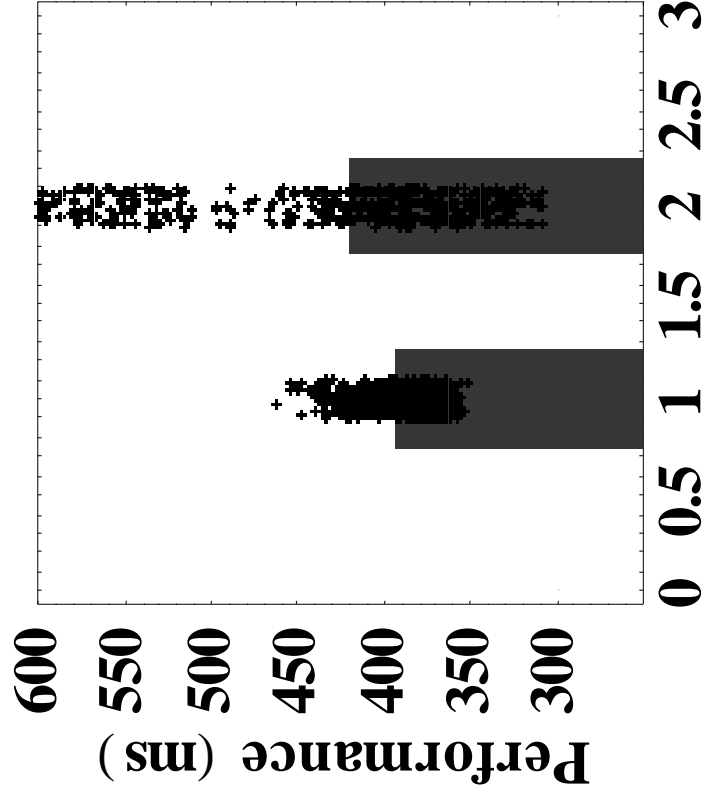


Analyses de la distribution des TR: Perspectives en largeur et en profondeur de la latence



**Denis Cousineau,
Indiana University**

Un TR \neq un TR moyen

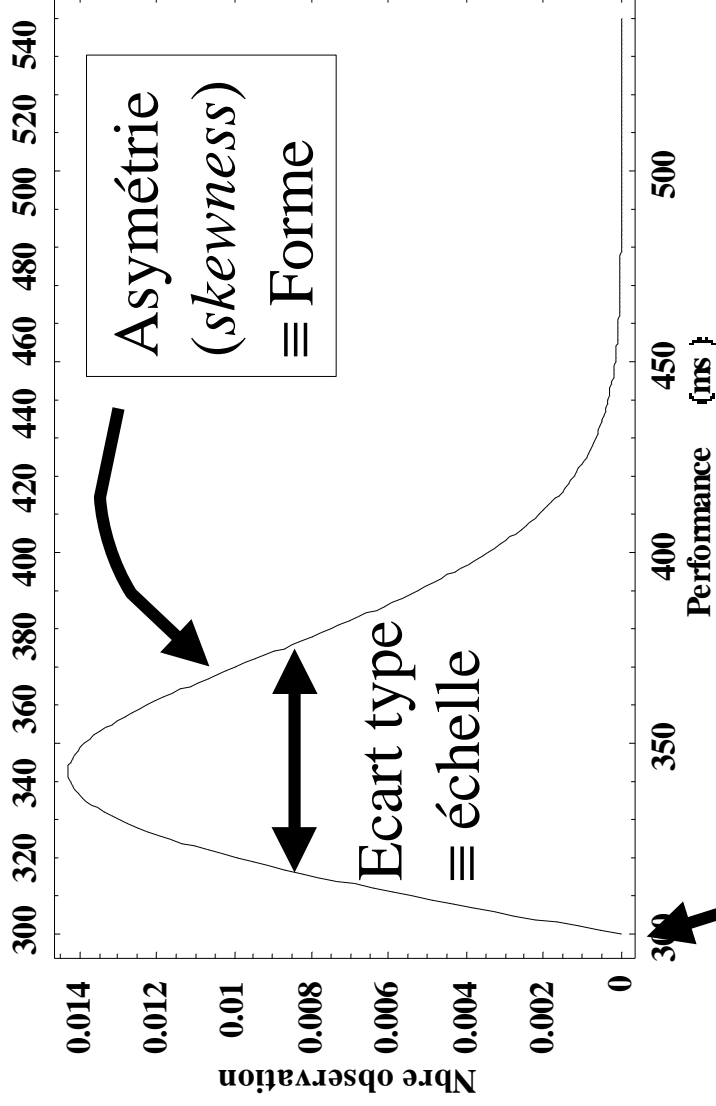


Quelle condition offre la meilleure performance?

Dépend de la moyenne, mais aussi:

- De la présence de TR extrêmes?
- De la variabilité?
- De données bimodales?

Représentation (1): Histogramme et fonction de densité (PDF)



La moyenne d'une distribution asymétrique n'est pas un très bon estimateur de la tendance centrale.

Asymétrie $> 0 \Rightarrow$ Centre de masse décalé vers la gauche;

Asymétrie $= 0 \Rightarrow$ symétrie.

Les 3 paramètres (position, échelle, forme) décrivent efficacement les données.

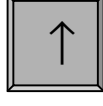
Exemples où (1) est plus diagnostic

1- Heathcote et al.(1991):

3 statistiques plus efficaces qu'une seule.

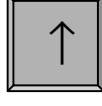
2- Hockley (1984): le débat exhaustif/autocessant

La forme est plus diagnostic que la moyenne.

