Introspection et métacognition :
Les mécanismes de la connaissance de soi

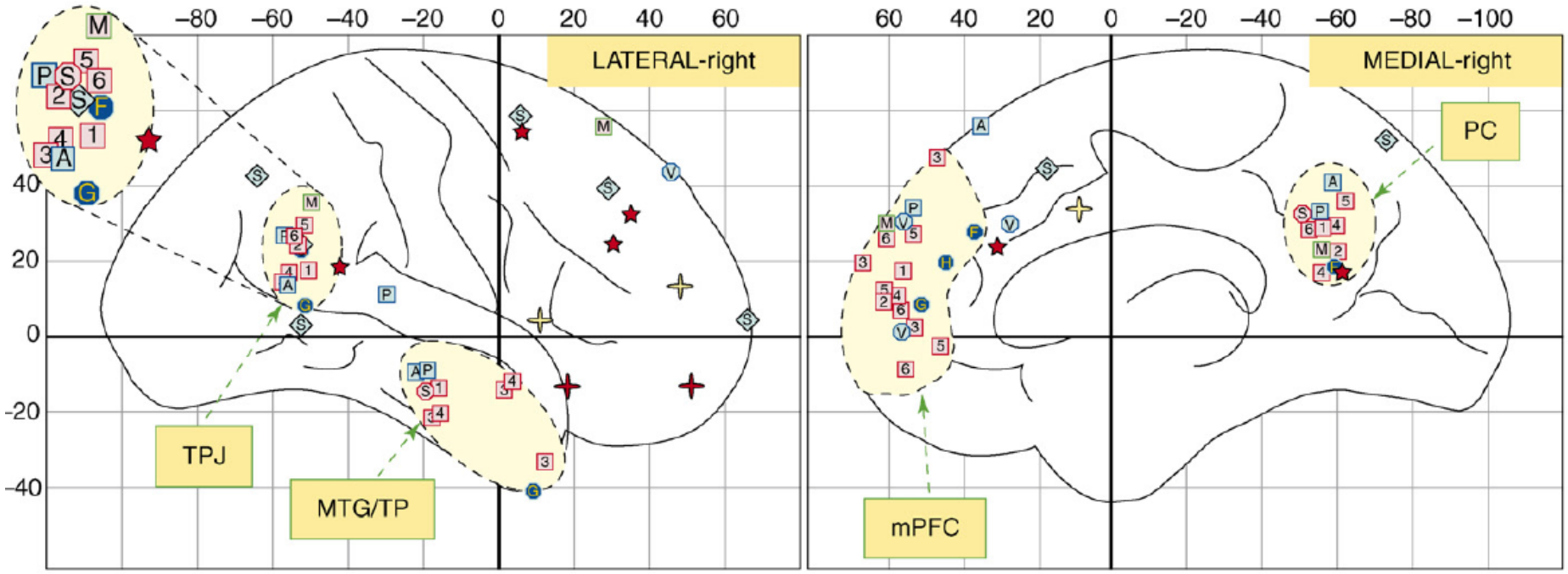
Stanislas Dehaene
Chaire de Psychologie Cognitive Expérimentale

Cours

Mécanismes cérébraux de la métacognition

Rappel du cours n°4: théorie de l'esprit et introspection de soi

Un réseau comprenant le cortex préfrontal antéro-mésial, le précuneus, la jonction temporo-pariétale (particulièrement à droite) et la partie antérieure du lobe temporal est impliqué simultanément dans la **théorie de l'esprit des autres** et dans la **représentation de soi**.



Perner et Aichorn, TICS 2008

Contribution de la neuropsychologie à l'étude de la métacognition: le rôle clé du cortex préfrontal

- Art Shimamura et Larry Squire (1986a,b) réalisent les premières études systématiques de la métamémoire chez les patients amnésiques.
- Les patients amnésiques peuvent avoir de profonds déficits de rappel en mémoire, et néanmoins un excellent « sentiment de savoir » (prédictif de la reconnaissance ultérieure de la bonne réponse)
- Selon Metcalfe (1993), c'est la préservation des indices de familiarité qui sous-tend les performances métacognitives
- Les amnésiques de Korsakoff, eux, ont à la fois des déficits mnésiques et des performances métacognitives anormales
- L'idée émerge que les lésions frontales pourraient être responsables de leur déficit métacognitif
- Cette hypothèse est testée directement par Janowsky et al (*Psychobiology*, 1989), dans une expérience bien contrôlée:

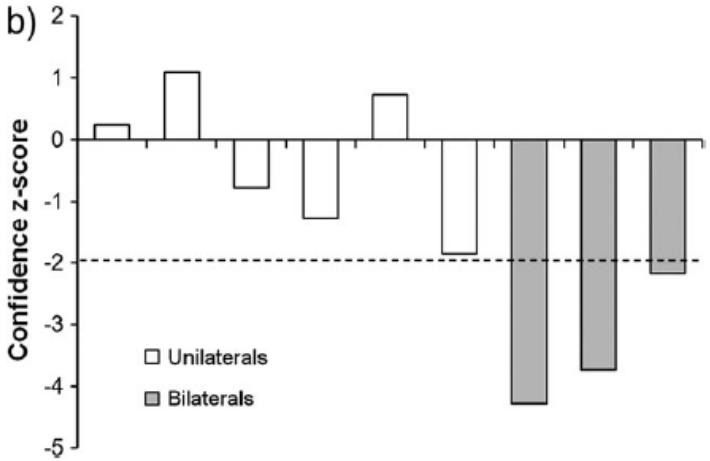
  - Présentation de phrases complètes. Il faut, plus tard, se souvenir du dernier mot.

  - Les performances de rappel sont égalisées entre contrôles, patients frontaux et temporaux (en faisant varier le délai de rappel)

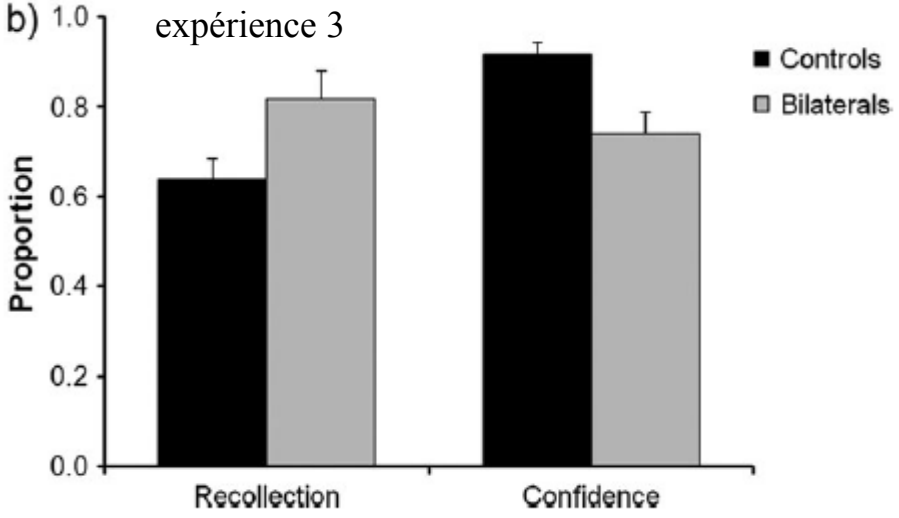
  - Les patients frontaux ont effectivement un important déficit du « sentiment de savoir »
- Le lien avec le cortex préfrontal est également confirmé par l'étude du vieillissement (Souchay et al., *Neuropsychology*, 2000): un déficit du « sentiment de savoir » existe chez les personnes âgées, et il est directement prédit par les tests des fonctions frontales.

Les lésions pariétales bilatérales affectent la méta-mémoire

Simons, J. S., Peers, P. V., Mazuz, Y. S., Berryhill, M. E., & Olson, I. R. (2010). Dissociation between memory accuracy and memory confidence following bilateral parietal lesions. *Cereb Cortex, 20(2), 479-485.*



Dans une tâche de mémoire épisodique, la personne doit se souvenir du contexte expérimental dans laquelle une phrase ou une image a été présentée, puis juger de la confiance dans sa propre réponse.
Les patients atteints de lésions pariétales bilatérales semblent avoir un déficit particulier dans le jugement de confiance (bien que celui ne soit pas mesuré, comme il le faudrait par la corrélation entre confiance et réussite)



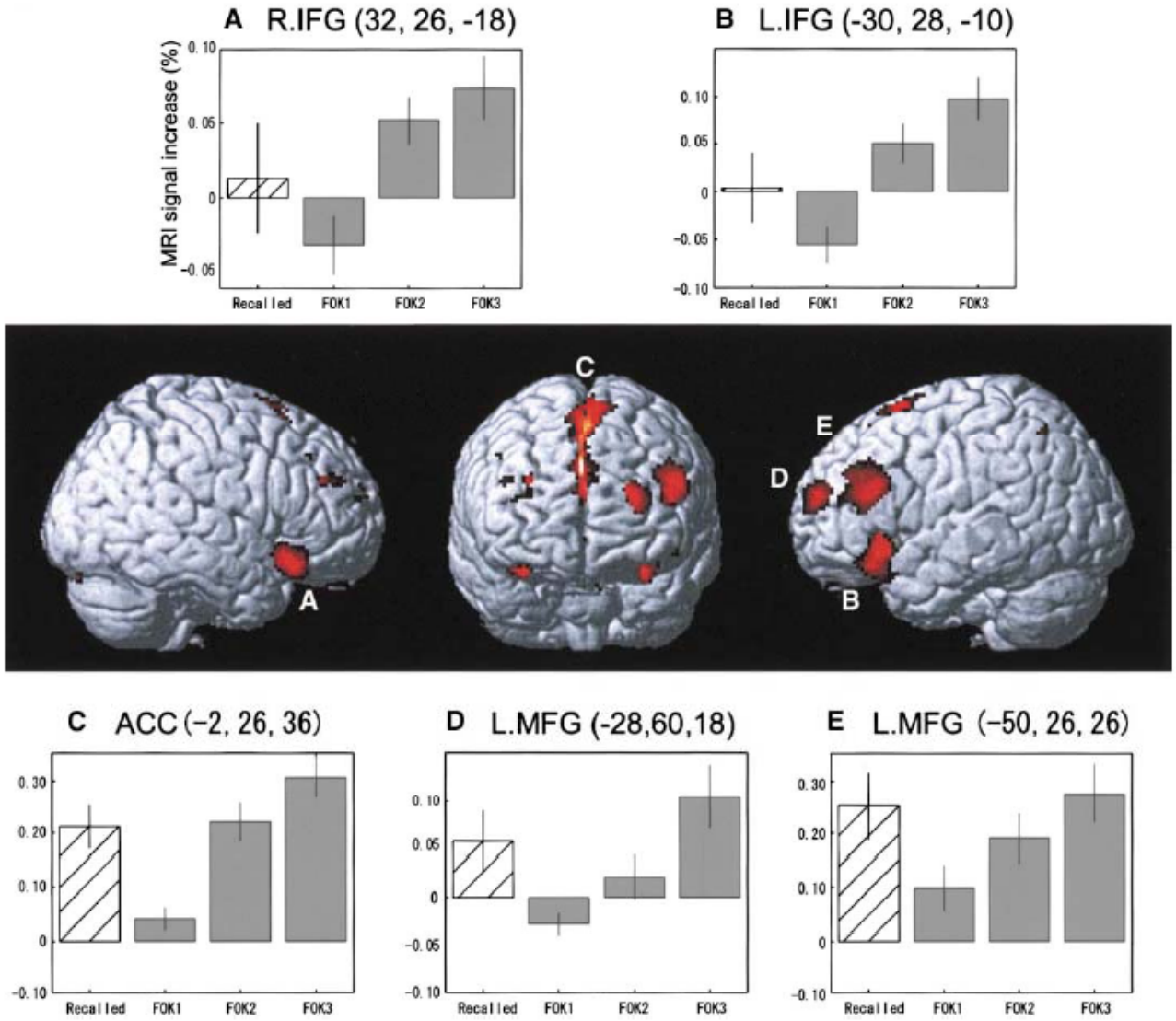
L'imagerie cérébrale du « sentiment de savoir »

Kikyo, H., Ohki, K., & Miyashita, Y. (2002).
Neural correlates for feeling-of-knowing: an fMRI parametric analysis. *Neuron, 36(1), 177-186.*

Etude en IRMf du sentiment de savoir (*feeling of knowing*) lorsqu'on ne sait pas répondre à une question.

Questions verbales, puis tri des essais selon qu'ils soient correctement rappelés, ou jugés avec un « sentiment de savoir » de niveau 1, 2 ou 3

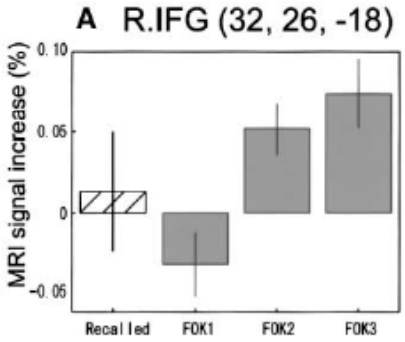
Un ensemble de régions, toutes préfrontales, augmentent leur activité en proportion du sentiment de savoir.

