L'annuaire du Collège de France
Cours et travaux

112 | 2013
Annuaire du Collège de France 2011-2012

Psychologie cognitive expérimentale

Stanislas Dehaene

Édition électronique
URL : https://journals.openedition.org/annuaire-cdf/1011
DOI : 10.4000/annuaire-cdf.1011
ISBN : 978-2-7226-0325-7
ISSN : 2109-9227

Éditeur
Collège de France

Édition imprimée
Date de publication : 1 avril 2013
Pagination : 339-362
ISBN : 978-2-7226-0198-7
ISSN : 0069-5580

Référence électronique
Stanislas Dehaene, « Psychologie cognitive expérimentale », *L'annuaire du Collège de France* [En ligne], 112 | 2013, mis en ligne le 28 août 2013, consulté le 22 août 2022. URL : http://journals.openedition.org/annuaire-cdf/1011 ; DOI : https://doi.org/10.4000/annuaire-cdf.1011

Tous droits réservés

Psychologie cognitive expérimentale

M. Stanislas DEHAENE, membre de l'Institut
(Académie des sciences), professeur

ENSEIGNEMENT

Cours : Le cerveau statisticien : la révolution bayésienne en sciences cognitives

L'organisation du cerveau humain résulte de multiples évolutions enchâssées, un « bricolage » (Jacob, 1977) *ad hoc* de systèmes les plus divers sélectionnés pour leur valeur de survie, sans garantie d'optimalité. C'est pourquoi bon nombre de neurobiologistes et de psychologues cognitifs estiment qu'il ne saurait y avoir de théorie mathématique universelle du cerveau, mais seulement des modèles restreints de compétences cognitives spécifiques.

À l'encontre de cette idée répandue, toutefois, un vaste courant récent des sciences cognitives s'appuie sur la théorie mathématique de l'inférence bayésienne pour modéliser une très grande diversité de phénomènes psychologiques : perception, inférence statistique, prise de décision, apprentissage, traitement du langage... La rapidité avec laquelle cette théorie envahit et unifie divers domaines de la cognition, la simplicité de ses fondements axiomatiques, et la profondeur de ses conclusions justifient de parler d'une véritable « révolution bayésienne » en sciences cognitives. Le cours de 2012 visait à en expliquer les principes.

Pour résumer, la théorie bayésienne fournit un modèle mathématique de la manière optimale de mener un raisonnement plausible en présence d'incertitudes. Dès la naissance, le bébé semble doté de compétences pour ce type de raisonnement probabiliste. L'inférence bayésienne rend également bien compte des processus de perception : étant donné des entrées ambigües, le cerveau en reconstruit l'interprétation la plus probable. La règle de Bayes indique comment combiner, de façon optimale, les *a priori* issus de notre évolution ou de notre mémoire avec les données reçues du monde extérieur. En cela, elle offre une nouvelle vision de l'apprentissage qui dépasse le dilemme classique entre théories empiristes et nativistes. Enfin, de nombreuses décisions humaines semblent résulter d'une approximation de la règle bayésienne d'accumulation d'évidence, combinée à une estimation de la valeur

attendue des conséquences de nos choix. Dans la mesure où les principes de l'inférence bayésienne sont ainsi partagés par de multiples domaines de la cognition, il se pourrait que l'architecture du cortex ait évolué pour approximer ce type de calcul probabiliste à grande vitesse, et de façon massivement parallèle. L'algorithme utilisé pourrait expliquer non seulement l'organisation du cortex en couches, mais aussi la manière dont notre cerveau anticipe sur le monde extérieur (codage prédictif) et dont il répond à la nouveauté (propagation des signaux d'erreur).

Introduction au raisonnement bayésien et à ses applications cognitives

Pasteur de l'Église presbytérienne et mathématicien britannique, le révérend Thomas Bayes (~ 1701-1761) étudie la logique et la théologie à l'université d'Édimbourg. Divers travaux, dont une introduction au calcul différentiel, le font élire à la *Royal Society* le 4 novembre 1742. Ce n'est qu'après la mort de Bayes, en 1761, que son ami Richard Price retrouve, dans ses papiers, un *Essai sur la manière de résoudre un problème dans la doctrine des risques*. Publié par la *Royal Society* en 1763, cet essai applique le principe d'inférence que nous connaissons aujourd'hui sous le nom de « règle de Bayes ». Sans nier l'inventivité de Bayes, on s'accorde aujourd'hui à penser que sa règle n'est qu'une application simple de la règle du produit en théorie des probabilités, déjà connue de Bernoulli et de De Moivre, et dont Laplace (1774) percevra, le premier, le très vaste champ d'applications.

De quoi s'agit-il ? Les principes du raisonnement bayésien étendent, à des valeurs continues de plausibilité, les principes de la logique classique avec des valeurs de vérité discrètes. Il est d'ailleurs possible de les déduire des axiomes que doivent vérifier ces plausibilités (Jaynes, 2003). Brièvement : on suppose que (1) les degrés de plausibilité sont représentés par des nombres réels ; (2) ces valeurs suivent des règles de bon sens, suivant ainsi la célèbre formule de Laplace (« la théorie des probabilités n'est au fond que le bon sens réduit au calcul ») ; (3) aucune des données disponibles n'est ignorée ; (4) des états de connaissance équivalents ont le même degré de plausibilité. Le théorème de Cox-Jaynes montre que ces règles suffisent à définir, à une fonction monotone près, des règles mathématiques universelles pour la plausibilité *p*. Celles-ci sont les règles habituelles de la probabilité et la règle fondamentale « de Bayes » :

$$p(A \,\&\, B) = p(A|B)\, p(B) = p(B|A)\, p(A)$$

Point important, dans la théorie bayésienne, ces probabilités *p* ne sont plus interprétées comme les fréquences relatives d'événements (point de vue « fréquentiste »), mais comme des mesures du degré de connaissance subjective. Cette acception est bien couverte par le terme de « plausibilité », mais étant donné l'équivalence que nous venons de présenter, dans ce qui suit, nous continuerons à utiliser également le terme usuel de probabilité.

La théorie des probabilités, telle qu'elle est enseignée, est surtout utilisée pour calculer la probabilité d'une observation D, étant donné certaines hypothèses H sur l'état du monde. Cependant, les données d'observation D et les hypothèses H jouant des rôles symétriques, rien n'empêche de poser la question inverse : étant donnée une observation D, quelle est la plausibilité de l'hypothèse H ? La règle fondamentale répond à cette interrogation :

$$p(H|D) = p(D|H)\, p(H) \,/\, p(D) \text{ ou } p(H|D) \propto p(D|H)\, p(H)$$

Dans cette équation, p(H) est appelée *probabilité a priori* de H (*prior* en anglais) (notons qu'il ne s'agit pas nécessairement d'une connaissance indépendante de l'expérience, au contraire, l'*a priori* résulte souvent d'expériences antérieures). p(H|D) est la *probabilité a posteriori* de H (ce terme n'étant pas nécessairement à prendre dans son sens temporel, mais plutôt au sein d'une déduction logique : après avoir pris connaissance des données D, nous révisons la plausibilité de H). Enfin, p(D|H), considéré comme une fonction de H, est la *vraisemblance* de H.

Le *facteur de Bayes* qui sépare deux hypothèses ou deux « modèles » M1 et M2, est une mesure de leur mérite relatif, le rapport de leurs vraisemblances : p(D|M1)/p(D|M2). Le logarithme décimal de cette valeur (*log likelihood ratio*), mesuré en décibans, est une mesure souvent plus intelligible. On appelle également « évidence » (en anglais *weight of evidence* ou, chez Turing, *log odds*) la valeur log(p/(1-p)), qui quantifie la vraisemblance d'une hypothèse par rapport à toutes les autres. Selon Jeffreys, 10 décibans, soit un rapport de 10 contre 1, constitue une évidence « substantielle », et 20 décibans ou plus (au moins 100 contre 1) une évidence « décisive ».

Le principe du *maximum de vraisemblance* consiste à choisir le modèle M_i pour lequel les données sont les plus probables ($p(D|M_i)$ est maximale). Le principe du *maximum a posteriori* (ou *MAP*) consiste à maximiser p(Mi |D) – ce qui prend également en compte les probabilités *a priori* des différents modèles M_i.

On attribue à Guillaume d'Ockham (1285-1347) un principe de raisonnement, en réalité énoncé depuis l'Antiquité, selon lequel « une pluralité ne doit pas être posée sans nécessité » : « Les entités ne doivent pas être multipliées au-delà du nécessaire » ; ou encore, toutes choses égales par ailleurs, les explications les plus simples doivent être préférées aux plus complexes. Le raisonnement bayésien, en tant que modèle optimal de la « logique de la science » (Jaynes, 2003), inclut automatiquement un rasoir d'Ockham. En effet, une vraisemblance plus petite est allouée aux hypothèses les plus faibles, celles qui font les prédictions les plus diffuses ou les moins spécifiques. De plus, si deux modèles atteignent la même vraisemblance, celui qui l'atteint avec le plus de paramètres libres est automatiquement pénalisé par les équations bayésiennes. Enfin, il est souvent naturel d'attribuer une probabilité *a priori* plus petite aux modèles plus complexes. Tous ces facteurs contribuent à la règle d'Ockham et font de la théorie bayésienne un modèle plausible du raisonnement scientifique (Jaynes, 2003).

Les mécanismes bayésiens de l'induction chez l'enfant

L'intérêt que suscite la théorie bayésienne en psychologie cognitive se fonde sur certaines observations qui suggèrent que le très jeune enfant, dès les premiers mois, serait capable d'un raisonnement plausible, probablement inconscient. Par exemple, un bébé de 12 mois, voyant le contenu d'une urne, peut anticiper le résultat d'un tirage aléatoire : si l'urne contient trois objets bleus et un jaune, l'enfant manifeste sa surprise en regardant plus longuement lorsque c'est l'objet jaune qui sort (Teglas, Girotto, Gonzalez & Bonatti, 2007 ; Teglas *et al.*, 2011). Et inversement, comme l'a décrit Laura Schulz dans le séminaire, les bébés de 8 mois peuvent inférer le contenu d'une urne à partir de quelques échantillons (Gweon, Tenenbaum & Schulz, 2010 ; Xu & Garcia, 2008).

