

Fondements cognitifs de l'arithmétique élémentaire

Stanislas Dehaene

Chaire de Psychologie Cognitive Expérimentale

Quatrième Cours

**Un modèle mathématique  
de la décision numérique**

# A la recherche de lois universelles en psychologie

Multiples niveaux de mise en relation « neuro - psychologique »

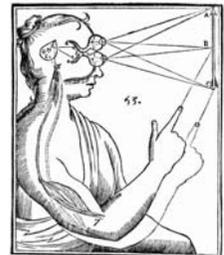
~~Psychologie =  
« Software »~~

~~Neurobiologie =  
« Hardware »~~

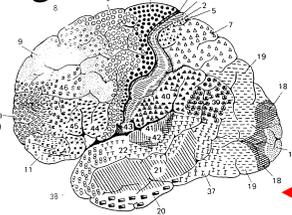
Culture et éducation



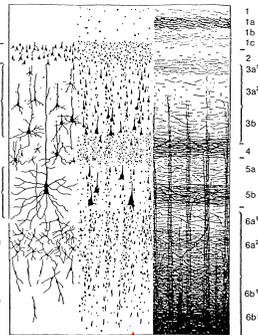
Comportements



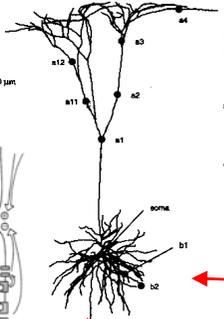
Régions et circuits



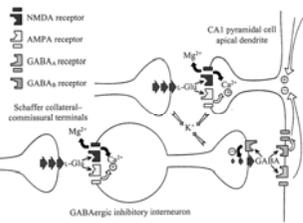
Colonnes corticales



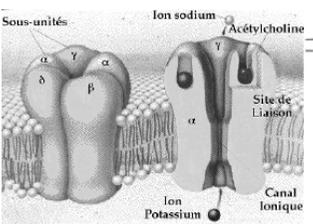
Neurones



Synapses



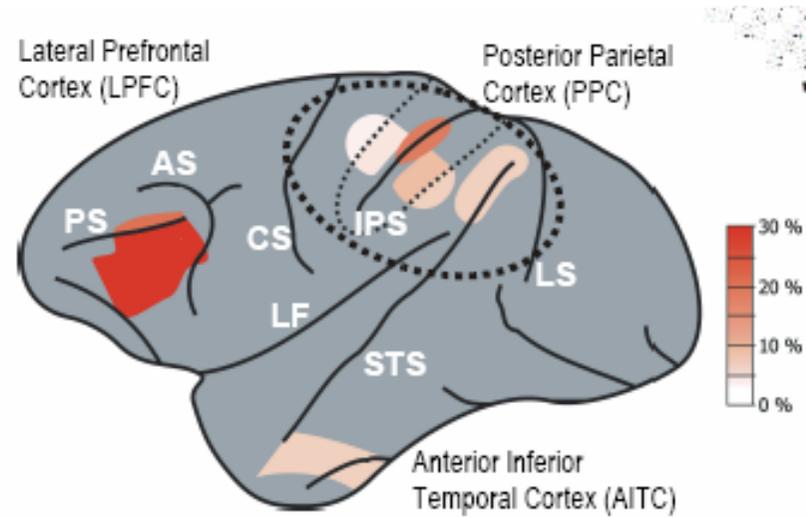
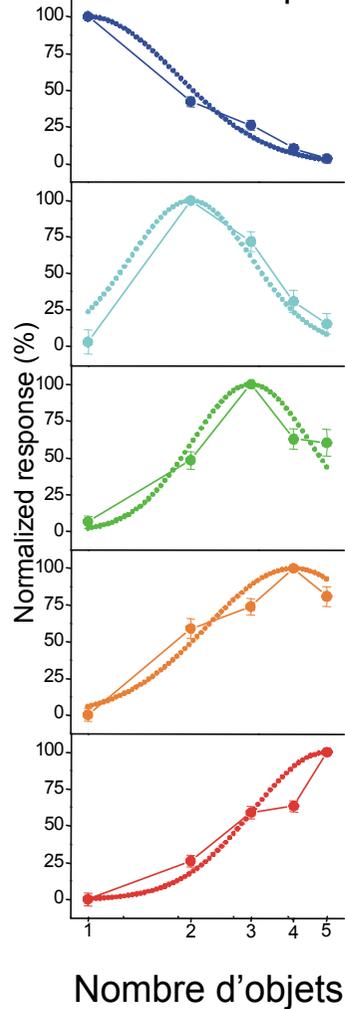
Récepteurs



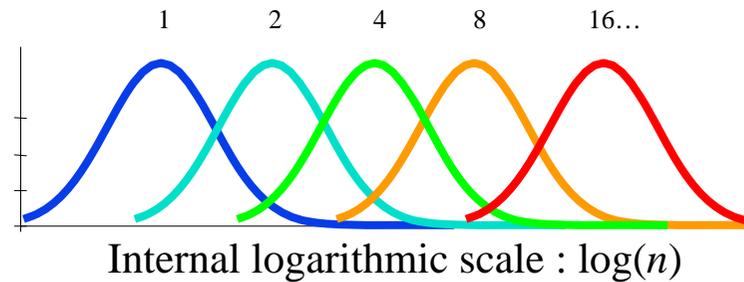
Recherche de lois de transition entre niveaux enchâssés

# Codage « log-gaussien » des nombres par une population de neurones chez le singe macaque

Courbes de réponse  
de neurones uniques



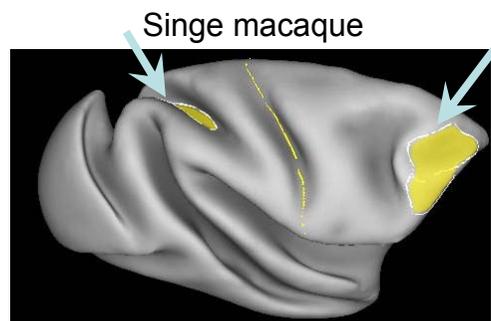
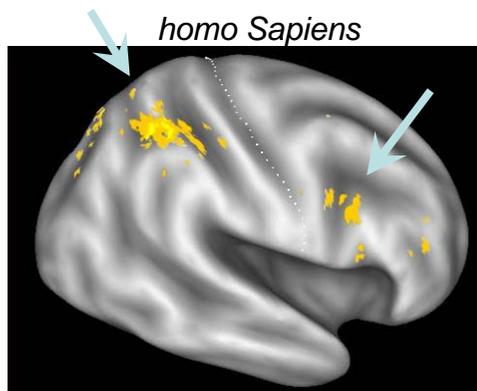
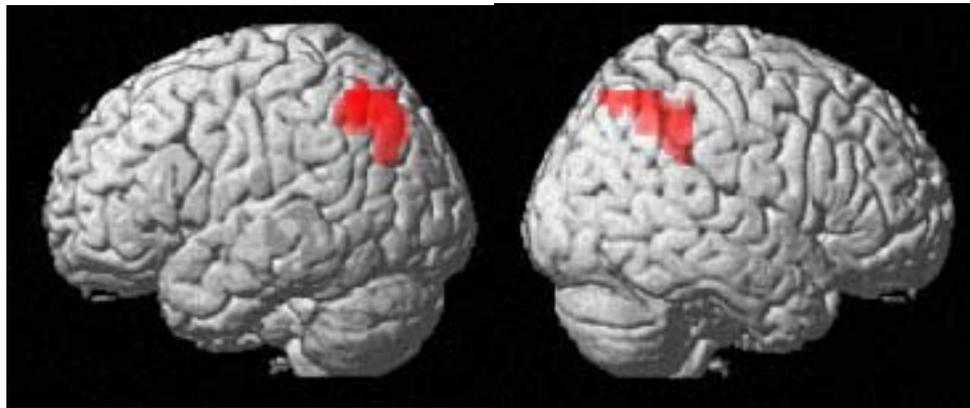
Modèle Log-Gaussien: chaque neurone  
possède une courbe d'accord gaussienne sur  
une échelle logarithmique



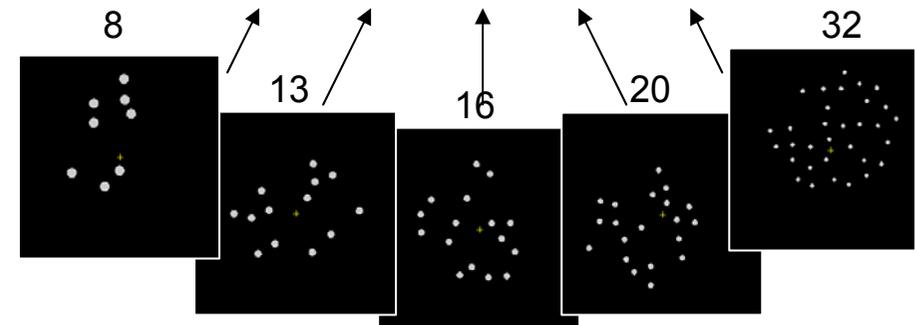
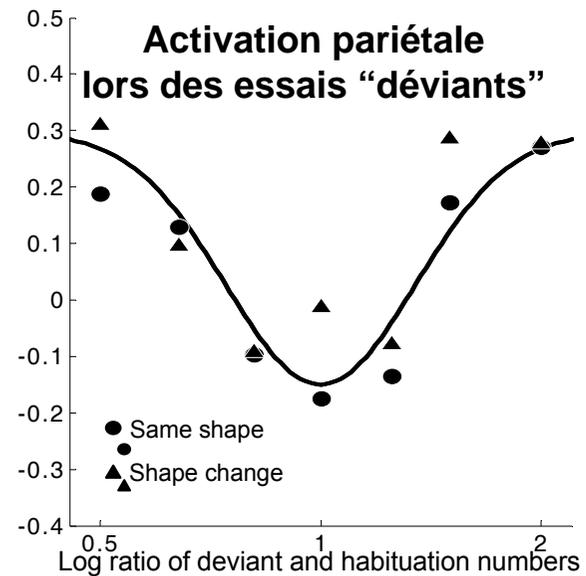
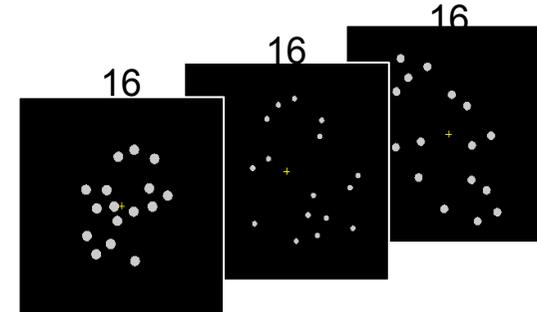
# Plausibilité d'un codage log-gaussien des nombres dans le sillon intrapariétal humain (hIPS)

