find really good approximations to integrals (an area called numerical integration).**

How to fit - 3

- **b.3) Almost analytical method:**
  - the program is a set of numerical integration ➔
  - this runs many times faster: No 1000 replications (near 0.5 sec. per simulation)
  - There is still the need for a minimization routine, but many are available (they are built-in in Mathematica, Matlab).

$$\Pr(\text{Hit}) = \Pr(RT_A = t) \times \Pr(RT_B > t) \text{ for all } t$$

$$= \int_0^\infty \Pr(RT_A = t) \times (1 - PR(RT_B < t))\, dt$$

$$\overline{RT}(\text{Hit}) = t \times \Pr(RT_A = t) \times \Pr(RT_B > t) \text{ for all } t$$

$$= \int_0^\infty t \times \Pr(RT_A = t) \times (1 - PR(RT_B < t))\, dt$$

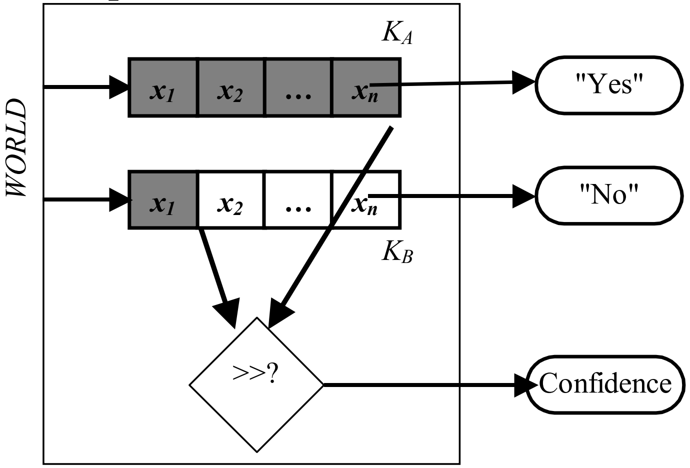
Where are we now?

We described some of the sampling models

- **The MsM (random walk/race models) differ by whether:**
  - the time between two sample is discrete (such as in the random walk model discussed earlier) or random (such as the Poisson race model). See Pike, 1970, Laberge, 1962,
  - the evidence accumulated is discrete (as in both examples) or continuous. See Smith and Vickers, 1980.

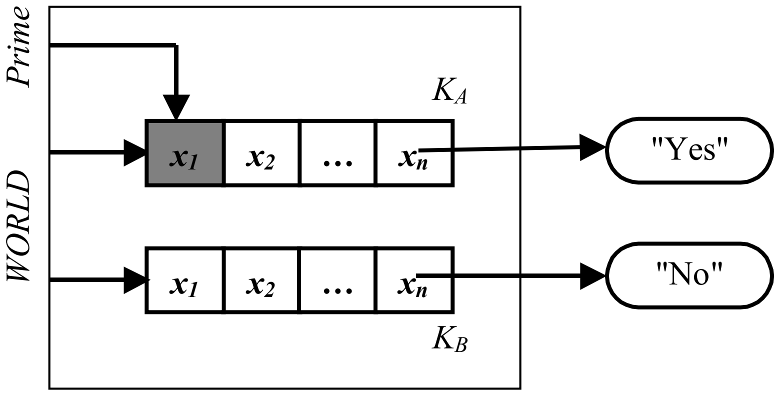
We described some of the sampling models

- **These models are successful in many situations:**
  - Confidence task (Cousineau et al.)
    - Was it a near miss? ➔ low confidence