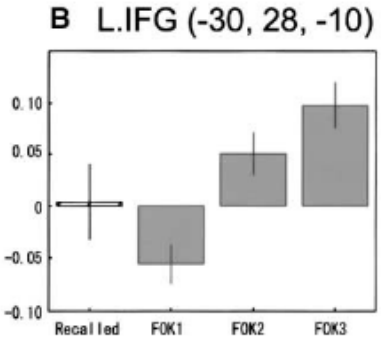


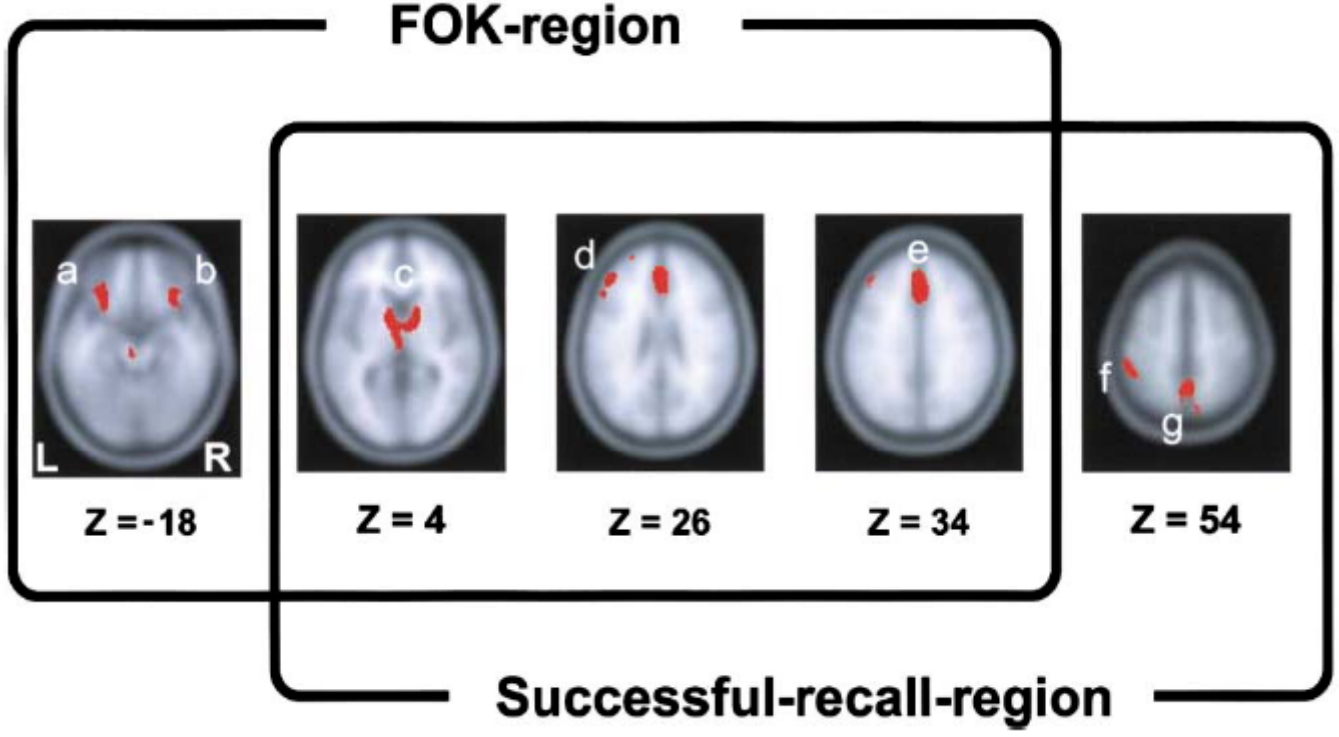
L’imagerie cérébrale du « sentiment de savoir »

Kikyo, H., Ohki, K., & Miyashita, Y. (2002).
Neural correlates for feeling-of-knowing: an fMRI parametric analysis. *Neuron, 36(1), 177-186.*

Certaines régions s’activent également lors de la récupération correcte en mémoire, mais d’autres semblent spécifiques du sentiment de savoir (IFG bilatéral)







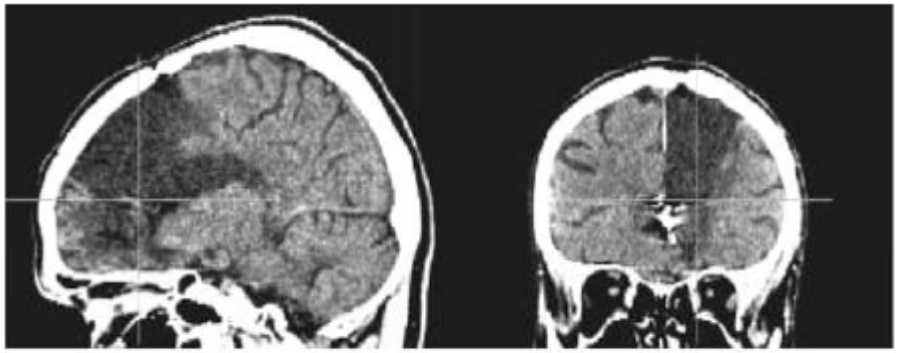
Dissociation entre l'effort et le sentiment de l'effort

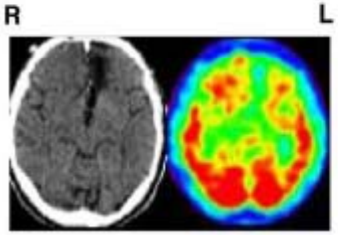
Naccache, L., Dehaene, S., Cohen, L., Habert, M. O., Guichart-Gomez, E., Galanaud, D., et al. (2005). Effortless control: executive attention and conscious feeling of mental effort are dissociable. *Neuropsychologia, 43(9), 1318-1328.*

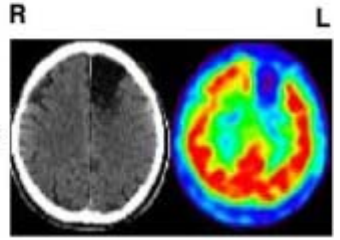
- Etude de cas unique d'une patiente atteinte d'une lésion ischémique massive de la région frontale mésiale gauche

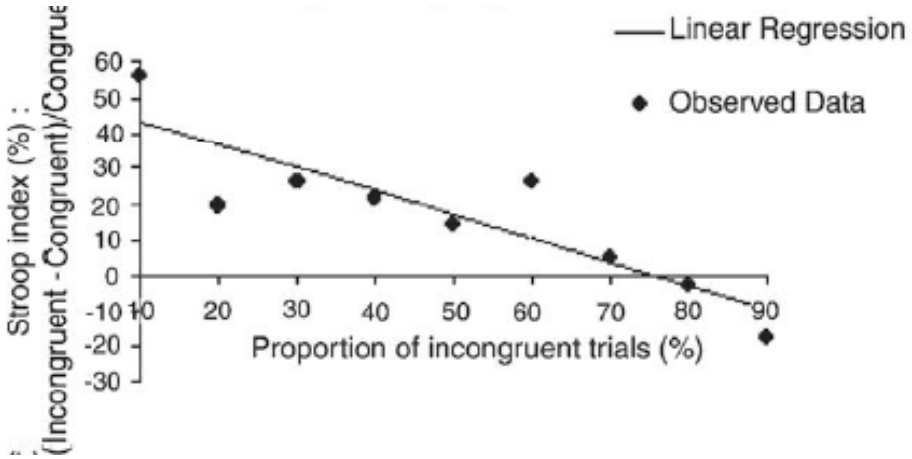
(avec possibilité d'un hypo-métabolisme des régions frontales inférieures)

- Dans une tâche de Stroop, la patiente reste capable de s'adapter à la difficulté de la tâche:
  - Elle ralentit après un essai difficile (non congruent)
  - La taille de l'effet Stroop varie avec la proportion d'essais non congruents
- Cependant elle est incapable de juger de son effort mental: tous les essais lui paraissent également faciles
  - Ses performances sont au niveau du hasard pour juger lequel de deux essais Stroop est le plus difficile
- Conclusion: déficit sélectif du « sentiment d'effort mental », avec préservation d'autres aspects de l'introspection (notamment la détection d'erreurs).



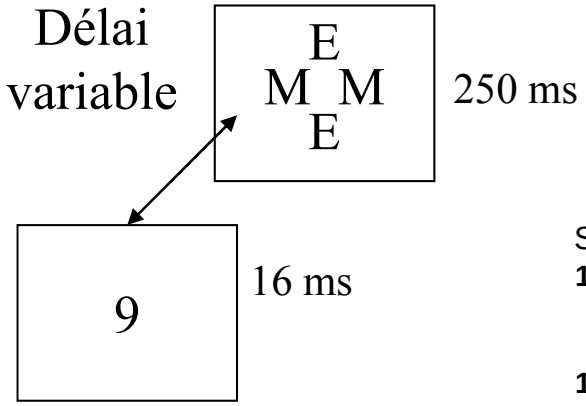






Le sentiment d'avoir vu est détérioré après lésion du cortex préfrontal

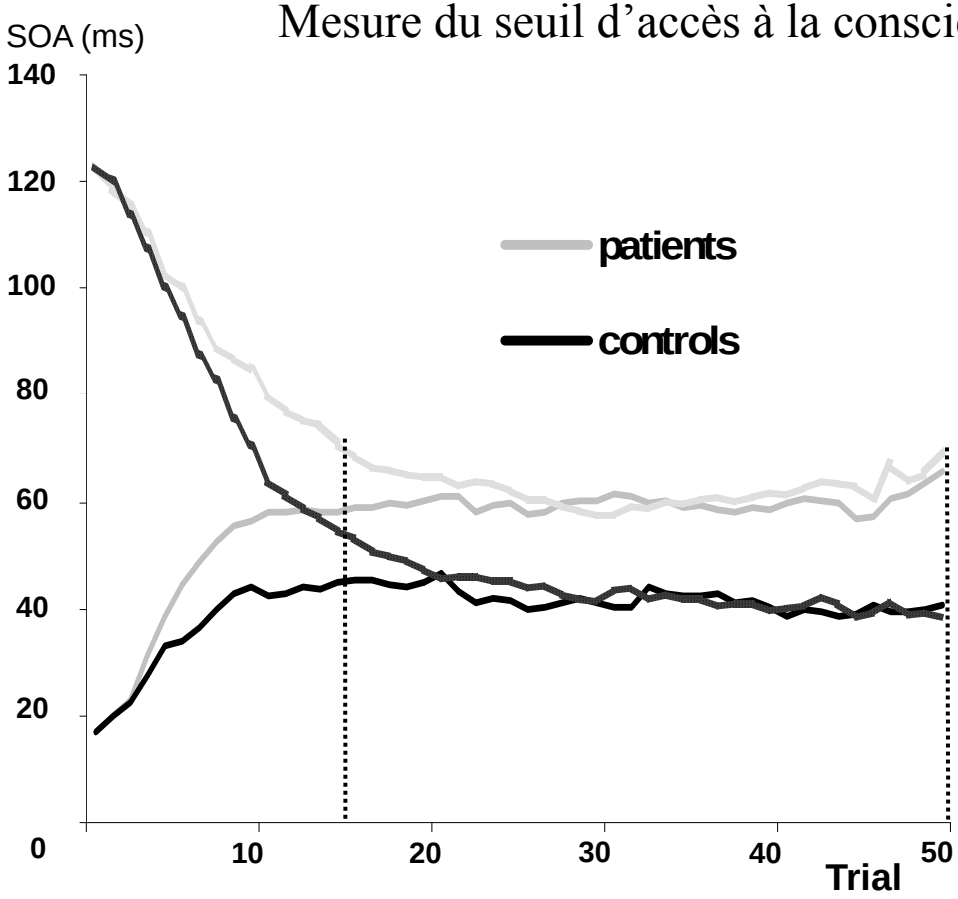
Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain, 132, 2531–2540.*



Présentation de chiffres masqués

Dans quatre conditions distinctes d'attention spatiale et temporelle

Mesure du seuil d'accès à la conscience



Le sentiment d'avoir vu est détérioré après lésion du cortex préfrontal

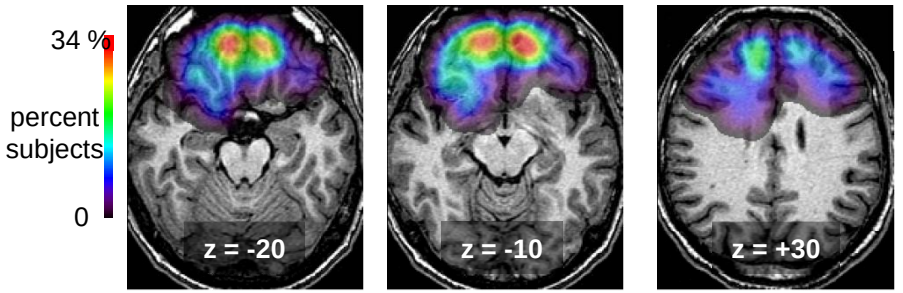
Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain, 132, 2531–2540.*

