

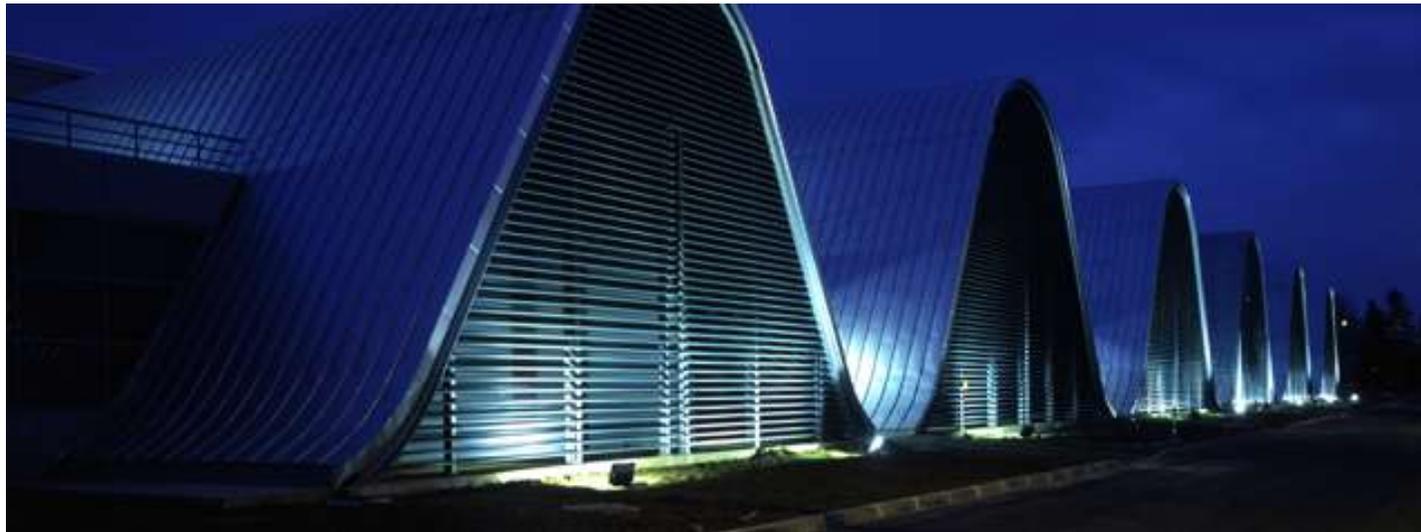
**Cours 2018-2019:**

**Progrès récents en imagerie cérébrale  
et décodage des représentations mentales**

Stanislas Dehaene  
Chaire de Psychologie Cognitive Expérimentale

Cours n°3

**Vers une cartographie de plus en plus fine des représentations mentales**



## Le concept d'isomorphisme de second ordre

Shepard, R. N., & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1, 1–17.

On pense parfois que, lorsqu'on imagine un carré vert, il doit y avoir une « image » analogue de ce carré dans le cerveau.

Mais ce type de raisonnement conduit rapidement à des absurdités

- Qui reconnaît l'image de ce carré dans le cerveau ? (régression à l'infini)
- Comment est codée la couleur verte? Par des neurones verts??
- Comment est codé le nombre? La liberté? et autres concepts abstraits
- Le temps: une durée n'est pas nécessairement codée par une durée !

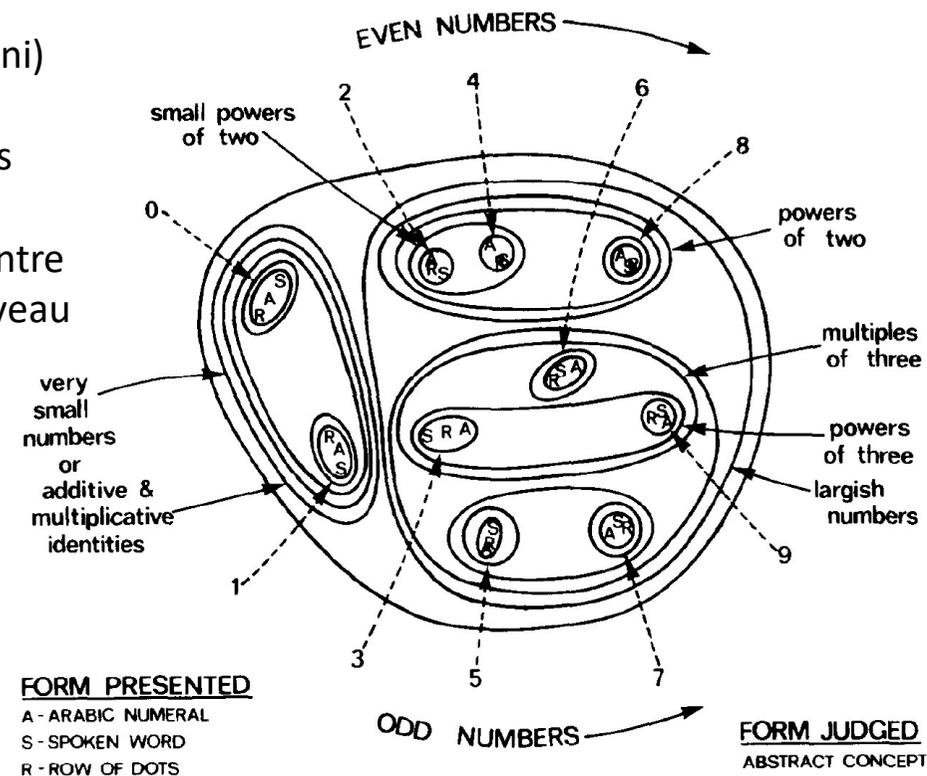
En général, il serait vain d'espérer trouver une correspondance directe entre les objets du monde extérieur et leur représentation interne dans le cerveau (un **isomorphisme de premier ordre**).

Par contre, on peut espérer trouver un isomorphisme **de second ordre**, c'est-à-dire une correspondance entre

- Les relations entre des objets du monde extérieur
- Et les relations entre leurs représentations internes

La représentation d'un rectangle doit ressembler plus à celle d'un carré qu'à celle d'un chou-fleur!

Shepard a défini une méthode de visualisation de la similarité subjective (*multidimensional scaling*) et réalisé de nombreuses études empiriques de différentes représentations mentales



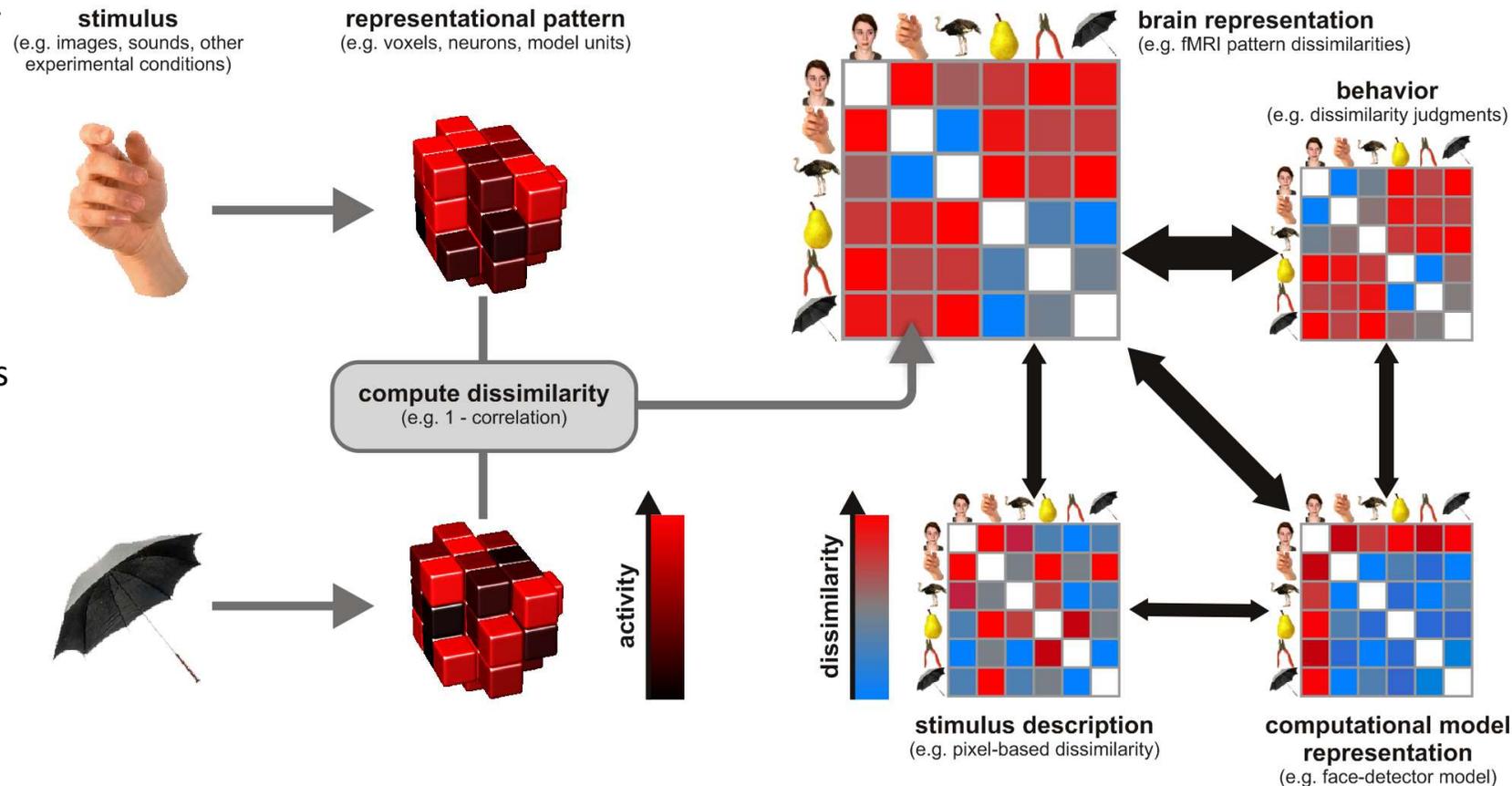
# L'analyse de la similarité des représentations

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2, 4.

Kriegeskorte et al. étendent l'idée de Shepard: ils proposent de tirer des inférences de la mesure de la similarité entre représentations cérébrales, dans chaque région du cerveau (RSA = *representational similarity analysis*)

Plusieurs étapes:

- Extraction des configurations d'activité pour chaque stimulus
- Mesure de la dissimilarité
- Comparaison avec les matrices prédites par différents modèles
- Etude de cette matrice (par exemple par *multi-dimensional scaling*)



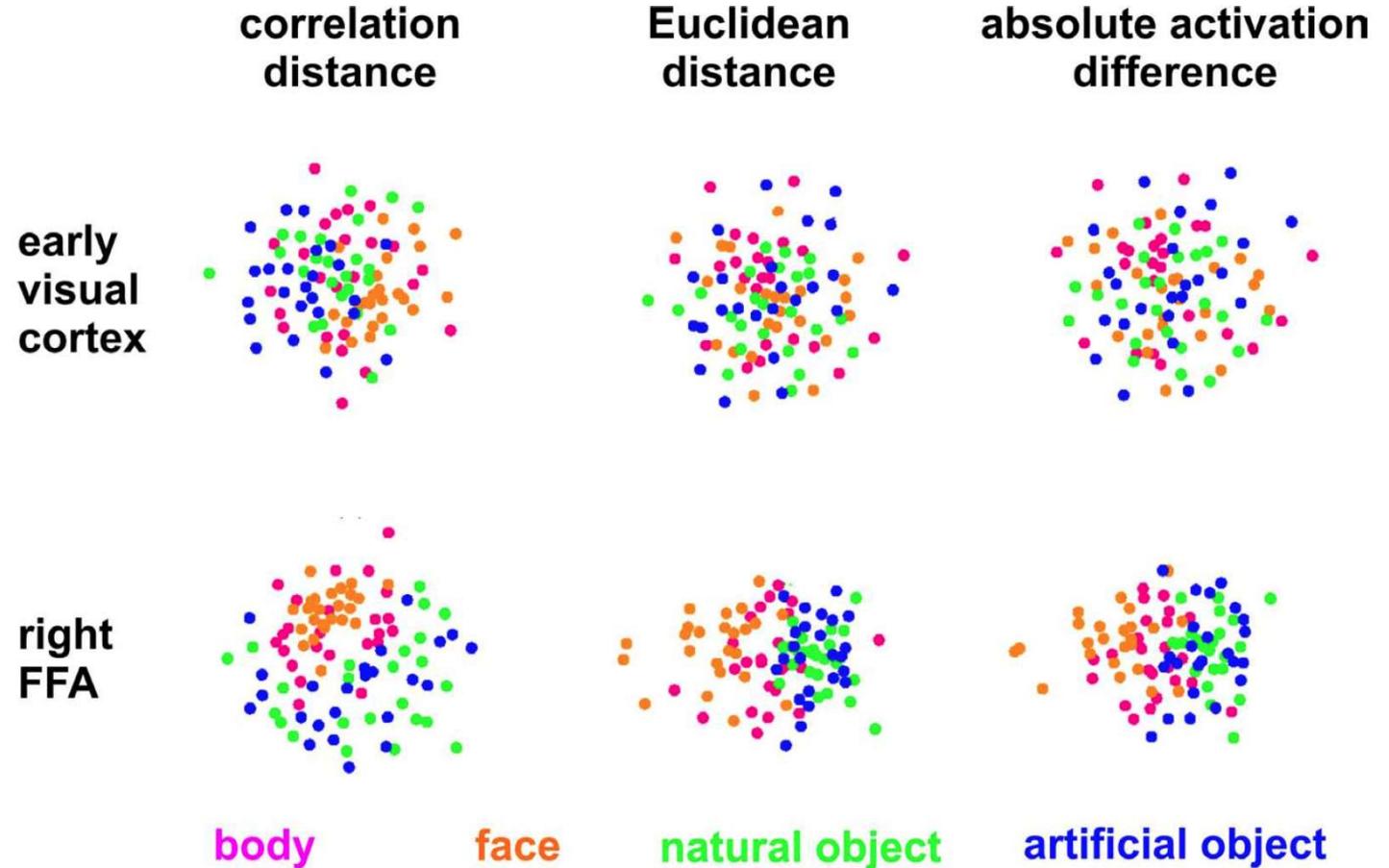
# L'analyse de la similarité des représentations

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2, 4.