  - Priming (Huber et al.)
    - Difficulty to monitor the source of the evidence

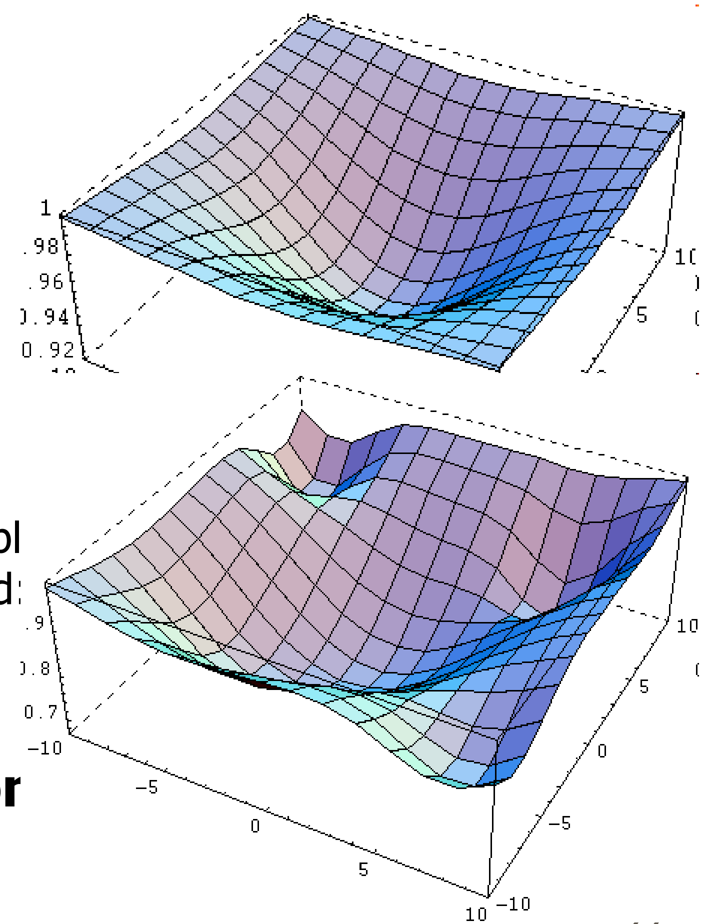
    or
    - Not reset to zero fast enough



  - Same-different (Cousineau), etc.

We described some of the sampling models

- **However:**
  - The race model and the diffusion model are difficult to distinguish
  - They involve many parameters…
  - They are difficult to fit
    - The almost analytical method is not too difficult (not too many local minima) with gradient descent
    - The simulation method is more difficult (a lot of randomness involved from replication to repl and may require an extension of the simplex method: the subplex method available for Matlab.



- **This explain the success of the connectior**
  - They are not fit, they are trained

A general critique addressed to Race models

- **Although they are powerful models apt to fit many situations, they have one limitation.**

- **This limitation is not in the decision mechanism (the trigger) which may have some equivalent at the neural level.**

- **This limitation is in the architecture:**
  - all the evidence in favor of one response must use the same channel
  - It therefore form a sort of "bottleneck"
  - Further, to send information on a single channel, a "spike" code must be used, alternating on and off states.
  - We have no physiological evidence that neurons uses subtle codes

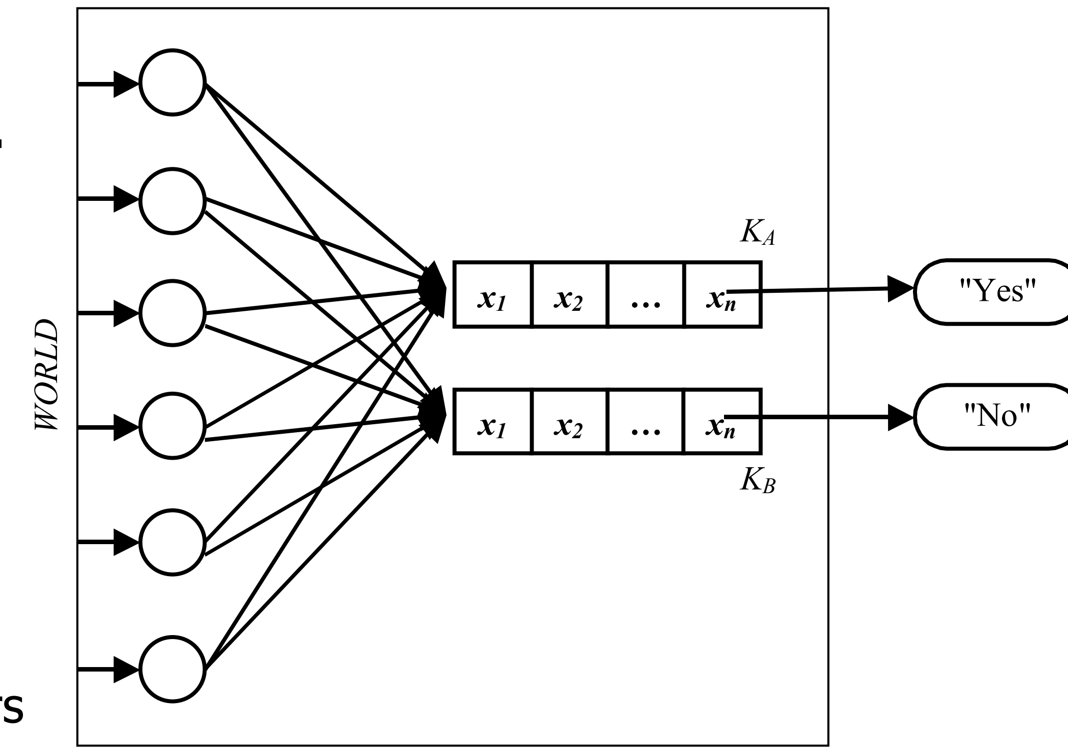
- **If Race models are to represents a useful simplification of the neural mechanisms, a parallel architecture must be introduced.**

