

## Psychologie cognitive expérimentale

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

### ENSEIGNEMENT

#### Cours : L'accès à la conscience

Ce cours est le second d'une série dédiée à l'examen des mécanismes du traitement conscient dans le cerveau humain. Alors que l'année 2009 était consacrée à la question de la profondeur du traitement non-conscient, nous nous penchons à présent sur les opérations cognitives et neurales qui permettent d'amener une information à la conscience.

La question de la nature et de l'origine de la conscience a joué un rôle central dans l'histoire de la philosophie de la psychologie. Comment un simple assemblage de neurones pourrait-il conduire à une expérience mentale ? Julian Jaynes, dans *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, souligne avec force et lyrisme la profondeur de cette ancienne interrogation :

O, quel monde de visions cachées et de silences entendus que cette contrée immatérielle de l'esprit ! Quelles essences ineffables que ces souvenirs irréels et ces rêveries invisibles ! Et l'intimité de tout cela ! Théâtre secret de monologues silencieux et de conseils anticipés, invisible demeure de tous les états d'âme, de toutes les songeries et de tous les mystères, séjour infini des déceptions et des découvertes. Un royaume entier sur lequel chacun de nous règne seul et replié sur soi, interrogeant ce que nous voulons, ordonnant ce que nous pouvons. Un ermitage caché dans lequel nous pouvons nous livrer à loisir à l'étude du livre agité de ce que nous avons fait et de ce qui nous reste à faire. Un monde intérieur qui est plus moi-même que tout ce que je peux trouver dans un miroir. Cette conscience qui est l'essence de tous mes moi, qui est tout, sans être cependant quoi que ce soit, qu'est-elle donc ?

Et d'où est-elle issue ?

Et pourquoi ?

(Traduction : Guy de Montjou, PUF, 1994)

La question de la nature de la conscience, longtemps bannie de la psychologie expérimentale, fait depuis des siècles l'objet de nombreuses spéculations, toutes plus étranges les unes que les autres. Rappelons brièvement que Descartes, tout en

affirmant explicitement que « Dieu a fabriqué notre corps comme une machine » et en concevant, le premier, « l'idée d'un mécanisme nerveux auto-suffisant et capable d'actions complexes et ayant toute l'apparence de l'intelligence » (William James), ne parvint pas à débarrasser sa philosophie du concept d'une âme immatérielle, qu'il attachait curieusement à la glande pinéale en raison de sa position centrale dans le cerveau. Plus près de nous, le physiologiste John Eccles imagine que l'interaction cartésienne entre corps et âme se produit par la mise en liaison des « dendrons » (unités fondamentales du cortex) et des « psychons » (unités de l'esprit conscient) par effet quantique. Sur ses traces, les physiciens Roger Penrose et Stuart Hameroff voient dans le théorème de Gödel la preuve que l'esprit humain dispose d'intuitions qui dépassent les calculs accessibles aux machines de Turing, et recherchent dans des propriétés inconnues de la matière (gravitation quantique) une forme nouvelle de calcul conscient. Pour le philosophe contemporain Dave Chalmers, il n'est pas incohérent d'imaginer deux cerveaux exactement identiques, atome par atome, mais dont l'un seul soit conscient... Ces crypto-dualismes contemporains ne sont guère plus cohérents que la position de Julian Jaynes lui-même, selon laquelle la conscience humaine est une invention culturelle qui n'existait pas du temps de *Illiade*, les hommes n'étant alors que des automates obéissant à la voix des dieux !

### *Les conditions d'une science expérimentale de la conscience*

Devant de telles propositions aussi bizarres que disparates, on peut légitimement se demander si une véritable science de la conscience est envisageable. Le caractère subjectif de l'expérience consciente interdit-il toute analyse scientifique objective, à la troisième personne, telle qu'on peut la mener par les méthodes de la psychologie cognitive ? L'expérience consciente, par nature subjective, échapperait-elle à l'expérimentation ? Bien entendu, notre cours défend la position contraire. Il entend montrer que la conscience est un objet d'étude légitime, « un phénomène réel, naturel, biologique, littéralement localisé dans le cerveau » (Antti Revonsuo), et que nous disposons de nombreux paradigmes expérimentaux pour l'étudier.

En liaison avec le cours de l'année dernière, soulignons d'abord que le problème de la conscience est bien délimité. Il ne s'agit pas d'analyser la totalité des opérations cérébrales, mais seulement celles qui vont au-delà du traitement non-conscient, pourtant très répandu dans le système nerveux. « Les neurosciences cognitives de la conscience visent à déterminer s'il existe une forme systématique de traitement de l'information, et une classe d'états d'activité du cerveau, qui distinguent systématiquement les états que les sujets identifient comme "conscients" des autres états » (Dehaene & Naccache, 2001).

La recherche des mécanismes cérébraux de la conscience passe donc par une mise en corrélation d'un état mental conscient avec ses mécanismes computationnels et cérébraux, une approche classique et partagée avec de nombreux autres domaines des sciences cognitives. Seule la distingue une condition spécifique : la nécessité de recueillir et de prendre au sérieux l'introspection des participants, qui définit le phénomène même que l'on cherche à étudier. Ainsi « la première étape, cruciale,

consiste à prendre au sérieux ce que les sujets rapportent de leur introspection et de leur phénoménologie. Ces rapports subjectifs sont les phénomènes clés qu'une neuroscience cognitive de la conscience vise à étudier. En tant que tels, ils constituent les données primaires que l'on doit mesurer et enregistrer en parallèle avec toutes les autres données physiologiques ». (Dehaene & Naccache, 2001).

Cette primauté du subjectif n'implique en rien que l'introspection soit un mode privilégié d'accès aux mécanismes intimes de l'esprit. L'introspection des participants peut être limitée, voire fautive, comme dans l'exemple de l'expérience subjective de « sortie du corps » (autoscopie). Elle n'en constitue pas moins une donnée primaire de la conscience dont il faut rendre compte (par exemple par la décharge anormale de neurones intégrateurs, à la jonction pariéto-temporale, impliqués dans la représentation multimodale du corps propre dans l'espace, voir Blanke, Landis, Spinelli, & Seeck, 2004).

En résumé, notre programme de recherche consiste à identifier des corrélats neuronaux objectifs des processus subjectifs accessibles à l'introspection consciente. Avec Francis Crick et Christoph Koch, on peut parler des « corrélats neuronaux de la conscience » (*Neural correlates of consciousness* ou NCC). Cependant, l'objectif va au-delà de la simple mise en corrélation psycho-neurale. Nous recherchons en réalité une liste de conditions nécessaires et suffisantes à l'établissement d'un état subjectif conscient, autrement dit des mécanismes causaux de la conscience.

