

Cours 2012:

**Le cerveau statisticien:
La révolution Bayésienne en sciences cognitives**

Stanislas Dehaene
Chaire de Psychologie Cognitive Expérimentale

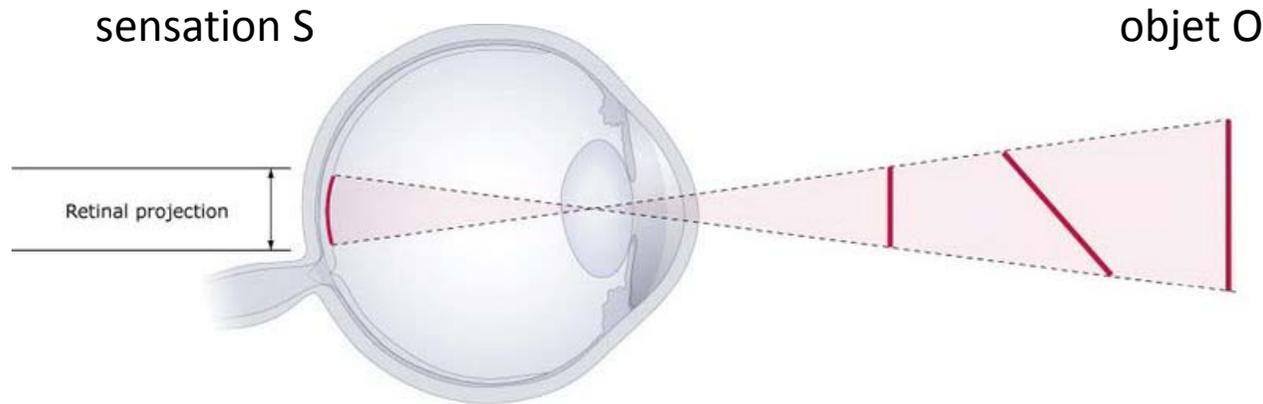
Cours n°4

**Combinaison de contraintes
et sélection d'un percept unique**

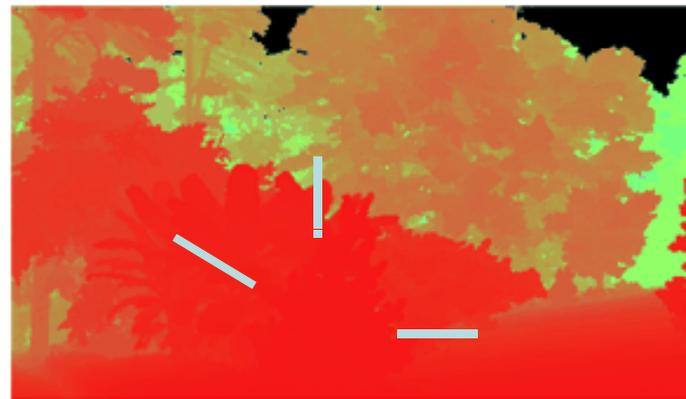
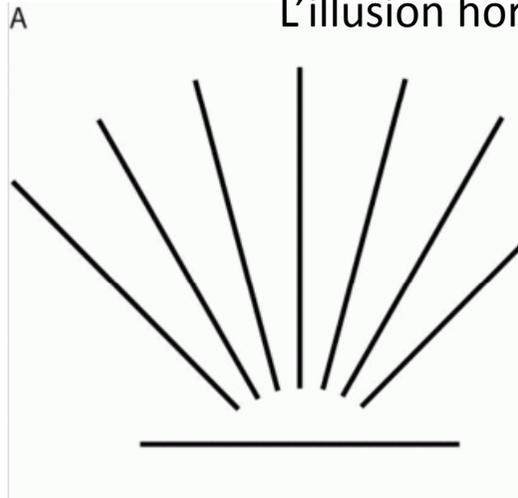
Pourquoi le cadre Bayésien s'applique-t-il bien à la perception?

Les entrées sensorielles sont *toujours* ambiguës.

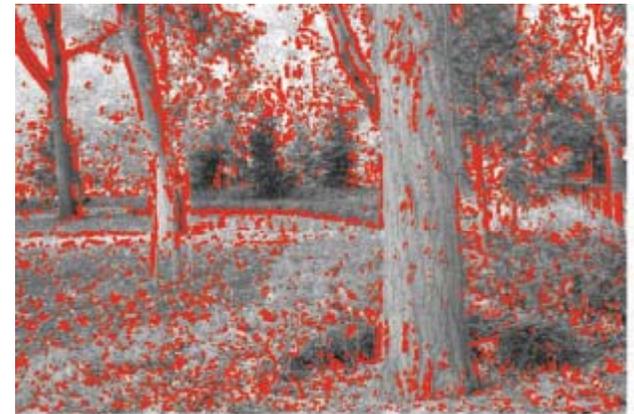
Notre système perceptif doit donc *sélectionner*, parmi une infinité de solutions possibles, celle qui est la plus *plausible*.



Perception de la taille:
L'illusion horizontal/vertical

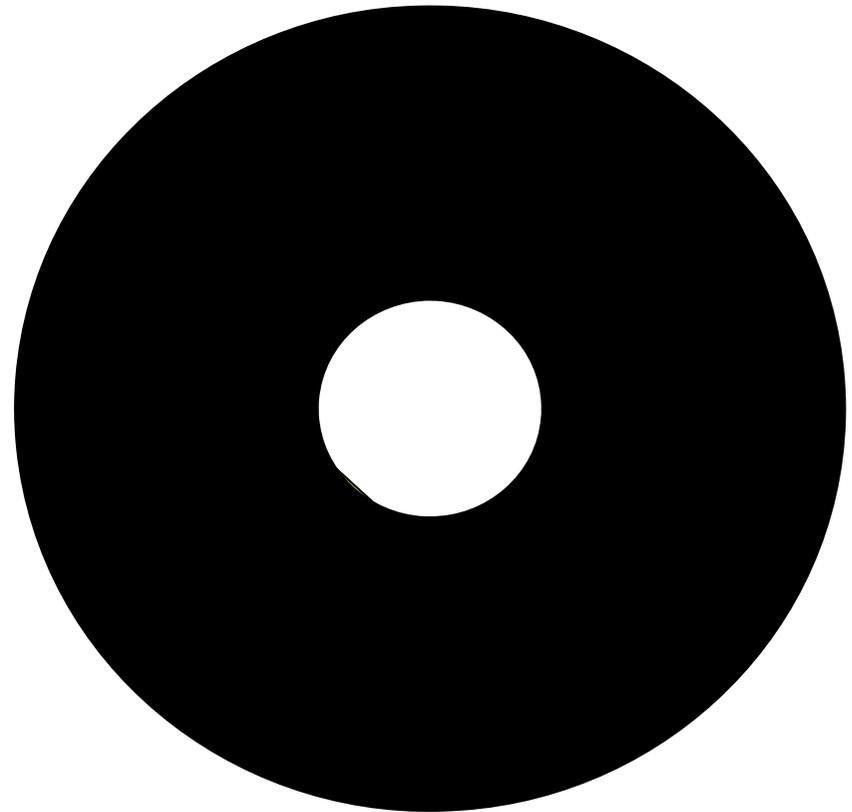
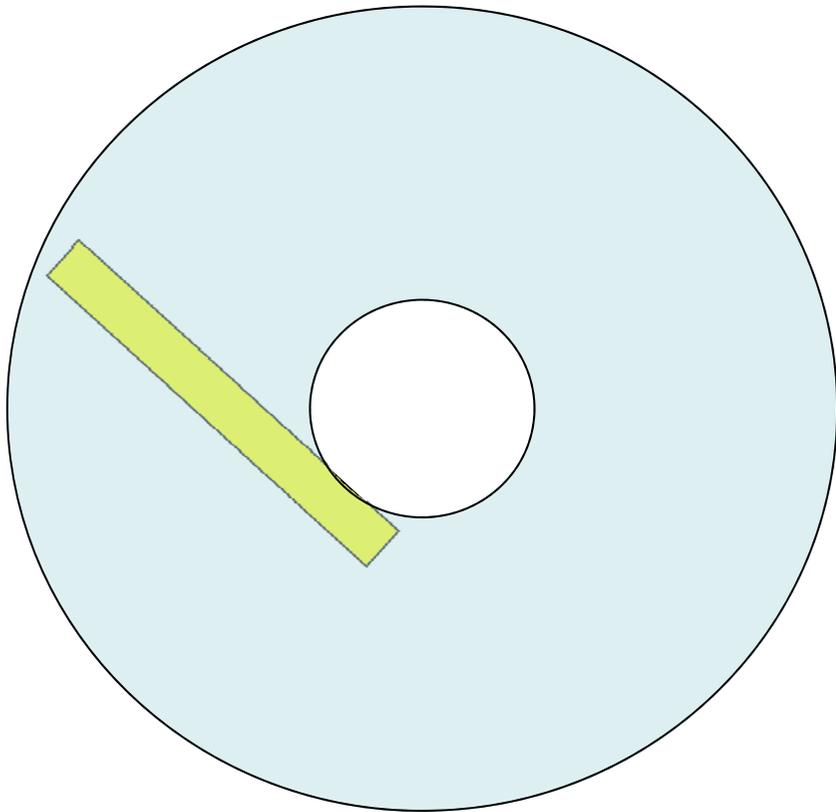


Perception de l'orientation: *a priori*
pour l'horizontale et la verticale



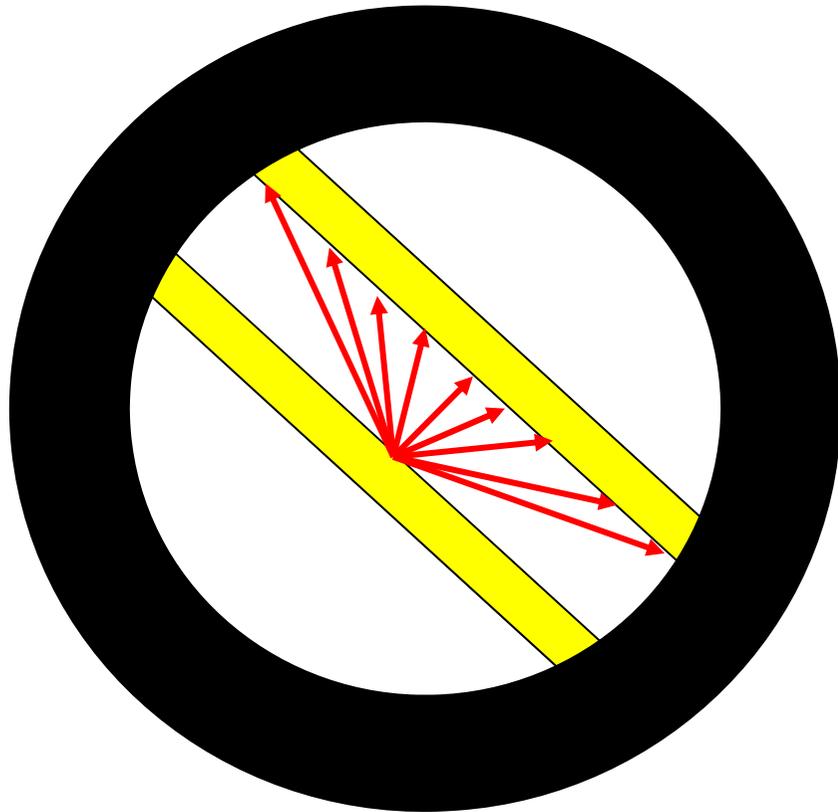
Images tirées de l'article de ScholarPedia « Visual_illusions:_An_Empirical_Explanation » par Dale Purves et coll.

Exemple 3: la perception du mouvement
et le problème de l'ouverture

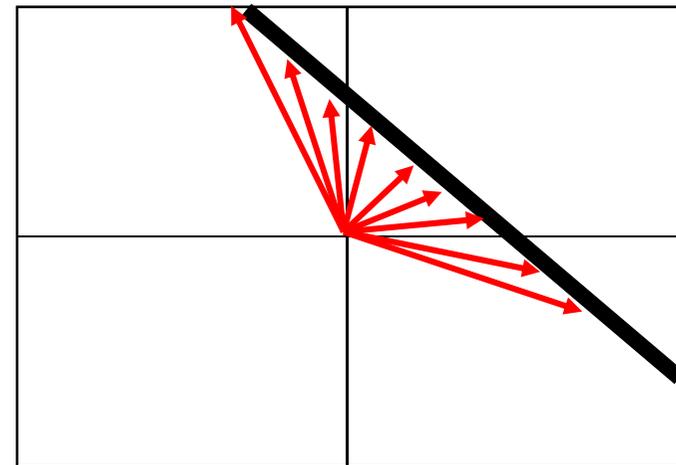


Le problème de l'ouverture

Chaque mouvement local, vu à travers un champ récepteur étroit, est compatible avec une infinité d'interprétations possibles



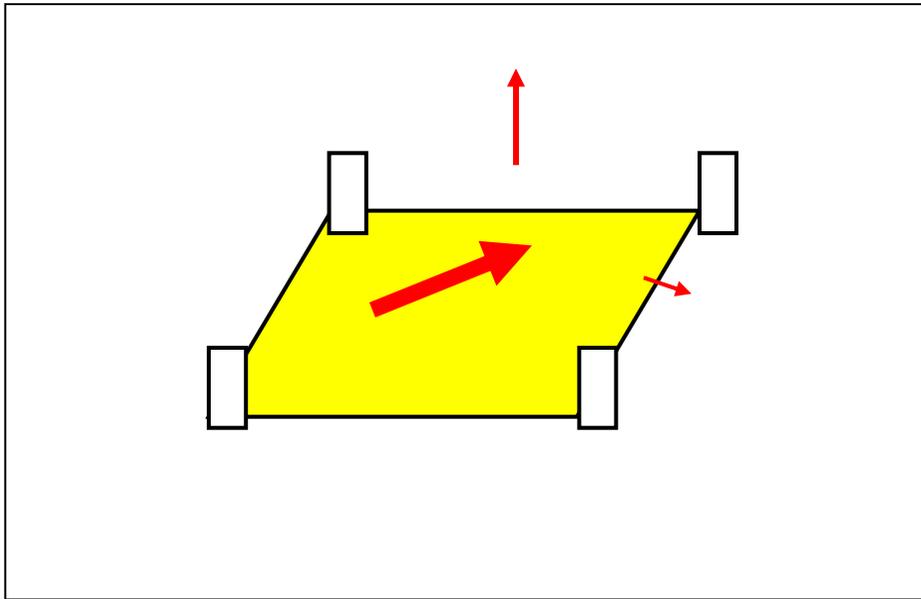
vitesse en y



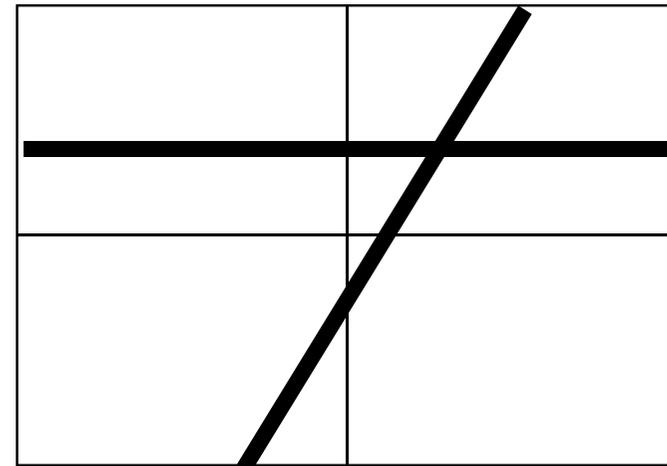
vitesse en x

Le problème de l'ouverture

Lorsqu'une surface est en mouvement, la combinaison des vitesses locales permet de calculer le mouvement global.