Exemple:

*Multidimensional scaling* de l'activité évoquée par 96 images, dans le cortex visuel précoce (V1) et dans l'aire fusiforme sensible aux visages (FFA)

Les visages sont nettement distingués des autres stimuli, uniquement dans l'aire FFA.

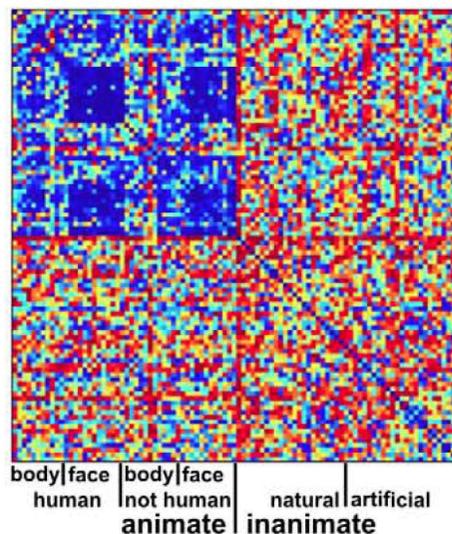


Voici les matrices complètes dans deux régions : *Fusiform Face Area* (FFA) et V1.

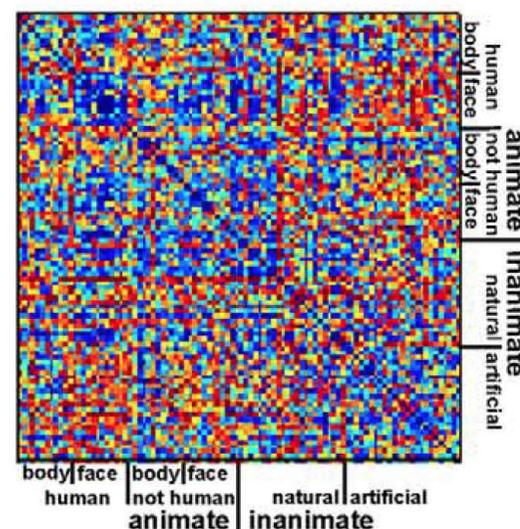
Pour comprendre ces matrices, on peut tenter de les comparer avec différentes matrices théoriques.

Il s'agit bien d'une analyse de 2<sup>nd</sup> ordre: Analyse de la similarité des matrices de similarité!

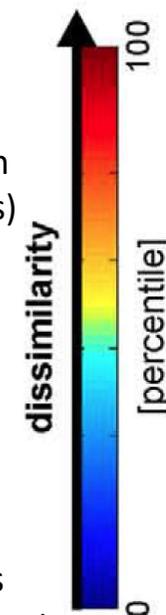
FFA



V1

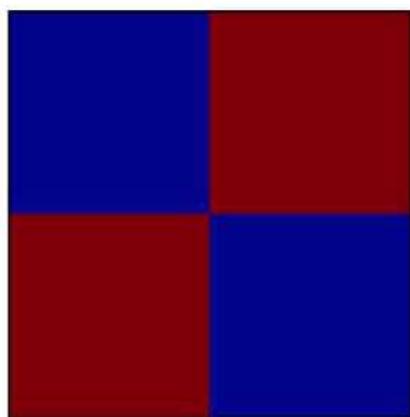


Les activités différent beaucoup (à noter l'échelle en percentiles)

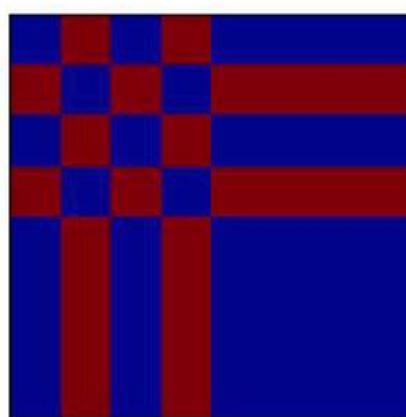


Les activités se ressemblent

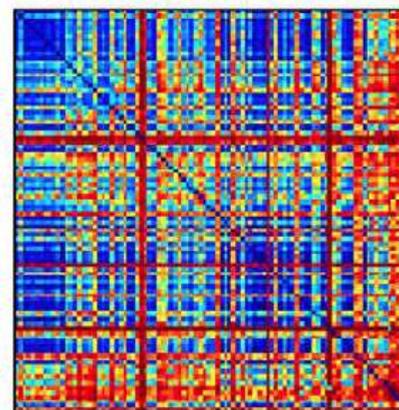
animate-inanimate



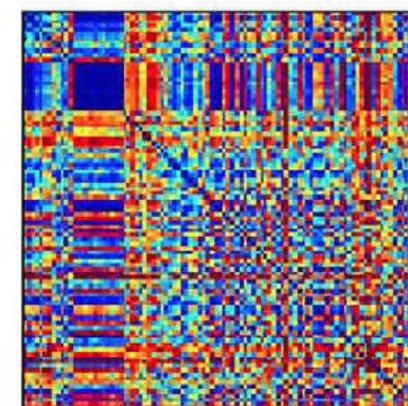
face-nonface



V1 model



silhouette image



# Détails techniques: quelle est la meilleure mesure de similarité?

Méthodologie : Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS Computational Biology*, 10(4), e1003553. <https://doi.org/10.1371/journal.pcbi.1003553> + VOIR AUSSI CosmoMOPA

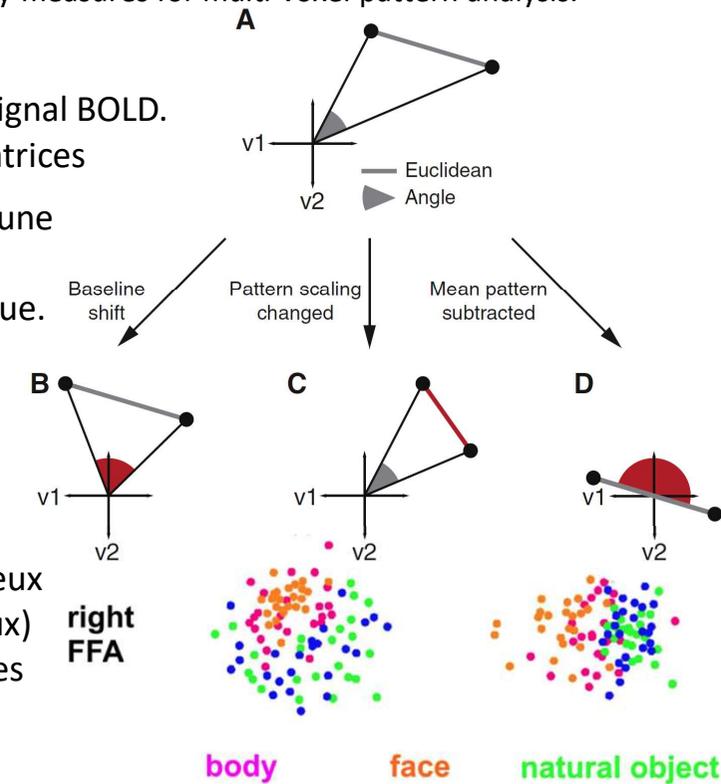
Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, 137, 188–200. <https://doi.org/10.1016/j.neuroimage.2015.12.012>

La **corrélation** fait abstraction à la fois des différences d'activation moyenne et de l'amplitude du signal BOLD.  
 La **corrélation des rangs (Spearman)** évite de faire l'hypothèse d'une relation linéaire entre les matrices

La **matrice de confusion d'un décodeur** n'est pas une mesure stable de distance (elle est discrète et sensible aux biais). Mieux vaut une mesure continue.

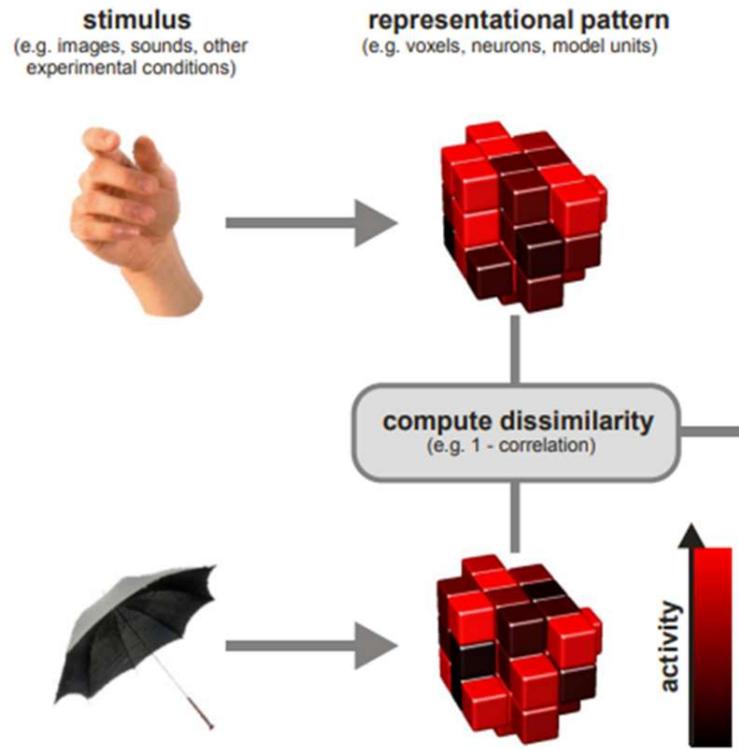
La **distance Euclidienne** est intéressante:

- Contrairement à la corrélation (mesure angulaire,  $1 - \cos \varphi$ ), elle est insensible aux changements de ligne de base.
- Elle vaut zéro SSI les vecteurs sont identiques, alors que la corrélation peut être nulle pour deux raisons (pas d'activité, ou vecteurs orthogonaux)
- Avec normalisation pour le bruit (e.g. utiliser les cartes t ou Mahalanobis): les matrices de dissimilarité sont moins bruitées.
- Et avec cross-validation sur des données indépendantes (sinon, à cause du bruit, la distance n'est pas nulle pour des représentations identiques mesurées 2 fois)



$$d_{Euclidean}^2(\mathbf{b}_k, \mathbf{b}_j) = \|\mathbf{b}_j - \mathbf{b}_k\|^2 = (\mathbf{b}_j - \mathbf{b}_k)(\mathbf{b}_j - \mathbf{b}_k)^T$$

$$d_{Euclidean, crossvalidated}^2(\mathbf{b}_k, \mathbf{b}_j) = (\mathbf{b}_j - \mathbf{b}_k)_A (\mathbf{b}_j - \mathbf{b}_k)_B^T$$

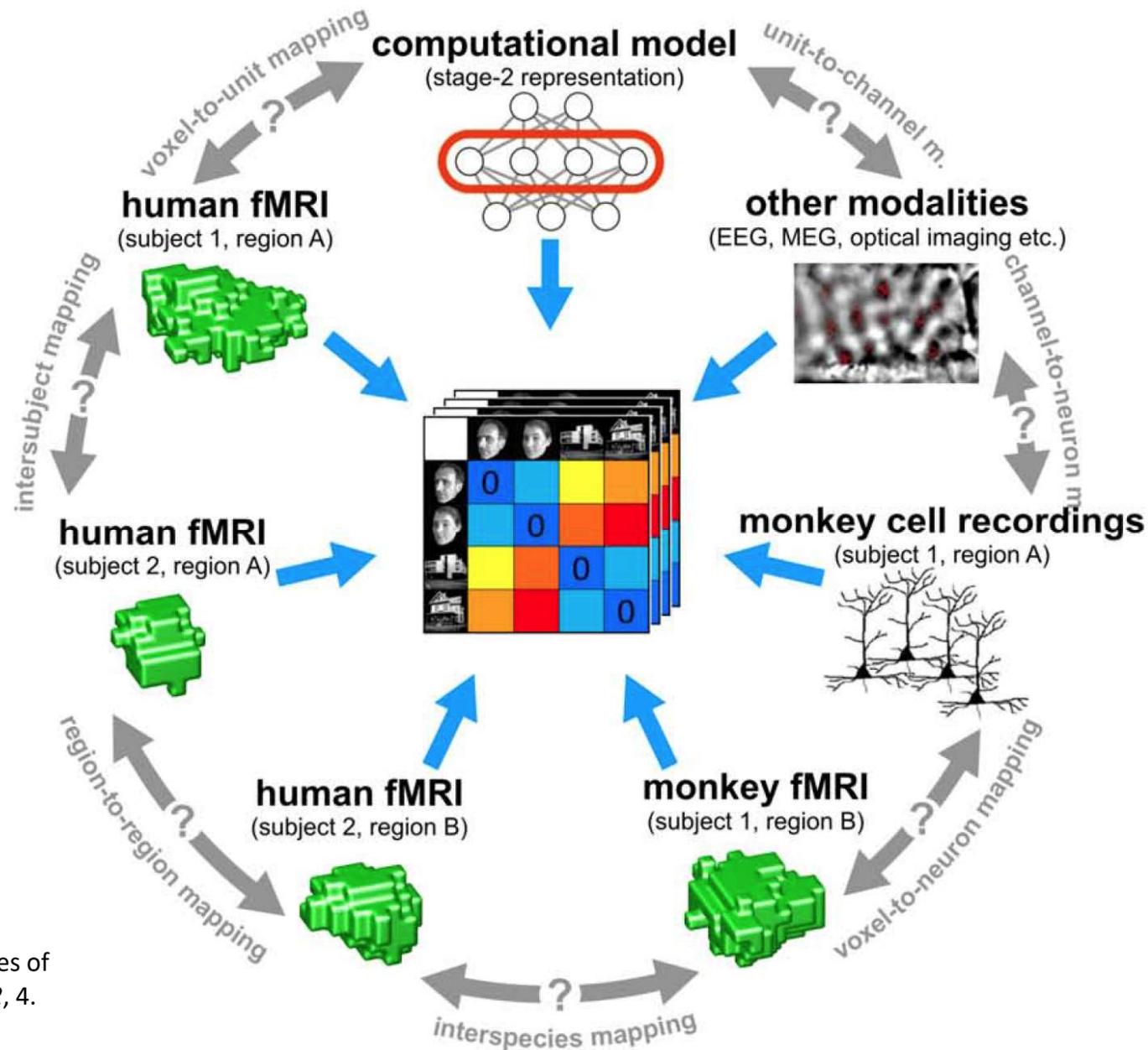


# L'analyse de la similarité des représentations: Une méthode unificatrice pour les neurosciences, quelle que soient les modalités de mesure de l'activité cérébrale.