### *Quelques distinctions théoriques*

Le mot « conscience » est polysémique, et l'introduction d'un vocabulaire précis et restreint est une étape importante de la théorisation. La plupart des chercheurs conviennent de distinguer, au minimum, l'état de conscience et le contenu de la conscience. Le premier usage, intransitif, du mot « conscience » renvoie aux variations graduelles de l'état de vigilance : veille, sommeil, anesthésie, coma, état végétatif... Le second usage, intransitif, fait référence à la prise de conscience d'une information particulière. On parle alors de « conscience d'accès », ou d'accès d'une information à la conscience. Selon le point de vue de Larry Weiskrantz (1997), le critère expérimental essentiel est la « rapportabilité » des stimuli : lorsque nous sommes conscients d'une information, il nous est toujours possible de la rapporter, verbalement ou par le biais de gestes de communication. L'information consciente est accessible : pratiquement tous les processeurs disponibles peuvent l'exploiter (rapport verbal, évaluation, mémorisation, intention, contrôle, stratégies...).

D'autres chercheurs privilégient des aspects différents du mot « conscience ». Pour certains (Antonio Damasio, Gerald Edelman), la référence au soi joue un rôle essentiel : pour devenir conscient, chaque percept doit être mis en relation avec un point de vue subjectif (le *self*), une représentation de l'organisme dans l'acte même de percevoir. D'autres mettent en avant des fonctions plus avancées de la conscience : la faculté de récursivité qui conduit à des pensées « d'ordre supérieur », le fait de représenter que nous pensons ce que nous pensons (David Rosenthal, Axel Cleeremans) ; le développement d'une narration explicite et parfois fictive de

nos actes (Michael Gazzaniga, Lionel Naccache, Dan Wegner) ; enfin la capacité de partager nos états mentaux avec autrui, qui suggère que l'émergence de la conscience humaine au cours de l'évolution pourrait être liée aux compétences sociales propres à notre espèce (Chris Frith).

Pour d'autres enfin (Ned Block, Victor Lamme), toute perception consciente s'accompagne d'une expérience subjective riche, personnelle, unique, qui peut sembler ineffable (la couleur unique d'un coucher de soleil, le frisson musical, etc.). Ces phénomènes sont désignés par Lewis (1929) sous le terme de *qualia* – les qualités particulières de l'expérience sensible ou « phénoménale ». La conscience phénoménale précéderait et dépasserait ce que nous sommes capables de rapporter (c'est-à-dire la conscience d'accès). Le philosophe David Chalmers distingue ainsi le problème « facile » de la conscience (qui consiste tout de même à expliquer l'intégralité des comportements, y compris les décisions volontaires, les rapports verbaux et les descriptions introspectives !), du problème « difficile », qui consiste à expliquer pourquoi ces opérations s'accompagnent d'une expérience consciente (*what it is like*). Un « gouffre explicatif » (*explanatory gap*) séparerait toute tentative d'explication matérialiste et externaliste des états mentaux (« à la troisième personne ») de l'expérience intime d'être conscient (« à la première personne »).

Bien que toutes ces facettes de la conscience présentent une certaine légitimité, le cours s'est largement limité à l'analyse des mécanismes de l'accès à la conscience. Deux raisons président à ce choix. D'une part, c'est sur cette question bien délimitée que se concentrent la plupart des résultats scientifiques récents. D'autre part, dans l'état actuel de nos connaissances, les concepts de *self* et de *qualia* s'appuient plus sur une intuition intime que sur une authentique modélisation analytique. Or, on peut penser, avec Daniel Dennett (2001), que la recherche conduit souvent à une révision radicale de telles croyances. Même si l'introspection semble valider l'idée d'une conscience phénoménale de la totalité d'une scène visuelle, il se peut qu'elle ne reflète qu'une « illusion de voir ». Peut-être ne voyons-nous réellement, à chaque instant, qu'une toute petite fraction du monde extérieur – notre sentiment de conscience globale résulterait du fait que, dès que nous y prêtons attention, chacune d'elles devient instantanément accessible (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006). Plusieurs expériences récentes menées sur les paradigmes de cécité au changement et de mémoire iconique soutiennent ce point de vue (de Gardelle, Sackur, & Kouider, 2009 ; Simons & Ambinder, 2005).

### *Théories de la conscience d'accès*

Il existe un début de convergence, sinon vers un « modèle standard » de la conscience d'accès, du moins en faveur d'un ensemble d'idées proposées depuis les années 1950-1960, et de mieux en mieux acceptées aujourd'hui.

*Un système de supervision centrale.* Selon cette idée déjà présente chez William James (1890), les traitements conscients interviennent à un niveau hiérarchiquement supérieur où ils régulent les autres opérations « afin de diriger un système nerveux devenu trop complexe pour se réguler lui-même ». Michael Posner distingue ainsi

les processus mentaux automatiques et ceux dits « contrôlés ». Les premiers démarrent sans intention, n'interfèrent pas avec les autres, et échappent à la conscience. Les seconds font appel à un système central à capacité limitée, qui ne permet pas l'exécution de plusieurs opérations simultanées sans interférence ; ils sont dépendants de nos intentions et conduisent à une expérience consciente. Selon Tim Shallice, le comportement volontaire conscient résulte d'un système de supervision, hiérarchiquement supérieur aux processeurs automatiques, et chargé de leur contrôle et de leur inhibition.

*Une étape sélective à capacité limitée.* Selon Broadbent, la perception comprend un filtre qui ne laisse entrer qu'une fraction des informations dans un « canal à capacité limitée ». Cette architecture répond à une nécessité algorithmique : tout organisme est bombardé en permanence de stimuli sensoriels qui excèdent sa capacité de traitement et surtout d'action.

L'attention sélective joue donc le rôle de filtre qui ne retient qu'une fraction des entrées. La conscience serait associée au traitement plus approfondi de ces données choisies pour leur pertinence vis-à-vis des buts de l'organisme. L'hypothèse d'une limite centrale revient, sous des formes distinctes, dans plusieurs théories du « goulot d'étranglement central » (Pashler, 1994), des « deux étapes de la perception » (Chun & Potter, 1995), ou d'une conscience perceptive limitée à une « carte Booléenne » (Huang, Treisman, & Pashler, 2007).

Il faut noter que ces théories n'impliquent pas une identité entre les processus d'attention et de conscience. Si l'on s'en tient à définition de l'attention comme la sélection d'un objet ou d'une position spatiale, et l'amplification de ses attributs sensoriels, de nombreuses données expérimentales montrent que ces processus peuvent se déployer sans pour autant conduire à une perception consciente. En effet, des effets d'alerte et de réorientation de l'attention peuvent être induits par des indices subliminaux (Bressan & Pizzighello, 2008 ; Woodman & Luck, 2003). L'attention peut amplifier le traitement de stimuli qui restent cependant subliminaux (Naccache, Blandin, & Dehaene, 2002). Enfin les corrélats cérébraux de l'attention sélective et de l'accès à la conscience de stimuli présentés au seuil sont radicalement différents (Wyart & Tallon-Baudry, 2008). Ainsi, attention et conscience sont dissociables – mais en présence de stimuli multiples, l'attention sélective apparaît comme un préalable essentiel, nécessaire mais pas suffisant – la « porte d'entrée » du traitement conscient (Mack & Rock, 1998).