Piazza, Izard, Pinel, Le Bihan & Dehaene, Neuron 2004

Régions qui répondent à un  
changement de nombre

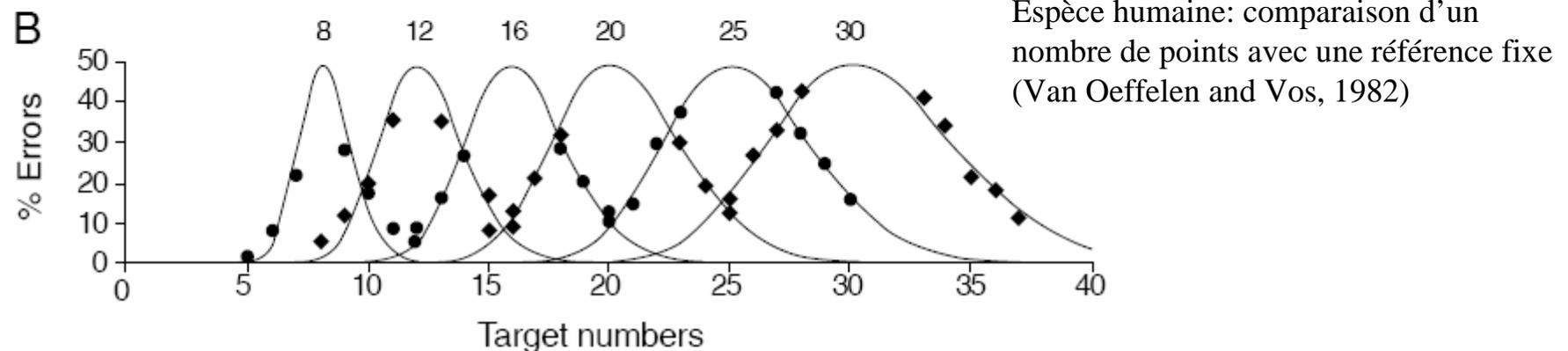
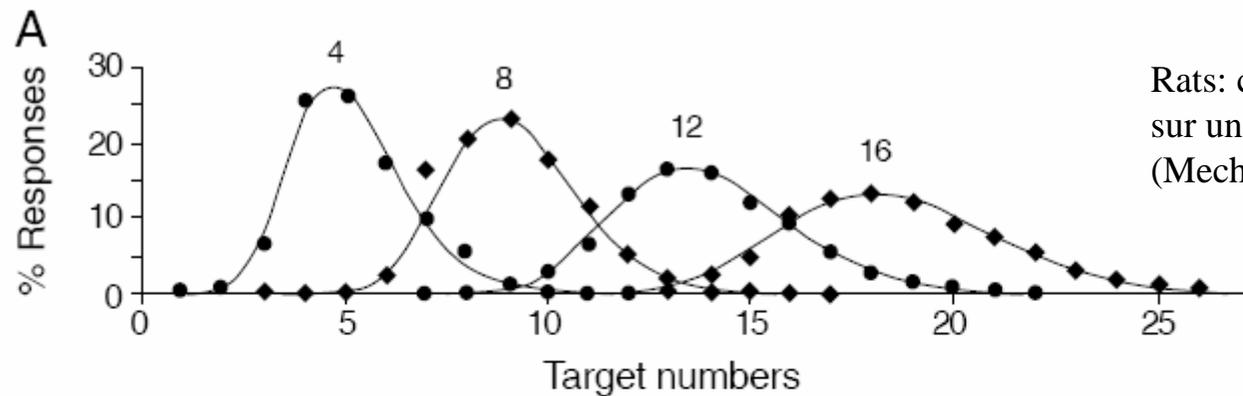


Adaptation à un nombre  
fixé dans un bloc donné



Problème abordé aujourd'hui:  
peut-on mettre en relation directe, par le biais d'une théorie  
mathématique, le codage neural des nombres  
avec les réponses comportementales animales et humaines?

Exemples de comportements numériques remarquablement réguliers, décrits dans le premier cours.



# 1. De la préférence des neurones individuels au codage par population d'une quantité

Problème du décodage neuronal:

Etant donné la décharge d'un groupe de neurones, peut-on déterminer quelle quantité est codée?

L'une des solutions consiste à définir le « vecteur de population », qui est la somme (ou la moyenne) des préférences des neurones pondérées par leur taux de décharge.

$$X = \sum_i p_i f_i(n)$$

où  $f_i(n)$  est le taux de décharge du neurone  $i$ , et  $p_i$  sa quantité préférée (ici le logarithme du nombre d'objets)

Dayan and Abbott (*Theoretical Neuroscience*, 2001, pp. 106–108) démontrent que, dans le cas de neurones avec un taux de décharge distribué suivant une loi de Poisson et dont la moyenne suit une courbe d'accord Gaussienne, le “vecteur de population” est le meilleur estimateur de la quantité codée par une population de neurones (au sens du maximum de vraisemblance).

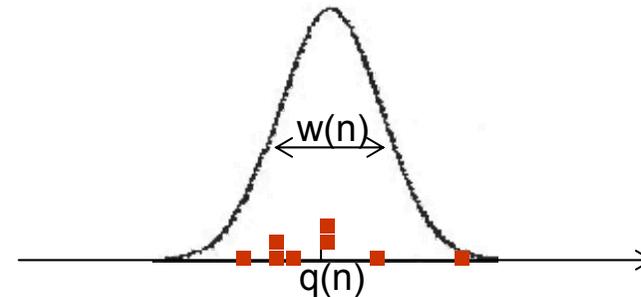
Leur analyse montre que, sous certaines conditions,  $X$  est également une variable Gaussienne de moyenne  $q(n) = \text{Log}(n)$  et de déviation standard  $w(n) = w =$  constante

Autrement dit...

Chaque nombre perçu est codé par une quantité interne  
*analogique, fluctuante, et Gaussienne*

C'est le cadre théorique classique de la psychophysique Thurstonienne :

La perception d'un stimulus d'intensité  $n$  est codée par une variable aléatoire Gaussienne  $X$  sur un continuum interne.



- Quantité évoquée par un stimulus de numerosité  $n$  lors d'un essai donné
- Distribution de ces quantités

Remarques importantes:

- Un nombre *entier* est codé par une quantité continue (un *réel*) et variable d'un essai à l'autre.
- Le meilleur modèle permettant de passer d'un code neural à une variable psychophysique et à une décision comportementale reste très débattu (Shadlen et al., 1996; Parker and Newsome, 1998; Shadlen and Newsome, 1998)
- Selon Alex Pouget et coll., une population de neurones ne représente pas seulement une valeur, mais également la distribution d'incertitude sur cette valeur.

## 2. Du codage d'une quantité fluctuante à la prise d'une décision élémentaire

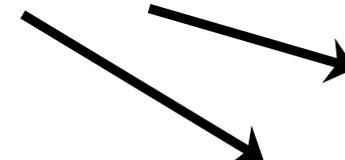
Sachant que les quantités sont représentées par des variables aléatoires fluctuantes, entachées d'incertitude, quelle stratégie adopter pour minimiser les erreurs (ou maximiser le renforcement)?

- Partant d'une représentation interne sous forme d'une variable aléatoire ( $X$ )
- Nous souhaitons choisir, parmi un ensemble de réponses possibles  $R_i$ , la réponse  $R$  qui a la plus grande probabilité d'être correcte, étant donné l'échantillon  $X$
- Il nous faut donc maximiser la probabilité que  $R_i$  soit correcte, étant donné  $X$ .  
Il faut donc calculer  $P(R_i | X)$  et diviser le continuum en sous-régions où le maximum est  $R_1, R_2, \text{etc...}$

- Appliquons la règle de Bayes:  $P(R_i | X) = P(X | R_i) P(R_i) / P(X)$



Quantité connue: c'est la moyenne des Gaussiennes de tous les stimuli qui peuvent conduire à la réponse  $R_i$



Constante indépendante de  $i$

- Souvent toutes les réponses sont équiprobables
- Ou leur probabilité peut être estimée à travers les essais

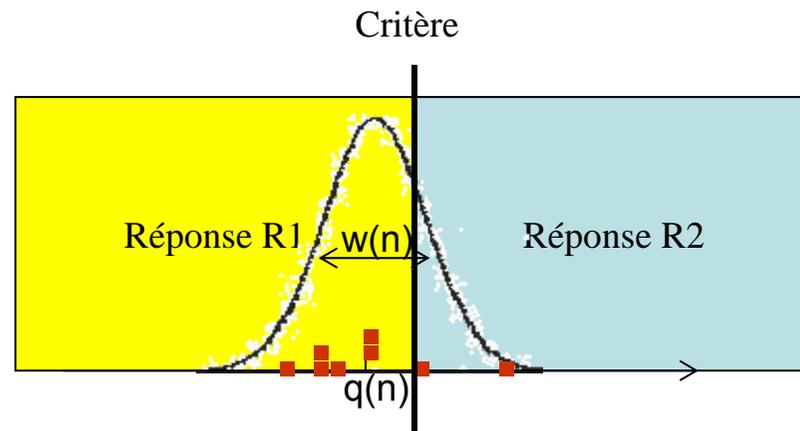
## 2. Du codage d'une quantité fluctuante à la prise d'une décision élémentaire

- Cas particulier des décisions binaires: on veut simplement savoir lequel de  $P(R1 | X)$  et de  $P(R2 | X)$  est le plus grand. On peut se contenter de calculer une seule quantité, le log du rapport de vraisemblance, et examiner s'il est positif ou négatif

$$LLR(X) = \text{Log} ( P(R1 | X) / P(R2 | X) ) =$$

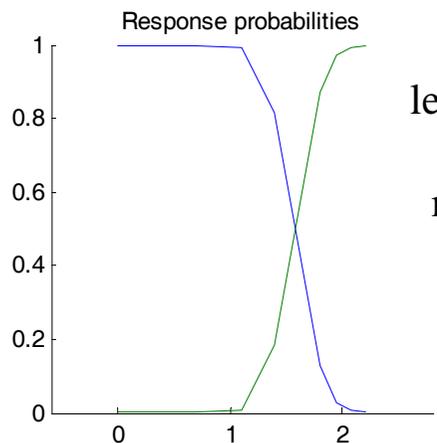
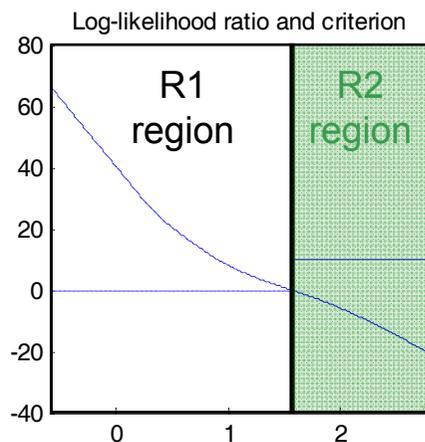
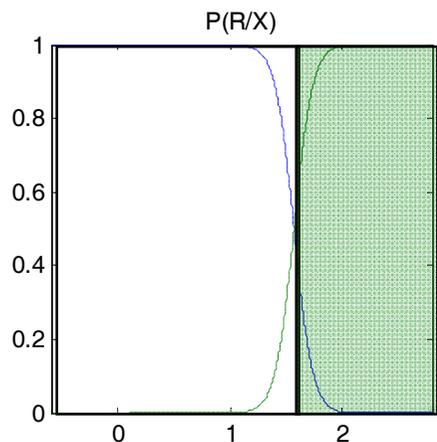
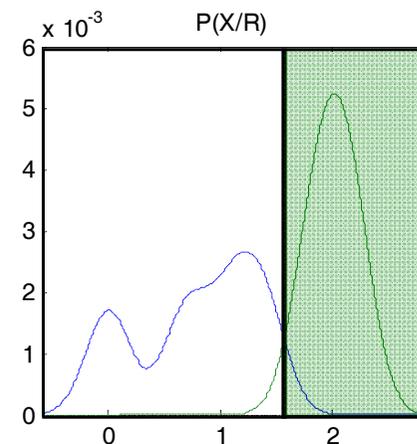
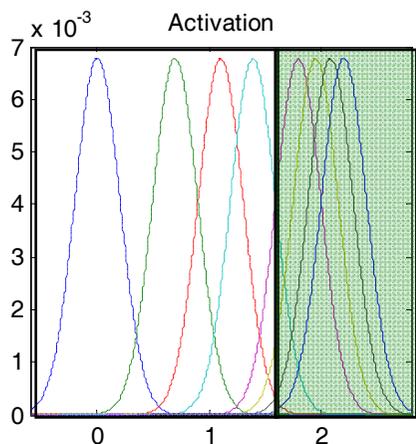
$$\text{Log} ( P(R1)/P(R2) ) + \text{Log} ( P(X | R1) / P(X | R2) )$$

