Introduction

Notre compréhension du fonctionnement cérébral est encore balbutiante. Confronté à la multitude des neurones savamment agencés en un réseau inextricable, nous butons sur la barrière de la complexité. Autrement dit nous nous trouvons face à un système dont le comportement est souvent reconnu comme déterministe mais non prédictible (chaotique).

Comme pour toutes les approches scientifiques, nous nous sommes adaptés en fragmentant le problème. Ainsi il existe de nombreuses échelles anatomo-fonctionnelles : la molécule, la synapse, le neurone, le groupe local (la colonne corticale), l'aire cérébrale, ou le cerveau. Malheureusement ils sont devenu autant de niveaux de description autonomes. Cet enfermement provenant du fait qu'aucune règle ne semble permettre de déduire le comportement d'un niveau n+1 sur la base de la connaissance que nous avons du comportement des éléments du niveau n.

Pourtant l'objet de cette thèse, "l'intégration fonctionnelle cérébrale" (IFC) cherche bien à décrire la coordination de l'activité des éléments de base du cerveau et l'activité mentale cohérente qui en résulte ainsi que les mécanismes en jeu.

Etant donné ces remarques préliminaires, il va de soit que ce domaine ne peut être abordé vierge de tout instrument conceptuel. Ce sont de ces instruments dont nous débattrons dans la première partie de ce travail. L'approche sera donc épistémologique en partant de l'étude du courant de pensée dominant en imagerie fonctionnelle qu'est le cognitivisme. Nous verrons les problèmes qu'il pose, tout particulièrement pour l'étude de l'IFC, et nous tenterons d'y répondre en proposant une alternative "neurodynamique".

Dans un second temps, nous aborderons les instruments d'enregistrement de l'activité cérébrale, puis les méthodes d'analyse de ce type de données dans le cadre de l'IFC. L'objectif est à la fois d'illustrer les concepts développés en première partie et de faire l'inventaire critique des méthodologies employées. Nous verrons combien l'interprétation des résultats issus de ces instruments d'analyse est dépendante d'un cadre conceptuel qu'elle ne semble pas permettre de tester.

J'aimerais soulever de suite 2 limites au développement qui va suivre.

En premier lieu, l'IFC sera surtout abordée dans le cadre de l'imagerie fonctionnelle cérébrale. C'est en effet dans ce domaine que cette question a réémergé. Une des raisons en est que les techniques d'imagerie fonctionnelle peuvent réaliser l'enregistrement concomitant

de l'ensemble de l'activité cérébrale. Elles nous permis alors de prendre la mesure du nombre de régions cérébrales impliquées dans une tâche aussi élémentaire soit elle. Il est devenu évident que les réponses à la question du "qui fait quoi" (ségrégation fonctionnelle), soulevait la question de savoir "comment tout cela marche ensemble" et peut rendre compte du comportement observé (intégration fonctionnelle). Enfin c'est parce que nous disposons à présent de techniques d'enregistrement de l'ensemble du cerveau que l'étude de l'IFC devient possible.

La seconde limite tient au rapprochement qui sera fait entre IFC et conscience. Il nous a semblé que de ne pas l'évoquer n'aurait pas permis de saisir toute la portée du débat. L'étude épistémologique en particulier aurait été dénaturée puisque la plupart des avancées semblent venir des équipes intéressées par l'étude de la conscience.

Enfin, et bien que ce point ne soit pas développé plus avant au cours de cette thèse, soulignions l'intérêt qu'il peut y avoir à l'étude de l'IFC pour le clinicien. En effet, les processus sous-tendant cette intégration peuvent probablement être défaillants et occasionner des troubles. Au rang des maladies susceptibles de concerner les processus intégratifs, citons :

- La maladie d'Alzheimer, pour laquelle on pense que la perte neuronale est trop modeste pour expliquer l'importance des troubles. En revanche ces troubles se comprennent mieux si on se rappelle que la perte en neurones mais surtout en connexions concerne pour l'essentiel les liaisons longues distance permettant l'IFC à large échelle. Le fait que les premières connexions à souffrir soient celles du cortex entorhinale sur le gyrus dentelé (les perforantes de l'hippocampe) explique que la première fonction à décliner soit la mémoire épisodique.
- La confusion mentale est une entité syndromique qui répond à de multiples causes. Pour la plupart celles-ci concernent le système ponto-meso-diencéphalique impliqué dans la coordination de l'activité cérébrale. L'atteinte de ce système central est à l'origine de la désorganisation générale de l'activité mentale dans la confusion. Ce tableau clinique nous semble être particulièrement illustratif de l'importance des processus intégratifs dans le fonctionnement cérébral normal.
- La schizophrénie est aussi une entité syndromique pour laquelle une atteinte plus limitée des processus intégratifs pourrait sous-tendre l'expression symptomatique. L'intégration entre l'activité des régions frontales et des régions temporo-pariétales serait particulièrement susceptible d'être touchée et pourrait expliquer en particulier l'émergence d'hallucinations auditivo-verbales.

- Enfin les atteintes de la substance blanche (maladies démyélinisantes, lésions axonales diffuses post traumatique…), ont toute probabilité d'interférer avec les processus intégratifs de façon plus ou moins localisée. Une bonne intégration passant par une bonne communication entre les différentes régions cérébrales. Une part de l'expression symptomatique de ces maladies découlerait aussi d'un dysfonctionnement des processus intégratifs .

Remarque pour le lecteur :

Nous avons décidé de faire en sorte que les connaissances nécessaires à la compréhension du texte soit disponibles. Cependant il ne fallait pas qu'elles gênent le développement des idées.

Ainsi pour première partie nous feront référence à certains concepts épistémologiques qui nous a semblé préférable de préciser. Il faut souligner combien certains termes doivent être compris dans une acceptation bien précise parfois différente du sens commun. Une définition est alors proposée en encart au fur et à mesure que ces termes apparaîtrons dans le corps du texte. En revanche, les remarques non essentielles à la compréhension du texte seront renvoyée en notes de bas de page.

Dans la seconde partie, l'étude mathématique précise des instruments analytiques nous a paru superflue d'autant que cette thèse ne s'adresse pas à des spécialistes de l'analyse du signal. Nous avons donc plutôt opté pour une présentation du concept de l'outil et de l'idée qu'il y a derrière en rejetant en annexe la partie mathématique. Celle-ci ne sera par ailleurs développée que si l'instrument est peu connu.

Pour une approche épistémologique de l'intégration fonctionnelle cérébrale

Pourquoi cette approche épistémologique ?

Peut-être parce que novice dans le domaine de l'imagerie fonctionnelle, certains de ses aspects me posaient problème. Ces vagues impressions de malaise pouvaient aller jusqu'à de véritables difficultés de compréhension. Celles-ci découlaient souvent de mon attachement à traduire ce nouveau langage en des termes neurobiologiques qui m'étaient plus familiers, déformation sans doute liée à ma formation de neurologue.

Ainsi que l'avait proposé l'épistémologiste Donald Norman (1), le seul recours pour dépasser ces difficultés était de revenir sur les concepts de base, aussi bien les miens que ceux que j'essayais de comprendre. Un travail dans lequel l'épistémologie s'est révélée être d'une aide incontournable, par la mise en perspective qu'apporte une étude historique et ses instruments conceptuels.

Une de ces difficultés a émergé alors que je tentais de rapprocher deux instruments d'analyse de l'intégration fonctionnelle cérébrale (IFC) qu'étaient les synchronies et les modèles d'équations structurelles. Là encore, le problème venait des concepts de bases. Ces instruments n'étaient pas comparables car ne procédant pas du même courant de pensée (cognitiviste ou neurodynamique).

Ce retour aux concepts m'a permis de transformer ces "malaises" en critiques de la démarche cognitiviste majoritaire dans le domaine de l'imagerie fonctionnelle. Celles-ci prennent toute leur ampleur dans le domaine de l'intégration fonctionnelle cérébrale, d'où le choix du sujet. J'espère montrer que l'abord neurodynamique est par certains égards plus satisfaisant sans que le niveau de description cognitiviste perde pour autant tout son sens. J'insisterai non seulement sur les aspects pratiques, mais aussi sur le problème théorique de l'abord de la conscience que cette approche permet de simplifier.

Bien que présenté comme une alternative, l'approche neurodynamique doit être plutôt comprise comme une description qui se veut plus complète, mais qui englobe la description cognitiviste.

Ce qui va suivre est issu de cette réflexion qui débouche directement sur des implications pratiques. Elle n'a malheureusement pas pu s'appuyer sur des références adressant cette question. La seule réflexion épistémologique en imagerie fonctionnelle ne concerne que

l'approche ségrégationniste classique et non son approche intégrative trop récente. Ce sera ma seule excuse pour les éventuelles erreurs qui pourront être relevées à par la suite.

Le courant cognitiviste

De la cognition au cognitivisme

L'imagerie fonctionnelle cérébrale s'inscrit dans la mouvance des sciences cognitives.

Mais qu'est-ce que la cognition ?

Ce terme de cognition était déjà utilisé par la philosophie scolastique où il désignait la faculté de connaître. Il nous est revenu dans son acceptation actuelle du monde anglo-saxon. Le terme de cognition dérive du mot latin *cognitio -onis* (connaissance). Il désigne la connaissance comme objet de recherches. Cela comprend aussi bien la connaissance en tant que telle, que comme l'ensemble des fonctions qui la réalisent (2;3).

Plus concrètement les objets de la cognition concernent des activités mentales supérieures comme la perception, la mémoire, le langage, l'attention, le raisonnement, la planification ou la coordination motrice etc... En psychologie, le domaine de la cognition est opposé à celui de l'affectivité.

La cognition est un objet recherche transdisciplinaire. Ce fut d'ailleurs historiquement le premier domaine de recherche où l'approche interdisciplinaire a réussi (2). Les principales disciplines concernées sont les neurosciences, l'intelligence artificielle, la logique, la psychologie cognitive, la linguistique, l'anthropologie et plus récemment sous l'influence des neuro-philosophes[1], la philosophie de l'esprit (4).

Selon cette définition des sciences cognitives par leur objet, le terme pourrait tout aussi bien s'appliquer à des recherches menées bien avant le milieu du XXe siècle. Nous prendrons pour exemple l'étude du langage par les premiers aphasiologistes dès le XIXe siècle. Pourtant ce terme n'est pas employé pour désigner les recherches de cette période (5). Pourquoi ?

Il existe en fait une certaine confusion entre l'objet de recherche "cognition" et le courant de recherche "cognitiviste". Or un courant scientifique signifie bien plus que l'objet auquel il s'intéresse. Il a une histoire, il a un programme de recherche et il utilise des paradigmes. Autant de points que nous allons aborder dans la perspective de l'IFC.

---

[1] Les neuro-philosophes cherchent à appliquer les découvertes des neurosciences aux questions philosophiques classiques (*"neurophilosophy"*). C'est une démarche différente de la philosophie des neurosciences qui est la démarche épistémologique qui s'intéresse au fondements conceptuels des neurosciences (*"philosophy of neuroscience"*).

“Cognitivisme” est le néologisme formé à la fin des années 70 par le philosophe John Haugeland pour désigner une certaine doctrine concernant la nature des phénomènes mentaux constituant un programme de recherche (6).

Les fondements du courant cognitiviste

Ce courant de pensée est né dans les années 50. Il prend ses racines dans un contexte socioculturel particulier : le *Massachusetts Institute of Technology* (MIT), l’*Institute of Advanced Study* de Princeton où sont rassemblés les élites américaines de l’ingénierie et de la technologie (2). De ce vivier bouillonnant sortiront en quelques années seulement de nouvelles théories initialement appliquées aux avancées de l’automatique et des communications issues de l’effort de guerre. Citons quelques unes des avancées qui influenceront le courant cognitiviste :

La cybernétique, ancêtre de la théorie des systèmes, est la source principale d’inspiration du cognitivisme qui en a conservé les schémas (cf. les théories du contrôle de l’action par Christopher Frith). C’est la science de l’action orientée vers un but. Elle se fonde sur l’étude des processus de commande et de communication entre les systèmes technologiques. Son développement initial concernait l’asservissement du tir des canons par les systèmes radars. Ce qui est important c’est la diffusion de ses concepts. Ceci a été entre autre l’œuvre de Norbert Wiener qui mettait en évidence l’identité structurelle des processus contrôlés. Ces concepts ont alors pu trouver de nouveaux terrains d’application comme en biologie avec les systèmes de régulation.

La diffusion de ces idées à la psychologie, la physiologie, la linguistique, la sociologie ou à l’anthropologie s’est surtout faite au cours de conférences multidisciplinaires comme les conférences Macy (1946-53). L’instigation de ces conférences a été un phénomène raisonné en particulier de la part de Wiener et McCulloch qui espéraient fonder un projet matérialiste du mental au physique (la “première cybernétique” – 1943).

Autre source d’inspiration a été la **théorie de l’information,** développée au même moment mais indépendamment par Claude Elwood Shannon (7). Un des apports de cette théorie a été de dissocier le concept de signification de celui d’information. Ainsi une information peut être traitée par un système même s’il ne la comprend pas comme un ordinateur par exemple.

Enfin il faut mentionner toute l’importance qu’a prise la **logique**, ainsi que son héritière directe qu’est l’informatique. Alan Mathison Turing propose en 1936 la première machine programmable théorique. Cette machine de Turing, est supposée capable d’effectuer tout type

de traitement logique qu'elle effectue par une manipulation de symboles. Il n'était plus nécessaire de concevoir une machine par type de traitement, c'était une machine "universelle". La fonction logique devient ainsi indépendante de son support physique et Turing proposera dès 1948 l'utilisation de sa machine comme moyen de simuler les fonctions mentales (8). Il a même proposé un test pour estimer "l'intelligence" de cette machine. Ce test de Turing était un jeu d'imitation. Une machine était reconnue comme intelligente si un observateur humain était incapable de la distinguer d'un autre humain après avoir répondu à ses questions.

C'est donc à l'interface de plusieurs disciplines que se construit le courant cognitiviste. Il ne s'agit pas cependant d'une véritable mise en commun des concepts. C'est plus une appropriation des concepts et instruments de la logique et de la cybernétique par les sciences "moles" comme la psychologie, les neurosciences ou l'anthropologie.

Le programme de recherche cognitiviste

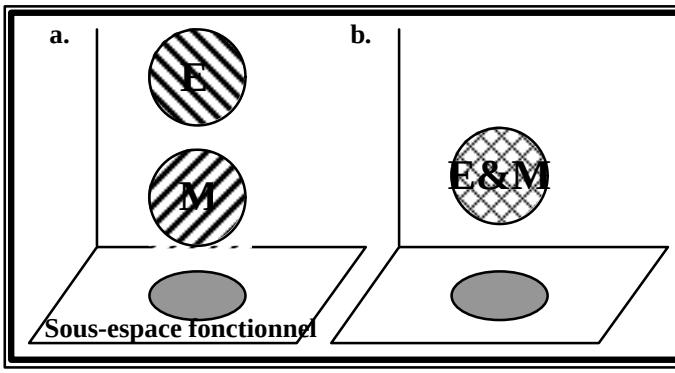
Selon Larry Laudan (1977), toute recherche scientifique s'inscrit dans un "programme de recherche". Celui-ci est un ensemble d'hypothèses de travail partagées par la communauté des chercheurs qui y participent, qui ne sont pas directement testables dans le cadre du programme. Ces hypothèses concerne la "substance" des objets étudiés (métaphysique) et la méthodologie pour les aborder sur le plan théorique et empirique. Les théories n'existent pas par elles même, mais seulement dans le cadre d'un programme de recherche. Il en découle que pour proposer une théorie qui est incompatible avec un programme de recherche, le chercheur n'a d'autre possibilité que de rejeter le programme.

Il est malheureusement assez fréquent que ce programme de recherche soit implicite. Notre démarche cherche donc à expliciter le programme de recherche cognitiviste pour mettre en évidence son incompatibilité avec une certaine vision de l'intégration fonctionnelle cérébrale. Mais cette dernière ne peut exister hors d'un nouveau programme de recherche qu'il nous faudra définir.

Pour Francisco Varela "le programme de recherche cognitiviste peut se résumer ainsi : la cognition est un traitement d'informations, c'est-à-dire une manipulation de symboles selon un ensemble de règles. Elle peut être effectuée par n'importe quel dispositif capable d'entretenir et de manipuler des éléments discrets appelés symboles. Le système n'appréhende que la forme des symboles et non leur signification... (Il) doit déboucher sur une bonne solution au problème posé."(3)

Reprenons les principales propositions :

1) Le programme s'éloigne de l'idée du monisme matérialiste de la "première cybernétique" (le mental est un produit du monde physique). Le cognitivisme reconnaît une difficulté à rapprocher des substances (mentale et physique) de natures si différentes. Il propose de ne s'intéresser qu'à leur produit commun : la cognition. C'est donc un point de vue fonctionnaliste : seule la fonction est prise en compte. On ne peut qu'admirer la pirouette philosophique qui permet, par un changement de cadre, le rapprochement du mental et du physique à la manière d'une projection sur un même plan (cf. encadré). On ne résout pas la question philosophique, mais on aboutit à un concept opératoire. Une belle illustration de cette vision fonctionnaliste est le test de Turing. La projection sur le sous-espace fonctionnel entre un homme et une machine intelligente y tient la même place, mais le retour sur la sphère du mental est imprécis puisqu'elle ne permet pas de connaître par exemple la part consciente ou inconsciente de ces opérations. Notons enfin que dans le cadre fonctionnaliste, le seul sens reconnu à la causalité est celui de causalité efficiente[2].

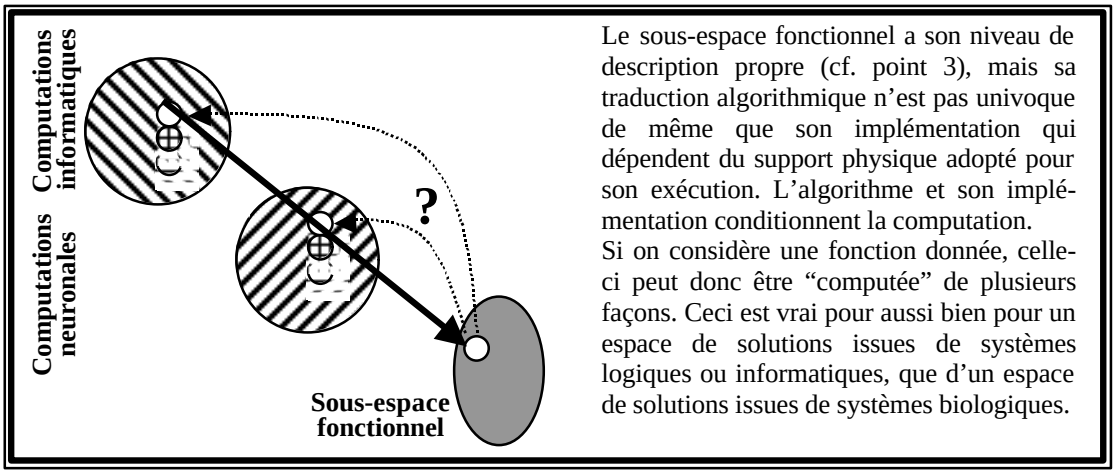


La positon dualiste (a) est illustrée dans un espace imaginaire séparant esprit et matière, alors que la position moniste (b) illustrée dans le même espace ou esprit et matière ne sont envisagés que comme deux niveau de description différent d'un même objet. La position fonctionnaliste limite son raisonnement à un sous espace commun (~projection) qui est la même que ce soit pour une position moniste ou dualiste.

2) Le traitement est indépendant de son support physique (Turing), il en découle que les fonctions cognitives peuvent être réalisées sur n'importe quel support. Si on considère que le cerveau est une machine à traiter de l'information, il est donc possible de reproduire ce traitement à l'aide d'un ordinateur par exemple. Projeté dans le sous-espace fonctionnel le cerveau et l'ordinateur pourraient ne pas être différenciables. Cependant cela ne veut pas dire qu'ils sont identiques. En effet la façon dont ils résolvent les problèmes est bien différente.

[2] Il existe 4 types de causalité définit par Aristote : pour une statue par exemple il y a une cause matérielle (la pierre), une cause formelle (sa forme), une cause efficiente (le sculpteur) et une cause finale (dans quelle but elle a été faite). Celle que nous utilisons ici en est l'acceptation la plus fréquente, celle de causalité efficiente. Elle répond à la question "pourquoi…" par "parce que…". La cause est extérieure à l'effet et le précède. Elle est par nature mal adaptée à la dépendance circulaire (un système influence l'autre).

La “computation” est la façon dont la fonction est réalisée, alors que la fonction c’est l’objectif que réalise cette computation. Il est théoriquement possible de parvenir au même objectif en utilisant des stratégies (des computations) différentes. C’est la notion de dégénérescence. Ce concept a été importé par Edelmann du système de traduction ARNm – protéine, mais était déjà connu des physiciens. De la même façon qu’un acide aminé est codé par plusieurs triplets, une même fonction peut être réalisée par différents systèmes. On parle de systèmes dégénérés parce que le lien qui les unis avec la fonction n’est pas une bijection mais une surjection : si connaissant la “computation” on peut en inférer la fonction, à l’inverse, connaissant la fonction on ne peut pas en inférer la “computation”. Celle-ci s’opère dans le monde physique, il faut donc faire une sorte de projection inverse (cf. shéma).



Le sous-espace fonctionnel a son niveau de description propre (cf. point 3), mais sa traduction algorithmique n’est pas univoque de même que son implémentation qui dépendent du support physique adopté pour son exécution. L’algorithme et son implémentation conditionnent la computation.
Si on considère une fonction donnée, celle-ci peut donc être “computée” de plusieurs façons. Ceci est vrai pour aussi bien pour un espace de solutions issues de systèmes logiques ou informatiques, que d’un espace de solutions issues de systèmes biologiques.

Il faut remarquer que le programme cognitiviste a une inégalité fondamentale. En effet, s’il s’intéresse à la projection inverse de la fonction sur l’espace des systèmes physiques qui la réalise, il n’en est pas de même pour la projection sur l’espace du mental qui en fait l’expérience. Tout se rapporte au monde physique, et le monde mental n’est défini que par les fonctions qu’il supporte. Comme nous le verrons, cette position n’est pas sans poser problème pour l’étude de la conscience. A moins d’en faire une entité fonctionnelle (module ou représentation cf. plus bas), l’absence de retour dans l’espace de l’activité mentale l’exclue du programme de recherche.

3) Le mental possède une autonomie conceptuelle ou logique qui permet et nécessite de l’étudier en tant que tel à un niveau propre de description qui porte sur la fonction des entités mentales (6). Cette nature fonctionnelle du mental relève de la compétence d’une nouvelle discipline, sorte d’extension de la logique, qui étudie les univers

informationnels régis par des transformations effectives c'est-à-dire algorithmiquement réalisables. Ce traitement s'effectue de façon séquentielle. Comme nous avons vu qu'il n'y a pas de projection en retour sur l'espace du mental, celui-ci n'aura que ce seul niveau de description. L'introspection ou la phénoménologie ne sont pas retenus comme paradigmes méthodologiques. S'ils restent certainement utilisés au niveau de l'intuition, ils ne font pas l'objet d'une analyse.

Le programme cognitiviste a donc deux objectifs :

i) Développer les paradigmes aptes à décrire l'espace fonctionnel.
ii) Explorer l'espace des solutions aussi bien informatiques (c'est le domaine de l'intelligence artificielle) que biologiques. Ce dernier point relève du domaine des neurosciences dont l'imagerie fonctionnelle cérébrale fait partie.

Ces deux points sont en fait liés. Malheureusement, le langage algorithmique apte à décrire les fonctions n'est pas indépendant du système physique qui la réalise. Par exemple, l'algorithme pour une machine à traitement sériel n'est pas le même que pour une machine à traitement parallèle. Il n'est donc pas étonnant que les paradigmes utilisés ne se recouvrent pas. C'est ainsi que la démarche en intelligence artificielle utilise directement le langage de la logique, alors que le langage utilisé pour appréhender les fonctions biologiques est plus inspiré du modèle cybernétique.

Les paradigmes cognitivistes

Dans la "structure des révolutions scientifiques" (1970), Thomas Kuhn appelle ***"paradigmes"***, des découvertes scientifiques universellement reconnues qui pour un temps fournissent à une communauté de chercheurs des problèmes types et des solutions.
Ultérieurement l'auteur reconnaîtra deux sens :

i) Dans une première acceptation, le paradigme fixe le manière correcte de poser et d'entreprendre la résolution de problèmes. C'est ce qu'il l'appellera encore "la matrice disciplinaire". Elle comprend aussi bien des généralisations symboliques que les métaphores.
ii) Dans une seconde acceptation, le paradigme désigne non pas ces entités représentatives elles-mêmes mais la manière dont celui qui est éduqué dans la discipline apprend à les reconnaître, à les isoler, à les distinguer. Ce second sens renvoie à la phase pratique de la formation du scientifique.