*L'importance de l'amplification tardive et descendante.* Gerald Edelman voit dans la « réentrée », c'est-à-dire l'échange permanent, dynamique et récursif de signaux parallèles entre cartes cérébrales, un mécanisme indispensable de la perception unifiée. Certaines formes de réentrée globales au cortex conduiraient à la conscience. L'idée est développée par Giulio Tononi, selon lequel l'information consciente est à la fois *intégrée* (unifiée en un tout) et *différenciée* (le nombre de contenus de conscience est gigantesque, et chacun d'eux est très informatif). La différenciation résulte de la multiplicité des groupes neuronaux codants, distribués dans de nombreuses régions du cortex, tandis que l'intégration émerge des nombreuses boucles réentrantes qui les lient. Elle peut être mesurée mathématiquement par le

paramètre  $\Phi$ , qui évalue l'information mutuelle minimale entre deux sous-parties du système complet. Notons que cette théorie conduit à prédire une absence de localisation cérébrale précise : la conscience serait associée à un « noyau dynamique » (*dynamic core*) thalamo-cortical distribué dont les contours ne cesseraient de changer et qui varierait considérablement d'une personne à l'autre. Le neurophysiologiste Victor Lamme, lui, développe un modèle plus précis sur le plan neuronal. Il postule qu'un traitement récurrent local, résultant de la conjonction d'activations ascendantes (*feedforward sweep*) et descendantes (*top-down amplification*), permet la conscience phénoménale d'une information sensorielle. L'accès à cette information se ferait dans un second temps seulement, par le biais de connexions récurrentes globales au cortex.

*Un espace interne de synthèse, de maintien et de partage des données.* La métaphore du théâtre conscient est proposée par Taine dès 1870 : « On peut comparer l'esprit d'un homme à un théâtre d'une profondeur indéfinie, dont la rampe est très étroite, mais dont la scène va s'élargissant à partir de la rampe. Devant cette rampe éclairée, il n'y a guère de place que pour un seul acteur... Au delà, sur les divers plans de la scène, sont d'autres groupes d'autant moins distincts qu'ils sont plus loin de la rampe. » Critiquée par Dennett (voir notamment Dennett, 1991), cette métaphore est néanmoins développée par Bernard Baars (1989) dans son modèle de l'espace de travail conscient. Selon celui-ci, de très nombreux processeurs spécialisés opèrent en parallèle de façon non-consciente, tandis qu'à un instant donné, une seule coalition de processeurs dominants envoie son résultat dans l'espace de travail global où il est diffusé à l'ensemble du système. La conscience joue donc le rôle d'un espace de synthèse, de maintien et de partage des données.

En 1998, Jean-Pierre Changeux et moi-même avons présenté une synthèse de ces quatre idées sous forme d'une architecture cérébrale plausible d'un « espace de travail neuronal global » (*global neuronal workspace* ou GNW). Notre théorie distingue deux espaces computationnels : les processeurs spécialisés, modulaires, encapsulés et automatiques, et l'espace global formé de neurones à axones longs capables de diffuser l'information à des régions cérébrales distantes. L'accès à la conscience correspondrait à l'ignition soudaine de ces neurones à axones longs, distribués dans l'ensemble du cortex cérébral, mais particulièrement denses dans les cortex préfrontaux et pariétaux. Selon la théorie, durant la perception consciente, le GNW se stabilise, pour une durée minimale, dans un état réverbérant métastable, qui permet de garder l'information « en ligne » pour une durée arbitraire. Ainsi, nous appelons « consciente » toute information qui, ayant atteint ce niveau de traitement central et global, est devenue accessible à tous les processeurs et donc, en particulier, aux processeurs verbaux qui permettent un rapport verbal.

Notons que, selon le modèle, un même processeur peut participer à des traitements conscients et non-conscients, le mode conscient se caractérisant par une amplification descendante et une mise en synchronie avec les neurones du GNW. Par ailleurs, les neurones du GNW sont supposés être le siège d'une activité spontanée permanente modulée par (1) des signaux ascendants d'éveil ; (2) des signaux sélectifs de renforcement. La série temporelle d'états métastables qui en

résulte correspondrait au « flux de conscience » décrit par William James. Ces états d'activité ont été simulés dans des modèles semi-réalistes de réseaux neuronaux (Dehaene & Changeux, 2005 ; Dehaene, Kerszberg, & Changeux, 1998 ; Dehaene, Sergent, & Changeux, 2003).

### *Signatures expérimentales de l'accès à la conscience*

Afin de mettre à l'épreuve ces hypothèses théoriques, un nombre croissant d'expériences ont comparé le comportement et l'activité cérébrale, mesurée en IRM fonctionnelle et en potentiels évoqués, selon qu'un même stimulus est présenté au-delà ou en deçà du seuil de conscience. Ces recherches ont permis de découvrir plusieurs « signatures » de l'accès à la conscience, c'est-à-dire des mesures de l'état d'activité du cerveau qui distinguent le traitement conscient et non-conscient et signalent l'entrée d'une information dans la conscience perceptive.

*Processus précoces non-conscients.* Quelle que soit la modalité de présentation visuelle, auditive ou tactile, les résultats sont similaires : l'IRM démontre que les cortex sensoriels continuent de s'activer fortement, parfois à des niveaux strictement inchangés, lorsqu'un stimulus n'est pas perçu consciemment (Boly *et al.*, 2007 ; Dehaene *et al.*, 2001 ; Sadaghiani, Hesselmann & Kleinschmidt, 2009). Ainsi l'entrée d'une information dans le cortex, même à haut niveau, tant dans la voie visuelle ventrale que dorsale, ne constitue pas une condition suffisante d'un traitement conscient.

*Réseau préfrontal, pariétal et cingulaire.* Le contraste entre stimuli conscients et non-conscients, en IRM, fait apparaître un réseau distribué mais néanmoins restreint qui isole les régions le plus souvent bilatérales du cortex préfrontal dorsolatéral, pariétal inférieur et cingulaire antérieur (Dehaene *et al.*, 2001). Lau et Passingham (2006) introduisent un contrôle important en égalisant les performances cognitives dans les deux conditions, en s'appuyant sur la possibilité d'obtenir une performance meilleure que le hasard sans conscience. Ils observent une corrélation spécifique de l'accès à la conscience avec l'activité du cortex préfrontal dorsolatéral gauche. De même, dans une expérience récente, Van Gaal *et coll.* (2010) contrastent la performance consciente ou non-consciente d'une tâche d'inhibition « go-no go ». De petites régions spécialisées du cortex frontal (pré-SMA, insula antérieure) sont susceptibles de s'activer non-consciemment, mais la prise de conscience s'accompagne d'une amplification de ces réseaux et l'extension de l'activation à de vastes régions préfrontales et pariétales inférieures bilatérales.