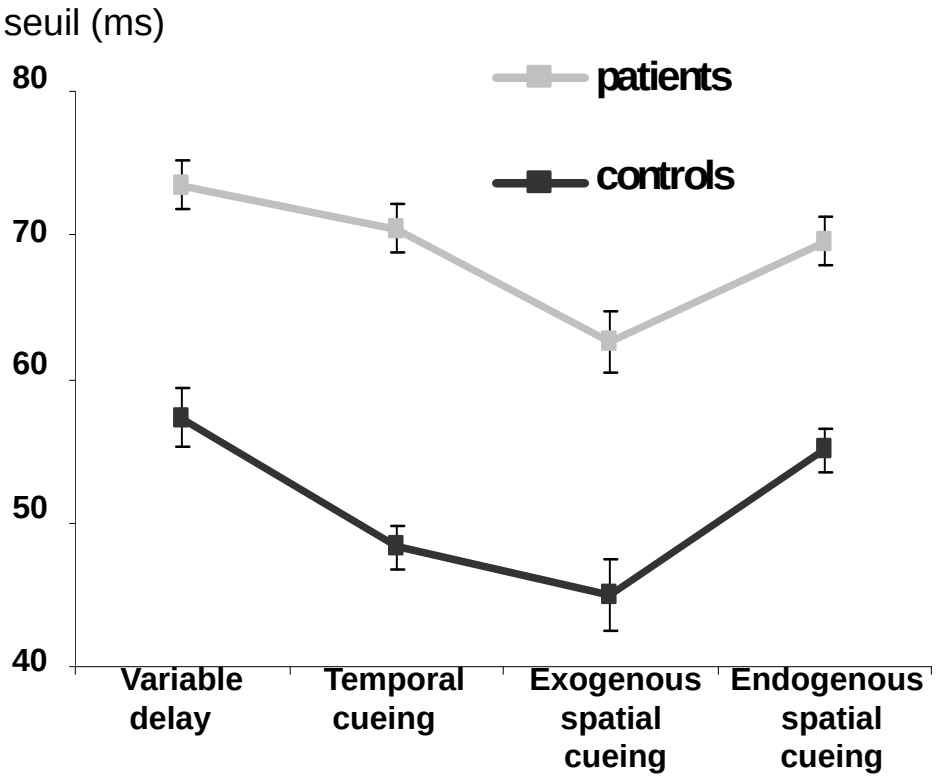
-Le seuil est effectivement plus élevé chez les patients que chez les contrôles

-L'effet est additif avec les effets de l'attention

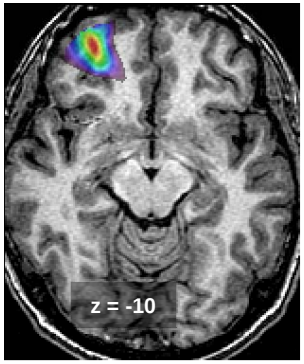
-Il corrèle particulièrement avec les lésions du cortex fronto-pôlaire

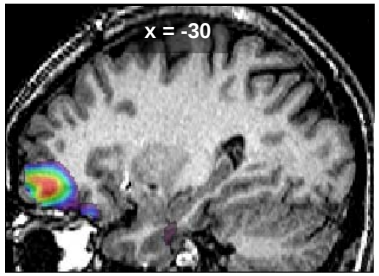
Distribution des lésions



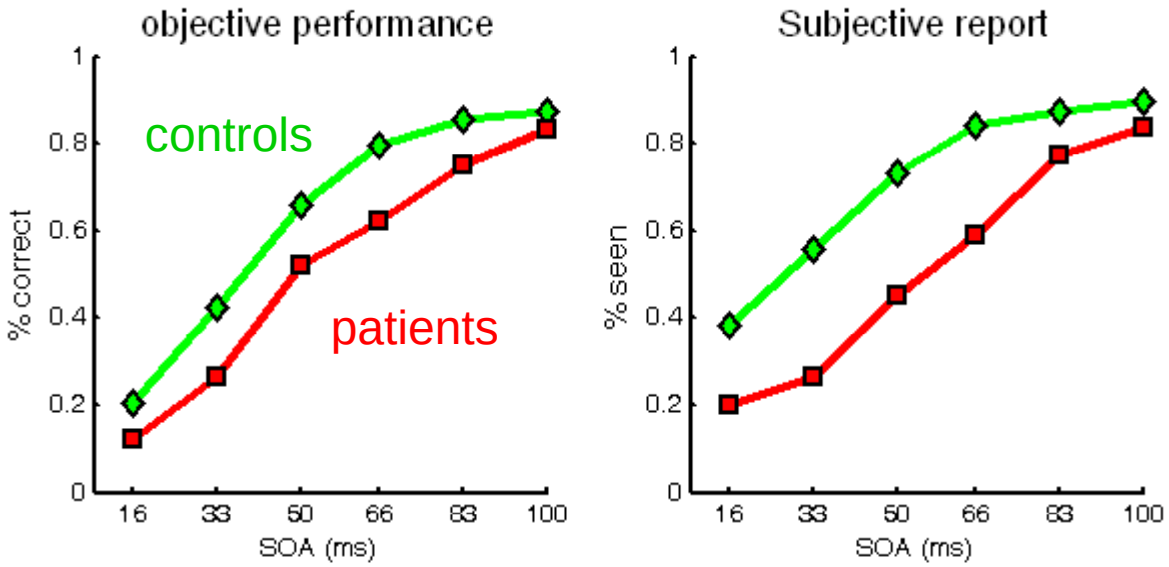


Pic de corrélation des lésions avec le seuil de masquage



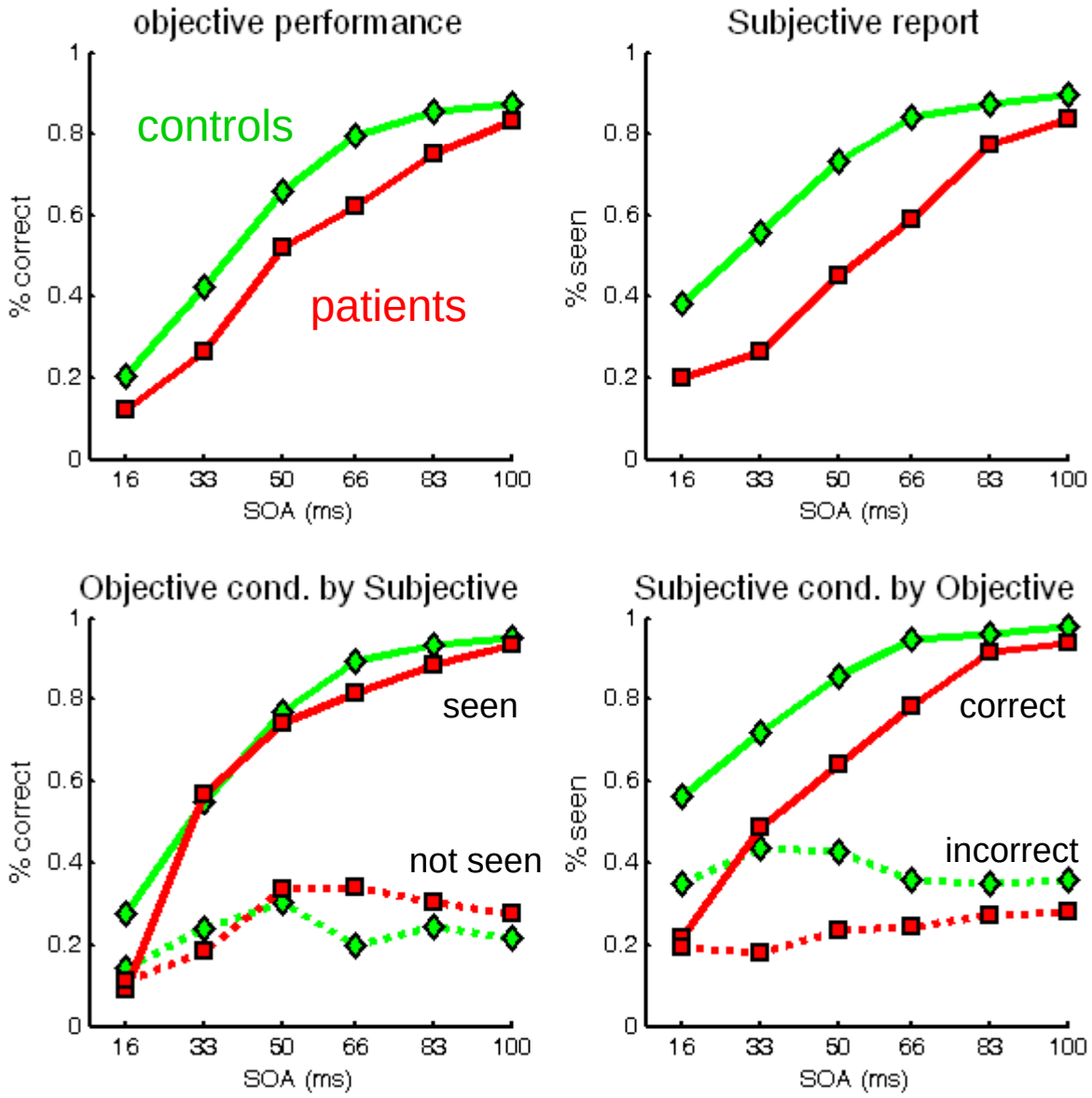


Dissociation partielle entre performance objective et rapport subjectif après lésion préfrontale



Les lésions frontales affectent les deux mesures

Dissociation partielle entre performance objective et rapport subjectif après lésion préfrontale



Les lésions frontales affectent les deux mesures

Toutefois, la performance objective est **normale** lorsque l'on trie les essais selon la performance subjective.

Inversement, à stimulus et performance identique, les patients conservent un déficit de rapport subjectif.

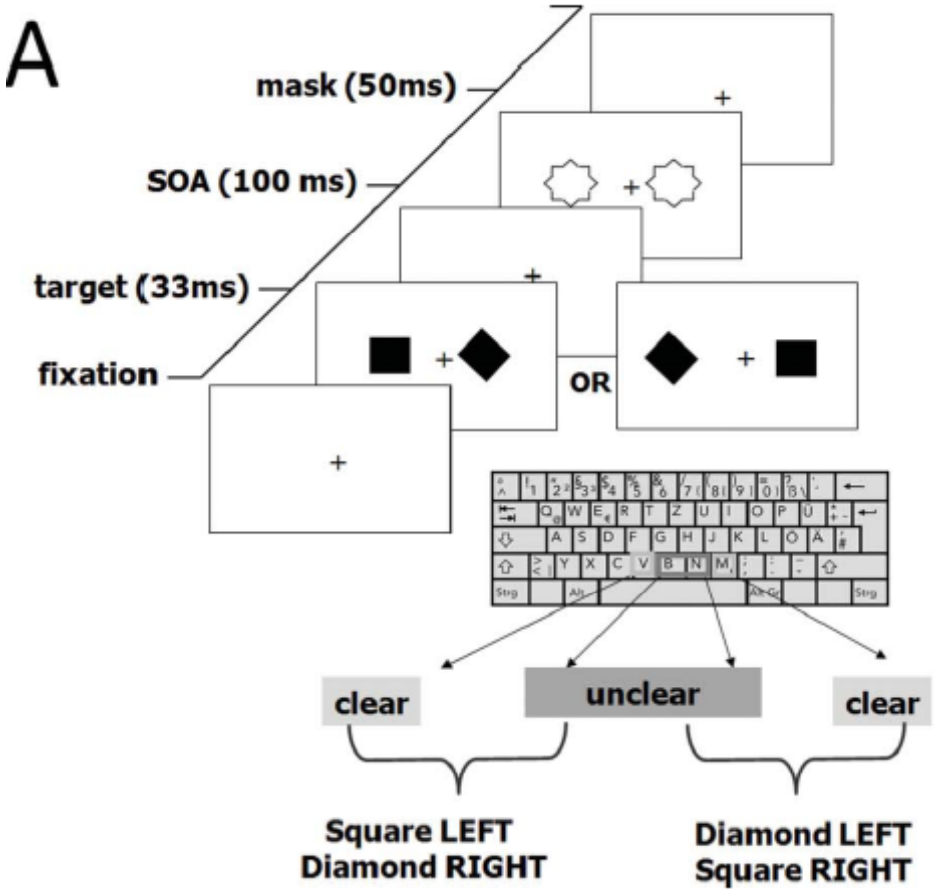
Dans de nombreux essais, ils nomment le chiffre correctement, mais disent ne pas le voir.

→ déficit métacognitif?