A parallel race model

Description of the architecture

(more precisely: a multiple-channel race model, sorry Jim;-)

- **The inputs are either**
  - on or off
- **The connections are either**
  - on of off
- **This rings a bell?**



- **This model is necessarily**
  - a continuous time model
  - a Weibull race model since the distributions of the fastest trigger out of many competitors is Weibull.
- **This model could be either**
  - discrete or continuous evidence accumulator

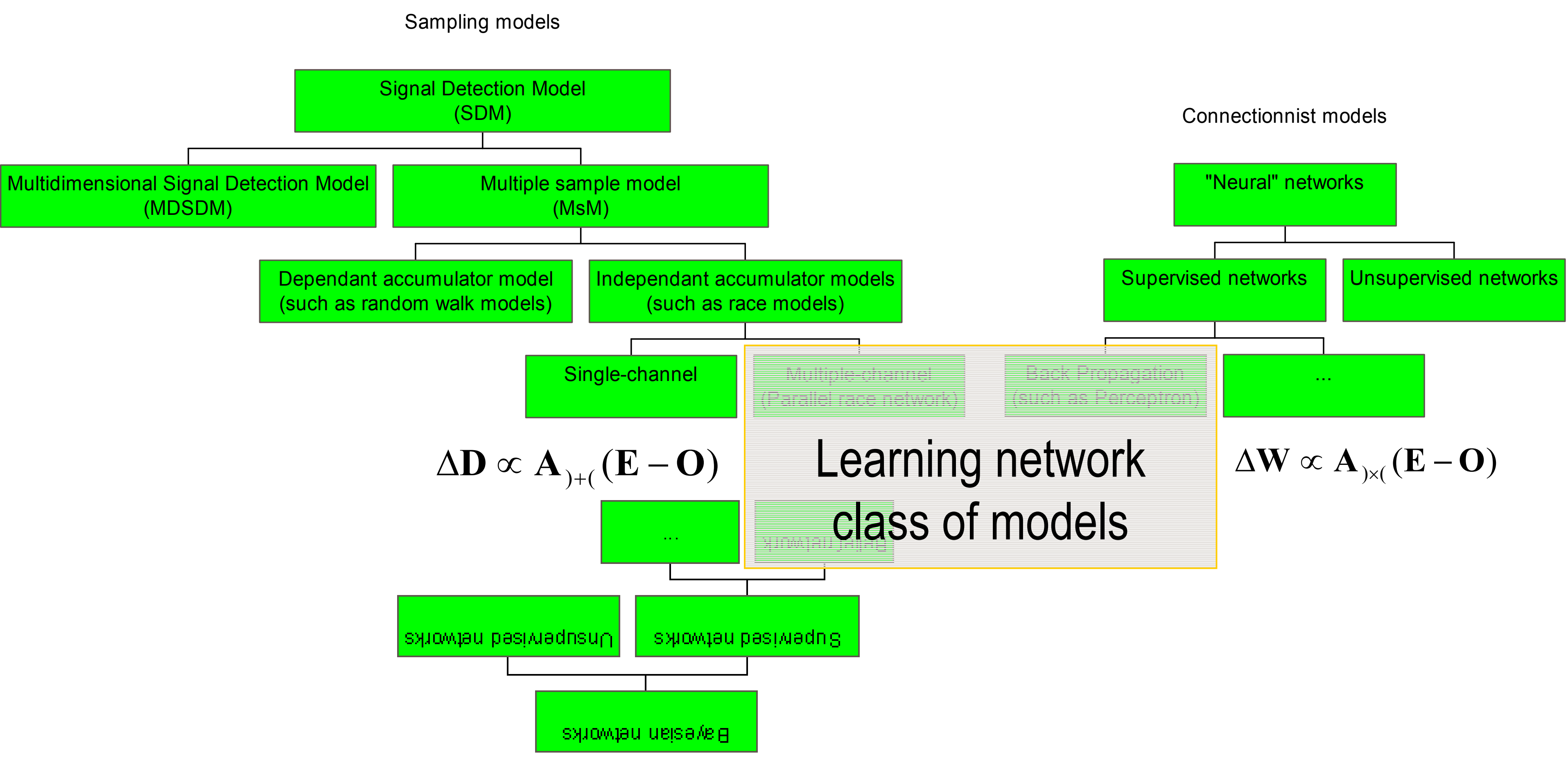