Joshua Tenenbaum et ses collaborateurs suggèrent que l'apprentissage du langage fournit l'un des exemples les plus convaincants d'apprentissage bayésien par induction

(Kemp & Tenenbaum, 2009 ; Perfors, Tenenbaum, Griffiths & Xu, 2011 ; Tenenbaum, Kemp, Griffiths & Goodman, 2011). En effet, la théorie apporte des éléments de réponse au « scandale de l'induction » débattu depuis l'Antiquité (Platon, Aristote, Kant, Peirce, Quine, Russell), et que Tenenbaum résume ainsi : « comment nos esprits parviennent-ils à inférer autant à partir de si peu? » Les enfants et les adultes réalisent quotidiennement des inférences très sophistiquées, alors qu'il semble évident que les données disponibles ne le justifient pas. Par exemple, tout scientifique sait que « corrélation n'est pas causation » – et pourtant, les humains infèrent régulièrement des relations causales, sur la base de données en nombre si faible qu'elles ne suffiraient même pas à calculer un coefficient de corrélation.

Tenenbaum approfondit l'exemple, discuté par Quine (1960), de l'apprentissage des mots. Comment un enfant ou un adulte infèrent-ils l'extension d'un mot sur la base de quelques exemples, qu'il s'agisse d'une instance (« Fido ») ou d'une catégorie plus ou moins étendue (« basset », « chien », « animal ») ? Cette observation peut être modélisée dans le cadre bayésien en supposant que le monde perceptif est organisé sous forme d'un arbre d'hypothèses. Une vraisemblance plus faible est automatiquement accordée aux catégories les plus vastes. Le modèle converge en deux ou trois exemples : il attribue la plausibilité la plus forte au sens correct du mot, et généralise correctement à des instances nouvelles. De fait, le modèle bayésien fournit une théorie universelle de la manière dont un apprentissage se généralise (Shepard, 1987), modèle qui capture un très grand nombre de données empiriques sur la psychophysique de la généralisation perceptive.

La théorie bayésienne offre une vision de l'apprentissage qui attribue une place bien définie aux connaissances innées : l'apprentissage est contraint par l'espace d'hypothèses envisagées, la fonction de probabilité *a priori* et la fonction de vraisemblance. De plus, un volet récent de la théorie introduit le concept fondamental de modèle bayésien hiérarchique : l'espace des hypothèses au niveau n peut être, à son tour, engendré par un *modèle génératif* de niveau supérieur $n + 1$, et ses probabilités *a priori* déterminées par un petit nombre d'*hyper-paramètres*. Une fois le modèle hiérarchique spécifié en détail, la théorie bayésienne fournit alors une manière simple de mettre à jour les plausibilités à chaque niveau.

De tels modèles bayésiens hiérarchiques possèdent une propriété intéressante que Tenenbaum appelle « bénédiction de l'abstraction » (*blessing of abstraction*) : l'apprentissage est souvent très rapide au niveau le plus élevé, celui qui concerne les principes mêmes d'organisation d'un domaine, car chaque observation nouvelle contribue à sélectionner le modèle le plus pertinent, et celui-ci se généralise à toutes les observations nouvelles. Cette propriété s'apparente à ce que l'on appelle « le transfert d'apprentissage » ou « apprendre à apprendre ». Pour reprendre un exemple dû à Nelson Goodman (1955), supposons que l'on tire d'une seule urne une unique balle bleue. Que peut-on en conclure sur les tirages suivants ? Rien… sauf si, par le passé, on a découvert que chaque sac ne contient que des billes d'une seule couleur, auquel cas on peut avoir la quasi-certitude que toutes les balles suivantes seront bleues. Tenenbaum (2011) montre comment modéliser cette situation par un modèle bayésien hiérarchique (modèle *Dirichlet-multinomial*).

Le modèle d'induction bayésienne s'applique directement aux expériences d'apprentissage des mots chez l'enfant de 17 à 19 mois réalisées par Linda B. Smith (2002). Dans cette étude, un enfant voit un adulte appeler des formes par des noms nouveaux (wif, dax, etc.). Systématiquement, l'enfant voit que la même syllabe s'applique à deux objets de même forme, mais dont la texture, la taille et la couleur

varient. Par la suite, exposé à une seule instance d'un mot nouveau (« lug ») apparié à un objet nouveau, il généralise correctement l'usage de ce mot à ces objets nouveaux de même forme, tandis qu'il refuse d'appliquer le mot à des objets de même texture, taille ou couleur. L'enfant a donc induit un principe général qui gouverne la plupart des noms communs : chacun d'eux se réfère à un objet de forme donnée. L'apprentissage bayésien est capable d'expliquer et de modéliser en détail cette capacité de généralisation : « Each bit of individual learning changes the learner, and thus progressively changes what the learner finds easy to learn. » (Smith *et al.*, 2002).

Sur la base d'un algorithme informatique plus complexe, Kemp et Tenenbaum (2008) parviennent à modéliser l'induction automatique de la structure profonde de nombreux domaines : l'arborescence de la famille des êtres vivants, la structure circulaire du globe terrestre, ou même le principe de causalité lui-même. Nous ne sommes donc pas loin d'une modélisation des grands principes de la découverte scientifique, même si les données de psychologie cognitive manquent encore pour le valider.

En résumé, les modèles bayésiens incluent à la fois des hypothèses *a priori* très riches (sur l'espace des représentations mentales considérées et leur probabilité *a priori*), et un mécanisme simple et puissant d'apprentissage par inférence statistique. Jusqu'à récemment, la psychologie cognitive était très divisée sur le rôle de l'apprentissage dans l'organisation des compétences de l'enfant. Le connexionnisme, d'une part, proposait des mécanismes sophistiqués d'apprentissage, mais portant sur des représentations non-structurées des connaissances (matrices d'associations). Les théories « représentationnelles », d'autre part, faisaient l'hypothèse de représentations symboliques structurées et sophistiquées des connaissances, mais étaient dépourvues d'un réel dispositif d'apprentissage. L'approche bayésienne présente l'avantage de réconcilier ces deux camps et de dépasser ainsi de nombreuses oppositions caricaturales et stériles (inné/acquis ; nativisme chomskyen contre constructivisme piagétien ou associationnisme skinnérien).

Sommes-nous de mauvais statisticiens ?

L'idée d'un « cerveau statisticien », présent dès la naissance, conforte la vision de Laplace, selon lequel « la théorie des probabilités […] fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte ». Selon cette hypothèse, nous serions tous pourvus d'une intuition de la plausibilité, fondée sur de complexes calculs bayésiens non-conscients. Toutefois, cette perspective s'oppose directement aux travaux d'Amos Tversky et Daniel Kahneman pour qui les humains sont de bien piètres statisticiens. « On s'accorde généralement à penser que les choix rationnels doivent satisfaire des critères élémentaires de cohérence », écrivent-ils dans leur célèbre article de *Science* (1981) ; « Dans cet article nous décrivons des problèmes de décision dans lesquels les gens violent systématiquement ces critères. »

Étendant notamment le paradoxe d'Allais (1953), Tverky et Kahneman montrent que les jugements humains s'écartent parfois massivement de la théorie du choix rationnel. Ils rendent compte de leurs observations empiriques en proposant une théorie des perspectives (*prospect theory*), qui s'écarte de l'optimalité bayésienne en plusieurs points : (1) les gains et les pertes ne sont pas absolus, mais rapportés à un cadre de référence (*frame*) ; (2) les décisions maximisent l'espérance de la

valeur subjective ; (3) la fonction de valeur subjective est concave pour les gains, convexe pour les pertes ; (4) la probabilité est pondérée selon une fonction en S inversé qui surestime les petites probabilités et sous-estime les grandes probabilités.

La réconciliation des travaux de Kahneman et Tversky avec les nombreux travaux qui soutiennent la perspective bayésienne fait partie des axes importants de recherche des prochaines années. Dans le cours, sans résoudre cette contradiction apparente, nous avons abordé quelques pistes de recherche. Tout d'abord, notons que bon nombre des difficultés qu'observent Kahneman et Tversky ne proviennent pas nécessairement de l'évaluation des probabilités, mais plutôt de la prise de décision. La fonction d'utilité, mais aussi la prise en compte du risque, pourraient expliquer certains des biais observés. En second lieu, le cerveau, qui n'est pas un ordinateur digital, pourrait n'implémenter qu'un algorithme bayésien imparfait. Il se pourrait ainsi que la représentation des toutes petites probabilités, ou la multiplication de deux probabilités ou de deux distributions, posent des difficultés particulières au cerveau.

En troisième lieu, Kahneman et Tversky utilisent des tâches conscientes, verbales, de haut niveau. Or notre espace de travail conscient ne prend en compte que peu de données : le traitement conscient fonctionne selon un mode sériel et fondamentalement limité (voir le cours de 2009-2010). Au contraire, il se pourrait que les traitements non-conscients qui sous-tendent la perception, la décision motrice, l'apprentissage du langage… fassent appel à des algorithmes massivement parallèles et plus proches de l'optimalité bayésienne. Un axe de recherche récent consiste ainsi à comparer des problèmes formellement identiques sur le plan mathématique, mais qui sont résolus soit par une réflexion verbale consciente, soit inconsciemment par le système sensori-moteur. Effectivement, des biais très différents sont observés (Braun, Nagengast & Wolpert, 2011).