- En résumé: la stratégie optimale consiste à diviser le continuum interne en sous-régions, chacune associée à une réponse optimale. L'analyse bayésienne nous dit où placer les frontières (les « critères ») de façon optimale



- Quantité évoquée par un stimulus de numérosité  $n$  lors d'un essai donné
- Distribution de ces quantités

C'est la théorie classique de la détection du signal



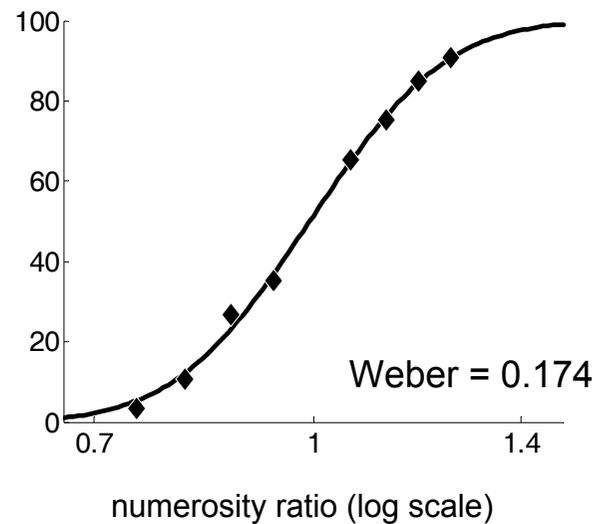
Prédiction:  
 le taux d'erreur est une fonction sigmoïde  
 de la distance entre les nombres,  
 mesurée sur une échelle logarithmique:  
 $f(\text{Log}(n) - \text{Log}(\text{ref}))$

Où  $f$  est l'intégrale d'une Gaussienne

# Modélisation de la comparaison d'un nombre avec une référence fixe

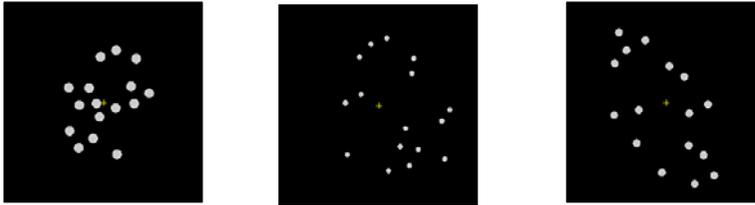
Réponse R1 = plus petit  
 Réponse R2 = plus grand  
 que la référence

Exemple: comparaison des nombres  
 1 à 9 par rapport à 5

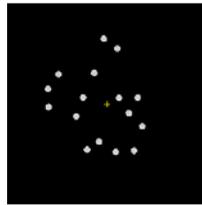


# L'expérience de Manuela Piazza et coll.

Trois exemples de la référence (16 or 32)



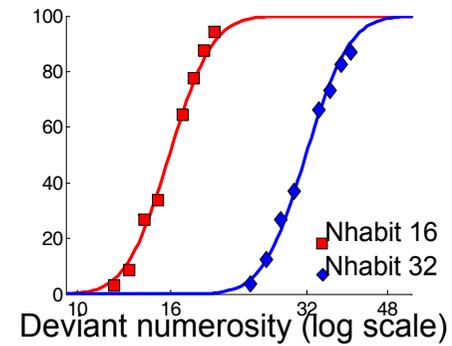
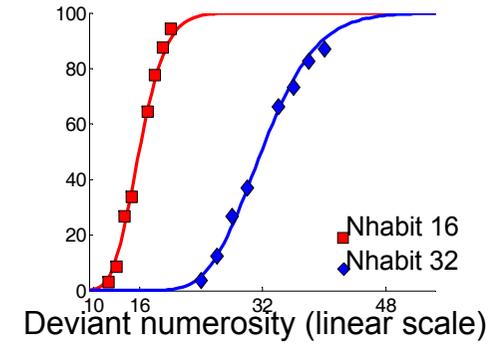
Suivi d'un nombre cible:



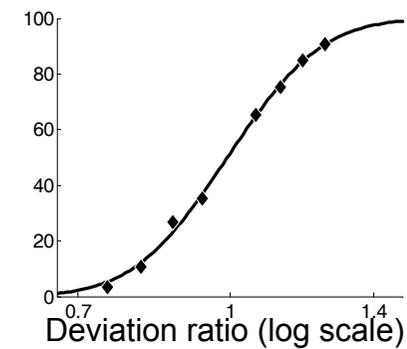
Plus grand ou petit?

### Larger-Smaller Task

% "larger" responses



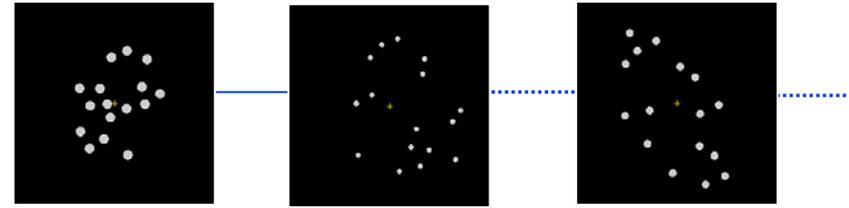
$w = 0.174$



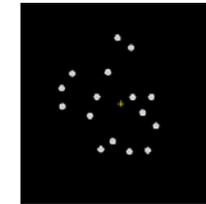
# La précision du codage numérique évolue avec l'âge et distingue les enfants en difficulté

(Piazza, Zorzi, Dehaene et al., travaux en cours)

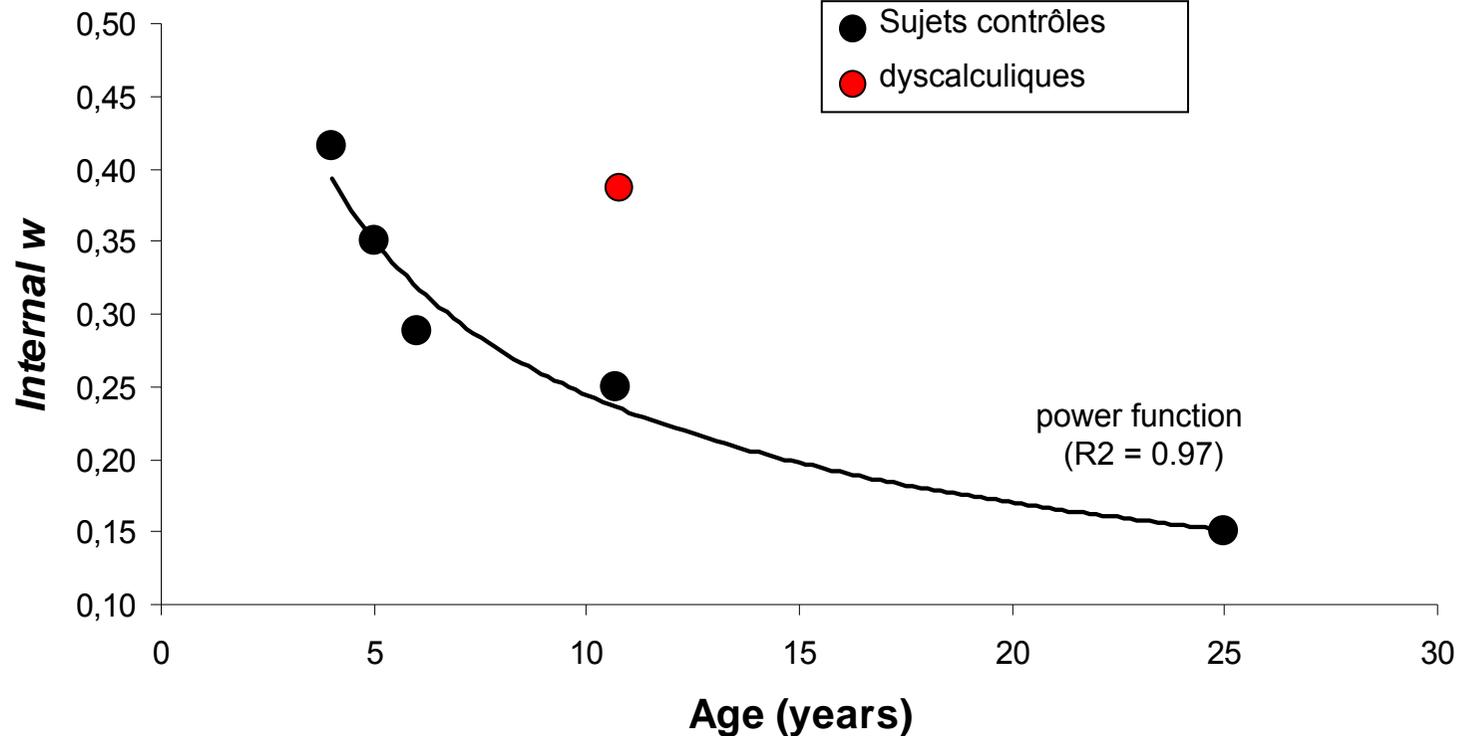
Trois exemples d'un certain nombre (16 or 32)

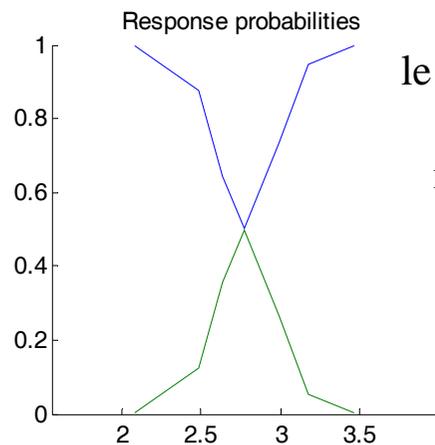
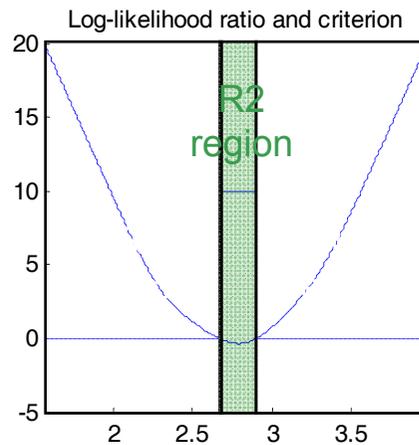
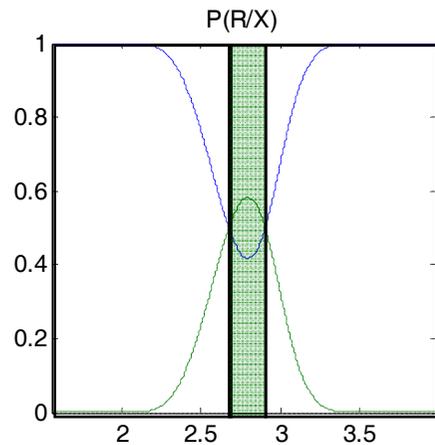
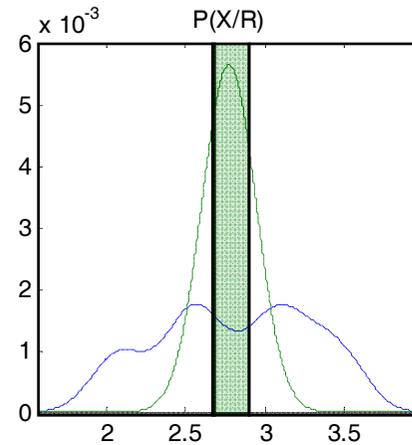
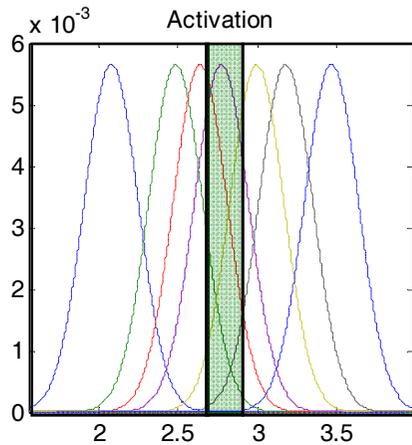


Le quatrième nombre est-il  
plus grand ou plus petit?