Dans ce cours, nous verrons 4 applications

1. « Hyper-alignement » de plusieurs cerveaux humains
2. Comparaison entre les représentations neuronales chez l'homme et chez le singe
3. Comparaison entre représentation cérébrale et comportement
4. Analyse de représentations cognitives de haut niveau (par ex. celle du nombre)

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2, 4.



# Exemple 1: Améliorer l'alignement des représentations inter-sujets par hyper-alignement

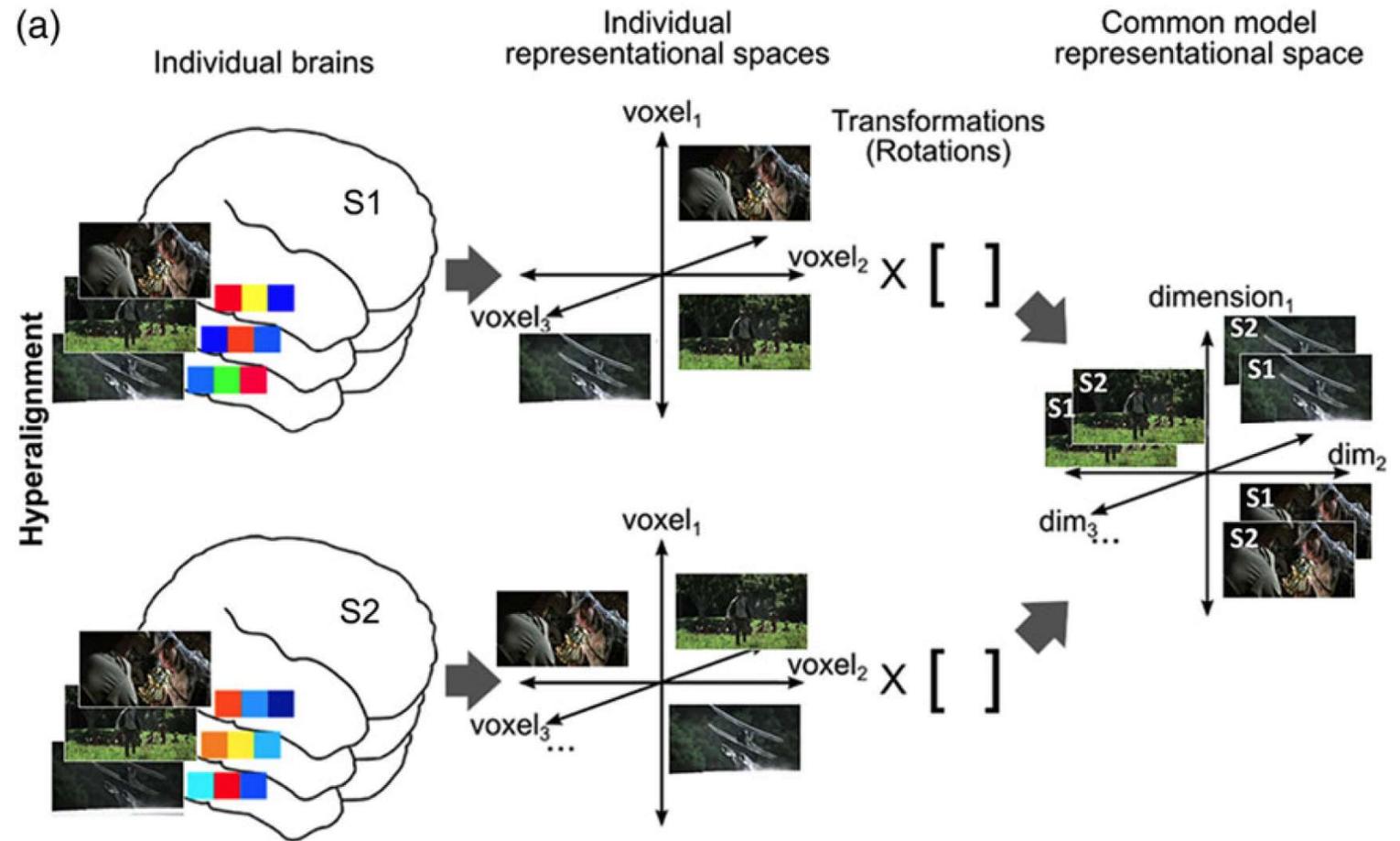
Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A Model of Representational Spaces in Human Cortex. *Cerebral Cortex (New York, N.Y.: 1991)*. <https://doi.org/10.1093/cercor/bhw068>

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., ... Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72(2), 404–416. <https://doi.org/10.1016/j.neuron.2011.08.026>

Comment normaliser les données d'IRM fonctionnelle issues de plusieurs sujets?

Idée:

- Les sujets ont des activations différentes, mais leurs matrices de similarité se ressemblent, indiquant que leurs représentations sont isomorphes (MDS)
- On va donc mesurer la matrice de réponse de deux sujets à des stimuli identiques (un film par exemple)
- Et aligner, non seulement les surfaces corticales, mais les **hyper-espaces** engendrés par les variations d'activité à travers tous les voxels d'une région d'intérêt (analyse multivariée).



# Améliorer l'alignement des représentations inter-sujets: L'hyper-alignement

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A Model of Representational Spaces in Human Cortex. *Cerebral Cortex (New York, N.Y.: 1991)*. <https://doi.org/10.1093/cercor/bhw068>

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., ... Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72(2), 404–416. <https://doi.org/10.1016/j.neuron.2011.08.026>

Au niveau du cerveau tout entier:

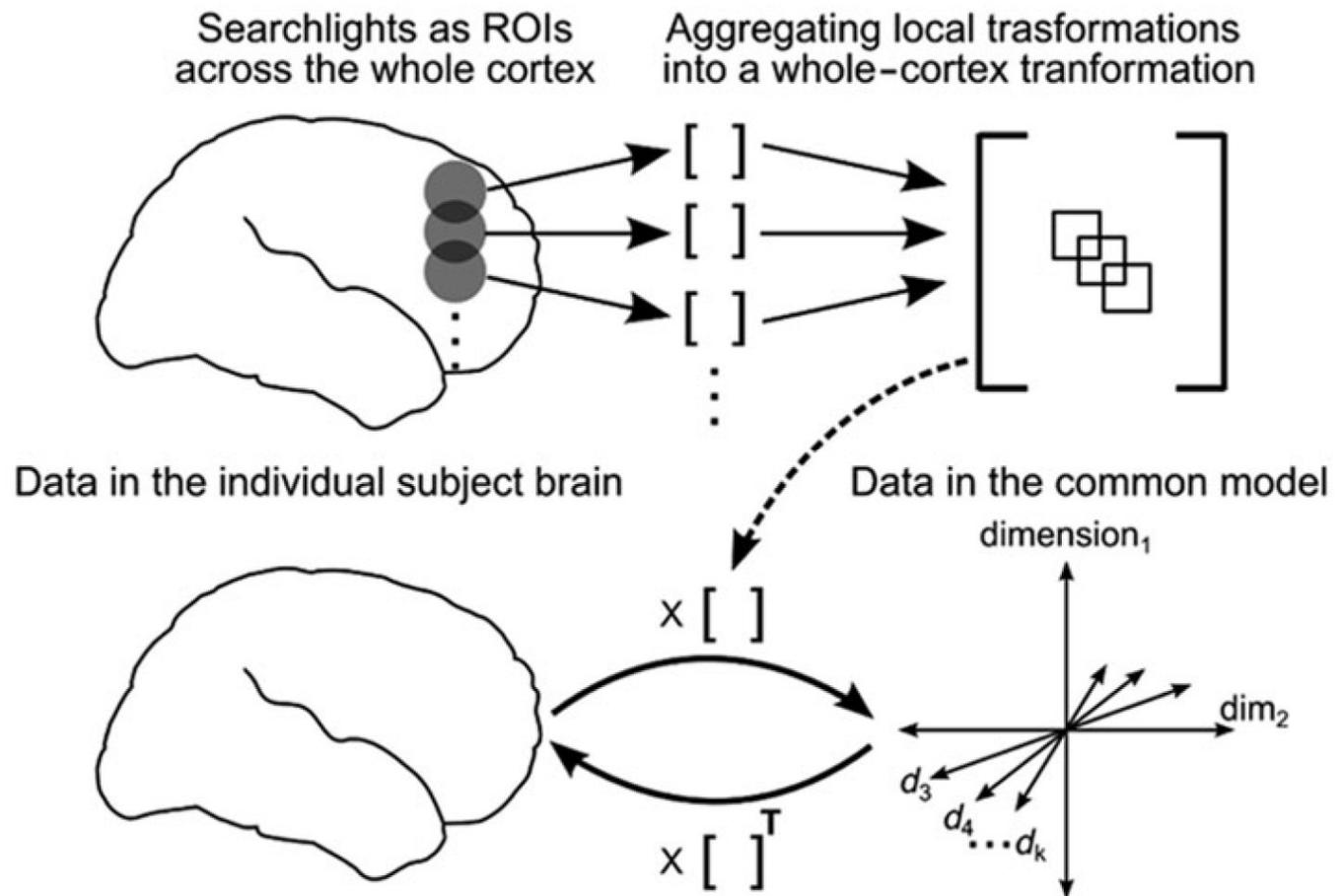
D'abord, hyper-aligner toute une série de sphères (de 20 mm de rayon).

Puis agréger toutes ces transformations pour hyper-aligner l'ensemble du cortex.

Enfin, appliquer cette transformation à d'autres données que celles de départ (le film).

(b)

Whole-cortex hyperalignment

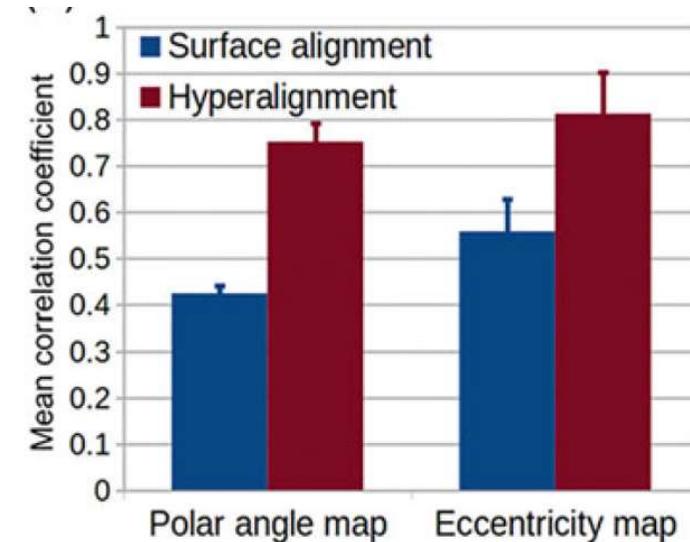
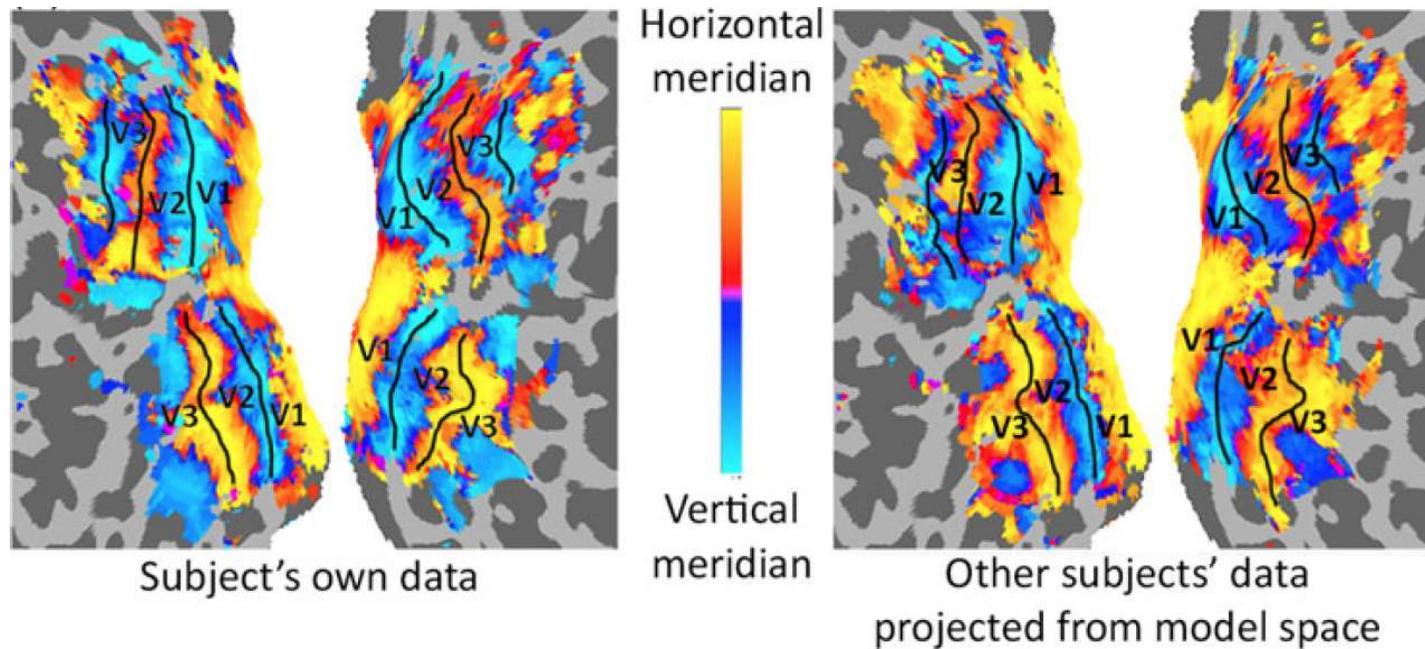