Enfin, le contexte et les mots utilisés pour présenter un problème statistique peuvent avoir un effet important. Notre cerveau pourrait traiter plus facilement certains types de problèmes, soit parce qu'ils font spontanément appel à des domaines de connaissances pour lesquels l'évolution nous a dotés de solutions spécifiques (Cosmides, 1989), soit parce que le format de présentation des données du problème facilite leur recodage dans le format interne que nous utilisons pour l'inférence bayésienne (Gigerenzer & Hoffrage, 1995). Gigerenzer conteste ainsi les travaux de Kahneman et Tversky en montrant, par exemple, que le biais de négligence de la probabilité *a priori* disparaît largement lorsque le problème est formulé à l'aide de « fréquences naturelles », c'est-à-dire avec des nombres entiers.

Krynski et Tenenbaum (2007) proposent une hypothèse distincte, également très intéressante. Selon eux, le cadre théorique approprié aux « jugements sous incertitude » de Kahneman et Tversky ne serait pas celui de l'inférence statistique traditionnelle, mais celui de l'inférence causale. Notre cerveau n'aurait pas évolué pour compiler des statistiques sur de complexes probabilités conditionnelles P(A|B,C,D…) (problème qui devient d'ailleurs très vite impossible à évaluer), mais plutôt pour tirer des inférences à partir de modèles causaux des observations effectuées. L'interprétation d'un problème statistique se ferait donc en trois étapes : (1) construction d'un modèle mental des relations causales, (2) attribution d'une valeur aux différents paramètres du modèle, (3) inférence bayésienne. La construction d'un modèle causal inapproprié expliquerait les erreurs des sujets. Krynski et Tenenbaum testent leur hypothèse en montrant qu'il suffit de changer une phrase dans l'énoncé du problème pour changer le modèle causal qu'utilisent les participants. Immédiatement, la capacité d'inférence s'améliore : même en l'absence

de valeurs numériques précises, les sujets testés montrent des intuitions correctes des probabilités mises en jeu.

Il semble donc bien, *contra* Kahneman et Tversky, que le cerveau humain comprenne des mécanismes intuitifs et approximatifs d'inférence bayésienne. Ceux-ci ne sont toutefois déployés de façon appropriée que si l'exposé du problème conduit l'auditeur à se représenter le modèle causal adéquat.

Les illusions visuelles : des inférences optimales ?

La perception visuelle fournit un excellent exemple d'application non-consciente de statistiques bayésiennes. Dès 1867, dans son célèbre *Traité d'optique physiologique* (*Handbuch der physiologischen Optik*), Hermann von Helmholtz introduit le concept d'« inférence inconsciente » : les entrées sensorielles (« *Perception* ») se combineraient à nos attentes (« *Vorstellung* ») pour former le contenu de l'expérience consciente (« *Anschauung* »).

De fait, la notion d'inférence s'applique bien au problème posé par la perception. En effet, toutes nos entrées sensorielles sont ambigües – chacune d'elle pourrait résulter d'une infinité de causes possible. La tâche de notre système perceptif est donc de sélectionner, parmi une infinité de solutions, celle qui est la plus plausible. La théorie bayésienne explique ce processus de choix optimal sur la base de (1) la connaissance *a priori* des probabilités des objets dans le monde extérieur p(O) (qui peut résulter d'une accumulation de connaissances au cours de l'apprentissage), (2) la connaissance de la fonction de vraisemblance p(S|O) (qui peut résulter d'un modèle interne du comportement des objets), et (3) l'application de la règle de Bayes : $p(O|S) \propto p(S|O)\, p(O)$.

Selon cette théorie, de nombreuses illusions visuelles s'expliqueraient par un processus d'inférence optimale qui prend en compte les statistiques de l'environnement. Si la perception est optimale, pourquoi engendre-t-elle une illusion ? Parce que le stimulus qui crée l'illusion est très appauvri en comparaison de notre expérience usuelle du monde extérieur. Le système perceptif l'enrichit alors d'informations statistiques accumulées au fil de l'expérience passée. De l'intégration de l'*a priori* et des entrées sensorielles résulte une perception consciente qui s'écarte des données sensorielles réelles. Bien qu'elle conduise ainsi à une illusion, il n'est pas inapproprié de qualifier cette inférence perceptive de « rationnelle », dans la mesure où, dans un contexte plus naturel, elle serait utile ou même optimale pour l'organisme.

Simoncelli et ses collègues analysent dans cette perspective les biais dans la perception de l'orientation d'une barre (Girshick, Landy & Simoncelli, 2011). La perception varie avec l'orientation, à la fois dans sa précision (la variabilité interne est minimale pour la verticale et l'horizontale) et dans son biais (les réponses sont biaisées en direction des orientations horizontale et verticale). L'analyse théorique montre que ces données s'expliquent intégralement, dans le cadre bayésien, par l'*a priori* d'une fréquence plus élevée de ces directions cardinales dans le monde extérieur.

Une interprétation très semblable permet de rendre compte des illusions de mouvement (Weiss, Simoncelli & Adelson, 2002). À partir d'un cadre théorique apparenté, Dale Purves et ses collègues analysent de nombreuses autres illusions visuelles et auditives, en montrant qu'elles peuvent s'expliquer par une internalisation des statistiques du monde extérieur. Pour ne prendre qu'un exemple, à cause de la

perspective, une barre verticale sur la rétine reflète souvent une plus grande distance dans le monde réel qu'une barre horizontale – c'est pourquoi, selon Purves, elle est également perçue comme plus longue (Howe & Purves, 2002; Purves, Lotto, Williams, Nundy & Yang, 2001).

Combinaison de contraintes et sélection d'un percept unique

L'inférence perceptive peut être bien plus complexe que la simple reconstruction de la valeur optimale d'un paramètre (orientation, vitesse, distance). En effet, les entrées sensorielles résultent fréquemment d'une combinaison complexe de nombreuses variables et sources de bruit. Une scène visuelle résulte par exemple d'un modèle génératif multi-niveau qui prend en compte les sources de lumière, la probabilité que tel ou tel objet soit présent, sa forme, les matériaux et leur réflectance, etc. Les réseaux bayésiens permettent de modéliser cette propagation hiérarchique des contraintes et d'inverser le modèle afin de reconstruire, à partir d'une image sur la rétine, la distribution de probabilité de tous les paramètres (Kersten, Mamassian & Yuille, 2004).

Ces modèles bayésiens hiérarchiques rendent compte d'illusions telles que la perception de la troisième dimension à partir de l'éclairement. Le cerveau infère la présence d'une source de lumière, en s'aidant de l'*a priori* qu'elle vient probablement du haut, et cette information inférée, à son tour, est utilisée pour résoudre l'ambiguïté sur la forme concave ou convexe d'une demi-sphère (Morgenstern, Murray & Harris, 2011).

La théorie bayésienne clarifie également la façon dont le cerveau intègre plusieurs indices sensoriels. En présence d'informations visuelles et tactiles, plus ou moins bruitées, comment décidons-nous, par exemple, de la taille d'un objet ? Si les distributions sont gaussiennes, la théorie fait des prédictions précises : la perception doit être une moyenne pondérée des valeurs suggérées par chaque indice, tandis que sa fiabilité (*reliability*, c'est-à-dire l'inverse de la variance) doit être la somme des fiabilités de chaque indice pris isolément. Une étude psychophysique démontre que le comportement humain se conforme très précisément à ces prédictions (Ernst, 2007 ; Ernst & Banks, 2002). La théorie bayésienne s'applique également à l'intégration séquentielle d'indices multiples (Petzschner & Glasauer, 2011).

Quoique la théorie prescrive comment les distributions de probabilité sont combinées et mises à jour, elle ne spécifie pas quel paramètre accède finalement à la conscience. Nous ne percevons jamais l'ambigüité d'une distribution statistique, mais seulement une valeur unique (éventuellement assortie d'une incertitude). Percevons-nous systématiquement la représentation la plus plausible (maximum *a posteriori* ou MAP) ? Pas nécessairement. Une hypothèse intéressante pose que notre conscience ne nous donne à voir qu'un échantillon de la distribution de probabilités *a posteriori* (Moreno-Bote, Knill & Pouget, 2011). Les processus perceptifs non conscients manipuleraient l'ensemble de la distribution, tandis que l'accès à la conscience consisterait en un tirage aléatoire d'un échantillon de la distribution *a posteriori*. Moreno-Bote *et coll.* notent avec intérêt qu'un tel tirage se produit spontanément dans certaines implémentations mathématiques des réseaux bayésiens, qui se fondent sur un échantillonnage aléatoire des différentes possibilités en jeu (*sampling*).

Moreno-Bote et ses collaborateurs testent l'hypothèse de l'échantillonnage conscient en présentant à des observateurs un stimulus ambigu constitué de deux grilles superposées. Dans cette situation typique de rivalité perceptive, la perception

consciente alterne : la grille A est perçue tantôt devant la grille B, tantôt derrière. La dominance d'un percept sur l'autre peut être manipulée par des indices visuels : la longueur d'onde et la vitesse de déplacement des grilles (expérience 1) ou bien la longueur d'onde et la disparité binoculaire (expérience 2). Le modèle bayésien prédit que, si les indices sont conditionnellement indépendants, la perception par échantillonnage doit suivre une loi multiplicative. Effectivement, la mesure de la proportion du temps pendant lequel les observateurs rapportent percevoir l'une des deux possibilités suit précisément cette loi. « Dans la perception bistable », concluent les auteurs, « la fraction de dominance se comporte comme une probabilité » – un résultat parfaitement compatible avec l'hypothèse d'un échantillonnage conscient de distributions calculées non consciemment.

Les travaux de Vul *et coll.* sur l'attention temporelle et spatiale renforcent encore ce point de vue (Vul, Hanus & Kanwisher, 2009 ; Vul, Nieuwenstein & Kanwisher, 2008). Ils indiquent que la sélection attentionnelle opère de manière graduelle et statistique, mais que, là encore, les réponses rapportées se comportent comme des échantillons d'une distribution interne. Bien que le cerveau fonctionne d'une manière probabiliste, nous ne prenons conscience que d'échantillons discrets (Aly & Yonelinas, 2012 ; Sergent, Baillet & Dehaene, 2005 ; Sergent & Dehaene, 2004).