Trajectoire développementale du paramètre  $w$   
qui estime la précision  
de la discrimination des nombres





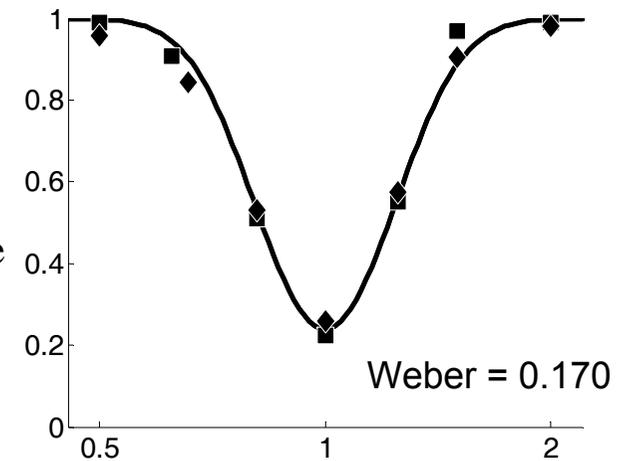
# Modélisation de la comparaison pareil / différent

Réponse R1= différent  
Réponse R2 = pareil  
que la référence

Prédiction:

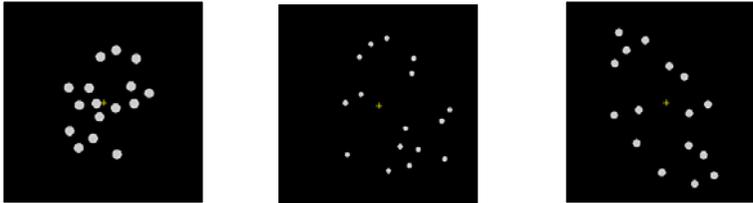
le taux d'erreur est une fonction symétrique  
de la distance entre les nombres,  
mesurée sur une échelle logarithmique:  
 $f(\text{Log}(n) - \text{Log}(\text{ref}))$

Où  $f$  est l'intégrale d'une Gaussienne  
*entre deux bornes*

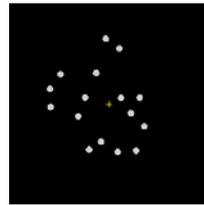


# L'expérience de Manuela Piazza et coll.

Trois exemples de la référence (16 or 32)

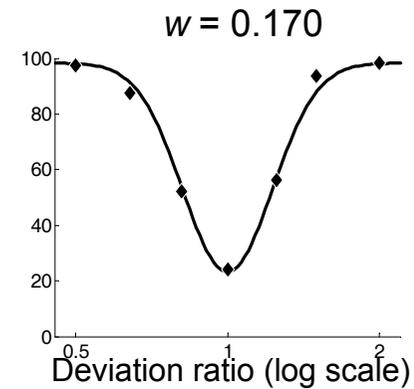
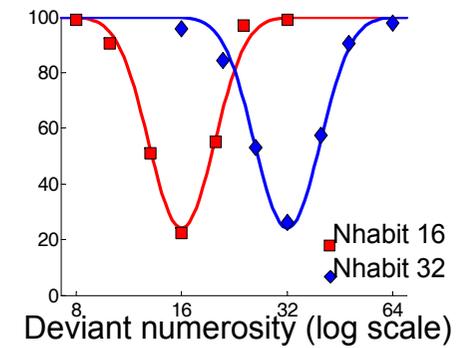
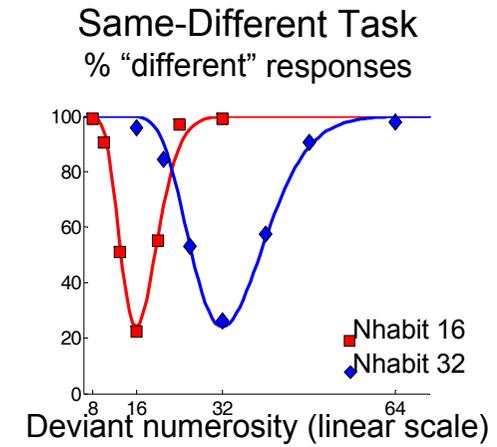


Suivi d'un nombre cible:



Nouveau jugement:

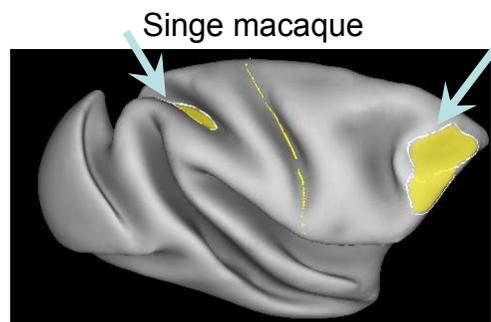
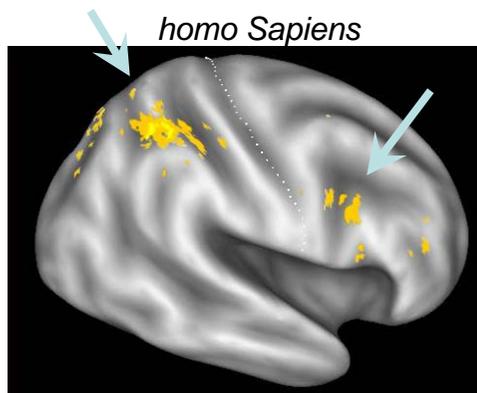
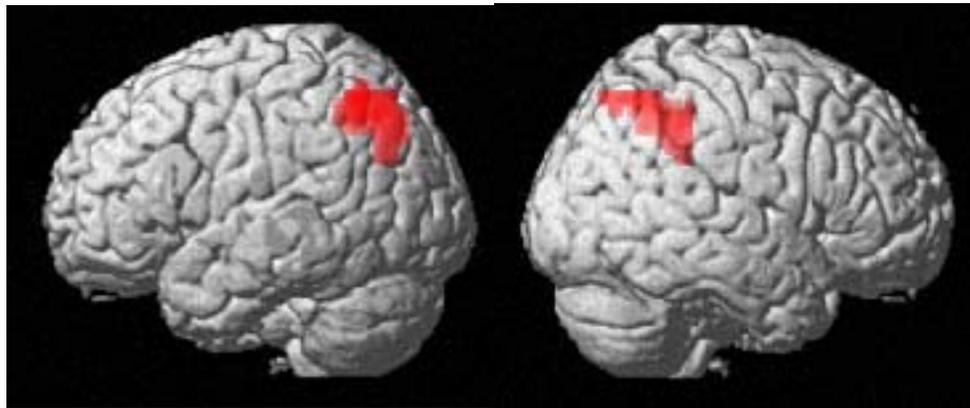
Pareil ou différent?



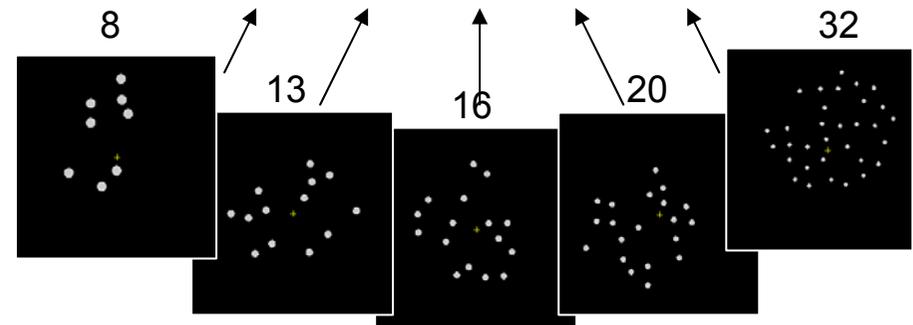
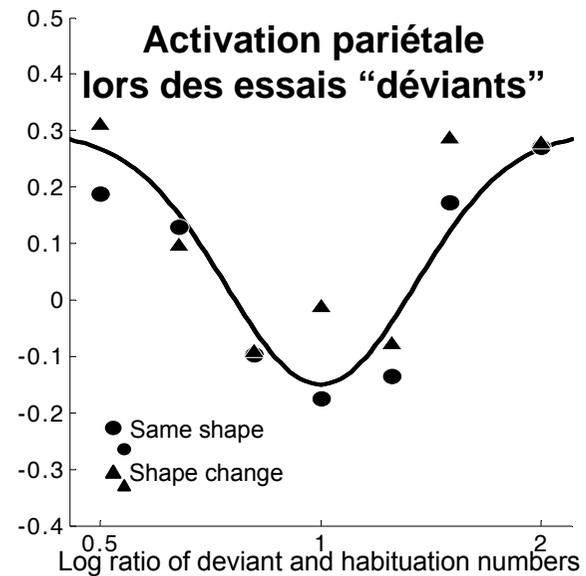
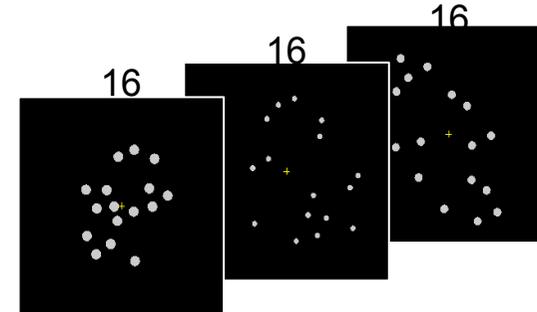
# Plausibilité d'un codage log-gaussien des nombres dans le sillon intrapariétal humain (hIPS)

Piazza, Izard, Pinel, Le Bihan & Dehaene, Neuron 2004

Régions qui répondent à un changement de nombre

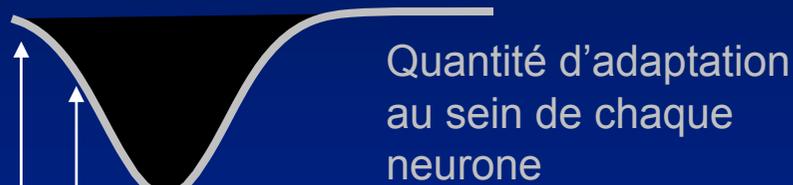
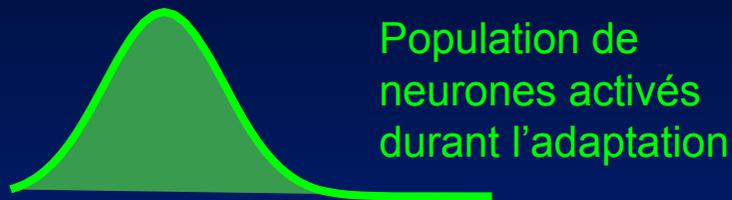
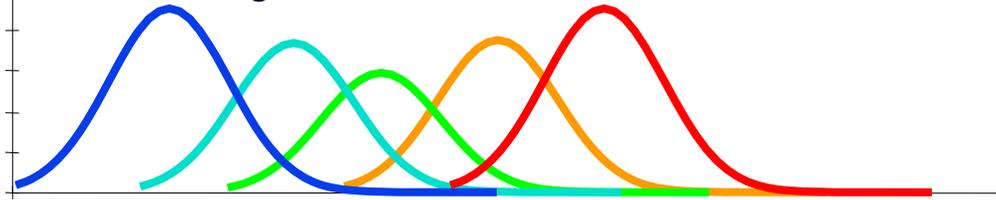