# Améliorer l'alignement des représentations inter-sujets: L'hyper-alignement

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A Model of Representational Spaces in Human Cortex. *Cerebral Cortex (New York, N.Y.: 1991)*. <https://doi.org/10.1093/cercor/bhw068>

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., ... Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72(2), 404–416. <https://doi.org/10.1016/j.neuron.2011.08.026>

Exemple: l'alignement des cartes rétinotopiques



# Améliorer l'alignement des représentations inter-sujets: L'hyper-alignement

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A Model of Representational Spaces in Human Cortex. *Cerebral Cortex (New York, N.Y.: 1991)*. <https://doi.org/10.1093/cercor/bhw068>

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., ... Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72(2), 404–416. <https://doi.org/10.1016/j.neuron.2011.08.026>

Exemple: le décodage de six espèces d'animaux (searchlight; niveau du hasard = 16%)

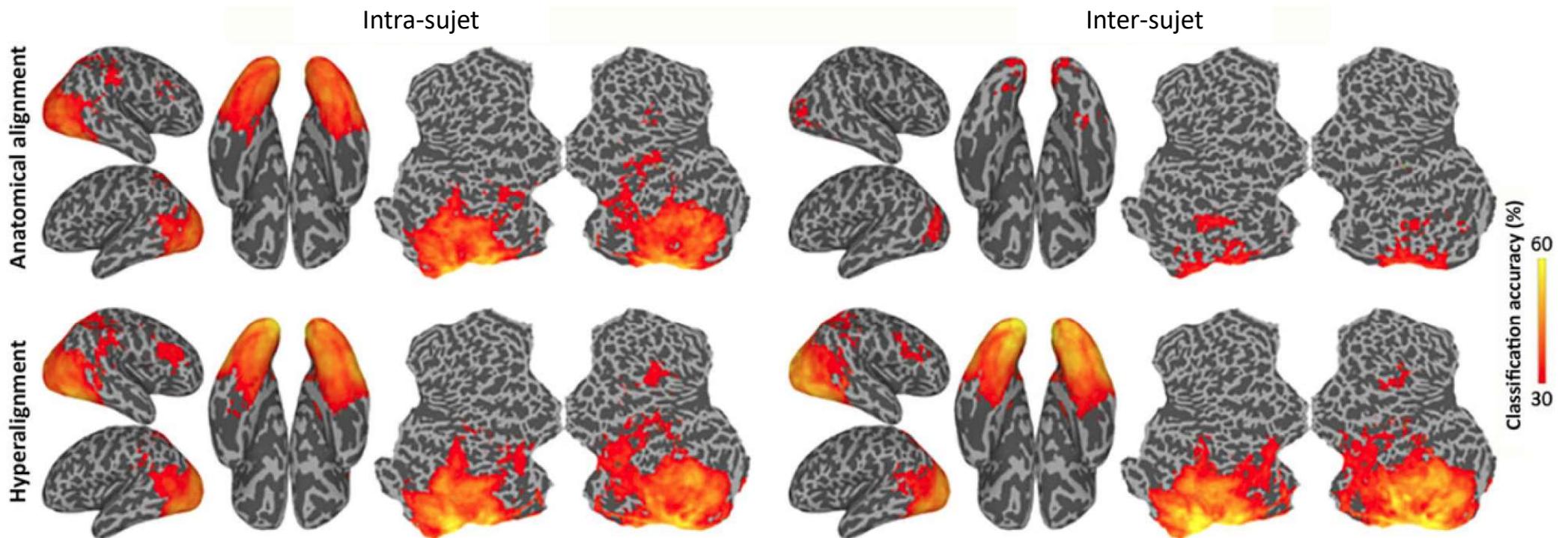


Figure 6. MVPC of responses to animal species. Searchlight-based classification accuracies of 6 animal species within-subjects (left) and between-subjects (right). Classification is performed on the data after anatomical alignment (top row) and whole-cortex hyperalignment (bottom row). Whole-cortex transformation matrices were derived from the movie data.

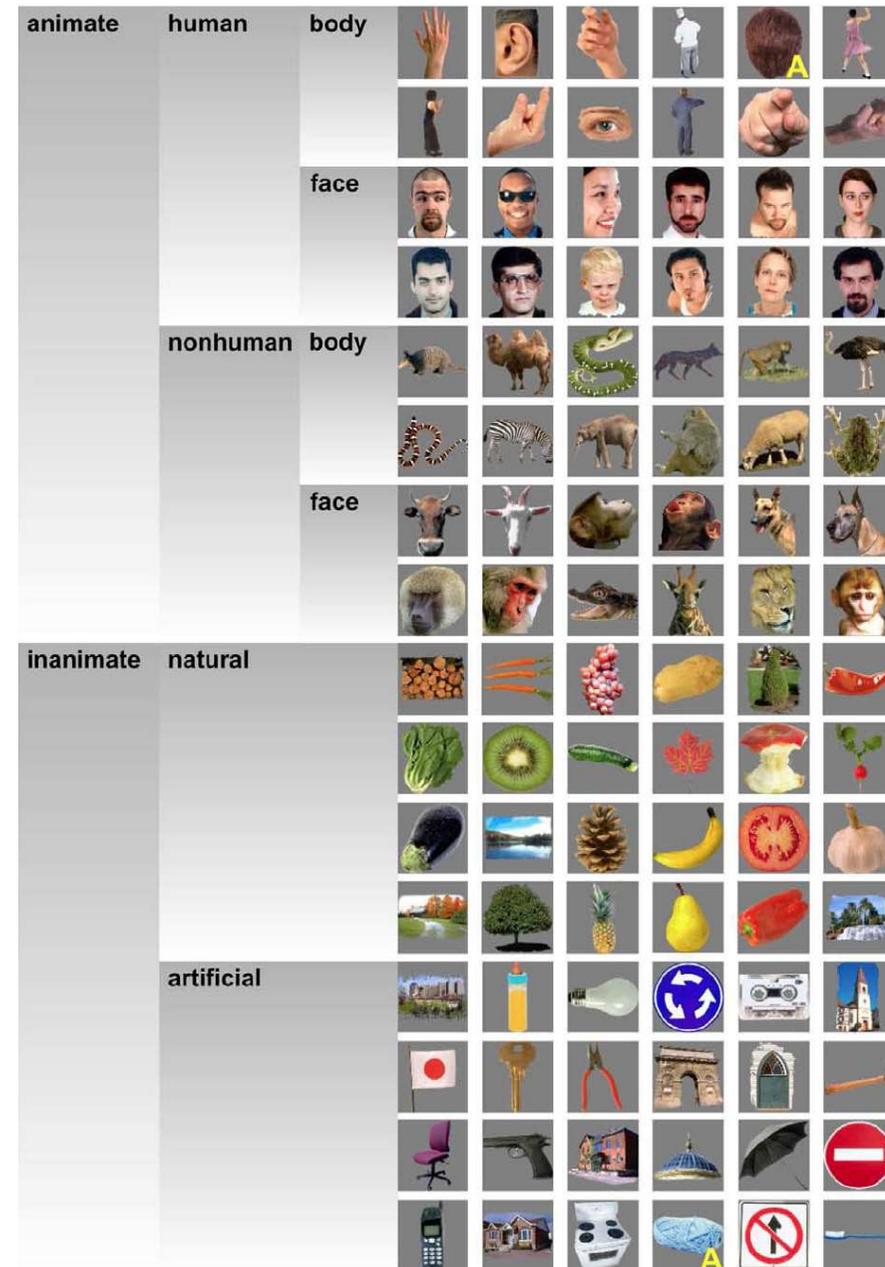
## Exemple 2: Comparaison de la représentation des objets dans le cortex inféro-temporal des humains et des singes macaques

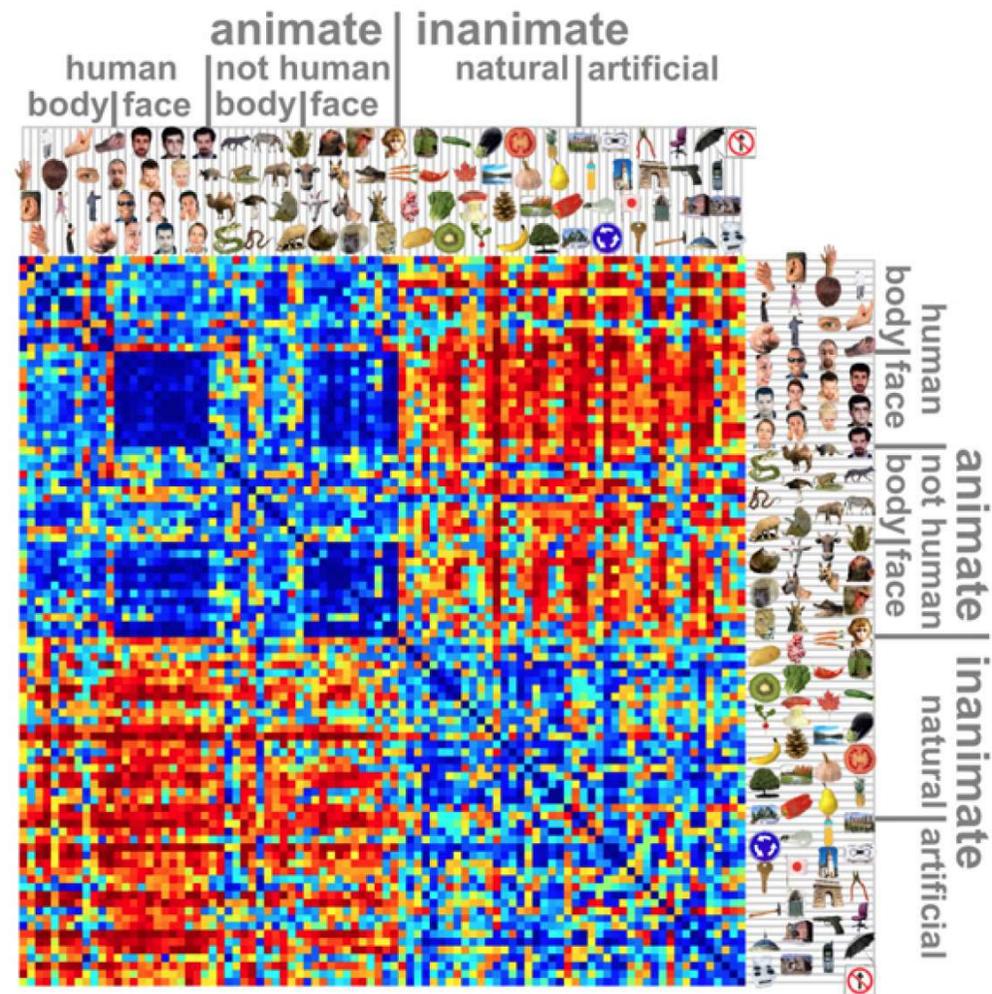
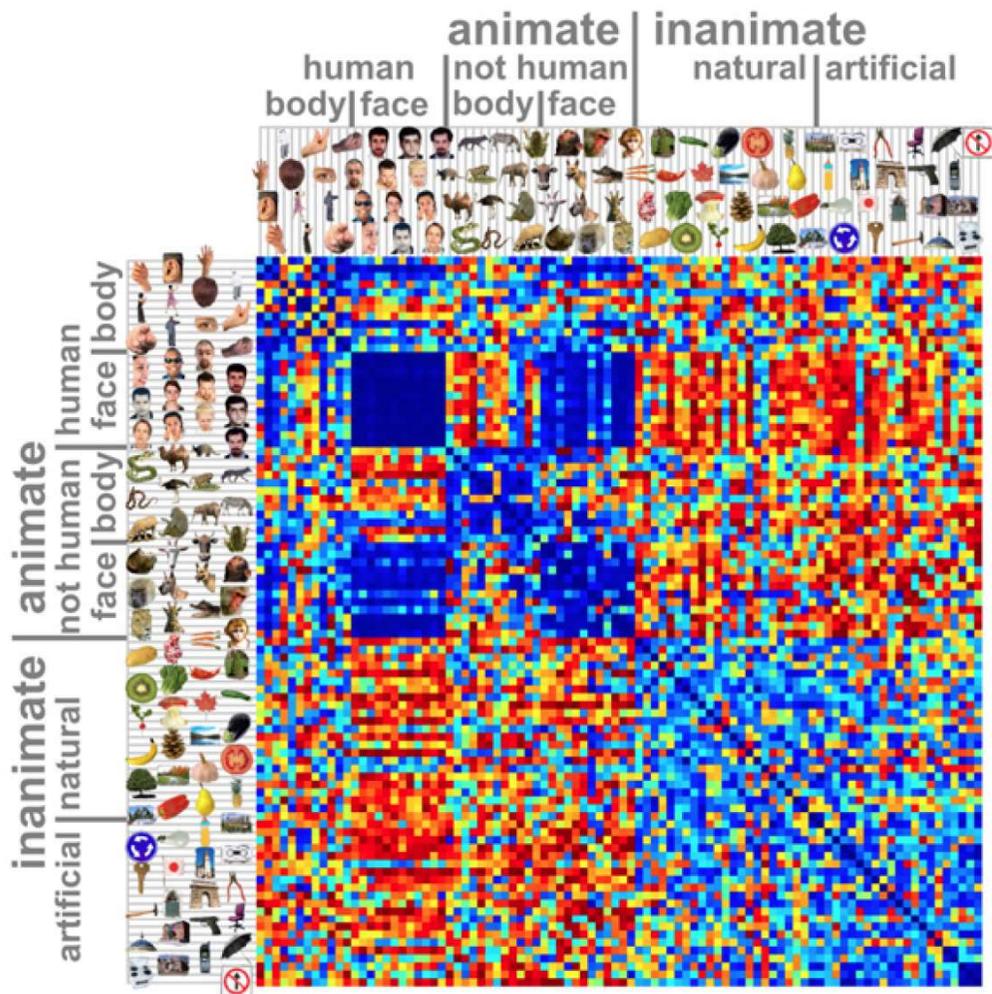
Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... Bandettini, P. A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*, 60(6), 1126–1141.