*Décodage d'un contenu conscient.* John-Dylan Haynes (2009) fait observer que, pour isoler les mécanismes cérébraux de la conscience en imagerie cérébrale, on ne peut se contenter de détecter un signal présent lors des essais conscients et absent lors des essais non-conscients. Il faut également montrer que ce signal constitue un code neural de l'état de conscience du sujet, tel que l'on puisse, en principe, concevoir un algorithme de décodage qui, partant de l'activation observée, reconstitue précisément l'état de conscience du sujet. Le code observé doit (1) présenter le même degré d'invariance que le contenu conscient du sujet,

(2) préserver les relations de similarité subjective entre les contenus conscients, (3) ne pas coder des différences physiques que le sujet ne perçoit pas, mais (4) coder l'objet perçu, même lorsqu'il s'agit d'une perception illusoire. Seules quelques expériences récentes d'IRM répondent à cet exigeant programme dans la modalité visuelle (Schurger, Pereira, Treisman, & Cohen, 2009 ; Sterzer, Haynes, & Rees, 2008 ; Williams, Dang, & Kanwisher, 2007). Leurs résultats suggèrent que les aires rétiniotopiques les plus précoces ne covarient pas toujours avec le contenu subjectif, mais qu'une telle covariation s'observe dans des aires plus avancées telles que le cortex occipital latéral ou la région fusiforme. Les résultats sont compatibles avec l'existence d'un code neural partiellement activé par les stimuli invisibles, mais amplifié, étendu spatialement, et rendu plus reproductible d'essai en essai lors de l'accès à la conscience. L'absence d'analyse en cerveau entier limite toutefois l'importance de ces travaux.

Dans la modalité auditive, une intéressante expérience d'IRM porte sur l'illusion de McGurk, dans laquelle la perception d'une syllabe est influencée par la vision d'un locuteur (Hasson, Skipper, Nusbaum & Small, 2007). De façon remarquable, l'activité cérébrale qui est en corrélation avec le contenu perçu n'est pas issue des régions auditives précoces, mais d'un traitement plus avancé et multimodal qui mobilise conjointement des régions temporales et préfrontales.

*Décours temporel de l'accès à la conscience.* Les potentiels évoqués, la magnéto-encéphalographie et l'électrophysiologie chez l'animal permettent de préciser le point de divergence temporelle entre stimuli conscients et non-conscients. Dès les tout premiers travaux, il apparaît clairement que les premières étapes du traitement cortical peuvent être parfaitement préservées lors du traitement visuel non-conscient (travaux de Schiller et Chorover et de Bridgeman). Chez l'animal, dans le cortex visuel primaire comme dans le cortex inféro-temporal et les champs frontaux oculogyres, une amplification tardive des décharges neuronales est associée à la rapportabilité des stimuli et est interrompue lors du masquage (Kovacs, Vogels, & Orban, 1995 ; Super, Spekreijse, & Lamme, 2001 ; Thompson & Schall, 1999). Cependant, seules les études chez l'homme permettent à la fois d'enregistrer l'activité électro-magnétique du cerveau entier et de la corréler avec un authentique rapport subjectif de visibilité plutôt qu'une mesure comportementale. Les potentiels évoqués isolent une cascade d'étapes successives d'accumulation de l'évidence sensorielle, mais seule l'activité tardive, au-delà de 270 ms (onde P3), présente un profil de réponse étroitement corrélé à celui de la perception rapportée par les sujets (Del Cul, Baillet, & Dehaene 2007). Une modélisation des sources corticales montre que la non-linéarité tardive de l'onde P3 est associée à une activation distribuée soudaine, notamment de la région frontale inférieure, conjointement à une deuxième vague d'activité tardive dans les aires visuelles. Il faut noter que, même lorsque le stimulus est unilatéral, lorsqu'il est perçu consciemment, l'activité qu'il évoque envahit un vaste ensemble de régions corticales bilatérales. Ces résultats répliquent très précisément un travail antérieur mené sur la condition expérimentale de clignement attentionnel, qui présente par ailleurs l'avantage de présenter des stimuli rigoureusement identiques mais perçus tantôt consciemment,

tantôt inconsciemment (Sergent, Baillet & Dehaene, 2005). Dans cette étude, l'activation sensorielle est identique jusqu'à  $\sim 200$  ms, puis diverge brutalement d'une manière compatible avec l'hypothèse d'une activation soudaine et auto-amplifiée de régions distantes, préfrontales et pariétales, ainsi que d'une ré-activation des cortex sensoriels initialement activés. Le caractère soudain et pratiquement « tout-ou-rien » de l'activité évoquée par un stimulus conscient est confirmé par l'enregistrement de neurones uniques dans le cortex temporal médian de personnes épileptiques (Quiroga, Mukamel, Isham, Malach & Fried, 2008).

*Quatre signatures électrophysiologiques.* Dans un travail récent du laboratoire, Gaillard *et al.* (2009) comparent chez dix patients épileptiques le déroulement des enregistrements intracrâniens de potentiels locaux en réponse à un mot conscient ou non-conscient. Leurs résultats démontrent qu'au cours du masquage, toute une série d'étapes précoces continuent de se dérouler sans conscience. Vers 270-300 ms, on commence à observer une corrélation forte entre l'activité cérébrale (P3) et la visibilité des stimuli. Dans la même fenêtre temporelle, l'amplification des oscillations dans la bande gamma ( $> 30$  Hz), la mise en synchronie de régions distantes dans la bande beta (13-30 Hz), et l'établissement d'un dialogue causal bidirectionnel entre ces régions, mesurable par la causalité de Granger, apparaissent comme autant de signatures d'un état cortical synchrone et distribué qui apparaît exclusivement lorsque le seuil de conscience est franchi (voir également Fisch *et al.*, 2009 ; Gross *et al.*, 2004). L'accès à la conscience est donc un phénomène cortical tardif, global, soudain et synchrone. Particulièrement intéressant est le fait que près de 70 % des sites corticaux enregistrés présentent une modulation du signal liée à la présence d'un mot conscient (contre 24,4 % dans le cas non-conscient). Cette observation est compatible avec l'hypothèse que l'accès à la conscience s'accompagne d'une diffusion de l'information à de très vastes secteurs du cortex (*broadcasting* proposé par Baars 1989, Dehaene et Changeux, 1998).

#### *Applications cliniques aux patients non-communicants*

Notre compréhension croissante des mécanismes neuronaux qui sous-tendent la prise de conscience d'une information laisse espérer des applications cliniques.

Un premier champ d'application concerne l'hypothèse que certaines pathologies cérébrales affectent plus sélectivement les réseaux corticaux à longue distance et descendants (*top-down*) associés à l'accès conscient, tout en laissant largement intacts les systèmes corticaux modulaires responsables de traitements automatisés non-conscients. À l'exact opposé de la neuropsychologie classique, qui se focalise sur les dissociations fonctionnelles affectant des compétences spécifiques et résultant de lésions cérébrales focales, le modèle de l'espace de travail neuronal global prédit que des atteintes diffuses, soit des corps cellulaires des neurones à axones longs, soit des longs faisceaux de la substance blanche correspondante, doivent entraîner des déficits dans une grande variété de tâches avec effort, de synthèse consciente ou de contrôle exécutif, tout en laissant intactes le traitement spécialisé non-conscient. La schizophrénie et l'autisme peuvent être analysés sous

cet angle. En dépit de l'étendue de leurs troubles, les patients schizophrènes présentent une remarquable préservation de l'amorçage subliminal et d'autres phénomènes non-conscients ou implicites (Danion, Meulemans, Kauffmann-Muller & Vermaat, 2001 ; Dehaene, Artiges *et al.*, 2003). Par contre, leur seuil de perception consciente lors du masquage est anormalement élevé (Del Cul, Dehaene & Leboyer, 2006). De ce point de vue, la schizophrénie présente des parallèles étroits avec d'autres catégories de pathologies : d'une part, les patients atteints de sclérose en plaque au stade débutant, où l'étendue de la pathologie des faisceaux transcorticaux à longue distance prédit la sévérité du trouble de perception consciente (Reuter *et al.*, 2007 ; Reuter *et al.*, 2009) ; et d'autre part, les patients atteints de lésions cérébrales focales du lobe frontal, particulièrement lorsque le cortex frontal rostral est atteint (aire 10 de Brodmann), qui présentent également un trouble de perception subjective de stimuli masqués (Del Cul, Dehaene, Reyes, Bravo & Slachevsky, 2009).