Adaptation à un nombre fixé dans un bloc donné

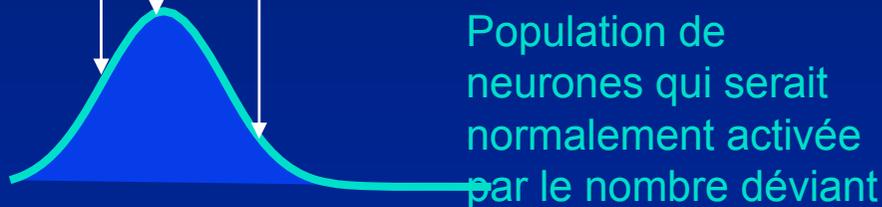


## Adaptation de l'activité neuronale

Taux de décharge des neurones



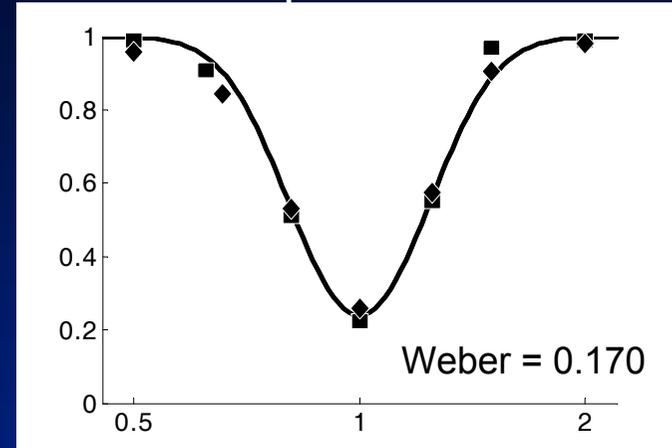
Multiplication point à point



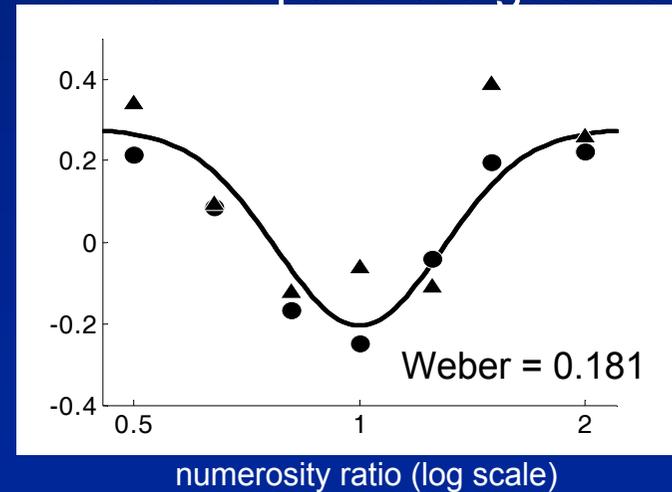
Le signal IRMf résulte de la somme de tous ces signaux neuronaux. Cela correspond à la convolution de deux Gaussiennes, dont une Gaussienne plus large d'un facteur 1.41

## Modélisation de l'adaptation en IRMf

Comportement



Activation pariétale gauche

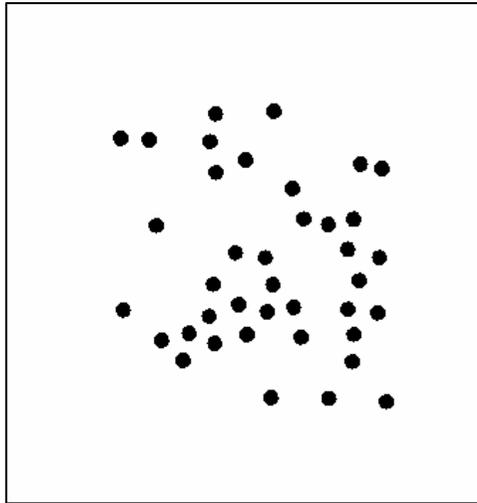


# Modélisation d'une tâche à plus de deux réponses: la dénomination des numérosités

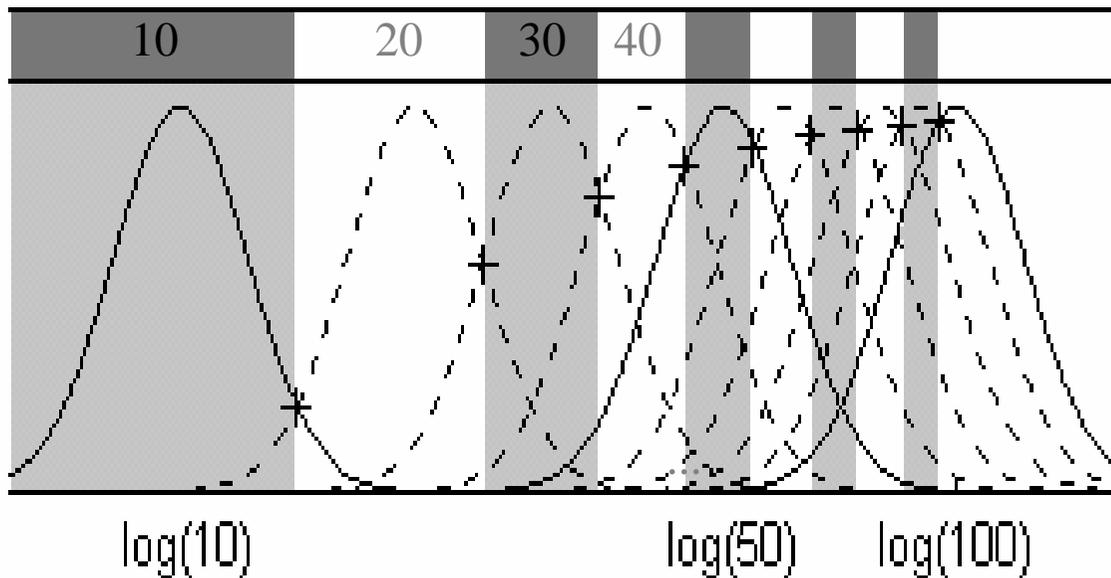
(Izard & Dehaene, *Cognition* 2007)

Combien de points?

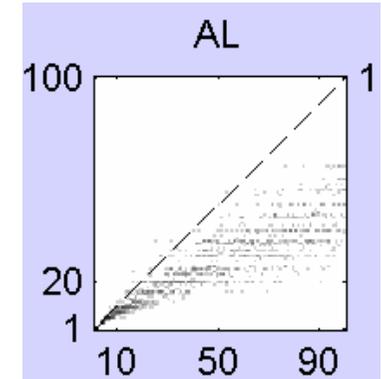
Y a-t-il plus ou moins que 40?



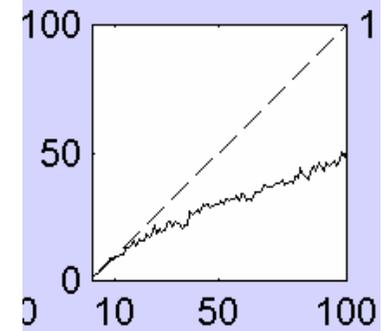
- La dénomination peut se modéliser comme l'application d'une grille de critères définissant un intervalle de quantités internes approprié à chaque mot.
- Curieusement, la grille de critères semble déformée de façon affine par rapport à la grille optimale.
- Cette déformation explique la **loi de puissance** qui relie les stimuli et les réponses (Stevens)



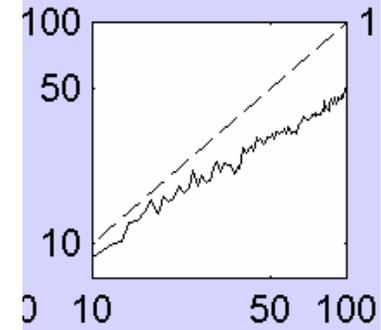
Distribution des réponses (échelle linéaire)



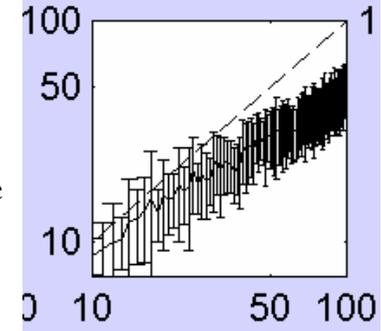
Réponse moyenne (échelle linéaire)



Réponse moyenne (double échelle logarithmique)

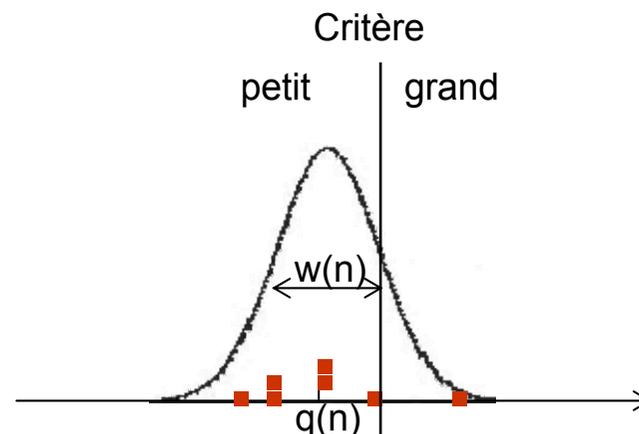


Ecart type des réponses (double échelle logarithmique)



### 3. Comprendre les temps de réponse

La perception d'un stimulus d'intensité  $n$  est codée par une variable aléatoire Gaussienne  $X$  sur un continuum interne.



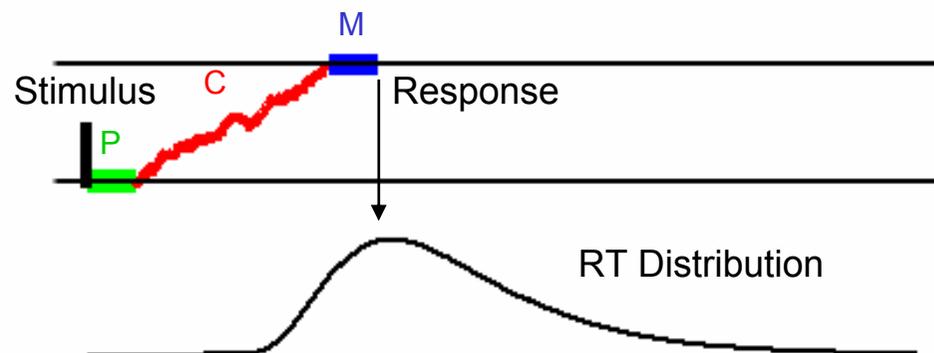
- Si l'on dispose de plusieurs échantillons successifs et indépendants,  $X_a, X_b, X_c \dots$  alors le théorème de Bayes fournit une façon simple d'accumuler ces données:

$P(R_i | X_a, X_b, X_c \dots)$  est proportionnel à  $P(R_i) * P(X_a | R_i) * P(X_b | R_i) * \dots$

- Dans le cas d'une réponse binaire, le log du rapport de vraisemblance est une fonction additive des échantillons successifs:

$LLR(X) = \text{Log} ( P(R1)/P(R2) ) + \text{Log} ( P(X_a | R1) / P(X_a | R2) ) + \text{Log} ( P(X_b | R1) / P(X_b | R2) ) + \dots$

- Il se produit donc une **marche aléatoire interne**:
  - Chaque échantillon fait évoluer, en positif ou en négatif, les « votes » en faveur de R1 ou de R2
  - Une décision optimale peut être prise lorsque le « vote » atteint un seuil

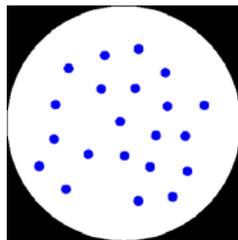


## Retour aux mécanismes neuronaux:

Selon Gold et Shadlen, on peut obtenir une quantité proportionnelle au LLR en calculant simplement la différence d'activité de deux populations de neurones.