vitesse en y

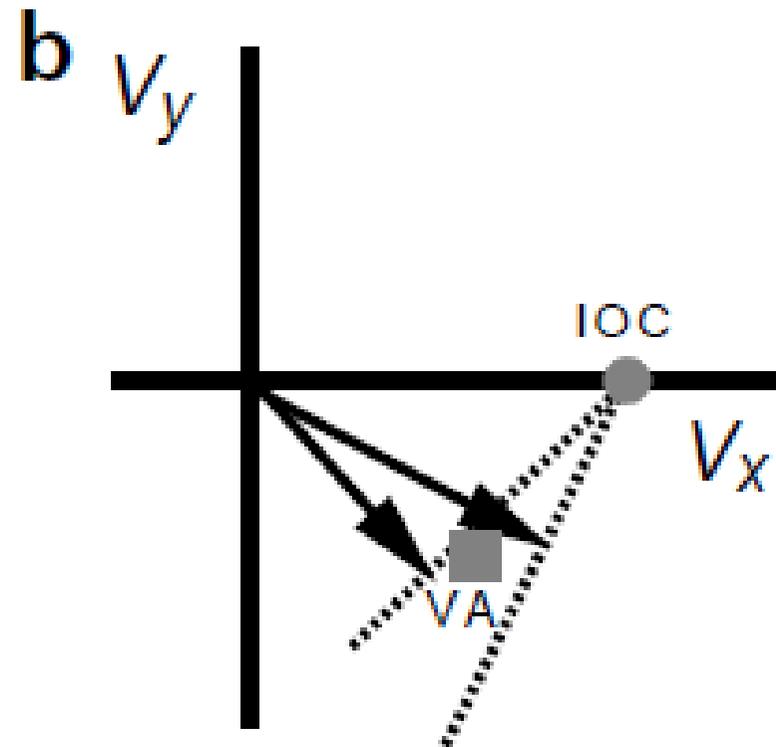
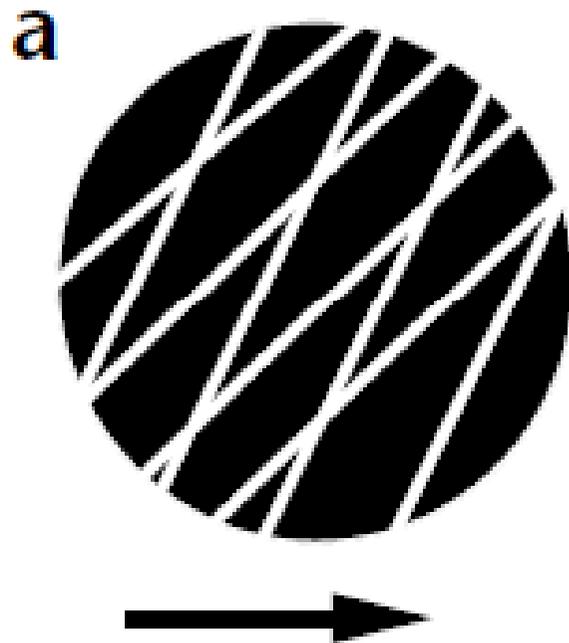


vitesse en x

Motion illusions as optimal percepts

Yair Weiss¹, Eero P. Simoncelli² and Edward H. Adelson³

Plusieurs mécanismes ont été proposés:
intersection de contraintes (*intersection of constraints*)
moyenne de vecteurs (*vector average, VA*)
suivi de traits élémentaires (*feature tracking, FT*).



Les mécanismes VA, IOC et FT ne suffisent pas à expliquer la perception du mouvement

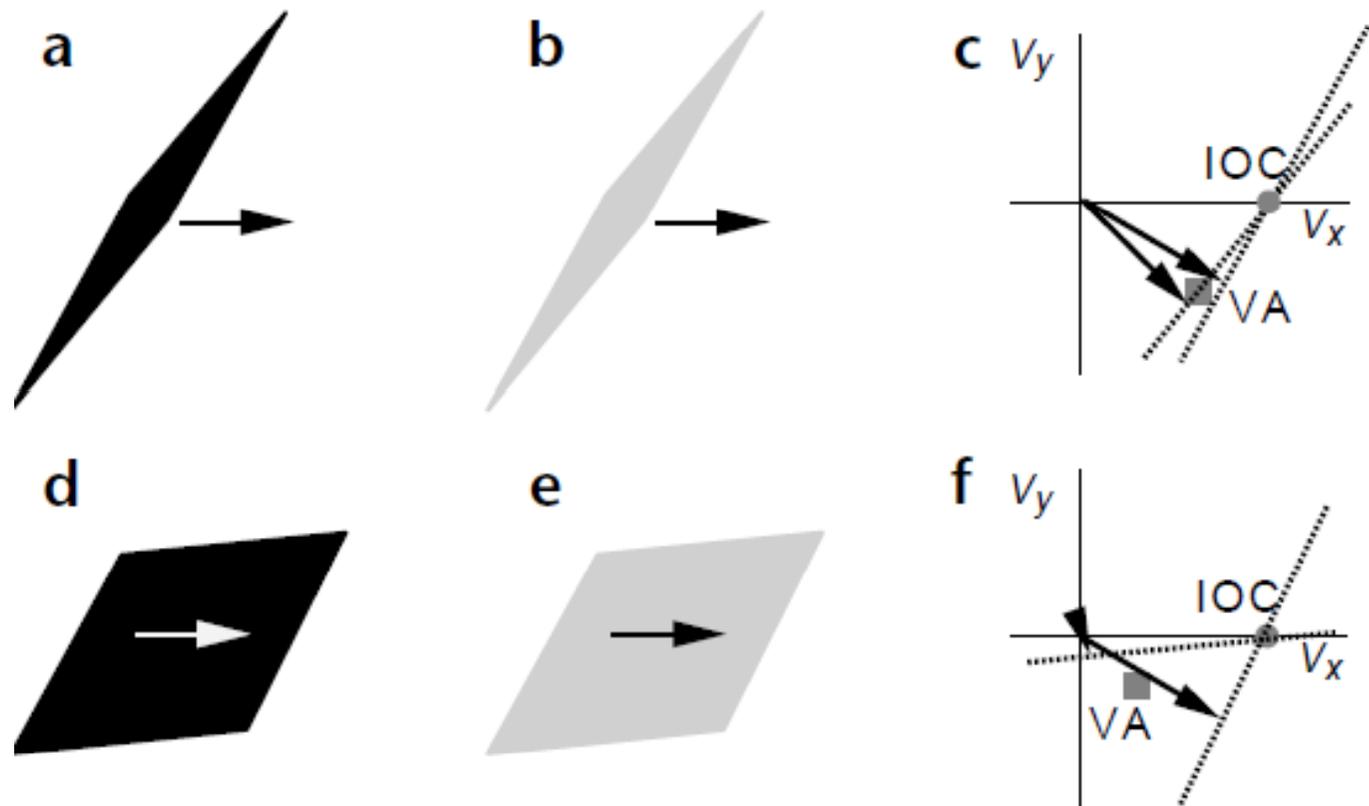
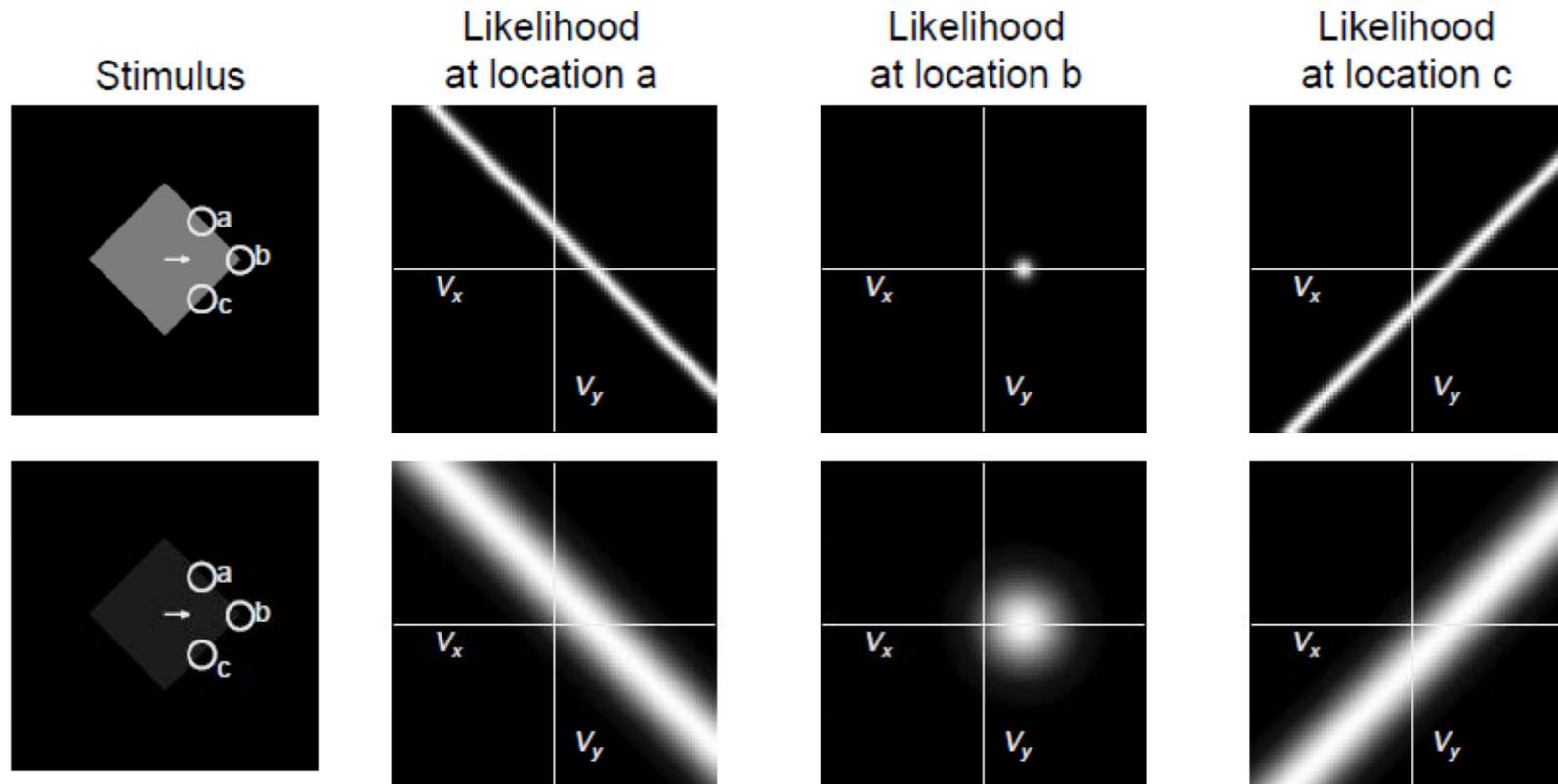


Fig. 2. Insufficiency of either VA, IOC or FT rules as an explanation for human perception of a horizontally moving rhombus. (a) A 'narrow' rhombus at high contrast appears to move horizontally (consistent with IOC/FT). (b) A narrow rhombus at low contrast appears to move diagonally (consistent with VA). (c) Velocity space constraints for a narrow rhombus. (d,e) A 'fat' rhombus at low or high contrast appears to move horizontally (consistent with IOC/FT). (f) Velocity space constraints for a fat rhombus.