Nous pouvons cependant ré-échantillonner plusieurs fois la même distribution, et obtenir à chaque fois de l'évidence supplémentaire. Cet effet d'échantillonnage répété affecte également la récupération d'informations en mémoire. Ainsi, lorsque nous répondons à des questions qui demandent une approximation numérique (par exemple : quel pourcentage des aéroports mondiaux se trouvent aux États-Unis ?), le fait de demander à la même personne deux réponses successives améliore le résultat : la moyenne de la première et de la seconde réponse s'avère plus proche de la vérité que chacune de ces réponses prises séparément, particulièrement si les réponses sont faites à plusieurs semaines d'intervalle (Vul & Pashler, 2008). Ces résultats sont compatibles avec l'hypothèse que chaque réponse échantillonne la distribution de probabilité de la réponse. D'une manière analogue, l'échange verbal d'informations sur la confiance permet à deux personnes de répondre mieux qu'une seule (Bahrami *et al.*, 2010), même si cet effet présente des limites (Lorenz, Rauhut, Schweitzer & Helbing, 2011).

La prise de décision bayésienne

Percevoir le monde extérieur ne suffit pas à agir de façon optimale. Comment passer d'une perception à une décision ? Le problème est exposé de façon simple dans une revue récente (Maloney & Zhang, 2010). Chaque état du monde (w) se traduit, après inférence bayésienne, par une distribution d'états sensoriels inférés (x). Le problème de la décision consiste à choisir l'action $a = d(x)$ en fonction des états sensoriels x. Les actions ont des conséquences positives ou négatives qui dépendent des états réels du monde, selon une fonction de Gain (ou de Coût) G(a,w). Une stratégie rationnelle consiste à choisir l'action qui maximise l'espérance du gain G.

De ce modèle très général, on peut déduire, par exemple, la totalité de la théorie de la détection du signal (*signal detection theory*). Les états du monde sont binaires (w ou ~ w) et la fonction de gain, très simple, vaut + 1 si la réponse correspond à l'état du monde, -1 sinon. Cependant, la théorie permet de modéliser des situations bien plus complexes. Par exemple, une tâche intéressante consiste à demander au sujet de pointer très vite du doigt vers une cible (un disque vert) tout en imposant une pénalité

si le doigt touche un disque rouge (Trommershauser, Maloney & Landy, 2008). Les résultats indiquent que de telles décisions motrices, en dépit de leur rapidité et de leur automaticité, s'approchent de l'optimalité. Les sujets prennent convenablement en compte l'incertitude associée à leurs propres mouvements, la distribution des stimuli et les coûts qui en résultent. Ils parviennent également à arbitrer entre le temps consacré à la prise de décision et le temps consacré au mouvement, toujours en optimisant la fonction de coût imposée par l'expérimentateur (Battaglia & Schrater, 2007). Enfin, ils prennent en compte, de façon quasi-optimale, les informations sensorielles intermédiaires, plus ou moins bruitées, qu'ils reçoivent sur la trajectoire de leur mouvement en cours (Kording & Wolpert, 2004).

Dans de nombreuses tâches, la prise d'information sensorielle n'est pas instantanée, mais séquentielle : plusieurs stimuli peuvent s'échelonner dans le temps, et même un stimulus unique fournit, au niveau neuronal, une série temporelle d'échantillons bruités. La théorie bayésienne prédit alors une *accumulation d'évidence*, c'est-à-dire l'addition de l'évidence apportée par chaque échantillon successif (l'addition des logarithmes correspond au produit des distributions) – éventuellement pondérée par les gains attendus. Ce modèle prédit que la prise de décision repose sur une marche aléatoire interne, la décision étant prise lorsque l'évidence atteint un seuil préalablement déterminé. Effectivement, une telle évolution progressive des taux de décharges neuronales, semi-aléatoire et dépendante de l'évidence disponible, est observée dans les neurones du cortex préfrontal et intrapariétal du singe macaque (Gold & Shadlen, 2002).

Dans un travail récent, Yang et Shadlen (2007) démontrent la capacité du modèle d'accumulation d'évidence à rendre compte de la prise de décision, tant au niveau comportemental que neuronal. Ils présentent à des singes quatre formes successives, chacune donnant une information partielle sur la probabilité de recevoir un renforcement après une saccade vers un rond rouge ou vert. Après plusieurs milliers d'essais d'entraînement, les singes connaissent la valeur symbolique de chaque forme et réalisent une intégration statistique quasi-optimale. Leur taux de réponse est une fonction sigmoïdale de l'évidence objective totale, et peut être modélisé par une régression logistique où les poids subjectifs attribués à chaque forme sont étroitement corrélés aux poids objectifs. Le taux de décharge des neurones de l'aire latérale intrapariétale reflète, à chaque instant, la somme de l'évidence apportée par chacune des formes observées jusqu'alors. Ainsi, la décharge des neurones de LIP constitue une approximation de la marche aléatoire postulée par de nombreux modèles de la prise de décision chez l'homme. Le taux de décharge de certains neurones « accumulateurs » serait directement proportionnel au logarithme du rapport de vraisemblance en faveur de telle décision plutôt que telle autre. Yang et Shadlen vont jusqu'à montrer qu'« une variation d'un potentiel d'action par seconde d'un neurone unique est équivalente à un déciban d'évidence ».

L'implémentation neuronale des mécanismes bayésiens

Toutes ces données comportementales et électro-physiologiques suggèrent que le cerveau des primates, et sans doute de nombreuses autres espèces, abrite des mécanismes qui approximent la statistique bayésienne. Au minimum, ces mécanismes doivent représenter plusieurs distributions de probabilité ; représenter et stocker leurs *a priori* ; combiner plusieurs distributions selon la règle du produit (ou par

addition de leur logarithme) ; et enfin, en identifier le maximum *a posteriori*. Quels circuits neuronaux remplissent ces fonctions ?

Selon Alex Pouget et ses collaborateurs, le calcul sur des distributions fait partie du répertoire naturel de toute population de neurones dont les taux de décharge sont aléatoires selon une certaine loi de probabilité (Beck *et al.*, 2008; Ma, Beck, Latham & Pouget, 2006). En effet, dans la plupart des neurones, la variabilité des décharges neuronales suit une loi « de type Poisson », appartenant à une famille exponentielle de distributions, c'est-à-dire que la probabilité d'observer un certain nombre de potentiels d'actions dans un intervalle de temps donné suit une courbe dont la variance est proportionnelle à la moyenne. Chaque neurone possède une courbe d'accord (*tuning curve*) qui le rend apte à décharger en réponse à certains stimuli (par exemple, en fonction de la direction d'un mouvement sur la rétine). Chaque stimulus extérieur (par exemple la vitesse d'un ensemble de points en mouvements) est donc représenté par un vecteur de décharges neuronales à travers une population de neurones. Ma *et coll.* (2006) montrent alors que la règle de Bayes permet de considérer ces décharges comme la représentation d'une distribution de probabilité sur l'espace des stimuli. L'intensité des décharges correspond directement à la largeur de la distribution de probabilité sur les stimuli.

Lorsque la variabilité des décharges neuronales appartient à la famille exponentielle, la somme arithmétique des décharges de deux populations de neurones réalise automatiquement le calcul du produit des distributions qu'elles représentent. Ce type de réseau offre également une solution simple au problème de marginalisation, qui consiste à calculer les probabilités en s'abstrayant de variables de non-intérêt (par exemple, inférer la vitesse d'un stimulus en s'abstrayant du contraste de l'image). Enfin, l'addition de connections latérales permet d'identifier rapidement la valeur approximative du maximum *a posteriori* (réseau dit « *winner take all* »). Tous les ingrédients d'une machine bayésienne neuronale sont donc réunis.

Beck et ses collaborateurs (2008) montrent que ces principes très généraux suffisent à capturer l'ensemble des données du groupe de Mike Shadlen sur la prise de décision optimale par accumulation d'évidence. Le réseau proposé calcule exactement, et à chaque instant, la distribution de probabilité *a posteriori* de la vitesse du stimulus, même si le contraste varie, sans avoir besoin de connaître ou d'estimer ce contraste, et sans savoir combien de temps s'est écoulé depuis le début de l'essai. Cette implémentation des mécanismes d'inférence bayésienne est donc très séduisante, même si elle n'est peut-être pas le seul modèle compatible avec les données observées.

Le cerveau vu comme un système prédictif

Il nous reste à aborder l'une des dernières notions qui appartiennent à la sphère des modèles bayésiens : l'idée de « codage prédictif » (*predictive coding*) (Mumford, 1992 ; Rao & Ballard, 1999). En effet, l'hypothèse du « cerveau bayésien » postule que notre cerveau infère, à partir des entrées sensorielles, un modèle interne du monde extérieur. À son tour, ce modèle interne peut être utilisé pour créer des anticipations sur les entrées sensorielles. L'hypothèse du codage prédictif suppose que le cerveau génère en permanence de telles anticipations, et génère un signal de surprise ou d'erreur lorsque ces prédictions sont violées par des entrées sensorielles inattendues.

L'idée que le cerveau ne fonctionne pas comme un dispositif passif d'entrée-sortie, mais comme un système actif capable de générer des prédictions et d'en vérifier la validité, possède une longue histoire en éthologie, psychologie, et neurosciences. Elle a été proposée sous des formes les plus diverses : concepts de copie efférente (von Helmholtz, von Holst), de critique interne (Sutton & Barto) ou de prédiction de la récompense (Schultz). Les avantages en sont nombreux. Prédire, c'est gagner du temps en disposant d'informations à l'avance, parfois avant même qu'elles n'atteignent les récepteurs sensoriels. Utiliser le passé pour prédire le présent peut également aider à interpréter des entrées sensorielles bruitées, voire à remplacer totalement un stimulus masqué, manqué ou absent (notion de filtre optimal de Kalman). La prédiction conduit également à simplifier l'architecture et le traitement des données, en comprimant l'information : comme les systèmes de compression de son ou d'image (JPEG, MPEG), notre cerveau n'a peut-être pas besoin de se représenter ou de transmettre ce qu'il sait déjà prédire, seule compte pour lui l'erreur de prédiction. Enfin, les algorithmes prédictifs fournissent une implémentation efficace de l'inférence bayésienne, dans la mesure où la maximisation de la vraisemblance d'un modèle des entrées sensorielles est équivalente à la minimisation de l'erreur de prédiction sur ces entrées.