Ceci permet à Kuhn de distinguer 2 régimes de l'activité scientifique :

i) Le régime de ***"science normale"*** durant lequel le scientifique n'a pas à reprendre à chaque fois les questions à leur commencement. Il bénéficie d'un certain état des problèmes. Mais le paradigme guide aussi l'attention des chercheurs, et leurs indique quels faits sont importants et quels autres négligeables.
ii) Mais lorsque les problèmes résistent aux efforts répétés s'installe alors une situation de crise. Elle ouvre sur le deuxième régime de l'activité scientifique : ***"la révolution scientifique"***. Les paradigmes sont alors changés pour prendre en compte les phénomènes inexpliqués. Les exemples classiques sont tirés de l'histoire de la physique avec les révolutions quantique et relativiste.

L'idée n'est pas de reprendre tous les paradigmes du cognitivisme, mais simplement ceux qui continuent d'influencer l'imagerie fonctionnelle et tout particulièrement l'intégration fonctionnelle. Nous ne développerons en particulier que ceux nécessaires à la description d'une algorithmique biologique.

Cette description fait appel à deux concepts (6) :

i) Les états, qui sont encore appelés représentations ou symboles. C'est la façon dont l'information se présente à l'entrée d'une unité de traitement et dont elle en sort. Dans la conception cognitiviste, cette représentation ne signifie rien pour le système. Mais en fait, elle garde le plus souvent une signification pour un observateur humain. Il persiste donc un lien sémantique entre l'information et l'objet représenté. Ce n'est pas le divorce proposé par la théorie de l'information.
ii) Les processus, qui sont encore appelés modules ou fonctions ou opérateurs. Leur rôle est de réaliser un calcul sur la représentation entrante pour délivrer une nouvelle représentation qui doit pouvoir améliorer l'extraction de l'information pertinente. Ces modules sont organisés hiérarchiquement. Jerry Fodor avait parlé de modules, et leur avait attribué un certain nombre de propriétés (cf. encadré). Ces propositions ont

suscité de vives réactions et nombreux sont ceux qui ont rejeté ce modèle, malheureusement sans en proposer un autre.

Ce modèle peut néanmoins servir de base moyennant l'apport de quelques corrections. Ainsi les modules ne seraient pas fixes, ni déterminés génétiquement. Par exemple, au cours de son développement, l'enfant utilise pour le langage des régions cérébrales très larges avant de se latéraliser et de se restreindre à quelques sites bien précis. L'encapsulation n'est plus considérée comme étant aussi rigoureuse. Par exemple, les processus attentionnels peuvent influencer le traitement du module. Enfin les modules les plus élevés dans la hiérarchie sont considérés comme beaucoup moins indépendant. Ainsi le vocable de "processus" semble se substituer à celui de "module", tout particulièrement lorsqu'on aborde les fonctions exécutives.

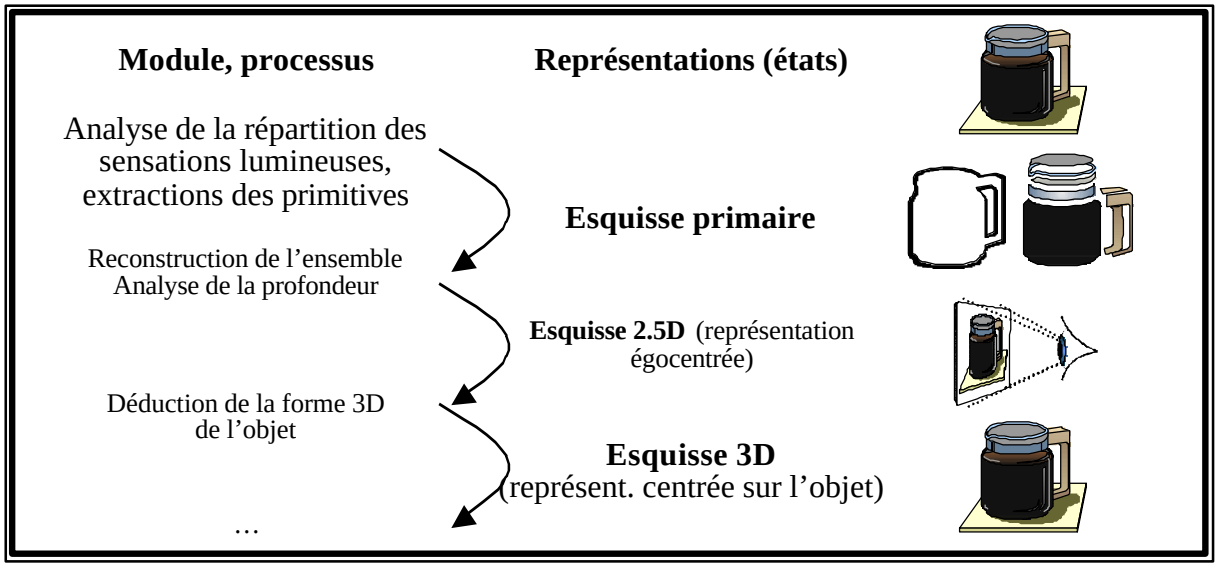
C'est le philosophe Jerry Fodor qui a décortiquer le paradigme du module cognitif ("*the modularity of mind*" – 1983). Il a proposé certaines propriétés devant être remplies pour pouvoir parler de module :

1. Ce sont les plus petites unités de traitement, non décomposable en unité plus petite.
2. Le traitement est automatique dès qu'un état approprié se présente à l'entrée.
3. Le traitement s'effectue toujours dans le même sens, d'un état de bas niveau vers l'état de haut niveau.
4. Ils sont spécialisés, connectés en dur (hardwired) et ne peuvent travailler que sur des entrées déterminées. Ils doivent donc avoir une localisation cérébrale précise, et leur "computation" est rapide.
5. Leur architecture est fixe, déterminée génétiquement.
6. Ils sont informationnellement encapsulés, leur "computation" n'est pas accessible aux autres parties du système et ne peuvent pas être influencée par elles. Ils sont donc indépendant.

Ces deux catégories d'objets cognitifs que sont représentation et processus se combinent pour réaliser des fonctions mentales complexes. Comprendre comment le cerveau réalise une tâche, c'est donc réaliser une décomposition fonctionnelle de celle-ci en terme de processus et de représentation devant s'enchaîner de manière sérielle avec néanmoins la possibilité de rétrocontrôle. La démarche idéale voudrait en fait que plusieurs modèles soient produits, et que l'on conçoive l'expérience qui permette de rejeter l'un d'entre eux.

Prenons l'exemple du modèle cognitif de David Marr pour la vision (cf. encadré). Dans notre exemple, une cafetière est présentée aux yeux du sujet. Il s'agit déjà d'une première représentation qui serait analogue aux pixels de l'image. Celle-ci passe dans le premier module d'extraction des primitives, d'extraction des surfaces, d'analyse de la luminosité et de

la couleur à une échelle aussi bien locale que globale (traitement de la forme). Il en ressort une nouvelle représentation appelée esquisse primaire. Celle-ci correspond aux éléments de base de l'image qui peuvent être considérés comme autant de symboles. Cette représentation passe en entrée d'un second module de reconstruction de l'ensemble et d'analyse de la profondeur. La nouvelle représentation qui en découle s'appelle l'esquisse 2.5D et correspond à la perception de l'orientation et de la profondeur des surfaces visibles du seul point de vue de l'observateur. Celle-ci se présente à l'entrée du module suivant qui va déduire les volumes de l'objet à la façon dont nous les reconstruisons à partir d'un dessin en perspective. La représentation qui en est issue, dite esquisse 3D est centrée sur l'objet. Le modèle veut qu'elle se présente alors à un module de reconnaissance. Le modèle est donc sériel. Les représentations se présentent en entrée des modules qui donnent de nouvelles représentations etc… Il s'agit d'une suite de processus pour passer d'une représentation à une autre. Il est possible d'inclure des boucles de rétroaction menant souvent à des comparateurs comme dans le modèle du contrôle moteur de Christopher Frith.



La notion de modules a connu une évolution. En effet l'approche logique de modules avait abouti à des implémentations de type système expert. Il est rapidement apparu que ceux-ci étaient inadaptés tout particulièrement pour les modules de bas niveau. Par exemple aucun système logique n'a pu réaliser les transformations représentationnelles du modèle cognitif de David Marr. Constatant les insuffisances de l'approche logique, les sciences cognitives se sont alors tournées vers les réseaux neuronaux (approche connexionniste)(3;9). Le premier réseau neuronal avait été proposé dès 1958 par Franck Rosenblatt (le "perceptron"). Malheureusement Marvin Minsky et Sigmour Pappert démontrèrent ses limites dans son application à la reconnaissance visuelle (sensibilité aux transformations géométriques de type

translation, rotation, variation de taille etc..) ce qui tua l'idée et renforça le règne de l'approche logique en intelligence artificielle pour les années 50 à 80. C'est la publication en 1982 d'un travail sur les verres de spin fait par un physicien (John Hopfield) qui donna le coup d'envoi à un renouveau de l'approche connexionniste. De nouveaux réseaux neuronaux furent inventés corrigeant certains défauts des précédents. Ainsi Net-Talk programmé par Terence Sejvnowski arrivait à réaliser une conversion grapho-phonologique (1985). Mathématiquement parlant ce que font ces réseaux est une simple transformation vectorielle : à partir d'un vecteur d'entrée ils calculent un vecteur de sortie. Paul Churchland a souligné l'analogie entre réseau neuronal et module et entre vecteur et représentation (4). L'approche connexionniste a ainsi amené une alternative à la notion de manipulation de symbole, tout en conservant la notion du découpage d'une fonction complexe en sous-fonctions réalisées en série par un ensemble de modules (réseaux neuronaux spécialisés). Le modèle effectue un traitement parallèle par morceaux qui s'enchaînent de façon sérielle.

L'imagerie fonctionnelle cérébrale
Instruments et interprétation

L'imagerie fonctionnelle cérébrale est une discipline récente. Elle regroupe les techniques d'enregistrement de l'activité cérébrale électrique, hémodynamique ou métabolique. En raison de leur caractère minimalement invasif, elles sont utilisables chez l'homme. Son objectif est de décrire les modifications de l'activité cérébrale en relation avec le comportement, aussi bien sur le plan spatial que temporel.

Il existe de nombreux instruments de mesure en imagerie fonctionnelle cérébrale humaine. Nous nous limiterons à l'exposé de deux d'entre eux car ils représentent les instruments les plus fréquemment utilisés et surtout les plus adaptés à l'étude de l'intégration fonctionnelle cérébrale.

L'IRM fonctionnelle – aspects techniques

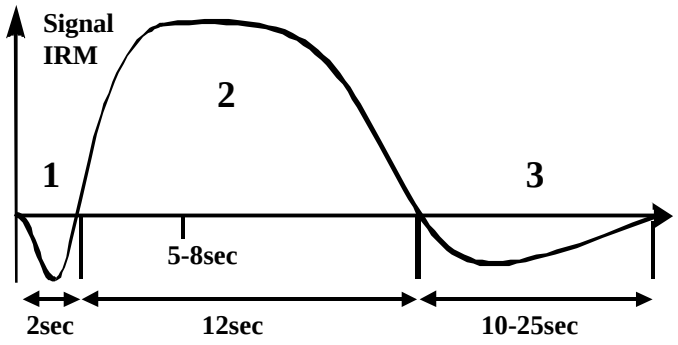
L'IRMf (Imagerie par résonance magnétique fonctionnelle) permet de mettre en évidence des modifications vasculaires en rapport avec l'activité cérébrale. On utilise le contraste BOLD (*Blood Level Oxygen Dependant*) mettant à profit les propriétés magnétiques de l'hémoglobine (Hb). Celle-ci se comporte différemment par rapport au champ magnétique en fonction de son état d'oxydation. L'Hb oxygénée est diamagnétique alors que les électrons libres de l'Hb désoxygénée lui confèrent des propriétés paramagnétiques à l'origine d'une perturbation du champ magnétique local (10).

Lors d'une augmentation de l'activité neuronale, on observe une augmentation de l'extraction d'oxygène, mais celle-ci est largement contrebalancée par une augmentation parallèle du débit et du volume sanguin cérébral local. Il en résulte une diminution de la quantité d'hémoglobine désoxygénée. Le champ magnétique local étant moins perturbé, il en résulte une augmentation du signal en T2* (contraste T2 avec rappel de l'écho par les gradients). Cette réponse BOLD survient ~2 sec après l'activité neuronale et atteint son maximum vers 8 sec pour les aires auditives (cf. figure). Il y a toutes les raisons de penser que certaines de ces caractéristiques sont variables en fonction des régions et des sujets.

La variation de signal observée est de faible amplitude, typiquement de l'ordre de 0.2 à 5 %. Il faut ainsi accumuler un nombre suffisant d'images au rythme d'1 toutes les 2 à 5 sec, alors que le sujet réalise une succession de tâches cognitives. Cette tâche peut être présentée par

blocs de plusieurs essais alternant avec des blocs d'une tâche contrôle, c'est le mode en créneaux, ou être présentée essai par essai, c'est le mode événementiel.

Décours temporel de la réponse BOLD



La dynamique de la réponse BOLD a pu être explorée en imagerie photonique dans le proche infrarouge :

1.Initialement, l'activité neuronale entraîne une augmentation de l'extraction d'oxygène d'où une élévation de la concentration en Hb désoxygénée et une diminution du signal T2* (*"initial dip"*).

2.Puis le débit et le volume sanguin augmentent, dépassant les besoins en O2. L'Hb désoxygénée diminue et le signal T2* augmente (signal BOLD proprement dit).

3.Le débit et le volume sanguin se normalisent dans un délais de 8 à 10 sec. Mais la glie et les neurones récupèrent d'une activité partiellement anaérobie d'où une extraction d'O2 élevée et une diminution du signal T2* (*"undershoot"*).

On parle d'activation lorsque le ou les voxels concernés présentent une variation de signal corrélée avec la tâche. Cette détermination est loin d'être triviale, puisqu'il faut fixer un seuil de significativité et parfois un seuil d'extension spatiale. Nous pouvons retenir l'hypothèse d'une implication pour les voxels dépassant ces seuils, mais ne nous pouvons formuler aucune conclusion sur les autres. Cette notion de régions impliquées sans être "apparemment" activées lors d'une tâche, pose problème dans des approches intégratives ou il est nécessaires de connaître tous les nœuds du réseau (cf. connectivité effective).

Il n'existe à ce jour aucune méthode de contrôle des faux négatifs. Tout au plus est-il possible de réaliser un test de puissance statistique donnant le pourcentage maximum de la variation de signal si le voxel était concerné.

IRM fonctionnelle – signification des modifications de signal

Cette technique ne donne donc qu'un indice indirect de l'activité cérébrale. Elle ne renseigne que sur des variations d'un effet BOLD qui est l'intégration de l'indice d'extraction d'oxygène, du débit et du volume sanguin. Il semble qu'il suive une dynamique grossièrement similaire à ce qu'on observe en PET (*Positron Emission Tomography*) à l'eau marquée ($H_2{}^{15}O$*) qui est le standard des techniques de perfusion. Il semble que cet indice vasculaire soit corrélé au nombre de synapses en activité intégré sur une période de 4 à 8 secondes. Cette fenêtre d'intégration correspond à la dynamique de la réponse BOLD, qui réalise l'équivalent d'une convolution de l'activité synaptique. Il se pourrait que le véritable déterminant de la réponse BOLD soit la quantité de neuromédiateurs libérés par le pool synaptique et métabolisé par les astrocytes (toujours intégré sur 4 à 8 sec). Dans le cadre de cette

interprétation, se pose la question de possibles différences d'efficacité entre les neuromédiateurs. La dopamine par exemple semble entraîner une réponse d'intensité égale voir supérieure à celle du GABA au niveau des noyaux gris, tout en étant libérée en quantité très inférieure (11). Enfin, la fonction qui unit la réponse BOLD à l'activité synaptique n'est pas connue, mais semble grossièrement linéaire lors d'une activité cognitive normale.

L'interprétation doit prendre en compte 2 facteurs :

i) il n'y a pas de distinction entre synapses excitatrices et synapses inhibitrices, les deux entraînent une activation,
ii) il s'agit bien d'une activité synaptique et non d'une activité neuronale.

Il faut prendre en compte l'anatomie. Si les neurones se trouvent dans une région A, et qu'ils projettent sur une région B, alors l'effet BOLD enregistré en B correspond à l'activité des neurones de A. C'est ce qui se passe au niveau des noyaux gris centraux par exemple qui ont une connectivité très hiérarchisée, 90% des connexions venant de régions distantes. Ainsi l'effet BOLD enregistré dans le noyau ventro-latéral du thalamus reflète l'activité du globus pallidus interne. Comme ses projections sont inhibitrices, les neurones thalamiques sont en fait inhibés.

Dans le cortex près de 90% des connexions sont locales, et toutes les projections longues distances sont excitatrices (mais peuvent activer des interneurones inhibiteurs). On peut donc considérer que l'effet BOLD enregistré reflète bien leur activité.

Dernière remarque sur la technique d'IRMf, elle ne donne aucune information sur le niveau de base. Cette remarque est de peu d'intérêt pour l'étude du sujet normal, mais dans le cas de l'étude d'une pathologie ou d'une drogue cela peut poser problème. En effet, la seule interprétation que l'on puisse donner à une diminution d'activation c'est une difficulté à recruter une plus grande quantité de neurones. Cela n'indique pas si le phénomène est lié à une moindre utilisation de la région, à une perte synaptique, ou à un niveau d'activité déjà élevé en situation de référence (saturation).

IRM fonctionnelle – quelle interprétation pour une activation ?

Le congrès international d'imagerie fonctionnelle cérébrale ainsi que la première revue du même domaine porte le nom de "*Humain Brain Mapping*" (cartographie du cerveau humain). C'est littéralement la tâche à laquelle se consacre cette nouvelle communauté. Assigner à chaque portion du cortex l'étiquette de la fonction qu'elle supporte, c'est un programme localisationniste au sens fort.

Dans le cadre de l'IRMf, les variations d'activité d'une région cérébrale liées à une tâche peuvent être de 2 ordres :

i) une modification de son extension spatiale
ii) une modification de son intensité.

Ces deux variations ne sont pas indépendantes l'une de l'autre. Il est à noter que la notion d'intensité d'activation n'est pas équivalente à la valeur statistique du ou des voxels considérés en raison d'une distribution inhomogène du bruit. Nous avons vu que l'intensité de la réponse était proportionnelle à l'intégration de l'activité synaptique sur une fenêtre temporelle de l'ordre de 4 à 8 sec. Cela signifie qu'une élévation de la réponse BOLD peut tout aussi bien refléter une activité synaptique plus intense qu'une activité identique, mais plus soutenue dans le temps.

Les variations de temps de survenue en revanche sont rarement prises en compte. La résolution temporelle de la technique n'y est que pour une part. En effet, si on pouvait compter sur une réponse hémodynamique stable et homogène des délais aussi brefs que 100 ms pourraient avoir du sens. Malheureusement le délai de la réponse hémodynamique est fonction de sa localisation. A titre d'exemple, les régions du gyrus de Heschl sont en avance de plus d'1 seconde sur les régions frontales (Broca) lors d'une tâche verbale, et les régions préfrontales dépassent parfois les 4 secondes. De plus ce délai et son inhomogénéité s'accroissent avec l'âge.

<u>Interprétation dans le cadre de tâches différentes</u>

A quoi correspondent les modifications d'activité en termes cognitifs de représentation et de processus ? Il est difficile d'élaborer des règles.

Interprétation en terme de	**Représentation**	**Processus**
Cortex	Primaire/secondaire	Tertiaire
Approche	Soustractive	Paramétrique
Modification	Extension spatiale	Intensité d'activation

Un cortex primaire reçoit les informations de l'organe sensoriel (ou les émet vers l'organe effecteur), un cortex secondaire se spécialise dans l'extraction de certaines informations issues d'une seule modalité sensorielle (par exemple la forme ou la couleur pour la vision), un cortex tertiaire est un cortex polymodale, intégrant les informations de plusieurs sens.
L'approche soustractive consiste à regarder la différence d'activation entre deux catégories de stimuli, l'approche paramétrique cherche à faire varier l'intensité d'une stimulation.

De façon générale, les auteurs favorisent une interprétation “représentationnelle” lorsque cela concerne des aires primaires dont l’organisation est proche d’une “représentation”, en cas d’une modification de l’extension spatiale et d’une approche de type soustractive. À l’inverse l’interprétation a d’autant plus de chance d’être “processuelle” qu’elle concerne du cortex tertiaire, qu’il s’agit d’une modification de l’intensité d’activation et que l’approche est de type paramétrique.

Par exemple dans une expérience d’imagerie mentale en PET à $H_2{}^{15}O^*$, Kosslyn demandait aux sujets de visualiser mentalement un objet de petite taille et un objet de grande taille. Il s’intéressait à l’activité de l’aire visuelle primaire (V1). Il a décrit, pour les grands objets, un maximum d’activation s’étendant plus en arrière le long des lèvres de la scissure calcarine (projection du champ visuel périphérique). Il en a déduit que les grands objets étaient imagés plus largement sur V1 que les petits objets comme ce serait le cas lors d’une stimulation réelle (12). C’est donc une interprétation représentationnelle des variations d’activité, et on retrouve les arguments avancés : aires primaires, modification d’extension spatiale, approche soustractive.

A l’inverse, dans l’exemple suivant, c’est une interprétation processuelle qui a été retenue. Il s’agissait d’une tâche de mémoire épisodique dans laquelle on comparait en IRMf mémoire du mot vs. mémoire de la position du mot. Ce contraste mettait en évidence une plus forte activité dans les régions frontales droites. Richard Henson a interprété cette activation comme un surcroît de charge de travail des processus exécutifs (13). C’est donc une interprétation processuelle avec comme arguments : cortex tertiaire, modification de l’intensité d’activation.

Mais bien souvent, l’interprétation ne tranche pas entre processus et représentations. On parle alors simplement de spécialisation fonctionnelle de telle ou telle région cérébrale. L’analogie entre réseau neuronal et module proposé par Paul Churchland permet de comprendre cette absence de distinction. L’expérience des réseaux neuronaux de type Hopfield nous montre que le même support assure aussi bien le traitement que la représentation. Ce qui change en fonction de l’item, c’est le nombre d’itérations (≡ processus = Δ activité synaptique) pour parvenir à un état stable (≡ représentation). Ainsi dans cette interprétation, processus et représentation sont indissociables. Dans ce cadre, on voit que seul un processus peut expliquer une modulation de son activité (approche paramétrique), alors qu’une représentation ne peut être que présente ou absente (approche catégorielle).

Interprétation dans le cadre d'une même tâche

L'étude de l'IFC (Intégration Fonctionnelle Cérébrale) en l'IRMf repose essentiellement sur des changements d'activations au cours de la réalisation d'une même tâche. Nous ne sommes plus dans le cadre de l'étude comparative de l'activité d'une aire entre deux tâches, mais dans celle des variations d'activité d'une aire au cours d'une même tâche. Si on admet que la même région supporte représentation et processus. Quelle est la part respective de chacun ?

Si on comprend bien qu'une représentation puisse ou ne puisse pas être impliquée au cours d'une tâche (approche catégorielle), le concept d'une modulation de cette implication est plus difficile à cerner (approche paramétrique). Cela signifierait que deux représentations différentes pourraient i) soit impliquer un nombre de décharges synaptiques significativement différent, ii) ou que l'une soit active plus longtemps que l'autre. La première hypothèse est peu probable, la seconde n'est évoquée que dans le cadre de la mémoire de travail.

En revanche, un processus peut requérir une activité synaptique plus ou moins importante au cours d'une même tâche. Cette modulation de la charge de travail correspond soit à une activité synaptique plus intense, soit à une activité synaptique plus soutenue dans le temps. C'est ainsi que l'on explique le phénomène d'inhibition par le priming (le même traitement réalisé une deuxième fois entraîne une activation moindre), le renforcement de l'activité par l'attention ou la réponse diphasique de l'activité au cours d'une phase d'apprentissage (diffusion spatiale et renforcement de l'activité initiale puis diminution). En reprenant l'analogie avec un réseau neuronal de type Hopfield, moins le stimulus est connu (≡ inhibition par le priming, apprentissage), et/ou plus le niveau de performance demandé est élevé (≡ attention), plus le réseau demandera d'itérations (≡ activité synaptique) pour aboutir à la configuration finale (≡ représentation).