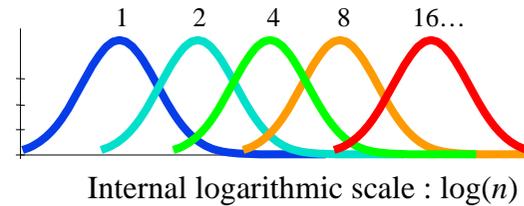
D'où un modèle complet de la chaîne de décision numérique:

**Stimulus de numérosité  $n$**

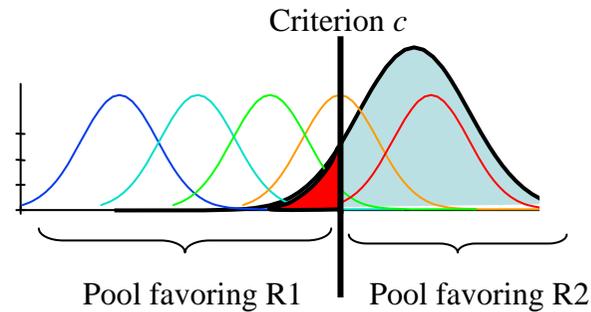


**Décision numérique simple**

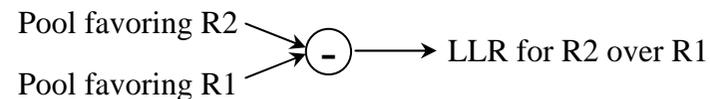
### 1. Coding by Log-Gaussian numerosity detectors



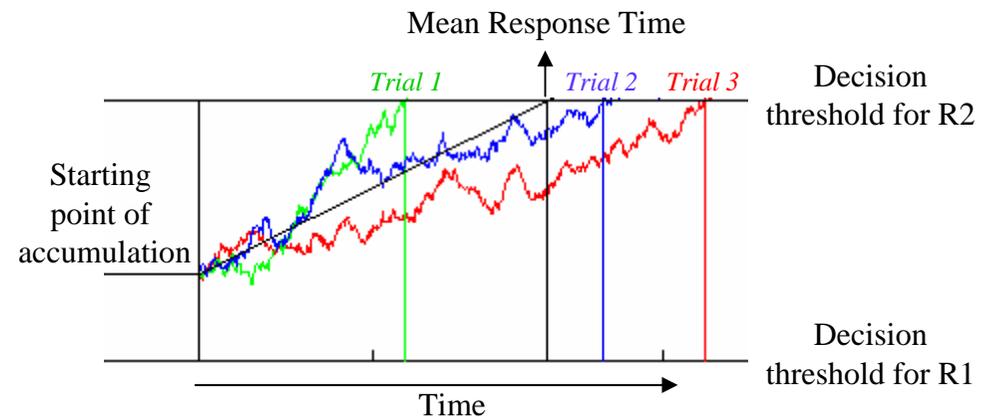
### 2. Application of a criterion and formation of two pools of units



### 3. Computation of log-likelihood ratio by differencing



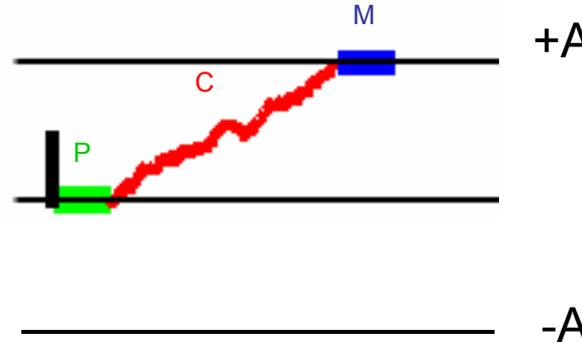
### 4. Accumulation of LLR, forming a random-walk process



# Prédictions du modèle. 1: la proportion d'erreurs

Décision non-biasée  
(seuils symétriques)

$$D_o = 0$$

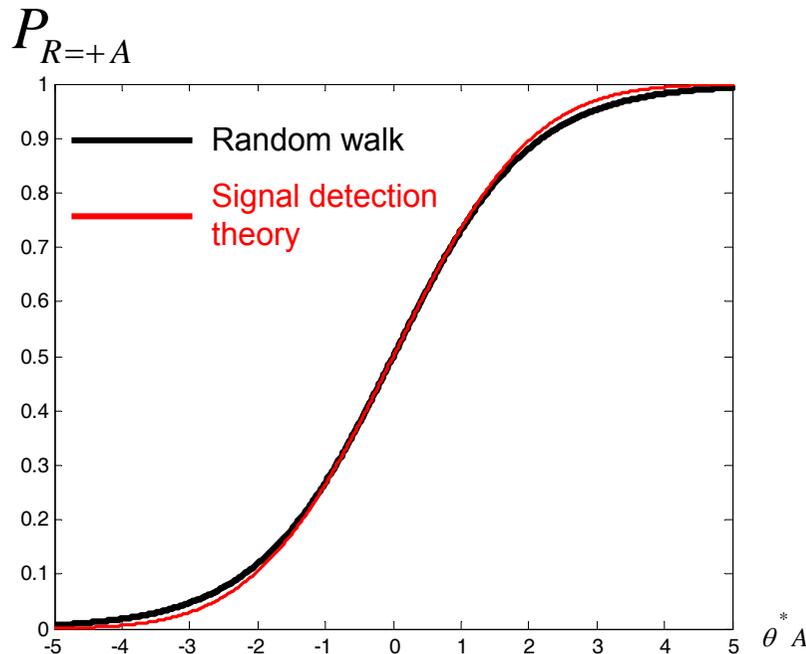


Pente  $\mu$   
Ecart-type  $\sigma$

$$P_{R=+A} = \frac{1}{1 + e^{-\theta^* A}}$$

où  $\theta^*$  est une fonction de  $\mu$  et  $\sigma$  qui dépend de la distribution du pas de la marche aléatoire

$$\theta^* = \frac{2\mu}{\sigma^2} \quad \text{pour une marche aléatoire Gaussienne}$$



## Conclusions:

- Les prédictions qualitatives sont virtuellement inchangées par rapport à la théorie de la détection du signal (théorie « à un échantillon »)
- Cependant, la pente de la fonction d'erreur dépend du seuil choisi. Le taux d'erreur n'est plus simplement fonction de la précision interne, mais également de la stratégie de réponse (et du temps de réponse)

## Prédictions du modèle. 2: Relation entre erreurs et temps de réponse

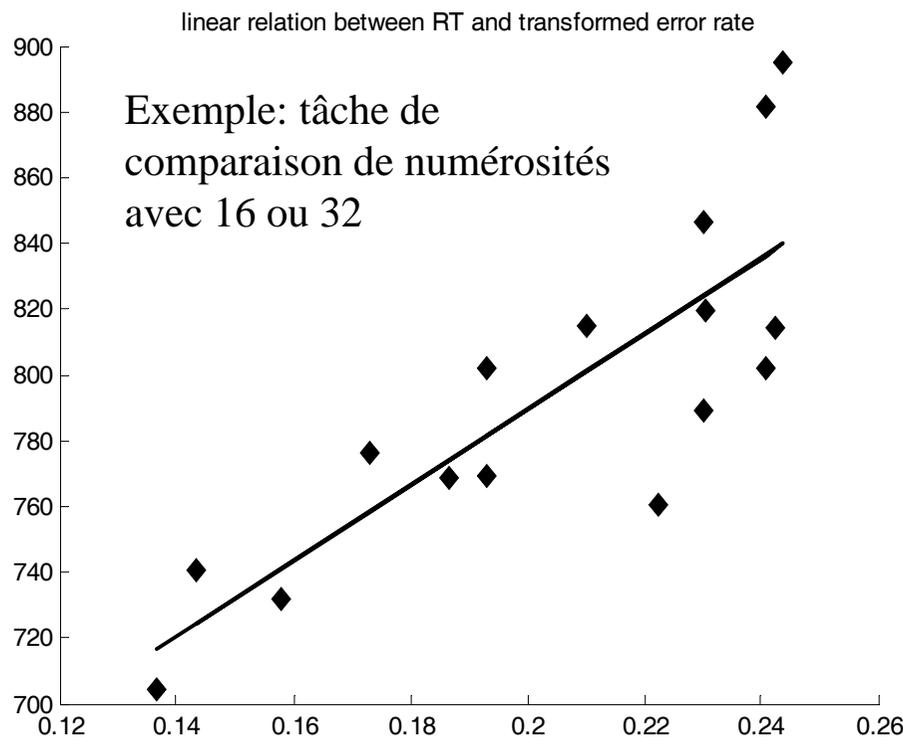
$$RT = T_0 + \frac{A}{\mu} (2P_{R=+A} - 1)$$

Temps de réponse moyenné à travers les réponses corrects et incorrectes

Noter que RT est proche de  $1 / \mu$  (alors que le taux d'erreurs décroissait exponentiellement avec la distance  $\mu$ )

Quand A et  $\sigma$  sont fixes, et que seul le stimulus varie (faisant ainsi varier  $\mu$ ), cette équation établit une relation systématique entre RT et taux d'erreurs.

Pour une marche aléatoire Gaussienne:



$$RT = T_0 + \frac{4A^2}{\sigma^2} \left( \frac{P_{R=+A} - 0.5}{\text{Log} \frac{P_{R=+A}}{1 - P_{R=+A}}} \right)$$

$$T^0 = 560 \text{ ms}$$

$$A = 16.9$$

(en supposant  $\sigma=1$ )

Note: these equations implicitly incorporate  $\Delta t=1$  time unit (1 ms)  
Thus, the random walk advances by time steps of 1 ms,  
with a large noise ( $\sigma=1$ )  
and a small mean step size (maximum in the order of 0.1)

# Estimation du pas moyen de la marche aléatoire

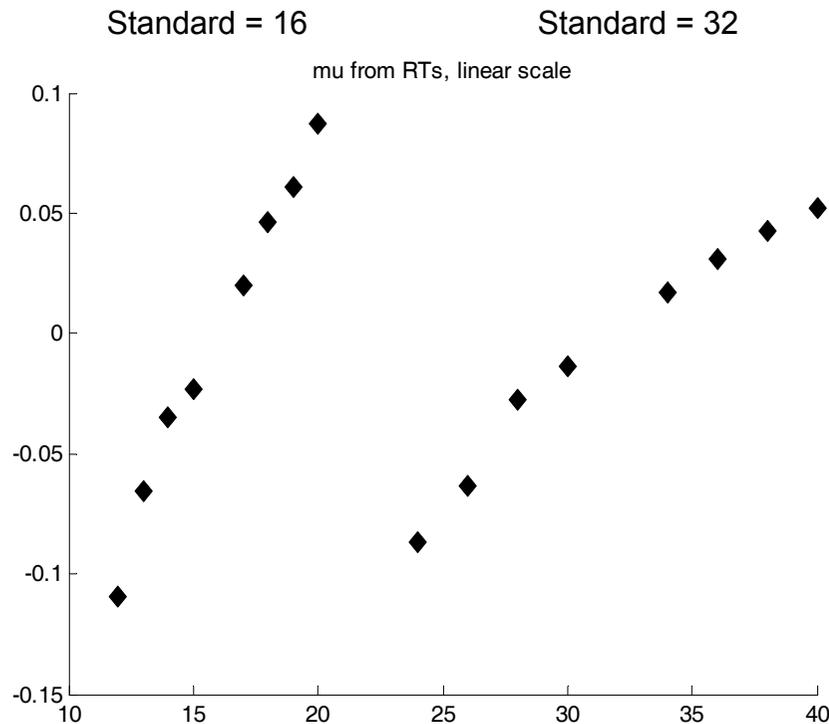
$$RT = T_0 + \frac{4A^2}{\sigma^2} \left( \frac{P_{R=+A} - 0.5}{\text{Log} \frac{P_{R=+A}}{1 - P_{R=+A}}} \right)$$

Par régression linéaire, cette relation permet d'estimer  $T_0$  et  $A$  (en supposant  $\sigma=1$ ).