Motion illusions as optimal percepts

Yair Weiss¹, Eero P. Simoncelli² and Edward H. Adelson³

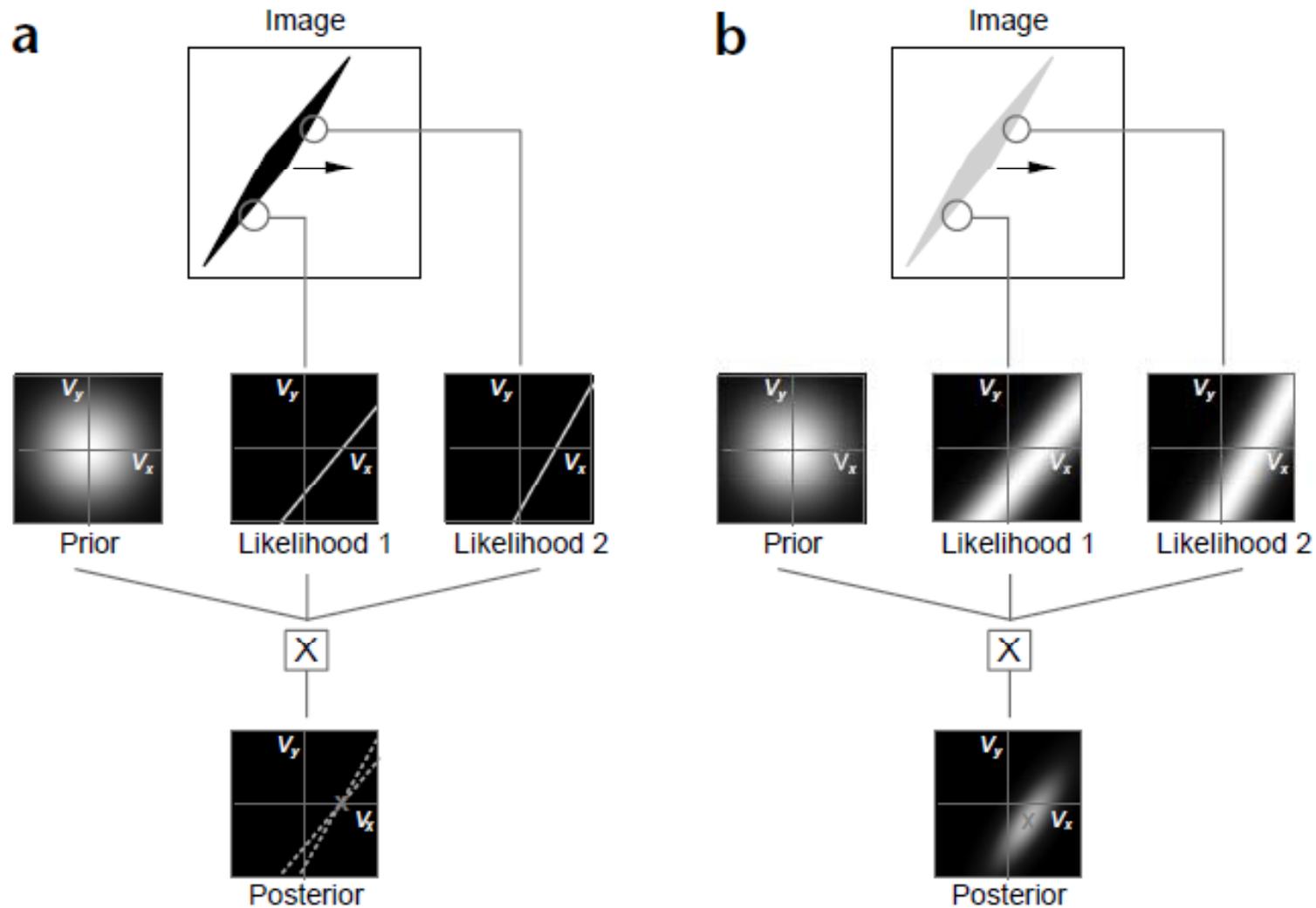
La perception du mouvement peut être vue comme un problème Bayésien:
Quel est le mouvement le plus probable, étant donné les indices disponibles?
Le modèle postule des fonctions de vraisemblance qui dépendent du contraste:
Plus un *a priori* centré sur zéro (le plus probable est l'absence de mouvement).



Motion illusions as optimal percepts

Yair Weiss¹, Eero P. Simoncelli² and Edward H. Adelson³

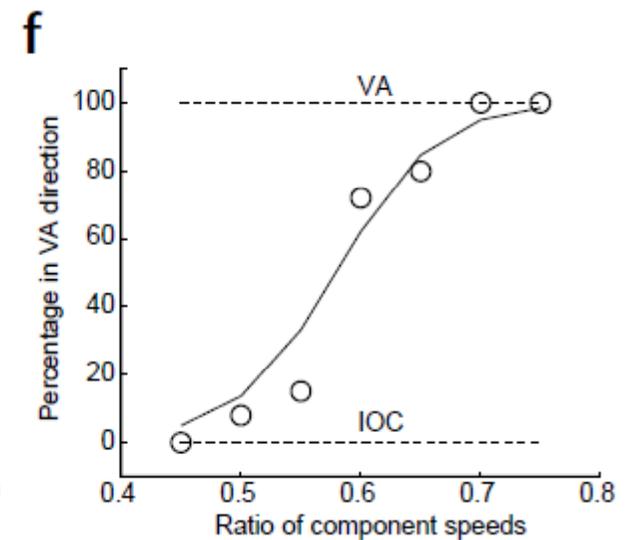
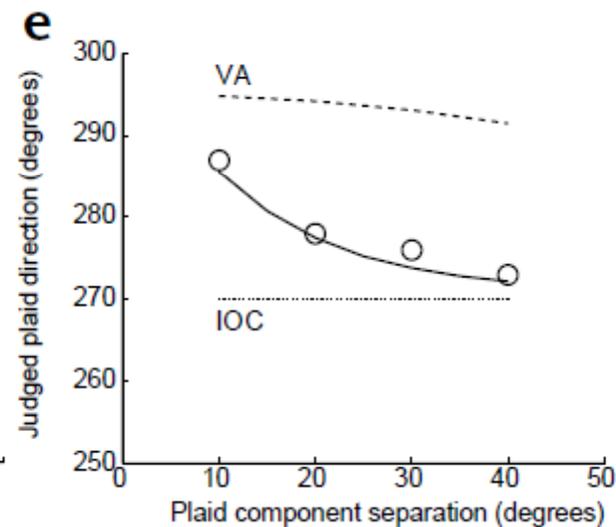
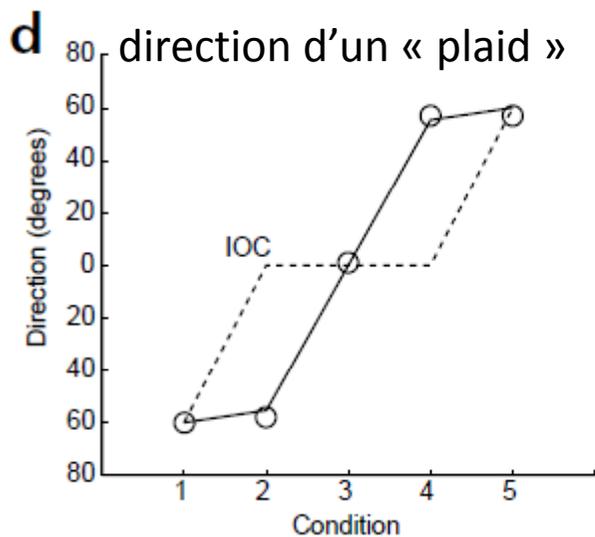
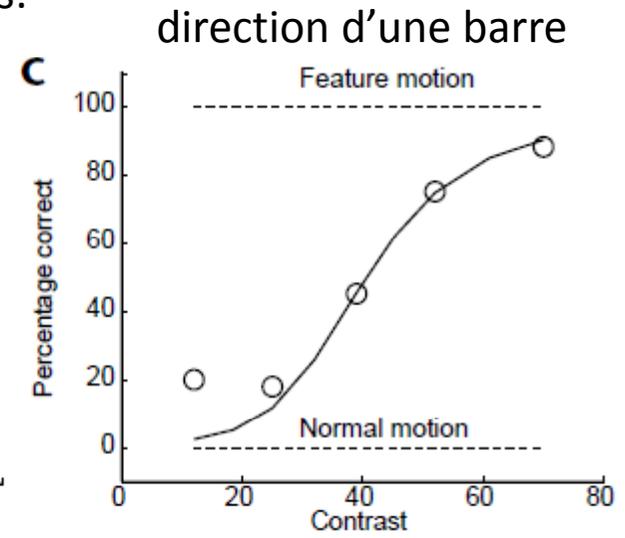
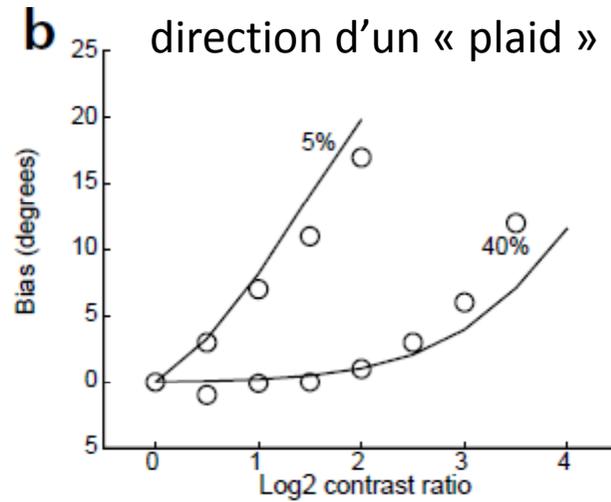
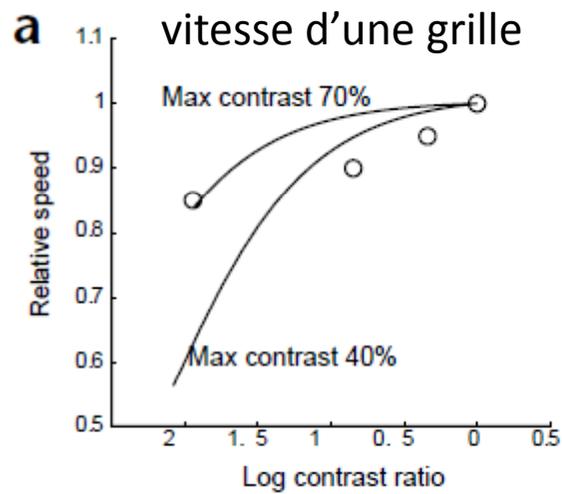
Le modèle peut expliquer l'illusion du losange:



Motion illusions as optimal percepts

Yair Weiss¹, Eero P. Simoncelli² and Edward H. Adelson³

Le modèle rend compte d'une très grande variété d'illusions:



Motion illusions as optimal percepts

Yair Weiss¹, Eero P. Simoncelli² and Edward H. Adelson³

Le modèle peut expliquer d'autres illusions:

- pourquoi percevons-nous que notre mouvement se ralentit lorsque nous roulons dans le brouillard?
- pourquoi percevons-nous parfois deux grilles superposées, en mouvement indépendant, plutôt qu'un seul « plaid »?

Pour une revue accessible des nombreuses régularités qui sont internalisées par notre système de perception visuelle, voir

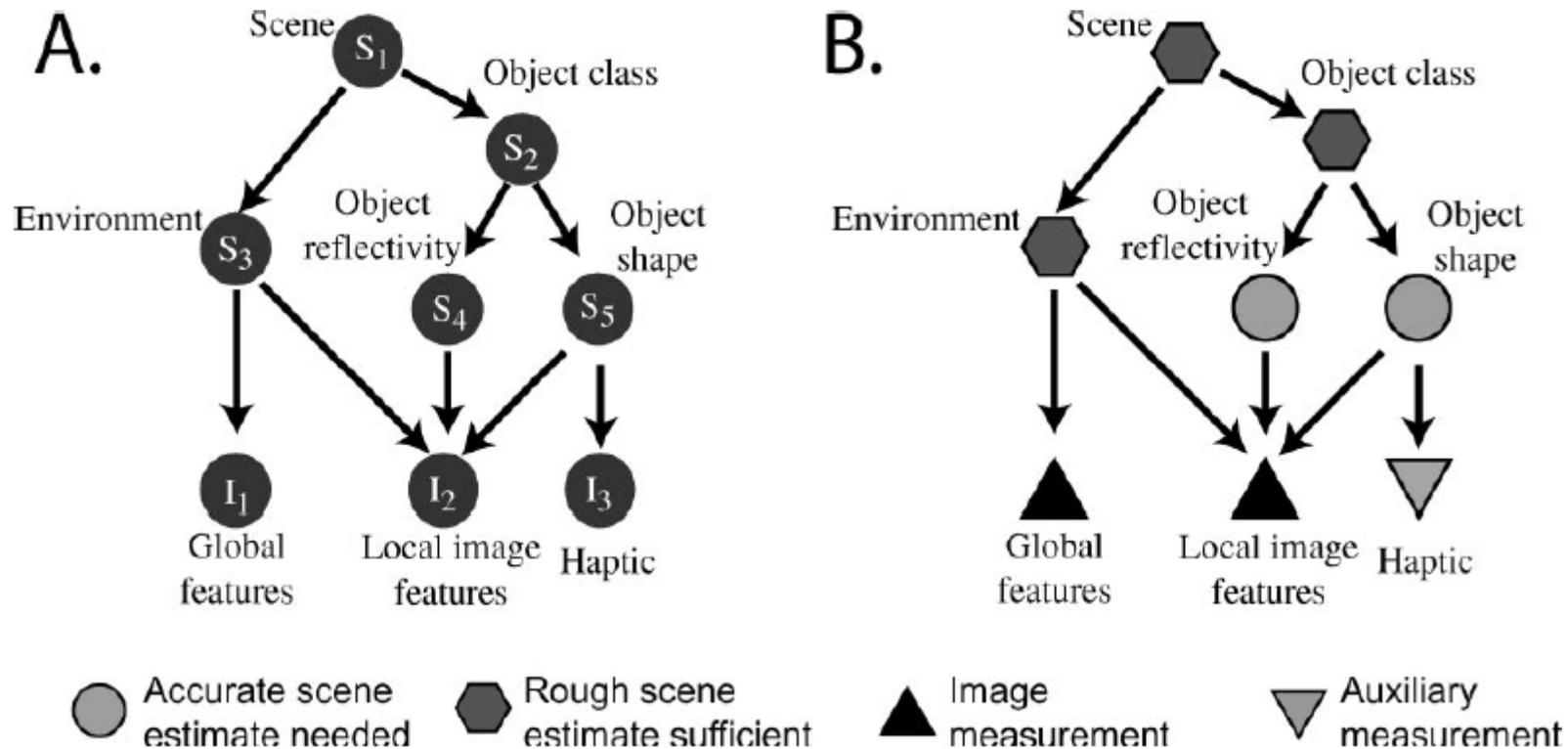
Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annu Rev Psychol*, 55, 271-304.

L'inférence perceptive peut être nettement plus complexe que dans l'exemple précédent

Les entrées sensorielles résultent fréquemment d'une combinaison complexe de variables et de sources de bruit.

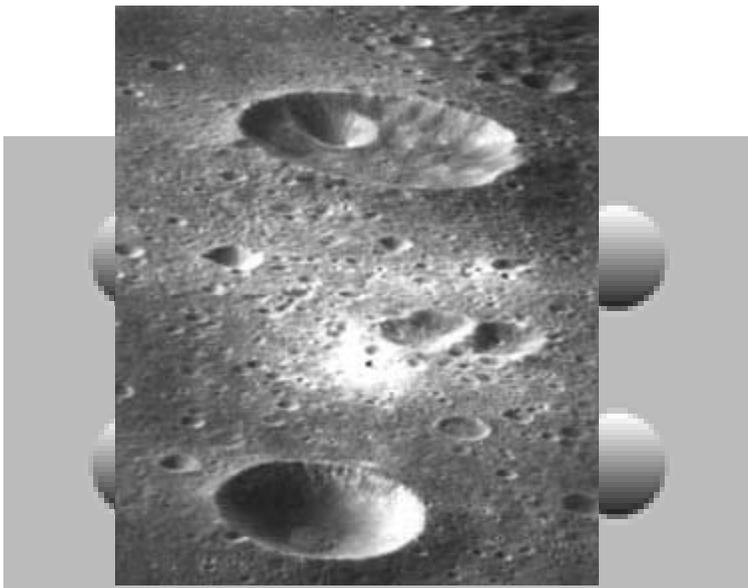
Les réseaux Bayésiens permettent de modéliser la propagation des contraintes.