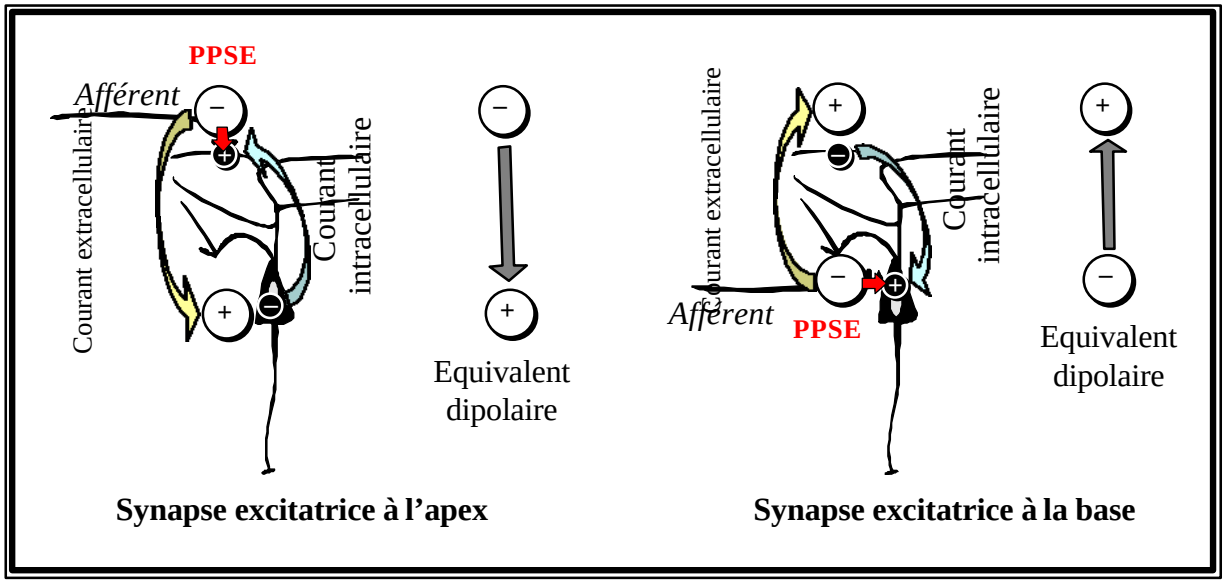
Cependant la relation qui unit charge de travail et activation pourrait ne pas être linéaire. C'est le cas de l'activité du gyrus de Heschl lors de l'audition de mots dont on fait varier le rythme de présentation. On observe bien une augmentation initiale quasi linéaire de l'activité avec l'augmentation du rythme de présentation. Mais au-delà de 50 mots par minute la tendance s'inverse (14). Une étude similaire a été réalisée sur la vitesse de présentation pour des mots lus. Elle a retrouvé en revanche une relation linéaire entre vitesse de présentation et activation dans les régions inféro-temporales bilatérales, mais aussi au niveau du gyrus de Heschl gauche (les sujets lisaient à haute voix) (15). On pourrait imaginer que ce qui a été décrit en auditif soit lié au caractère automatique ou contrôlé du traitement. Les sujets passeraient progressivement d'un mode à l'autre en fonction du temps dont ils disposent. De même, c'est

par un changement de stratégie que l'on interprète l'activation non linéaire dans les régions frontales lors d'une tâche de mémoire de travail (*"n-back task"*). Dans ce type de tâche le sujet est confronté à un défilement continu de stimuli et ne doit répondre à une cible que si celle-ci est précédée d'un signe convenu "n" présentations avant. Par exemple, dans un défilement de stimuli, ne répondre à la lettre "P" que si elle est précédée d'un "X" 2 stimuli avant (*"2-back"*). Plus le "n" augmente, et plus la charge de travail s'accroît. On observe bien dans un premier temps la croissance en intensité et en diffusion spatiale de l'activation avec l'augmentation de "n". Mais au-delà d'une certaine valeur la tendance s'inverse à nouveau. On imagine que dans ce cas, le sujet traite l'information par morceaux (*chuncking*) entraînant une réduction de la charge computationnelle (16).

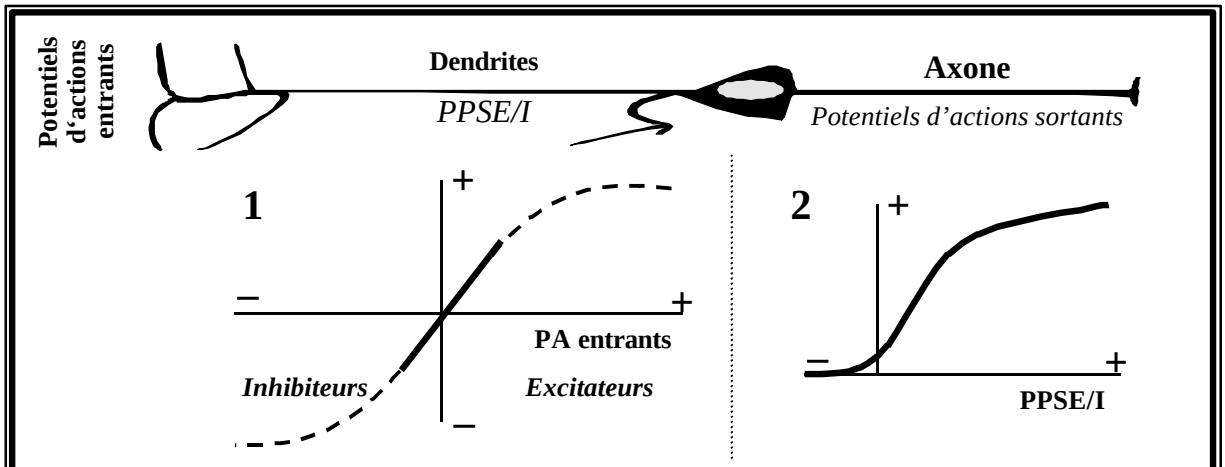
En IRMf, on estime donc que les variations d'intensité d'activation au cours d'une même tâche sont liés à des modification de charge de travail (effet de nouveauté, attention …).

L'EEG-MEG – aspects de la technique et signification des modifications de signal

Les techniques d'EEG-MEG (Electro et Magnéto-EncéphaloGraphie) permettent l'enregistrement des potentiels électriques (EEG) ou des variations du flux magnétiques (MEG) liés à l'activité cérébrale. Il faut l'activité synchronisée d'un grand nombre de neurones dit à champ ouvert pour obtenir un signal enregistrable en surface. Les neurones à champ ouvert sont essentiellement les cellules pyramidales du cortex. Leur conformation anatomique fait qu'elles se comportent comme un dipôle disposé tangentiellement à la surface. A nouveau, le signal ne correspond pas véritablement à leur activité, mais à l'activité synaptique. Plus exactement le signal correspond aux potentiels post-synaptique excitateurs (PPSE) et inhibiteurs (PPSI). Il est malheureusement difficile de déduire la part de l'une et de l'autre sur le simple sens de la déflexion. Il faudrait pour cela connaître la position de la synapse sur la cellule. Par exemple, une synapse excitatrice disposée au sommet du neurone va entraîner une source de potentiel en haut (pour les synapses inhibitrices, on parle de "puits de potentiel"). Le courant externe qui en résulte circule de haut en bas (du – vers le +), ce qui est aussi l'orientation de l'équivalent dipolaire du système. On enregistre donc en surface une déflexion négative. C'est l'inverse bien entendu pour une synapse inhibitrice située au même endroit, mais c'est aussi l'inverse pour une synapse excitatrice disposée cette fois ci à la base du neurone(cf. schéma)(17).



La mesure ne renseigne donc pas sur l'activité du neurone lui-même, mais plutôt sur l'état de ses entrées. Or si l'intensité des PPSE/I est linéairement corrélée aux nombre de potentiels d'actions parvenant aux synapses (à l'échelle de la population), le nombre de potentiels d'actions qui résulte des variations de potentiels post synaptiques ne suit pas une réponse linéaire mais saturable même en condition physiologique (cf. schéma)(18).



Au niveau d'une population de neurones, le potentiel dendritique est linéairement corrélé à la somme et à la pondération (excitatrice ou inhibitrice) des potentiels d'actions (PA) entrants. En théorie la relation correspond à une sigmoï de, mais en pratique seule la portion grasse de la courbe correspond au fonctionnement physiologique (courbe 1). En revanche, le nombre de PA sortants n'est pas linéairement corrélé à l'intensité du potentiel dendritique (courbe 2).

L'intensité du signal en surface est dépendante de 4 facteurs (17) :

i) Le nombre de synapses et de neurones à champ ouvert. En pratique seules les cellules pyramidales participent au signal.

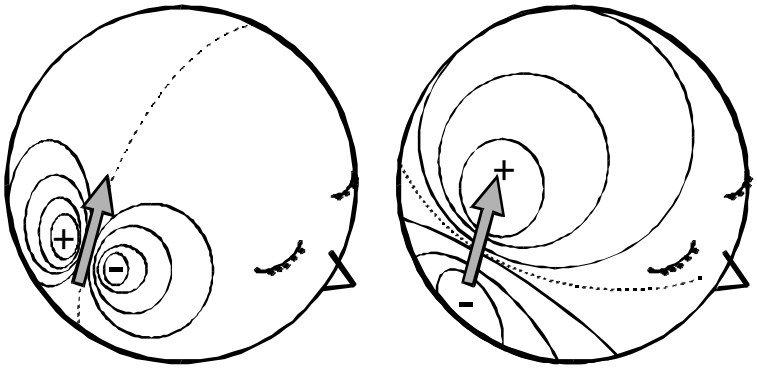
ii) Leur profondeur : l'intensité du signal diminue avec le carré de la distance de la source en EEG et avec les MEG n'utilisant pas la technique gradiométrique. Cette technique destinée à réduire les artefacts dus aux variations de flux magnétique lointain, entraîne

une réduction de l'intensité du signal approximativement en fonction du cube de la distance à la source.

iii) L'homogénéité et l'axe de leur orientation : si tous les dipôles élémentaires étaient orientés différemment, la résultante serait nulle et aucun signal ne serait enregistrable. Il faut donc une certaine homogénéité dans l'orientation des dipôles élémentaires pour réaliser une sommation spatiale. Un autre point est l'inclinaison de l'axe moyen par rapport au cortex. En fait, s'il change peu de chose pour l'EEG en raison de la diffusion du signal électrique dans les milieux biologiques, il n'en est pas de même pour la MEG, qui n'est sensible qu'aux dipôles d'orientation tangentielle au scalp à ±20° près. Pour cette raison, on estime que la MEG est insensible à environ 60% des dipôles, alors que ce n'est le cas que de 2 à 5% pour l'EEG.

iv) La synchronisation de leurs afférences : pour un même nombre de potentiels post synaptiques, l'intensité du signal dépendra de la synchronisation ou de l'étalement temporel de leurs occurrences. Pour que des potentiels post synaptiques puissent se sommer temporellement, il faut qu'ils puissent se recouvrir, sachant que leur durée est de l'ordre d'une dizaines de millisecondes. Leur étalement temporel va définir la période de l'onde résultante. De cette connaissance découle l'utilisation de filtres appropriées. Par exemple l'entrée dans une crise partielle est signée par ce qui est décrit comme un aplatissement tonique de l'EEG (essentiellement visible sur les enregistrements intracrâniens). Aucune déflexion n'apparaît, l'EEG s'aplatit, les rythmes normaux disparaissent, et pourtant un grand nombre de synapses sont en train de décharger. Aucune déflexion n'est visible parce que l'activité synaptique n'est pas synchronisées. La déflexion existe, mais elle a une période très longue qui n'apparaît pas car elle est filtrée par les filtres basses fréquences des systèmes classiques. Lorsque l'on peut bénéficier d'un appareillage dit DC (courant continu, c'est à dire sans filtre basse fréquence), on observe alors un décalage DC. C'est une déflexion très lente, mais de grande amplitude signant une très grande activité synaptique.

La distribution spatiale des potentiels ou des modifications de flux magnétiques renseigne sur la position anatomique des équivalents dipolaires. Cependant le lien n'est pas direct. En EEG, l'équivalent dipôle est orienté sur l'axe qui unit le pic de négativité avec celui qui uni le pic de positivité, il va du – vers le +, et se trouve sur la ligne de potentiel nul. En MEG, l'équivalent dipôle est orienté perpendiculairement à cette même ligne, sa direction est telle qu'il a le flux positif sur sa gauche et il se trouve sur la ligne de potentiel nul (cf. figure)(19).



MEG **EEG**

Différences MEG-EEG

Il existe des avantages de la MEG par rapport à l'EEG (19). Le premier concerne la diffusion. Elle est très forte et anisotrope pour le courant électrique dans les tissus biologiques, alors qu'elle est infiniment moindre pour les flux magnétiques. Il en résulte une bien meilleure localisation spatiale du signal MEG par rapport au signal EEG. Un autre avantage de la MEG est de ne pas nécessiter de référence. En EEG, on mesure une différence de potentiel entre deux points, l'électrode de scalp et la référence. Or la position de celle-ci n'est pas anodine. Il lui faut être la plus éloignée possible d'une source cérébrale commune aux électrodes de scalp. Si elle devait capter le même signal que celles disposées sur le scalp, il n'y aurait aucune différence de potentiel mesurable. Il lui faut aussi être éloignée de toute source éventuelle d'artefact, sinon celui-ci parasiterait l'ensemble des dérivations. A l'inverse, la MEG n'enregistre pas des différences de courant magnétique, mais des variations de flux magnétique[3]. Il faut que le flux magnétique qui traverse la bobine d'enregistrement varie pour y induire un courant. C'est ce courant induit qui constitue le signal. Il n'y a donc aucun besoin de référence.

Mais la MEG n'a pas que des avantages. En dehors des difficultés pratiques, elle souffre aussi d'un mauvais rapport signal/bruit et n'est sensible qu'à une fraction de la surface corticale (cf. ci-dessus).

La résolution temporelle en EEG-MEG, une force et une faiblesse

En raison de sa résolution temporelle, cette technique est essentiellement utilisée en mode événementiel (potentiels évoqués, puissance/synchronie/cohérence évoquées etc…). Dans ce mode, le résultat va dépendre de la maîtrise du temps de survenue de l'événement électrique.

[3] Le signal MEG est donc la dérivée temporelle première du flux magnétique. S'il y a des flux puissants, mais constants, aucun signal n'est enregistré. Si on considère que la différence de potentiel est à l'origine du flux magnétique, le signal MEG est la dérivée temporelle première du signal électrique.

En effet, les événements cérébraux sont brefs. On peut estimer que dans leur plus grande majorité, ils ne dépassent pas 200 à 400 ms. Pour pouvoir les extraire du bruit de fond de l'activité cérébrale normale, on moyenne les événements de sorte que seuls ceux qui sont corrélés temporellement au stimulus (ou à la réponse) s'additionnent. Mais cette technique ne fonctionne correctement que si l'événement électrique survient à délais à peu près constant après la présentation du stimulus (ou avant la réponse). Dans le cas contraire le moyennage ne permet plus de les faire ressortir du bruit de fond. Or plus la tâche est élaborée, plus le délai varie entre le stimulus et certaines étapes de son traitement. Par exemple, le potentiel le plus tardif qui reste correctement défini est la P300[4]. Au-delà, les N400-600, voir N1000 ne sont plus que des déflexions informes, probablement très différentes dans la réalité de ce que la technique de moyennage permet de ressortir. La haute résolution temporelle de la technique pose ainsi certaines difficultés dès lors qu'on s'intéresse à des fonctions élaborées.

L'EEG-MEG – quelle interprétation pour une activation ?

Dans le cadre de l'électrophysiologie, les variations d'activité liées à une tâche peuvent être de 3 ordres : i) une modification de la diffusion spatiale et ii) une modification de son intensité, iii) mais aussi de son temps de survenue. Le problème est encore compliqué par la multiplicité des événements en question. Le plus souvent il s'agit de potentiel, mais il peut aussi s'agir d'une augmentation de puissance, de cohérence etc…

Le cognitivisme a beaucoup moins marqué le milieu de l'électrophysiologie qu'il n'a marqué celui de l'imagerie. Il faut dire que la technique reste encore mal adaptée à l'exploration des fonctions cognitives complexes. Nous avons vu le problème de la variabilité de l'intervalle entre le stimulus et la réponse cérébrale. Il faut y ajouter des difficultés de localisation spatiale. Les tâches complexes activent plusieurs régions et parfois sur de vastes surfaces. Cela, ajouté aux déformations temporelles du potentiel explique que le problème inverse[5] ne permet pas encore de donner des résultats convaincants. Ainsi le plus souvent ce ne sont pas des régions, mais des ondes auxquelles on accroche une fonction. Par exemple, on parle de P3a pour parler d'une P300 automatique, d'une P3b pour parler d'une P300 attentionnelle. Et

---

[4] Dans la définition d'un potentiel évoqué, la lettre indique le sens de la déflexion P pour positive, N pour négative, et le chiffre la latence moyenne de survenue du pic après la présentation du stimulus (en millisecondes) pour une population normale. P300 signifie : onde positive survenant 300ms après la présentation du stimulus.

[5] Le problème direct consiste à calculer la répartition des potentiels en surface à partir de la connaissance que l'on a de la localisation, de l'orientation et de l'intensité des dipôles équivalents. Pour chaque configuration dipolaire, il existe une répartition de potentiel et une seule. Le problème inverse consiste à retrouver la localisation, l'orientation et l'intensité des dipôles équivalents qui pourraient donner une distribution de potentiels identique à celle qui est observée. Ce problème est dit mal posé, car il admet de très nombreuses

quand il s'agit de région, on fait souvent l'amalgame entre zone de projection de la négativité (plus rarement de la positivité) et région activée. C'est ainsi que dans une étude de la mémoire épisodique en potentiels évoqués, les auteurs ont interprété la négativité frontale droite à 600ms comme un signe d'activité de cette région. Pourtant on sait combien cette répartition est trompeuse. Ainsi le potentiel auditif tardif N100 est-il maximal au vertex (~SMA), alors que ses sources sont au niveau du planum temporal. Il faut au moins faire une reconstruction des sources pour pouvoir désigner une région, et encore, le résultat reste douteux en raison de sa sensibilité aux hypothèses faites.

C'est en espérant intégrer la résolution spatiale des techniques basées sur la réponse vasculaire et la résolution temporelle des techniques basées sur la réponse électrique, que plusieurs équipes s'intéressent à la multimodalité. L'espoir est de pouvoir assigner à chacune des aires activées en IRMf le décours temporel de son signal électrique. Il serait ainsi théoriquement possible de connaître l'ordre dans lequel elles s'activent ce qui donnerait tout son sens à l'idée d'un simple passage de l'information d'une aire à une autre. On retrouve derrière ce projet la conception cognitiviste du traitement de l'information. Pour l'instant la seule étude à notre connaissance qui ait pu démontrer un délai entre le passage du potentiel d'une aire à une autre, ne l'a fait que sur la N100 entre V1 et V2 pour un potentiel évoqué visuel standard (vision d'un damier où les cases noires et blanches alternent).

---

solutions. Il faut donc introduire des contraintes supplémentaires pour en sélectionner une. Celle-ci est donc éminemment dépendante des hypothèses formulées.

L'intégration fonctionnelle cérébrale
Définition, paradigmes et méthodes classiques

L'intégration fonctionnelle cérébrale – une définition

> *"Cognitive functions are mediated by coordinated linkages between multiple brain areas that are transiently bound for the purpose of creating a unified percept and purposeful action"*
> (Marcel Mesulam, 1990)

Parler d'intégration fonctionnelle cérébrale c'est reconnaître que dans l'exécution d'une tâche le cerveau fait collaborer un grand nombre d'éléments. Collaborer sous-entend que cette activité est coordonnée et régulée. L'étude de l'intégration fonctionnelle cérébrale concerne donc la collaboration élémentaire et le fonctionnement d'ensemble. Elle cherche à décrire et à expliquer la coordination de l'activité cérébrale en une activité mentale cohérente.

L'intégration fonctionnelle cérébrale – histoire d'un vocable

Le vocable d'IFC semble apparaître sous la plume de Giulio Tononi dans un article de 1994 (20). L'intérêt pour une vision intégrée du fonctionnement cérébral émerge de sa collaboration avec Gerald Edelmann, directeur de l'Institut de neurosciences de La Jolla en Californie. Celui-ci est lauréat du prix Nobel de médecine en 1972. Un prix qu'il a obtenu, non pour ses travaux en neurobiologie, mais pour ses travaux sur la structure des immunoglobulines. Fort de cette reconnaissance et de sa position, il a pu consacrer une bonne partie de l'effort de recherche de son Institut à l'étude d'un sujet tabou et donc à risque à l'époque : la conscience. Il annonçait cette intention dans une série de livres entre 87 et 90 (citons : *"Neural Darwinism"* et *"The remembered present"*). Certes, il ne fut pas le seul scientifique à s'intéresser à ce problème, à titre d'exemple, rappelons les travaux novateurs menés par Benjamin Libet quelques années plus tôt. Mais c'était là les efforts de francs tireurs travaillant souvent isolément. L'Institut de Neurosciences s'est offert deux luxes : celui de lancer non pas un homme, mais une équipe dans la bataille, et celle de prendre le temps de conceptualiser la conscience en terme biologique (nous reverrons plus tard toute la différence avec une position cognitiviste). Or la conscience possède cette propriété intégratrice, et dans l'esprit de l'article princeps de Tononi et collaborateurs, intégration fonctionnelle rime avec

conscience. Nous verrons que si ces auteurs semblent avoir promulgué le vocable, ils se sont largement inspiré de leurs prédécesseurs pour ce qui est du concept.

Cette idée va faire écho quelques années plus tard dans la communauté de l'imagerie fonctionnelle cérébrale ou elle sera importée par Karl Friston. Celui-ci a en effet passé plusieurs mois à l'Institut de neurosciences de La Jolla en 1993-94 d'où il rapportera de nombreux concepts qu'il appliquera au domaine de l'imagerie fonctionnelle. Plus qu'une idée neuve, il apportera à la communauté de l'imagerie fonctionnelle une synthèse des instruments pour l'étude de l'IFC (21). Bien que toutes les méthodes aient souvent été abordées par d'autres, en particulier par le groupe de Toronto avec Randy MacInstosh (22), c'est Karl Friston qui donnera un cadre de référence à ces différentes approches : approches corrélatives comme l'analyse en composante principale, ou approches causales comme celles des modèles d'équations structurelles. Il développera le concept de connectivité issu de l'électrophysiologie. Cependant comme nous le verrons, bien que le vocable d'intégration fonctionnelle soit resté, le concept qu'il recouvre alors est très différent de celui de l'Institut de neurosciences. Il s'est "cognitivisé".

L'intégration fonctionnelle cérébrale – les paradigmes classiques

Dans la conception cognitiviste les éléments de base dont il faut intégrer l'activité sont les modules. Au niveau élémentaire cela équivaut à étudier la connectivité entre deux modules, donc leur dépendance en terme de causalité. Or nous avons vu que la seule causalité admise dans le cadre fonctionnaliste du cognitivisme est la causalité efficiente (cf. 1er chapitre). Au niveau du système, cela équivaut à décrire l'ensemble des connexions valides entre les modules impliqués par la tâche.

<u>La connectivité</u>

La notion de connectivité est directement inspirée de l'électrophysiologie. Elle y est définie dans les termes de la théorie de l'information comme la dépendance qu'il y a entre l'activité de l'unité A et celle de l'unité B (21). Cette dépendance est caractérisée par l'information mutuelle (en électrophysiologie, l'équivalent est le *"Joint Peristimulus Time Histogram"* – JPTH) ou la corrélation en cas de relations linéaires. Nous parlerons souvent de corrélation, puisque jusqu'à présent, l'essentiel des approches en intégration fonctionnelle en IRMf est restée dans ce cadre linéaire.

La résolution temporelle de l'électrophysiologie permet de mettre en évidence un décalage temporel constant entre l'émission d'un potentiel d'action de A et l'émission d'un potentiel

d'action de B. La cause devant précéder l'effet, on en déduit que A influence B. Malheureusement cette résolution temporelle fait défaut à l'IRMf, ce qui rend le lien causal difficile à démontrer.