Pour Karl Friston (Friston, 2005, 2010), l'hypothèse du codage prédictif s'inscrit dans un cadre théorique beaucoup plus large, le principe de minimisation de l'énergie libre. Il s'agit, dans les mots mêmes de son auteur, d'une « formulation mathématique de la manière dont les agents biologiques résistent à la tendance naturelle au désordre » : ils « maintiennent leur état dans un environnement changeant ». Les états de l'organisme doivent donc être de basse entropie : « les agents biologiques doivent minimiser la moyenne à long terme de la surprise ». L'énergie libre est une équation mathématique qui donne une borne supérieure sur la surprise : « si les agents minimisent l'énergie libre, ils minimisent implicitement la surprise ». Dans le domaine de la perception, le principe conduit à optimiser les inférences perceptives. « L'agent infère ou représente, de façon implicite, les causes de ses entrées sensorielles selon les principes de l'inférence bayésienne optimale ».

L'énergie libre peut être maximisée par l'algorithme EM (*Expectation-Maximisation*), une procédure itérative qui, en alternant deux étapes (E et M) de façon répétée, permet de trouver le maximum de vraisemblance ou le maximum *a posteriori* des paramètres d'un modèle probabiliste. Friston propose que le cerveau utilise un algorithme EM. Si le modèle générateur est gaussien, il montre que cela revient à propager des erreurs de prédiction. Chaque aire du cortex réaliserait une inférence causale spécifique, et le passage de messages d'erreurs entre aires permettrait leur mise à jour. L'architecture en couches du cortex correspondrait à une répartition naturelle du calcul. Les connections descendantes implémenteraient le modèle prédictif (*forward model*), c'est-à-dire la prédiction des signaux du niveau n, sur la base des représentations inférées au niveau $n + 1$. Inversement, les connections ascendantes, issues des couches supérieures du cortex, transmettraient l'erreur de prédiction, c'est-à-dire la différence entre l'entrée reçue et sa prédiction. Le signal d'erreur est utilisé pour mettre à jour le modèle de niveau supérieur, afin d'éviter que cette erreur ne se reproduise à l'avenir.

Bien que le modèle précis que propose Friston n'ait pas reçu de validation définitive, et ait été critiqué pour son ambiguïté (Spratling, 2008), l'idée plus générale d'un échange bidirectionnel de prédictions et de signaux d'erreurs rend compte d'un ensemble exceptionnellement vaste d'observations électro-physiologiques. En effet,

de nombreuses réponses cérébrales sont évoquées par des stimuli inattendus ou qui violent une règle précédemment inférée. Par exemple, la présentation de sons répétés, puis d'un son nouveau, élicite une onde (*mismatch response* [MMR] ou *mismatch negativity* [MMN]) facilement mesurable à la surface du scalp, à tous les âges de la vie. D'autres ondes similaires indexent la surprise dans d'autres modalités sensorielles ou à plus haut niveau : représentation sémantique des mots ou des images (N400), anomalies syntaxiques (ELAN), conscience de la nécessité d'une mise à jour de la mémoire de travail (P3b).

Bien que certains tentent d'expliquer ces réponses par un simple processus synaptique d'habituation/récupération (May & Tiitinen, 2010), de nombreuses données récentes militent en faveur de l'hypothèse du codage prédictif. Dans le cas de l'onde MMN, la réponse continue d'être observée lorsque les stimuli alternent (ABAB…) ou sont présentés par paires (AB) et que c'est donc la répétition (AA) qui constitue un stimulus nouveau (Horvath & Winkler, 2004). Analysant un paradigme de ce type au niveau neuronal dans la modalité visuelle, Meyer et Olson (2011) montrent clairement que la diminution des décharges neuronales au deuxième élément de la paire AB ne peut s'expliquer que par un signal de prédiction. La surprise observée lorsqu'une paire inattendue est présentée (AC) entraîne une décharge neuronale si précoce qu'elle indique nécessairement la violation d'une attente perceptive.

Le modèle prédictif suppose qu'à chaque instant, notre cerveau génère des attentes sensorielles. Avant même de recevoir un stimulus, un signal descendant tente d'en annuler les conséquences sensorielles. Seule l'erreur, c'est-à-dire la différence entre la prédiction et l'observation, est représentée dans les décharges neuronales transmises. S'en déduit une prédiction simple : si l'on supprime l'entrée attendue, une réponse neuronale devrait être évoquée par *l'absence* d'un son attendu. Cette prédiction est vérifiée (Joutsiniemi & Hari, 1989). Le paradigme d'omission inattendue d'un stimulus attendu offre une manière simple d'isoler un signal prédictif et d'en étudier les propriétés d'organisation hiérarchique (Wacongne *et al.*, 2011).

Le modèle de Friston rend compte de ces observations, mais il se cantonne au niveau d'une description mathématique phénomologique. Catherine Wacongne, Jean-Pierre Changeux et moi-même avons proposé un modèle neuro-réaliste du codage prédictif, qui capture la grande majorité des observations sur la *mismatch negativity* (Wacongne, Changeux & Dehaene, 2012).

Conclusion

L'hypothèse du « cerveau bayésien » s'avère remarquablement productive dans de nombreux champs des neurosciences cognitives. Le comportement animal et humain suggère que les adultes et les enfants possèdent une vaste capacité d'inférence statistique à de multiples niveaux (perception, action, langage…). L'architecture même du cortex pourrait s'expliquer par la réplication, à plusieurs niveaux hiérarchiques, d'un même circuit neuronal bayésien. Dans le cours de 2012-2013, nous étudierons comment l'hypothèse bayésienne s'applique aux tout premiers apprentissages, lorsque l'enfant apprend à ajuster ses modèles internes en fonction de l'environnement qui l'entoure et de l'éducation qu'il reçoit.

Références bibliographiques

Aly M. & Yonelinas A.P., « Bridging consciousness and cognition in memory and perception: evidence for both state and strength processes », *PLoS ONE*, 7(1), 2012, e30231 [doi : 10.1371/journal.pone.0030231].

Bahrami B., Olsen K., Latham P.E., Roepstorff A., Rees G., Frith C.D., « Optimally interacting minds », *Science*, 329(5995), 2010, 1081-85.

Battaglia P.W. & Schrater P.R., « Humans trade off viewing time and movement duration to improve visuomotor accuracy in a fast reaching task », *J. Neurosci.*, 27(26), 2007, 6984-94.

Beck J.M., Ma W.J., Kiani R., Hanks T., Churchland A.K., Roitman J. *et al.*, « Probabilistic population codes for Bayesian decision making », *Neuron*, 60(6), 2008, 1142-52.

Braun D.A., Nagengast A.J., Wolpert D.M., « Risk-sensitivity in sensorimotor control », *Front. Hum. Neurosci.*, 5, 2011 [doi : 10.3389/fnhum.2011.00001].

Cosmides L., « The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task », *Cognition*, 31(3), 1989, 187-276.

Ernst M.O., « Learning to integrate arbitrary signals from vision and touch », *J. Vis.*, 7(5), 2007, 1-14.

Ernst M.O. & Banks M.S., « Humans integrate visual and haptic information in a statistically optimal fashion », *Nature*, 415(6870), 2002, 429-33.

Friston K., « A theory of cortical responses », *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 360(1456), 2005, 815-36.

Friston K., « The free-energy principle: a unified brain theory? », *Nat. Rev. Neurosci.*, 11(2), 2010, 127-38.

Gigerenzer G. & Hoffrage U., « How to improve Bayesian reasoning without instruction: frequency formats », *Psychological Review*, 102(4), 1995, 684-704.

Girshick A.R., Landy M.S., Simoncelli E.P., « Cardinal rules: visual orientation perception reflects knowledge of environmental statistics », *Nat. Neurosci.*, 14(7), 2011, 926-32.

Gold J.I. & Shadlen M.N., « Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward », *Neuron*, 36(2), 2002, 299-308.

Gweon H., Tenenbaum J.B., Schulz L.E., « Infants consider both the sample and the sampling process in inductive generalization », *Proc. Natl. Acad. Sci. USA*, 107(20), 2010, 9066-71.

Horvath J. & Winkler I., « How the human auditory system treats repetition amongst change », *Neurosci. Lett.*, 368(2), 2004, 157-61.

Howe C.Q. & Purves D., « Range image statistics can explain the anomalous perception of length », *Proc. Natl. Acad. Sci. USA*, 99(20), 2002, 13184-88.

Jacob F., « Evolution and tinkering », *Science*, 196(4295), 1977, 1161-66.

Jaynes E.T., *Probability theory: the logic of science*, Cambridge, Cambridge University Press, 2003.

Joutsiniemi S.L. & Hari R., « Omissions of auditory stimuli may activate frontal cortex », *Eur. J. Neurosci.*, 1(5), 1989, 524-28.

Kemp C. & Tenenbaum J.B., « The discovery of structural form », *Proc. Natl. Acad. Sci. USA*, 105(31), 2008, 10687-92.

Kemp C. & Tenenbaum J.B., « Structured statistical models of inductive reasoning », *Psychol. Rev.*, 116(1), 2009, 20-58.

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

Kording K.P. & Wolpert D.M., « Bayesian integration in sensorimotor learning », *Nature*, 427(6971), 2004, 244-47.

Krynski T.R. & Tenenbaum J.B., « The role of causality in judgment under uncertainty », *J. Exp. Psychol. Gen.*, 136(3), 2007, 430-50.

Lorenz J., Rauhut H., Schweitzer F., Helbing D., « How social influence can undermine the wisdom of crowd effect », *Proc. Natl. Acad. Sci. USA*, 108(22), 2011, 9020-25.