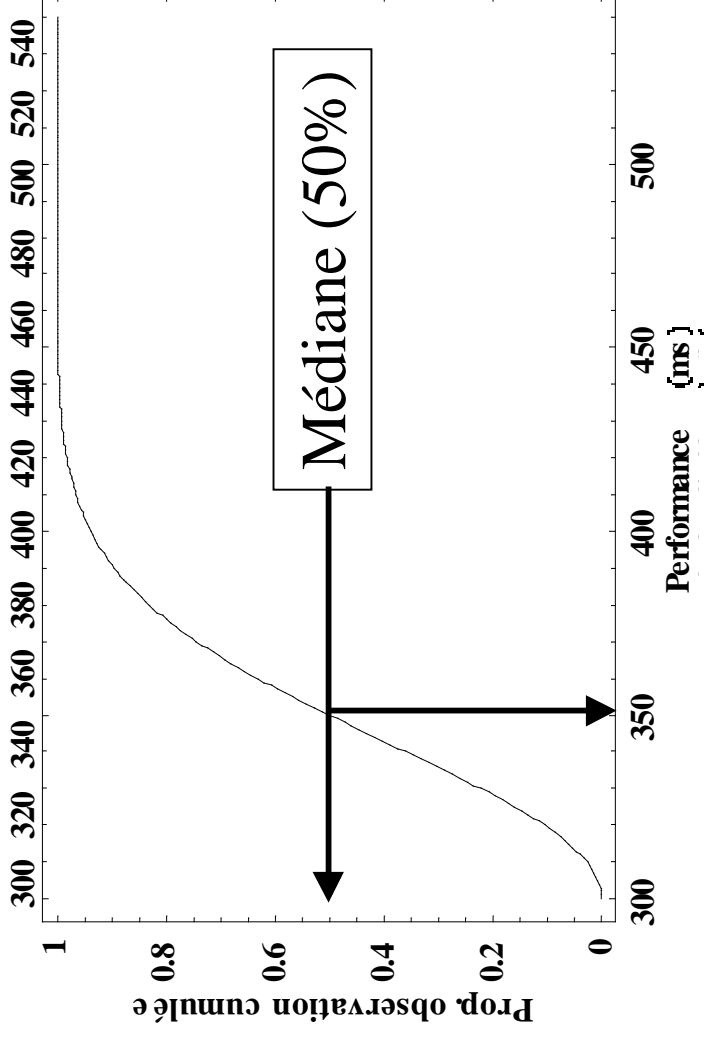


3- Cousineau et Shiffrin: le débat sériel/parallèle

La présence de multimodalité est très diagnostic.



Représentation (2): Fonction de distribution (CDF)



Généralement plus stable que la méthode par histogramme (1).

La médiane est facile à lire, mais pas l'échelle ni la forme.

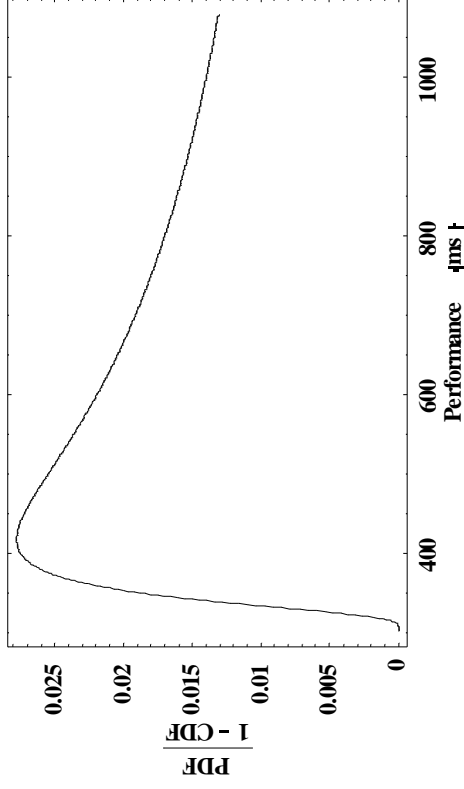
Exemples où (2) est plus diagnostic

1- Les modèles de canaux parallèles (« Separate activation channels ») 

L'inégalité de Miller et l'inégalité de Grice.

2- Les modèles d'échange vitesse-justesse
(speed-accuracy trade-off)

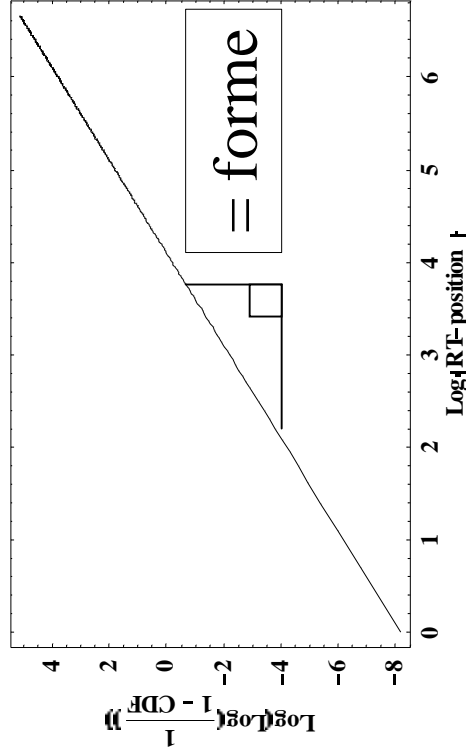
Représentation (3), (4): Fonction de hasard, linéarisation



Fonction de hasard \equiv probabilité que le sujet réponde maintenant s'il ne s'est pas encore arrêté.

Croissant \Rightarrow la réponse est toujours de plus en plus imminente;

Décroissant \Rightarrow la réponse peut ne jamais venir.



Linéarisation de la distribution

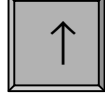
La pente sur ce graphe représente la forme de la distribution des TR.

Position et échelle sont impossible à lire.

Exemples où (3) et (4) est plus diagnostic

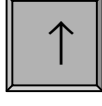
1- Burbeck et Luce: temps limite pour répondre

«La fonction de hasard reflète une quantité d'information accumulée en fonction du temps.»



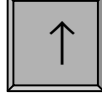
2- Townsend et Nozawa: notion d'énergie

La fonction de hasard est une quantité d'énergie investit dans le traitement.



3- Cousineau et Larochele: effet de pratique

changement de stratégie visible dans le changement dans la forme de la distribution avec la pratique.



Comment travailler avec des distributions de TR?

Les outils sont accessibles:

- Visualisation des données suivant (1), (2), (3) { ou (4) }
(SPSS, Mathematica, MathLab, etc.)
- Extraction de:
 - | statistiques (approche non-paramétrique):
moyenne/minimum, écart type/variance, asymétrie
 - | paramètres (approche paramétrique)
e.g. position, échelle et forme de la distribution de Weibull la plus semblable aux données
via minimisation de l'indice du maximum de vraisemblance (PASTIS).
- Affiner une courbe PDF/Hasard instable
Silverman ('86), Miller et Shingpurwalla ('86)
- Séparer des composantes lors de distributions bimodales
Yantis ('91), Dolan (en préparation)
- Tester des différences entre distributions
Test de Kolmogorov-Smirnov sur CDF (2)

Comment travailler avec des distributions de TR? (suite)

Difficultés:

- le nombre de données brutes n doit être $> > >$
 - supérieur à 100 par condition pour CDF (2) et Linéarisation (4);
 - supérieur à 400 par condition pour PDF (1);
 - supérieur à 2000 par condition pour fonction de hasard (3).
- préférable de procéder par sujet
 - Une technique pour « moyenner » les distributions de plusieurs sujets est de « Vincentiser » (*Vincentize* ou *Vincent Average*) (Dawson, '80). Cependant, cette méthode ne préserve pas la position de chacune des distributions.
 - Procéder par sujet multiplie le travail.

Conclusion (1)

L'analyse des distributions offrent 3 fois plus de puissance discriminatoire que la moyenne seulement

Elle permet de rejeter des modèles efficacement, sans recourir à des conditions extrêmes:

- Nouvelles prédictions - approche théorique vs.
- Nouvelles perspectives - approche analytique.

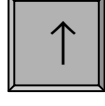
S.V.P.: au moins y jeter un coup d'œil.

Évitez le gaspillage de degrés de liberté!

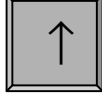
Est-ce tout?