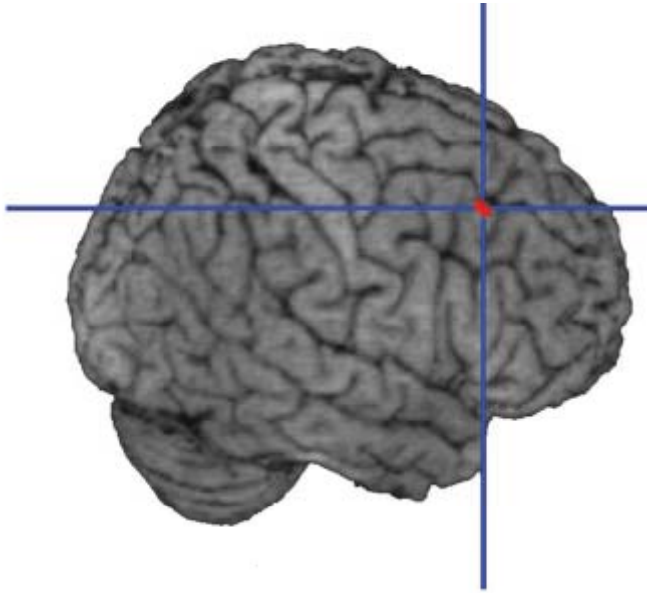
La stimulation magnétique du cortex préfrontal perturbe sélectivement les réponses métacognitives

Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience.*

Jugement simultané de l'identité d'une forme masquée (réponse de type I) et de la confiance dans ce jugement (réponse de type II)

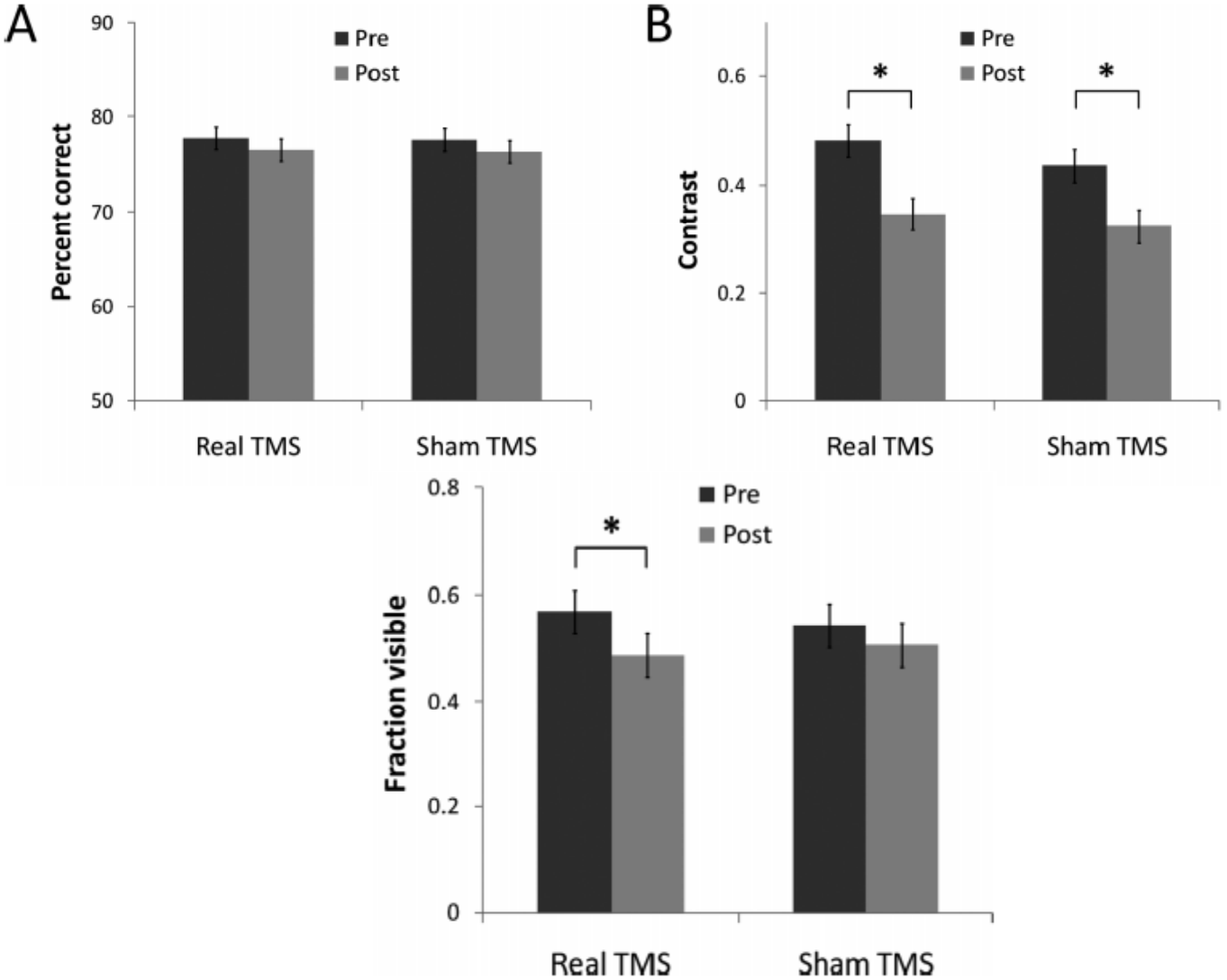


- Simulation magnétique transcrânienne en paquets dans la bande theta (ici 5 Hz), appliquée pendant 20 secondes successivement au cortex préfrontal dorsolatéral gauche puis droit, avec un intervalle d'une minute
- Cette simulation induit une inhibition qui peut durer jusqu'à 20 minutes.



La stimulation magnétique du cortex préfrontal perturbe sélectivement les réponses métacognitives

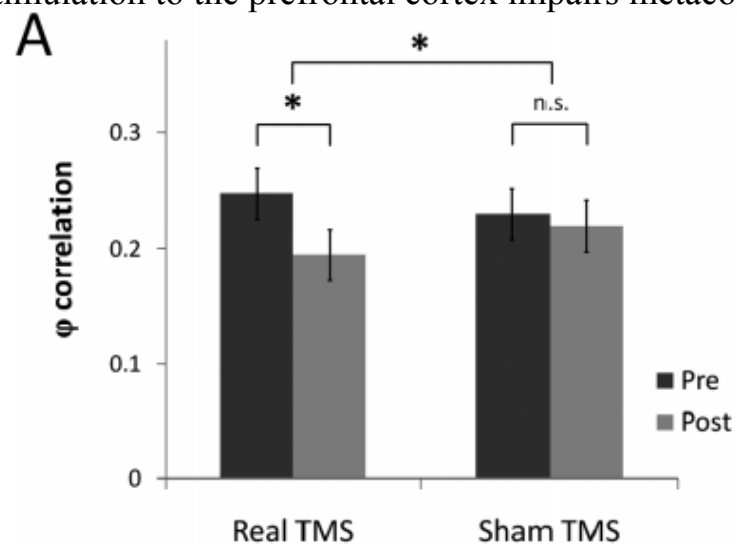
Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience.*

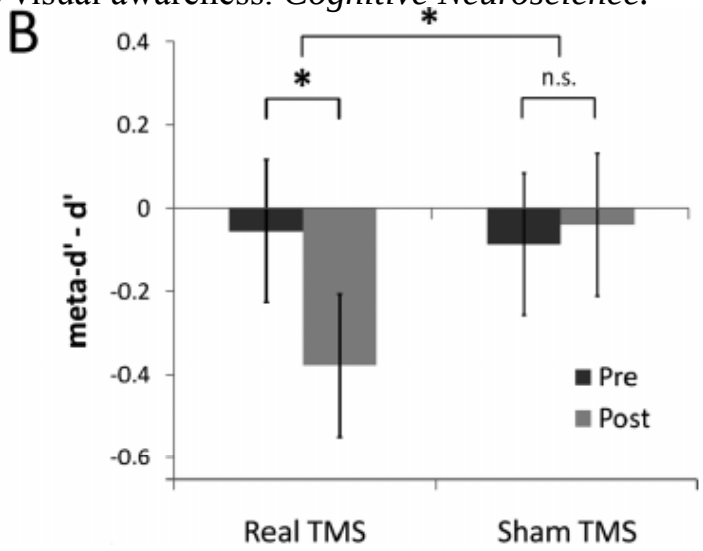


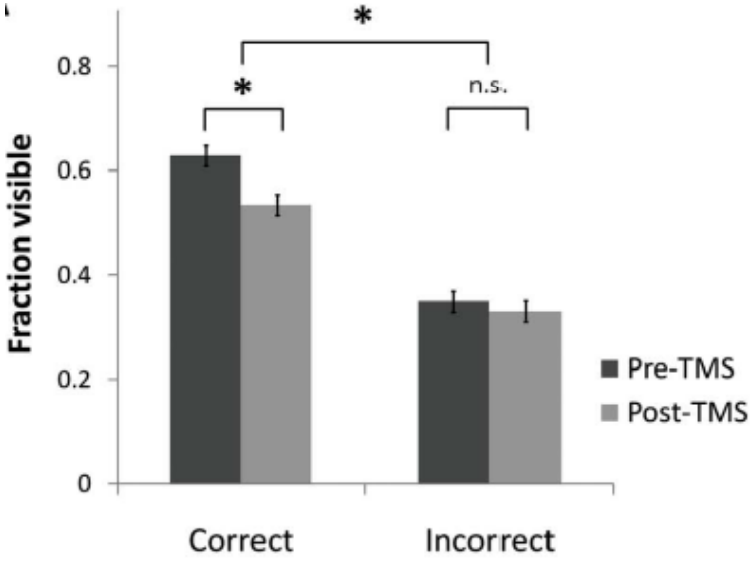
- Le contraste est ajusté en permanence pour que la performance objective approche 75% de réussite.
- La performance objective et le contraste présenté sont les mêmes, que la TMS soit ou ne soit pas appliquée.
- Par contre, la visibilité subjective des stimuli diminue, uniquement dans la condition de vraie TMS.

La stimulation magnétique du cortex préfrontal perturbe sélectivement les réponses métacognitives

Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience.*





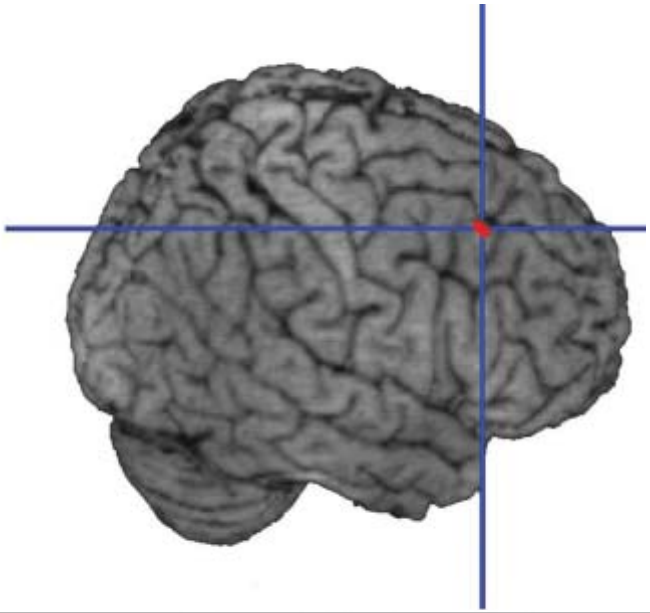


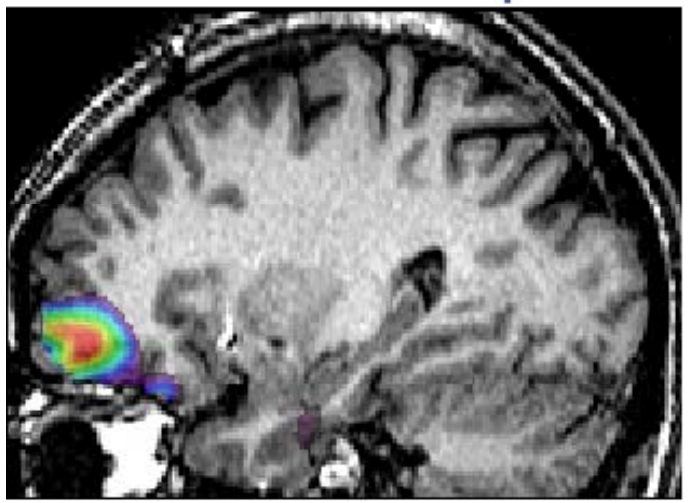
- Crucialement, la corrélation entre le jugement de type II et la réussite de type I diminue après TMS
- Normalement, la performance de type II est « alignée » sur la performance de type I (meta-d’ = d’, ce que signifie que les participants extraient exactement du stimulus autant d’information métacognitive que d’information sensorielle)
- Cependant après TMS, la performance métacognitive diminue sélectivement (méta-d’<d’)
- Ce déficit affecte particulièrement les réponses correctes.

Conclusion intermédiaire:
L'intégrité du cortex préfrontal semble indispensable aux jugements métacognitifs

Le cortex préfrontal contribue au jugement métacognitif de visibilité
Sa lésion ou son inactivation temporaire conduisent à une réduction des compétences métacognitives subjectives, sans changement de performance primaire (ou avec un changement plus modeste dans l'étude de Del Cul et al).

Au moins deux régions distinctes pourraient être impliquées:
-Le cortex préfrontal dorsolatéral (Rounis et al.)
-Le cortex préfrontal rostral, aire 10 de Brodmann (Del Cul et al.).