Un second champ naturel d'application des recherches sur la conscience concerne la vaste catégorie des patients non-communicants : comas, états végétatifs, états de conscience minimale ou états d'enfermement (*locked-in*). Ces états restent méconnus et près de 40 % des patients *locked-in* seraient initialement mal détectés (d'après Monti *et coll.*, 2010). Chez ces patients, l'imagerie cérébrale des « signatures » de l'état conscient pourrait aider à établir des critères diagnostics plus solides que le simple examen clinique. Un premier test d'IRM, développé par le groupe d'Owen à Cambridge consiste à simuler mentalement pendant trente secondes, soit un service de tennis, soit une promenade dans un lieu connu. Sur 54 patients (23 en état végétatif, 31 minimalement conscients), cinq montrent des signes de contrôle volontaire de leur activité cérébrale comparables à ceux d'un volontaire sain. Or quatre d'entre eux étaient classifiés comme végétatifs et, même après l'IRM, un test intensif ne montrait de signes de conscience que chez deux d'entre eux. Ainsi, cette méthode fondée sur l'IRM fonctionnelle semble détecter des traces de conscience avec une sensibilité meilleure que l'examen clinique. Cependant le test, strictement empirique, ne s'appuie sur aucune théorie. Par ailleurs, l'IRM présente de nombreuses limites pratiques qui contraignent son utilisation répétée en clinique. Avec Lionel Naccache et Tristan Bekinschtein, nous avons développé un test fondé sur l'enregistrement des potentiels évoqués, moins coûteux et plus simple (Bekinschtein *et al.*, 2009). Notre méthode vise à détecter, chez les patients, l'onde corticale P3 tardive et globale que nos travaux théoriques et pratiques associent à l'accès d'une information à la conscience. Les premiers résultats suggèrent qu'il est effectivement possible, dans un test auditif très simple, de séparer les étapes non-conscientes et conscientes du traitement d'un stimulus et de les détecter en quelques minutes chez toute personne normalement vigilante. Ces travaux se poursuivent actuellement afin de vérifier la spécificité et la sensibilité de cette technique chez une grande série de patients comateux, végétatifs ou minimalement conscients.

À l'avenir, le développement des technologies de stimulation cérébrale profonde et d'interface cerveau-machine pourrait permettre non seulement de détecter

la présence d'une conscience résiduelle, mais également d'intervenir à la fois sur la vitesse de récupération et sur la restauration de la communication interpersonnelle chez les patients non-communicants (Schiff *et al.*, 2008). Au laboratoire, ce programme de recherches fait l'objet d'un financement obtenu en 2010 auprès du Conseil européen de la recherche (ERC).

### *Bibliographie succincte*

Le cours s'est appuyé sur plusieurs ouvrages et articles de revue :

Bernard Baars, *A cognitive theory of consciousness*, Cambridge University Press, 1989.

Baars B.J., Banks W.P., Newman J.B., *Essential sources in the scientific study of consciousness*, MIT Press, 2003.

Susan Blackmore, *Conversations on consciousness*, New York, Oxford University Press, 2006.

Jean-Pierre Changeux, *L'homme de vérité*. Paris, Odile Jacob, 2002.

Stanislas Dehaene (éd.), *The cognitive neuroscience of consciousness*, Cambridge, MIT Press, 2002.

Daniel Dennett, *Consciousness explained*, London, Penguin, 1991.

Julian Jaynes, *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, New York, Houghton Mifflin Company, 1976 ; Traduction (Guy de Montjou), *La Naissance de la conscience dans l'effondrement de l'esprit*, PUF, 1994.

Christoph Koch, *The Quest for Consciousness: A Neurobiological Approach*, Roberts & Co, 2004 ; traduction française, *À la recherche de la conscience*, Odile Jacob, 2006.

Lionel Naccache, *Le nouvel inconscient*, Odile Jacob, 2006.

Anil K. Seth, *Models of consciousness*. *Scholarpedia*, 2(1), 2007, 1328.

Daniel Wegner, *The illusion of conscious will*, Cambridge, MIT Press, 2003.

Lawrence Weiskrantz (éd.), *Frontiers of consciousness*. *Chichele Lectures*, Oxford, Oxford University Press, 2008.

### *Principaux travaux cités*

Bekinschtein T.A., Dehaene S., Rohaut B., Tadel F., Cohen L. & Naccache L., « Neural signature of the conscious processing of auditory regularities » *Proc. Natl. Acad. Sci. USA*, 106(5), 2009, 1672-1677.

Blanke O., Landis T., Spinelli L. & Seeck M., « Out-of-body experience and autoscapy of neurological origin », *Brain*, 127(Pt 2), 2004, 243-258.

Boly M., Balteau E., Schnakers C., Degueldre C., Moonen G., Luxen A. *et al.*, « Baseline brain activity fluctuations predict somatosensory perception in humans », *Proc. Natl. Acad. Sci. USA*, 104(29), 2007, 12187-12192.

Bressan P. & Pizzighello S., « The attentional cost of inattention blindness », *Cognition*, 106(1), 2008, 370-383.

Chun M.M. & Potter M.C., « A two-stage model for multiple target detection in rapid serial visual presentation », *J. Exp. Psychol. Hum. Percept. Perform.*, 21(1), 1995, 109-127.

Danion J.M., Meulemans T., Kauffmann-Muller F. & Vermaat H., « Intact implicit learning in schizophrenia », *Am. J. Psychiatry*, 158(6), 2001, 944-948.

De Gardelle V., Sackur J. & Kouider S., « Perceptual illusions in brief visual presentations », *Conscious Cogn.*, 18(3), 2009, 569-577.

Dehaene S., Artiges E., Naccache L., Martelli C., Viard A., Schurhoff F. *et al.*, « Conscious and subliminal conflicts in normal subjects and patients with schizophrenia: the role of the anterior cingulate », *Proc. Natl. Acad. Sci. USA*, 100(23), 2003, 13722-13727.

Dehaene S. & Changeux J.P., « Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattention blindness », *PLoS Biol.*, 3(5), 2005, e141.

Dehaene S., Changeux J.P., Naccache L., Sackur J. & Sergent C. 2006 « Conscious, preconscious, and subliminal processing: a testable taxonomy », *Trends Cogn. Sci.*, 10(5), 204-211.

Dehaene S., Kerszberg M. & Changeux J.P., « A neuronal model of a global workspace in effortful cognitive tasks », *Proc. Natl. Acad. Sci. USA*, 95(24), 1998, 14529-14534.

Dehaene S. & Naccache L., « Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework », *Cognition*, 79, 2001, 1-37.