Quelques théories asymptotiques

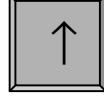
A) les modèles « normaux »
somme pondérée des entrants



B) les modèles de course
le plus rapide détermine la réponse



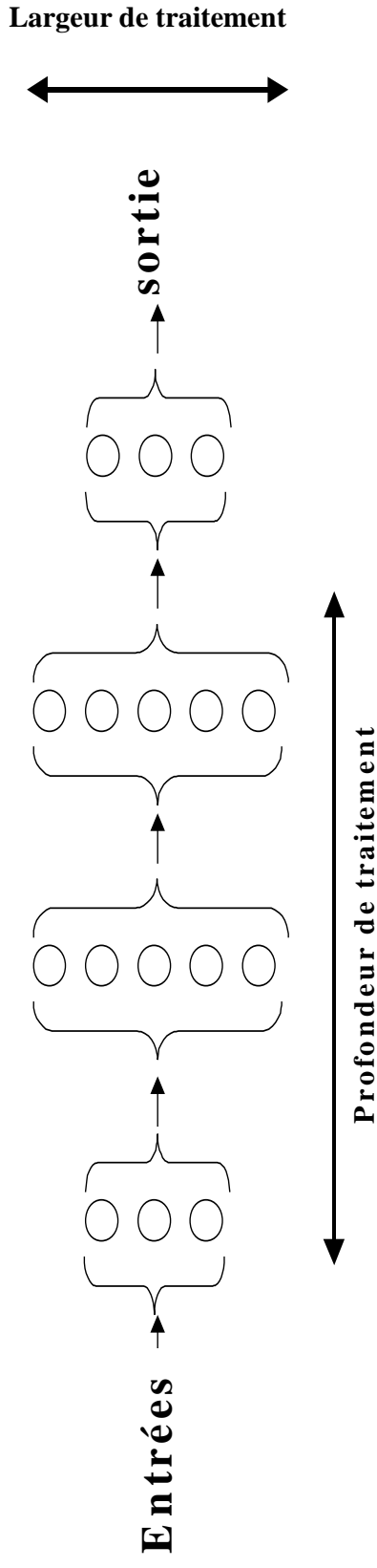
C) les modèles en cascade
séquence de traitements



D) autres modèles?

Compatibilité entre les théories asymptotiques

- I) Les modèles de course et les modèles séquentiels:
sont très semblables aux distributions obtenus empiriquement,
sont compatibles,
...et peut-être plausible biologiquement?



- II) Le modèle « normal » n'est pas compatible ni avec les modèles de course ni avec les modèles séquentiels
⇒ introduire dans une simulation du bruit normalement distribué est certainement la meilleure façon de procéder pour être incapable de reproduire des distributions de TR (très visible avec (4)).

Conclusion (2)

Avantages des théories asymptotiques:

- | Recouvrent de très grandes classes de modèles d'un coup
- | Sont basées sur un nombre très restreint de postulats (2 dans les cas étudiés); tout postulat supplémentaire n'a aucune incidence sur la distribution attendue.

Désavantages des dites théories:

- | Des postulats supplémentaires sont nécessaires pour modéliser l'apprentissage (validité psychologique faible)
- | La distribution de Weibull et la distribution LogNormale sont difficiles à distinguer.

Un réseau qui génère des TR

Objectif: Construire un réseau qui:

- apprend à catégoriser des stimuli simples
- produit une distribution semblable aux distribution de TR
- L'architecture de ce modèle est similaire aux réseaux connectionistes;
- La règle de décision est basée sur l'unité la plus rapide
- Une horloge prend le relais si aucun unité ne répond.

Nom temporaire: Réseau de course.

Idée générale

La règle de décision utilise l'écoulement du temps.

Par exemple, pour catégoriser une conjonction (T), la règle est:

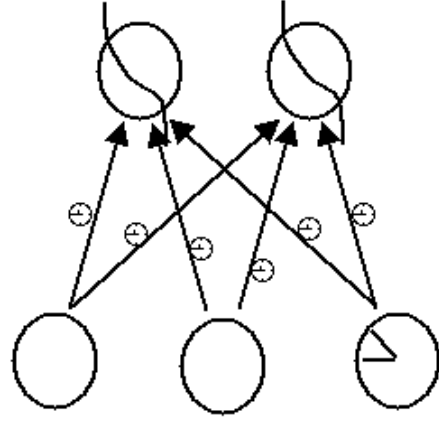
- I Répondre « T » aussitôt que les deux évidences sont détectées**
- I Ou alors, répondre « pas de T » si une seule évidence est détectée**
- I ou alors, répondre « pas de T » après un certain temps [horloge].**

Suppose qu'une absence d'information n'est pas codée.

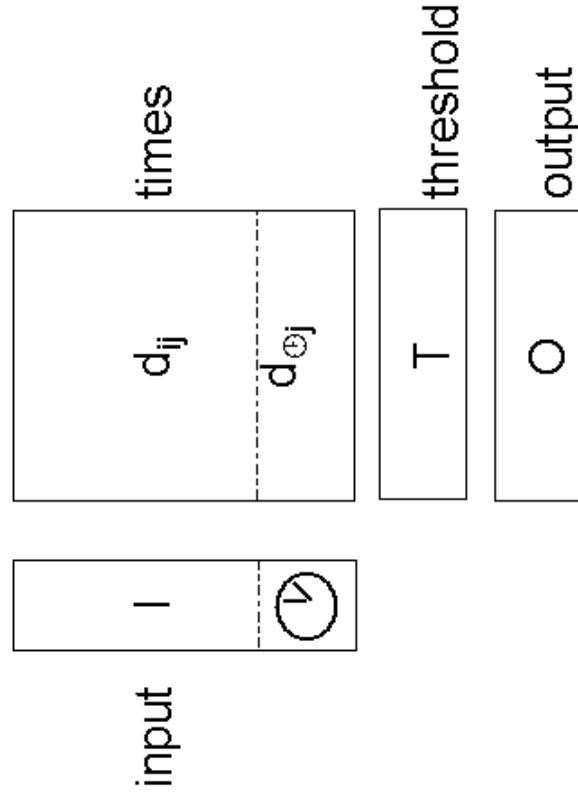
La difficulté réside dans l'apprentissage de délais: « Combien de temps faut-il attendre pour être sûr que la seconde évidence n'apparaîtra pas? »

Architecture

network representation
detectors deciders



matrix representation



\oplus delay d_{ij}

threshold t_j

clock node (time-out)