<https://doi.org/10.1016/j.neuron.2008.10.043>

Présentation de 92 images à 2 singes macaques et à 4 humains

- Singes: enregistrements électrophysiologiques de 674 neurones, mesure des décharges moyennes entre 0 et 140 ms
- Humains: réponses en IRM fonctionnelle à une image présentée pendant 300 ms, toutes les 4 secondes. Voxels de ~2 mm isotropes, slab de 5 cm, à 3 Tesla

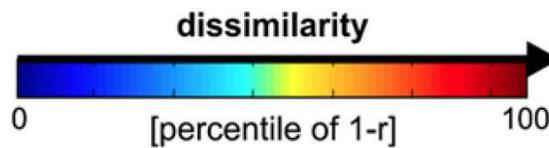




La catégorie (visage/non-visage, animé/inanimé) est un facteur dominant dans les deux espèces

**monkey IT**

**human IT**

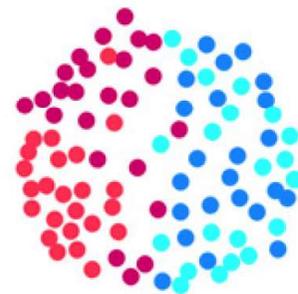
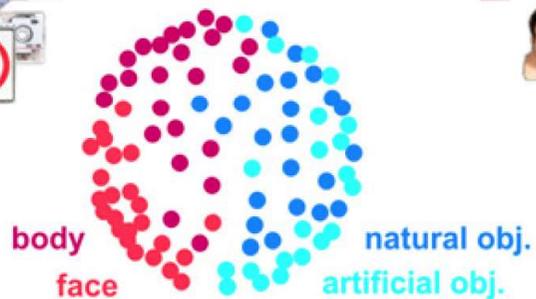




**monkey IT**



**human IT**



*Multi-dimensional scaling* suivi d'un alignement Procrustéen entre les deux espèces.

L'organisation est principalement catégorielle, mais il existe également une organisation intra-catégorie commune aux deux espèces.

Il existe une discrimination légèrement meilleure des visages de sa propre espèce (significative chez les humains seulement)

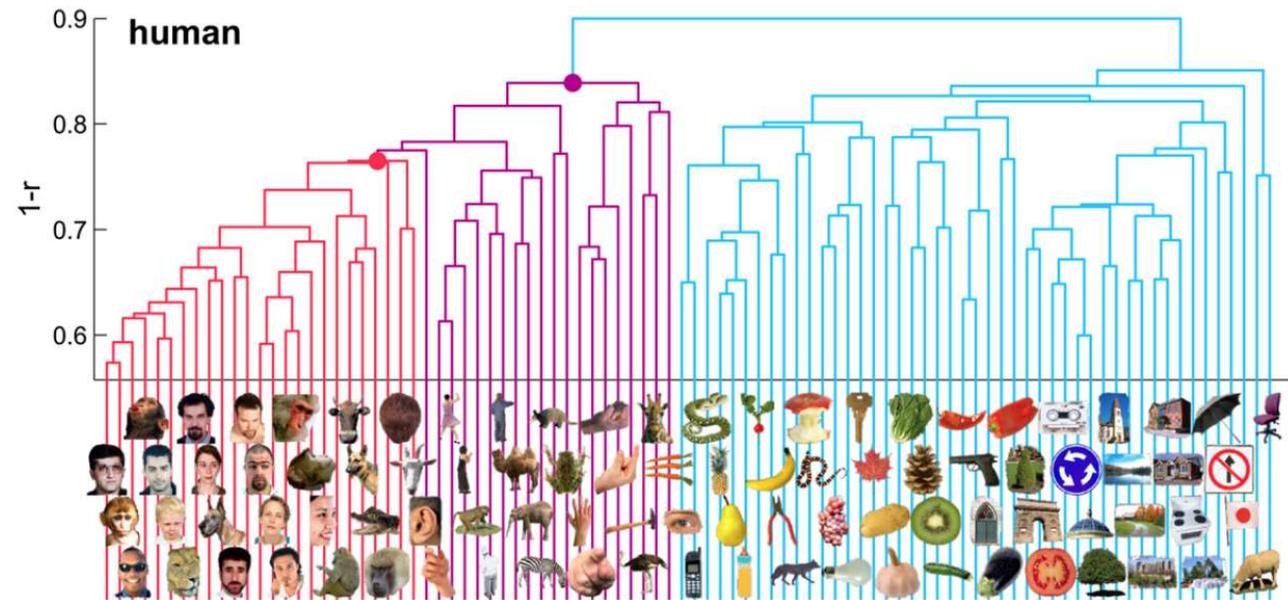
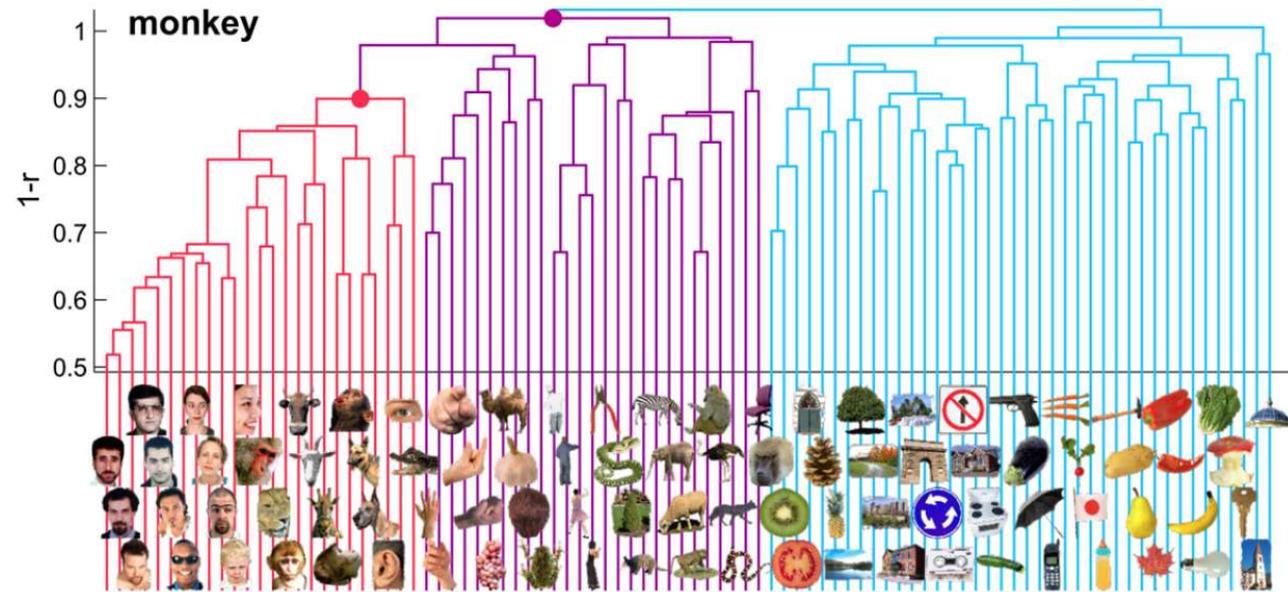
Conclusion:

en première approximation, il existe une forte similarité entre les réponses

- Des humains en IRMf
- Et des singes, en électrophysiologie

Critiques et nouvelles idées

- Peu ou pas de stimuli culturels
  - Lettres
  - Chiffres
  - Equations
- Pas de distinction entre les stimuli artificiels (outils) et naturels (pour lesquels l'évolution aurait pu doter le cerveau de tous les primates de représentations spécialisées: arbres, fleurs, fruits...)
- Pas de tentative d'être exhaustif ou représentatif de la diversité de ce que nous voyons.





# Lien avec le comportement: jugement de similarité

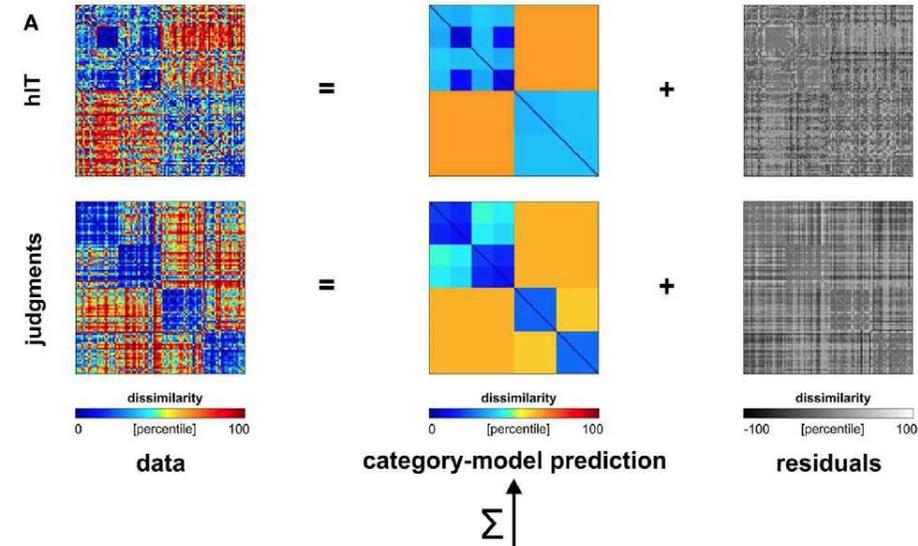
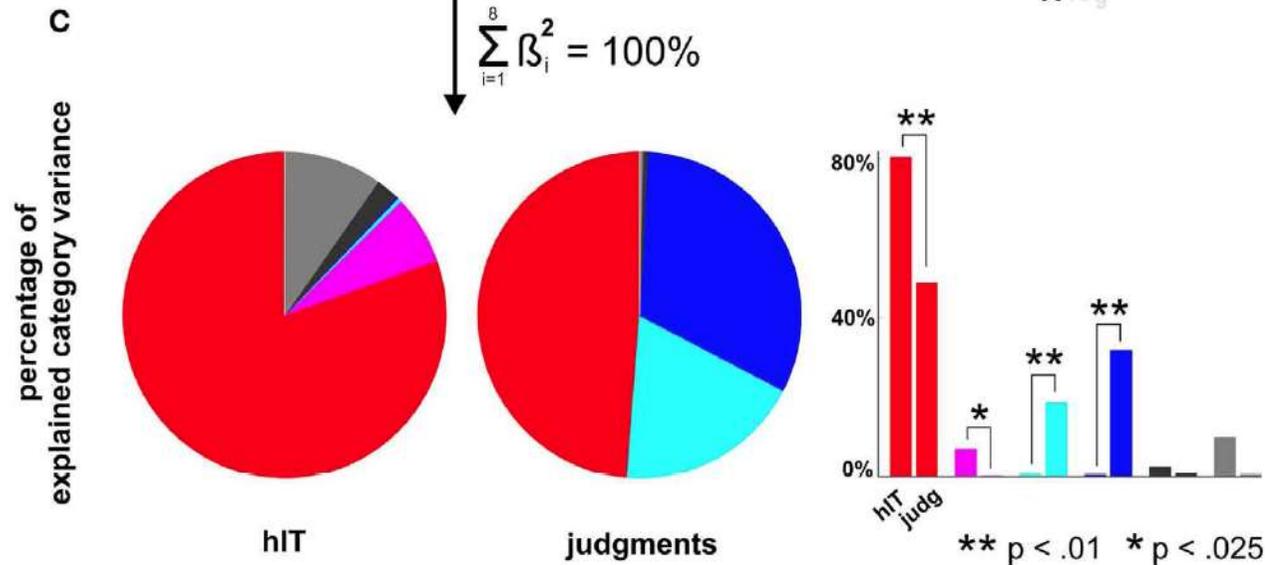
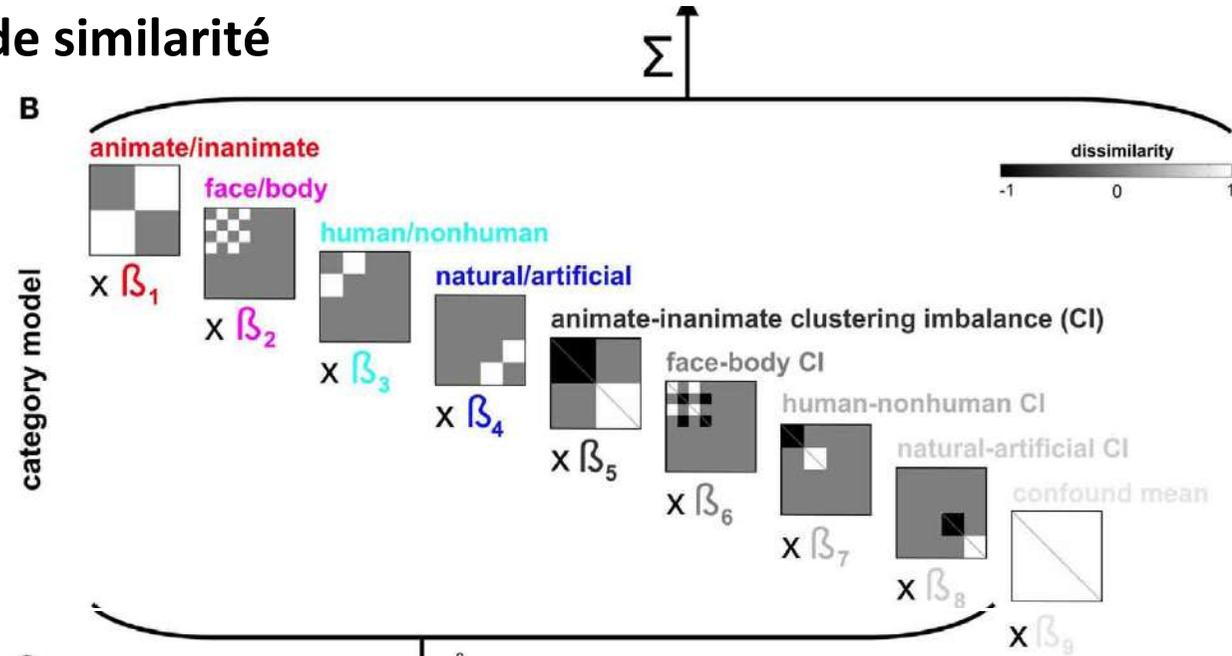
Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human Object-Similarity Judgments Reflect and Transcend the Primate-IT Object Representation. *Frontiers in Psychology*, 4, 128. <https://doi.org/10.3389/fpsyg.2013.00128>