La tâche exigée détermine l'importance relative des paramètres à estimer.



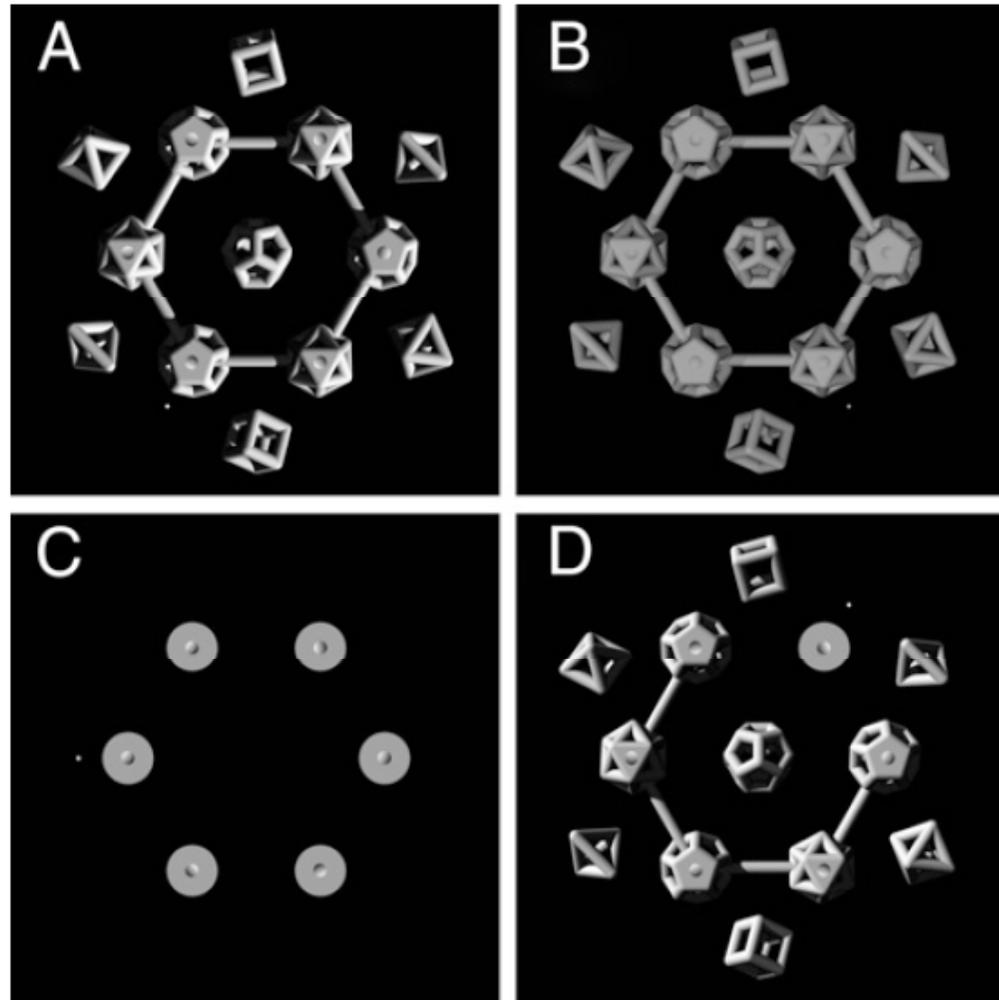
Exemple: la perception de l'illumination et son influence sur la perception de figures concaves ou convexes

En l'absence d'évidence, le système visuel « fait l'hypothèse » que la lumière vient d'en haut (et à gauche)



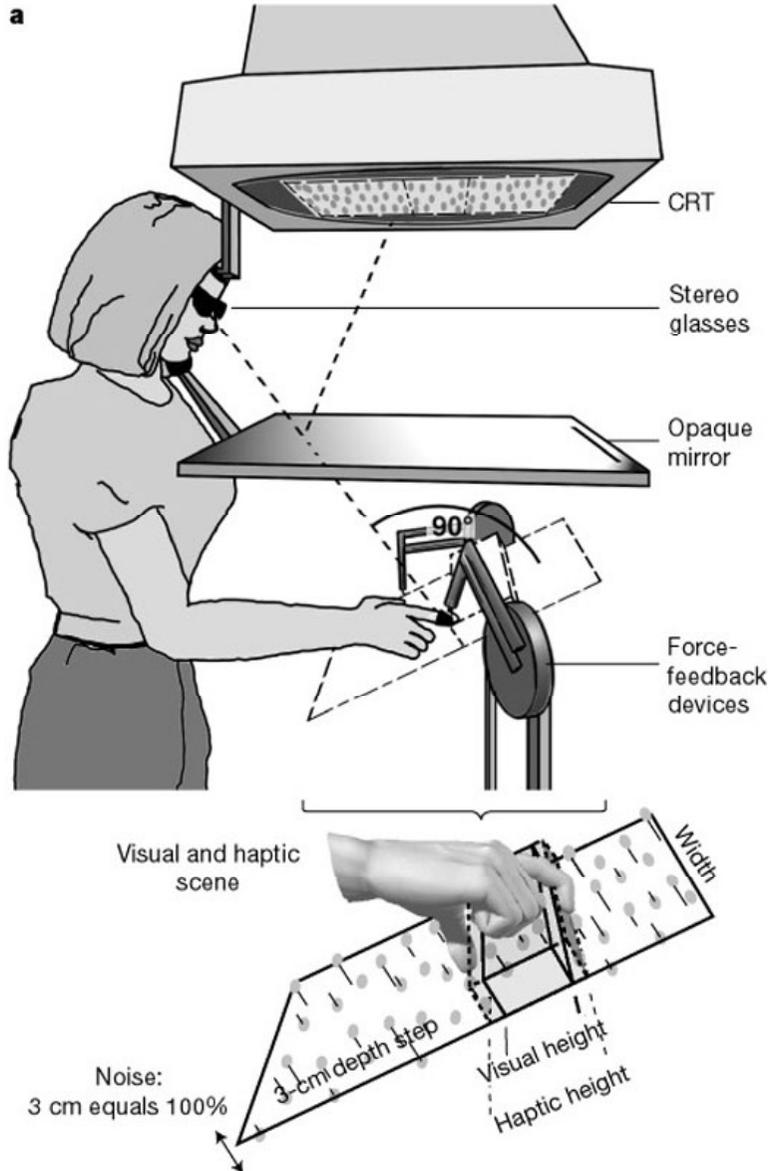
Ramachandran, V. S. (1988). Perceiving shape from shading. *Sci Am*, 259(2), 76-83.

Le système intègre efficacement l'*a priori* avec les indices perceptifs pour réviser son jugement.



Morgenstern, Y., Murray, R. F., & Harris, L. R. (2011). The human visual system's assumption that light comes from above is weak. *Proc Natl Acad Sci U S A*, 108(30), 12551-12553.

L'intégration Bayésienne de plusieurs indices sensoriels

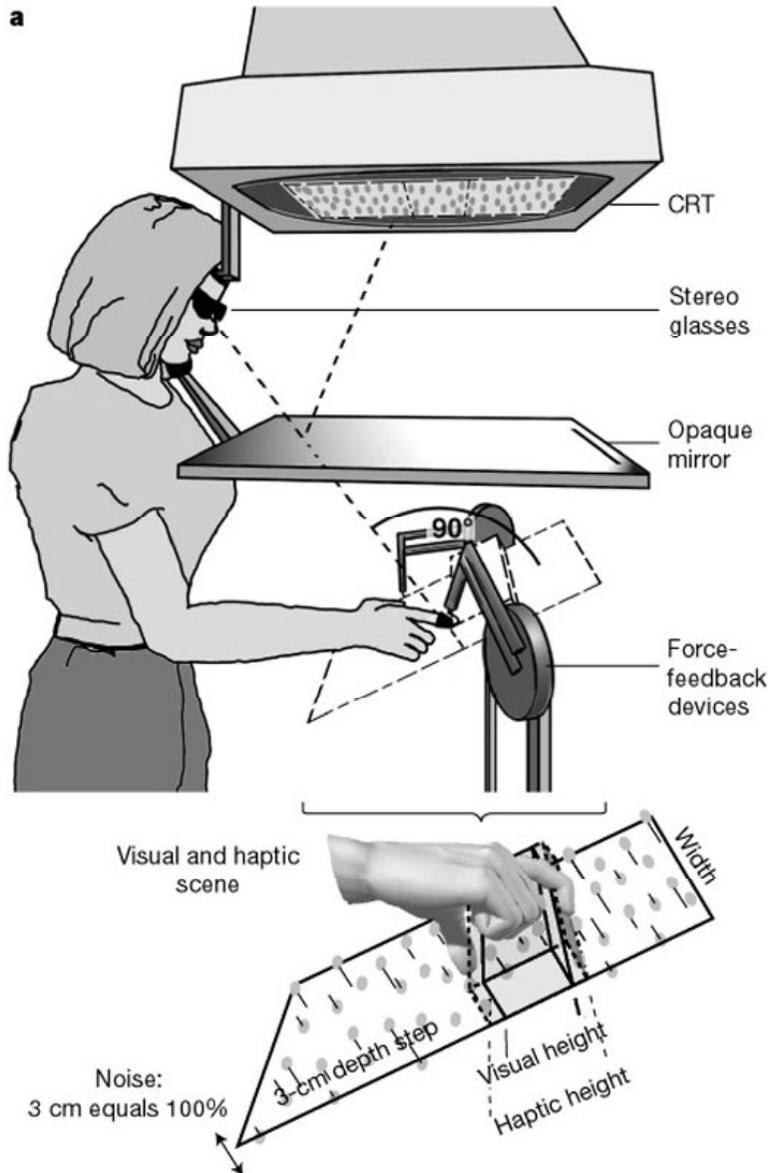


Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429-433.

Imaginons que nous recevions simultanément des indices visuels et tactiles sur la taille d'un objet.

Comment combiner ces deux informations?

L'intégration Bayésienne de plusieurs indices sensoriels



Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429-433.

Si ces indices sont conditionnellement indépendants, la théorie Bayésienne nous dit que la densité de probabilité de leur combinaison est le produit des deux densités.

$$\begin{aligned}
 P(w|t,v) &= \frac{P(t,v|w)P(w)}{P(t,v)} \\
 &= \frac{P(t|w)P(v|w)P(w)}{P(t,v)} \\
 &\propto P(t|w)P(v|w)P(w)
 \end{aligned}$$

L'intégration Bayésienne de plusieurs indices sensoriels

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429-433.

Le produit de deux Gaussiennes est une nouvelle Gaussienne

Le principe du maximum de vraisemblance prédit que

-la perception est une moyenne pondérée des valeurs suggérées par chaque indice

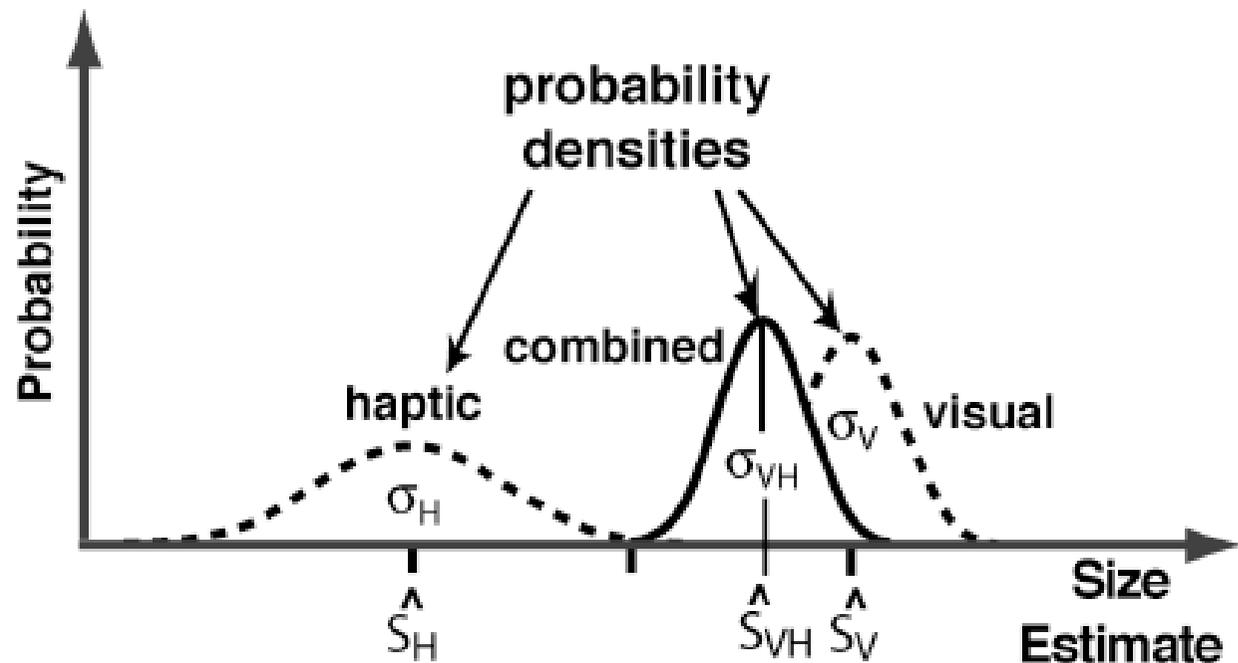
-les pondérations sont fonction de la fiabilité (*reliability*) des indices (l'inverse de la variance)