Ma W.J., Beck J.M., Latham P.E., Pouget A., « Bayesian inference with probabilistic population codes », *Nat. Neurosci.*, 9(11), 2006, 1432-38.

Maloney L.T. & Zhang H., « Decision-theoretic models of visual perception and action », *Vision Res.*, 50(23), 2010, 2362-74.

May P.J. & Tiitinen H., « Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained », *Psychophysiology*, 47(1), 2010, 66-122.

Meyer T. & Olson C.R., « Statistical learning of visual transitions in monkey inferotemporal cortex », *Proc. Natl. Acad. Sci. USA*, 108(48), 2011, 19401-06.

Moreno-Bote R., Knill D.C., Pouget A., « Bayesian sampling in visual perception », *Proc. Natl. Acad. Sci. USA*, 108(30), 2011, 12491-96.

Morgenstern Y., Murray R.F., Harris L.R., « The human visual system's assumption that light comes from above is weak », *Proc. Natl. Acad. Sci. USA*, 108(30), 2011, 12551-53.

Mumford D., « On the computational architecture of the neocortex. II. The role of cortico-cortical loops », *Biol. Cybern.*, 66(3), 1992, 241-51.

Perfors A., Tenenbaum J.B., Griffiths T.L., Xu F., « A tutorial introduction to Bayesian models of cognitive development », *Cognition*, 120(3), 2011, 302-21.

Petzschner F.H. & Glasauer S., « Iterative Bayesian estimation as an explanation for range and regression effects: a study on human path integration », *J. Neurosci.*, 31(47), 2011, 17220-29.

Purves D., Lotto R.B., Williams S.M., Nundy S., Yang Z., « Why we see things the way we do: evidence for a wholly empirical strategy of vision », *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 356(1407), 2001, 285-97.

Quine W.V.O., *Word and object*, Cambridge (Mass.), MIT Press, 1960.

Rao R.P. & Ballard D.H., « Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects », *Nat. Neurosci.*, 2(1), 1999, 79-87.

Sergent C., Baillet S., Dehaene S., « Timing of the brain events underlying access to consciousness during the attentional blink », *Nat. Neurosci.*, 8(10), 2005, 1391-1400.

Sergent C. & Dehaene S., « Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blin », *Psychol. Sci.*, 15(11), 2004, 720-28.

Shepard R.N., « Toward a universal law of generalization for psychological science », *Science*, 237(4820), 1987, 1317-23.

Smith L.B., Jones S.S., Landau B., Gershkoff-Stowe L., Samuelson L., « Object name learning provides on-the-job training for attention », *Psychological Science (Wiley-Blackwell)*, 13(1), 2002, 13-19.

Spratling M.W., « Reconciling predictive coding and biased competition models of cortical function », Front. Comput. *Neurosci*., 2, 2008 [doi: 10.3389/neuro.10.004.2008].

Teglas E., Girotto V., Gonzalez M., Bonatti L.L., « Intuitions of probabilities shape expectations about the future at 12 months and beyond », *Proc. Natl. Acad. Sci. USA*, 104(48), 2007, 19156-59.

Teglas E., Vul E., Girotto V., Gonzalez M., Tenenbaum J.B., Bonatti L.L., « Pure reasoning in 12-month-old infants as probabilistic inference », *Science*, 332(6033), 2011, 1054-59.

Tenenbaum J.B., Kemp C., Griffiths T.L., Goodman N.D., « How to grow a mind: statistics, structure, and abstraction », *Science*, 331(6022), 2011, 1279-85.

Trommershauser J., Maloney L.T., Landy M.S., « Decision making, movement planning and statistical decision theory », *Trends Cogn. Sci.*, 12(8), 2008, 291-97.

Tversky A. & Kahneman D., « The framing of decisions and the psychology of choice », *Science*, 211(4481), 1981, 453-58.

Vul E., Hanus D., Kanwisher N., « Attention as inference: selection is probabilistic; responses are all-or-none samples », *J. Exp. Psychol. Gen.*, 138(4), 2009, 546-60.

Vul E., Nieuwenstein M., Kanwisher N., « Temporal selection is suppressed, delayed, and diffused during the attentional blink », *Psychol. Sci.*, 19(1), 2008, 55-61.

Vul E. & Pashler H., « Measuring the crowd within: probabilistic representations within individuals », *Psychological Science* (Wiley-Blackwell), 19(7), 2008, 645-47.

Wacongne C., Changeux J.P., Dehaene S., « A neuronal model of predictive coding accounting for the mismatch negativity », *J. Neurosci.*, 32(11), 2012, 3665-78.

Wacongne C., Labyt E., van Wassenhove V., Bekinschtein T., Naccache L., Dehaene S., « Evidence for a hierarchy of predictions and prediction errors in human cortex », *Proc. Natl. Acad. Sci. USA*, 108(51), 2011, 20754-59.

Weiss Y., Simoncelli E.P., Adelson E.H., « Motion illusions as optimal percepts », *Nat. Neurosci.*, 5(6), 2002, 598-604.

Xu F. & Garcia V., « Intuitive statistics by 8-month-old infants », *Proc. Natl. Acad. Sci. USA*, 105(13), 2008, 5012-15.

Yang T. & Shadlen M.N., « Probabilistic reasoning by neurons », *Nature*, 447(7148), 2007, 1075-80.

SÉMINAIRE : LE NOURRISSON : CE QU'IL SAIT, CE QU'IL APPREND (CORE KNOWLEDGE AND LEARNING IN INFANTS)

En complément du cours, le séminaire a fait intervenir sept chercheurs dont les travaux portent sur l'origine des compétences cognitives du très jeune enfant. Dans les dix dernières années, en effet, la recherche sur la cognition du nourrisson a considérablement progressé. Des compétences insoupçonnées ont été mises en évidence, par exemple, dans l'anticipation des propriétés statistiques des ensembles d'objets (travaux d'Arno Teglas, Luca Bonatti, Fei Xu, Laura Schulz). Dans le même temps, le débat classique sur l'origine de ces compétences a été relancé par le développement de modèles théoriques bayésiens. Ces modèles proposent un mécanisme précis d'apprentissage qui pourrait expliquer la vitesse remarquable avec laquelle les compétences cognitives de l'enfant progressent dans les premières années de la vie (travaux de Joshua Tenenbaum). L'apprentissage du langage fournit un ensemble de données particulièrement pertinentes pour mettre à l'épreuve ce modèle. Dans ce domaine, en effet, les hypothèses nativistes des années 1990 sont en partie abandonnées. L'expérimentation et la théorie convergent pour suggérer que l'enfant pourrait, par des inférences bayésiennes simples, acquérir rapidement diverses compétences complexes sur la segmentation de la parole, la formation du lexique, l'extraction des règles syntaxiques et le sens des mots. Toutefois, cet apprentissage nécessite de supposer l'existence, chez le nourrisson, d'un espace contraint d'hypothèses linguistiques, ainsi que d'une machinerie complexe d'inférence.

Afin de faire la part des compétences précoces et de l'apprentissage rapide dans la cognition du très jeune enfant, le séminaire a présenté les exposés de sept spécialistes du développement cognitif :

– Anne Christophe (CNRS, Paris) : L'acquisition précoce de la syntaxe ;

– Patricia Kuhl (University of Washington, Seattle) : *The first stages of first and second-language acquisition* ;

– Benoist Schaal (CNRS, Dijon) : La cognition olfactive du nouveau-né et du nourrisson ;

– Francesca Simion (Université de Padoue) : *Babies in society: perception of faces and biological motion* ;

– Véronique Izard (CNRS, Paris) : L'intuition mathématique : du bébé à l'adulte ;

– Jenny Saffran (University of Wisconsin, Madison) : *Statistical learning and infant language acquisition* ;

– Laura Schulz (MIT) : *Causality and inference in infants*.

AUTRES ENSEIGNEMENTS

Deux autres cours et séminaires ont été donnés par Stanislas Dehaene à l'université Macquarie de Sydney (Australie).

ACTIVITÉS DE RECHERCHE DU LABORATOIRE

Mécanismes cérébraux de la conscience

Depuis les années 1990, avec Jean-Pierre Changeux, nous avons conçu une approche théorique et expérimentale permettant d'analyser les mécanismes cognitifs et neurobiologiques impliqués dans l'accès à la conscience chez l'homme. En 2011, nous avions publié une vaste synthèse de ces réflexions théoriques dans le journal *Neuron*.

Pour mesurer l'activité neuronale lors de la prise de conscience, plusieurs paradigmes expérimentaux permettent de comparer de l'activité cérébrale dans des conditions expérimentales minimalement différentes, mais dont l'une est consciente tandis que l'autre demeure non-consciente. La combinaison de ces travaux avec des mesures de l'activité cérébrale a permis d'établir la succession des événements qui se déroulent dans le cerveau lors de la prise de conscience. Lors du traitement conscient, entre 200 et 400 millisecondes après la présentation d'un stimulus auditif ou visuel, un vaste réseau cérébral qui inclut le cortex préfrontal montre soudainement une activité soutenue et de grande ampleur. Selon le modèle de l'espace de travail global, cet embrasement du réseau préfrontal reflète l'accès de l'information à un réseau neuronal richement interconnecté. L'accès d'une information à la conscience correspondrait à l'entrée dans un « espace de travail neuronal » constitué de neurones à axones longs, particulièrement denses dans le cortex préfrontal. Cette théorie, appuyée par des simulations de réseaux de neurones, rassemble une grande variété d'observations expérimentales, et notamment les situations cliniques où l'accès à la conscience est altéré ou empêché, par exemple dans le cas de l'anesthésie générale, du coma ou de maladies psychiatriques comme la schizophrénie.