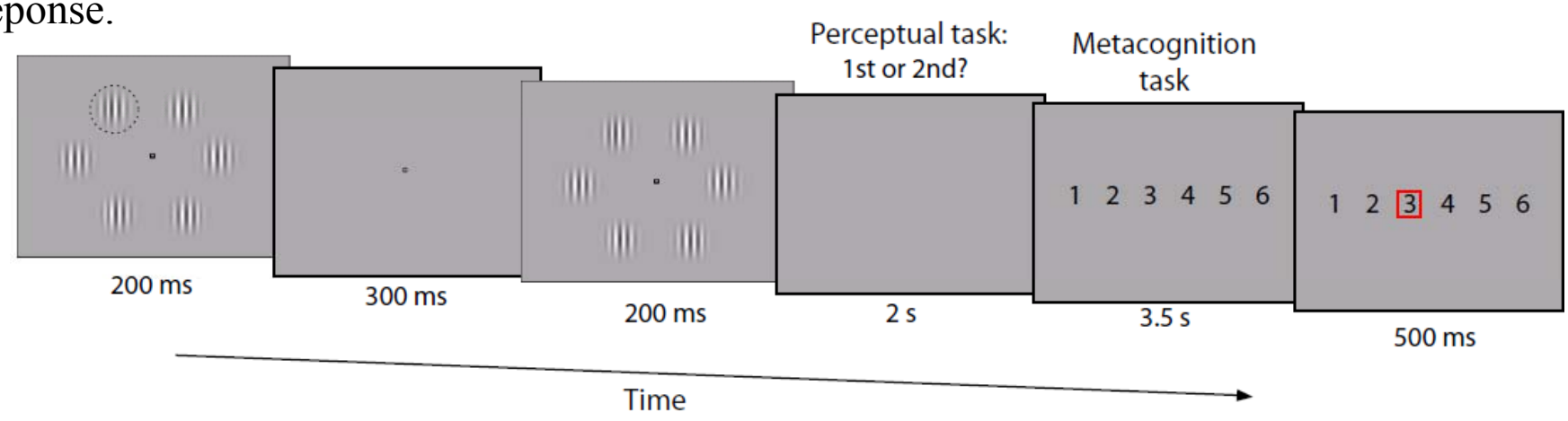




Les bases cérébrales des différences individuelles dans la capacité d'introspection

Fleming et al., *Science* 2010

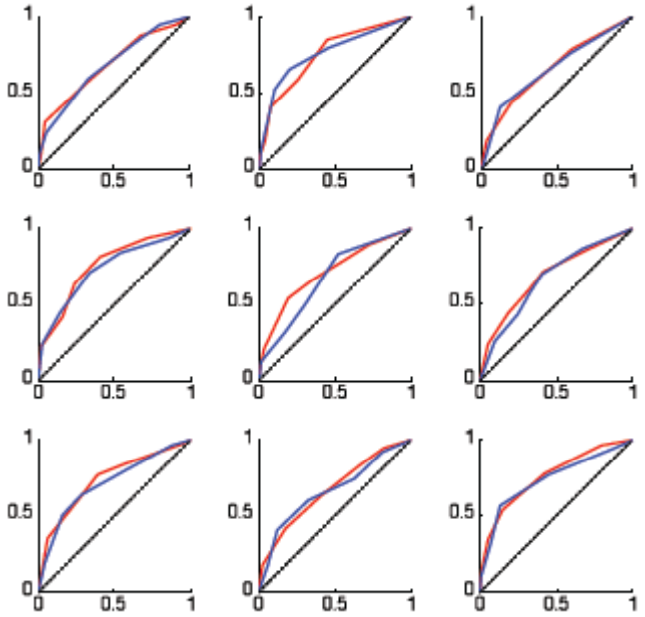
La capacité de porter un jugement métacognitif de second-ordre (qui porte sur la véracité d'un jugement de premier ordre) varie d'une personne à l'autre.
Elle est détériorée chez les patients atteints de lésions du lobe frontal, particulièrement dans sa partie très antérieure (rostrale).
Corrèle-t-elle avec des changements plus subtils d'organisation cérébrale chez le sujet normal?
Ici, les sujets sont engagés dans une tâche psychophysique difficile (détection d'un patch de contraste légèrement plus élevé, dans le premier ou dans le second écran).
Cette tâche est maintenue proche du seuil de manière à obtenir 71% de réussite chez tous les sujets.
Après chaque essai, les participants donnent leur degré de confiance dans leur première réponse.



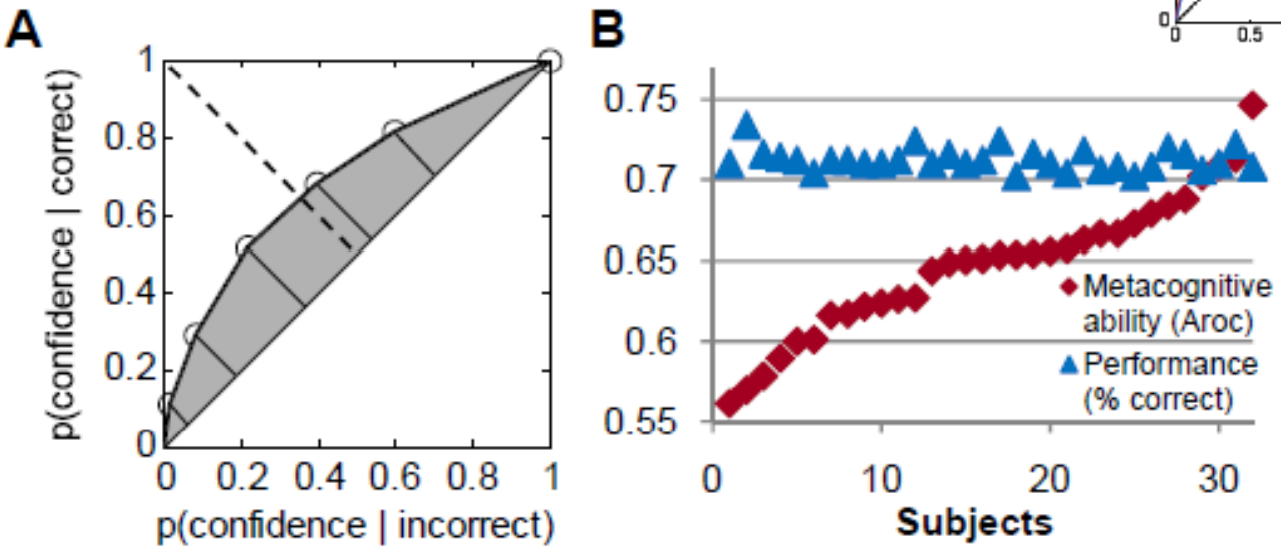
Les bases cérébrales des différences individuelles dans la capacité d'introspection

Fleming et al., *Science* 2010

Les résultats du jugement de confiance peuvent être bien modélisés par la théorie de la détection du signal de second ordre.
L'aire sous la courbe est une bonne mesure de la compétence métacognitive des sujets (toutes les mesures sont positives, ce qui indique une introspection non-négligeable).
L'aire est stable si l'on analyse séparément deux moitiés indépendantes de l'expérience.



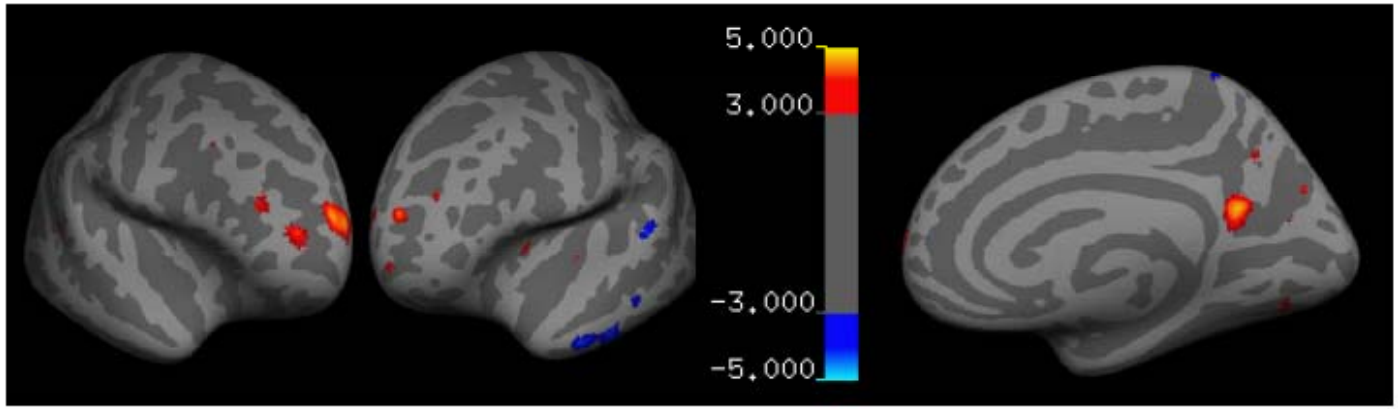
Exemple chez 9 sujets

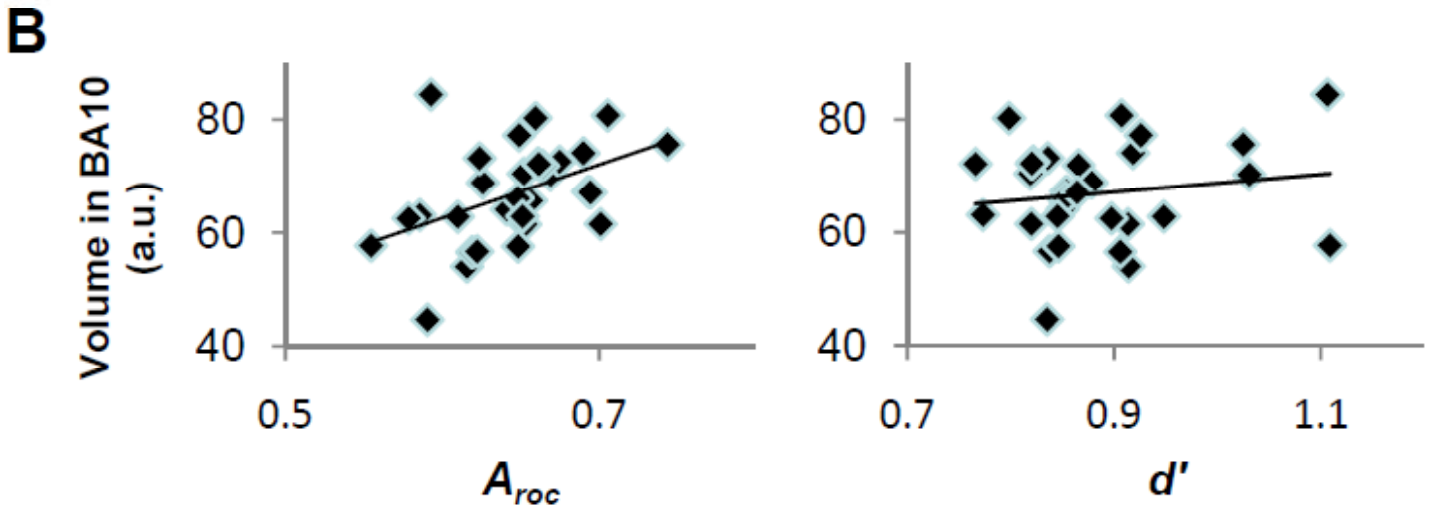


Les bases cérébrales des différences individuelles dans la capacité d'introspection

Fleming et al., *Science* 2010

- La compétence métacognitive est corrélée avec la quantité de matière grise des participants dans le cortex préfrontal rostral droit, aire de Brodmann 10.
- Pas de corrélation avec les résultats de la tâche primaire.

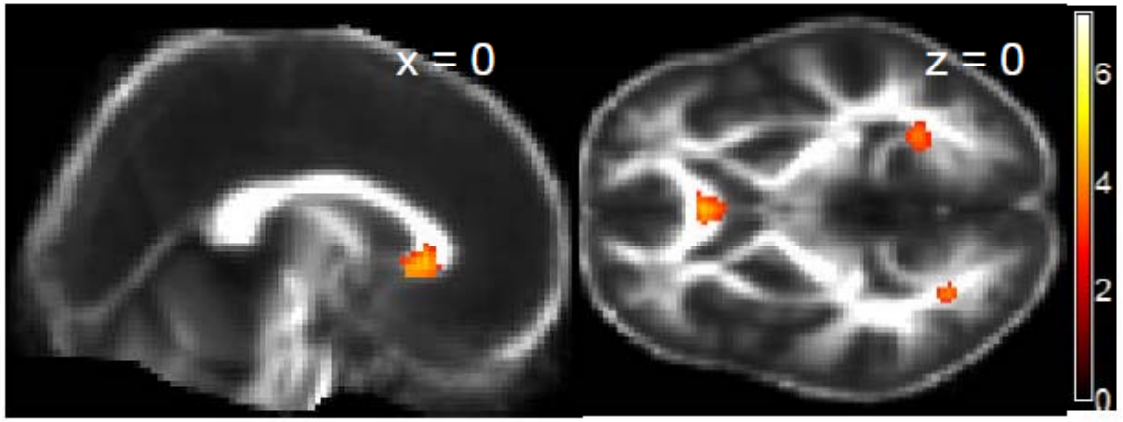


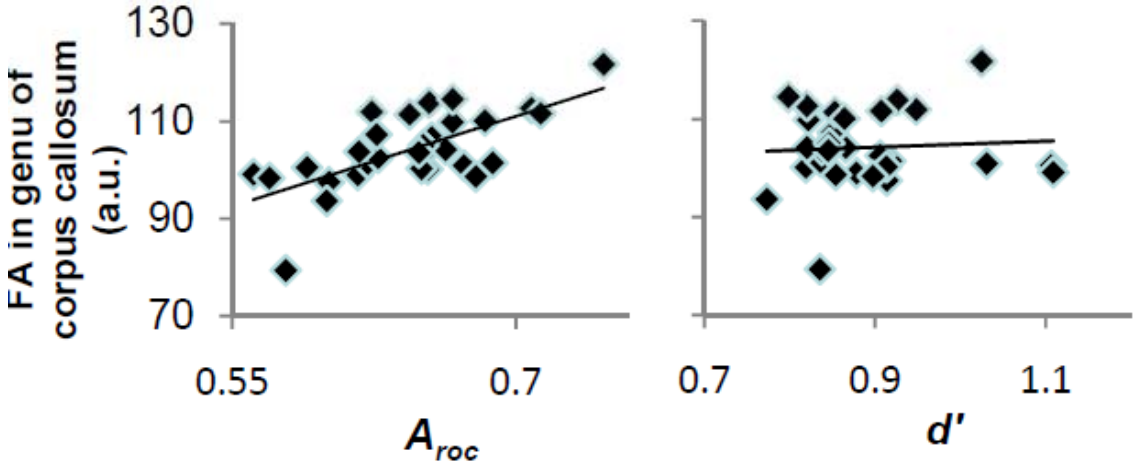


Les bases cérébrales des différences individuelles dans la capacité d'introspection

Fleming et al., *Science* 2010

La compétence métacognitive est également corrélée avec la densité de matière blanche des participants dans la région rostrale du corps calleux – là encore, sans corrélation avec la performance de premier ordre.