Régression multiple:

Les jugements de similarité accentuent des distinctions catégorielles autres que la distinction animé/inanimé, et qui ne sont pas présentes dans le cortex inféro-temporal:

- humain/non-humain
- Naturel/artificiel

→ Reflet d'autres aires cérébrales qui contribuent également à la décision?



# Exemple d'application cognitive de la similarité: le codage du nombre

Elisa Castaldi, Manuela Piazza, Stanislas Dehaene, Alexandre Vignaud, Evelyn Eger (INSERM, CEA, NeuroSpin), Article en préparation

Le nombre d'un ensemble d'objets est-il effectivement codé dans le cortex pariétal?

Ou bien seulement des dimensions non-numériques telles que la taille, la luminosité, la densité ou la surface occupée?

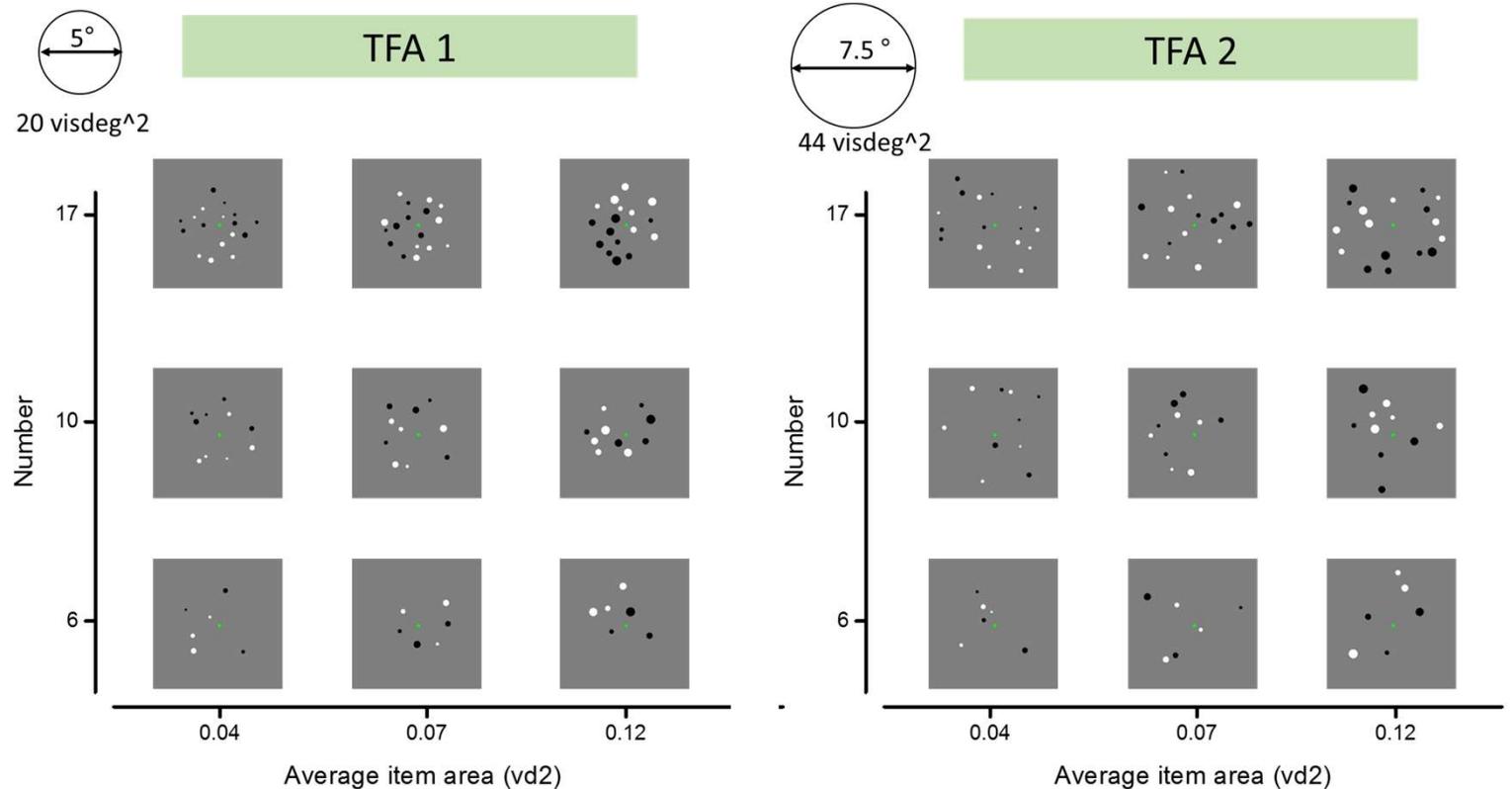
→ Création d'un nouvel espace de stimuli qui décorrèle ces paramètres

→ Utilisation de l'IRM à 7T et de l'analyse multivariée des configurations d'activité

18 conditions expérimentales.

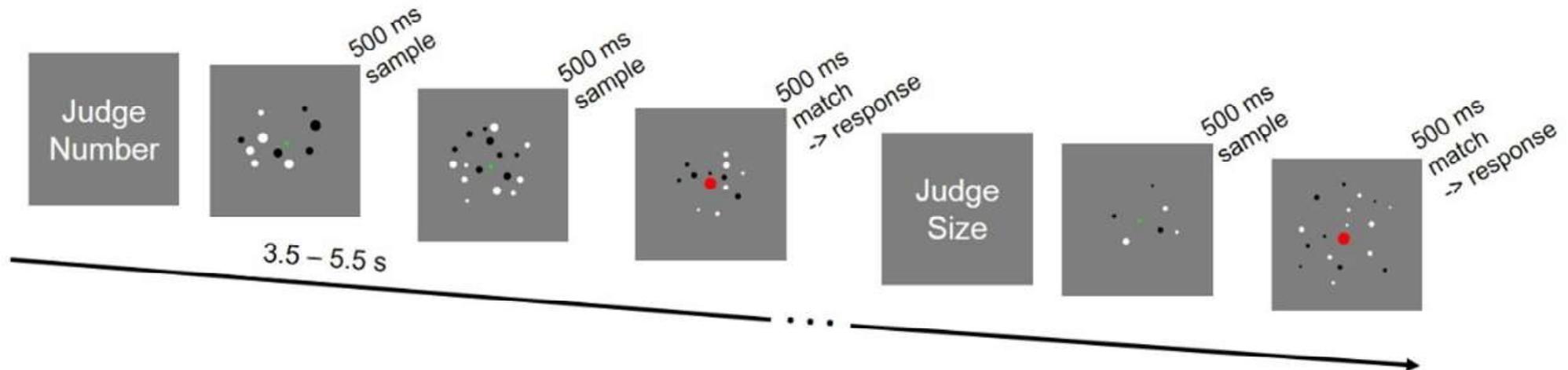
Variations orthogonales de 3 facteurs:

- Nombre: 6, 10, ou 17
- Taille moyenne des items: petit, moyen ou grand
- Surface totale occupée (TFA = *total field area*)



## Exemple 4. Application cognitive au codage du nombre

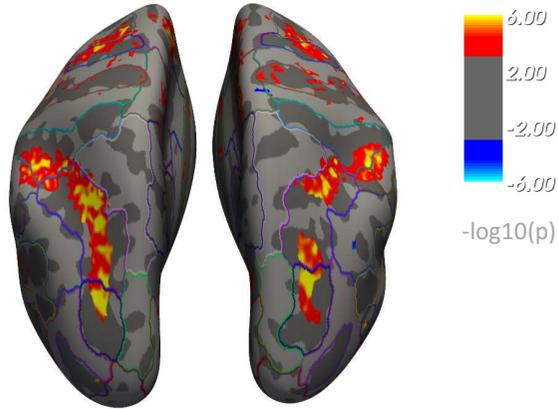
Elisa Castaldi, Manuela Piazza, Stanislas Dehaene, Alexandre Vignaud, Evelyn Eger (INSERM, CEA, NeuroSpin), Article en préparation



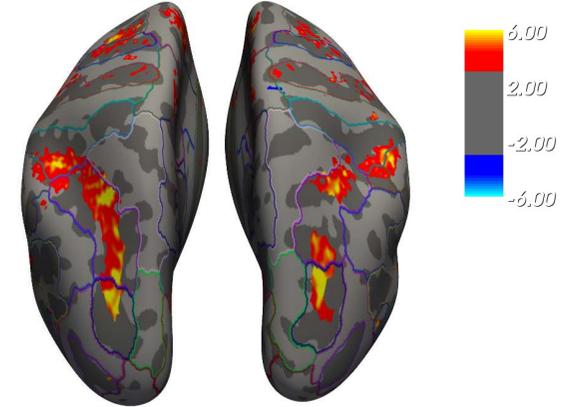
- Dans des blocs différents, les sujets font attention, soit au nombre, soit à la taille des items.
  - Pour garantir cette attention, lors de deux essais par bloc, les sujets voient le point de fixation devenir rouge, et ils doivent alors juger si le stimulus actuel est identique au précédent (dans la dimension pertinente)
  - La difficulté de cette tâche est soigneusement titrée, sujet par sujet, en fonction d'un test comportemental précédent, pour être identique dans les deux blocs.
  - 20 volontaires.
  - Aimant SIEMENS à 7T de NeuroSpin, avec insert de gradient tête seule, et antenne à 32 canaux
- IRM fonctionnelle: Résolution de 1.3 mm isotrope en utilisant la séquence multibande de Minneapolis (MB 2, IPAT 2, PF 7/8, matrix 150 x 188, 60 slices, TR 2 s, TE 22 ms, ES 0.71 ms, FA 68°, phase-encode direction left>>right)
- Images anatomiques: séquence MP2RAGE, images pondérées en T1, resolution de 0.75 mm isotrope

# Analyse univariée: pas de différence entre les deux types de jugement au niveau du groupe

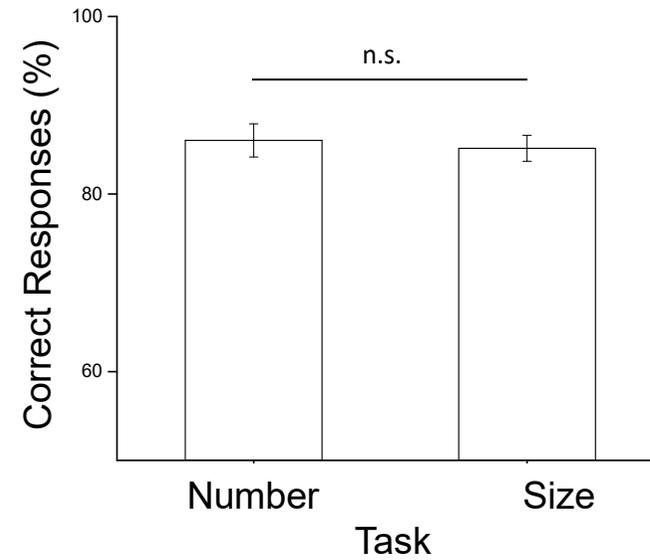
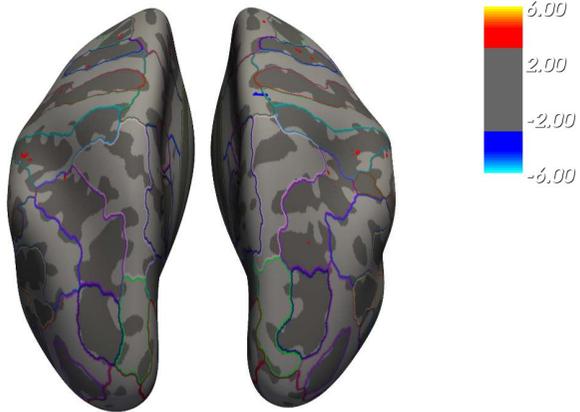
Judge Number > Baseline