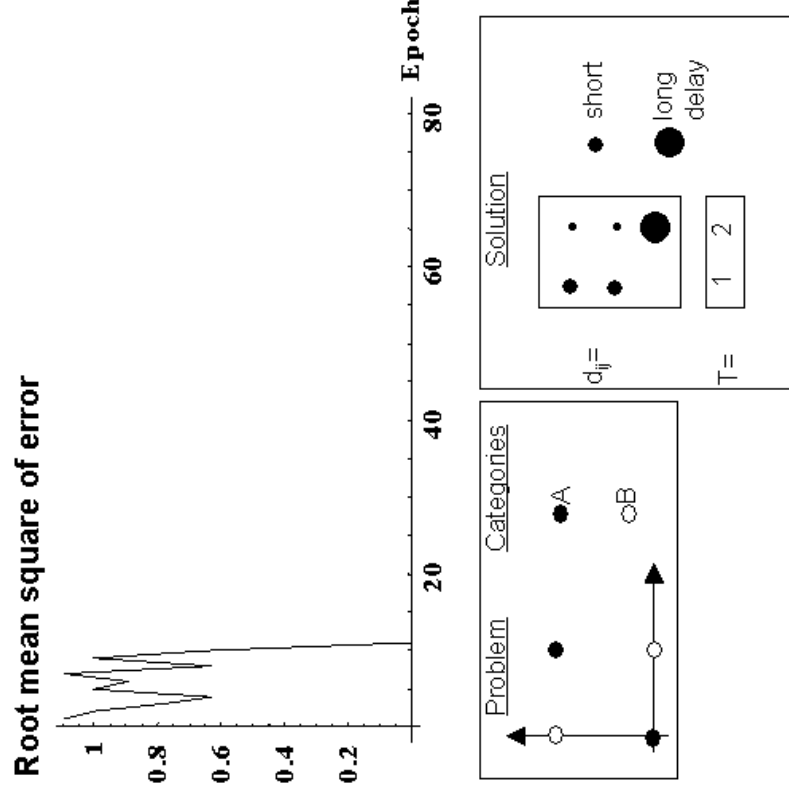
Exemples de fonctionnement

Propriétés du réseau:

a) Il détecte des conjonctions plus rapidement que des disjonctions ou des absences d'entrées (voir Fournier et al. '98, Trabasso et al. '71) ;

b) Il apprend aussi facilement des problèmes non-linéairement séparable que des problèmes linéairement séparables, sans nécessiter de nœuds cachés.

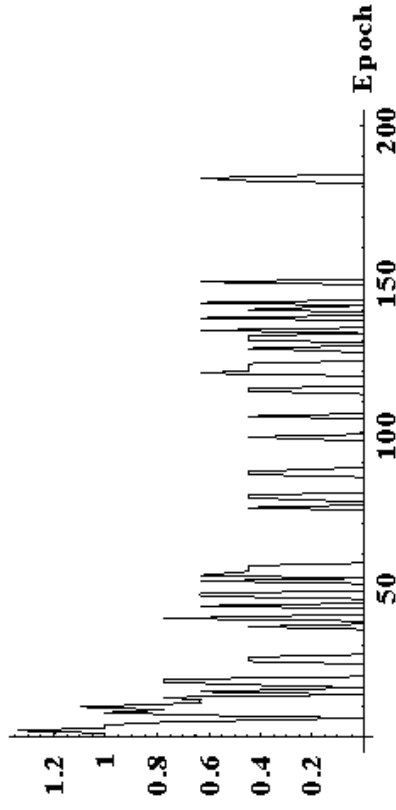
XOR problem



Introduction de bruit dans l'entrée

Noise = 0.025

Root mean square of error



$d_{ij} =$

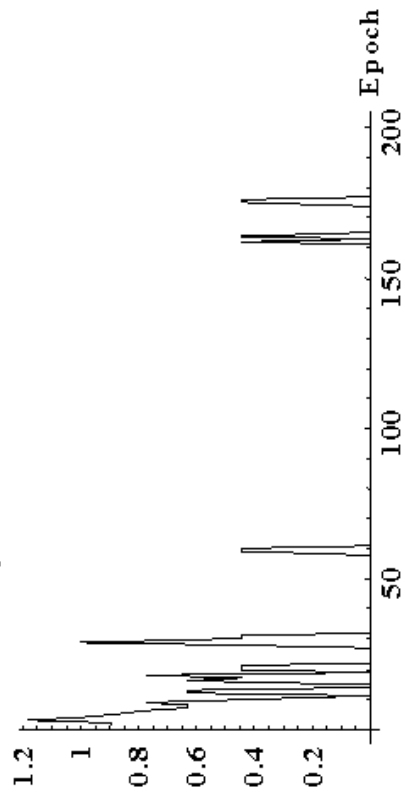
116.0	101.5
113.4	102.1
129.6	129.2

$T =$

1	2
---	---

Noise = 0.010

Root mean square of error



$d_{ij} =$

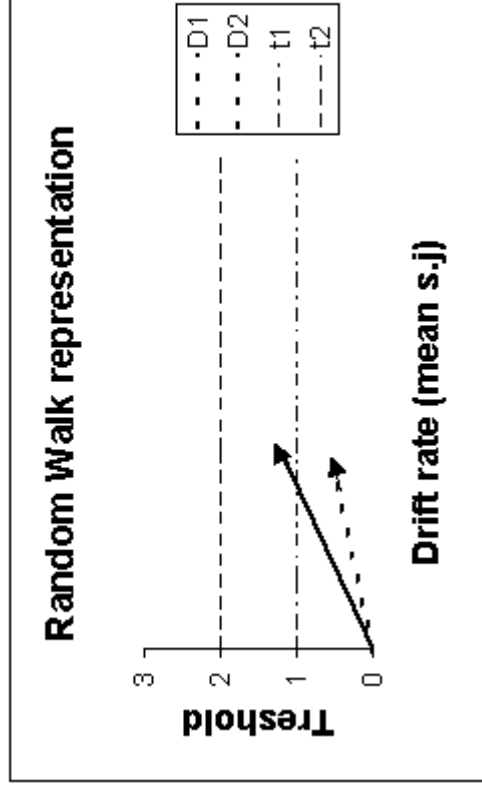
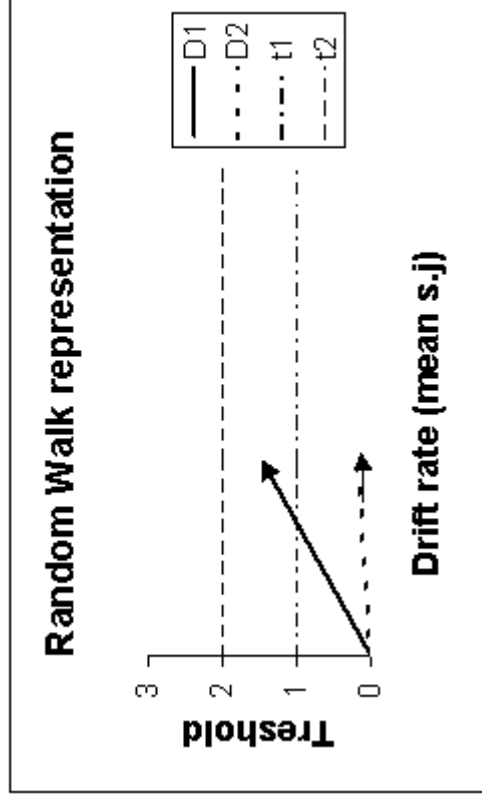
101.1	95.8
103.1	95.6
110.7	110.4

$T =$

1	2
---	---

L'introduction de bruit dans les entrants n'affecte pas la capacité du réseau à atteindre une configuration stable (bruit non-normal, bien entendu...)