Dehaene S., Naccache L., Cohen L., Bihan D.L., Mangin J.F., Poline J.B. *et al.*, « Cerebral mechanisms of word masking and unconscious repetition priming », *Nat. Neurosci.*, 4(7), 2001, 752-758.

Dehaene S., Sergent C. & Changeux J.P., « A neuronal network model linking subjective reports and objective physiological data during conscious perception », *Proc. Natl. Acad. Sci. USA*, 100, 2003, 8520-8525.

Del Cul A., Baillet S. & Dehaene S., « Brain dynamics underlying the nonlinear threshold for access to consciousness », *PLoS Biol.*, 5(10), 2007, e260.

Del Cul A., Dehaene S. & Leboyer M. « Preserved subliminal processing and impaired conscious access in schizophrenia », *Arch. Gen. Psychiatry*, 63(12), 2006, 1313-1323.

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

Fisch L., Privman E., Ramot M., Harel M., Nir Y., Kipervasser S. *et al.*, « Neural "Ignition": Enhanced Activation Linked to Perceptual Awareness in Human Ventral Stream Visual Cortex », *Neuron*, 64, 2009, 562-574.

Gaillard R., Dehaene S., Adam C., Clemenceau S., Hasboun D., Baulac M. *et al.*, « Converging intracranial markers of conscious access », *PLoS Biol.*, 7(3), 2009, e61.

Gross J., Schmitz F., Schnitzler I., Kessler K., Shapiro K., Hommel B. *et al.*, « Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans », *Proc. Natl. Acad. Sci. USA*, 101(35), 2004, 13050-13055.

Hasson U., Skipper J.I., Nusbaum H.C. & Small S.L., « Abstract coding of audiovisual speech: beyond sensory representation », *Neuron*, 56(6), 2007, 1116-1126.

Haynes J.D. « Decoding visual consciousness from human brain signals », *Trends Cogn. Sci.*, 13, 2009, 194-202.

Huang L., Treisman A. & Pashler H., « Characterizing the limits of human visual awareness », *Science*, 317(5839), 2007, 823-825.

Kovacs G., Vogels R. & Orban G.A., « Cortical correlate of pattern backward masking », *Proc. Natl. Acad. Sci. USA*, 92(12), 1995, 5587-5591.

Lau H.C. & Passingham R.E., « Relative blindsight in normal observers and the neural correlate of visual consciousness », *Proc. Natl. Acad. Sci. USA*, 103(49), 2006, 18763-18768.

Mack A. & Rock I., *Inattention blindness*, Cambridge, Mass., MIT Press, 1998.

Naccache L., Blandin E. & Dehaene S., « Unconscious masked priming depends on temporal attention », *Psychological Science*, 13, 2002, 416-424.

Pashler H. « Dual-task interference in simple tasks: data and theory », *Psychol. Bull.*, 116(2), 1994, 220-244.

Quiroga R.Q., Mukamel R., Isham E.A., Malach R. & Fried I., « Human single-neuron responses at the threshold of conscious recognition », *Proc. Natl. Acad. Sci. USA*, 105(9), 2008, 3599-3604.

Reuter F., Del Cul A., Audoin B., Malikova I., Naccache L., Ranjeva J.P. *et al.*, « Intact subliminal processing and delayed conscious access in multiple sclerosis », *Neuropsychologia*, 2007.

Reuter F., Del Cul A., Malikova I., Naccache L., Confort-Gouny S., Cohen L. *et al.*, « White matter damage impairs access to consciousness in multiple sclerosis », *Neuroimage*, 44(2), 2009, 590-599.

Sadaghiani S., Hesselmann G. & Kleinschmidt A., « Distributed and antagonistic contributions of ongoing activity fluctuations to auditory stimulus detection », *J. Neurosci.*, 29(42), 2009, 13410-13417.

Schiff N.D., Giacino J.T., Kalmar K., Victor J.D., Baker K., Gerber, M. *et al.*, « Behavioural improvements with thalamic stimulation after severe traumatic brain injury », *Nature*, 452(7183), 2008, 120.

Schurger A., Pereira F., Treisman A. & Cohen J.D., « Reproducibility distinguishes conscious from nonconscious neural representations », *Science*, 327(5961), 2009, 97-99.

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.

Simons D.J. & Ambinder M.S., « Change blindness: theory and consequences », *Curr. Dir. Psychol. Science*, 14(1), 2005, 44-48.

Sterzer P., Haynes J.D. & Rees G., « Fine-scale activity patterns in high-level visual areas encode the category of invisible objects », *J. Vis.*, 8(15), 2008, 10, 11-12.

Super H., Spekreijse H. & Lamme V.A., « Two distinct modes of sensory processing observed in monkey primary visual cortex (V1) », *Nat. Neurosci.*, 4(3), 2001, 304-310.

Thompson K.G. & Schall J.D., « The detection of visual signals by macaque frontal eye field during masking », *Nature Neuroscience*, 2, 1999, 283-288.

Van Gaal S., Ridderinkhof K.R., Scholte H.S. & Lamme V.A., « Unconscious Activation of the Prefrontal No-Go Network », *J. Neurosci.*, 2010, sous presse.

Weiskrantz L., *Consciousness lost and found: A neuropsychological exploration*. New York, Oxford University Press, 1997.

Williams M.A., Dang S. & Kanwisher N.G., « Only some spatial patterns of fMRI response are read out in task performance », *Nat. Neurosci.*, 10(6), 2007, 685-686.

Woodman G.F. & Luck S.J., « Dissociations among attention, perception, and awareness during object-substitution masking », *Psychol. Sci.*, 14(6), 2003, 605-611.

Wyart V. & Tallon-Baudry C., « Neural dissociation between visual awareness and spatial attention », *J. Neurosci.*, 28(10), 2008, 2667-2679.

### **Séminaire : Mécanismes de l'apprentissage et applications à l'éducation**

Dans la lignée des séminaires antérieurs consacrés à l'apprentissage de l'arithmétique et de la lecture, le séminaire 2010 a été consacré aux avancées récentes dans le domaine des mécanismes de l'apprentissage chez l'enfant et l'adulte, et à leurs conséquences pour l'éducation et la rééducation. Comment optimiser l'enseignement afin de maximiser la rétention en mémoire ? Faut-il préférer un enseignement strictement visuel de la lecture, ou plutôt un apprentissage multi-sensoriel qui mette également en valeur le toucher et le geste d'écriture ? Comment fonctionne l'apprentissage implicite, par exemple celui de l'orthographe, et quelles en sont les limites ? Quelles statistiques l'enfant compile-t-il pour

apprendre sa langue écrite ou orale ? L'apprentissage est-il modulé par le contexte de communication sociale et les signaux qu'envoie l'enseignant ? Le sommeil joue-t-il réellement un rôle important dans la consolidation des apprentissages ? Ces diverses questions ont été abordées à travers les exposés de sept spécialistes de l'apprentissage et du développement cognitif :

- Henry Roediger (Washington University, Saint Louis, USA) : « Mechanisms of memory: Enhancing retention through repeated retrieval » ;
- Édouard Gentaz (CNRS, Grenoble) : « L'importance de l'exploration multisensorielle dans les premiers apprentissages scolaires élémentaires » ;
- Sébastien Pacton (CNRS, Paris) : « L'apprentissage implicite du langage écrit » ;
- Axel Cleeremans (Bruxelles) : « Impact et limites de l'apprentissage implicite » ;
- Gyorgy Gergely (Budapest) : « Natural Pedagogy: The role of communicative cues in the transmission of generic knowledge in human infants » ;
- Jacques Mehler (Trieste, Italie) : « Le rôle des statistiques dans l'apprentissage du langage » ;
- Jan Born (Université de Luebeck, Allemagne) : « The memory function of sleep ».