Les recherches du laboratoire tentent de mettre à l'épreuve ce modèle théorique. Avec Sébastien Marti, en utilisant le nouvel équipement de magnéto-encéphalographie du centre NeuroSpin, nous avons vérifié une prédiction du modèle : lorsqu'on demande à un participant de réaliser deux tâches successives, l'espace de travail impose un goulot d'étranglement central qui ralentit le traitement du second stimulus (effet de « période psychologique réfractaire », *psychological refractory period* ou

PRP). Il empêche parfois totalement son accès à la conscience (effet de « clignement attentionnel », *attentional blink* ou AB). En accord avec le modèle, nos résultats montrent, pour la première fois, que ces deux phénomènes sont étroitement reliés : ils peuvent survenir au cours de la même tâche et leurs mécanismes cérébraux sont très similaires : l'activité préfrontal tardive, vers 350 millisecondes et au-delà, est différée dans le premier cas (PRP) et absente dans le second (AB).

Avec Guido Hesselmann, nous avons exploré les mécanismes de la PRP par des enregistrements simultanés en EEG et en IRMf, tant chez le sujet normal que chez un patient avec une lésion de la partie postérieur du corps calleux. Chez ce dernier, en réponse à un stimulus visuel présenté dans un seul hémichamp visuel, on n'observe pas un embrasement tardif bilatéral, mais seulement restreint à l'hémisphère controlatéral. Ces résultats sont compatibles avec l'hypothèse, initialement proposée par Michael Gazzaniga, d'une dissociation partielle de la conscience chez les patients atteints de lésions du corps calleux.

Enfin, une analyse de l'accès à la conscience au cours du masquage visuel a été réalisée en collaboration avec Valentin Wyart et Catherine Tallon-Baudry ; et une nouvelle démonstration de la profondeur du traitement non-conscient des nombres a été réalisée par Filip Van Opstal.

Codage prédictif et erreur de prédiction

L'hypothèse du codage prédictif postule que le cerveau comprend des mécanismes d'apprentissage bayésien qui internalisent les régularités des entrées sensorielles et anticipent sur leur survenue (voir le cours de 2011-2012). Cependant, la nature précise des circuits neuronaux qui permettent cette anticipation n'est pas encore connue. Avec Catherine Wacongne, thésarde, et Jean-Pierre Changeux, nous avons développé des simulations détaillées d'une architecture corticale dotée d'une capacité de codage prédictif. Fondé sur des mécanismes neuronaux réalistes, qui font notamment jouer au récepteur NMDA du glutamate un rôle important dans l'apprentissage des régularités, le modèle reproduit l'intégralité des phénomènes associés à la *mismatch negativity* (MMN) – l'émission d'un potentiel de grande amplitude lorsqu'un son inattendu est présenté au sein d'une séquence auditive régulière. Selon cette hypothèse, la MMN reflèterait directement l'erreur de prédiction (la différence entre la prédiction attendue et l'entrée sensorielle reçue) au sein d'un réseau qui internalise les probabilités de transition.

En nous appuyant sur ce modèle, nous avons étudié, avec la technique de magnéto-encéphalographie, les propriétés des réponses à la nouveauté auditive chez l'adulte. Les résultats montrent que la MMN traduit bien une prédiction et non pas une simple habituation (Wacongne *et al.*, PNAS, 2011). En particulier, l'omission d'un stimulus attendu se traduit par l'émission d'une réponse que l'on peut interpréter comme un signal de prédiction. Nous avons également montré que le cerveau humain abrite plusieurs systèmes prédictifs, vraisemblablement organisés de façon hiérarchique (Wacongne *et al.*, *J. Neurosci.*, 2012).

Nous avons également utilisé le traitement de la nouveauté comme index de la profondeur du traitement non-conscient et conscient. La mise en opposition des régularités auditives locales et globales, paradigme que nous avons développé en collaboration avec l'équipe de Lionel Naccache à l'hôpital de la Salpêtrière (Bekinschtein *et al.*, *PNAS*, 2010), montre que la première étape de détection de la

nouveauté (MMN) peut rester présente chez les patients comateux et reflète donc un traitement non-conscient de la nouveauté, tandis que la seconde (P3) indexe plus spécifiquement l'accès de l'information à la conscience. Nous étendons actuellement ces résultats afin de détecter une éventuelle conscience résiduelle chez certains patients en état végétatif (Faugeras *et coll.*, *Neurology*, 2012). Ces recherches font l'objet d'un vaste projet financé par l'*European Research Council*.

Génétique des réseaux du langage

L'année 2012 a vu l'aboutissement d'un projet étendu sur plus de cinq ans et portant sur l'héritabilité génétique des réseaux corticaux du langage. Dirigé par Philippe Pinel, il consistait à créer une grande base de données qui permette de relier imagerie cérébrale et polymorphismes du génome humain, afin d'examiner quels gènes modulent les réseaux cérébraux impliqués dans le traitement du langage parlé, de la lecture et du calcul.

Jusqu'à présent, l'identification de gènes impliqués dans le développement des bases cérébrales du langage et de la lecture était essentiellement fondée sur l'étude de famille de patients. Ainsi, un haplotype du gène TTRAP réduisant l'expression du gène de migration neuronale KIAA0319 avait été précédemment associé à un risque de dyslexie (trouble spécifique de la lecture), tandis qu'une mutation du gène de régulation FOXP2 était associée à une dyspraxie orofaciale ainsi qu'à des difficultés de compréhension du langage. En corrélant, dans une population de 94 adultes normaux, les variations ponctuelles de ces gènes (*single-nucleotide polymorphisms* ou SNPs) avec l'intensité et l'asymétrie de l'activité cérébrale mesurée par IRM, nous avons pu confirmer que l'asymétrie fonctionnelle du cortex temporal est associée à l'haplotype à risque de la région KIAA0319/TTRAP, alors que l'activation des régions frontales de l'hémisphère gauche varie en fonction du polymorphisme du gène FOXP2.

Cette première étude suggère donc que les polymorphismes de ces deux gènes sont associés à des modulations du réseau de la lecture dans des régions distinctes, y compris chez le sujet sain. Le résultat sur FOXP2 est particulièrement intéressant dans la mesure où ce gène, très stable, a subi une évolution récente dans la lignée humaine. Sa mutation pourrait, en combinaison avec d'autres gènes, expliquer l'émergence des compétences linguistiques dans la lignée humaine. À l'avenir, nous envisageons d'étendre la base de données à plusieurs centaines, voire plusieurs milliers d'examens d'IRMf, afin d'explorer l'ensemble du génome pour identifier et proposer aux généticiens et cliniciens de nouveaux gènes candidats pouvant être impliqués dans des pathologies de la cognition humaine.

Codage cérébral de la structure syntaxique des phrases

Au cours des dix dernières années, le laboratoire a clarifié les mécanismes de la lecture de mots isolés. Avec Christophe Pallier, nous commençons à présent l'exploration des réseaux cérébraux qui sous-tendent l'intégration de plusieurs mots au sein d'une phrase. Un financement de l'Agence nationale de la recherche nous a permis de mettre au point un nouveau paradigme expérimental visant à caractériser le codage neural de la structure en constituants des phrases (Pallier *et al.*, *PNAS*, 2011).

L'une des caractéristiques universelles des langues humaines, en effet, est l'organisation des phrases en constituants enchâssés, des groupes consécutifs de mots qui forment des structures arborescentes selon des règles récursives. Afin de caractériser le codage cérébral de ces structures, nous avons créé un jeu de stimuli dont le nombre de mots est fixe, mais dont la taille des constituants varie. En mesurant l'activité cérébrale en IRM fonctionnelle alors que les participants lisaient ces séries de mots, nous avons identifié un réseau cérébral dont l'activité croît avec la taille des constituants. Ce réseau comprend quatre régions situées le long du sillon temporal supérieur gauche, deux régions dans le gyrus frontal inférieur gauche (aire de Broca) ainsi que le putamen gauche. Une fraction de ce réseau continue de s'activer lorsque les mots à contenu sont remplacés par des pseudo-mots équivalents, formant une version expérimentale du célèbre poème *Jabberwocky* de Lewis Carroll – ce qui suggère que ces aires forment un noyau de compétence syntaxique qui continue à fonctionner même en l'absence d'informations lexico-sémantiques.

Le profil d'activation en IRM donne plusieurs indices importants sur la nature du code neural des phrases dans ces régions. L'activation varie de façon logarithmique avec la taille des constituants, aussi bien dans son intensité que dans sa phase. Le délai de l'activation, notamment dans la région frontale inférieure gauche, suggère que le réseau effectue des calculs pendant une longue durée qui s'étend probablement après la fin de la présentation de la phrase. Dans les années à venir, ce travail sera poursuivi en magnéto-encéphalographie et à l'aide d'enregistrements intracrâniens, afin de préciser la dynamique temporelle de l'activation dans ces régions, et la nature précise du code qui s'y trouve.

PUBLICATIONS 2011-2012

Articles originaux

Pegado F., Nakamura K., Cohen L., Dehaene S., « Breaking the symmetry: mirror discrimination for single letters but not for pictures in the Visual Word Form Area », *Neuroimage*, 55, 2011, 742-49.

Szwed M., Dehaene S., Kleinschmidt A., Eger E., Valabrègue R., Amadon A., Cohen L., « Specialization for written words over objects in the visual cortex », *Neuroimage*, 56, 2011, 330-44.

Kamienkowski J.E., Pashler H., Dehaene S., Sigman M., « Effects of practice on task architecture: combined evidence from interference experiments and random-walk models of decision making », *Cognition*, 119(1), 2011, 81-95.

Hesselmann G., Flandin G., Dehaene S., « Probing the cortical network underlying the psychological refractory period: a combined EEG-fMRI study », *Neuroimage*, 56(3), 2011, 1608-21.

Cantlon J.F., Pinel P., Dehaene S., Pelphrey K.A, « Cortical representations of symbols, objects, and faces are pruned back during early childhood », *Cerebral cortex*, 21, 2011, 191-99.

Van Opstal F., de Lange F.P., Dehaene S., « Rapid parallel semantic processing of numbers without awareness », *Cognition*, 120, 2011, 136-47.

Dehaene S., Cohen L., « The unique role of the visual word form area in reading », *Trends in Cognitive Science*, 15(6), 2011, 254-62.