De plus, le système nerveux est un systèmes complexe. L'activité de B peut sembler dépendante de A sans qu'il y ait un lien causal direct entre l'activité de A et de B. On peut imaginer qu'un facteur commun les influence tous deux. Par exemple, en électrophysiologie, la co-occurrence des potentiels d'actions de deux neurones peut être liée à l'augmentation parallèle de leurs potentiels d'actions sous l'influence du même stimulus sans qu'il y ait un lien de cause à effet (on introduit des corrections pour cet effet dans les JPTH). On peut aussi imaginer que A et B ne sont pas connectés directement, mais à travers une autre unité C. Ou encore, on peut surestimer l'influence de A sur B en négligeant les effets de C. Autant de raisons qui font que la corrélation ne peut pas être interprétée comme un lien de causalité. Or c'est justement ce contraste que Friston a introduit entre connectivité fonctionnelle (lien de corrélation simple), et connectivité effective (lien causal efficient).

La technique consiste à observer la corrélation de l'activité de différentes aires. Chacune des populations neuronales est considérée comme l'équivalent d'un module, l'information circulant de l'une à l'autre sous la forme d'une représentation. On observe donc une corrélation de la *charge de travail* entre les modules. Si un item est moins connu, plus difficile à traiter ou s'il s'est accompagné d'un surcroît d'attention, la charge de travail est supposée s'accroître pour l'ensemble des modules concernés. Les modules ont donc une activité dépendante les uns des autres, et c'est cette dépendance qui signe leur connectivité.

La connectivité fonctionnelle

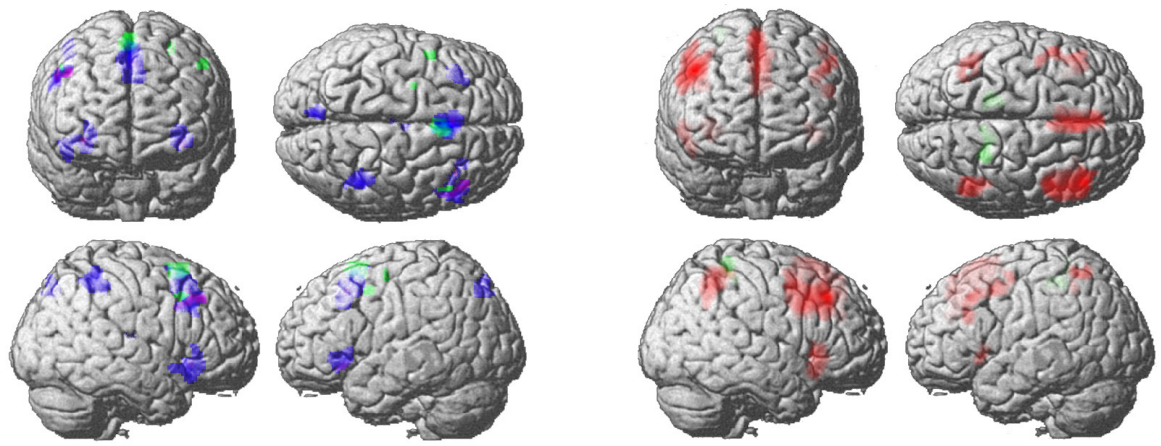
Techniques linéaires

L'approche univariée

La **corrélation** a été la première technique employée. Il suffit de prélever l'activité d'une aire (corrigée pour un certain nombre d'artefacts dépendants de la technique), et de regarder les différents voxels dont l'activité est en corrélation.

Prenons l'exemple d'une tâche de mémoire au cours de laquelle nous présentons aux sujets des mots soit à droite soit à gauche de l'écran. La tâche consiste soit à dire s'ils reconnaissent le mot présenté, soit à dire de quel coté il est (facteur récupération). A la sortie du scanner, nous leur demandons quel était le pourcentage de réponse pour lesquelles ils étaient certains (facteur doute). La région plus active lorsque le sujet doutait de ses réponses était une région dorso-latérale préfrontale droite (DLPF-D). Pour mieux la caractériser, nous avons entrepris

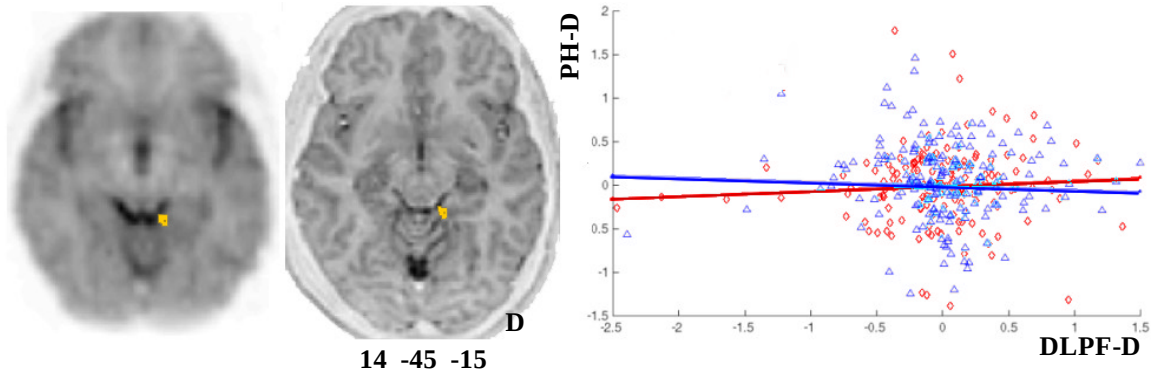
une étude de sa connectivité fonctionnelle durant la tâche de récupération. Nous avons donc gardé toutes les images en condition de récupération, extrait le décours temporel des voxels du cortex DLPF-D, et regardé quels étaient les autres voxels qui leur étaient corrélés. Dans cette exemple, les régions dont l'activité est positivement corrélée correspondent à un réseau fronto-cingulo-pariétal bilatéral. Ce résultat est conforme à ce que l'on connaît des liens anatomiques qu'entretiennent ces régions. Remarquons aussi la présence du cortex parahippocampique droit dans ce réseau. D'autres régions peuvent être corrélées négativement. C'est le cas d'une autre région du cortex pariétal disposée en avant des régions précédentes.



A gauche, activation durant la récupération du mot (en vert), de la position du mot (en bleu). La tache mauve en DLPF-D correspond à la région plus active lorsque le sujet doute de sa réponse. A droite, se trouvent les régions dont l'activité est corrélée avec celle de la région DLPF-D sus-citée (en rouge) ou anti-corrélée (en vert).

Plus que la connectivité fonctionnelle en tant que telle, sa modulation en fonction de la tâche réalisée peut être intéressante. Cette idée a été conceptualisée par Karl Friston sous l'appellation d'**interaction psychophysiologique** (23). Le patron de ce type d'étude est souvent factoriel c'est à dire faisant intervenir 2 facteurs de façon croisée (ab, Ab, aB, AB). Ce patron permet de mettre en évidence une interaction entre les deux facteurs psychologiques étudiés. Mais on peut lors de l'analyse remplacer l'un des facteur par le décours temporel d'une des aires qui lui est corrélé. On étudie donc une interaction entre un facteur psychologique et un facteur physiologique. Cela revient à comparer la corrélation de l'aire 1 avec l'aire 2 dans la condition a et dans la condition A. Si la différence est statistiquement significative, on peut en conclure que l'aire 1 voit sa connectivité fonctionnelle varier en fonction de la tâche avec l'aire 2. En reprenant l'exemple ci-dessus, on va s'interroger sur la modulation de la connexion du cortex DLPF-D avec l'ensemble des autres régions du cerveau par le facteur récupération. On s'aperçoit que le réseau fronto-

cingulo-pariétal bilatéral ne voit pas sa connectivité fonctionnelle changer, à l'inverse du cortex parahippocampique D (pH-D). Le pH-D est fonctionnellement plus fortement connecté avec le DLPF-D en situation rappel qu'en situation lecture seule.



Interaction psycho-physiologique. Le cortex parahippocampique droit (pH-D en orange) est corrélée positivement avec le cortex DLPF-D lorsque le sujet récupère une information (droite de régression en rouge), alors que la pente est voisine de 0 lorsque le sujet lit simplement le mot (en bleu).

Karl Friston a proposé deux interprétations causales à l'interaction psychophysiologique : i) DLPF-D est la cause de l'activité du pH-D dans la situation rappel mais n'a aucune influence en situation lecture, ii) la tâche de rappel est la cause de l'activité du pH-D, et son effet est d'autant plus important que l'activité du DLPF-D est importante. Nous pourrions nous limiter à cela si l'activité du cortex DLPF-D était véritablement un facteur que nous pouvions manipuler et si nous étions certains de l'implication directe de ces deux facteurs sur les aires activées. Cela n'est malheureusement pas le cas et les possibilités sont beaucoup plus nombreuses. Sans être exhaustif, les possibilités suivantes sont possibles : i) le pH-D est la cause de l'activité du DLPF-D dans la situation rappel mais n'a aucune influence en situation lecture, ii) pH-D et DLPF-D s'influencent mutuellement en condition rappel mais pas en condition lecture iii) Une aire X est la cause de l'activité du DLPF-D et du pH-D, iv) L'effet du rappel n'est pas directe, mais passe par une aire Y, v) L'effet du DLPF-D n'est que partiel, tout ou partie des autres composantes du réseau fronto-cingulo-pariétal bilatéral l'influence aussi...

Cette idée de remplacer un facteur psychologique par l'activité physiologique d'une aire qui lui est lié, peut parfaitement être étendue aux deux facteurs. Paul Fletcher a proposé de l'appeler **interaction physio-physiologique** (24).

Une autre façon d'approcher un changement de connectivité est d'explorer sa dynamique au cours du temps. Une méthode de **régression à coefficient variable** inspirée des filtres adaptatifs (filtre de Kalman) a été proposée par Christian Büchel et Karl Friston. Ils ont ainsi pu mettre en évidence une modulation de la connectivité entre V5 et le cortex pariétal par

l'attention au cours de la vision d'un champ d'étoiles en mouvement (*"star field stimulation"*). C'est la seule méthode qui soit adaptable en théorie à des paradigmes événementiels pour l'IRMf, mais cela n'a pas encore été fait.

Approche multivariée

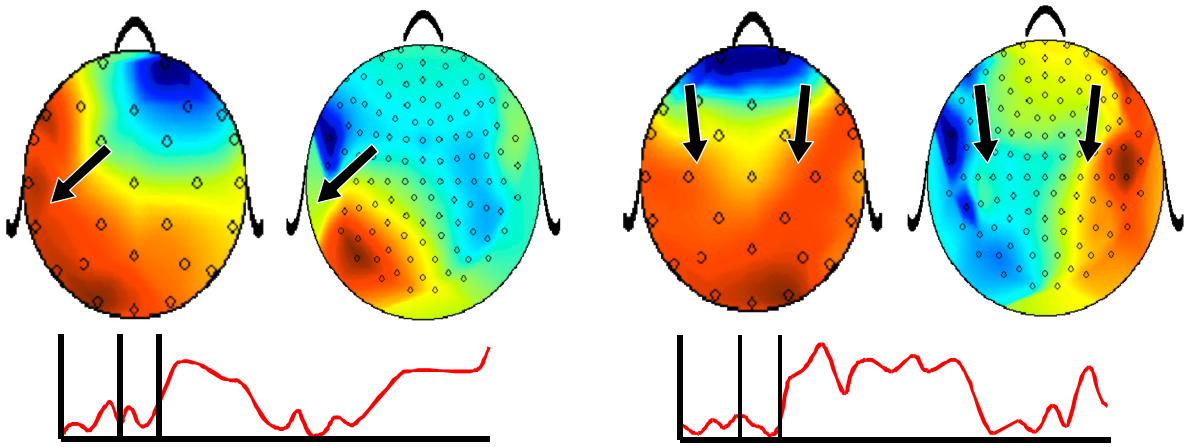
Pour réaliser une analyse en connectivité fonctionnelle, il n'est pas obligatoire de définir une région d'intérêt. Il est ainsi possible de s'interroger sur les régions dont l'activité semble covarier. Une première approche serait de faire l'hypothèse qu'un voxel ne peut participer de façon significative qu'à un seul réseau, ce qui est, à première vue, une hypothèse acceptable. L'analyse la plus adaptée ne va donc retenir qu'une seule composante par voxel. Cela correspond aux différentes techniques de *clustering* (regroupement).

A titre d'exemple, nous avons implémenté une technique de ***clustering*** dite mixte, qui allie la rapidité des techniques d'agrégation autour d'un nombre de composantes défini à priori (fuzzy *clustering*) à une technique d'agrégation hiérarchique (25). Cette dernière permet de ne pas avoir à faire d'hypothèse à priori concernant le nombre de composantes. Pour l'instant cette technique n'a été appliquée que pour des données simulées à partir d'une série de données réelles. Les résultats obtenus montrent que l'hypothèse de l'appartenance d'un voxel à une seule composante ne prend pas en compte la réalité expérimentale. En particulier cette hypothèse fait abstraction du bruit lié à la mesure, et c'est en partie l'explication de ses faibles résultats. Il faut rappeler qu'un grand nombre de sources d'artefacts ne sont pas corrigées en IRMf. C'est le cas des artefacts cardiaques que nous avons surtout retrouvé au niveau de la base du crâne (ils y expliquent jusqu'à 20% de la variance) ou des artefacts respiratoires surtout gênants au niveau de la convexité.

Pour prendre ce fait en considération, on peut faire l'hypothèse que l'activité d'un voxel est la somme pondérée de plusieurs composantes : certaines correspondent à son implication dans un ou plusieurs réseaux, d'autres correspondent à une ou plusieurs sources de bruit. Cette approche exploratoire des données dérive de la décomposition en valeurs propres. L'exemple le plus classique est l'**analyse en composantes principales** (26). Dans le cadre de l'imagerie fonctionnelle, cette technique pose quelques problèmes. En effet, elle fait l'hypothèse que la première composante se trouve le long de l'axe de la plus grande variance. Comme les composantes qui suivent doivent lui être orthogonales, la solution correspond simplement à la décomposition la plus "économique" (la maximum de variance doit être expliqué par les premières composantes). En intégration fonctionnelle, la question serait plutôt de savoir qui

travaille avec qui. Cela revient à identifier des groupes de voxels qui partagent la même composante.

La technique des **projections révélatrices** (*"projection pursuit"*), proche de l'analyse en composante principale parvient, en imposant une contrainte de robustesse à la matrice de variance/covariance, à trouver une décomposition qui permet de séparer au mieux les différents groupes de voxels (27). La technique d'**analyse en composantes indépendantes** dans son implémentation actuelle, réalise la même opération, mais impose en plus, une contrainte d'indépendance entre les composantes qui s'étend au-delà de la simple orthogonalité (minimisation des interactions linéaires et non linéaires entre composantes)(28;29). En pratique elle n'est utile qu'en cas d'écart significatif à la multinormalité (moments d'ordres supérieurs à 2 non nul : kurtosis et skewness). Dans le cas contraire, ses résultats ne devraient pas être différents d'une analyse basée sur une décomposition en valeurs propres. Les données d'IRMf semblent souvent avoir une distribution multinormale, ce qui n'est que rarement le cas des données électrophysiologiques. Ainsi dans l'exemple suivant, nous avons voulu analyser les potentiels évoqués lors d'une tâche de génération de verbe et de mémoire de travail. Nous avons décomposé le potentiel évoqué par une analyse en composantes indépendantes sur les données MEG et EEG. Tous deux ont mis en évidence des composantes cohérentes quant à la disposition et au décours temporel des cartes de potentiel. Notez que la composante n°2 semble être expliquée par deux équivalents dipôles disposés au niveau des deux lobes frontaux. La corrélation de leur décours temporel évoque une connectivité fonctionnelle entre eux.



Analyse en composant indépendante. Sont illustrées deux composantes issues des potentiels évoqués recueillis en EEG – MEG Durant une tâche de génération de verbe chez un même sujet. Celui-ci entend un mot (entre les deux dernière barres verticales ~500ms), et doit lui trouver un verbe sémantiquement relié. Une fois qu'il a trouvé celui-ci, il doit le garder en mémoire de travail jusqu'à la fin de l'épreuve (5000ms). On recherche les capteurs ayant une composante présentant des variations en fonction du temps qui sont identiques (courbe du bas). La pondération de ces composantes ainsi que leurs dipôles équivalent sont cartographiés pour l'EEG (à gauche) et la MEG (à droite).

Type	*Nbr. composantes non prédéfini*	*> 1 composante par voxel*	*Groupement*	*Non multi-normale*
Clustering	+/ –	–	+	
PCA	+	+	+/ –	–
Proj. révélatrices	+	+	+	–
ICA	+	+	+	+

Tableau récapitulatif des différentes **méthodes multivariées** utilisées pour l'analyse de la **connectivité fonctionnelle** et de leurs propriétés : i) nécessité ou non de prédéfinir le nombre de composantes recherchées, ii) possibilité pour un voxel d'avoir un décours temporel expliqué par une ou plusieurs composantes, iii) optimisation du regroupement, iv) prise en compte d'une distribution non multi-normale (en pratique surtout intéressante pour les données électrophysiologiques).

Techniques non linéaires

Rien ne nous permet de déterminer a priori si la fonction de dépendance entre 2 aires correspond à une fonction linéaire. Il y a même toutes les raisons de penser que cela n'est pas le cas. L'électrophysiologie a fréquemment confirmé que l'activité d'une aire était parfois loin d'être une fonction linéaire de la charge de travail. Cependant l'analyse non linéaire souffre de quelques difficultés. L'une d'elles est que la plupart des méthodes sont gourmandes en données nécessaires à l'obtention d'un résultat. En pratique, seules les techniques électrophysiologiques offrent suffisamment de points de mesure pour envisager de les utiliser. Pour contourner ce problème certaines techniques appliquées en IRMf reviennent à un système linéaire.

La première méthode de linéarisation consiste à ajuster l'activité d'une aire B sur un **polynôme** de l'activité d'une aire A.

$$B = \beta_n A^n + \beta_{n-1} A^{n-1} + \ldots + \beta_2 A^2 + \beta_1 A + \varepsilon$$

Il suffit alors de déterminer les coefficients de régression β. Une généralisation de ce type d'approche à plusieurs aires correspond aux séries de Volterra. Il faut alors inclure les termes d'interaction (30). Par exemple si on pense que l'activité de C est fonction de deux aires A et B, la série de Volterra d'ordre 2 correspondrait à :

$$C = \beta_1 A^2 + \beta_2 B^2 + \beta_3 AB + \beta_4 A + \beta_5 B + \varepsilon$$

Bien qu'intéressante, cette approche présente 3 difficultés : i) le nombre de coefficients de régression augmente avec l'ordre du polynôme, ii) il est nécessaire de déterminer a priori l'ordre du polynôme, en pratique, il est rare de dépasser l'ordre 2, iii) le problème est asymétrique, dans le cas ci-dessus, nous avons polymérisé la valeur de l'aire A, mais pourquoi ne pas avoir choisi B, en sachant que cela n'aboutirait pas au même résultat.

Une seconde méthode de linéarisation a été proposée par Fabrice Wendling, c'est le coefficient de corrélation h2. Il a appliqué cette méthode pour établir la dépendance entre deux plots d'enregistrement au cours de crises médio-temporales. Le **coefficient de corrélation h2** correspond en fait à une régression linéaire par morceaux (31). Les valeurs de A sont découpées en intervalle et A et B sont ajustés linéairement sur chacun des intervalles ainsi définis. Là encore il existe un certain nombre de problèmes : i) la définition des d'intervalles et leur nombre, ii) et encore cette notion d'asymétrie, puisque la découpe en intervalle de A ne donnera pas le même résultat que si c'est B qui était découpé. Fabrice Wendling a proposé d'utiliser cette asymétrie inhérente aux approches de linéarisation pour en inférer un sens : si A explique plus B que B n'explique A, alors B est asservi à A (31). Le problème est que cette asymétrie est extrêmement dépendante des choix quant au nombre et à la disposition des intervalles, ce qui en fait une donnée très artéfactuelle.

Toutes les techniques de linéarisation sont mal adaptées si la fonction qui unie A à B comporte des points de rupture. La théorie de l'information met à notre disposition un instrument plus universel pour mesurer la dépendance entre 2 signaux ou plus, c'est **l'information mutuelle** (cf. pages d'illustration)(32). A titre d'exemple, nous avons analysé le décours temporel de l'information mutuelle entre un plot d'une électrode située dans l'amygdale (Amg), et un autre dans le pole temporal (Pol), au moment de l'entrée en crise d'une patiente. Les états que pouvaient prendre les systèmes Amg et Pol étaient des valeurs discrètes correspondant à des intervalles de différence de potentiel par rapport à la référence. A titre de comparaison, le décours temporel de la corrélation est aussi présenté. On voit que cette dernière ne permet pas de différencier la dépendance entre Amg et Pol en situation de pré-crise par rapport à la phase de diffusion de l'activité critique, alors que la différence est très nette pour l'information mutuelle. Cela signifie que les deux systèmes deviennent très interdépendants quelques secondes après le début de la crise, mais que la relation qui les unis n'est pas simplement linéaire.

Un autre exemple a été proposé par Karl Friston. Il a proposé comme code de l'intégration cérébrale le "transient" (33). L'opérationnalisation de cette idée s'envisage en définissant comme état du système non un niveau de différence de potentiel, mais une puissance dans une certaine bande de fréquence. Il a analysé le signal MEG entre un senseur frontal et un senseur pariétal chez un sujet réalisant une tâche visuo-motrice. Il a observé un accroissement de l'information mutuelle des spectres de l'un et l'autre senseur entre deux bandes de fréquence différentes : autour de 20 et 40 Hz.

Une des difficultés de l'approche non linéaire a trait à la signification statistique. En effet, si pour des systèmes linéarisés, il reste possible d'avoir une idée de la probabilité de distribution de l'indice de dépendance, ce n'est plus le cas pour l'information mutuelle. La technique consiste alors à faire du *bootstrapping*, c'est à dire à refaire le calcul à plusieurs reprises en cassant la dépendance entre les données. Si ce qui nous intéresse est simplement la signification d'un écart par rapport à la normale comme dans le cas de la crise ci-dessus, nous pouvons effectuer une normalisation. On estime que la distribution de l'indice suit une loi gaussienne, et on compare sa distribution par rapport à une situation de référence.

Quelques éléments de théorie de l'information

L'entropie quantifie l'incertitude, c'est à dire le potentiel informatif d'un événement. Imaginons que nous avons un système dont tous les états sont équiprobables. Si le système prend un état, la réduction de l'incertitude est grande et donc l'information apportée est grande. A l'inverse, imaginons un système où la distribution des états correspond à un pic de Dirac, c'est à dire que le système ne peut prendre qu'un seul état. Cette fois ci, lorsque l'évènement survient, la réduction de l'incertitude est nulle, et donc l'information apportée est nulle. Mathématiquement, l'entropie H du système X se calcule comme suit :

$$H(X) = \sum_{x \in X} p(x) \,.\, \log (1/p(x))$$

Où x correspond à l'un des états de X et p(x) est sa probabilité. Notez que dans le cas décrit plus haut, d'une distribution équiprobable des états, l'entropie vaudrait log (1/p), ce qui est la définition de la quantité d'information.

Imaginons à présent que nous avons deux systèmes X et Y, et que nous voulions estimer l'entropie du système (X,Y), ou encore son potentiel informationnel. Son nombre d'états correspond au nombre d'états du système X multipliés par le nombre d'états du système Y. Il est possible de calculer une entropie conjointe :

$$H(X,Y) = \sum_{x \in X \text{ et } y \in Y} p(x,y) \,.\, \log (1/p(x,y))$$

A présent, nous aimerions disposer d'une mesure de dépendance entre deux systèmes X et Y que nous allons appeler l'Information Mutuelle (IM_{XY}). Cette mesure va nous permettre d'estimer la réduction de l'incertitude sur Y, qu'apporte la connaissance de X. Mathématiquement, l'IM_{XY} correspond à l'addition du potentiel informatif des systèmes X et Y diminué du potentiel informatif du système (X,Y) :

$$IM_{XY} = H(X) + H(Y) - H(X,Y)$$

Cette valeur prend 0 lorsque X et Y sont indépendants, et 1 lorsque X permet de prédire Y sans risque d'erreur.

Imaginons que le système X a une distribution quelconque, alors que le système Y a une distribution en pic de Dirac. Dans ce cas, la connaissance de Y n'informe en rien sur celle de X, l'$IM_{XY} = 0$ (et H(X,Y) = H(X)).

Imaginons à l'inverse que X et Y aient des distributions équiprobables, mais que le système (X,Y) ne se distribue que sur les couples (x,y) où x = y (relation linéaire), de façon équiprobable. Alors l'incertitude du système (X,Y) est faible, alors que l'incertitude sur les systèmes pris séparément est forte. Cela signifie que les systèmes X et Y sont dépendants, l'$IM_{XY} = H(X) = H(Y)$.