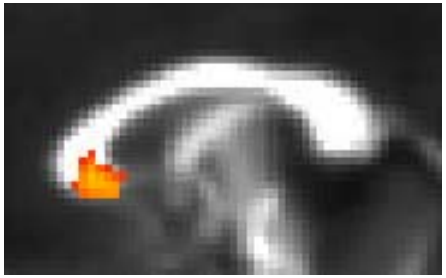


Les bases cérébrales des différences individuelles dans la capacité d'introspection

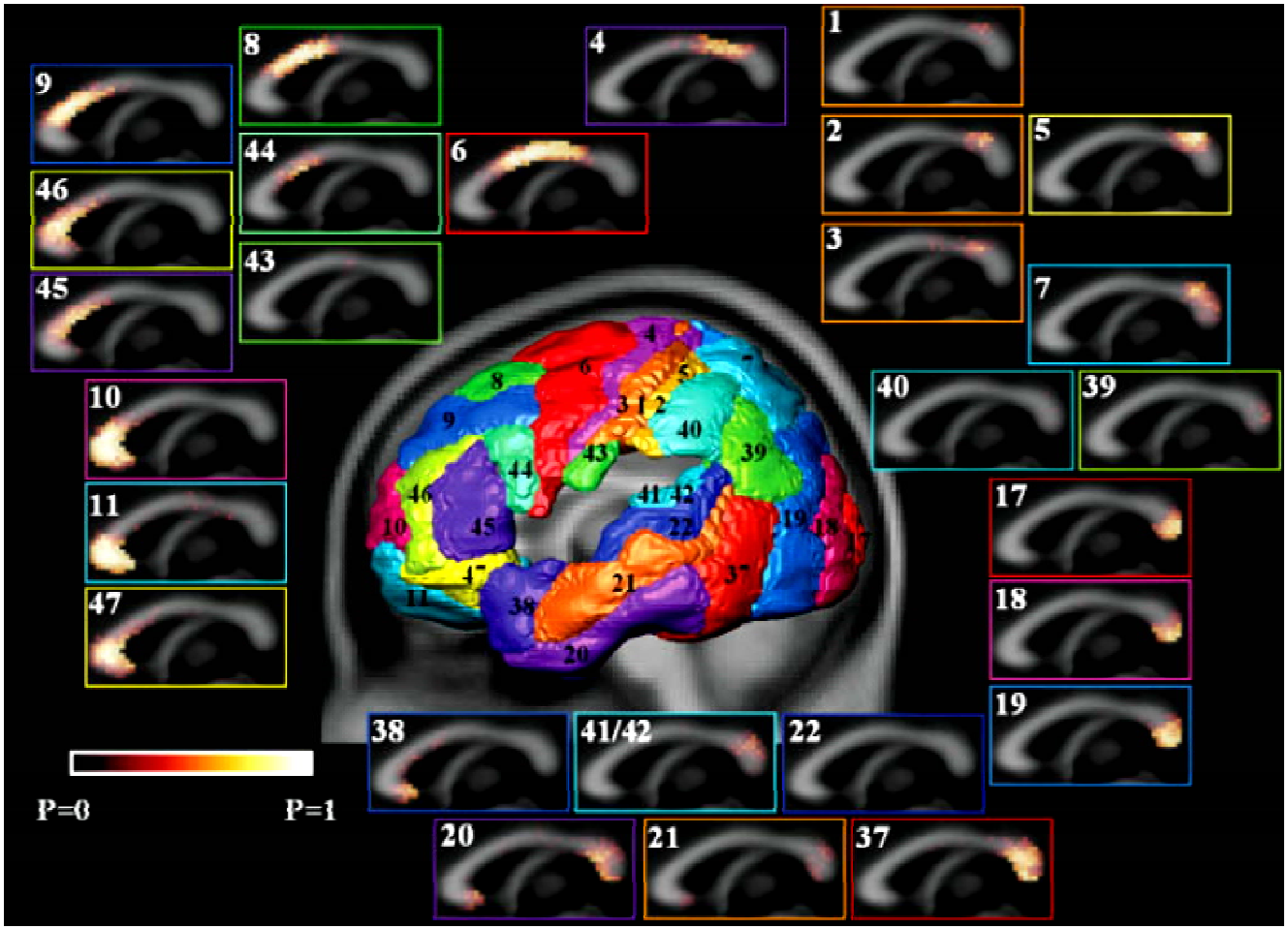
Fleming et al., *Science* 2010

Cette région du corps calleux connecte précisément les aires frontales rostrales, et notamment l'aire 10 de Brodmann (image de Chao et al, *Human Brain Mapping*, 2009)

image de Fleming et al.
(retournée)



antérieur postérieur



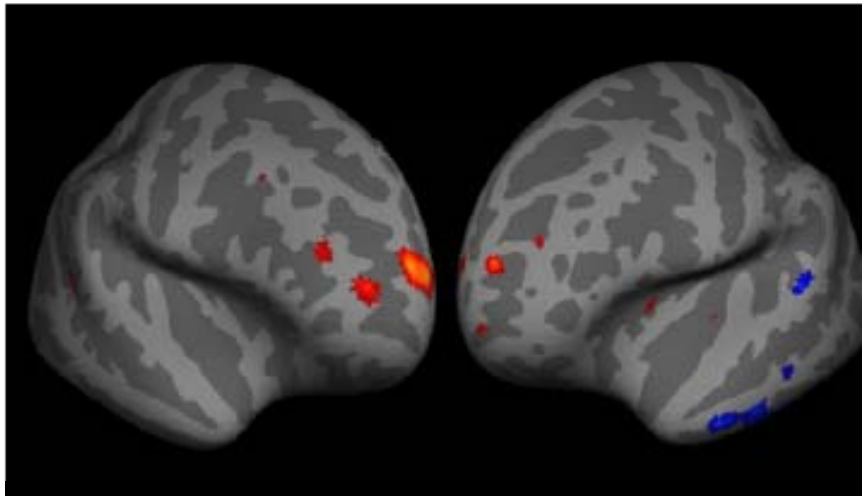
Les bases cérébrales des différences individuelles dans la capacité d'introspection

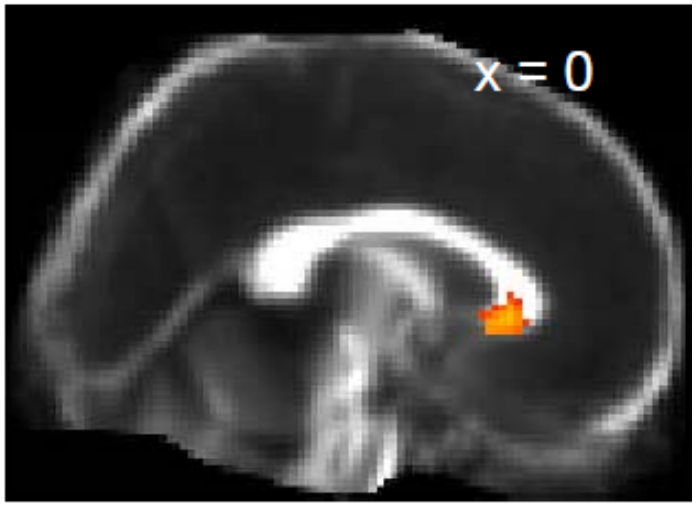
Fleming et al., *Science* 2010

Le cortex préfrontal rostral apparaît donc comme l'une des régions clés de l'introspection et de la métacognition.

Les auteurs soulignent que leurs résultats sont compatibles avec deux interprétations: différences anatomiques innées, ou bien entraînement différentiel de l'introspection.

Cette région se situe au plus haut niveau d'un système préfrontal hiérarchique (cf. travaux d'Etienne Koechlin)





Expansion de l'aire 10 dans l'espèce humaine

Semendeferi, K., Armstrong, E., Schleicher, A., Zilles, K., & Van Hoesen, G. W. (2001). Prefrontal cortex in humans and apes: a comparative study of area 10. *Am J Phys Anthropol, 114(3), 224-241.*

- L'aire 10 est une région du cortex préfrontal dont l'expansion différencie tout particulièrement l'espèce humaine des autres espèces de primates.
- Dans les couches supragranulaires, la densité cellulaire est moindre, et en conséquence plus d'espace est disponible pour la connectivité à longue distance.

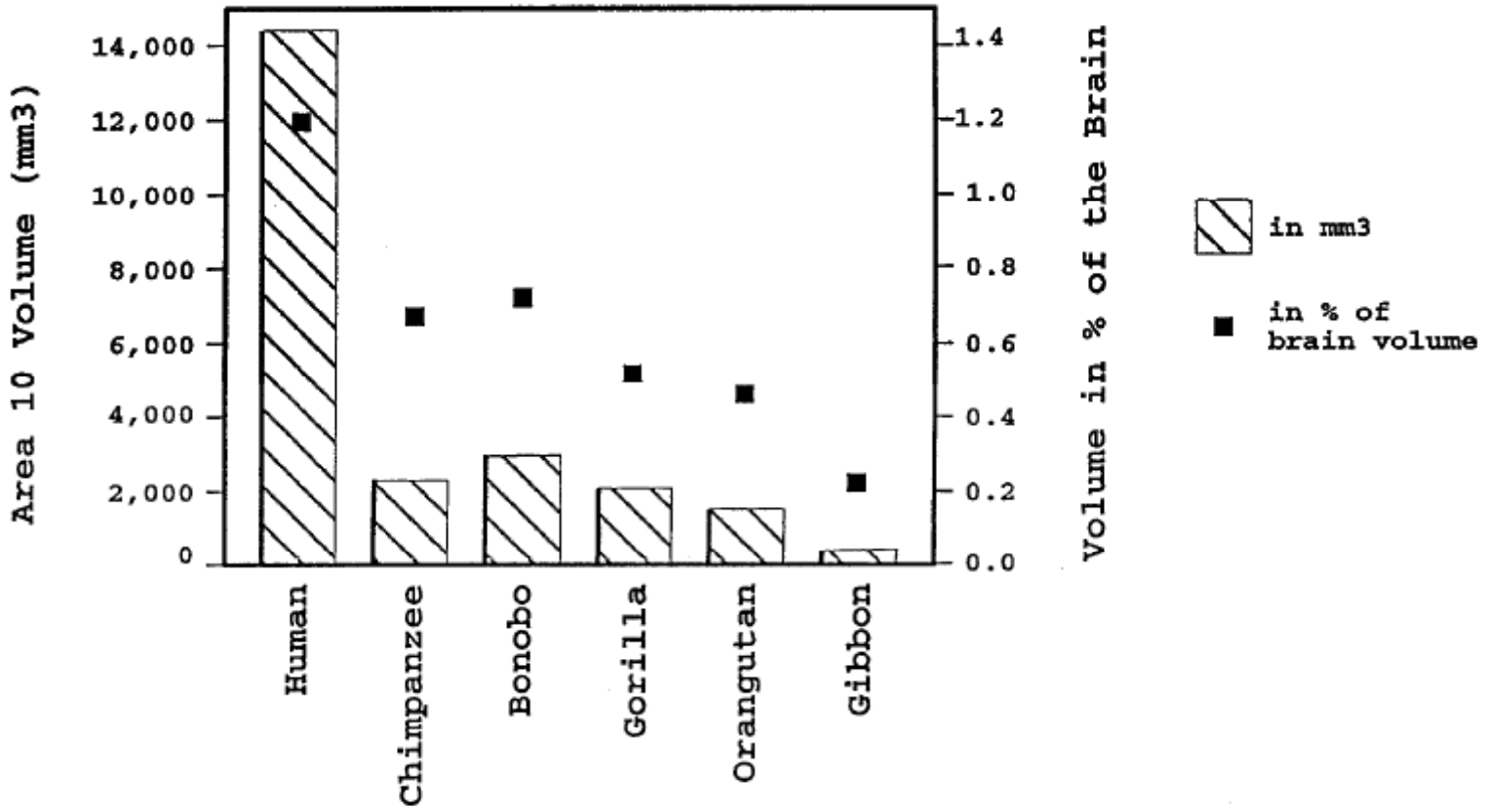


Fig. 7. Absolute and relative size of area 10 in right hemisphere. Columns represent absolute values in cubic millimeters, and squares represent relative values in percentage of total brain volume.