Faugeras F., Rohaut B., Weiss N., Bekinschtein T.A., Galanaud D., Puybasset L., Bolgert F., Sergent C., Cohen L., Dehaene S., Naccache L., « Probing consciousness with event-related potentials in patients who meet clinical criteria for vegetative state », *Neurology*, 77(3), 2011, 264-68.

Vinckier F., Qiao E., Pallier C., Dehaene S., Cohen L., « The impact of letter spacing on reading: a test of the bigram coding hypothesis », *Journal of Vision*, 11(6), 2011, 1-21.

Izard V., Pica P., Spelke E., Dehaene S., « Flexible intuitions of Euclidean geometry in an Amazonian indigene group », *PNAS*, 108, 2011, 9782-87.

Markram H., Meier K., Lippert T., Grillner S., Frackowiak R., Dehaene S., Knoll A., Sompolinsky H., Verstreken K., Defelipe J., Grant S., Changeux J.P., Sariam A., « Introducing the Human Brain Project », *Procedia Computer Science*, 7(1), 2011, 39-42.

Fan Z., Singh K., Muthukumaraswamy S., Sigman M., Dehaene S., Shapiro K., « The cost of serially chaining two cognitive operations », *Psychol. Res.*, 76(5), 2011, 566-578.

Dehaene S., Changeux J.P., « Experimental and theoretical approaches to conscious processing », *Neuron*, 70, 2011, 200-27.

Zylberberg A., Dehaene S., Roelfsema P.R., Sigman M., « The human Turing machine: a neural framework for mental programs », *Trends Cogn. Sci.*, 15, 2011, 293-300.

Pallier C., Devauchelle A.D., Dehaene S., « Cortical representation of the constituent structure of sentences », *PNAS*, 108, 2011, 2522-27.

Marti S., Sigman M., Dehaene S., « A shared cortical bottleneck underlying Attentional Blink and Psychological Refractory Period », *Neuroimage*, 59(3), 2011, 2883-98.

De Lange F.P, van Gaal S., Lamme V.A.F., Dehaene S., « How awareness changes the relative weights of evidence during human decision-making », *PLoS Biol.*, 9(11), 2011, e1001203 [doi : 10.1371/journal.pbio.1001203].

Wacongne C., Labyt E., van Wassenhove V., Bekinschtein T., Naccache L., Dehaene S., « Evidence for a hierarchy of predictions and prediction errors in human cortex », *Proc. Natl. Acad. Sci. USA*, 108(51), 2011, 20754-59.

Faugeras F., Rohaut B., Weiss N., Bekinschtein T.A., Galanaud D., Puybasset L., Bolgert F., Sergent C., Cohen L., Dehaene S., Naccache L., « Event related potentials elicited by violations of auditory regularities in patients with impaired consciousness », *Neuropsychologia*, 50(3), 2012, 403-18.

Pinel P., Fauchereau F., Moreno A., Barbot A., Lathrop M., Zelenika D., Le Bihan D., Poline J.P., Bourgeron T., Dehaene S., « Genetic Variants of FOXP2 and KIAA0319/TTRAP/THEM2 locus are associated with altered brain activation in distinct language-related regions », *J. Neurosci.*, 32(3), 2012, 817-25.

Wyart V., Dehaene S., Tallon-Baudry C., « Early dissociation between neural signatures of endogenous spatial attention and perceptual awareness during visual masking », *Frontiers in Human Neuroscience*, 6, 2012 [doi: 10.3389/fnhum.2012.00016].

Wacongne C., Changeux J.P., Dehaene S., « A neuronal model of predictive coding accounting for the mismatch negativity », *J. Neurosci.*, 32(11), 2012, 3665-78.

Monzalvo K., Fluss J., Billard C., Dehaene S., Dehaene-Lambertz G., « Cortical networks for vision and language in dyslexic and normal children of variable socio-economic status », *Neuroimage*, 61(1), 2012, 258-74.

Szwed M., Ventura P., Querido L., Cohen L., Dehaene S., « Reading acquisition enhances an early visual process of contour integration », *Developmental Science*, 15(1), 2012, 139-49.

Hesselmann G., Naccache L., Cohen L., Dehaene S., « Splitting of the P3 component during dual-task processing in a patient with posterior callosal section », *Cortex*, Available online 27 March 2012, 2012.

Livres

Dehaene S., *The number sense* (2e édition), New York, Oxford University Press, 2011.

Dehaene S. & Brannon E., *Space, time and number in the brain: searching for the foundations of mathematical thought*, London, Elsevier, 2011.

Dehaene S. & Christen Y. (éd.), *Characterizing consciousness: from cognition to the clinic?*, coll. « Research and Perspective in Neurosciences », Springer-Verlag, 2011.

Dehaene S., *Apprendre à lire. Des sciences cognitives à la salle de classe*, Paris, Odile Jacob, 2011.

Chapitres de livres

Sigman M., Dehaene S., « Why does it take time to make a decision? The role of a global workspace in simple decision making », *in* Vartanian O. & Mandel D.R. (éd.), *Neurosciences of decision making*, Hove, Psychology Press, 2011, 11-44.

Dehaene S., « Reading as neuronal recycling », *in* McCardle P., Miller B., Lee J.R., Tzeng O.J.L. (éd.), *Dyslexia across languages*, Baltimore, Paul H. Brookes Publishing, 2011, 102-16.

Dehaene S., Changeux J.P., Naccache L., « The global neuronal workspace model of conscious access: from neuronal architectures to clinical applications », in Dehaene S. & Christen Y. (éd.), *Characterizing consciousness: from cognition to the clinic?*, Paris, Springer-Verlag, 2011, 55-84.

Dehaene S., « Cuando el reciclaje neuronal prolonga la hominización », *in* Lapina S., Sigman M. (éd.), *La pizarra de Babel*, Buenos Aires, Libros del Zorzal, 2012, 91-105.

Revues, commentaires, diffusion des connaissances

King J.R., Bekinschtein T., Dehaene S., « Comment on preserved feedforward but impaired top-down processes in the vegetative state », *Science*, 334(6060), 2011, 1203.

Dehaene S., Pegado F., « O impacto da aprendizagem da leitura sobre o cérebro », *Grupo A, Revista Pátio*, févr.-avr. 2012.

AUTRES ACTIVITÉS

Principales conférences invitées

– « The visual word form area », conférence et discussion avec Cathy Price, Society for the Neurobiology of Language, San Diego, 12 novembre 2011.

– « Les neurones de la lecture », Collège de Belgique, 18 janvier 2011.

– « Quantity coding and computation in the animal and human brain », conférence COSYNE, Salt Lake City, 27 février 2011.

– « Physiological signatures of conscious access and the workspace model », conférence COSYNE, Snowbird, États-Unis, 28 février 2011.

– « Brain mechanisms of numeracy and consequences for education », The Latin American school for education, cognitive and neural sciences (James S. McDonnell Foundation & Universidad de Chile), San Pedro de Atacama, 14 mars 2011.

– « How math comes to mind: intuition, visualization, and teaching », discussion avec le mathématicien Stephen Strogatz, Princeton, 27 avril 2011.

– « Limits of subliminal processing and signatures of conscious access », Association for Psychological Science, Washington, 27 mai 2011.

– « How learning to read changes the visual system », Rank prize lecture, European Conference on Visual Perception (ECVP), Toulouse, 28 août 2011.

– « Reading in the brain », conférence invitée, Duke University, Durham, 7 octobre 2011.

– Enregistrement à New York de l'émission télévisuelle de Charlie Rose consacrée aux mécanismes cérébraux de la conscience (NBC), New York, 14 novembre 2011.

– Session plénière « Emerging fields of science », World Science Forum, Budapest, 17 novembre 2011.

– « Reading as neuronal recycling », The second Latin American school for education, cognitive and neural sciences (James S. McDonnell Foundation), El Calafate, Argentine, 7 mars 2012.

– « Qu'est-ce que lire ? », conférence au Salon du livre, Paris, 18 mars 2012.

– « Signatures of conscious processing in the human brain », Seattle Brain Salon, à l'occasion de l'inauguration de l'institut Paul Allen, Seattle, 19 avril 2012.

– « Education matters: literacy, numeracy and the developing brain », Robarts Research Institute, London, Ontario, 7 mai 2012.

– « The brain circuits for reading: how literacy alters our cortical representations and their connections », The annual Geschwind lecture, Harvard Institute of Medicine, Boston, 9 mai 2012.

– « How learning to read changes the brain », seminar series, Stanford Institute for Neuro-Innovation and Translational Neuroscience, Stanford University, 10 mai 2012.

– Série de trois conférences invitées sur l'alphabétisation et la lecture, à l'occasion de la publication du livre *Les Neurones de la lecture* au Brésil ; Rio, Florianopolis, Brasilia, 6, 13 et 17 juillet 2013.

– « From word to sentence: in search of the brain mechanisms for constituent structure », colloque en l'honneur de Noam Chomsky, Trieste (Italie), 18 septembre 2012.

Thèses soutenues dans le laboratoire

Francois Leroy (15/09/2011, directeur : Ghislaine Dehaene-Lambertz) : « Étude méthodologique et structurale du développement cérébral en IRM : Application aux aires du langage dans une population de nourrissons ».

Felipe Pegado (6/12/2011, directeur : Stanislas Dehaene) : « Impact de l'alphabétisation sur l'organisation cérébrale et la cognition: aspects de la perception visuelle et du langage parlé chez les adultes alphabétisés et non-alphabétisés ».

Participation aux programmes nationaux et internationaux

Responsabilité du pilier « Neurosciences cognitives » du *Human Brain Project*, dossier préparé en réponse à l'appel d'offre *Future Emerging Technologies* de la Communauté européenne.

Participation au comité de sélection du Prix international Kavli en neurosciences.

Participation au conseil scientifique de la direction générale de l'Enseignement scolaire (DGESCO) dirigée par Jean-Michel Blanquer.

Conseil scientifique du Max-Planck Institute de Francfort, dirigé par Pascal Fries.