Judge Size > Baseline



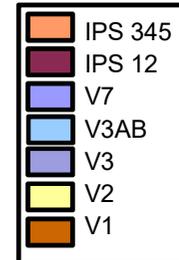
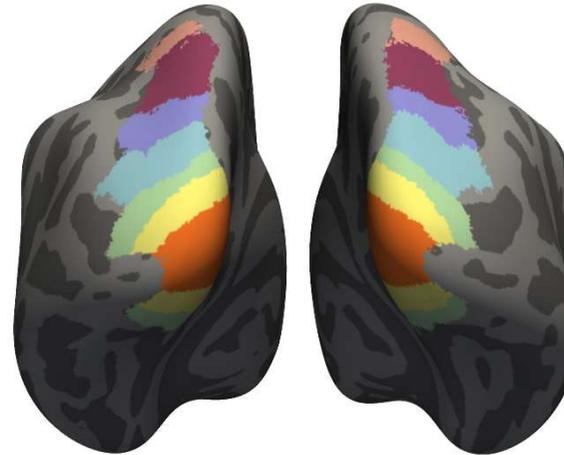
Judge Number > Judge Size



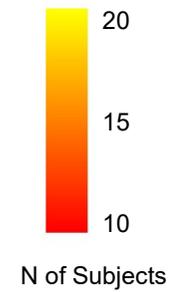
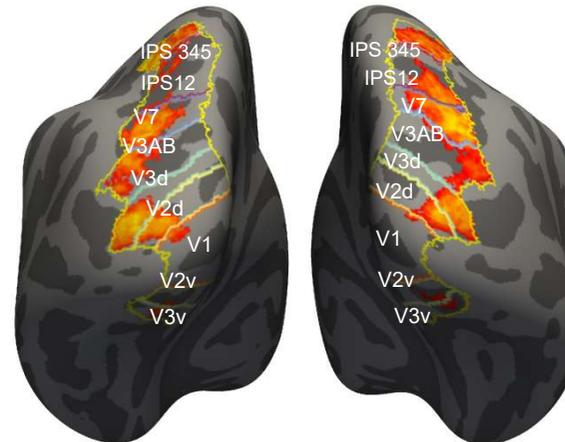
Random effects group analyses  
thresholded at  $p < .001$  uncorrected

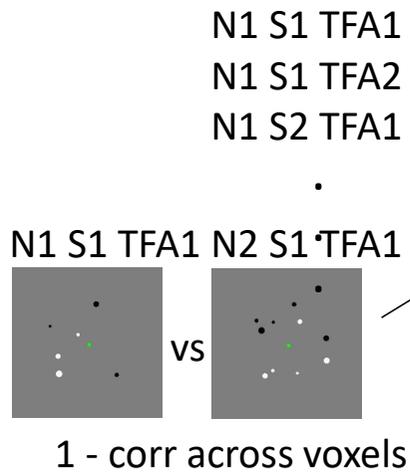
## Analyse multivariée par régions d'intérêt

Régions choisies sur la base d'un atlas probabiliste de la surface du cortex  
(Wang et al., 2015, *Cerebral Cortex*)

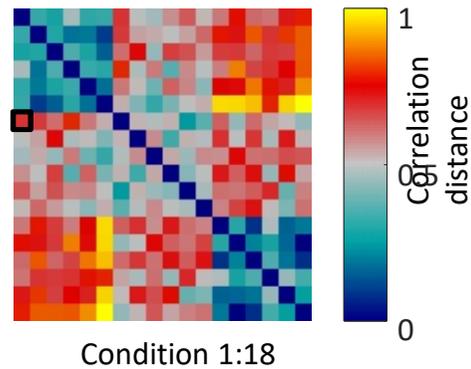


Choix des voxels, au sein de chacun de ces régions, sur la base d'un contraste orthogonal et propre à chaque sujet.  
(toutes les images > repos)

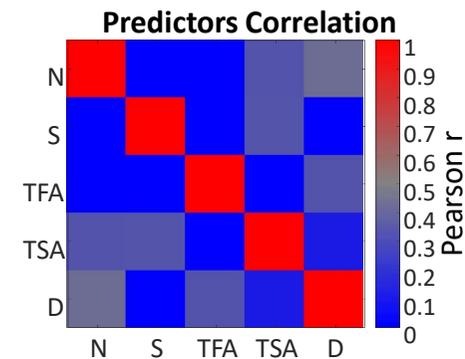
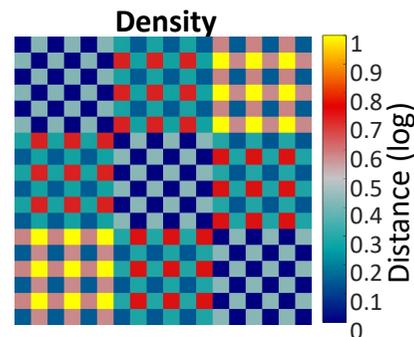
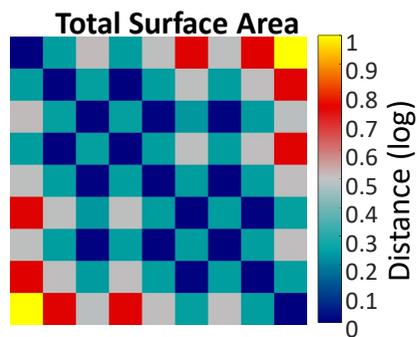
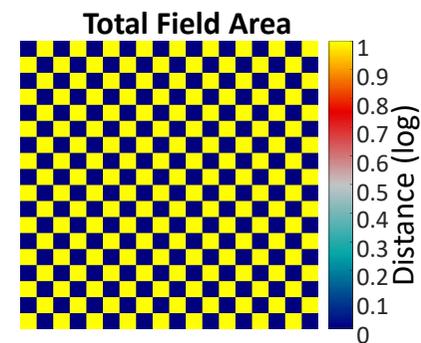
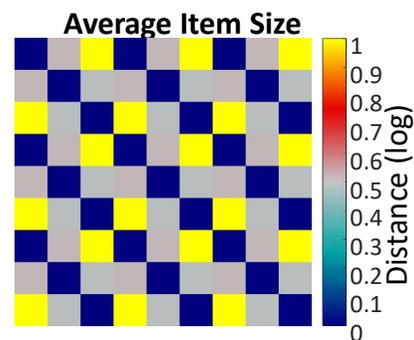
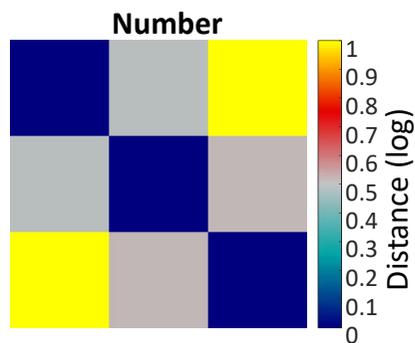




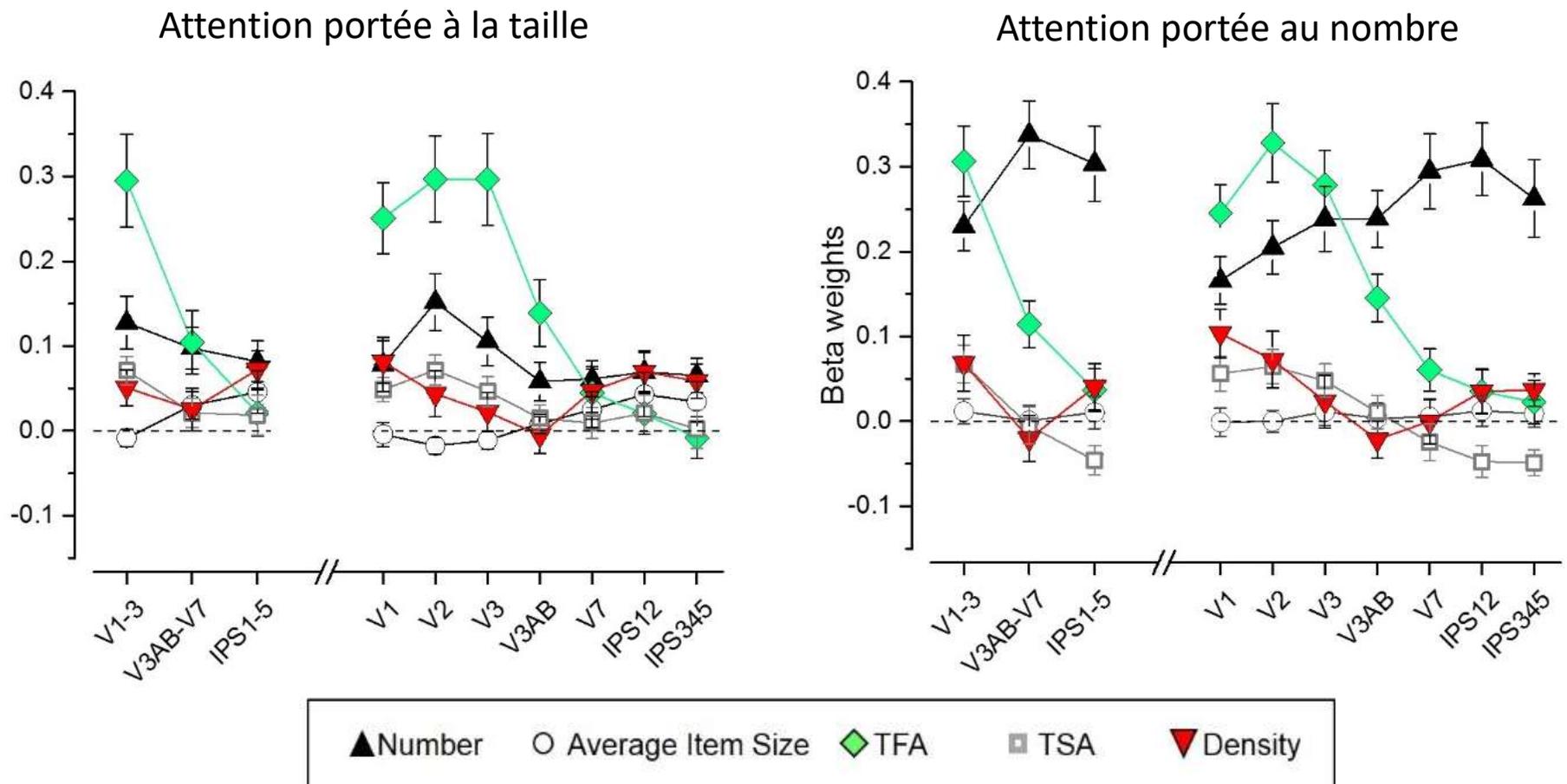
A l'intérieur de chaque région, on calcule la matrice de dissimilarité entre les 18 conditions



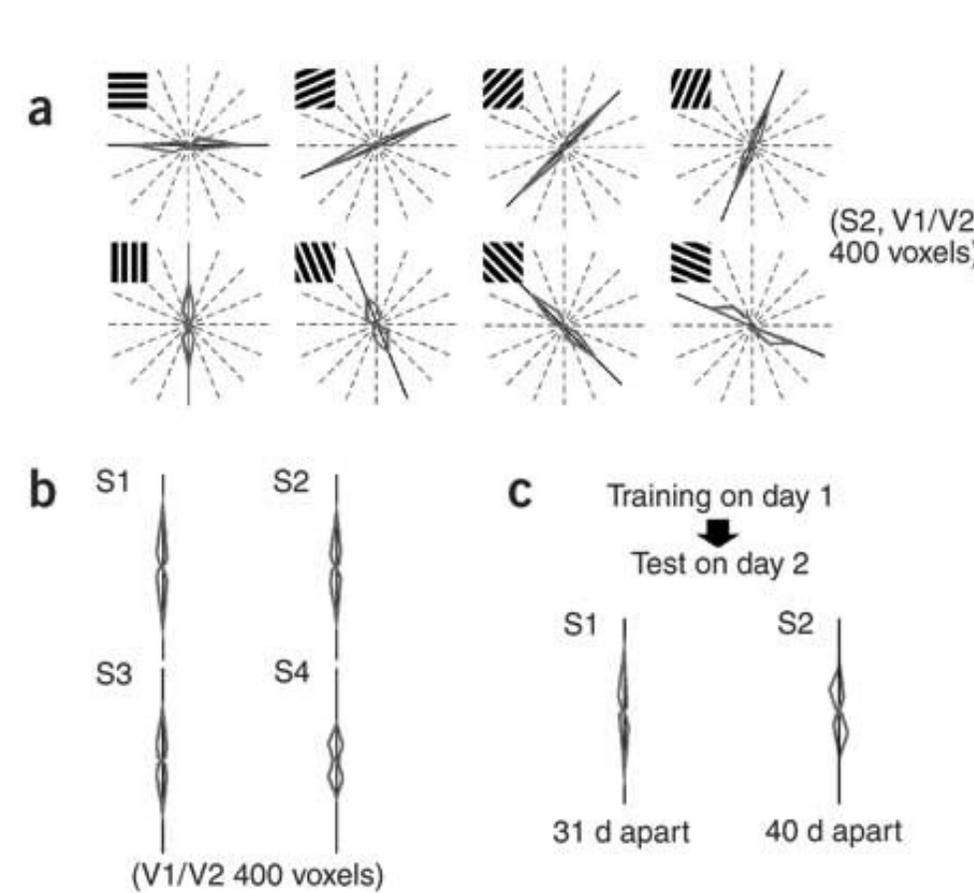
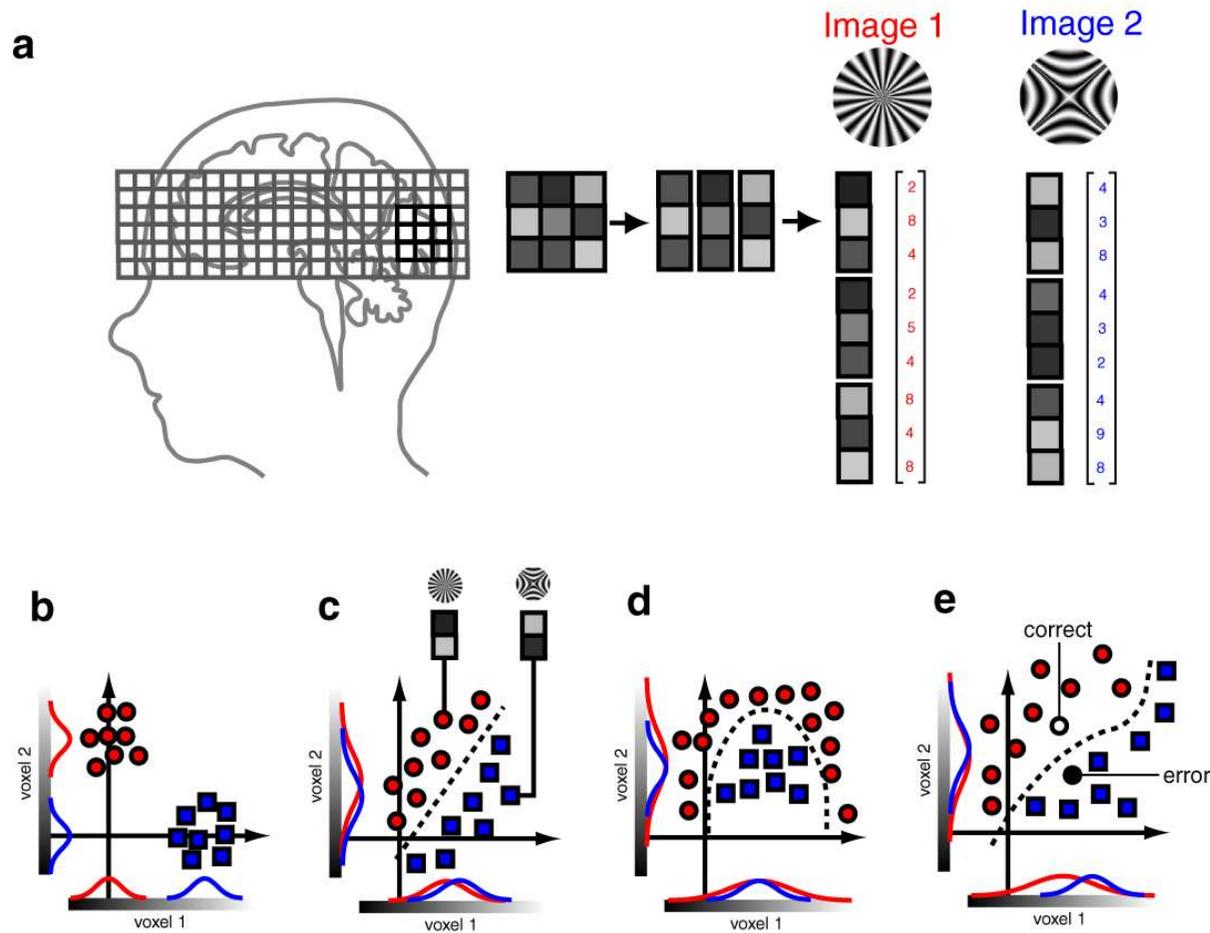
Et on la soumet à une régression multiple avec plusieurs descripteurs des stimuli