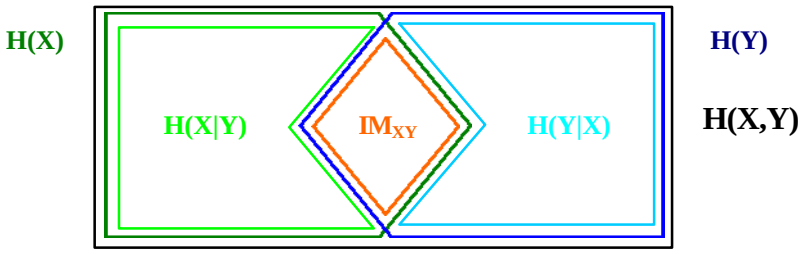


Diagramme de Venne des entropies de deux sources

Connectivité effective

A l'inverse de la connectivité fonctionnelle, la connectivité effective cherche à établir la part réelle de l'effet direct d'une aire A sur une aire B. C'est donc une approche causale de l'activité des aires cérébrales. Le concept est directement inspiré des JPTH utilisés en électrophysiologie et utilisant des techniques de corrections issues de la théorie de l'information (21).

En électrophysiologie, on déduit le sens de la connectivité par le délai entre l'activité de A et celle de B. La cause devant précéder l'effet, on estime que A influence B. L'IRMf ne dispose pas d'une telle résolution temporelle. Il faut alors se rabattre sur des techniques permettant des inférences causales moyennant un certain nombre d'hypothèses : i) d'avoir identifié l'intégralité des nœuds du réseau causal et de les avoir tous imagés, ii) le système est linéaire, donc s'il y a des non-linéarités, il faut les modéliser. Ainsi l'étude de la connectivité effective par ces méthodes nous amène toujours à étudier le système dans son ensemble.

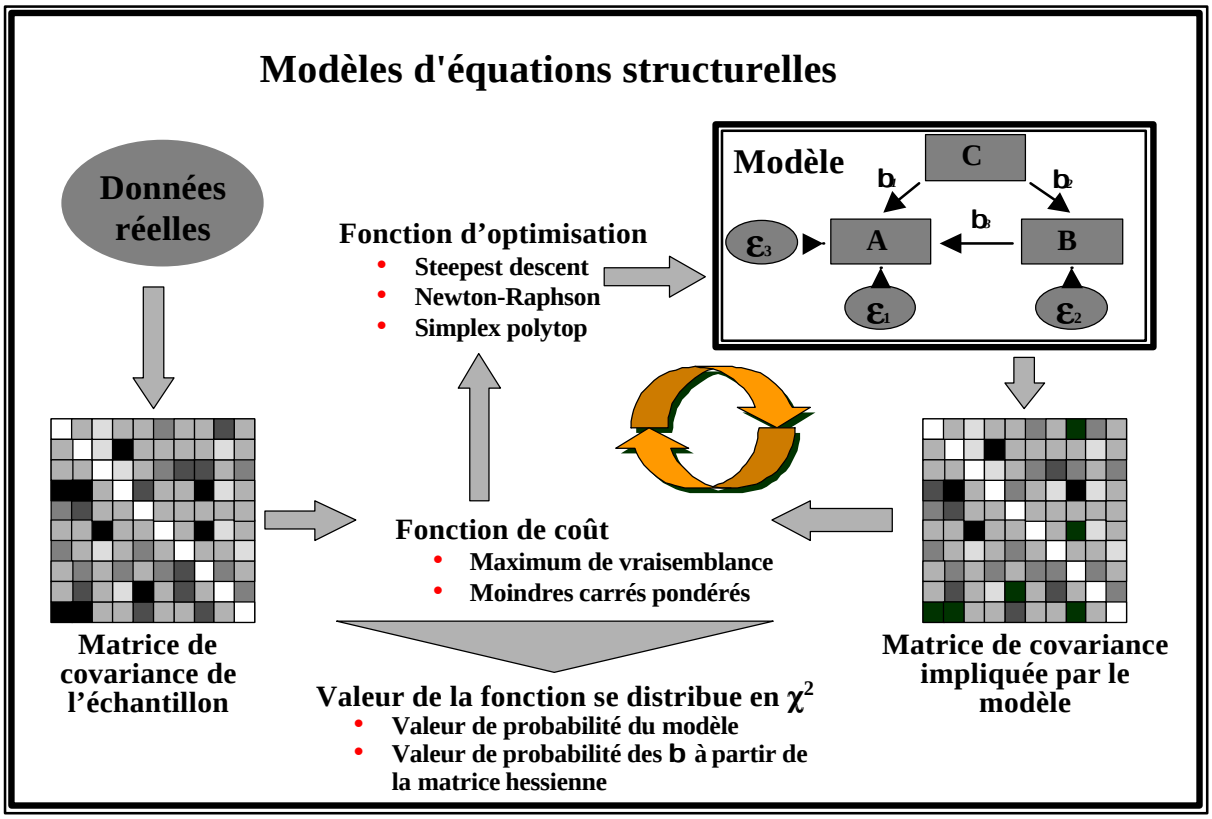
<u>Techniques linéaires</u>

L'approche la plus utilisée actuellement en IRMf utilise la technique des **modèles d'équations structurelles** (**SEM** pour Structural Equation Modelling). Le premier à l'avoir appliqué au cas de l'imagerie fonctionnelle est Randy MacIntosh (22). Dans les applications qu'il en a faites, c'était principalement pour pouvoir décrire le système durant l'exécution d'une tâche. C'est Karl Friston et Christian Büchel qui la conceptualiseront comme une technique de mesure de connectivité effective (21).

La technique est fort simple (cf. schéma, et pour plus de détail cf. appendice A)(34). Il faut avant tout définir un modèle causal entre les différentes aires. Jusqu'à présent seul des modèles utilisant des variables observées ont été utilisés (ce cas particulier des SEM s'appelle encore analyse des chemins ou *"path analysis"*). Puis il faut fixer une valeur de départ aux coefficients de régression partiels qui les lient les unes aux autres. A partir de cela il est possible de déduire la matrice de variance–covariance des différentes aires en fonction des

coefficients. On compare alors cette matrice théorique à la matrice de variance–covariance mesurée et on ajuste les coefficients jusqu'à trouver la solution qui soit la plus proche possible des données. On dispose alors de deux séries de résultats :

1) La première a trait à la validité du modèle dans son ensemble et nous permet de savoir si le modèle proposé doit être rejeté, ou peut être conservé. Cette technique ne permet pas de valider un modèle, seulement de l'invalider. Ces données sont surtout intéressantes pour pouvoir comparer deux modèles, en sachant que cela n'est possible avec un seuil de significativité que pour les modèles dit emboîtés (on passe d'un modèle à un autre en rajoutant ou en enlevant des liens) et récursifs (ne formant pas de cycle, donc pas de liens de causalité inverse). Pour un modèle pris isolément, son degré d'écart par rapport aux données est dépendant du nombre d'observations. Plus celles-ci sont nombreuses, plus le modèle risque d'être significativement différents des observations et plus il a de chance d'être rejeté.
2) La seconde série nous donne la valeur des coefficients de régression partiels, leur intervalle de confiance et leur signification. C'est la valeur de ces paramètres qui correspond à la connectivité effective.



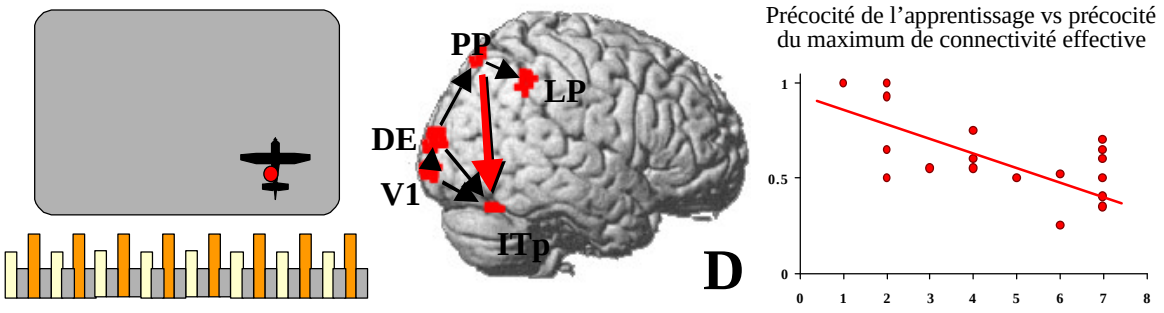
La difficulté tient souvent à la difficulté de proposer un modèle, et que celui-ci soit identifiable. En effet, on part souvent sans idée très précise sur un modèle, or la valeur des coefficients de régression partiels dépend de ce modèle. Cela signifie qu'il faut avoir spécifié

toutes les aires et tous les liens pouvant intervenir. Or rien ne dit que toutes les aires qui interviennent soient activées par la tâche. C'est pourtant l'hypothèse faite par l'ensemble des travaux à l'exception du travail de Jenny Coull et collaborateurs (35). Quant aux liens, on pourrait se dire qu'il suffit de les mettre tous et d'enclencher le programme. Malheureusement au-delà d'un trop grand nombre de paramètres libres, le modèle n'est plus identifiable. C'est à dire qu'il existe plus d'une solution qui permette de minimiser la différence entre la matrice de variance–covariance réelle et mesurée. Randy MacIntosh a contourné ce problème en faisant tourner le programme une première fois avec un nombre de liens réduit, puis de fixer les valeurs des coefficients ainsi trouvés pour refaire tourner le programme afin de découvrir celles d'autres coefficients (36). Malheureusement ces valeurs sont extrêmement dépendantes de l'ordre dans lequel la recherche a été effectuée.

Lorsqu'on ne dispose pas de modèle, il existe une technique exploratoire permettant de rechercher le meilleur modèle expliquant les données. Celle-ci a été appliquée par Ed Bullmore à des données d'IRMf dans une tâche de décision sémantique (37). Il a sélectionné les aires qui étaient activée par la tâche et il a débuté avec un modèle sans lien. Puis il a recherché le lien qui réduisait le plus la distance entre la matrice de variance–covariance réelle et mesurée. Ce lien était gardé et il recherchait le deuxième lien le plus important et ainsi de suite. Il a utilisé plusieurs critères d'arrêt de la procédure qui permettent de trouver le modèle le plus adapté tout en faisant une économie en nombre de liens. Cette approche pose deux types de problèmes : i) quel est le sens à donner au lien, en sachant que le même modèle a souvent la même significativité que le lien se trouve dans un sens ou dans l'autre, ii) la construction par morceaux pose un problème, car si on procédait deux liens par deux liens plutôt que 1 par 1, le résultat ne serait pas le même. Enfin le modèle trouvé ne peut servir à l'exploitation des données qui ont servi à sa construction. Il ne pourra servir de base que pour une autre série de données.

Voyons un exemple concret de l'utilisation de ce type d'approche en imagerie fonctionnelle, publié par Christian Büchel et Karl Friston (38). Ils se sont intéressés à la connectivité effective du cortex pariétal et du cortex infero-temporal au cours d'une tâche d'apprentissage objet–localisation. Pour cela ils ont défini un réseau dorsal pour le traitement de l'espace et un réseau ventral pour le traitement de l'objet, et exploré le lien qui les unissait. Ainsi la modulation de ce lien n'est pas liée au fait que l'activité de ce aires dépendent d'une même source (V1). On peut cependant contester le sens de ce lien. Ils ont mis en évidence au fur et à mesure des présentations, un déclin de l'activité de ces deux régions, alors que leur

connectivité effective augmentait. Ils ont même pu corréler l'importance de cette corrélation partielle avec la vitesse d'apprentissage des sujets.



Modification de la connectivité effective entre le cortex pariétal postérieur (PP) et inféro-temporal (ITp) au cours de l'apprentissage d'une association entre une position spatiale et un objet. Le modèle utilisé est illustré sur un hémisphère droit, en ayant mis en rouge la valeur du lien qui est testé. Le graphique de droite représente en abscisse la présentation pour laquelle ce lien était maximal, en fonction de la vitesse d'apprentissage en ordonnée.

Il existe d'autres approches que nous n'allons qu'évoquer puisque aucune d'elles n'a encore fait l'objet de publication en imagerie fonctionnelle. La première a été présentée par Jeffrey Binder au congrès HBM en juin 2001, elle consiste en l'utilisation d'un **réseau neuronal** de type Hopfield (c'est à dire totalement connecté, chaque nœud du réseau étant lié à tous les autres), dont les nœuds étaient les différentes régions présentant une activation durant une tâche verbale. Il faut remarquer qu'il n'est pas nécessaire de définir le réseau a priori puisque tous les liens sont explorés et sont susceptibles d'être réciproques. Au départ, les poids sont fixés au hasard à un seuil faible, puis les données sont présentées comme modèle et on cherche à ajuster les sorties du réseau avec ses entrées. L'apprentissage utilise une règle Delta[6]. La significativité des pondérations était estimée sur la base des distributions calculées en présentant des séries aléatoires. Cette méthode ne prend pas en compte le caractère autocorrélé des données IRMf, mais il existe des méthodes appropriées pour générer des séries plus réalistes. En scindant leur *pool* de données, ils ont même pu vérifier que les résultats étaient assez reproductibles entre la première et la seconde moitié.

Une seconde méthode vient d'être présentée par Habib Benali à la réunion SPM de juillet 2001. Elle est basée sur les **modèles graphiques d'association** encore appelés les champs de Markov aléatoires. Leur caractéristique est de ne pas être directionnel, mais le graphe est recherché automatiquement en se basant sur la loi de Wishart et sur la matrice de covariance.

[6] "Delta rule", ou règle des moindres carré, ou encore règle de Widrow-Hoff va optimiser les poids entre deux unités en utilisant comme fonction de coût entre l'état voulu et l'état actuel les moindres carrés ordinaires, et comme méthode d'optimisation une "steepest descente", c'est à dire en se déplaçant dans l'axe de la pente (calcul de la dérivé).

Des résultats intéressants, car assez proches de ceux qui étaient attendus, ont été obtenus sur une série de donnée IRMf dans une tâche de mémoire de travail.

Depuis quelques années, les statisticiens et les chercheurs en intelligence artificielle (formelle) ont réinvesti le concept de causalité. De nouvelles méthodes conjuguant approche statistique Bayesiennes et théorie des graphes (dont les modèles graphiques d'association font partie) permettent d'inférer une causalité sans nécessairement passer par une expérience contrôlée. Ce sont les **réseaux Bayesiens** (39). Ces méthodes vont entraîner un profond bouleversement sur la conception d'une expérience. Elles permettent d'inférer un lien causal entre A et B, sans avoir à maîtriser tous les autres paramètres (dont les liens causals pourront être estimés en même temps) et sans avoir à faire varier A, le paramètre supposé causal. Ce principe est tout à fait adapté à notre sujet, puisqu'en réalité nous ne contrôlons pas l'activité de A. Dans le cadre de l'intégration cérébrale fonctionnelle cette approche présente néanmoins l'inconvénient de ne pas pouvoir gérer les causalités cycliques. On parle de graphes directionnels acycliques (*DAG – Directed Acyclic Graphs*).

Ces techniques nécessitent un nombre important de données, et bien que les premières utilisations étaient effectuées sur des données PET, nul doute que l'IRMf par la quantité d'observation qu'elle génère soit plus adaptée. Bien entendu, rien ne s'oppose à ce que ces techniques s'appliquent aux données électrophysiologiques, moyennant la prise en compte d'éventuelles latences entre les aires.

Rappelons que la recherche automatique du meilleur graphe suppose pour toutes les méthodes que tous les nœuds aient été identifiés. Cela inclut toutes les formes de non-linéarité comme les interactions entre les nœuds qui doivent être explicitement définis sous forme de nouveaux nœuds. Pour cette recherche, les SEM sont mal adaptées à une étude exploratoire en raison de problèmes de construction et de directionnalité difficiles à définir. Sous la forme d'une recherche automatique du meilleur graphe, ils souffrent du même problème que les réseaux Bayesiens : ils ne permettent pas de mettre en évidence de liens cycliques.

La technique a priori la moins sensible à ce dernier problème semble être celle des réseaux neuronaux. C'est un problème important puisque beaucoup de modèles supposent des boucles de rétrocontrôle (expl. modèle du contrôle moteur de Frith), et la connectique cérébrale laisse supposer que la plupart des systèmes sont cycliques.

Technique	*Directionnalité*	*Recherche du graphe*	*Cyclique*
SEM	+	+/ –	+/ –
Réseau neuronal	+	++	+
Graphes d'association	–	+	
DAG	+	++	–

Tableau récapitulatif des techniques d'analyse de la **connectivité effective** et de quelques unes de leurs propriétés : i) l'évaluation de la directionnalité, autrement dit quelle est la cause et quel est l'effet, ii) la possibilité de laisser la technique rechercher le meilleur graphe sans avoir à le définir autrement qu'en fournissant la liste exhaustive de ses constituants, iii) la possibilité de prendre en compte des dépendances cycliques (par exemple A influence B qui influence C qui influence A).

Techniques non linéaires

Encore une fois, il est possible de **linéariser** le problème en utilisant les mêmes techniques que celles évoquées plus haut. A titre d'exemple, la décomposition en série de Volterra a été utilisée par Karl Friston pour une modélisation en SEM d'interactions entre différentes aires (30). Dans ce cas, chaque terme de la série de Volterra constitue un nouveau nœud du système. Cela complique l'élaboration du modèle. Mais rien n'empêche d'utiliser cette décomposition avec une méthode qui ne nécessite pas de définition préalable du modèle.

L'électrophysiologie simplifie le problème de la connectivité. Le critère temporel facilite l'établissement de la directionnalité du lien causal. A titre d'illustration, nous allons reprendre l'exemple de la crise décrit précédemment. Nous avons calculé **l'information mutuelle** en décalant les signaux de quelques millisecondes entre Amg et Pol (de –12 à +12 ms de décalage). On peut estimer que le **décalage** qui permet d'obtenir le coefficient le plus élevé correspond au temps nécessaire au passage de l'information d'une structure à une autre. On constate que par rapport à une ligne de base plutôt plate et centrée autour de 0 avant la crise, l'entrée dans la crise se caractérise par deux brèves périodes au cours desquelles Pol précède Amg de 6 ms, avant que la tendance ne s'inverse au moment de l'augmentation de l'IM, avec un délai autour de 2 à 4 ms entre Amg et Pol. Il faut souligner que les deux brefs instants ou Pol a précédé Amg semblent correspondre à des modifications du signal électrique.

Cette approche simpliste a été largement améliorée avec la technique du **transfert d'informations**, qui mêle les chaînes de Markov à l'information mutuelle pour tester plus systématiquement les différents délais entre les deux aires.

Les oubliés du cognitivisme

Le cognitivisme comme programme de recherche fonctionnaliste excluait certains objets de son programme de recherche. En particulier, émotion et conscience, deux pans de notre vie psychique étaient ainsi exclus par nature puisque non “utiles”. Certaines définitions même des sciences cognitives, comme celles que nous avons adoptées au départ, ont gardé la trace de cette vision exclusive. Nous verrons comment les modèles cognitivistes tentent de palier à ces défauts originels en soulevant la question de leur implication en imagerie fonctionnelle.

La “cognitivisation” des émotions

C’est vers les années 1980, que l’échec (relatif) de l’intelligence artificielle formelle, c’est à dire des systèmes logiques basés sur la manipulation de symboles, montre l’incapacité du cognitivisme à réaliser une description opératoire et adaptative des fonctions cérébrales. L’arrivé du connexionnisme et les premières simulations en milieu réel mettent à nu les limites d’une description du mental limitée aux “produits utiles”. Le simple problème de “Qu’est-ce qui pourrait faire bouger ces machines ?” débouche sur des notions aussi loin des préoccupations cognitivistes que la motivation. Pour que ces machines virtuelles ou réelles survivent dans un milieu simplifié, il est nécessaire d’implémenter une “appétence” pour tel objet.

Dans un même temps, mais indépendamment, sont décrits certains tableaux neurologiques secondaires à une lésion cérébrale qui ne peuvent s’expliquer par la disparition d’un module cognitif classique. Ainsi en est-il de la description du tableau de perte de l’auto activation psychique secondaire à une lésion bipallidale par Laplane (40) ou des troubles psychopathiques secondaires à une lésion fronto-basale par Damasio (41). Il ne s’agit pas de trouble des fonctions instrumentales ou exécutives objectivables par des tests classiques, mais des troubles du comportement dans sa motivation, sa programmation et son élaboration. Leur caractérisation est devenue l’objet d’une nouvelle spécialité : la neurologie comportementale (“*behavioral neurology*”).

La motivation et les émotions doivent donc réintégrer le champ d’une description de l’activité mentale. Damasio fonctionnalise les émotions, en leur assignant un rôle dans la prise de décision (41). En termes cognitivistes il décrit des modules et des représentations émotionnelles, avec la possibilité de faire une véritable imagerie mentale, une simulation des

émotions (la boucle “comme si”) impliquée dans la planification du comportement. C’est ainsi qu’on voit fleurir une nouvelle vague de travaux en imagerie fonctionnelle, qui d’une dizaine à HBM en 1997 sont passés à plus de 200 en 2001. Mais de nombreux problèmes se posent, à commencer celui de définir une émotion. Existe t’il des émotions primaires (~affect), combien sont-elles, qu’est-ce qu’une émotion secondaire, une émotion d’arrière plan[7] (~humeur), comment les caractériser, comment les provoquer ? Enregistre t’on les effets d’une émotion, les circuits de l’émotion ou le ressenti de l’émotion ? Comment les émotions interfèrent-elles avec la cognition ou avec le comportement … ? Ces dernières questions nous font glisser vers l’autre oubliée du cognitivisme : la conscience.

Cognition et conscience

La “conscience”, cesse d’être un objet non scientifique dès le début des années 1990. Cette “révolution” émerge hors du courant du cognitivisme, et même hors des sciences cognitives classiques. Les modèles cognitivistes n’ont pas été développés pour remplir ce rôle. Tant que les processus qu’ils décrivent sont assurés, le vécu phénoménologique pourrait tout aussi bien ne pas exister. Ce n’est pas pour autant que le cognitivisme ne cherche pas à incorporer ce nouvel objet.

La conscience du processus

Une première approche consistait à se demander si la computation était accessible à la conscience. Pour cela on pouvait proposer deux façons d’approcher le problème :

i) On peut s’intéresser à la part de computation consciente durant l’exécution d’une tâche (mais la conscience peut potentiellement intervenir après l’exécution de la tâche). Par exemple, lorsqu’une tâche est réalisée sur un mode automatique en comparaison à la même tâche réalisée sur un mode contrôlé et/ou attentionnel. Nous avons vu plus haut que l’apprentissage d’une tâche se manifestait par une augmentation initiale de l’activation (mode contrôlé/attentionnel) avant de décliner bien en dessous de sa valeur initiale une fois qu’elle est effectuée en mode automatique. Cette observation est malheureusement entachée d’un effet de répétition et de mémorisation. Mais d’autres expériences ont été réalisées en présentant le même stimulus au sujet, soit en

[7] Selon Damasio, les émotions primaires sont la joie, la tristesse, la peur, la colère, la surprise, et le dégoût. Les émotions secondaires ou sociales seraient l'embarras, la jalousie, la culpabilité, et l'orgueil. Enfin les émotions d'arrière plan seraient la sensation de bien-être, de malaise, de calme, de tension, de fatigue, d’énergie, d’exaltation, de relâchement, le sentiment d'émerger, de pesanteur, de stabilité ou d’instabilité, d’équilibre, de fragilité, d’harmonie ou de discorde.

condition attentionnelle, soit en condition non attentionnelle. Par exemple, la perception d'un champ d'étoiles en mouvement (*"star field stimulation"*) entraîne non seulement un renforcement de l'activation de V5 (aire qui traite les stimuli en mouvement), mais en plus elle renforce sa connectivité avec l'aire pariétale postérieure (30).

ii) On peut aussi faire en sorte que le sujet ne se rend pas compte des processus qu'il utilise en n'explicitant pas la tâche (que ce soit pendant ou après la tâche). Nous avons vu que l'effet de *priming* par la répétition (*"repetition priming"*) s'accompagnait d'une diminution d'activité (42). Il en est de même en mémoire implicite qui pourtant est supposée reposer sur les mêmes modules, et qui repose effectivement sur des aires identiques.

Il semble ainsi que l'accès conscient aux processus d'exécution de la tâche entraîne une plus forte activation des régions concernées. Mais il y a dans tous les cas un confondant qui est le niveau d'attention prêté au processus.