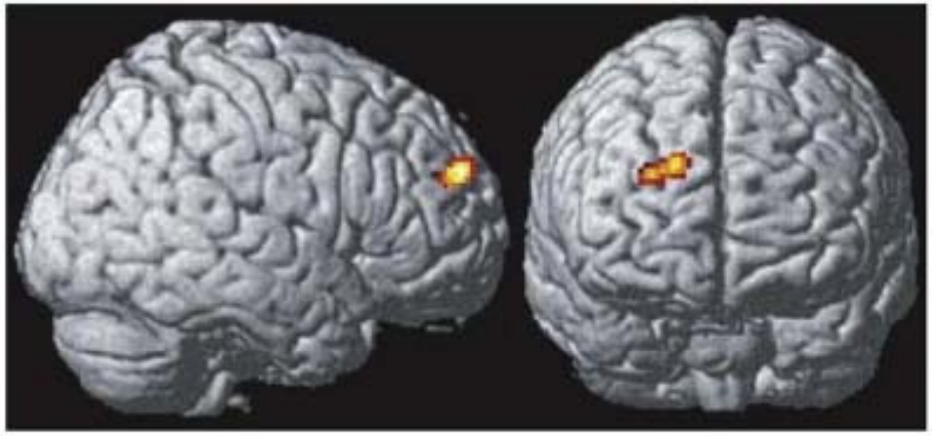
Les variations inter-individuelles de l'introspection corrèlent avec l'activation du cortex fronto-pôlaire droit

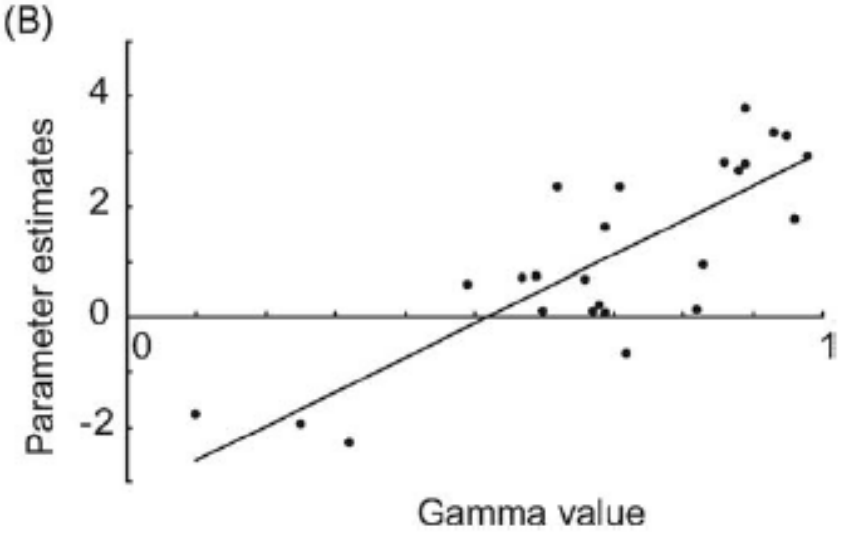
Yokoyama, O., Miura, N., Watanabe, J., Takemoto, A., Uchida, S., Sugiura, M., et al. (2010). Right frontopolar cortex activity correlates with reliability of retrospective rating of confidence in short-term recognition memory performance. *Neurosci Res, 68(3), 199-206.*

Tâche de mémoire à court terme suivie à chaque essai d'un jugement de confiance dans la réponse qui vient d'être faite.

Les participants sont classifiés en fonction de la corrélation de leur introspection avec leur réussite objective.

Ce score (gamma) corrèle fortement avec l'activité du cortex fronto-polaire droit en IRMf durant l'exécution de la tâche.





Une étude en IRM fonctionnelle de la décision et de la confiance qu'on peut lui accorder

Rolls, E. T., Grabenhorst, F., & Deco, G. (2010). Choice, difficulty, and confidence in the brain. *Neuroimage, 53(2), 694-706.*

Simulation de la prise de décision dans un réseau de neurones (travaux de Wong & Wang, *J Neuroscience* 2006)

Deux populations de neurones en compétition

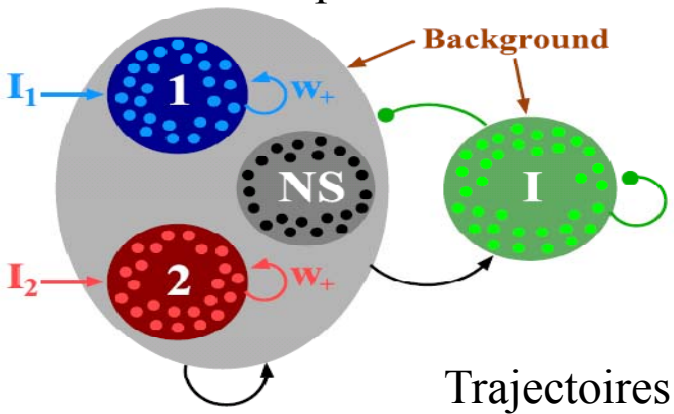
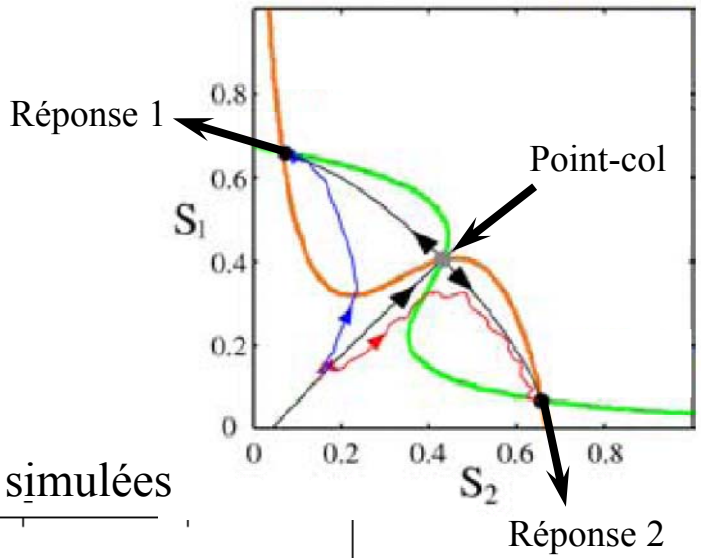
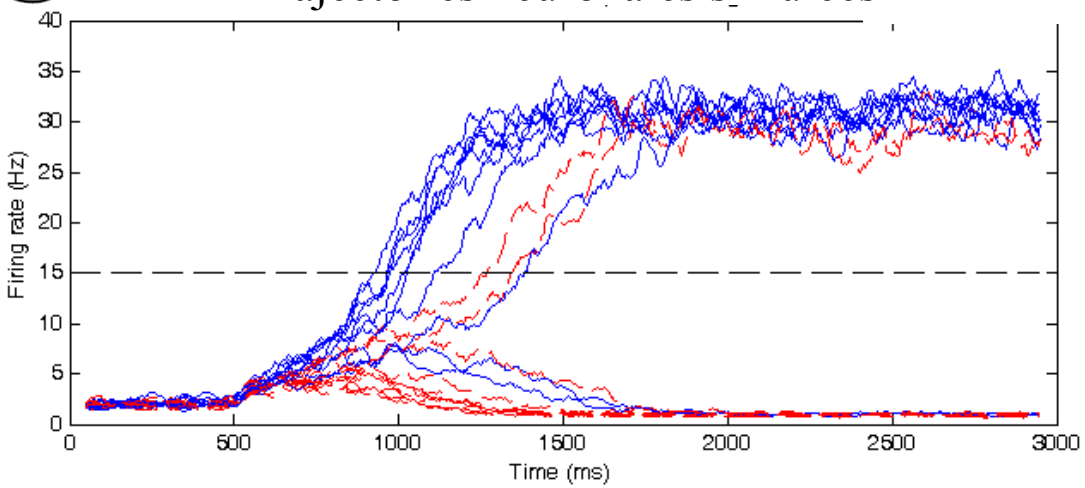


Diagramme de phase à deux points fixes



Trajectoires neuronales simulées



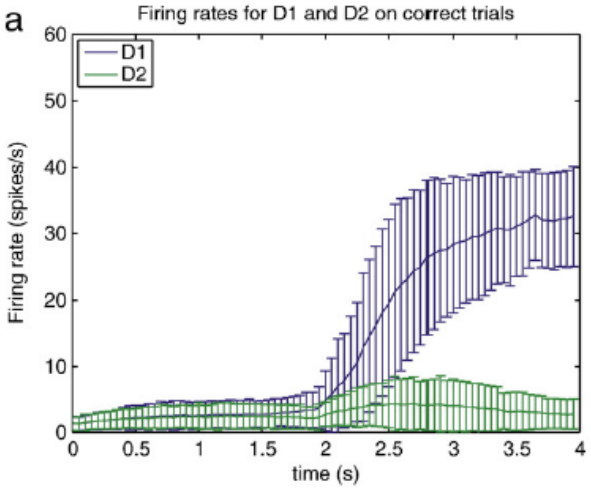
Réponses correctes

Erreurs

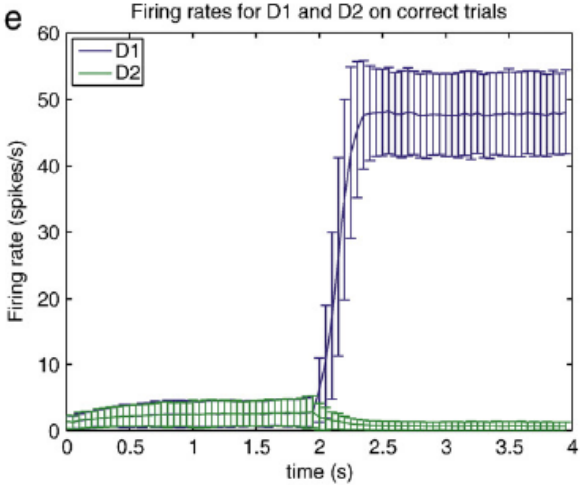
Une étude en IRM fonctionnelle de la décision et de la confiance qu'on peut lui accorder

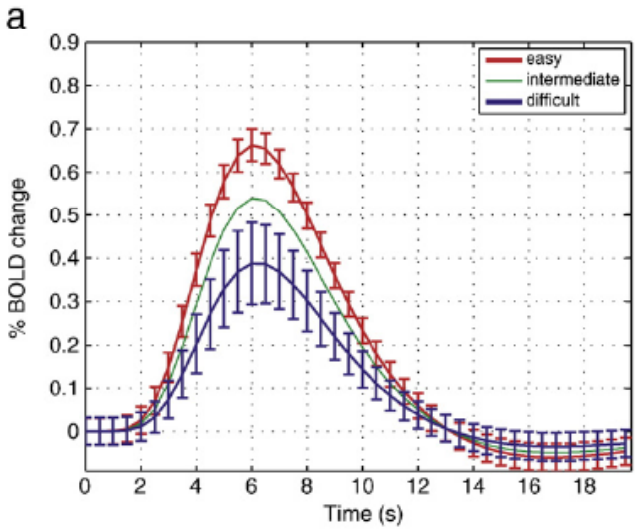
Rolls, E. T., Grabenhorst, F., & Deco, G. (2010). Choice, difficulty, and confidence in the brain. *Neuroimage, 53(2), 694-706.*