- Lorsqu'on n'y fait pas attention, le nombre est tout de même codé, mais faiblement
- Lorsqu'on y fait attention, le nombre est sélectivement amplifié par l'attention dans la voie visuelle dorsale
- le nombre est codé indépendamment des autres paramètres non-numériques
- Plusieurs régions codent le nombre, mais les régions pariétales semblent être particulièrement impliquées dans le jugement numérique, car leur représentation est sélectivement amplifiée en fonction de l'attention.



# Peut-on extraire le code plus directement? Le décodage multivarié



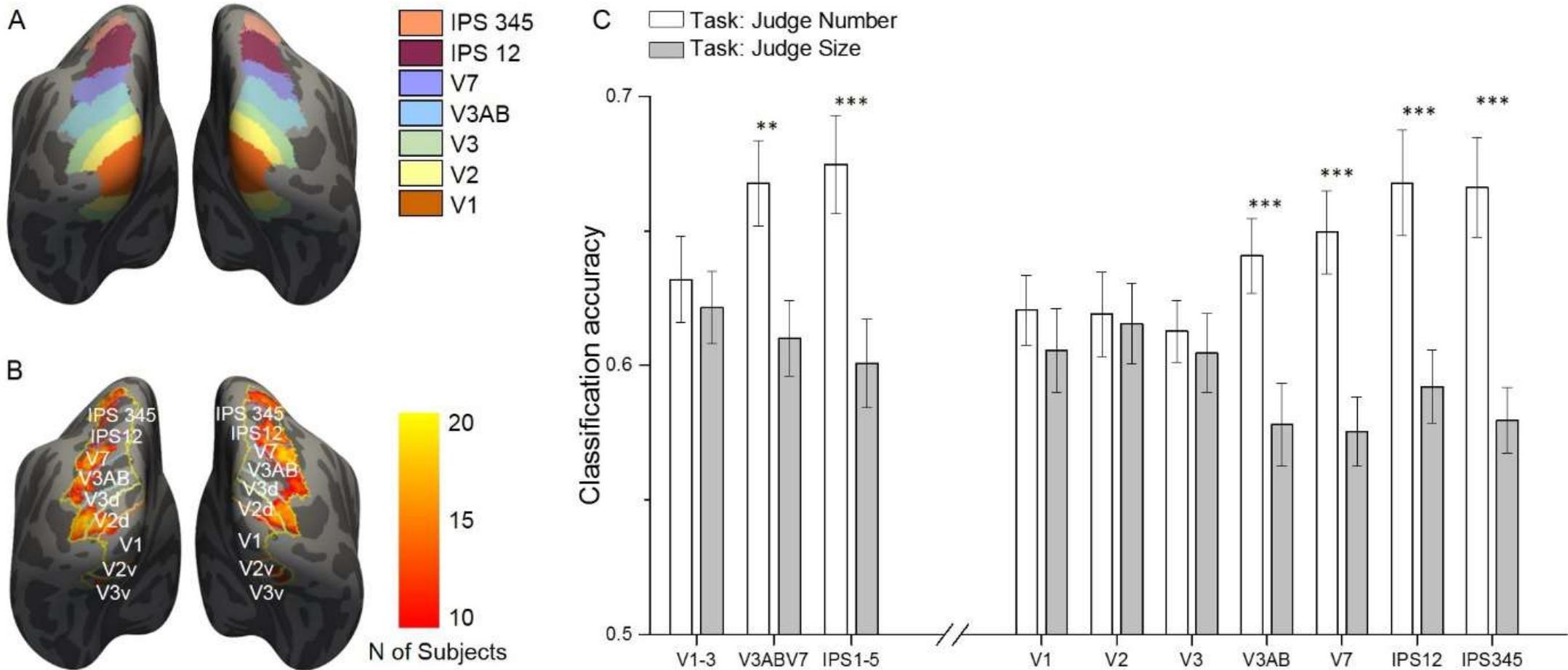
Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat Rev Neurosci*, 7(7), 523–534.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nat Neurosci*, 8(5), 679-685.

# Effet de l'attention sur le codage des nombres à 7T:

Des conclusions très semblables sont obtenues par décodage multivarié

Mais la méthode ne permet pas de décorréler si facilement les différentes variables (surface, etc)

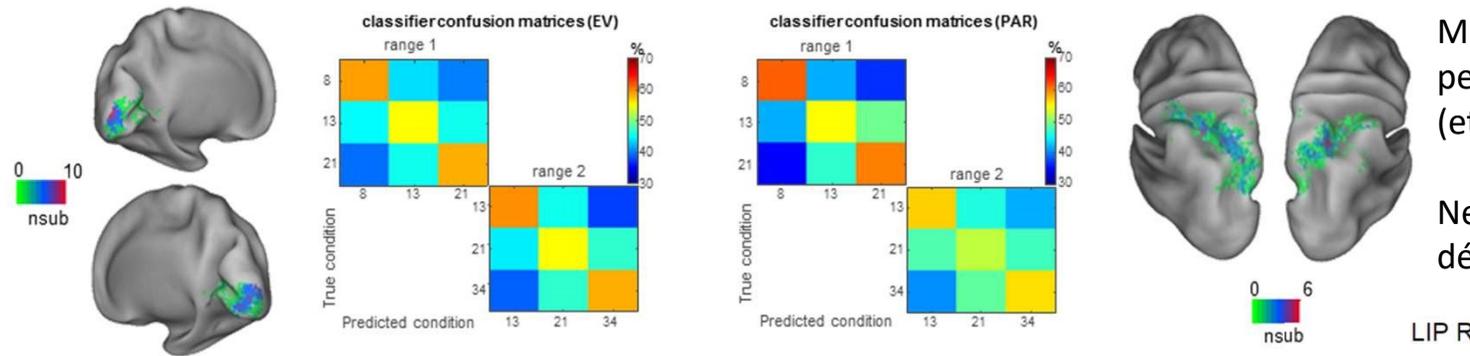


# La performance d'un décodeur multivarié peut être directement comparée au comportement

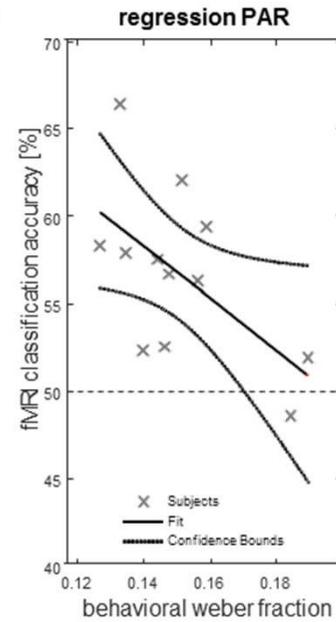
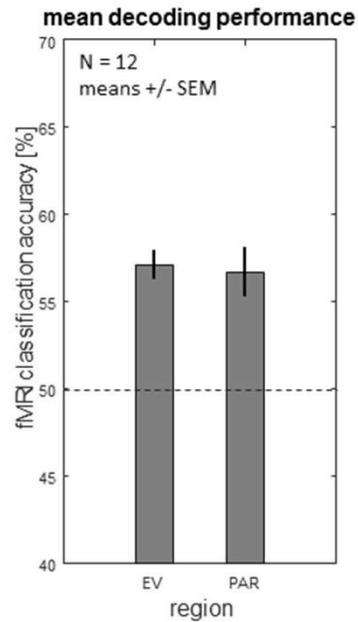
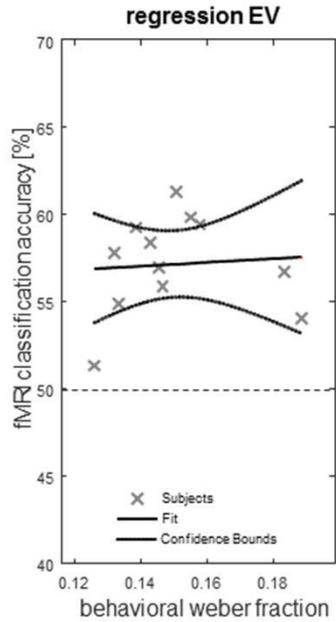
Lasne, G., Piazza, M., Dehaene, S., Kleinschmidt, A., & Eger, E. (2018). Discriminability of numerosity-evoked fMRI activity patterns in human intra-parietal cortex reflects behavioral numerical acuity. *Cortex*. <https://doi.org/10.1016/j.cortex.2018.03.008>

On peut décoder les quantités dans V1 et dans IPS. Mais seul le décodage dans IPS corrèle avec la performance des sujets. (et particulièrement dans l'aire LIP droite)

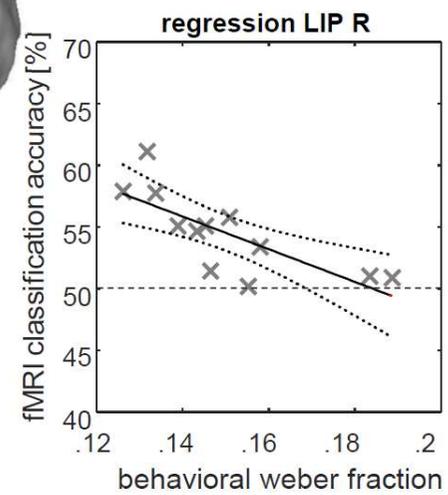
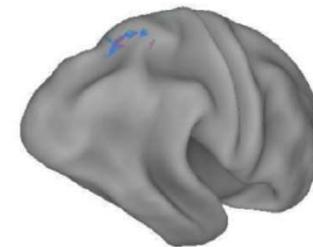
Ne jamais oublier que, ce que nous parvenons à décoder, le cerveau ne l'utilise pas forcément !



$r = -0.07$   
n. s.



$r = -0.59$   
 $p < .05$



$r = -.74$   
 $p < .01$

# Méthodes pour la mesure des représentations cérébrales en IRM fonctionnelle

Davis, T., & Poldrack, R. A. (2013). Measuring neural representations with fMRI: practices and pitfalls. *Annals of the New York Academy of Sciences*, 1296(1), 108–134. <https://doi.org/10.1111/nyas.12156>

## Visualisation globale des circuits impliqués:

La méthode de soustraction

- Mais problème de spécificité: difficulté de l'inférence inverse

## Visualisation indirecte des représentations et des codes neuronaux

- L'analyse multivariée (MVPA = *multivariate pattern analysis*)
- L'analyse de la similarité des représentations (RSA = *representational similarity analysis*)
  - Et l'une de ses applications: l'hyper-alignement
- La méthode de décodage multivarié

## Visualisation de plus en plus fine et directe des codes

- La méthode d'amorçage ou d'adaptation
- La modélisation du profil (*tuning curve*) de chaque voxel
- L'imagerie de haute résolution
  - Colonnes corticales
  - Couches du cortex (*laminar or layer-based fMRI*)