description of the learning rule

- **If we assume that**
  - feedback is provided on errors
  - the connections can be changed with practice
- **Then**
  - this network can learn to associate inputs with outputs using a rule very similar to the Delta rule:

$$\Delta \mathbf{D} \propto \mathbf{A}_{)+(}(\mathbf{E} - \mathbf{O})$$

➔This model is a cousin of the standard connectionist model

Comparison with neural networks

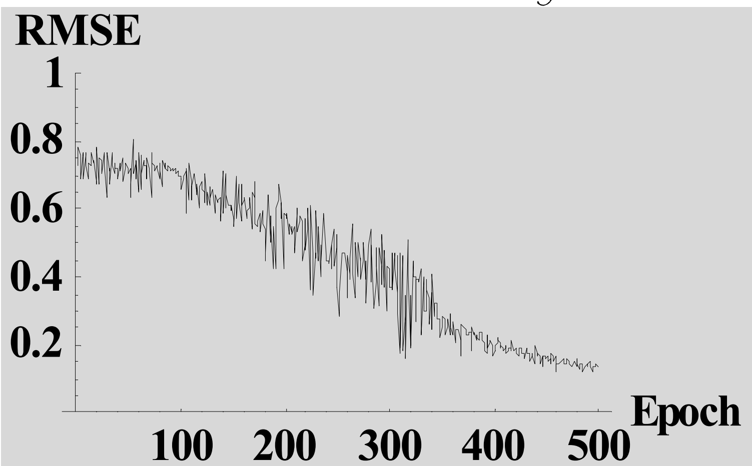


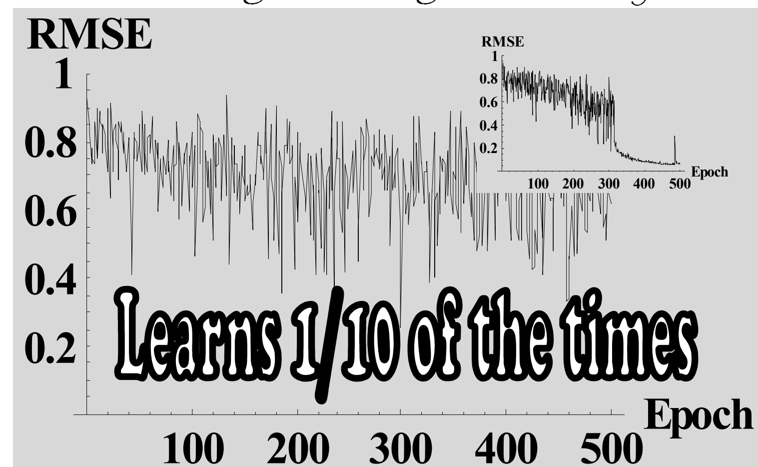
Comparison between race and sum networks