#### ACTIVITÉS DE RECHERCHE DU LABORATOIRE

Nous mettons ici en valeur quelques résultats qui nous paraissent importants. Vient ensuite une liste complète des publications du laboratoire.

#### **Codage et décodage des nombres dans la région intrapariétale**

Les travaux antérieurs du laboratoire ont démontré que le sillon intrapariétal bilatéral s'active systématiquement lors du calcul mental. Dès 1993, Jean-Pierre Changeux et moi-même avons émis l'hypothèse que cette région pourrait abriter un code neuronal des quantités numériques. En 2003, Andreas Nieder, Earl Miller et leurs collaborateurs identifiaient, chez le singe macaque, des neurones accordés aux quantités numériques dans la profondeur du sillon intrapariétal, dans une région qui constitue un homologue potentiel de la région activée dans le cerveau humain. Toutefois, jusqu'à présent, nous ne disposions que de données indirectes concernant l'existence d'un code similaire dans le cerveau humain. En 2004, Manuela Piazza et moi-même avons utilisé l'IRM fonctionnelle pour montrer que l'adaptation du cerveau humain à une quantité fixe d'objets suit une loi log-Gaussienne étroitement liée aux courbes d'accord neuronal observées chez le singe. Ces données, ainsi que leur extension aux chiffres arabes en 2007, suggéraient fortement que la région intrapariétale humaine devait contenir un code neuronal du nombre perçu ou représenté. Cependant, il demeurerait possible d'interpréter certains de ces résultats en termes d'attention, d'effort ou de mémoire de travail.

Les nouvelles recherches menées par Evelyn Eger ont donc visé à établir directement l'existence d'un code neural du nombre, chez l'homme, à l'aide de l'IRM fonctionnelle de haute résolution. Des images de haute résolution spatiale couvrant les régions dorsales bilatérales du cortex pariétal et frontal ont été acquises alors que la personne volontaire réalisait une tâche de mémorisation à court terme de nombres présentés, soit sous forme d'un ensemble de points (expérience 1), soit sous forme d'un chiffre arabe (expérience 2). Les cartes locales d'activation ont ensuite été soumises à un algorithme de classification multivariée qui apprenait à détecter si les configurations d'activation observées permettaient de discriminer entre les nombres mémorisés. Effectivement, après entraînement, il était possible de repérer, dans les images d'IRM, des informations subtiles qui indiquaient si la personne avait reçu 4, 8, 16 ou 32 points à mémoriser, et de généraliser cet apprentissage à des images nouvelles. Conformément à nos prédictions, l'information numérique se trouvait concentrée sur les flancs du sillon intrapariétal, comme chez le singe macaque. De plus, il était également possible d'utiliser les images d'IRM pour distinguer quels chiffres arabes (2, 4, 6, ou 8) la personne avait mémorisé – et le classificateur qui avait appris à distinguer ces chiffres généralisait ses performances aux ensembles concrets de même cardinal.

Ces résultats établissent de façon directe que la région intrapariétale humaine contient un code neural du nombre, et que celui-ci fait appel à un secteur cortical suffisamment large pour être lisible par IRM fonctionnelle. Il importe de souligner que ces résultats n'ont qu'une portée théorique et non pratique, car ils ne peuvent s'obtenir que dans des conditions extrêmement contraintes, avec la collaboration du sujet, et avec des performances qui ne s'écartent que faiblement du niveau du hasard (~ 70 % de réussite pour les ensembles concrets, ~ 57 % pour les chiffres arabes, là où le niveau du hasard est de 50 %).

L'intérêt principal de cette étude est d'établir qu'il existe un code neural conjoint pour les nombres présentés sous forme symbolique (les chiffres arabes) et non-symbolique (les ensembles d'objets). Ce résultat conforte une des hypothèses de la thèse du recyclage neuronal, selon laquelle la capacité humaine de manipulation des symboles fait appel, au moins en partie, au recrutement de circuits neuronaux anciens dans l'évolution, également présents chez les autres primates, et codant pour des objets mentaux suffisamment proches pour être reconvertis à ce nouvel usage.

### **Les mécanismes cérébraux de la lecture de l'écriture manuscrite**

Le laboratoire s'est également intéressé aux mécanismes cérébraux de l'identification des mots. La grande majorité des travaux antérieurs dans ce domaine porte sur l'identification des mots imprimés. Or, l'apprentissage de la lecture implique avant tout la maîtrise de l'écriture manuscrite, qui pose des problèmes particuliers. Les caractères manuscrits ne sont pas standardisés comme le sont les polices de caractères, et le cerveau du lecteur doit donc résoudre un difficile problème d'invariance perceptive.

Avec Émilie Qiao et Laurent Cohen, nous avons commencé par quantifier la difficulté de la reconnaissance de plusieurs écritures manuscrites. Nous avons montré que le temps de lecture constituait une bonne mesure de cette difficulté, surtout lorsque l'on prenait en considération sa variation en fonction de la longueur des mots. Les écritures manuscrites difficiles se caractérisent en effet par un ralentissement massif de la lecture accompagné d'un effet de longueur. Contrairement à ce que l'on observe pour des polices de caractères standard et des écritures manuscrites très lisibles, le temps de lecture d'un mot écrit dans une écriture manuscrite difficile à lire est proportionnel au nombre de lettres présentes dans le mot, ce qui indique un déploiement sériel de l'attention visuelle afin d'identifier certaines lettres ou groupes de lettres.

Émilie Qiao a ensuite utilisé l'IRM fonctionnelle afin d'examiner les mécanismes cérébraux de la reconnaissance de l'écriture manuscrite. Les résultats ont démontré que cette compétence recrutait particulièrement la région occipito-temporale droite, en complément de la région occipito-temporale gauche que nos travaux antérieurs impliquaient déjà dans l'identification visuelle des mots imprimés. De plus, dès que l'écriture devenait difficile à identifier, on observait une activation supplémentaire des régions pariétales et précentrales bilatérales systématiquement impliquées dans l'effort attentionnel.

Au sein de ce réseau, l'utilisation de l'amorçage par répétition nous a permis de repérer les aires responsables de l'invariance perceptive. Seule la région de l'« aire de la forme visuelle des mots », dans le sillon occipito-temporal latéral de l'hémisphère gauche, présentait une réduction d'activation lorsque le même mot était présenté deux fois, et ce même lorsque le format de présentation différait, un exemplaire du mot étant présenté en lettres d'imprimerie et l'autre en caractères manuscrits. Cet effet suggère que cette aire contient un code neuronal remarquablement invariant, capable de normaliser pour les variations considérables qui distinguent les formes manuscrites et imprimées du même mot.