Autre représentation



Parameters

	Panel a	Panel b
Free		
F	Exponential	Exponential
η	0.025	0.010
Learned		
$d_j (D_j)$	12.88505	6.4297
t_j	1	1
Thresholds	$t_1 - t_2$	$D_1 - D_2$

Le réseau se représente sous la forme d'une marche aléatoire (random walk model). Le taux d'accumulation des évidences s'accroît avec le bruit dans l'entrée, et se réduit avec la redondance dans les détecteurs.

Conclusion (3)

- L'utilisation du minimum permet de créer un modèle de course en réseau qui:
 - apprend bien,
 - n'a que deux paramètres libres,
 - prédit une distribution des TR de type Weibull
(donc très semblable aux TR des sujets humains)
- Différences entre les types de réseaux:

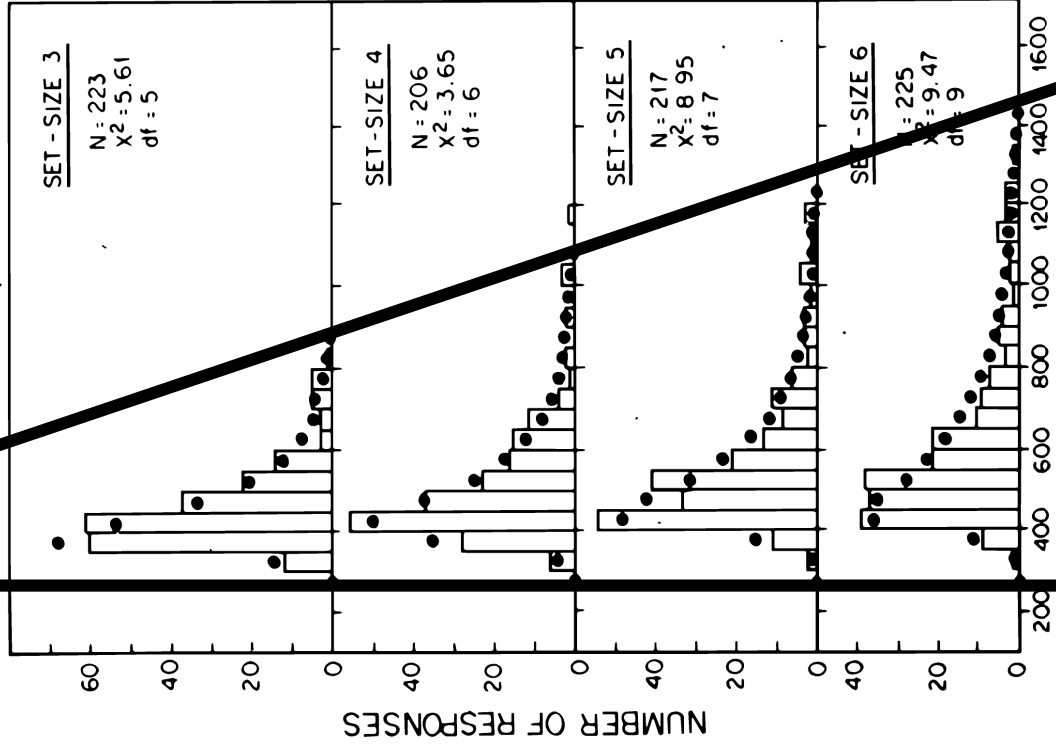
<u>mode d'apprentissage</u>	→	<i>perceptron</i>	←	<u>mode de décision</u>
compétitif	→		←	coopératif
coopératif	→		←	modèle de course
				←
				compétitif

Cette présentation est aussi disponible sur:
<http://Prelude.Psy.Umontreal.ca/~cousined/talks>

Vous pouvez contacter l'auteur à:
DeCousin@Indiana.Edu

Merci de votre attention.

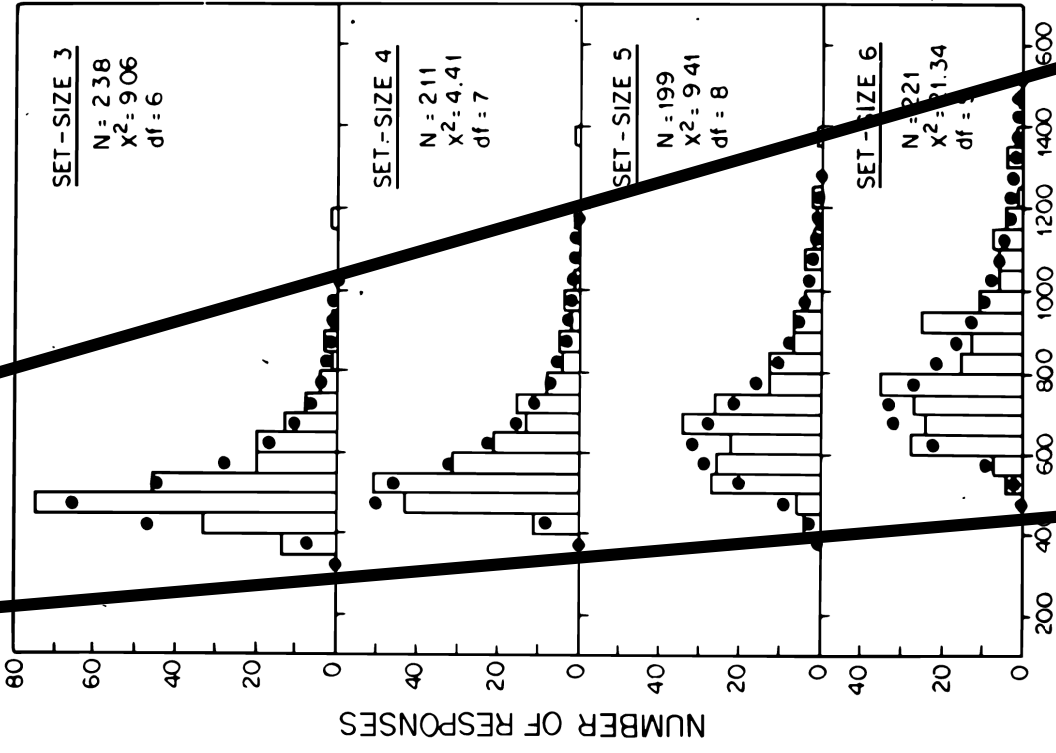
MEMORY SEARCH = exhaustif?