-la fiabilité totale est la somme des fiabilités (l'information de Fischer est additive pour des signaux indépendants)

$$\hat{S} = \sum_i w_i \hat{S}_i$$

$$w_j = \frac{r_j}{\sum_i r_i}$$

$$r = \sum_i r_i$$

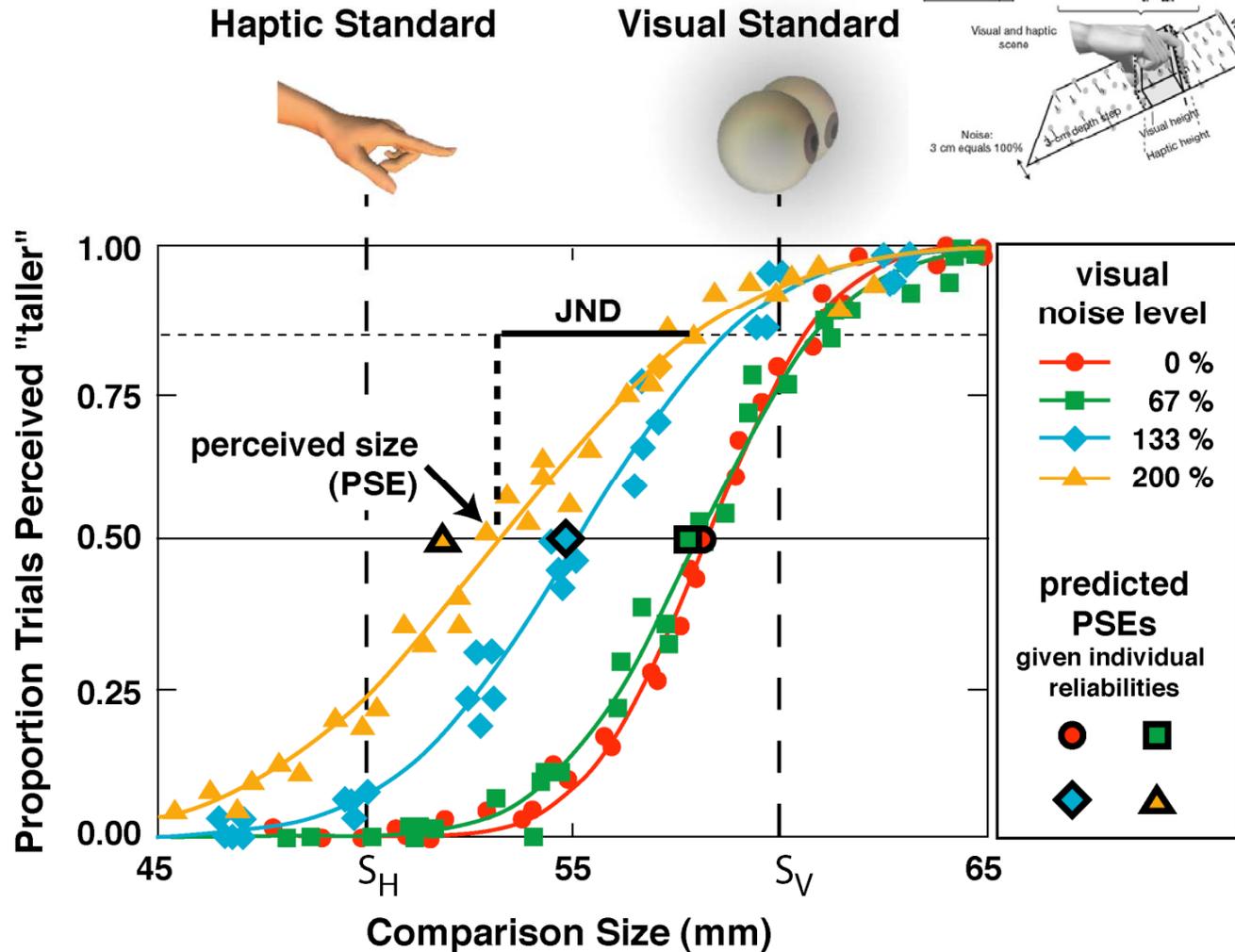
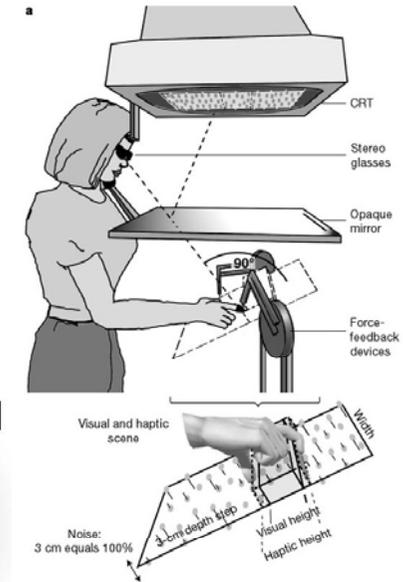


L'intégration Bayésienne de plusieurs indices sensoriels

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429-433.

Résultats:

- la taille perçue se déplace en direction de celle suggérée par la vision
- en proportion directe de la fiabilité des indices visuels
- avec un niveau de bruit qui décroît lorsque la fiabilité augmente



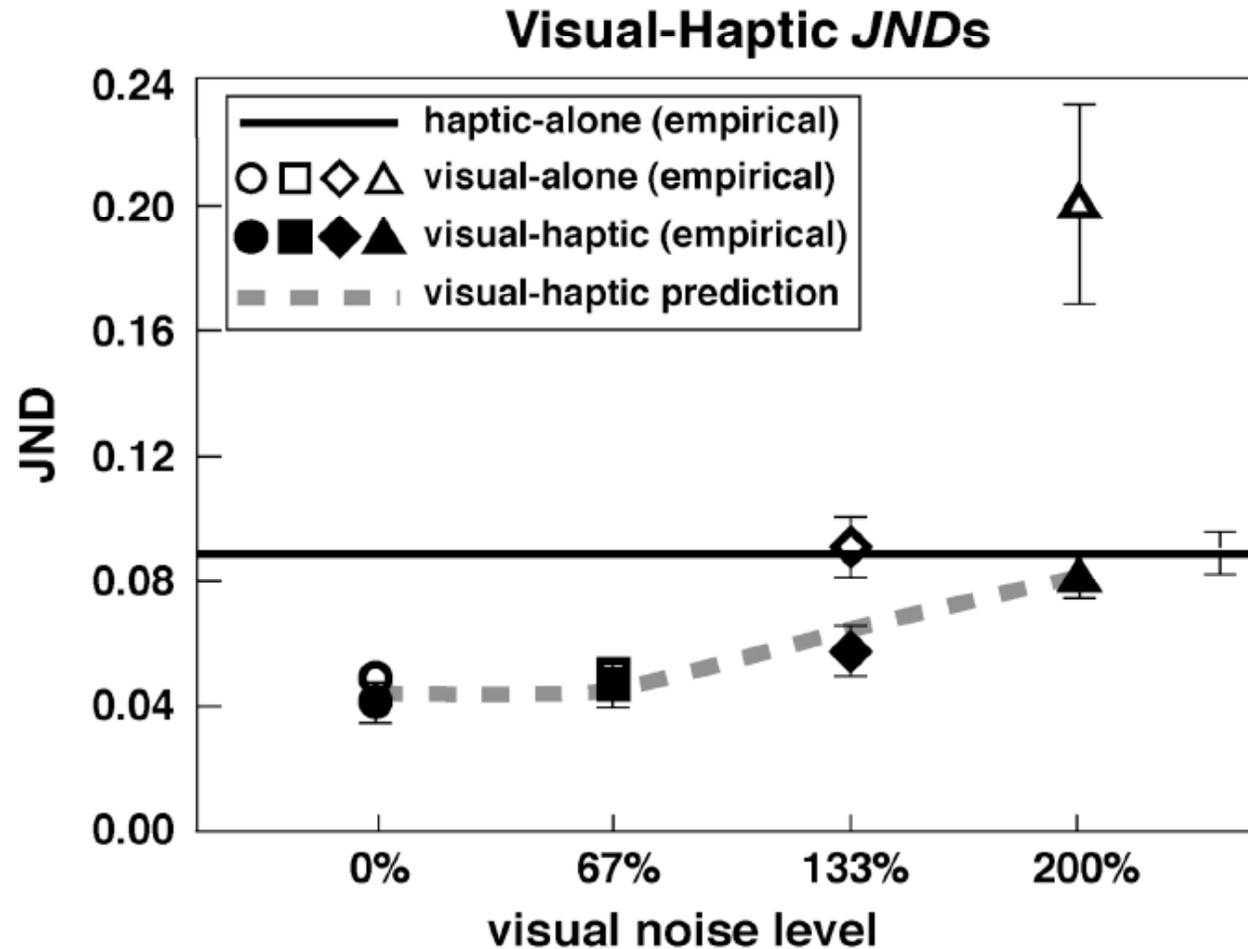
L'intégration Bayésienne de plusieurs indices sensoriels

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429-433.

La précision de la réponse (JND = *just noticeable difference*) est quantitativement conforme aux prédictions de la théorie Bayésienne

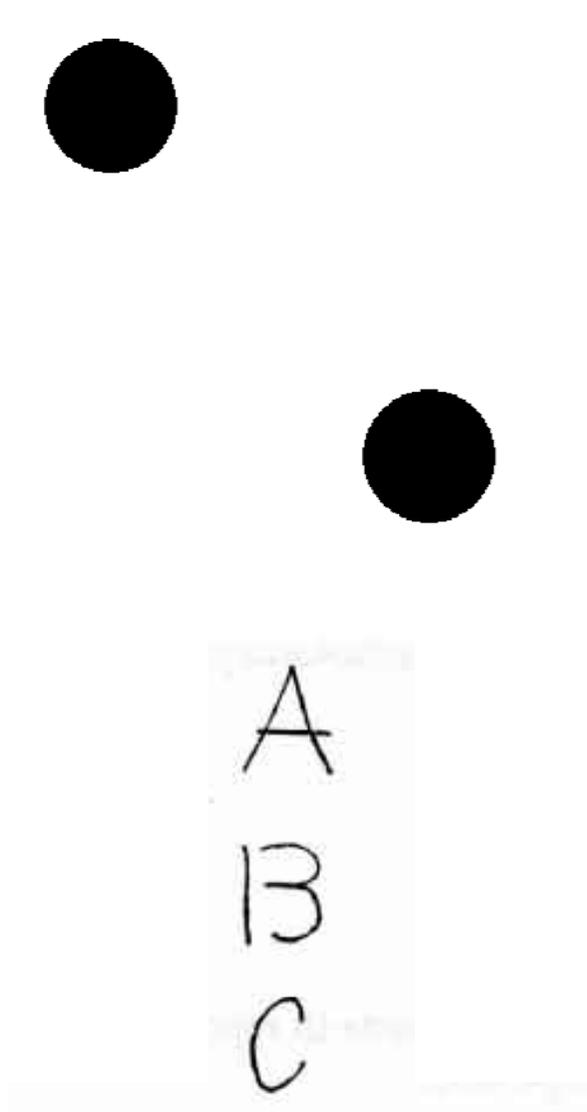
Conclusion:

Le système perceptif intègre les sensations issues de deux modalités sensorielles selon les lois de l'inférence Bayésienne.



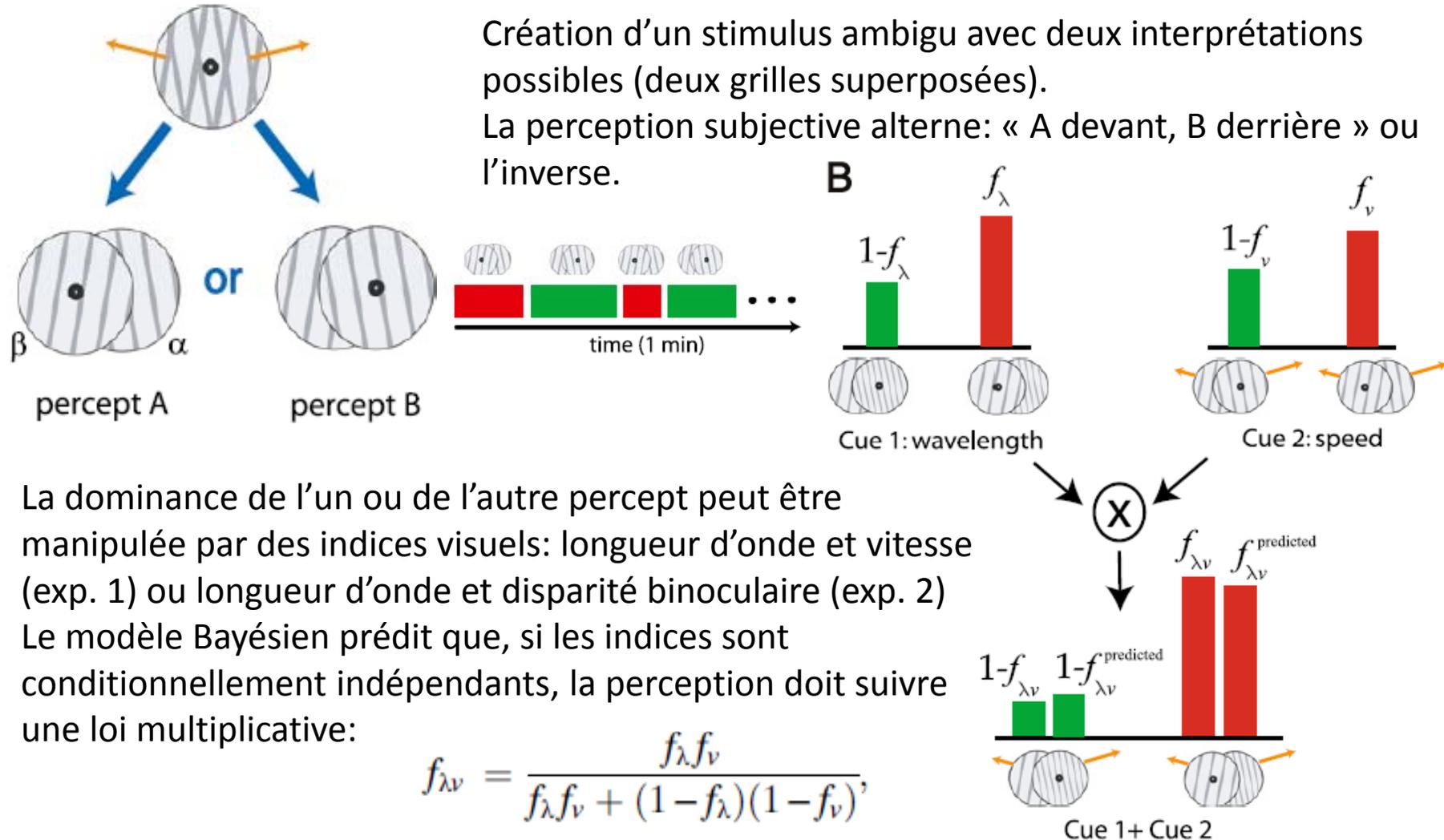
L'ambiguïté et la conscience:

Notre perception ne nous donne souvent à voir qu'un **échantillon** de la distribution de probabilités sous-jacente



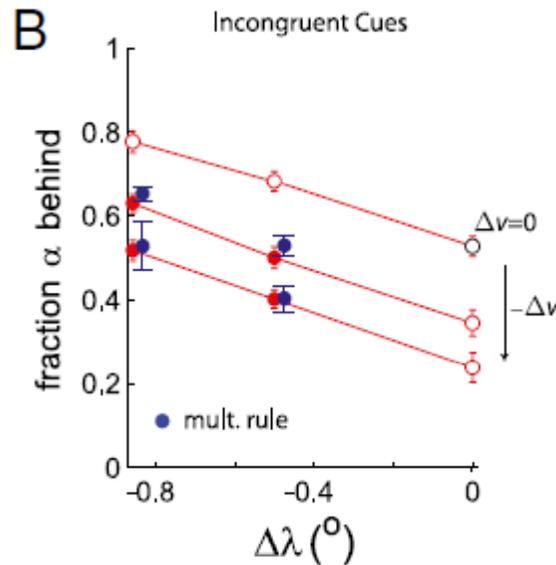
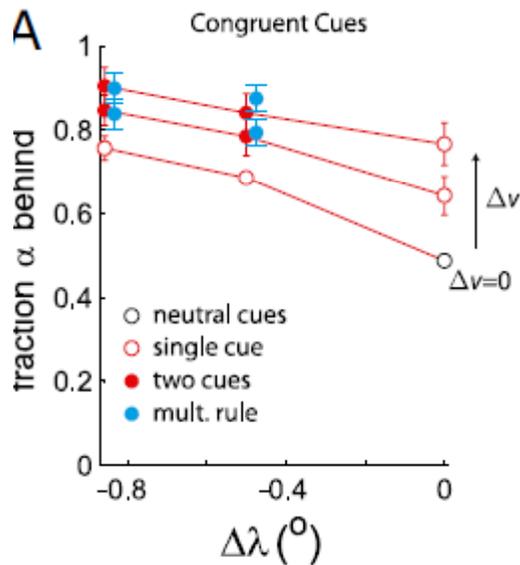
La perception de figures ambiguës reflète un échantillonnage Bayésien (*sampling*)

Moreno-Bote, R., Knill, D. C., & Pouget, A. (2011). Bayesian sampling in visual perception. *Proc Natl Acad Sci U S A*, 108(30), 12491-12496.



La perception de figures ambiguës reflète un échantillonnage Bayésien (*sampling*)

Moreno-Bote, R., Knill, D. C., & Pouget, A. (2011). Bayesian sampling in visual perception. *Proc Natl Acad Sci U S A*, 108(30), 12491-12496.

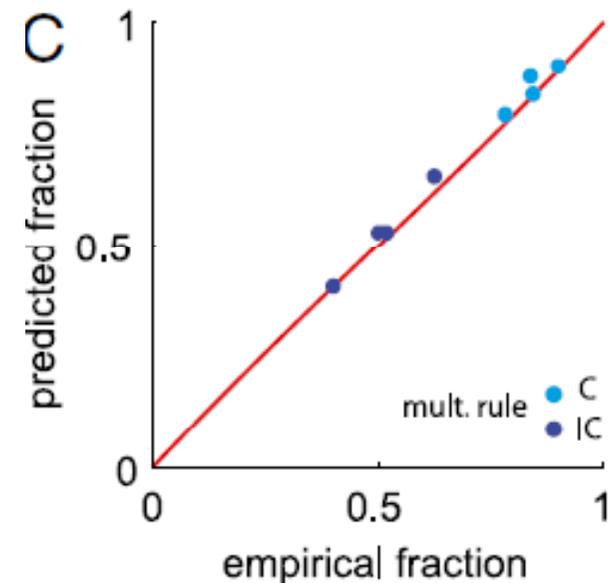


Résultats:

Que les indices soient congruents ou incongruents, la fraction de perception des items peut être prédite selon la règle du produit, à partir des distributions observées quand un seul indice est manipulé

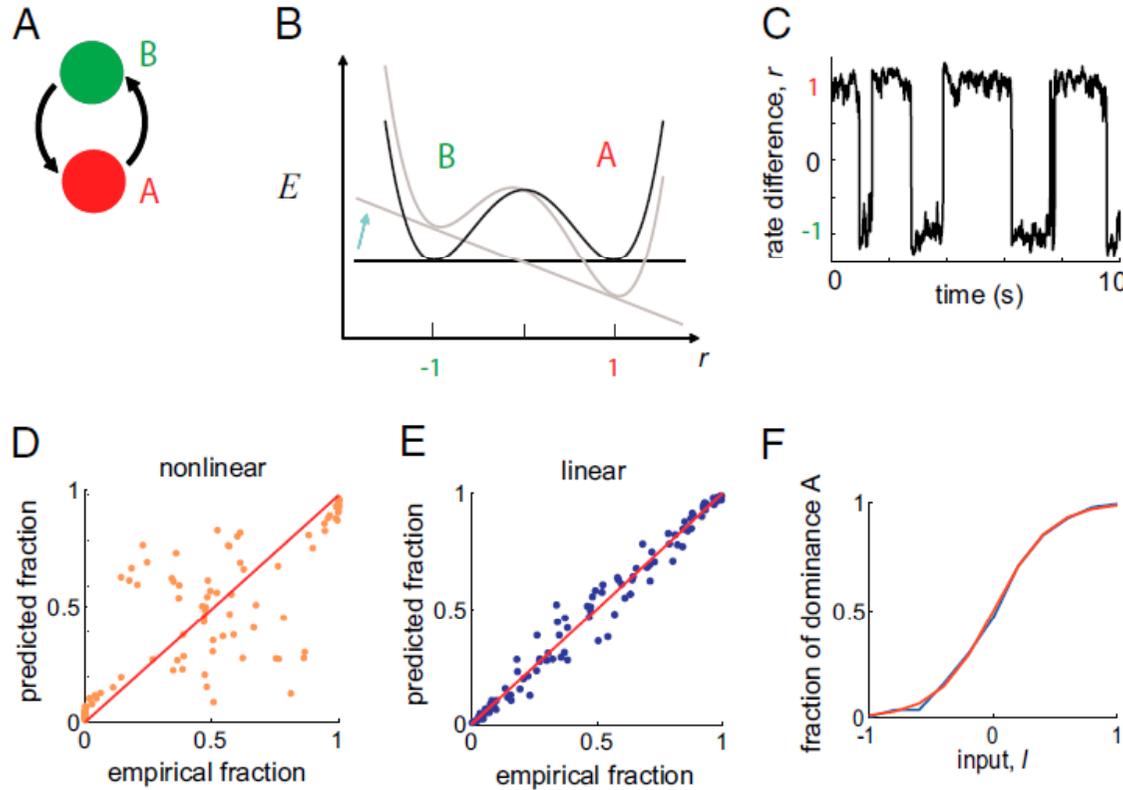
L'ajustement des courbes expérimentales est presque parfait, ce qui n'est pas le cas pour d'autres modèles, par exemple si l'on suppose que l'un des indices domine l'autre.

Conclusion: "the fraction of dominance in bistable perception behaves as a probability"



Comment modéliser l'échantillonnage Bayésien?

Moreno-Bote, R., Knill, D. C., & Pouget, A. (2011). Bayesian sampling in visual perception. *Proc Natl Acad Sci U S A*, 108(30), 12491-12496.



Les données peuvent s'expliquer par un modèle à deux attracteurs. Une entrée I biaise la fonction d'énergie vers l'attracteur A ou B. Du bruit permet de basculer le système d'un attracteur à l'autre (B).

Le système oscille entre deux états (C)

La fraction du temps passé dans chaque état obéit à la règle du produit (E), seulement si l'entrée est une somme des entrées I_1 et I_2 . Pourquoi?

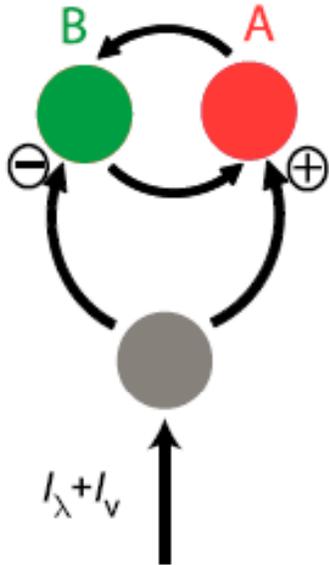
La fraction de temps passé dans l'un des états est une fonction sigmoïde, donc proche d'une exponentielle $\exp(I/\sigma^2)$

Si l'entrée est une fonction linéaire de deux indices $I = I_1 + I_2$, alors ce modèle possède exactement la propriété multiplicative requise, car:

$$\exp((I_1 + I_2)/\sigma^2) = \exp(I_1/\sigma^2) \exp(I_2/\sigma^2)$$

Comment modéliser l'échantillonnage Bayésien?

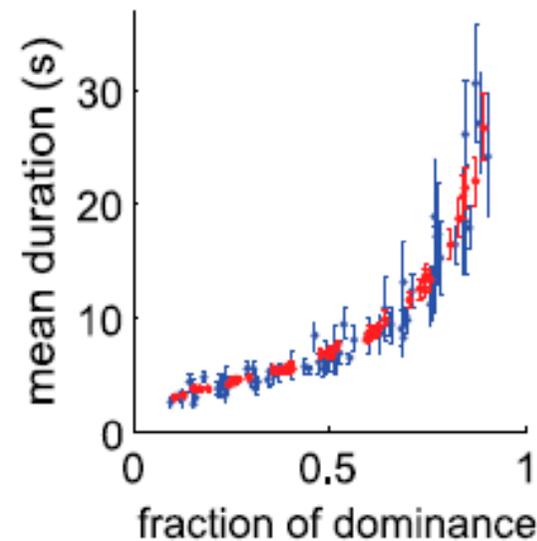
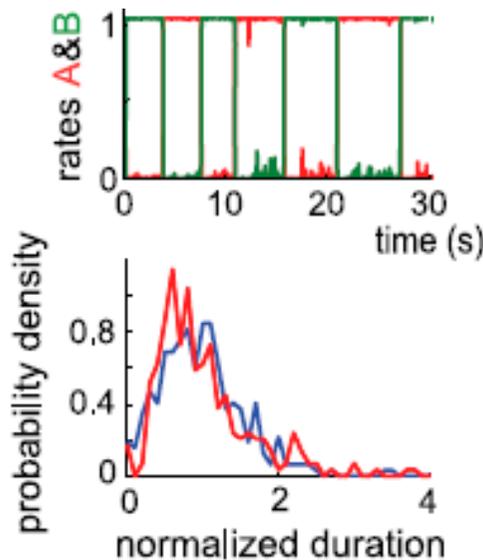
Moreno-Bote, R., Knill, D. C., & Pouget, A. (2011). Bayesian sampling in visual perception. *Proc Natl Acad Sci U S A*, 108(30), 12491-12496.



Ce modèle théorique peut être simulé par un réseau de neurones formels (et même implémenté avec des neurones à décharges réalistes; voir aussi Wong et Wang, *J Neuroscience* 2006).

La simulation conduit à des propriétés nouvelles et importantes:

- la durée de vie d'un attracteur suit une loi gamma (en rouge les prédictions du modèle, en bleu les données empiriques)
- lorsqu'on varie les paramètres du stimulus, il existe une relation systématique qui relie durée moyenne et fraction de dominance:



L'hypothèse d'un échantillonnage conscient

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: selection is probabilistic; responses are all-or-none samples. *J Exp Psychol Gen*, 138(4), 546-560.

Selon Vul et al., tous les calculs perceptifs et attentionnels seraient probabilistes. Cependant, au moment de répondre, le cerveau échantillonnerait la distribution de probabilité résultante.

L'échantillonnage serait une réponse à la difficulté de réaliser des calculs Bayésiens complets, avec l'ensemble de la distribution.

Dans sa thèse, Vul montre que travailler à partir de quelques échantillons suffit souvent à prendre la décision appropriée – et que s'il y a une pression temporelle importante, prendre un seul échantillon optimise le renforcement par unité de temps.

Test du modèle:

Sélection et rapport
d'une lettre.