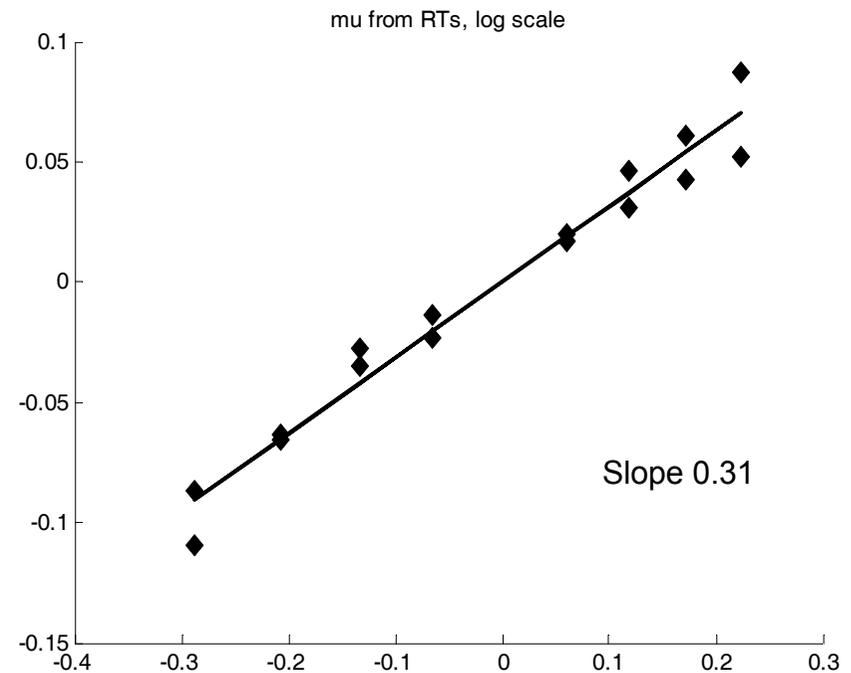
On peut alors calculer  $\mu = \frac{A(2P_{R=+A} - 1)}{RT - T_0}$

séparément pour chaque stimulus présenté

Exemple: tâche de comparaison de numérosités avec 16 ou 32



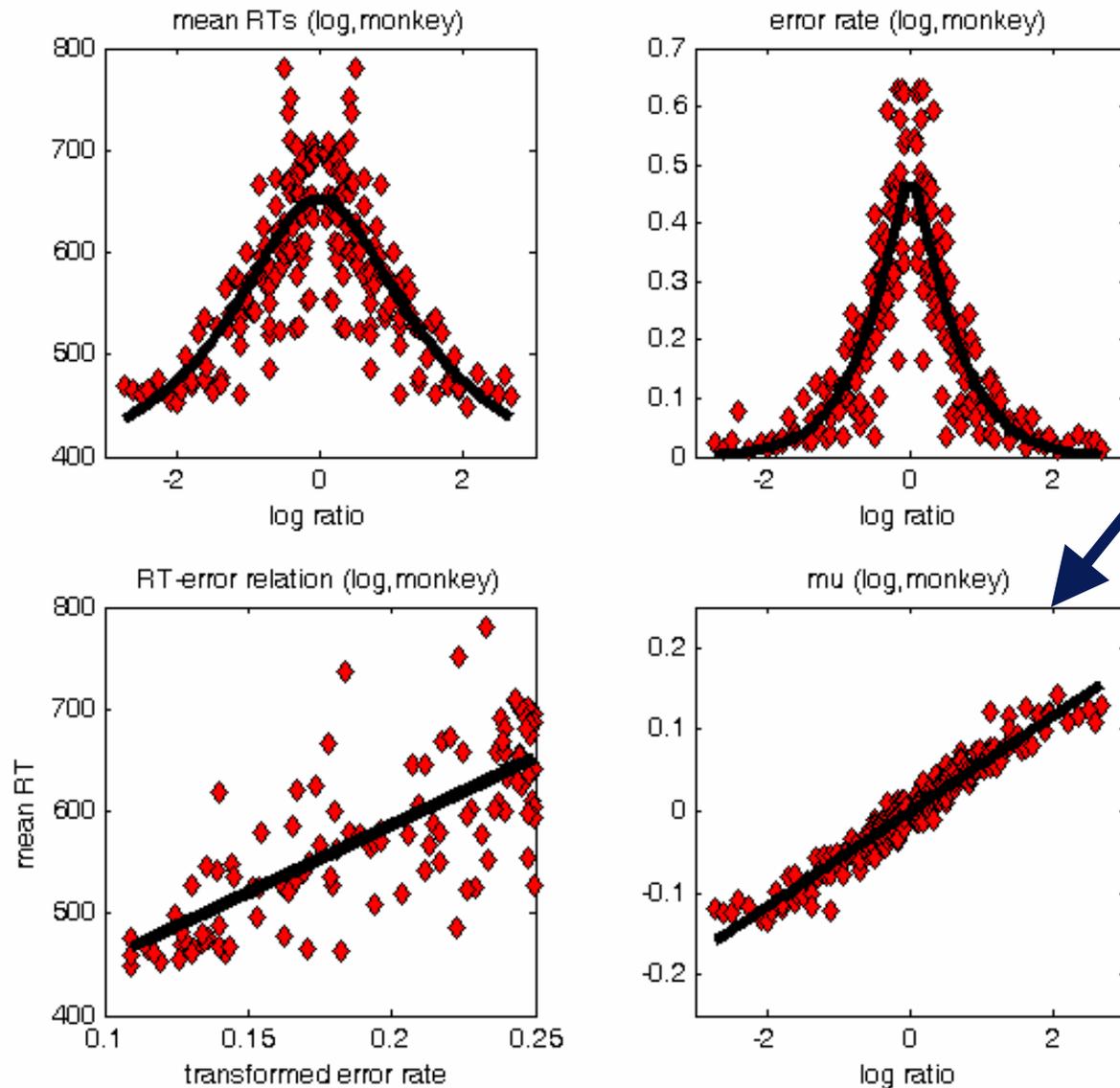
Stimulus sur une échelle linéaire



Distance sur une échelle logarithmique

# Exemple: comparaison de deux numérosités chez le singe et l'homme

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychol Sci*, 17(5), 401-406.

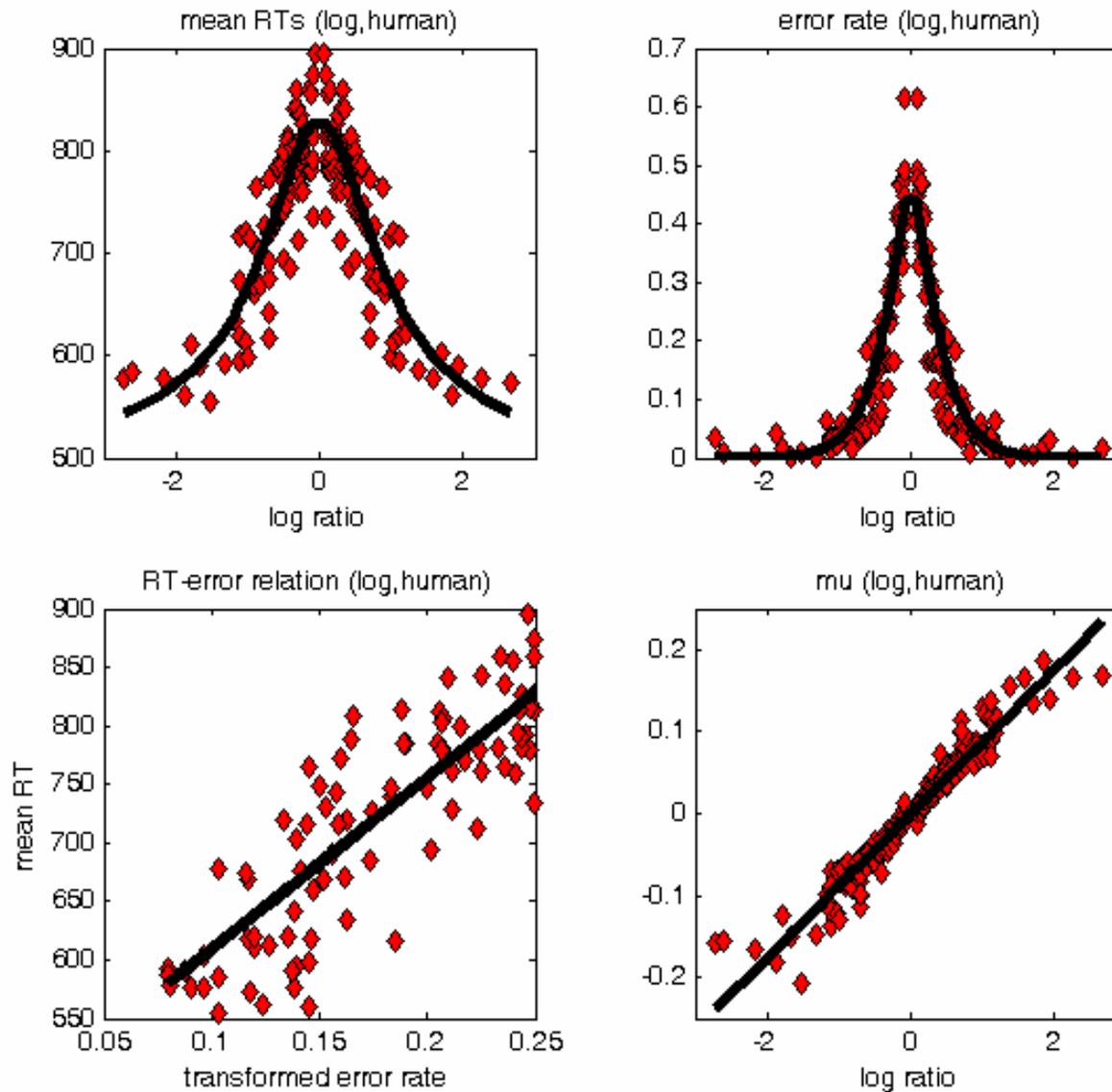


La théorie de la marche aléatoire permet de recouvrer cette variable cruciale, le pas moyen de la marche aléatoire ou **quantité moyenne d'information accumulée par unité de temps**

On voit que, conformément à la théorie log-gaussienne, cette quantité est proportionnelle au log du rapport des nombres à comparer (autrement dit, à la distance sur une échelle logarithmique)

## Exemple: comparaison de deux numéros chez le singe et l'homme

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychol Sci*, 17(5), 401-406.



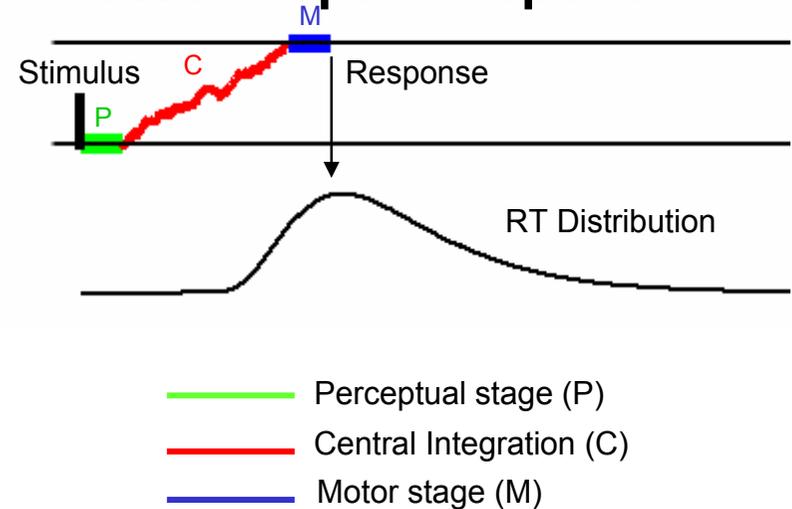
- Les données chez les sujets humains sont remarquablement proches – les décisions sont plus lentes mais un peu plus précises.