LATENCY (MSEC)

Seul l'échelle
change

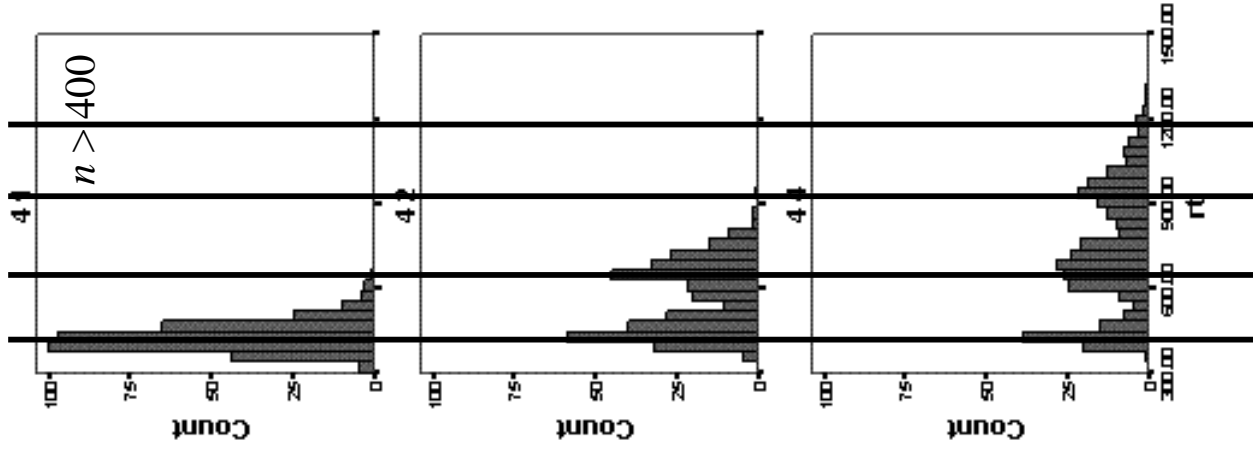
VISUAL SEARCH = autoceasant?



LATENCY (MSEC)

l'échelle ET la position
ET la forme changent

RT distribution for hit



Sujet très entraînée dans une tâche de recherche visuelle; taux d'erreur très bas (<3%).

Le même pic se répète d'une condition à l'autre, à la même position.

Cette mixture de distributions va à l'encontre des modèles parallèles qui postulent une division de l'attention (de l'énergie) entre les différents stimuli.

Inégalité de Miller

Inégalité de Grice

$$CDF_x(t) + CDF_y(t) \geq CDF_{xy}(t) \geq \text{Max}[CDF_x(t), CDF_y(t)]$$

Ces inégalités définissent un intervalle prédisant les performances quand deux stimuli apparaissent sur deux canaux différents.

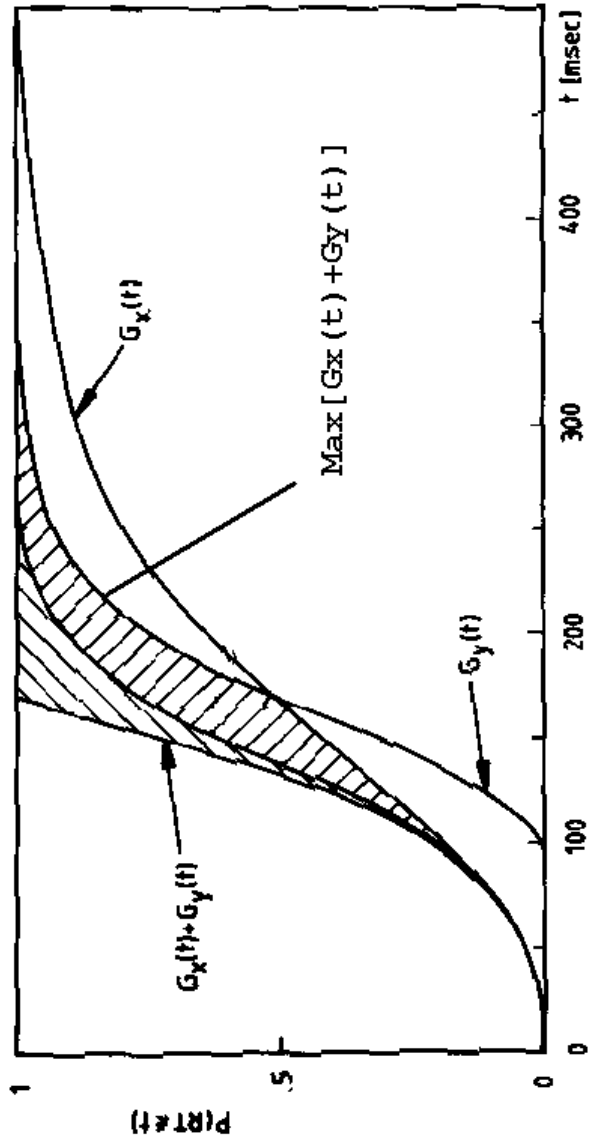
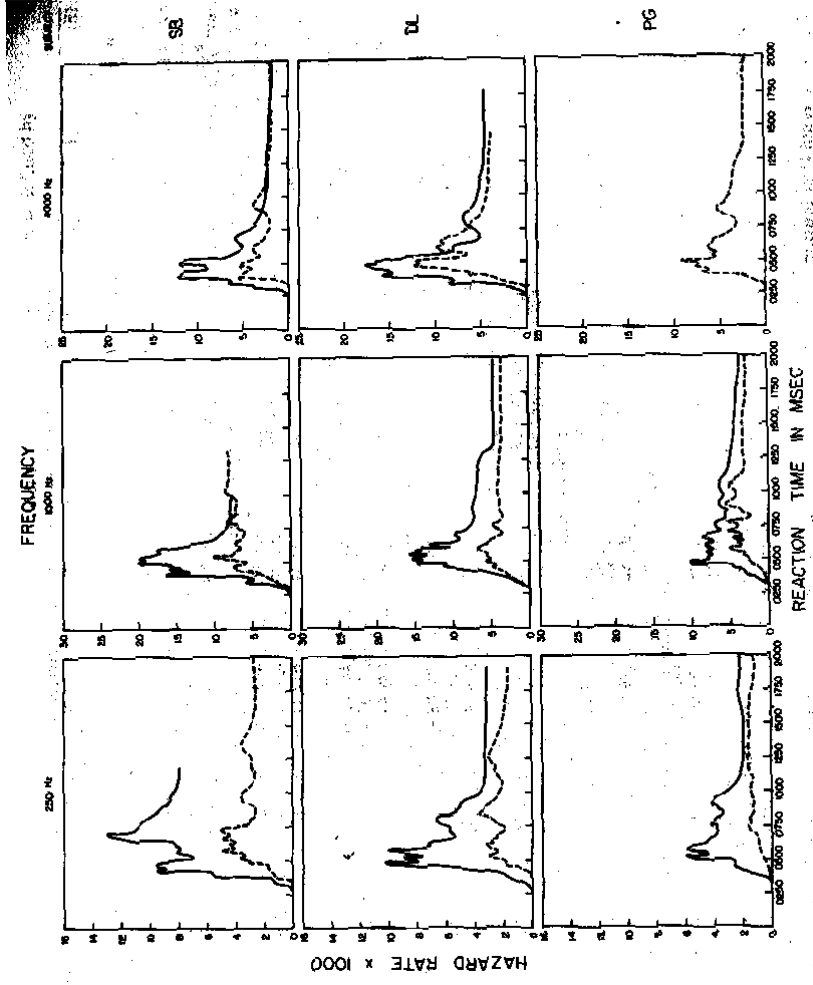


Figure 2. Illustration of the tests for the separate-activation model. Hatched regions (L and U) indicate permissible values of $G_{rr}(t)$. These regions can be constructed using $G_x(t)$ and $G_y(t)$. See text for a more detailed explanation.