L'approche expérientielle

Une seconde approche dans laquelle pourrait s'inscrire l'étude des émotions est dite approche "expérientielle". On ne s'intéresse plus qu'aux seules performances, mais aussi au vécu du sujet. Cette forme de phénoménologie[8] simplifiée s'est illustrée en mémoire épisodique avec le paradigme R/K (*Remember/Know*, je me rappelle/je sais – Tulving, Gardiner) ou la position d'observateur ou d'acteur dans l'évocation d'un souvenir, en imagerie mentale avec sa vivacité (*"vividness"* – Baddley) ou sa précision qui pourrait conditionner l'implication de V1 (*"depictive"* – Kosslyn) … Mais toutes ces approches ont en commun de retenir une interprétation cognitiviste classique.

Attardons-nous sur le paradigme R/K (*Remember/Know*, je me rappelle/je sais) en reconnaissance (43). Sur le plan expérientiel, prenons l'exemple d'un sujet qui voit un visage sur lequel il ne parvient pas à mettre un nom, mais qu'il sait connaître. C'est l'état "je sais" (*Know*). En cherchant un peu il retrouvera le nom et tout le contexte qui l'accompagne comme « quand » et « où » il l'a vu pour la dernière fois. C'est l'état "je me rappelle" (*Remember*). L'interprétation de ces états de conscience (et non pas conscience tout court) se fait en terme de distinction entre un rappel en mémoire épisodique évoquant tout le contexte

[8] Le terme de phénoménologie a ici une acceptation limitée à la description des phénomènes conscients et à leur succession. A la différence de l'introspection pure, elle prend en compte les phénomènes conscients dans leur rapport au monde, joignant ainsi "l'extrême subjectivisme à l'extrême objectivisme" (Merleau-Ponty – "La phénoménologie de la perception").

du souvenir (*Remember*), et un rappel en mémoire sémantique n'évoquant qu'un simple sentiment de familiarité (*Know*). Et lorsque le substratum organique de ces deux états est abordé en imagerie fonctionnelle, l'interprétation ne se fait pas plus en termes de conscience, mais en termes de processus. Par exemple, Henson et coll. ayant observé une plus forte implication de la région frontale droite lors d'une réponse "K" vs. "R", en ont déduit que l'état "K" impliquait plus de monitoring que l'état "R" (44). On ne parle pas du corrélat neuronal de l'état de conscience, mais de la conséquence cognitive de celui-ci.

La conscience comme un fait

La conscience ne serait-elle pas opérationnalisable en termes cognitivistes ? Il faut admettre que de telles tentatives restent assez rares. Sur le plan théorique Jerry Fodor avait proposé qu'au-dessus des modules encapsulés, il puisse y avoir un système superviseur recevant en entrée les informations des différents modules, et dont l'une des caractéristiques serait justement d'être conscient. Une décision très arbitraire qui n'est pas éloignée des théories de l'homoncule et qui souffre des mêmes incohérences[9].

La solution simple qui consiste à dire « la conscience est là parce que c'est ce qu'on observe » est fréquemment utilisée : ce module est conscient, et celui-ci ne l'est pas parce que telle performance implique la conscience et telle autre ne l'implique pas. Il n'y a pas d'élaboration générale sur le problème de la conscience, mais une adaptation au cas par cas. Prenons l'exemple de l'hypothèse de Milner et Goodale (45;46). Ces auteurs ont décrit le cas d'une patiente DF, qui après une lésion infero-temporale bilatérale (voie visuelle ventrale), gardait certaines performances conservées, sans être capable de les faire émerger à la conscience. Par exemple, elle ne rapportait pas avoir conscience d'une orientation horizontale ou verticale, mais pourtant, elle adaptait l'orientation d'une lettre correctement lorsqu'il s'agissait de la glisser à travers une fente horizontale ou verticale. L'interprétation retenait que le système dorsal utilisait des coordonnées égocentrées, adaptées à l'élaboration du mouvement, et fonctionnait sur un mode inconscient. A l'inverse, le système ventral travaillant en coordonnées allocentrées permettait de reconnaître consciemment une orientation (mais à l'inverse il ne permettait pas une programmation motrice efficace comme dans l'ataxie optique[10]).

---

[9] Dans le XVIIIème siècle de l'homme machine, on imaginait qu'un petit homoncule dirigeait toute la machinerie humaine. Comme cela ne faisait que déplacer le problème sans l'expliquer, on aboutissait à une régression sans fin : un autre homoncule dans le tête de l'homoncule etc…

[10] L'ataxie optique se manifeste par une perte de la précision des gestes visuo-guidés en l'absence de troubles sensitivo-moteurs, cérébelleux ou extrapyramidaux. La patient reste capable de designer des parties de son corps,

La conscience comme une fonction

Quelle fonction la conscience pourrait-elle jouer dans le système cognitiviste ? En 1990 Francis Crick avait proposé un rapprochement entre la conscience et le problème de liaison (47), considérant qu'ils relevaient peut-être d'une solution commune. Le problème de liaison (*"binding problem"*), a été évoqué par Milner en 1974. La question était de savoir comment l'information spatiale, de forme, de texture, de couleur etc… a priori traitée par des modules spécifiques, était synthétisée en une vision unique et cohérente d'un objet[11]. Francis Crick considérant que cette intégration des informations était une propriété de la conscience, proposait d'appliquer les solutions trouvées au problème de liaison à celui de la conscience. Une de ces solutions stipulait que l'intégration des informations de l'ensemble des modules était réalisée au niveau d'une aire en aval sur laquelle les projections des différents modules convergeaient.

Les théories cognitivistes sur la conscience reposent toutes sur cette *"higher order theory of conscious states"* (théorie d'un ordre élevé des états de conscience – Rosenthal)(48;49). Celle-ci stipule par exemple, que pour être conscient de regarder un arbre, il faut se forger une image mentale de soi regardant un arbre. Autrement dit, il faut se forger une méta-représentation, une représentation des représentations de soi et de l'arbre. Cette méta-représentation est le résultat de processus intégratifs.

Voyons les deux exemples les plus élaborés qui s'inscrivent dans un cadre cognitiviste. Le premier historiquement est le modèle de Frith, qui s'intéresse surtout à la partie intentionnelle de la phénoménologie, alors que le modèle de Damasio s'adresse plus à sa partie phénoménale.

Christopher Frith se consacre à l'étude des hallucinations et tout particulièrement au syndrome d'influence. Il a développé pour cela une théorie de la conscience dans le cadre du modèle cognitiviste essentiellement appliqué à la conscience de l'action. Autrement dit, il s'interroge sur ce qui affleure à la conscience lors d'un acte moteur. Il part du constat suivant : nous ne sommes pas conscients de la préparation et de l'exécution motrice, sauf lorsque le résultat obtenu ne correspond pas au résultat attendu. Ainsi, ce qui devient

---

puisque cette fonction est sous dépendance du système proprioceptif. Ce trouble est liée à une lésion pariétale postérieure qui fait parti de la voie visuelle dorsale travaillant en coordonnées égocentrées.

[11] Cette formulation est une simplification du problème de liaison. Celui-ci a émergé comme solution au problème de superposition. Si un sujet voit une pomme et une banane proches l'une de l'autre, comment les neurones de l'aire de la forme, codant pour la banane, savent que les neurones codant pour le jaune code pour le même objet qu'eux, et ne confondent pas avec les neurones codant pour le rouge de la pomme ? Sa résolution par l'idée d'une méta représentation résout le problème en imposant des contraintes spatiales fortes. Mais la résolution spatiale quelle suppose dépasse celles des neurones codants pour des attributs de forme ou de couleur.

conscient est une méta-représentation de deux représentations : celle d'un état final objectif de la position des membres, et celle d'un état prédit après affinement de l'état voulu par un calcul inverse et sa résolution par le calcul direct (cf. schéma). Le processus qui réalise cette méta-représentation est un processus de monitoring basé sur un comparateur à priori frontal (49;50).

Le modèle d'Antonio Damasio sépare deux types de conscience, la conscience noyau, sorte de conscience minimale préservée dans le mutisme akinétique[12], et la conscience étendue permettant l'intégration de l'individu dans un continuum temporel (51). La conscience noyau est considérée comme une méta-représentation d'une représentation de l'état physique de l'organisme et de celle de l'objet perçu. La conscience étendue est conceptualisée comme une nouvelle méta-représentation de la conscience noyau et des représentations mnésiques intégrant en particulier la mémoire autobiographique (cf. schéma). Damasio n'a pas élaboré de processus susceptibles d'aboutir à ces méta-représentations, qu'il nomme cartes de deuxième ordre, mais il fait l'hypothèse que chacune est sous-tendue par des régions cérébrales spécifiques.

Ces modèles présentent certaines faiblesses :

1) Ils n'expliquent pas en quoi une représentation et une méta-représentation sont dissemblables pour être phénoménologiquement si différentes. On peut se demander si la conscience rempli véritablement une fonction computationnelle, puisque c'est juste l'hypothèse "méta-représentation = conscience", qui permet au modèle de lui reconnaître un autre objectif qu'elle-même. Une falsification de cette hypothèse serait de mettre en évidence des méta-représentations sans conscience, ce qui semble être le cas lors de l'exécution de certaines tâches automatiques.
2) Ils n'expliquent pas pourquoi une et une seule méta-représentation est présente en même temps, la conscience étant considérée comme une et indivisible.
3) Il en découle que ces modèles n'expliquent pas non plus comment s'opère ce passage de l'une à l'autre méta-représentation pour l'accès à la conscience.

Cette définition s'est enrichie, puisqu'il y a actuellement de multiples acceptations du problème de liaison qui ne sont pas toujours superposables.

[12] Suite à la lésion de certaines structures cérébrales (principalement l'aire cingulaire antérieure de façon bilatérale), le sujet reste les yeux ouverts, le regard vide, sans parler ni bouger. Il persiste néanmoins la capacité de prêter une attention fugace et de faible niveau à un évènement ou à un objet saillant.

Sur le plan de l'imagerie fonctionnelle, ce type de modèle amène deux prédictions :

1) La méta-représentation et/ou le processus intégratif seraient liés à l'activation d'une aire spécifique. Spécifique, car pour une autre méta-représentation, une autre aire serait concernée. On remarque que les processus intégratifs reposent souvent sur des processus exécutifs[13] (cf. modèle de Frith), particulièrement difficiles à cartographier puisqu'on cherche encore le critère de ségrégation (par type de processus, type d'objet traité …).
2) Il devrait y avoir une connectivité effective entre les aires de représentation "primaires" et l'aire supportant la méta-représentation. Cette connectivité devrait aller dans le sens d'une convergence sur la méta-représentation où des interactions entre représentation "primaires" seraient possibles. En revanche un retour sur les représentation et une interaction entre elles sont exclues de ce schéma.

[13] Une fonction exécutive est une fonction dédiée à la mise en œuvre et au contrôle des actions. Elles permettent l'anticipation, la définition des buts, la sélection et la planification de la réponse appropriée, l'allocation de l'attention dans le traitement de l'information liée à ces objectifs, puis la surveillance du bon déroulement de l'execution de l'action. Elles exercent donc un contrôle sur les fonctions dites instrumentales comme la perception sensorielle, la motricité, le langage etc…

L'alternative neurodynamique
Un nouveau paradigme pour l'intégration fonctionnelle cérébrale

> *"Le cognitivisme, en tant que programme de recherche ... a bel et bien rendu possible l'essor des sciences cognitives. Mais celles-ci ne sont pas en réalité tributaire du cognitivisme ... en sorte qu'on peut rejeter le cognitivisme en tant que doctrine positive sans pour cela estimer vains les efforts accomplis actuellement pour donner corps à une science de l'esprit".* (Daniel Andler 1999 (6))

Il n'y a pas de courant neurodynamique aussi "lisible" que le courant cognitiviste. En fait, le terme même de "neurodynamique" n'est pas reconnu comme qualificatif d'un nouveau courant en neurosciences, mais nous l'adopterons par commodité en tentant de le définir. Nous verrons que cette conception du fonctionnement cérébral reste compatible avec la plupart des apports du cognitivisme, mais qu'elle apporte un renouveau dans la conception de l'intégration fonctionnelle cérébrale (IFC).

Les fondements du courant neurodynamique

Dès 1937 Gaston Bachelard lançait un appel pour un abord scientifique de la complexité (52), tenue comme incongrue depuis 3 siècles gouvernés par le discours de la méthode prônant la simplicité (René Descartes – 1637). En 1948, Warren Weaver, un professeur du MIT qui s'illustrera dans la promotion de la théorie de l'information, publiait un article *"science and complexity"*. Il proposait de lire l'histoire de la science moderne en y reconnaissant les émergences successives de trois paradigmes : i) le paradigme de la simplicité qui s'était développé de 1600 à 1800 en proposant des modèles objectifs causalistes quantitatifs et certains, ii) le paradigme de la complexité désorganisée qui s'était formé à partir de 1850 avec le développement de la thermodynamique et de la cinétique chimique, iii) et le paradigme de la complexité organisée que Weaver voyait émerger à la fin des années 1940.

Le behaviorisme a été la première étape dans l'abord du système complexe qu'est le cerveau. Celui-ci était considéré comme inabordable (concept de la "boite noire") et on n'envisageait que ses entrées et ses sorties. Le cognitivisme était la seconde façon d'aborder les systèmes

complexes dont on reconnaissait le caractère organisé. Les systèmes cybernétiques, à l'époque modèles d'une complexité organisée, illustraient une méthode de décomposition possible. Herbert Simon qui appliquait le paradigme de la complexité organisée à l'économie, proposait des règles pragmatiques à la décomposition d'un système complexe en sous système : i) les interactions entre sous-systèmes devaient être relativement peu nombreuses et aisément identifiables, ii) et les interactions au sein d'un même sous-système devaient être plus nombreuses et relativement stables au fil du temps. L'approche cognitiviste est la parfaite illustration de l'application de ces principes. Elle a ainsi pu proposer un niveau de description de l'activité mentale doté de concepts et de méthodes opératoires. Cependant si le cerveau peut, comme tout système complexe être décrit en terme de sous-systèmes, il ne faut pas oublier qu'il ne s'agit que d'une approximation.

D'autres approches de la complexité vont émerger dans les années qui suivrent. En suivant un découpage similaire à celui proposé par Warren Weaver, on peut reconnaître actuellement trois branches au concept de complexité organisée : i) une complication du paradigme de la simplicité dont le cognitivisme est une illustration, au même titre que les réseaux d'automates, ii) un ordonnancement de la complexité désorganisée qui nous fait passer du hasard au chaos déterministe, ouvrant la voie à l'étude de la dynamique des systèmes non linéaires (théorie des états critiques, théorie de la bifurcation, théorie des catastrophes, théorie des orbite périodique, notion d'attracteurs etc…), iii) une approche "systémique" (interactionniste, synergétique – "*synergetics*") qui se développe à l'interface des deux premières approches pour mettre en avant le comportement coopératif des systèmes complexes comme les réseaux neuronaux (52).

C'est surtout le milieu de l'électrophysiologie qui s'est saisi de ces nouveaux instruments chaos et synergétique pour les appliquer à l'interprétation de leurs données. Mais au-delà de l'application technique, l'élaboration d'un véritable programme de recherche basé sur ces principes s'ébauche à peine. En plus de cette approche commune de la complexité, les "neurodynamiciens" ont en commun l'ambition de parvenir à une description complète de l'activité mentale. Celle-ci doit en particulier inclure la conscience dans la conception même du programme de recherche.

Le programme de recherche neurodynamique

Comme il n'existe pas de programme neurodynamique "reconnu", celui qui va suivre est construit à partir des éléments plus ou moins communs des approches suivantes : le programme neurophénoménologique de Francisco Varela (53), l'approche pragmatique de

Walter Freeman (18), ainsi que celle de Guilio Tononi et Gerald Edelman (54). Cette liste parait courte, mais il faut considérer qu'elle n'adresse que les auteurs ayant formalisé un programme de recherche. Il est probable que de nombreux autres neuroscientifiques se reconnaissent dans cette approche (55).

Le "credo" du programme de recherche neurodynamique est que le cerveau est un système complexe auto-organisé et adaptatif. Son objectif est de parvenir à une description des rapports entre le physique et le mental dans un cadre écologique. La phénoménologie mentale, intentionnelle et comportementale émerge de l'organisation des états physiques du système soumis aux stimulations de l'extérieur.

Le programme de recherche neurodynamique fait donc un certains nombre d'hypothèses :

1) Il pose l'identité des substances physiques et mentales. Ces dernières sont comprises comme des niveaux de description différents du même objet (monisme). Cela signifie aussi qu'il n'y a pas d'essence psychique sans support physique, ce qui est une position matérialiste. Le fonctionnalisme n'est pas exclu, mais il ne correspond plus qu'à un sous-espace de "projection" parmi d'autres. Un autre sous-espace possible serait le finalisme (la téléologie). Cela n'est pas sans importance sur le plan de la causalité. En effet, la causalité finaliste devient aussi intéressante à décrire que la causalité efficiente. C'est d'un intérêt tout particulier pour l'étude du corrélat neuronal de la conscience. On peut se contenter d'une description de l'état du système associé à une phénoménologie particulière sans devoir démontrer comment la phénoménologie est expliquée par l'état du système. C'est peu satisfaisant pour des scientifiques élevés dans le culte du cartésianisme, mais c'est le fondement même de l'épistémologie non cartésienne dont Gaston Bachelard faisait l'apologie. Cependant, il ne s'agit pas d'une nouvelle pirouette philosophique pour esquiver la "pierre d'achoppement" (le "*hard problem*" de David Chalmers), c'est à dire de définir le support nécessaire et suffisant de la conscience. Le programme n'exclu pas une compréhension en terme de causalité efficiente.
2) L'idée de l'indépendance du traitement par rapport à son support physique reste, mais conditionnée par une équivalence de "computation". Autrement dit, l'ordinateur peut théoriquement disposer d'un "équivalent d'esprit", mais il faudrait pour cela que la façon dont il réalise les opérations (l'algorithmique) reproduise parfaitement celle du cerveau. En pratique, le programme n'a plus l'ambition de développer des solutions technologiques "intelligentes" exploitables, mais de simuler la computation cérébrale.

L'intelligence artificielle a laissé la place aux neurosciences computationnelles, qui aident à la compréhension de la complexité du système.

3) Le cerveau est un système intelligible en sous-systèmes fonctionnels variants en fonction des états antérieurs du système et de la tâche à réaliser. Ainsi dans certains cas une découpe fonctionnelle de type cognitiviste est une bonne approximation, alors qu'elle peut s'avérer erronée dans d'autres où les neurones des différentes aires participent au traitement en un ensemble cohérent non décomposable (unitaire). On passe alors d'une computation ségrégée, parallèle par morceaux s'enchaînant en série à une computation entièrement parallèle et intégrée. En pratique, cela signifie que l'aire n'est plus le niveau de description approprié, mais qu'il faudrait envisager un système comprenant plusieurs aires. On fait l'hypothèse que la conscience est le corrélat de cette intégration de l'activité d'une large population de neurones parce qu'elle a les même propriétés d'unicité et d'intégration. Cette proposition est exclusive, et donc suppose que les opérations automatiques ou implicites ne reposent pas sur la coordination de larges populations neuronales ségrégées mais correspondent plus à une description de type cognitiviste. Cette approche pourrait s'avérer particulièrement pertinente pour des fonctions intégrées comme les fonctions executives qui ne répondent pas à une découpe claire. Enfin, il faut remarquer que cette conception apporte une nouvelle solution au problème de liaison (*"Binding problem"* cf. plus haut). La synthèse de l'objet en un percept unique et cohérent ne correspond pas à l'activité d'une aire de méta-représentation, mais à l'intégration des multiples composantes de l'objet, codées par des populations neuronales distinctes et ségrégées.

   Le passage du niveau physique élémentaire au niveau phénoménologique fait intervenir le concept d'emergence. Broad l'a conceptualisé dans son livre en 1925 en parlant des "propriétés d'un système qui ne peuvent être déduites de la connaissance complète de celles de ses composants pris séparément ou combinés et de leur organisation"(56).

4) Le cerveau est un système autonome plongé dans un monde de sensation qu'il doit interpréter. Comme tout système complexe, son état n'est pas seulement dépendant des ses entrées, mais aussi et même surtout des états qui les ont précédés. Il faut l'envisager comme un système ayant une dynamique interne, propre, influencée par les sensations, mais pas mené par elles. Dans le système intact, cette dynamique est tout entière tournée vers l'action (3;9), et les sensations ne prennent de signification

que par l'action qu'elles pourraient entraîner[14]. C'est cette dynamique tournée vers l'action qui est le moteur du développement ontogénique. Enfin, le cerveau ne peut pas être considéré comme un système stable établi une bonne fois pour toute. Au fur et à mesure qu'il apprend, ses connexions changent ce qui modifie sa dynamique. Cette transformation du système au cours de son évolution est un niveau de complexité supplémentaire qui n'a pas été abordé par les modèles physiques dont sont issus les paradigmes actuels.

Ainsi le programme de recherche neurodynamique n'a pas la même vertu fédératrice que le programme cognitiviste. Il intéresse avant tout les neuroscientifiques, mais se sert des instruments mis au point par d'autres scientifiques : mathématiciens (simulation sur ordinateurs), physiciens (dynamique des système complexes) ou ingénieurs (théorie de l'information)… Il inclut toute la part de l'approche cognitiviste qui concerne l'étude du cerveau et est particulièrement adapté au domaine de l'imagerie fonctionnelle puisqu'il adresse directement la question des rapports entre le physique et le mental. Ses implications sont majeures pour l'étude de l'intégration fonctionnelle cérébrale, car c'est sans doute ce qui tranche le plus par rapport au cognitivisme.

Les paradigmes neurodynamiques

L'approche neurodynamique est basée sur les paradigmes de la complexité organisée. Nous allons dans un premier temps définir la notion de complexité. Puis nous aborderons la notion d'organisation dans un système complexe avec les paradigmes synergétiques proposés par Guilio Tononi. Dans ce qui va suivre, nous mettrons surtout l'accent sur la notion d'intégration fonctionnelle qui fait intégralement partie des paradigmes neurodynamiques. C'est pourquoi nous n'aborderons pas les paradigmes issus des méthodes d'analyse non linéaire des systèmes dynamiques.

La notion de complexité

Un système comme le cerveau peut prendre un grand nombre d'états. Guilio Tononi définit cet état comme une combinaison des valeurs prises par les éléments du système à un instant donné. Nous pouvons par exemple définir les éléments de base du cerveau comme étant les

[14] Ce point reprend en le généralisant, le concept d'affordance proposé par le psychologue James Gibbson. Les affordances sont les propriétés réelles des objets qui peuvent avoir une valeur utile pour leur observateur. Elle porte sur ce que l'on perçoit en fonction de ce sur quoi on peut agir. Par exemple, nous percevons qu'un petit objet est préhensible, alors qu'un grand ne l'est pas. Ce concept avait bien antérieurement été développé par les philosophes comme Thomas d'Aquin sous le nom d'assimilation.

neurones auxquels nous reconnaissons 2 valeurs : repos (0) ou excitation (1). Nous allons pouvoir observer les différents états que prendra notre système au cours des différents instant qui suivront. Si le nombre d'états est important, qu'ils sont tous équiprobables, et non dépendants des états précédents, il est impossible de prédire le prochain état que prendra le système. Cela signifie que l'apparition d'un nouvel état apporte une réduction majeure de cette incertitude et donc une quantité d'informations élevée. Or la complexité serait cette "imprévisibilité essentielle" (Paul Valéry). John von Neumann avait donc naturellement proposé de mesurer la complexité à l'aune de l'information apportée par chaque nouvel état. Plus exactement la mesure pourrait être la probabilité de réduction de l'incertitude ou le potentiel informatif d'un nouvel état, autrement dit l'entropie H(X) (cf. plus haut)(32).

En conclusion, un système est d'autant plus complexe que le nombre d'états est élevé et qu'ils sont équiprobables. Nous avons introduit plus haut une troisième condition concernant la dépendance du système aux états précédents[15]. Dans ce qui va suivre ce point ne sera pas pris en compte, et c'est le principal reproche que l'on pourrait faire aux paradigmes tel qu'ils ont été proposés par Tononi. On peut contourner cette difficulté de deux façon différentes : soit en intégrant la dépendance temporelle dans les formules même (système markovien d'ordre n[16]), soit en introduisant la dimension temporelle à la définition de l'état (configurations de X à plusieurs instants). Ce dernier point soulève la question de la définition d'un état. L'élément de base aurait tout aussi bien pu être la synapse, ou la valeur d'un voxel cortical (où il faudrait prendre en compte la dépendance spatiale des données). Les résultats, et leur interprétation doivent donc être fonction du choix opéré.