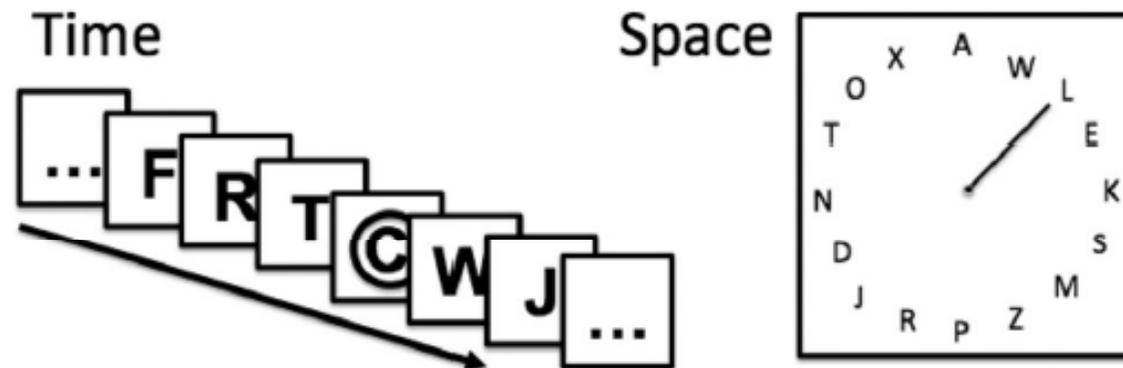
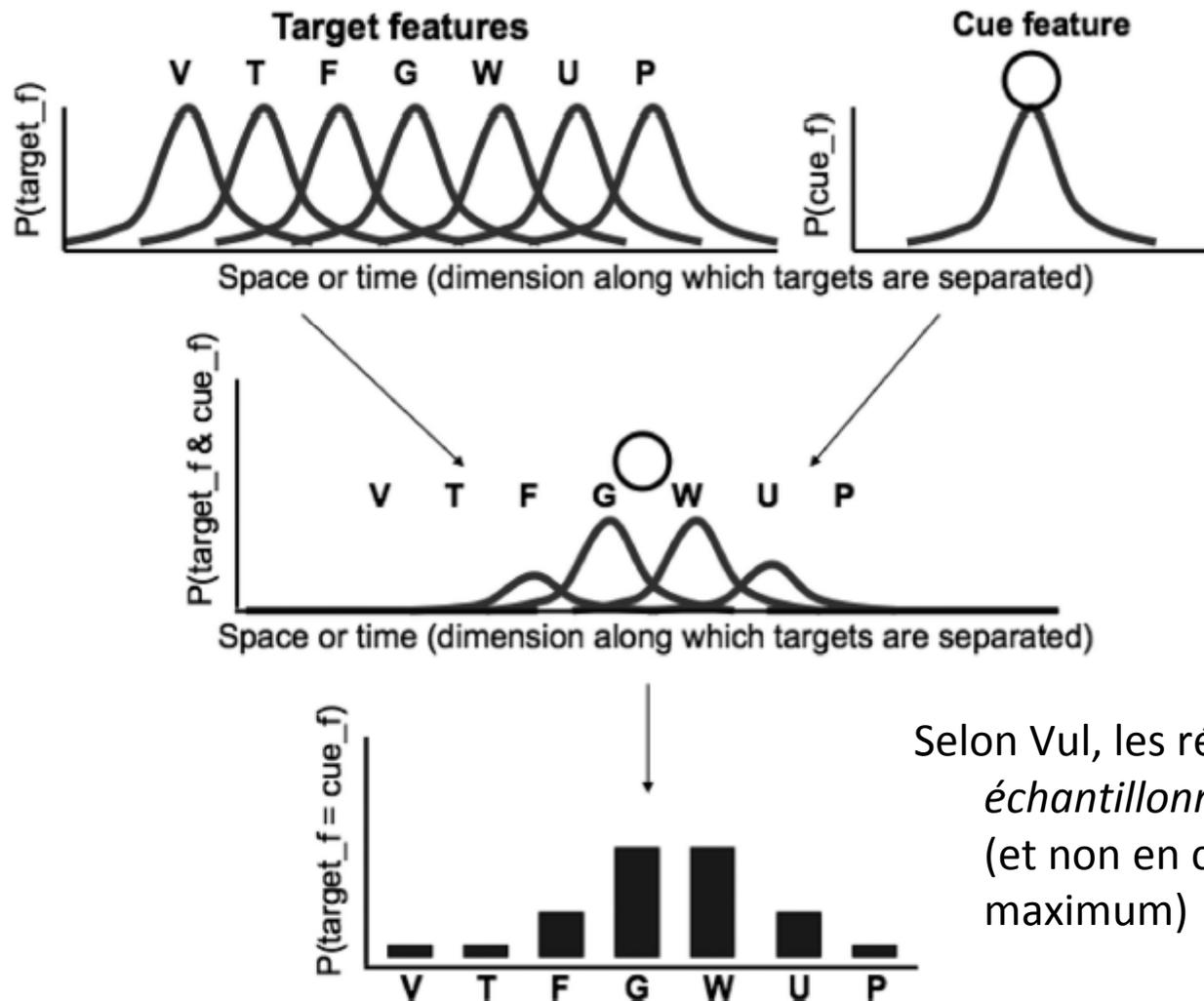


Figure 1. Prototypical experimental paradigms used in attentional selection tasks in time and space. One of many items is cued, and the subject must report that item.

L'hypothèse d'un échantillonnage conscient

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: selection is probabilistic; responses are all-or-none samples. *J Exp Psychol Gen*, 138(4), 546-560.

Modèle Bayésien proposé:

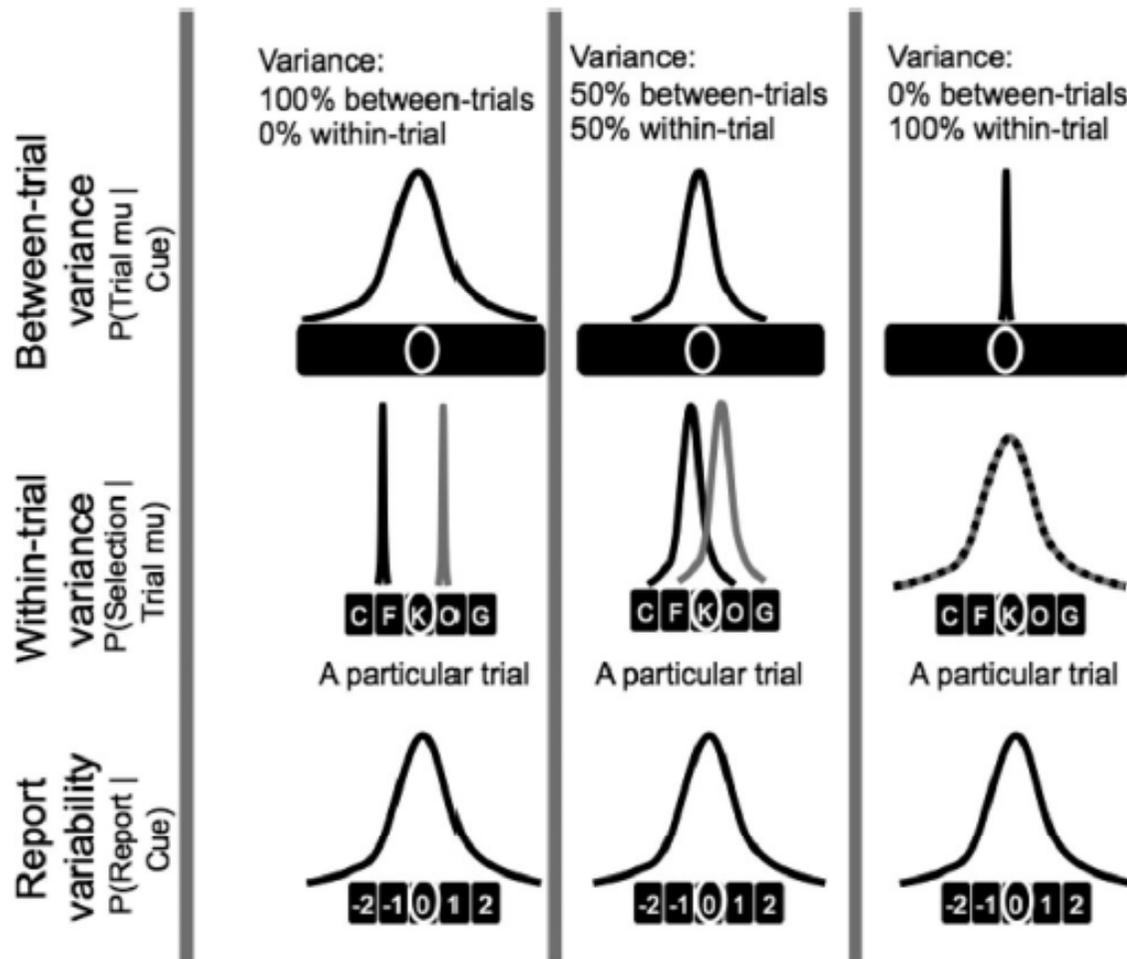


Selon Vul, les réponses du sujet *échantillonnent* cette distribution (et non en choisissent le maximum)

L'hypothèse d'un échantillonnage conscient

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: selection is probabilistic; responses are all-or-none samples. *J Exp Psychol Gen*, 138(4), 546-560.

Le fait que les erreurs, à travers les essais, soient distribuées autour de la vraie cible, ne suffit pas à établir la véracité de ce modèle.



En effet, il se pourrait que la variabilité des réponses provienne d'une variabilité *inter-essais* dans la sélection de la lettre cible.

La sélection attentionnelle serait tout-ou-rien, mais son point d'application varierait d'essai en essai.

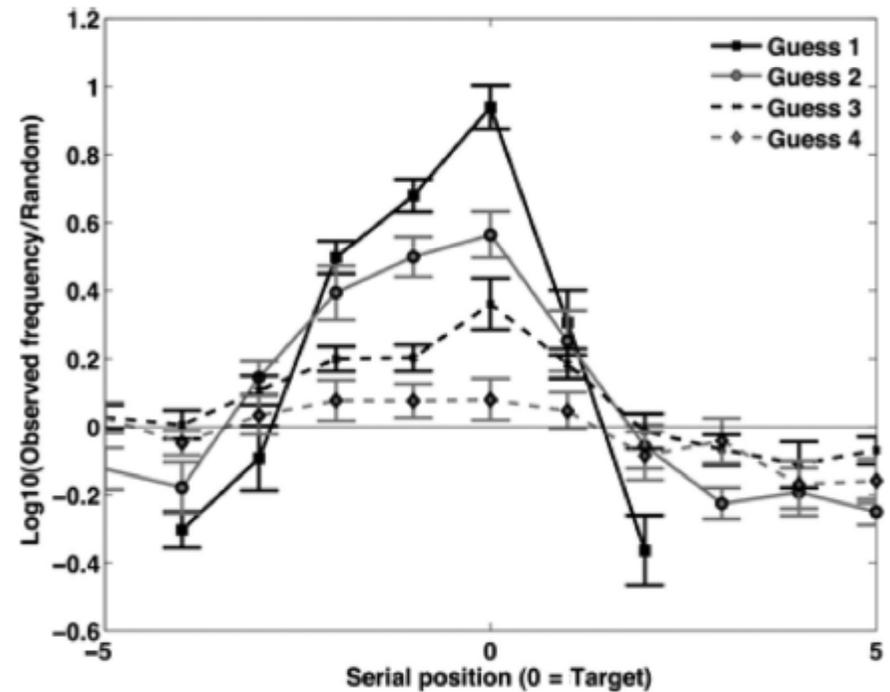
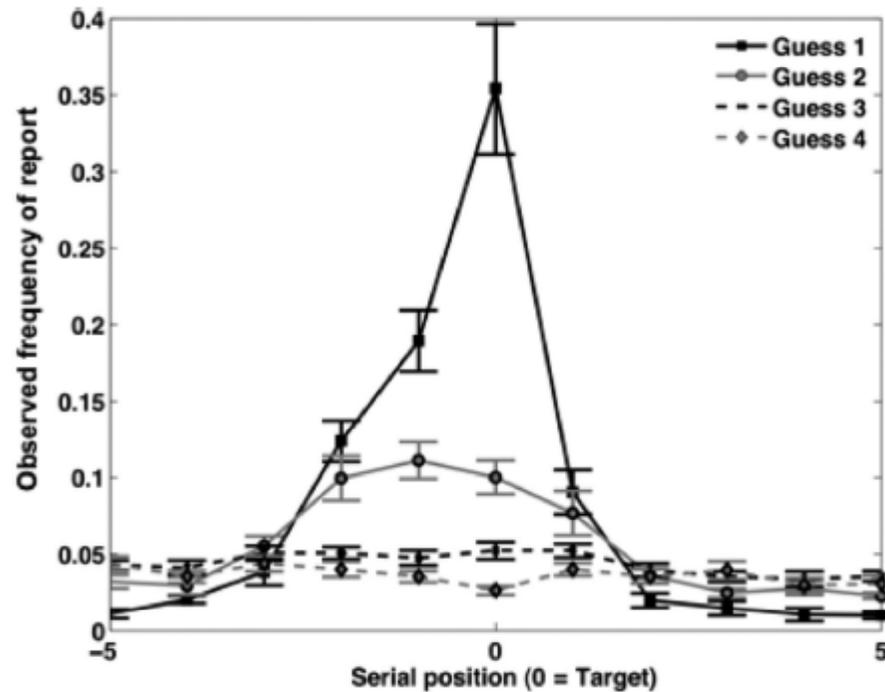
Comment rejeter cette interprétation?

L'hypothèse d'un échantillonnage conscient

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: selection is probabilistic; responses are all-or-none samples. *J Exp Psychol Gen*, 138(4), 546-560.

Solution = demander plusieurs réponses à *chaque essai*:

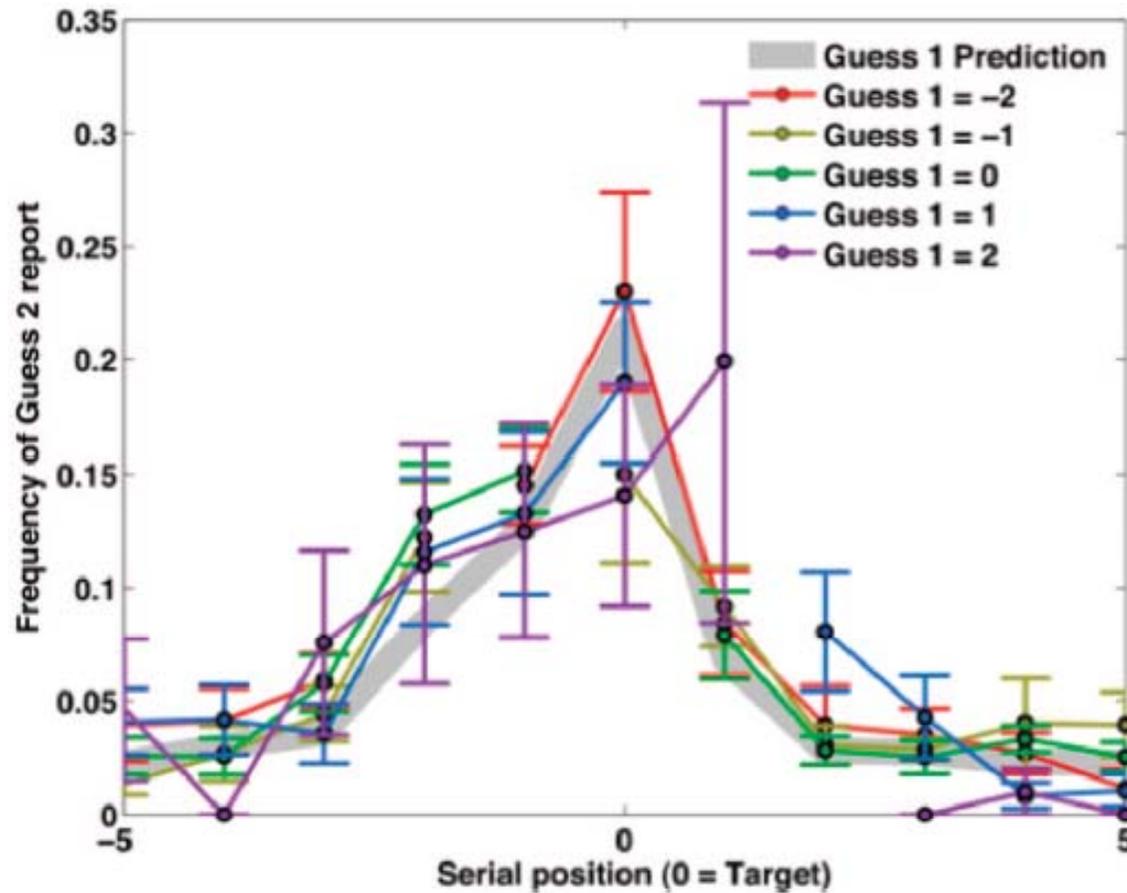
La seconde réponse n'est pas au hasard, ni la troisième; et leurs fréquences ne changent pas



L'hypothèse d'un échantillonnage conscient

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: selection is probabilistic; responses are all-or-none samples. *J Exp Psychol Gen*, 138(4), 546-560.