L'information ne peut que s'accroître avec le temps. Tout pic dans la fonction de hasard représente l'effet d'un second type de traitement qui interrompt prématurément le traitement normal du stimulus.

Notez que la fonction de hasard est très instable, même quand n est $\gg \gg$ (ici, n est supérieur à 1000).

$$C(t) = \frac{H_{xy}(t)}{H_x(t) + H_y(t)}$$

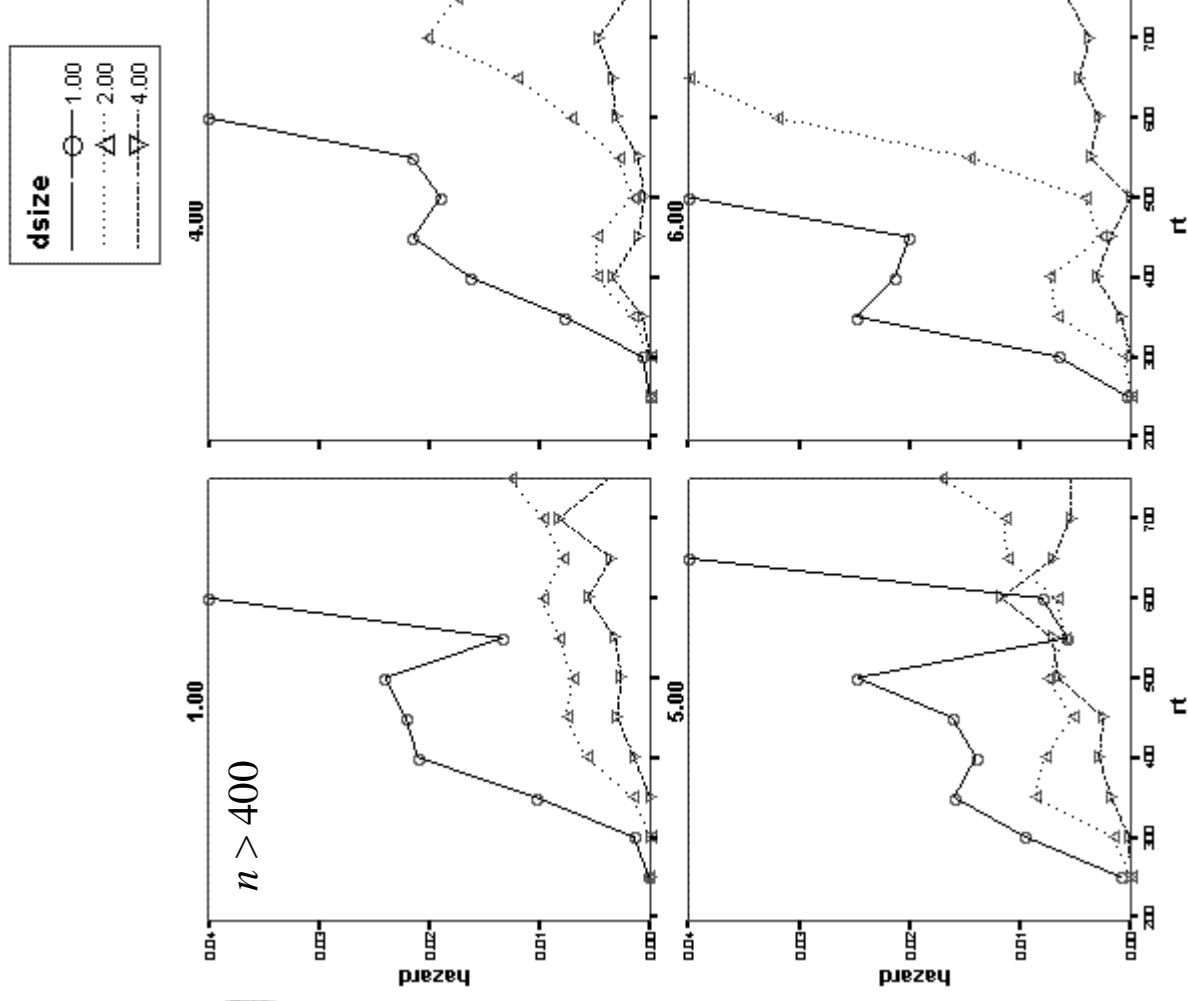
La capacité de traitement reçoit une quantification formelle par le biais de la fonction de hasard (3):

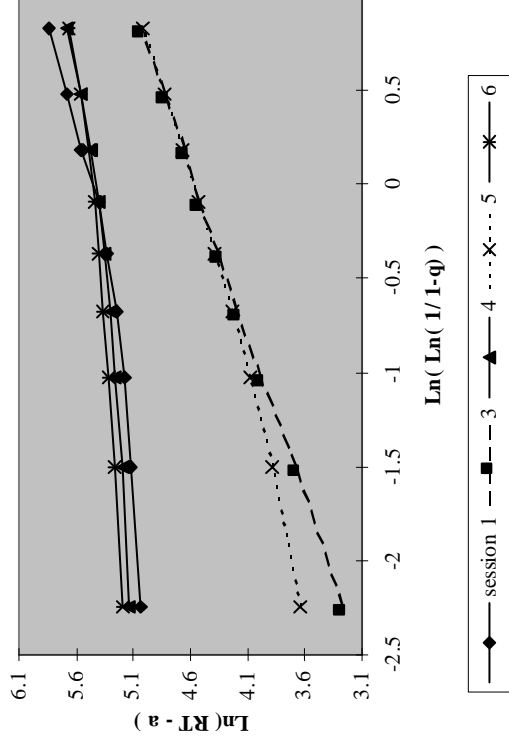
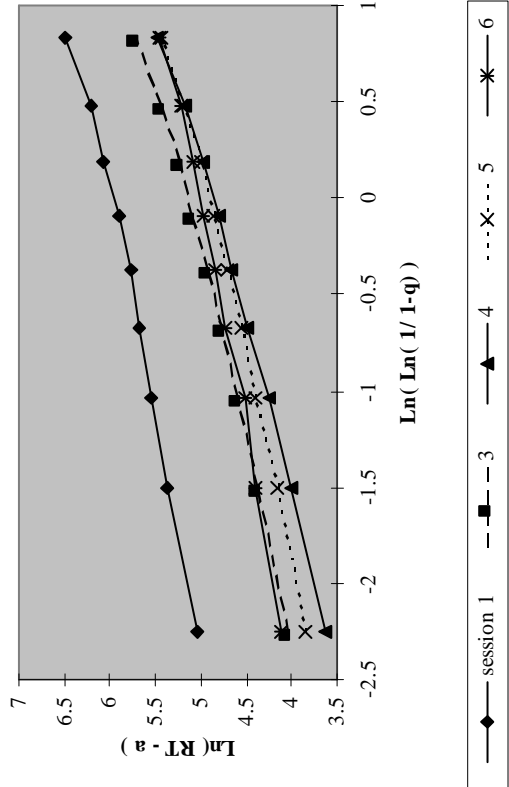
$C(t) > 1 \Rightarrow$ supercapacité

$C(t) = 1 \Rightarrow$ sans limite

$C(t) < 1 \Rightarrow$ limité

$C(t) < 1/2 \Rightarrow$ fixe





- Pour 19 des 24 sujets étudiés, la forme est constante avec la pratique
- Pour 5 sujets, la forme alterne entre deux valeurs distinctes.

