

# Corrélations fantômes en Neurosciences sociales

Soumis par Stephane Desbrosses

Depuis une dizaine d'années, l'observation des corrélations entre mesures obtenues à l'IRM et traits de personnalité ou émotions constitue un champ de recherche en expansion. Les plus grands journaux ont participé à la publication de telles études de neurosciences sociales, études dont une partie conséquente comporte un biais jetant le doute sur leurs résultats, selon une récente recherche à la critique acerbe.

Pire que cela, les conclusions de l'étude de Vul et ses collaborateurs[1], semblent bien s'étendre à d'autres champs de recherche pour lesquels l'IRM est utilisé. Cette étude, sous presse, est partie du constat simple selon lequel de nombreuses recherches, mettant en lien l'activation de régions cérébrales localisées avec des mesures psychologiques plus traditionnelles, présentent comme résultats des indices de corrélations particulièrement élevés... A tel point que cela pourraient bien en être trop beau pour être vrai. Neurosciences sociales et cognitives : un champ prolifique Si la recherche en neurosciences sociales utilise de nombreuses méthodes dans ses études (études de cas/lésions, électrophysiologie, recherches sur l'animal...), l'une de ces méthodes a la faveur d'une grande partie des recherches. Elle consiste à établir la corrélation entre des mesures d'activité cérébrale via l'IRM fonctionnelle, et les différences individuelles observées dans les mesures d'entités psychologiques, de personnalité, d'émotions ou encore de comportements sociaux. Vul et ses collaborateurs désignent trois études pour illustrer leur propos : Eisenberger et ses collaborateurs (2003), montrent dans Science que la détresse ressentie après un rejet, est corrélée à .88 à une suractivation du cortex cingulaire antérieur (CCA) ; une corrélation "parfaite" serait de 1. Un an plus tard et dans le même journal, Singer et al trouvaient une corrélation s'élevant à .52 et .72, entre l'activité du CCA et de l'insula gauche, avec les niveaux d'empathie mesurés à l'aide de deux échelles psychométriques. Dans NeuroImage, Sander et al (2005) confirmaient quant à eux une corrélation record s'élevant à .96, entre l'activité du cuneus (aire 17 de Brodmann dans le lobe occipital!) et une tendance à des réactions anxieuses. Statistiques de la fiabilité Le problème, c'est que les mesures obtenues via les tests psychométriques, autant qu'avec l'IRM, n'expriment elles mêmes que des résultats statistiques, non des résultats sûrs à 100%... Par exemple, les fidélités test-retest d'échelles psychométriques, dépendent largement du nombre d'items, de la fiabilité de chacun d'entre eux. Même une échelle très utilisée et analysée de nombreuses fois présente rarement des corrélations de mesure test-retest supérieures à .80. Quant à l'IRM, celui-ci, selon les auteurs, dépasse difficilement .70. En conséquence et en toute logique, la valeur de corrélation maximum[2] que l'on devrait observer dans toute étude liant l'activité IRM à un test psychométrique, ne devrait pas s'élever à plus de  $\sqrt{0.7 * 0.8}$ , soit .74, même si les différences aléatoires entre deux mesures étaient nulles. Nota : voir Nunally, 1970 : la mesure d'une corrélation entre deux mesures dont les fidélités test-retest ont elles mêmes une corrélation inférieure à 1 peut se calculer de la façon suivante :  $r(A, B) = r(A \text{ mesuré}, B \text{ mesuré}) * \sqrt{r(A) * r(B)}$  ; Selon Vul, Pashler et leurs collaborateurs, de nombreuses études souffrent de cette confusion, suffisamment pour présenter des résultats douteux. Lors d'une méta-analyse portant sur 52 articles, présentant 256 corrélations significatives, ces chercheurs montrent que 54% de ces corrélations se basent sur une méthodologie présentant une erreur de non-indépendance : les auteurs de ses études ont choisi de se focaliser sur les voxels (unités de volumes d'une image 3D IRM), précédemment calculés comme probablement actif (selon un seuil spécifique standard ou choisi) pour établir la correspondance entre les voxels "allumés" et la mesure psychométrique.