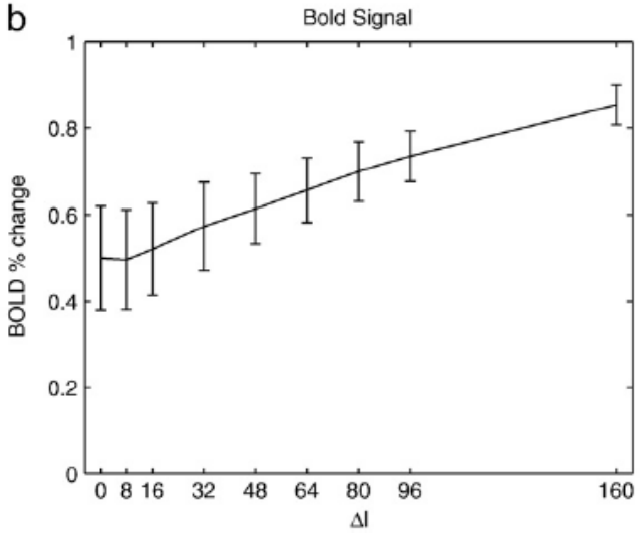
Essais difficiles



Essais faciles





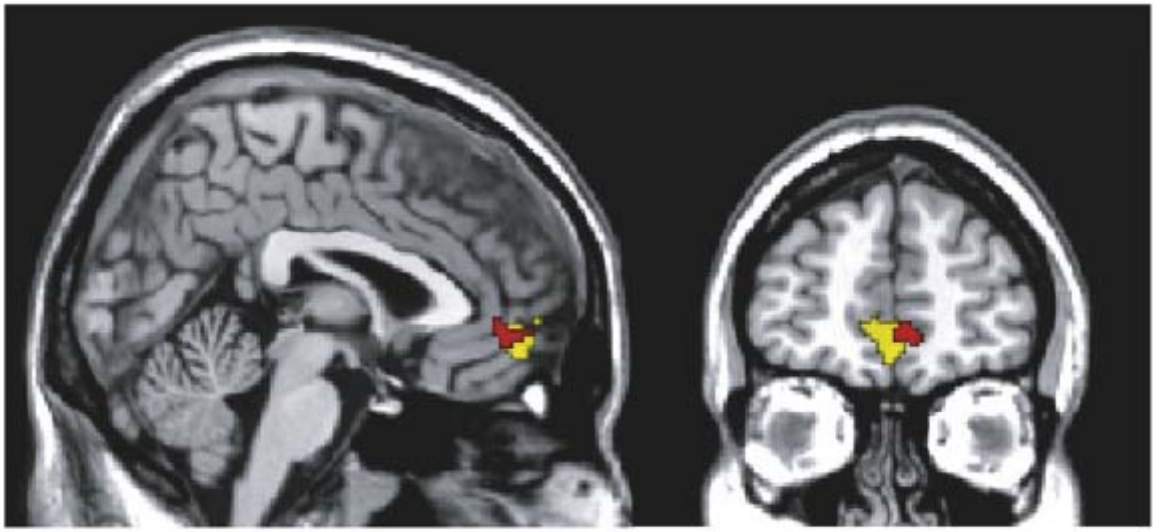


La simulation du réseau de neurones montre que l'activité neuronale doit monter plus vite, en moyenne, lorsque l'évidence sensorielle (ΔI) est plus grande.

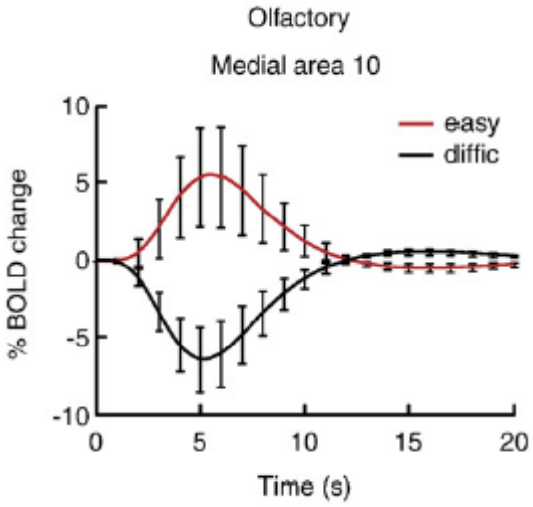
En IRM fonctionnelle, en moyenne à travers les deux populations de neurones, cela devrait se traduire par une activation dont l'amplitude croit linéairement avec ΔI

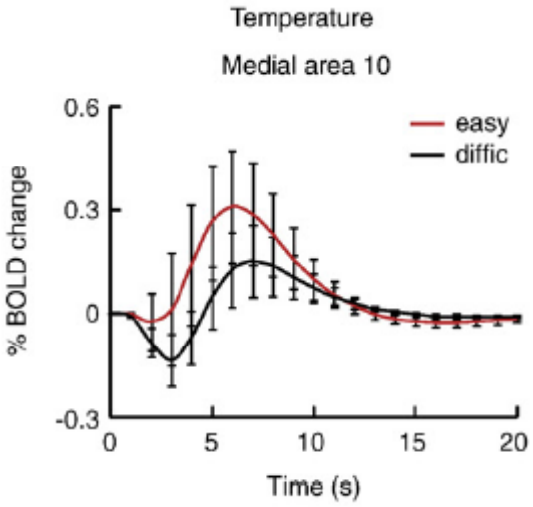
Une étude en IRM fonctionnelle de la décision et de la confiance qu'on peut lui accorder

Rolls, E. T., Grabenhorst, F., & Deco, G. (2010). Choice, difficulty, and confidence in the brain. *Neuroimage, 53(2), 694-706.*



Medial area 10





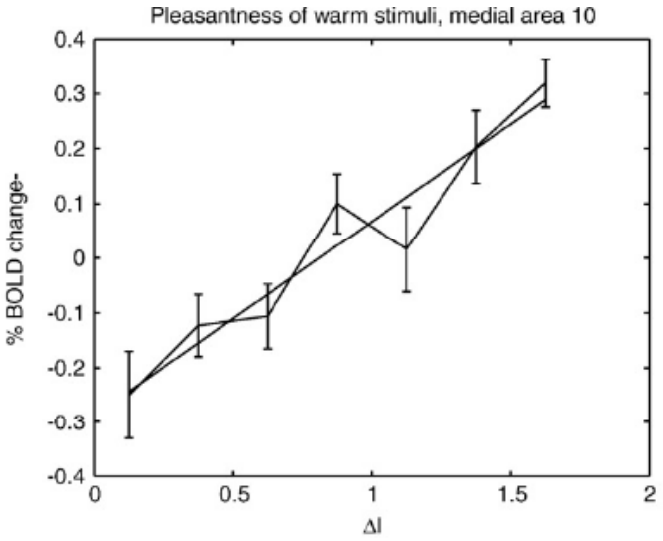
Données empiriques:

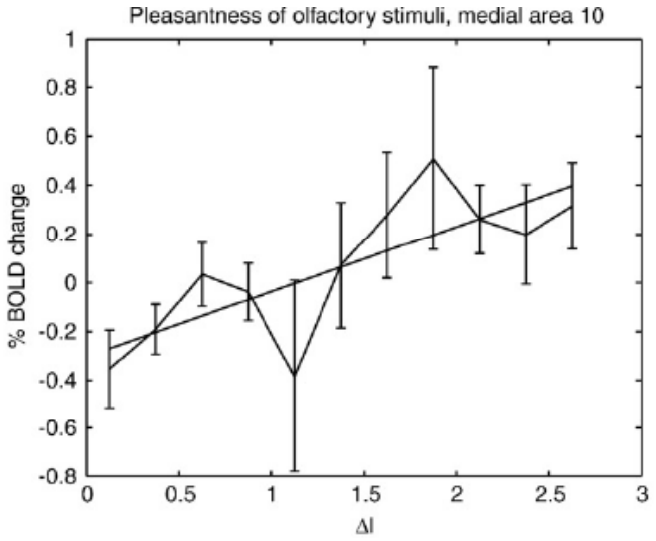
Deux tâches psychophysiques (juger laquelle de deux odeurs, ou laquelle de deux températures, est la plus agréable)

L'IRMf montre une activation du cortex préfrontal rostromésial (±partie mésiale de l'aire 10 de Brodmann), plus forte lors des essais faciles que lors des essais difficiles.

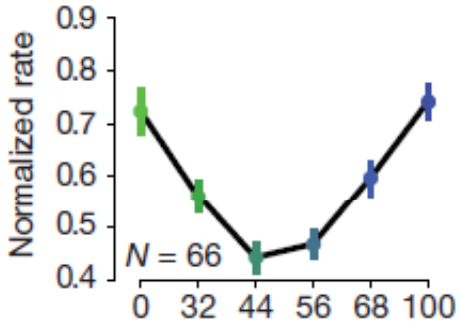
Une étude en IRM fonctionnelle de la décision et de la confiance qu'on peut lui accorder

Rolls, E. T., Grabenhorst, F., & Deco, G. (2010). Choice, difficulty, and confidence in the brain. *Neuroimage, 53(2), 694-706.*





Rappel des données de Kepecs et al (2008) : certains neurones orbitofrontaux déchargent en proportion de la facilité de la tâche

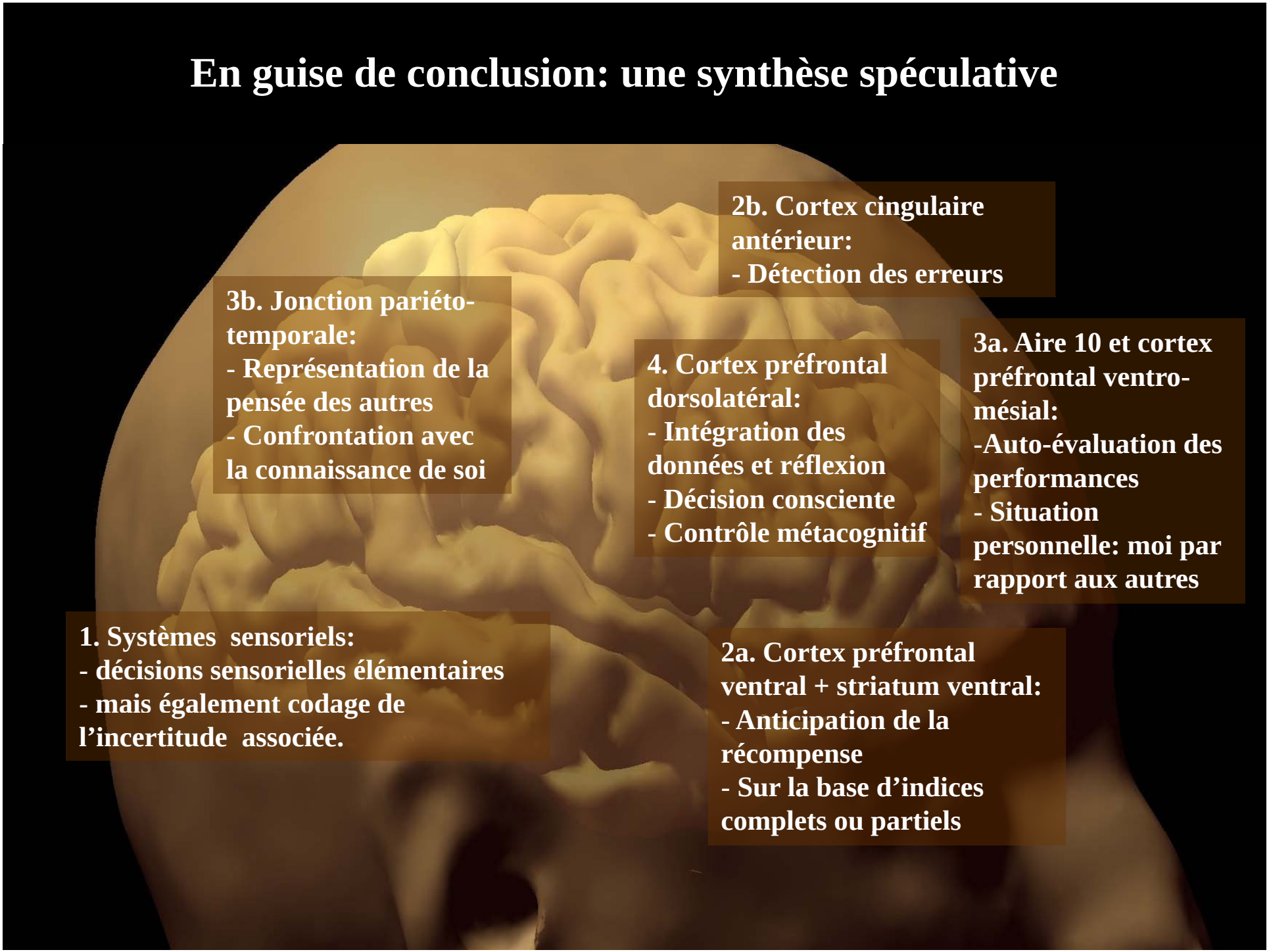


- L'activation croit linéairement avec la valeur absolue de la différence des données sensorielles (jugements numériques du plaisir éprouvé, mesurés dans des essais distincts).
- Le cortex préfrontal rostro-médian pourrait coder la probabilité que la réponse soit correcte (un signal proche de la confiance en soi ou du sentiment de savoir).

Cependant, cette étude souffre de plusieurs limites évidentes:

- La courbe présentée pourrait souffrir d'un problème de sélection (*double dipping*)
- L'article présenté ne comporte aucune donnée introspective ou métacognitive, mais uniquement un jugement de niveau 1. Rien ne prouve donc que le signal mesuré dans cette région ait à voir avec le jugement de confiance.

En guise de conclusion: une synthèse spéculative



2b. Cortex cingulaire antérieur:
- Détection des erreurs

3b. Jonction pariéto-temporale:
- Représentation de la pensée des autres
- Confrontation avec la connaissance de soi

4. Cortex préfrontal dorsolatéral:
- Intégration des données et réflexion
- Décision consciente
- Contrôle métacognitif

3a. Aire 10 et cortex préfrontal ventro-mésial:
-Auto-évaluation des performances
- Situation personnelle: moi par rapport aux autres

1. Systèmes sensoriels:
- décisions sensorielles élémentaires
- mais également codage de l'incertitude associée.

2a. Cortex préfrontal ventral + striatum ventral:
- Anticipation de la récompense
- Sur la base d'indices complets ou partiels