- Les temps de réponse et les erreurs ont exactement la forme prédite par la théorie.

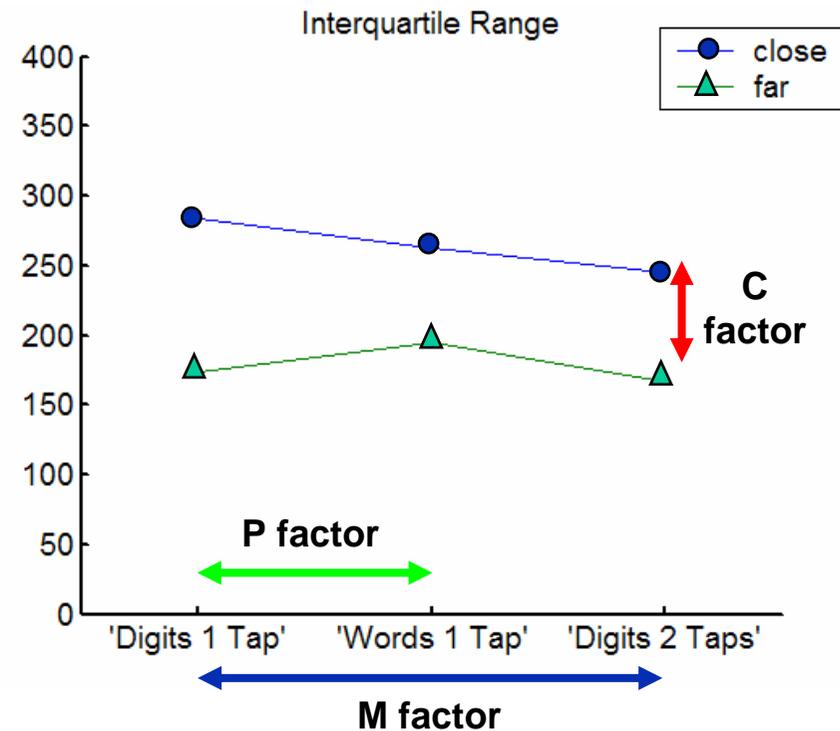
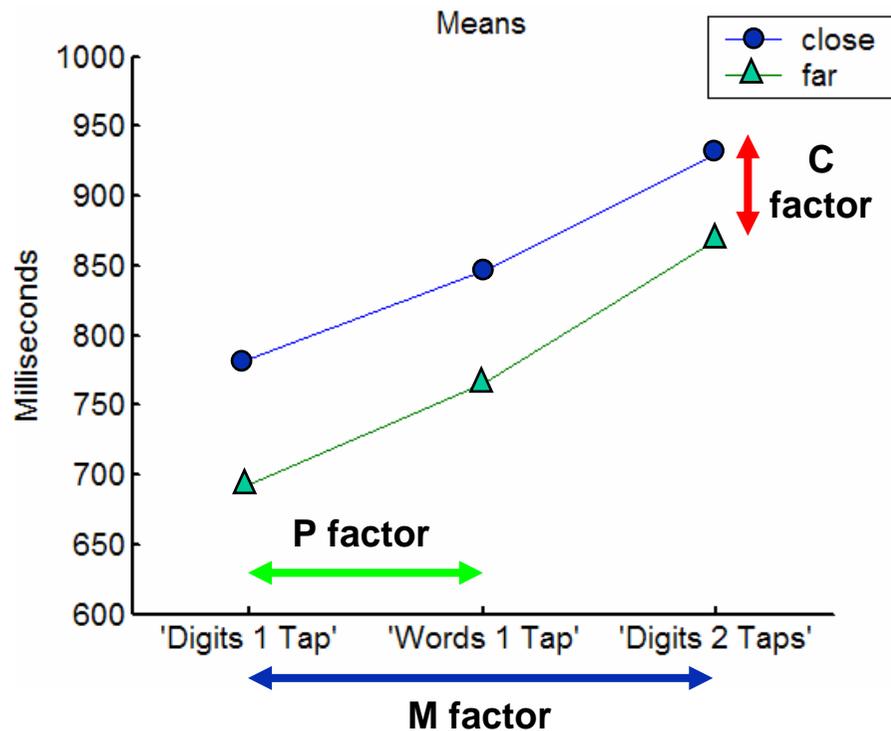
### Prédictions. 3: Sources de la variabilité dans les temps de réponse

L'étape de décision est responsable de la grande variabilité des temps de réponse.

- Les facteurs qui affectent la pente de l'étape de décision centrale C (par exemple la distance) doivent simultanément affecter la moyenne et l'écart type des temps de réponse.
- Les autres facteurs qui affectent P ou M (par exemple la notation des nombres, la complexité de la réponse) doivent affecter la moyenne mais pas nécessairement l'écart type.



Exemple: comparaison de nombres en notation Arabe ou Verbale (Sigman & Dehaene, PLOS: Biology, 2004)

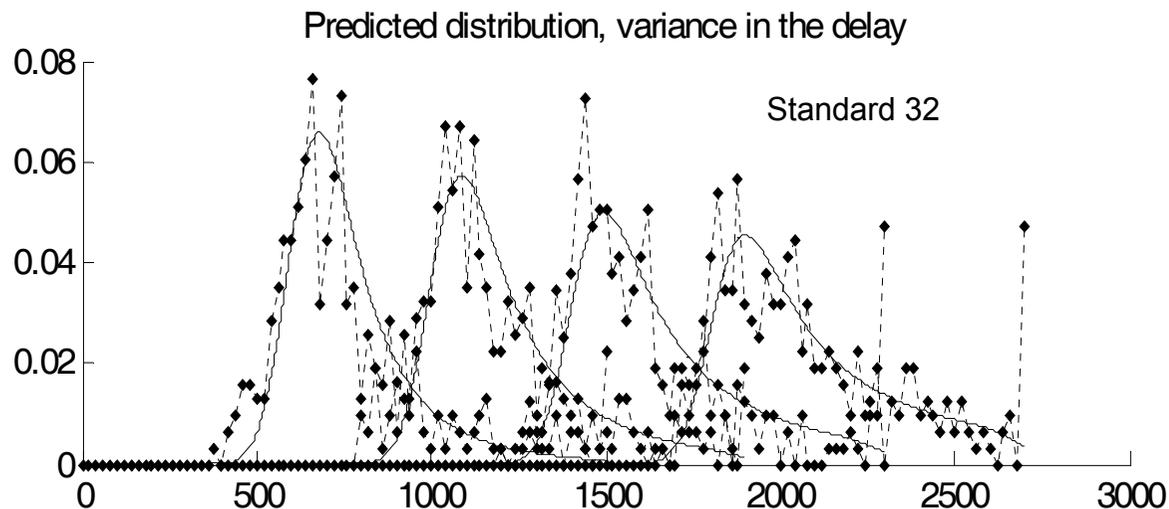
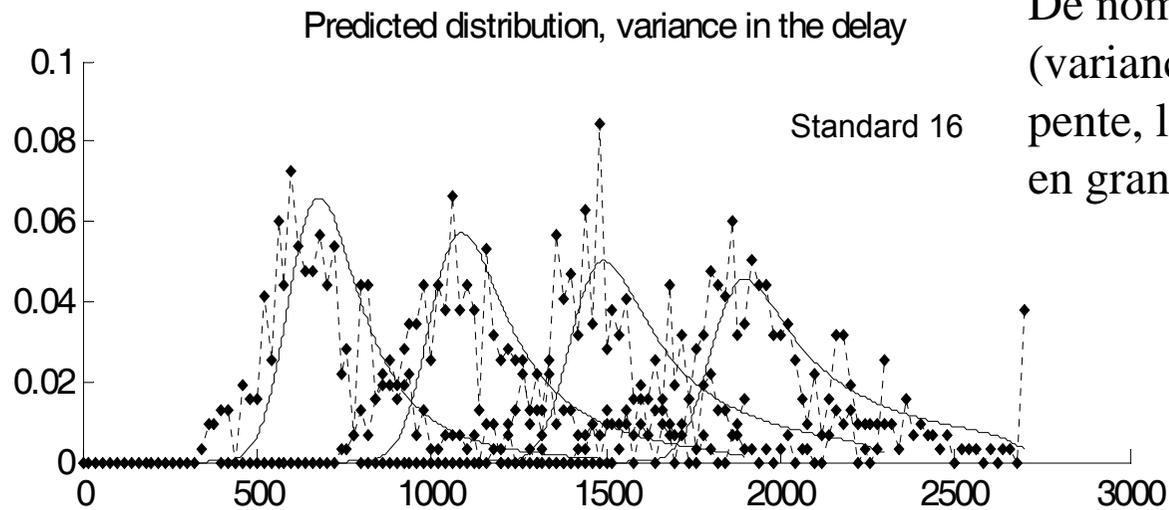


## Prédictions. 4: Forme de la distribution des temps de réponse

$$g(t) = \frac{1}{\sigma \cdot \sqrt{2\pi \cdot (t - t_0)^3}} \cdot \exp\left(-\frac{(1 - \alpha \cdot (t - t_0))^2}{2 \cdot \sigma^2(t - t_0)}\right)$$

Approximation la plus simple (en considérant une marche aléatoire avec une seule borne fixe)

De nombreuses améliorations existent (variance dans le point de départ, la pente, les bornes...) et sont discutées en grand détail par Roger Ratcliff



Distributions des temps de réponse pour la comparaison de deux numérosités.

Données catégorisées en 4 niveaux de distance (décalés arbitrairement sur l'axe des x pour plus de lisibilité)

# Conclusions

- Si l'on adjoint au code neural log-gaussien des mécanismes optimaux de prise de décision, on parvient à un modèle mathématique relativement simple de la décision numérique dans des tâches simples: comparaison, dénomination... mais aussi addition, soustraction
- Ces mécanismes de prise de décision ne sont pas propres au domaine numérique
- On voit se dessiner une théorie générale de la décision, dont les bases neurales se clarifient également (travaux de M. Shadlen; X.J. Wang)

# Quelques références pour approfondir le sujet

Dehaene, S. (2007). Symbols and quantities in parietal cortex : Elements of a mathematical theory of number representation and manipulation. *Attention & Performance XXII*, pp 527-574. ([www.unicog.org](http://www.unicog.org))

- **Psychologie mathématique de la décision et du temps de réponse**

- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84, 279-325.
- Green, D., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Krieger Publishing Company.
- Laming, D. R. J. (1968). *Information theory of choice-reaction times*. London: Academic Press.
- Link, S. W. (1992). *The wave theory of difference and similarity*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Page, R., Izquierdo, E., Saal, A., Codnia, J., & El Hasi, C. (2004). A response time model for judging order relationship between two symbolic stimuli. *Percept Psychophys*, 66(2), 196-207.
- Ratcliff, R., & Rouder, J. (1998). Modelling response times for two-choice decisions. *Psychological Science*, 9, 347-356.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. *Psychon Bull Rev*, 9(3), 438-481.

- **Modèles neuronaux**

- Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, Mass.: MIT Press.
- Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, 36(2), 299-308.
- Lo, C. C., & Wang, X. J. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat Neurosci*, 9(7), 956-963.
- Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nat Neurosci*, 9(11), 1432-1438.
- Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: Probing the physiology of perception. *Annu Rev Neurosci*, 21, 227-277.
- Salinas, E. (2006). How behavioral constraints may determine optimal sensory representations. *PLoS Biol*, 4(12), e387.