Cela se complique, et il est nécessaire de rappeler le fonctionnement classique de l'IRM pour bien comprendre ce qu'il se passe : pour obtenir une image IRM, deux mesures sont réalisées, l'une en situation de repos, normale, l'autre en situation de tâche cognitive ou lors d'une réaction affective. Chaque image est enregistrée sous forme de volume composé de voxels approchant le mm<sup>3</sup>. Le postulat est le suivant : si on "enlève" à la seconde image, la première, on obtient une cartographie des zones cérébrales qui se sont activées spécifiquement pendant la tâche ou la réaction concernée. Pour cela, chaque voxel est comparé à son double et subit un test statistique répondant à la question "ce voxels s'est-il significativement plus "allumé" lors de l'activité cognitive que lors du repos?". Il s'agit bien d'un test statistique, non d'une mesure à 100% fiable! Par la suite, l'image est recomposée en utilisant les voxels dont le seuil statistique est suffisamment élevé. Sur les quelques centaines de milliers de voxels observés, seule une partie est conservée, en fonction du seuil standard ou choisi, puis "greffée" sur une image anatomique (par ailleurs, le plus souvent, standard...). A la suite de cette première mesure, et principalement si l'on souhaite tenter de corréler une région cérébrale avec, par exemple, une mesure psychométrique, il est nécessaire de regrouper ou sélectionner les voxels "pertinents". On peut choisir d'observer les voxels en fonction d'une anatomie précise (ceux qui se trouvent dans la région du cortex cingulaire, par exemple), ou en fonction de leur activité à l'IRM fonctionnelle, ou bien encore, les filtrer selon les deux critères. Ces voxels sélectionnés sont ensuite testés afin de déterminer la corrélation qu'ils entretiennent, pour chaque sujet, avec la mesure psychométrique. Cette pratique souffre d'un biais conséquent : à partir d'une simulation IRM totalement aléatoire, si l'on sélectionne les voxels sur un critère fonctionnel, selon un seuil de corrélation avec la mesure psychométrique étudiée, qu'on les regroupe ensuite afin de confirmer cette corrélation, il n'est alors pas étonnant que celle-ci se montre très élevée. Nous n'aurons gardé, virtuellement, presque que les voxels qui confirment l'hypothèse de départ... Ce biais nommé erreur de non-indépendance, comme le montre l'équipe à l'origine de cette étude, peut aboutir à découvrir des corrélations là où il n'y en a aucune. En tous les cas, il augmente sensiblement cette corrélation. Les résultats obtenus dépendent des critères de sélection ; cette erreur classique n'est pas spécifique aux mesures IRM, et on la retrouve dans d'autres types de mesures, dans d'autres domaines de la psychologie, dans d'autres disciplines que la psychologie. Sur 52 articles étudiés, et publiés, 28 se sont révélés comporter cette erreur de non-indépendance. Le schéma ci-contre, produit par l'équipe, montre les corrélations selon un code couleur correspondant à leur méthodologie. On voit aisément que les corrélations découvertes à partir d'une

méthodologie comprenant l'erreur de non-indépendance sont bien plus fortes qu'avec une méthodologie plus rigoureuse. On pourrait rétorquer que, si ces fortes corrélations sont surestimées, elles n'en traduisent pas moins des corrélations réelles. Vul et ses collaborateurs estiment quant à eux, que de tels résultats sont virtuellement vides de signification. Si à partir d'un échantillon aléatoire composé de bruit, on peut arriver à une corrélation forte, alors chaque corrélation présentant l'erreur de non indépendance pourrait tout aussi bien surgir d'une coïncidence, et se révéler en définitive fausse.

Comment se prémunir de cette erreur? Les auteurs proposent plusieurs solutions : choisir les voxels que l'on va sélectionner avant de calculer leur corrélation avec la mesure psychométrique, ou choisir les voxels hautement corrélés sur la moitié des sujets, puis vérifier cette corrélation à partir des données issues de l'autre moitié, sur les voxels précédemment choisis (méthodologie split-half).

Cet article, plutôt critique, n'a pas tardé à susciter des réponses plus ou moins vindicatives. Matthew Lieberman, l'un des collègues d'Eisenberger, et dont l'étude fut reprise en exemple, prépare actuellement une réponse à Vul et ses collaborateurs en vue de diminuer l'impact de Voodoo correlations in Social Neurosciences, L'article de Vul et ses collaborateurs. Ed Vul lui même a pris part (voir la réponse de Jabbi et al, ainsi que la seconde réponse de Vul) à ce qui s'annonce comme une mini bataille sur le terrain de la méthodologie. Quoi qu'il en soit, l'article fait déjà beaucoup parler de lui chez nos confrères anglophones, nul doute qu'il va entretenir une bénéfique réflexion sur l'importance de la méthodologie dans le champ de la recherche, en neuroscience et en psychologie. Sources : [1] Edward Vul, Christine Harris, Piotr Winkielman, Harold Pashler (2009). Voodoo Correlations in Social Neuroscience. Perspectives on Psychological Science. In Press. (voir aussi l'article dédié de Edward Vul)[2] Nunally JC. 1970 Introduction to psychological measurement. New-York ; McGraw-Hill. 1970.