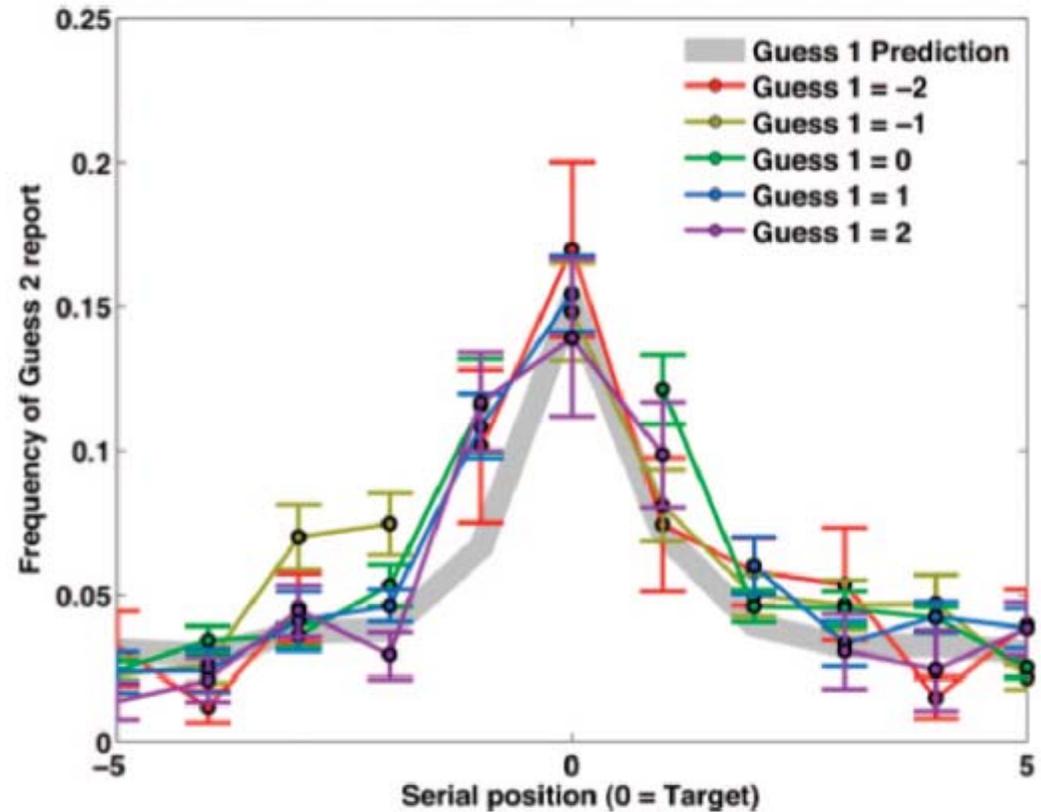
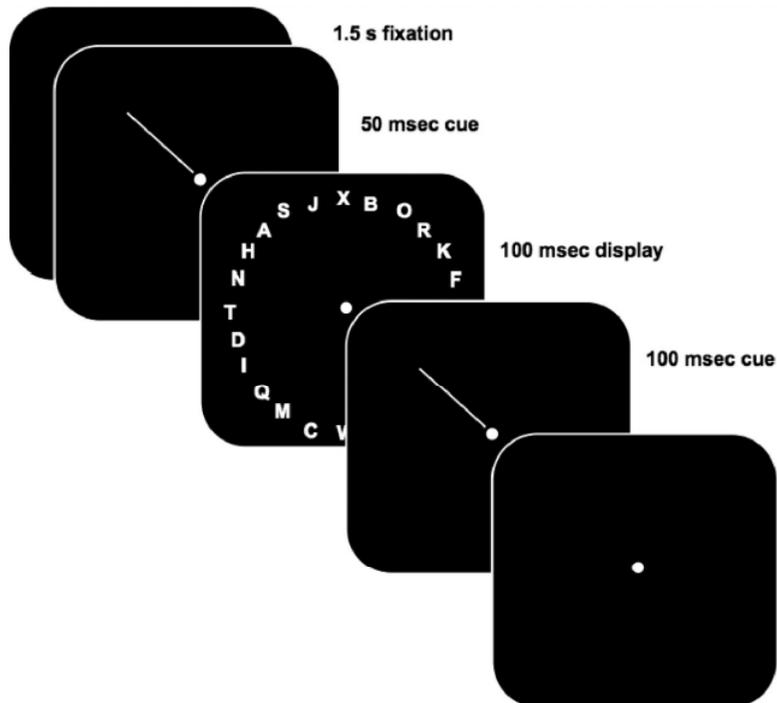
La réponse 2 se comporte comme un second échantillon de la même distribution, totalement indépendant de la position de la réponse 1.



L'hypothèse d'un échantillonnage conscient

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: selection is probabilistic; responses are all-or-none samples. *J Exp Psychol Gen*, 138(4), 546-560.

Les résultats sont répliqués dans une seconde expérience où la sélection est spatiale et non temporelle.



L'hypothèse d'un échantillonnage conscient

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: selection is probabilistic; responses are all-or-none samples. *J Exp Psychol Gen*, 138(4), 546-560.

Conclusions:

- les résultats indiquent que la sélection attentionnelle, temporelle et spatiale, opère de manière graduelle et statistique.
- Il existe une variabilité des réponses entre les essais, mais *toute* cette variabilité semble provenir de la variabilité *intra*-essai dans la fonction de sélection
- Les réponses rapportées se comportent comme des échantillons d'une distribution interne qui obéit aux règles d'inférence Bayésienne [bien que ce dernier point ne soit pas testé très systématiquement par Vul et al.].
- Bien que le cerveau fonctionne d'une manière probabiliste, nous ne prenons conscience que d'échantillons discrets:
 - L'attention sélective opère de façon non-consciente et continue
 - L'accès à la conscience opère de façon tout-ou-rien
 - Ces résultats sont compatibles avec ceux de Sergent et Dehaene, 2004, 2005

Un échantillonnage dans les inférences cognitives?

Vul, E., & Pashler, H. (2008). Measuring the Crowd Within: Probabilistic Representations Within Individuals. *Psychological Science (Wiley-Blackwell)*, 19(7), 645-647.

« La sagesse des foules »: Galton (1907) montre que, lorsqu'on demande à des personnes de juger du poids d'un bœuf, l'erreur sur la moyenne des réponses est moindre que la moyenne des erreurs de chaque personne.

Il y a du sens à demander « l'avis du public »...

Dans l'expérience, 428 sujets répondent à des questions du genre « quel pourcentage des aéroports mondiaux se trouvent aux Etats-Unis? »

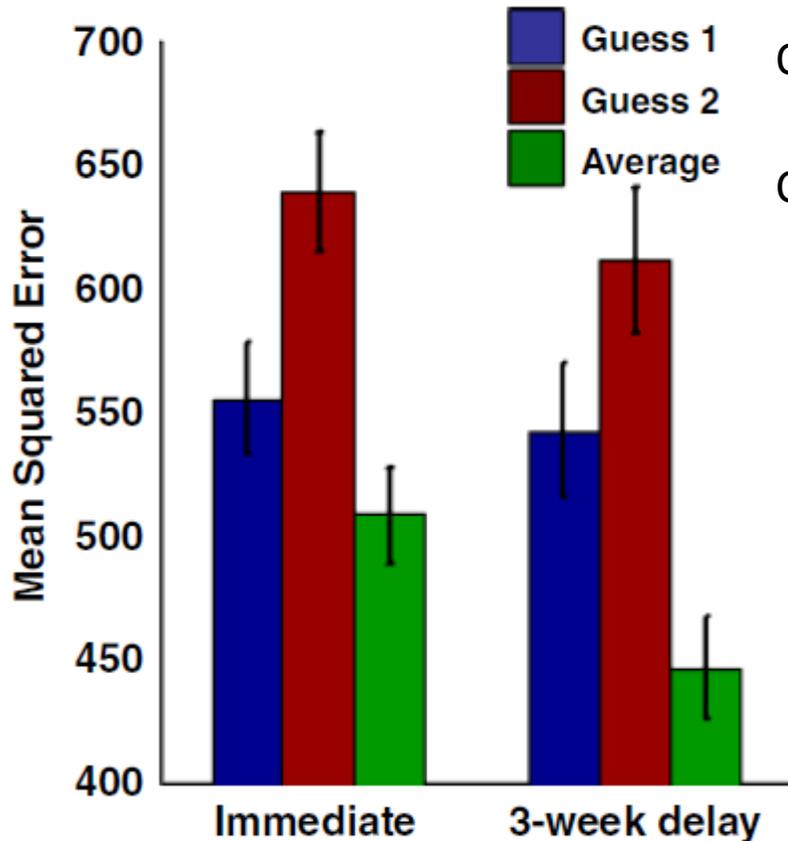
La moitié donnent immédiatement une seconde réponse, tandis que l'autre moitié est recontactée 3 semaines plus tard.

Etonnamment, le fait de se poser deux fois la même question améliore le résultat:

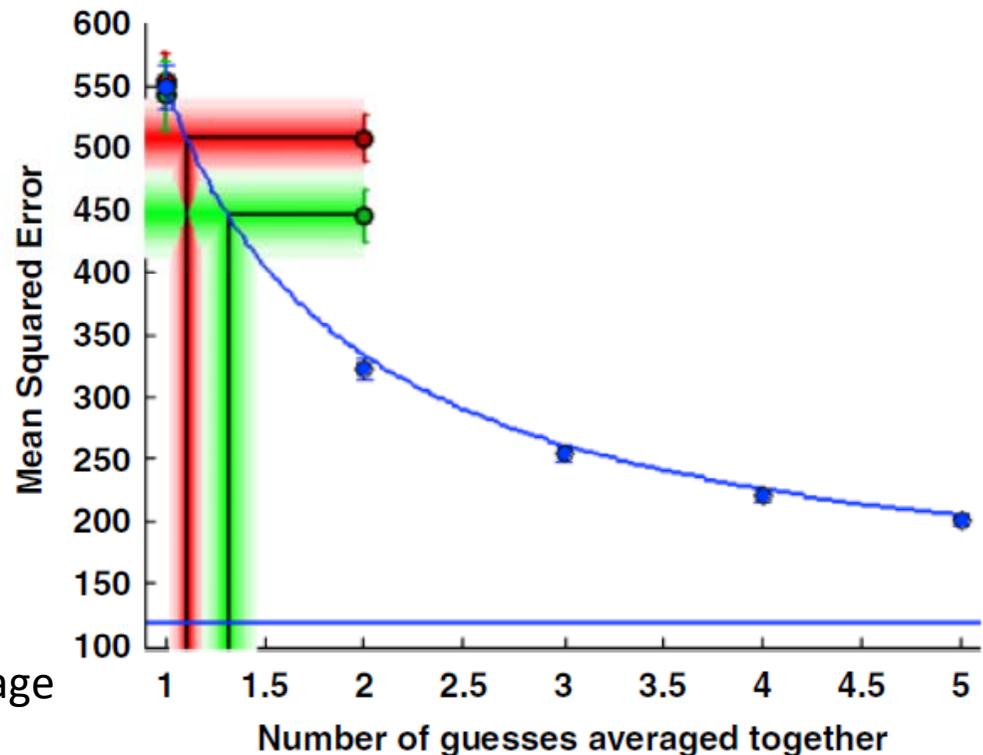


Un échantillonnage dans les inférences cognitives?

Vul, E., & Pashler, H. (2008). Measuring the Crowd Within: Probabilistic Representations Within Individuals. *Psychological Science (Wiley-Blackwell)*, 19(7), 645-647.



On peut comparer l'amélioration observée à celle qui consisterait à demander à une seconde personne. C'est environ 1/10 pour une seconde réponse immédiate, et 1/3 pour une réponse trois semaines plus tard.



Conclusion: même les raisonnements plausibles, fondés sur la recherche en mémoire, s'appuieraient sur l'échantillonnage d'une distribution de probabilité.

Conclusion: inférence et échantillonnage

Lorsque nous percevons un stimulus ou que nous prenons une décision, la valeur perçue ou choisie se comporte comme un **échantillon** de la distribution a posteriori, étant données les indices dont nous disposons.

Cette distribution combine, selon les règles de l'inférence Bayésiennes, les *a priori* issues de notre connaissance préalable du monde extérieur, ainsi que les différentes indications ou sensations que nous recevons à un instant donné.

Pourquoi échantillonner une distribution?

L'échantillonnage n'est pas nécessairement la meilleure solution: il pourrait être plus adapté de travailler avec la distribution toute entière, ou bien de choisir toujours la solution la plus probable (*maximum a posteriori*, MAP).

Plusieurs possibilités:

- Au-delà d'un certain niveau de complexité, l'échantillonnage est la seule manière de faire le calcul
- Le système nerveux est capable de représenter des distributions, mais doit les réduire à un échantillon au moment de prendre une décision.
- Dans un monde changeant, l'échantillonnage permet d'explorer et de rester ouvert au changement (dilemme entre exploitation et exploration).

Le niveau de bruit ou l'utilisation d'une fonction puissance avant l'addition ($I = I_1^\alpha + I_2^\alpha$) permet de « régler » finement la dynamique du réseau sur l'axe exploitation-exploration. La présence de bruit peut également permettre de trouver l'optimum (algorithme de recuit simulé [*simulated annealing*]).