Modèles normaux

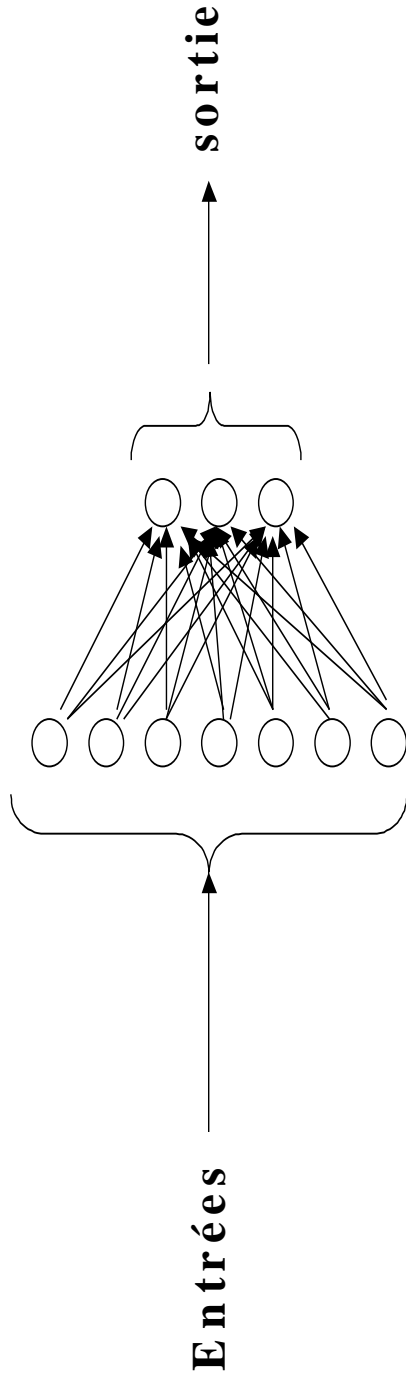
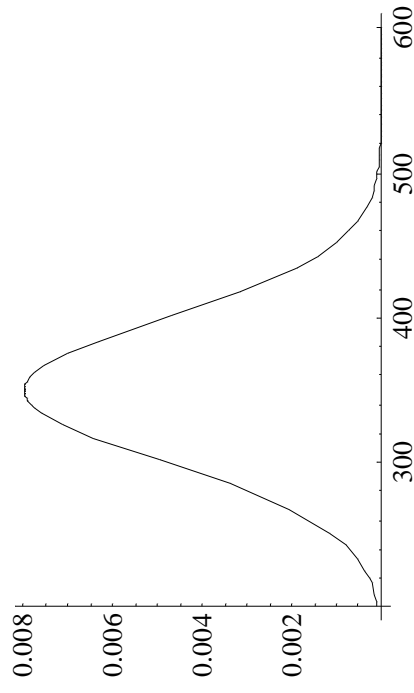
La réponse est effectuée après avoir pondéré l'ensemble des n entrants.

[$n > 30$]

Distribution attendue: distribution normale

Formellement:

Le théorème de la limite centrale...

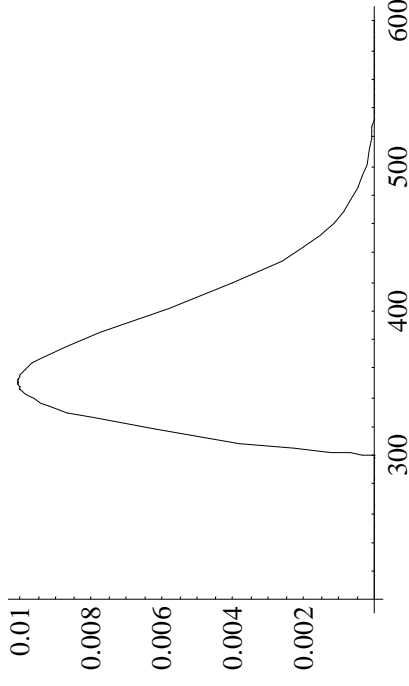


Modèles de course

La réponse a lieu **dès** que la plus rapide des n unités s'active.

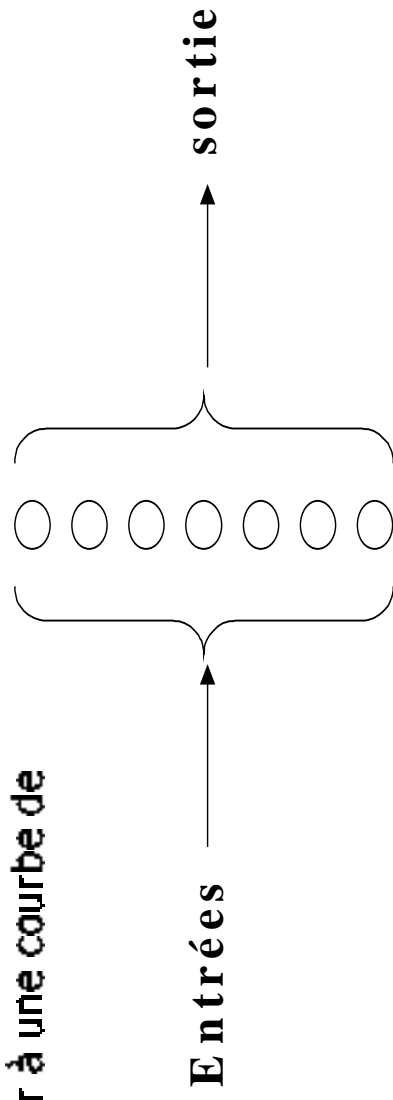
[$n > 50$]

Distribution attendue: distribution de Weibull.



Formellement (Gumbel, '58, Cousineau et al.):

- | chaque unité doit avoir un temps minimum
- | l'accroissement de la CDF dans la partie droite doit ressembler à une courbe de puissance.



Modèles en cascade

La réponse a lieu aussitôt que les n étapes de traitement ont été complétées.
[$n > ?$]

Distribution attendue: distribution LogNormale.

Formellement (Ulrich et Miller, '93):

- | chaque unité doit avoir un temps minimum
- | l'accroissement de la CDF dans la partie droite doit ressembler à une courbe de puissance.