Une mesure de l'intégration

Comme nous parlons de systèmes complexes *organisés*, cela signifie qu'il y a un élément de structuration. Cet élément est lié à l'interaction des éléments du système entre eux. Si les éléments étaient indépendants, l'entropie du système H(X) serait simplement la somme des entropies de ces éléments individuels $\Sigma_{x\in X}$ H(x). A l'inverse si les éléments présentent un certain degré d'interdépendance, l'entropie du système H(X) sera plus faible que la somme des entropies des différents éléments $\Sigma_{x\in X}$ H(x). Autrement dit, connaissant l'état d'une partie

[15] La profondeur de prédictibilité d'un système a été utilisée comme mesure de complexité. Par exemple les équipes de Roland Jouvent et de Francisco Varela ont mis en évidence une augmentation de cette profondeur de prédictibilité chez les sujets déprimés lors d'enregistrement EEG au cours d'une tâche cognitive. Celle-ci a été interprétée comme une réduction de la complexité chez les sujet déprimés (essentiellement localisé en frontal).

[16] Encore faut-il déterminer la profondeur de cette dépendance aux états antérieurs du système, c'est à dire combien de temps il est nécessaire de regarder en arrière pour déterminer l'état actuel du système.

des éléments, il serait possible d'en déduire l'état des autres. Moins il faut d'éléments pour déduire l'état des autres et plus le système est considéré comme intégré (54;57).

$$I(X) = \Sigma_{x \in X} H(x) \quad - \quad H(X)$$

Plus I(X) est élevé, plus le système est interdépendant, donc plus l'ensemble aura un fonctionnement cohérent. Il faut remarquer qu'en gagnant en intégration le système gagne en prévisibilité et donc perd en complexité puisque H(X) diminue. En conséquence, cette mesure n'est pas directement utilisable comme estimation de l'intégration d'un système différencié.

Le regroupement fonctionnel – Une mesure de spécialisation et de différenciation

A présent plutôt que de considérer le système vu de l'extérieur, nous allons le considérer vu de l'intérieur. Nous allons nous demander s'il est possible de fractionner le système en sous-systèmes fonctionnels. Autrement dit en sous-systèmes fortement liés entre eux et plus faiblement liés avec les autres. D'une certaine façon, nous retombons sur la question de la ségrégation, mais à la différence du cognitiviste, l'unité de base n'est plus forcément l'aire, mais un réseau d'aires.

Nous avons vu qu'il était possible de mesurer la dépendance statistique entre un sous-ensemble (X_k) et le reste du système (son complément $X-X_k$) en terme d'information mutuelle (IM – cf. plus haut). Pour une sous partition X_k choisie, cela correspond à :

$$IM_{Xk,X-Xk} = H(X_k) \; + \; H(X-X_k) \; - \; H(X)$$

Cet indice est d'autant plus élevé qu'il existe une interaction entre X_k et le reste du système $X-X_k$.

Nous disposons donc d'une mesure d'intégration d'un sous-système $I(X_k)$, et d'une mesure de sa dépendance au reste du système $IM_{Xk,X-Xk}$. Nous avons donc les bases pour construire une indice de regroupement fonctionnel (RF) pour chaque sous-ensemble du système :

$$RF(X_k) = I(X_k) \; / \; IM_{Xk,X-Xk}$$

Plus le RF est fort, et plus le sous-système X_k peut être considéré comme une unité de découpage fonctionnel valide. A l'inverse plus il est faible et moins le sous-ensemble est un niveau de description fonctionnel pertinent(54;57;58).

Bien entendu ce regroupement fonctionnel doit être considéré comme dynamique et non pas comme une donnée établie et invariante. Il est donc supposé être modulé par la tâche ou plutôt par la façon dont le système la résout.

Ce concept a été directement testé en utilisant des données PET obtenues aux cours de plusieurs tâches cognitives dans une population de sujets schizophrènes et une population de sujet sains(58). En raison d'un nombre de mesure restreint (le PET génère un faible nombre

d'images), le problème de la dépendance n'a été envisagé que dans le cadre linéaire. Pour les mêmes raisons, mais aussi parce que les images étaient auto corrélées spatialement, le nombre de voxels utilisé était limité (à peine plus d'une dizaine). Sur cet ensemble de plusieurs tâches cognitives, les auteurs ont pu mettre en évidence la plus grande difficulté des sujets schizophrènes à forger des réseaux fonctionnels différenciés par rapport au reste du système. Cela peut s'interpréter de deux façons : i) soit les différentes aires ne parviennent pas à intégrer leur activité en un tout cohérent, ii) soit les schizophrènes ne parviennent pas à faire émerger un regroupement fonctionnel suffisamment indépendant pour éviter que le reste du système le parasite.

Cette interprétation nous fait toucher un point important de l'approche neurodynamique : le système doit être pris en compte dans son ensemble. Il est en effet tout aussi important de connaître les parties du système impliquées que celles qui ne le sont pas, c'est la notion de différenciation. Prenons un exemple réducteur : imaginons qu'un sujet regarde une pomme de couleur rouge. Pour qu'il puisse voir la couleur rouge, il faut que son sous-système "pomme rouge" implique le neurone codant pour le rouge et exclut ceux qui codent le jaune ou le bleu. Si tous les neurones codant pour la couleur étaient actifs en même temps, il n'y aurait plus d'information de couleur. Un exemple classique de disparition de la différenciation est la crise épileptique. Si celle-ci ne touche que certaines régions, comme les aires du langage par exemple, ces aires cessent d'être fonctionnelles. Autrement dit, plutôt que d'être à l'origine d'une symptomatologie positive, l'activité critique produit une symptomatologie négative. Aucune information ne peut émerger d'un système non différencié[17]. A l'extrême limite, l'implication de l'ensemble du système dans une activité cohérente, comme dans une crise généralisée, entraîne une perte de la conscience. Une définition opératoire de la conscience pourrait donc être l'intégration transitoire d'éléments différenciés.

L'instrument décrit ci-dessus est lourd à manier, ce qui explique qu'il n'a été utilisé que dans une seule publication en imagerie fonctionnelle. Cependant nous avons vu dans l'étude de la connectivité fonctionnelle des instruments adaptés si nous acceptons de nous restreindre au cadre linéaire. Ainsi les méthodes de *clustering*, de projection révélatrice ou d'analyse en composante indépendante donnent-elles des résultats assez semblables. De même, l'approche par les graphes d'association pourrait s'avérer pertinente pour peut que l'ensemble du système soit pris en compte. Cependant, bien que les instruments mathématiques soient les mêmes, les

[17] Chacun sait pourtant que les crises partielles produisent souvent aussi une part de symptomatologie positive. Pourtant celle-ci n'est pas à rapporter aux aires impliquées par la crise, mais à celles qui sont influencées par l'activité critique, c'est ce que Fabrice Bartolomei a appelé le réseau séméiotique.

conclusions qu'on en tire sont différentes. Dans un cas c'est l'illustration d'une connectivité de sens indéterminé, dans l'autre c'est la démonstration d'une coopération exclusive des éléments de l'ensemble du sous-système. Il faut par ailleurs remarquer que l'idée de connectivité effective n'est pas pertinente dans le cas d'un système intégré selon la définition neurodynamique. Lorsque tous des éléments collaborent, savoir si le neurone A est directement connecté au neurone B ou s'il y a un ou deux neurones intermédiaires n'apportent rien à la description du système. Dans le cerveau, tout neurone est supposé être à moins d'une vingtaine de synapses de n'importe quel autre, et cette distance est probablement inférieure pour les systèmes travaillant souvent ensemble.

Nous avons jusqu'à présent surtout évoqué l'intégration en imagerie PET ou IRMf. Pour ce faire nous avons implicitement reconnu le voxel comme unité du système. Nous avons donc fait abstraction du fait qu'un voxel représente l'activité d'une population différenciée, certains des neurones participants au noyau d'intégration, d'autres n'y participants pas. En électrophysiologie, il est possible d'user d'un artifice pour ramener cette étude à l'échelle du neurone et ne s'intéresser qu'à ceux qui participent à un sous-système intégré. En effet, pour qu'il y ait un potentiel de surface mesurable en EEG-MEG, il faut qu'une large population neuronale décharge en même temps. Or une synchronisation sans décalage de phase est une forme d'intégration : si les neurones A et B déchargent de façon synchrone, il suffit de savoir que A est activé, pour pouvoir dire que B l'est aussi. A et B sont interdépendants et donc participent à un sous-système intégré. Si A, B et les autres neurones synchrones sont proches, leur activité s'additionne et il est possible d'enregistrer un potentiel oscillant sur la fréquence de synchronisation F_s. Si le nombre de neurones participant au réseau augmente localement, on va observer une augmentation de l'énergie pour la fréquence F_s.

En pratique, il s'agit de mettre en évidence une augmentation de puissance pour une fréquence donnée après exposition à un stimulus. C'est le paradigme de réponse spectrale évoquée[18]. Il consiste simplement à calculer la puissance spectrale sur de courtes fenêtres temporelles pour chaque instant, et à utiliser la période pré stimulus comme référence pour détecter des écarts dans la période post stimulus. Cette décomposition dite temps fréquence se fait pour chaque électrode et pour chaque essai (cf. appendice B et figure 2). Pour faire ressortir l'événement pertinent, on effectue la somme les décompositions temps fréquences à travers les essais. Par exemple dans une tâche cognitive complexe (génération de verbe – cf.

[18] La technique du potentiel évoqué ne peut pas être utilisée. En effet, le potentiel évoqué correspond à une synchronisation neuronale déclenchée par le stimulus, alors que ce qui nous intéresse ce sont les rythmes induits, c'est à dire non temporellement liés au stimulus. Dans ce cas, c'est le système qui génère la synchronisation, qui émerge comme une caractéristique propre du système.

figure 1), il nous a été possible de mettre en évidence deux régions cérébrales où l'augmentation de puissance signait une plus forte intégration : une région frontale et une région pariétale gauche (cf. figure 4). La résolution temporelle de la technique a permis de montrer que cette augmentation correspondait à des étapes d'utilisation de la mémoire de travail (cf. figure 5). Il est donc possible de dire qu'il existe une activité localement plus intégrée, mais cela ne donne aucun élément concernant la différenciation locale. En revanche cela donne une idée de la différenciation pour l'ensemble du cortex, puisque les autres régions cérébrales ne semblent pas présenter de renforcement de leur intégration.

L'approche synergétique demande à prendre en compte le système dans son ensemble. Il faut donc se demander si les deux aires mises en évidence ne participeraient pas à un même système intégré, autrement dit, si elles n'étaient pas synchrones l'une avec l'autre. Pour cela il est possible de calculer la synchronisation des signaux entre deux capteurs (59), et de comparer une synchronisation de base en prenant comme référence la période pré-stimulus. Les écarts par rapport à la normale que l'on observe, seraient liés à la résolution de la tâche (cf. appendice B et figure 2). Dans cette même étude, il nous a été possible de montrer que ces deux aires se synchronisent lors des phases de la tâche utilisant la mémoire de travail (cf. figure 4). De façon surprenante eu égard aux temps de conduction, il n'y a aucun décalage de phase entre les deux aires, c'est à dire que les neurones de l'aire frontale semblent décharger en même temps que les neurones de l'aire pariétale (cf. figure 6). Dans ce contexte, il est impossible de parler de connectivité effective, mais plutôt de sous-système intégré durant la résolution de la tâche de mémoire de travail. Encore une fois il est impossible de segmenter le système entre une région frontale qui réactiverait le souvenir et une aire de stockage pariétale. La fonction est considérée comme intégrée et repose sur le sous-système fronto-pariétal envisagé dans un ensemble non fragmentable.

Conclusions

L'apport de techniques d'investigations capables d'enregistrer l'activité de l'ensemble du cerveau a offert des réponses à la question "qui fait quoi". Elles ont souvent montré en plus des aires attendues sur la base des données lésionnelles, un réseau d'aires dont on ne soupçonnait parfois pas l'implication. Bien qu'ancienne, la question de comprendre le fonctionnement d'ensemble s'est à nouveau posée à la nouvelle communauté des "cartographes du cerveau" (*"brain mappers"*).

L'approche dominante dans ce domaine s'inspire du courant cognitiviste. L'aire cérébrale y est l'unité de base. Elle est censée opérer un traitement de l'information entrante avant de passer l'information ainsi transformée à l'aire suivante. Cette conception repose sur une équivalence entre aire et module cognitif ("localisationisme" au sens fort), qui bien que décriée et parfois mise en défaut, reste la base du raisonnement. L'intensité de l'activation de cette aire est supposée refléter sa charge de traitement. L'intégration fonctionnelle se réduit dans ce contexte aux flux d'informations entre les modules, ou aires, impliquées. Ceci correspond à la notion de connectivité effective, issue de la neurophysiologie, c'est à dire la part de variance de l'activité d'une aire qui serait dépendante de l'activité d'une autre. On la distingue de la connectivité fonctionnelle qui décrit simplement un lien corrélatif entre deux aires pour lequel on ne sous entend pas de causalité. Cette vision de l'intégration fonctionnelle par connectivité effective a surtout été mise en avant dans le cadre de l'imagerie fonctionnelle dite d'activation (PET, SPECT et IRMf) ou l'adhésion au programme de recherche cognitiviste est la plus forte. Nous avons vu un exemple d'analyse de connectivité fonctionnelle sur des données d'IRMf utilisant la corrélation simple (interaction psychophysiologique), ou une mesure analogue mais susceptible de varier dans le temps inspirée des filtres auto régressifs, ainsi qu'une méthode non linéaire : l'information mutuelle. Nous avons évoqué les approches multivariées comme la PCA, les projections révélatrices et l'ICA. L'analyses en connectivité effective a été abordée par les modèles d'équations structurelles permettant de prendre en compte les coefficients de régression partiels. Nous avons également mentionné les techniques alternatives récemment proposées que sont les graphes d'association et les réseaux neuronaux de type Hopfield. Enfin, nous avons proposé des approches à ces différentes méthodes pour l'analyse de données événementielles en IRMf, un domaine ou elles n'ont pas encore été utilisées.

Bien que moins clairement définit comme courant de pensée, l'approche "neurodynamique" est une alternative au courant cognitiviste. Elle est surtout soutenue par les neurophysiologistes, tout particulièrement par la branche théorique et modélisatrice de cette discipline. Leur niveau d'analyse serait plutôt le neurone sur le plan anatomique, et/ou la population neuronale sur le plan dynamique. La notion d'information est dissociée de celle de signification. Ainsi l'activité d'un seul neurone si elle n'a pas de sens en soit, est pourtant porteuse d'information. Cette dissociation n'est pas le fait de la tendance dure des localisationistes pour lesquels l'activité d'une aire est porteuse de sens. L'intégration n'y est plus vue comme le simple passage d'une information d'un module à un autre, mais comme la cohérence de l'activité des diverses unités permettant l'apparition de fonctions nouvelles dites "fonctions émergentes". La notion même de localisation fonctionnelle devient relative et c'est le réseau dans son ensemble qui possède cette propriété, non divisible en sous-unités autonomes (modules cognitifs). Ces caractéristiques d'indivisibilité et d'émergence ont été rapprochées de celles de la conscience. Rien d'étonnant donc à ce que la plupart des équipes qui s'attaquent au corrélat physiologique de la conscience soient aussi les fers de lances de ce courant "neurodynamique" de l'intégration fonctionnelle. Ce dernier point est une nouvelle différence avec le courant cognitiviste dans lequel la conscience est exclue du programme de recherche. Nous avons vu un exemple de ce type d'approche chez l'homme en utilisant les signaux EEG et MEG par une analyse des synchronies et de la cohérence. Nous avons de plus envisagé une ouverture vers les interactions entre bandes de fréquence et les méthodes non linéaires dont les apports récents issus de la théorie de l'information.

Le courant "neurodynamique" n'exclut pas la possibilité que le cerveau dispose de plusieurs modes de fonctionnement et que plusieurs formes d'intégration fonctionnelles puissent coexister. Ainsi face à une situation connue entraînant une routine comportementale, un mode de fonctionnement proche du modèle cognitif pourrait être à l'œuvre. L'information (toujours dissociée d'un éventuel contenu sémantique) passerait d'un neurone à l'autre et les premiers potentiels d'actions arrivant dans les aires effectrices généreraient la réponse. A l'inverse en cas de situation nouvelle, d'apprentissage, d'évaluation d'une réponse ou de correction d'erreur (de façon générale lors de la mise en œuvre des fonctions exécutives et/ou de la conscience), c'est en terme de population neuronales transitoires que l'intégration fonctionnelle devrait être envisagée.

L'intégration fonctionnelle est un domaine de recherche amené à s'épanouir. Pour le psychiatre et le neurologue ce concept permettra peut-être de mieux appréhender la physiopathologie de maladies comme la schizophrénie, la confusion mentale, les maladies

démyélinisantes ou certaines formes de démences. Souhaitons que cette connaissance puisse inspirer de nouvelles approches thérapeutiques.

La progression dans la compréhension de l'intégration fonctionnelle viendra de la clarification des concepts très liés comme on l'a vu aux courants de pensées qui les sous-tend. Cela devra accompagner voir précéder l'amélioration des instruments tant techniques que d'analyse. Idéalement la technique d'acquisition devrait être capable d'enregistrer l'activité des unités de base (sans doute le neurone) de tout le système (le cerveau), avec une résolution temporelle adaptée (de l'ordre de la milliseconde pour le neurone). Les techniques d'acquisition répondant à ces critères dites d'enregistrement à large échelle ("large array recording") ne sont applicables pour l'instant que chez l'animal mais ne permettent pas de couvrir la totalité des neurones, loin s'en faut. L'instrument d'analyse idéal devrait pouvoir extraire de cette énorme masse de données bruitées, les interactions liées à un comportement donné. Face aux limitations techniques actuelles, la simulation serait un moyen puissant pour tester nos hypothèses. Elle mérite d'être développée moins dans sa forme de simulation fonctionnelle que dans sa forme de simulation d'un système biologique vraisemblable.

VU

Strasbourg, le ……

Le président du jury de thèse

Professeur Daniel GRUCKER

VU et approuvé

Strasbourg, le ……

Le doyen de la faculté de médecine de Strasbourg I

Professeur Pierre GERLINGER

Appendice A
Les modèles d'équations structurelles
Principes

Source : "*Structural equation with latent variables*"
KA Bollen, John Willey & Sons, 1989

Propriété de la covariance :

- cov(x,y) = E((x-E(x)) (y-E(y))) = E(x*y) - E(x)*E(y)
- c est une constante : cov(c,x) = 0
- cov(c*x,y) = c*cov(x,y)
- cov(x+y,z) = cov(x,z) + cov(y,z)
- REM : beta = cov(x,y)/var(x)

La causalité

La causalité (effective) repose sur 3 principes :

- L'***isolation*** toutes les autres variables étant maîtrisées et/ou randomisée, seule la variable x a un effet sur y. En fait en SEM il s'agit plutôt d'une ***pseudo-isolation***, en effet on suppose que dans la relation y = ax + e, le terme d'erreur ne soit pas corrélé avec x (cov(x,e) = 0). Cette condition est violée lorsque **1**) une variable explicative n'a pas été prise en compte. Ainsi, l'absence de prise en compte d'une variable ayant un effet directe sur y (appelée z) entraîne une absence de lien effectif entre x et y (l'effet indirecte masquant l'effet directe). Si la corrélation n'implique pas la causalité, on devrait rajouter que l'absence de corrélation ne permet pas d'exclure la causalité. **2**) Lorsqu'une (ou des) variable exogène se trouve affectée d'une erreur aléatoire ou lorsque la variable endogène influence de retour la variable exogène (lien en double sens non pris en compte). Dans ce cas, le terme d'erreur est aussi affecté du coefficient et la condition cov(x,ζ) peut être différente de 0. **3**) Lorsqu'une relation entre 2 variables est mal spécifiée : linéaire alors qu'elle ne l'est pas. Cela va entraîner une corrélation entre x et ζ. **4**) Lorsque l'on prend une variable autocorrélée dans le temps, et que sa valeur au temps t-1 sert de variable exogène, on va avoir une corrélation entre x et ζ.
- L'***association***, ou la corrélation entre y et x (sans influence d'une autre variable cf. isolation). Plusieurs problèmes peuvent émerger : **1**) La valeur statistique peut être biaisée

si le terme d'erreur (ζ) est *hétéroscédastique (variance inégale)* ou auto corrélée (l'erreur associée à l'estimation d'un coefficient peut biaiser la statistique). **2**) La multicolinéarité est la dépendance linéaire qui existe entre les différentes variables explicatives *d'une même équation*. Elle va augmenter l'erreur associée aux coefficients des variables colinéaires. **3**) Souvent le modèle est trouvé par tâtonnement : débute par un petit modèle, puis le modèle est complexifié. Le problème est que pour véritablement tirer une conclusion, il faudrait utiliser un nouvel échantillon.

- La ***direction*** du lien causal. Pour Hume (1977), la préséance de la cause est une condition nécessaire à la causalité. Le problème est qu'il faut alors connaître la relation temporelle entre la cause et son effet, et surtout disposer de l'échantillonnage suffisant. Si l'échantillonnage est suffisant, il est proposé d'utiliser un modèle explorant les dépendances temporelles multiples :

$$Y_t = \sum_{i=1-k} \beta_i Y_{t-i} + \sum_{j=1-l} \gamma_j X_{t-j} + \zeta_t$$

Malheureusement, le terme d'erreur (ζ) est autocorrélé avec les valeur des x et y avec délais violant la condition de pseudo-isolation..

On peut noter que la manipulabilité d'un facteur n'est ni une condition suffisante, ni une condition nécessaire à la causalité. Enfin rappelons que le modèle explique les données, pas la réalité. Pour pouvoir extrapoler à la réalité, il faudrait que plusieurs jeu de données confirment la validité d'un modèle par rapport à un autre (on ne peut que falsifier un modèle, pas le prouver). Enfin le principe de parcimonie (rasoir d'Okham) veut qu'entre 2 modèles expliquant les données de la même façon, ce soit le plus simple qui soit gardé.

Variables

Description des différentes variables :

- Variables : latentes endogènes (η eta), latentes exogènes (ξ xi dont la matrice de covariance est Φ phi), observée ou manifeste (proxies, indicator, measures) soit exogène (x dont le nombre est souvent représenté par q et dont la matrice de covariance est avec comme Φ éléments ϕ), soit endogènes (y dont le nombre est p).
- Path coefficient de var exogènes à var endogènes (γ ou Γ gamma) ou entre var endogènes (β ou B bêta) : Coefficient standard = coef non standardisé * std(var explicative)/sdt(var affectée). Ce coefficient exprime le changement de la variable affectée en unité de déviation standard, après la modification de la variable explicative d'une unité de

déviation standard en gardant les autres valeurs constantes (Bollen p.13, p.69). Attention il est plus délicat à interpréter si il s'agit d'un terme d'interaction.

- Terme d'erreur (ζ zêta dont la matrice de covariance est Ψ d'élément ψ psi). Normalement les termes d'erreur ne devraient pas être corrélé les uns aux autres et donc la matrice est diagonale.

  Rem : si on tiens à prendre en compte les erreurs de mesure, il est nécessaire d'utiliser alors un passage par des variables latentes qui seraient la valeur exacte, influençant la variable observée dont la mesure est entachée d'une erreur (le coefficient ainsi calculé est souvent *supérieur* au coefficient calculé sur des variables observées).

L'analyse de chemin

L'analyse de chemin n'utilise que des variables manifestes

$$y = By + \Gamma x + \zeta.$$

avec $B = p \times p$, $\Gamma = p \times q$, $y = p \times 1$, $x = q \times 1$, $\zeta = p \times 1$.

Il en existe de 2 types : **1**) Les modèles récursifs sont des modèles qui n'admettent pas de lien de causalité inverse, que ce soit directement ou indirectement. Alors B peut s'écrire comme une matrice triangulaire basse, et la matrice de covariance de l'erreur Ψ est diagonale (le terme d'erreur d'une équation n'est pas corrélé avec les autres). Ces modèles sont identifiables. **2**) Les modèles non récursifs.