No noise, no redundancy

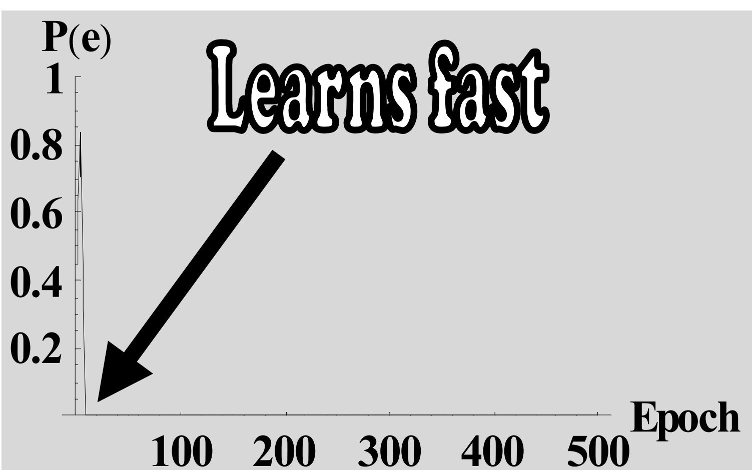
High noise, high redundancy

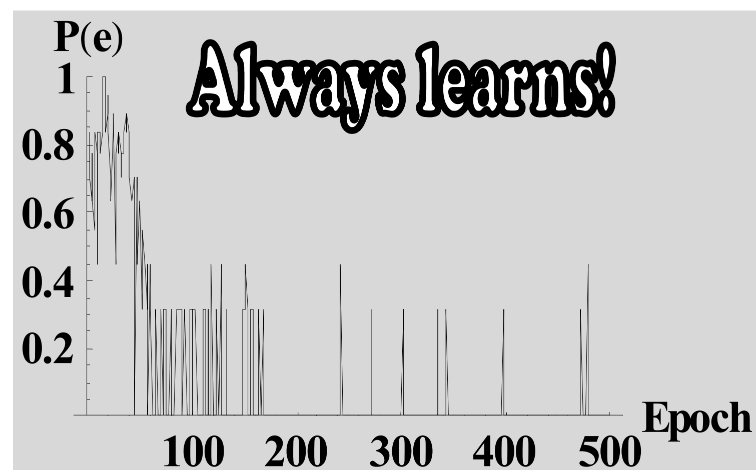
Back Propagation





Race Network





In conclusion

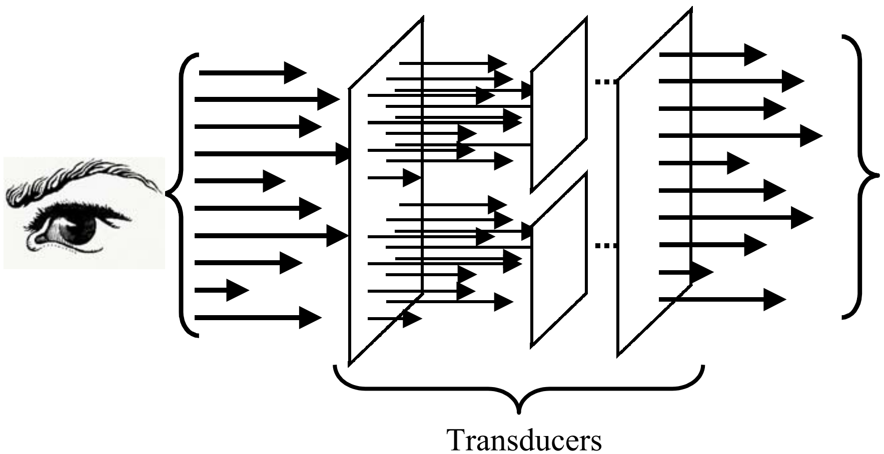
If you can't fit your model, train it.

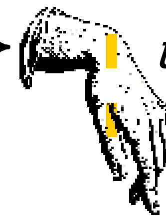
Thank you.

This talk is available at
http://mapageweb.umontreal.ca/cousined

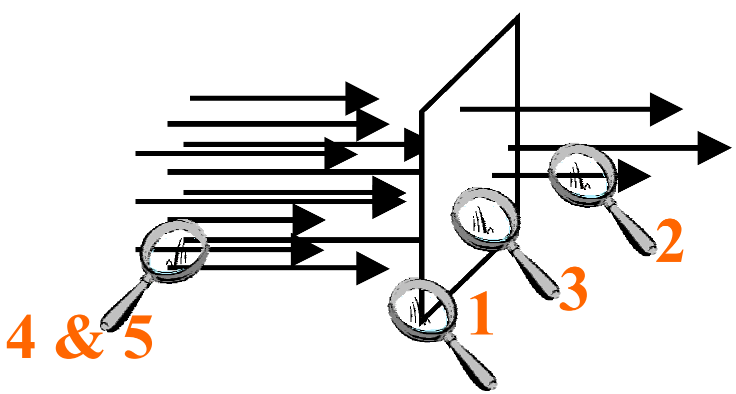
or using
mailto:Denis.Cousineau@Umontreal.CA

Characteristics of networks





- **More than likely:**
  - *the brain is massively parallel*
  - *the brain is slightly serial*
  - *This is an hybrid architecture.*



- **What are the possible characteristics of this network:**
  1. *Is processing all-or-none (cascade model)?*
  2. *Is evidence discrete or continuous and is there noise added to it?*
  3. *What is the decision rule (exhaustive, self-terminating, trigger)?*
  4. *Is there facilitation across channels?*
  5. *Is there redundancy?*
  6. *Is there lateral inhibition on an area?*