L'hypothèse H0 est de dire que :

$$S = \Sigma(\theta)$$

S est la matrice de covariance de l'échantillon (mesurée, dont le nombre d'élément non redondant est de $n*(n-1)/2$ avec $n = p+q$) et $\Sigma(\theta)$ est la matrice de covariance fonction des paramètres libres du modèle (θ). Elle est décomposable en 3 parties :

- La matrice de covariance des y : $\Sigma yy(\theta) = E(yy') = (I-B)-1\ (I-B)-1'\ (\Gamma\Phi\Gamma' + \Psi)$
- La matrice de covariance des x et des y : $\Sigma xy(\theta) = E(xy') = (I-B)-1'\ \Phi\Gamma'$
- La matrice de covariance des x : $\Sigma xx(\theta) = E(xx') = \Phi$

Soit au final :

$$\Sigma(\theta) = \begin{bmatrix} (I-B)-1\ (I-B)-1'\ (\Gamma\Phi\Gamma' + \Psi) & (I-B)-1\ \Phi\Gamma \\ (I-B)-1'\ \Phi\Gamma' & \Phi \end{bmatrix}$$

Le modèle est identifiable si le nombre de paramètre inconnu (θ) reste inférieur au nombre de valeur connues (donc d'équation - t-rule : $t \leq n*(n+1)/2$) et surtout qu'il soit organisé pour que les inconnues soient éliminables (règle de rang : dans une matrice composée de

[(I-B) | -Γ], pour vérifier la solvabilité de chaque équation, effacer toutes les colonnes qui n'ont pas de 0 dans la ième ligne, il faut que la matrice résultante soit de rang = p-1 - le rang est le nombre de colonne et de ligne indépendante linéairement). Dans la pratique la résolution algébrique bien que possible est replacée par une méthode d'optimisation. il est donc nécessaire de contraindre le modèle : **1**) mettre la diagonale de B = 0 (pas d'influence des facteurs sur eux-même, appelé convention de normalisation), **2**) les coefficients d'erreur sont mis à 1 (ceci à pour avantage de leur donner un ordre de grandeur).

L'identification débute par un choix de valeur pour ces paramètre qu'il s'agira d'optimiser. La fonction d'optimisation doit avoir les propriété suivantes : 1) être un scalaire, 2) être positive, 3) ne valoir 0 que quand S = Σ(θ), 4) être continues. Expl : maximum de vraisemblance, moindre carré non pondérés, moindre carré généralisés, moindre carrés pondérés.

Les fonctions de coût

<u>Le maximum de vraisemblance :</u>

$$FML = \log |\Sigma(\theta)| + tr(S\Sigma\text{-}1(\theta)) - \log |S| - (p+q)$$

Cette fonction bénéficie de plusieurs avantages :

- Elle est non biaisée asymptotiquement même pour de petit échantillon
- C'est un estimateur consistent (plim θest = θ)
- Elle a la variance asymptotique la plus faible
- La distribution de cet estimateur est gaussienne avec l'augmentation des paramètres. Ainsi, le ratio du paramètre estimé sur sa déviation standard suit une distribution en z. La variance de θest correspond aux éléments de la diagonale de l'inverse de la matrice d'information (covariance asymptotique de l'estimateur ML de θ, de taille t × t) :

  $$(2/(N\text{-}1)) \{E[\partial 2FML / \partial\theta\, \partial\theta\, '] \}^{-1}$$

- Invariance d'échelle (à quelques exceptions près). Ainsi si D est une matrice diagonale consistant en l'inverse des std des déviations standard de la matrice des variances observées S, alors DSD est une matrice de corrélation, et DΣ(θ)D est son estimée (p.109).
- Pour les modèles suridentifiés, la distribution asymptotique de (N-1)FML est une distribution du χ2 avec n*(n-1)/2)-t ddl.

- Elle peut s'appliquer même si la distribution des x est multinormale et donc même en cas d'utilisation de terme d'interaction (P.127 et 129).

<u>Les moindres carrés pondérés</u>

$$FWLS = [S - \sigma(\theta)]' \; W^{-1} \; [S - \sigma(\theta)]$$

Lorsque l'hypothèse de multinormalité n'est plus respectée, le maximum de vraisemblance ou les moindres carrés généralisés ne permettent plus de donner une estimation correcte de la significativité. On peut alors avoir recours :

- A une correction des données pour retourner à la multinormalité
- Utiliser des facteurs de correction applicables au ML, ULS, GLS
- Utiliser une approche par bootstrapping
- Mais la plus courante est d'utiliser un estimateur adapté à la non normalité.

Attention, seule l'estimation de la significativité pose un problème, l'estimation des paramètres reste valide quelque soit la méthode.

- S est un vecteur de (p+q)(p+q+1)/2 éléments, et $\sigma(\theta)$ leur correspondant dans $\Sigma(\theta)$.
- W^{-1} est une matrice $[(p+q)(p+q+1)/2]^2$ de poids définit positive (càd valeur propres toutes positives, c'est le cas des matrices de variance-covariance et de corrélation). Si il se réduit à une matrice identité, la formule équivaut aux moindres carrés ordinaires.

  W doit être un estimateur consistant de la matrice asymptotique de covariance de S avec S.

  $$N\text{-}1 \; (\partial \, \sigma(\theta) / \partial\theta) \;\; (\partial \, \sigma(\theta) / \partial\theta)']^{-1}$$

- Pour les modèles suridentifiés, la distribution asymptotique de (N-1)FWLS est une distribution du $\chi 2$ avec n*(n-1)/2)-t ddl

Optimisation

Le principe algorithmique est de résoudre l'équation différentielle suivante :

$$\partial FML(\theta_i) / \partial\theta_i = 0 \qquad \text{pour } i = 1 \ldots t$$

1) Déterminer des valeurs de départ pour les β, γ et ψ. Pour β_{ij}, γ_{ij} et ψ_{ij} avec $i \neq j$: a*std(y_i)/std(x_j) pour β et γ, et a*sqrt($\psi_{ii}\psi_{jj}$) pour ψ_{ij}. Le coefficient a peut être fixé à ±0.9 si le lien est supposé fort, ±0.4 si modéré, ±0.2 si faible. Enfin pour les ψ_{ii} : a*var(y_i), avec a fixée à 0.2 si le fit est supposé bon, 0.4 si moyen, et 0.9 si faible.
2) La fonction d'ajustement doit avoir une valeur progressivement décroissante. Pour cela on calcule son gradient : $\partial FML(\theta_i) / \partial\theta_i$ (avec i = 1 à t). En cas de valeur de

gradient négative le paramètre doit être augmenté, et diminué si la valeur du gradient est positive. Le paramètre est mis à jour de la façon suivante : $\theta = \theta - Cg$ (g est le vecteur de valeur du gradient pour chaque paramètre, et C est une matrice $t \times t$ définit positive). C'est la matrice identité dans la méthode de "steepest descent", ou l'inverse de la matrice de la dérivée partielle (matrice hessienne inverse) : $[\ \partial 2FML / \partial\theta\, \partial\theta\ '\ ]^{-1}$ dans la méthode de Newton-Raphson.

3) Enfin il reste à déterminer des critères d'arrêt : a) nombre d'itération maximum, b) une différence d'estimation entre θ_l et $\theta_{l-1} < 0.001$ par exemple. Pour être certain qu'il ne s'agit pas d'un minimum local, il est possible de recommencer la procédure avec des valeurs de départ très différentes.

D'autres méthodes d'optimisation sont utilisables. Celles décrites plus haut ne peuvent pas différencier un minimum local du minimum réel pour l'ensemble des solutions. Ainsi des techniques comme la méthode simplex polytop additionnée d'un recuit simulé seraient certainement plus adaptées.

Appendice B

Analyse des synchronies en EEG – MEG et extraction des réseaux

Source : *"Comparison of Hilbert transform and Wavelet methods for the analysis of neuronal synchrony"* M. Le Van Quyen, J. R. Foucher, J. P. Lachaux, E. Rodriguez, A Lutz, J. Martinerie, and F. J. Varela. *J.Neurosci.Methods* 2001 (In Press)

Traitement des données EEG-MEG (fig. 2a)

Les signaux électro-physiologiques ont été analysés en puissance (synchronies locales) et en phase-locking (synchronies longues distances)(60;61).

On estimait que les signaux $s_1(t)$ et $s_2(t)$ étaient synchrones, si leurs phases instantanées $\phi_1(t)$ et $\phi_2(t)$ conservaient un angle constant sur un intervalle de temps donné :

$$|\phi_1(t) - \phi_2(t)| \approx constant \quad pour\ t = T-\delta\ à\ T+\delta.$$

Les signaux ont été préalablement filtrés par un passe bande à ±2 Hz autour de la fréquence d'intérêt. En EEG celles-ci s'étendaient de 14 à 70 Hz par intervalle de 2 Hz. En MEG, seule la fréquence d'intérêt dépistée en l'EEG a été analysée. Le signal de chaque senseur était transformé en signal analytique par convolution avec une ondelette de Gabor complexe (encore appelée ondelette de Morlet) de formule :

$$\Psi_{\tau,f}(u) = \sqrt{f}.\exp(i2\pi f(u-\tau)).\exp(-\frac{(u-\tau)^2}{2\sigma^2})$$

Où $\Psi_{\tau,f}(u)$ était le produit d'une fonction sinusoï de de fréquence *f* avec une fonction Gaussienne centrée sur τ, de déviation standard σ proportionnelle à l'inverse de *f*. Le paramètre σ, permettait de régler le nombre de cycles de l'ondelette ($nco = 6f\sigma$) que nous avions fixer à 7 pour nos analyses afin de préserver une bonne résolution temporelle (FWHM = 140 ms à 20 Hz), la résolution fréquentielle nous étant donnée par le préfiltrage.

L'étape de convolution permettait d'obtenir le signal analytique $Sa_1(t)$. Elle consistait en une multiplication terme à terme des vecteurs $s_1(t+u)$ et $\Psi_{\tau,f}(u)$ effectuée pour tous les temps t sur un intervalle de temps correspondant à la taille l de l'ondelette, ce qui donnait sur un espace continu :

$$Sa_1(t) = \int_0^l s_1(t+u) \,.\, \Psi_{t,f}(u)\, du$$

On extrayait alors la phase instantanée $\phi_1(t)$ et la puissance instantanée $\omega_1(t)$:

$$\phi_1(t) = angle\,(Sa_1(t)) \qquad et \qquad \omega_1(t) = Sa_1(t) \,.\, conj\,(Sa_1(t))$$

Il était alors possible de calculer la différence de phase instantanée $\Psi_{1,2}(t)$ ou encore décalage de phase instantané :

$$\Psi_{1,2}(t) = |\,\phi_1(t) - \phi_2(t)\,|$$

L'indice de synchronisation (PLV pour Phase Locking Value) pouvait alors se calculer comme la somme vectorielle des angles $\Psi_{1,2}(t)$ sur un intervalle de temps [0,u] de n éléments que nous avions défini en nombre de cycle (donc adaptable pour chaque fréquence) et fixé à 7 pour nos analyses (résolution temporelle de 350 ms à 20 Hz). Ceci était formulé mathématiquement sur un espace discret comme suit :

$$PLV_{1,2}(t) = \frac{1}{n} \sum_{t \in [t,t+u]} e^{i\Psi_{1,2}(t)}$$

La PLV était comprise entre 0 (absence complète de synchronisation) et 1 (synchronisation parfaite des deux signaux), et le décalage de phase α entre les deux signaux était calculé comme la moyenne des $\Psi_{1,2}$ sur un intervalle de temps [0,u].

Chaque senseur devait avoir un niveau de puissance de base différent. De plus, en raison de la diffusion du signal électrique et magnétique, deux senseurs proches avaient une probabilité plus élevée d'être synchronisés que deux senseurs éloignés. On calculait donc les valeurs de référence de la puissance ω et de la PLV sur la période de repos pour chaque senseur (ω) ou chacune des paires de senseur (PLV) soit entre -2200 et le début de présentation du mot. Pour éviter des erreurs statistiques liées à l'autocorrélation des signaux, la distribution de référence réunissait toutes les périodes de repos du même sujet pour ne prélever que des valeurs distantes de plus d'une seconde (chute de la fonction d'autocorrélation au-delà des 500 ms pour tous les signaux). On a vérifié que la distribution de référence est bien une courbe gaussienne, ce qui permettait d'exprimer les valeurs suivant la présentation du mot en déviation standard δ par rapport à cette distribution.

Analyse statistiques de l'EEG-MEG – Extraction des réseaux (fig. 2b)

La démarche consistait alors pour chaque sujet et pour chaque essai à sélectionner les fréquences pour lesquelles les variations étaient les plus importantes ($\delta > \pm 3.6$ soit $p < 0.0006$). Les réseaux de capteurs et leur temps de survenu étaient extraits par une décomposition en valeur propre en conservant le premier vecteur propre pour chacune des

fréquences sélectionnées. Ainsi on ne retenait que les réseaux les plus représentés au cours des 5000 ms qui suivaient la présentation du mot. Typiquement la variance capturée était de l'ordre de 40 à 60% en EEG (378 paires de capteurs) et de 25 à 35 % en MEG (7140 paires de capteurs).

Nous avons fait l'hypothèse que ces réseaux devaient se retrouver au travers des différents essais. Comme nous supposions qu'il puisse y avoir plusieurs réseaux pertinents, nous avons tenté de les regrouper en classe homogène par une technique d'agrégation hiérarchique (ou encore clustering hiérarchique). La distance entre chaque réseau était mesurée par le critère de Ward généralisé qui est une ultramétrique (conservation des rapports de distance après agrégation) :

$$d(a,b) = m_a \cdot m_b / (m_a + m_b) \cdot d^2 (g_a, g_b)$$

Où d (a,b) était la distance calculée selon le critère de Ward généralisé entre les classes a et b, m_a et m_b leur poids respectif (ici le nombre d'élément de chaque classe) et $d^2 (g_a,g_b)$ était le carré de la distance entre les centres de gravité des classes a et b (la norme du vecteur $g_a - g_b$ soit $||g_a - g_b||^2$). L'algorithme était itératif, avec agrégation de la paire séparée de la plus petite distance à chaque itération. On minimisait ainsi l'inertie intra-classe, tout en maximisant l'inertie interclasse à chaque itération (25). La différence d'inertie entre deux agrégations servait d'indice de niveau.

Le résultat était présenté sous la forme d'un dendrograme, dont les branches séparées par les distances les plus grandes servaient de repères visuels pour le découpage en classe.

Bibliographie

(1) Norman D. The Psychology of Everyday Things. HarperCollins, 1988.

(2) Vignaux G. Les sciences cognitives : une introduction. Paris: La découverte, 1992.

(3) Varela FJ. Cognition et sciences cognitives. In: Lecourt D, editor. Dictionnaire d'histoire et philosophie des sciences. Paris: PUF, 1999: 185-191.

(4) Bickle J, Mandik P. The Philosophy of Neuroscience. http://plato.stanford.edu/entries/neuroscience/ . 1999.

(5) Jeannerod M. De la physiologie de l'esprit. Paris: Editions Odile Jacob, 1996.

(6) Andler D. Cognitivisme. In: Lecourt D, editor. Dictionnaire d'histoire et philosophie des sciences. Paris: PUF, 1999: 191-193.

(7) Dion E. Invitation à la théorie de l'information. Paris: Seuil, 1997.

(8) Michie D. Turing, Alan Mathison. In: Wilson RA, Keil FC, editors. The MIT encyclopedia of the cognitive sciences. Cambridge: Bradford - MIT press, 1999: 847-849.

(9) Varela FJ. Connaître: Les Sciences Cognitives, tendences et perspectivess. Paris: Editions du Seuil, 1988.

(10) Foucher JR, Meyer ME, Gounot D, Scheiber C, Namer IJ, Grucker D. Cartographie cérébrale fonctionnelle préopératoire par IRMf. Médecine Nucléaire 1999; 23(5):259-269.

(11) Marota JJ, Mandeville JB, Weisskoff RM, Moskowitz MA, Rosen BR, Kosofsky BE. Cocaine activation discriminates dopaminergic projections by temporal response: an fMRI study in Rat. Neuroimage 2000; 11(1):13-23.

(12) Kosslyn SM, Thompson WL, Kim IJ, Alpert NM. Topographical representations of mental images in primary visual cortex. Nature 1995; 378(6556):496-498.

(13) Henson RN, Shallice T, Dolan RJ. Right prefrontal cortex and episodic memory retrieval: a functional MRI test of the monitoring hypothesis. Brain 1999; 122(Pt 7):1367-1381.

(14) Price C, Wise R, Ramsay S, Friston K, Howard D, Patterson K et al. Regional response differences within the human auditory cortex when listening to words. Neurosci Lett 1992; 146(2):179-182.

(15) Mechelli A, Friston KJ, Price CJ. The effects of presentation rate during word and pseudoword reading: a comparison of PET and fMRI. J Cogn Neurosci 2000; 12 Suppl 2:145-156.

(16) Carpenter PA, Just MA, Reichle ED. Working memory and executive function: evidence from neuroimaging. Curr Opin Neurobiol 2000; 10(2):195-199.

(17) Speckmann EJ, Elger CE. Introduction to the neurophysiological basis of the EEG and DC potentials. In: Niedermeyer E, Lopes da Silva F, editors. Electroencephalography, basic principles, clinical applications and related fields. Williams & Wilkins, 1999: 15-27.

(18) Freeman WJ. How brains make up their minds. Weidenfeld & Nicolson, 1999.

(19) Lopes da Silva F, van Rotterdam A. Biophysical aspects of EEG and magnetoencephalogram generation. In: Niedermeyer E, Lopes da Silva F, editors. Electroencephalography, basic principles, clinical applications and related fields. Williams & Wilkins, 1999: 93-109.

(20) Tononi G, Sporns O, Edelman GM. A measure for brain complexity: relating functional segregation and integration in the nervous system. Proc Natl Acad Sci U S A 1994; 91(11):5033-5037.

(21) Buchel C, Friston KJ. Characterising functional integration. In: Frackowiak R, Friston KJ, Frith CD, Dolan RJ, Mazziotta JC, editors. Human brain function. Academic Press, 1997: 127-140.

(22) McIntosh AR, Gonzalez-Lima F. Structural modeling of functional neural pathways mapped with 2- deoxyglucose: effects of acoustic startle habituation on the auditory system. Brain Res 1991; 547(2):295-302.

(23) Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ. Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 1997; 6(3):218-229.

(24) Fletcher P, McKenna PJ, Friston KJ, Frith CD, Dolan RJ. Abnormal cingulate modulation of fronto-temporal connectivity in schizophrenia. Neuroimage 1999; 9(3):337-342.

(25) Lebart L, Morineau A, Piron M. Statistique exploratoire multidimensionnelle. Paris: Dunod, 1997.

(26) Friston KJ. Characterising distributed functional systems. In: Frackowiak R, Friston KJ, Frith CD, Dolan RJ, Mazziotta JC, editors. Human brain function. Academic Press, 1997: 107-126.

(27) Caussinus H. Projections révélatrices. In: Droesbeke JJ, Fichet B, Tassi P, editors. Modèles pour l'analyse des données multidimentionnelles. Economica, 1992.

(28) McKeown MJ, Makeig S, Brown GG, Jung TP, Kindermann SS, Bell AJ et al. Analysis of fMRI data by blind separation into independent spatial components. Hum Brain Mapp 1998; 6(3):160-188.

(29) Makeig S, Jung TP, Bell AJ, Ghahremani D, Sejnowski TJ. Blind separation of auditory event-related brain responses into independent components. Proc Natl Acad Sci U S A 1997; 94(20):10979-10984.

(30) Friston KJ, Buchel C. Attentional modulation of effective connectivity from V2 to V5/MT in humans. Proc Natl Acad Sci U S A 2000; 97(13):7591-7596.

(31) Bartolomei F, Wendling F, Vignal JP, Kochen S, Bellanger JJ, Badier JM et al. Seizures of temporal lobe epilepsy: identification of subtypes by coherence analysis using stereo-electro-encephalography. Clin Neurophysiol 1999; 110(10):1741-1754.

(32) Auger F. Introduction à la théorie du signal et de l'information. Paris: Editions Technip, 1999.

(33) Friston KJ. Brain function, nonlinear coupling, and neuronal transients. Neuroscientist 2001; 7(5):406-418.

(34) Bollen KA. Structural equations with latent variables. New York: Wiley, 1989.

(35) Coull JT, Buchel C, Friston KJ, Frith CD. Noradrenergically mediated plasticity in a human attentional neuronal network. Neuroimage 1999; 10(6):705-715.

(36) Cabeza R, McIntosh AR, Tulving E, Nyberg L, Grady CL. Age-related differences in effective neural connectivity during encoding and recall. Neuroreport 1997; 8(16):3479-3483.

(37) Bullmore E, Horwitz B, Honey G, Brammer M, Williams S, Sharma T. How good is good enough in path analysis of fMRI data? Neuroimage 2000; 11(4):289-301.

(38) Buchel C, Coull JT, Friston KJ. The predictive value of changes in effective connectivity for human learning. Science 1999; 283(5407):1538-1541.

(39) Pearl J. Causality, models, reasoning and inference. Cambridge: Cambridge University Press, 2000.

(40) Laplane D, Levasseur M, Pillon B, Dubois B, Baulac M, Mazoyer B et al. Obsessive-compulsive and other behavioural changes with bilateral basal ganglia lesions. A neuropsychological, magnetic resonance imaging and positron tomography study. Brain 1989; 112(Pt 3):699-725.

(41) Damasio AR. Descartes' error, emotion, reason and the human brain. Grosset & Putnam Books, 1994.

(42) Buckner RL, Goodman J, Burock M, Rotte M, Koutstaal W, Schacter D et al. Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. Neuron 1998; 20(2):285-296.

(43) Gardiner JM, Java RI. Forgetting in recognition memory with and without recollective experience. Mem Cognit 1991; 19(6):617-623.

(44) Henson RN, Rugg MD, Shallice T, Josephs O, Dolan RJ. Recollection and familiarity in recognition memory: an event-related functional magnetic resonance imaging study. J Neurosci 1999; 19(10):3962-3972.

(45) Goodale MA, Milner AD, Jakobson LS, Carey DP. A neurological dissociation between perceiving objects and grasping them. Nature 1991; 349(6305):154-156.

(46) Schmidt T. Visual perception without awareness: priming responses by color. In: Metzinger T, editor. Neural correlates of consciousness. Cambridge MA: MIT press, 2000: 157-169.

(47) Crick F, Koch C. Towards a neurobiological theory of consciousness. Semin Neurosc 1990; 2:263-275.

(48) Sperber D. Metarepresentation. In: Wilson RA, Keil FC, editors. The MIT encyclopedia of the cognitive sciences. Cambridge: Bradford - MIT press, 1999: 541-543.

(49) Proust J. Awareness of agency: three levels of analysis. In: Metzinger T, editor. Neural correlates of consciousness. Cambridge MA: MIT press, 2000: 307-324.

(50) Frith CD, Blakemore S, Wolpert DM. Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. Brain Res Brain Res Rev 2000; 31(2-3):357-363.

(51) Damasio AR. The feeling of what happens : body and emotion in the making of consciouness. New York: Harcourt Brace & Compagny, 1999.

(52) Le Moigne JL. Complexité. In: Lecourt D, editor. Dictionnaire d'histoire et philosophie des sciences. Paris: PUF, 1999: 205-215.

(53) Varela FJ. Neurophenomenology: A Methodological remedy to the hard problem. J Consc Studies 1996; 3:330-350.

(54) Edelman GM, Tononi G. A universe of consciousness. How matter becomes imagination. Basic Books, 2000.

(55) van Gelder T. Dynamic approaches to cognition. In: Wilson RA, Keil FC, editors. The MIT encyclopedia of the cognitive sciences. Cambridge: Bradford - MIT press, 1999: 244-246.

(56) Beckermann A. The perennial problem of the reductive explainabilité of phenomenal consciousness: C.D. Broad on the explanatory gap. In: Metzinger T, editor. Neural correlates of consciousness. Cambridge MA: MIT press, 2000: 41-55.

(57) Tononi G, Edelman GM. Consciousness and complexity. Science 1998; 282(5395):1846-1851.

(58) Tononi G, McIntosh AR, Russell DP, Edelman GM. Functional clustering: identifying strongly interactive brain regions in neuroimaging data. Neuroimage 1998; 7(2):133-149.

(59) Le Van Quyen M, Foucher JR, Lachaux JP, Rodriguez E, Lutz A, Martinerie J et al. Comparison of Hilbert transform and Wavelet methods for the analysis of neuronal synchrony. J Neurosci Methods. In press.

(60) Lachaux JP, Rodriguez E, Martinerie J, Varela FJ. Measuring phase synchrony in brain signals. Hum Brain Mapp 1999; 8(4):194-208.

(61) Rodriguez E, George N, Lachaux JP, Martinerie J, Renault B, Varela FJ. Perception's shadow: long-distance synchronization of human brain activity. Nature 1999; 397(6718):430-433.