### **L'origine des erreurs en miroir au cours de la lecture**

Avec Kimihiro Nakamura, nous avons étudié un autre aspect de la lecture : les erreurs de lecture et d'écriture en miroir. Comme je l'ai souligné dans le chapitre 9 des *Neurones de la lecture* (O. Jacob, 2007), il est remarquable que les jeunes enfants, au cours des premières étapes de l'apprentissage de la lecture, fassent fréquemment des erreurs en miroir, écrivant par exemple leur nom de la droite vers la gauche sans s'en rendre compte. La compétence de lecture et d'écriture en miroir n'est jamais enseignée explicitement, et à l'âge adulte les lecteurs experts ne la possèdent plus, alors pourquoi apparaît-elle spontanément chez la majorité des jeunes enfants ? La théorie du recyclage neuronal fournit une hypothèse : la reconnaissance des objets quelque soit leur orientation dans l'espace serait une propriété fondamentale du système de reconnaissance visuelle invariante de tous les primates. L'espèce humaine hériterait donc de cette capacité, mais celle-ci

devrait être « désapprise » afin de traiter avec efficacité l'alphabet latin, qui exige de distinguer des paires minimales telles que b/d ou p/q.

Selon cette hypothèse, la capacité de reconnaissance en miroir et sa disparition pour les lettres et les mots constituent donc des indices importants de l'origine évolutive de la lecture. On pourrait les qualifier d'indices vestigiaux dans la mesure où ils traduisent l'état initial d'un système visuel qui n'a jamais évolué *pour* la lecture, mais plutôt pour la reconnaissance invariante d'objets dans un monde naturel où la distinction droite-gauche n'est pas essentielle à la survie.

Nous avons donc examiné, à l'aide de l'IRM fonctionnelle, les mécanismes cérébraux de l'invariance perceptive en miroir pour les objets, et de son absence pour les mots. Afin d'examiner l'invariance cross-culturelle de ces mécanismes, treize adultes français et treize adultes japonais, tous excellents lecteurs, ont été scannés alors qu'ils effectuaient une classification sémantique sur des mots écrits (dans leur langue) et sur des dessins au traits d'animaux et d'objets. Les stimuli étaient présentés par paires, et pouvaient être répétés à l'identique, répétés en miroir, ou différents. Nous avons ainsi répliqué l'existence d'un effet de répétition pour les images dans l'ensemble du cortex visuel ventral bilatéral, et pour les mots spécifiquement dans le sillon occipito-temporal latéral gauche (aire de la forme visuelle des mots). Crucialement, pour les mots répétés en miroir, cette région ne montrait aucun effet de réduction d'activation, alors qu'elle présentait un effet clair d'invariance en miroir pour les dessins. Ainsi, dans la même région existe une invariance en miroir pour les images mais pas pour les mots. Ces résultats sont compatibles avec l'hypothèse d'un désapprentissage de cette compétence à mesure que l'on apprend à lire. Plus remarquable encore, de tout le cortex inféro-temporal ventral, cette région gauche est celle qui présente l'invariance perceptive la plus prononcée pour les images en miroir. Ainsi n'est-il peut-être pas étonnant que les enfants confondent initialement les lettres et leurs images en miroir : pour des raisons que nous ne comprenons encore qu'imparfaitement, l'apprentissage de la lecture recycle une aire cérébrale qui est précisément celle qui présente, au départ, l'invariance en miroir la plus forte.

### **La vitesse d'intégration des lettres dans les mots**

Avec Joaquim Forget, nous avons étudié un dernier aspect de la lecture : l'intégration rapide et parallèle des lettres au sein des mots. En effet, notre système visuel présente une compétence remarquable : pour les petits mots (jusqu'à environ 8 lettres), il détermine la position relative des lettres et leurs combinaisons en graphèmes, syllabes et morphèmes, avec une vitesse élevée et essentiellement indépendante de leur nombre. Cette compétence permet, par exemple, de distinguer rapidement un mot de ses anagrammes (lapin/alpin).

Afin d'étudier cette compétence, nous avons conçu un paradigme original, qui consiste à présenter sur un écran d'ordinateur alternativement les lettres paires et impaires d'un mot, à leur position normale (par exemple C M A N alterne avec P G N E). Si l'alternance est suffisamment rapide, la perception est celle d'un mot unique (ici

CHAMPAGNE), alors que, dans le cas contraire, les fragments de lettres sont perçus mais ne sont pas intégrés en un tout. Nos expériences ont démontré l'existence d'une fenêtre d'intégration temporelle non-linéaire de l'ordre de 80 millisecondes. En deçà de cette limite, il est facile de fusionner le stimulus oscillatoire en un percept unique, et la perception du sujet présente toutes les caractéristiques d'une lecture normale. Au-delà, par contre, les temps de lecture présentent un ralentissement soudain et linéaire en fonction du nombre de lettres. L'enregistrement des potentiels évoqués démontre que, même à la fréquence la plus élevée, l'intégration n'a pas lieu à bas niveau, réfutant l'idée que ces effets sont dus à la « persistance rétinienne ». Au contraire, les fragments alternés continuent d'évoquer des oscillations synchrones dans les cortex visuels postérieurs, et seuls les composantes tardives des potentiels évoqués, sensibles au statut lexical du stimulus, disparaissent en coïncidence avec le seuil de fusion perceptive. Les résultats imposent une borne supérieure au processus d'intégration temporelle des lettres, qui se déroule en moins de 300 millisecondes. L'aire de la forme visuelle des mots est l'un des candidats plausibles pour le site de cette intégration.

Dans la mesure où ils montrent qu'un mot reste parfaitement visible même lorsque ses lettres apparaissent avec des délais de l'ordre de 50 millisecondes, nos données contraignent également les modèles du processus de reconnaissance visuelle. Les théories qui postulent que le liage perceptif des lettres survient par le biais d'un codage temporel ou d'oscillations neuronales dans la bande gamma (> 30 Hz) ne sont guère compatibles avec la lenteur de l'intégration observée, ni avec l'absence d'impact important de cette asynchronie du stimulus sur le comportement de lecture.

### **Mise en évidence d'une limite de l'introspection**

Dans le domaine de l'accès à la conscience, le laboratoire a continué ses investigations de l'introspection et de ses limites. Sébastien Marti a ainsi repris le paradigme d'introspection du temps de réaction initialement étudié par Guido Corallo, Jérôme Sackur, Mariano Sigman et moi-même. La question centrale que pose ce travail est celle des limites de l'introspection. Depuis la révolution behavioriste, les psychologues refusent de considérer l'introspection comme une mesure appropriée des processus mentaux. Cependant, les sujets humains fournissent spontanément des descriptions introspectives détaillées des étapes de leur réflexion, par exemple dans le jeu d'échec. Ces rapports introspectifs sont-ils vraiment systématiquement erronés ? Sur la base du modèle de l'espace de travail neuronal conscient, nous faisons l'hypothèse que l'introspection n'est pas entièrement dépourvue d'à-propos, mais qu'elle est sévèrement limitée à la seule description de l'étape de traitement sériel qui, à un instant donné, occupe l'espace conscient.

Nous avons donc demandé aux volontaires, à chaque essai, de réaliser deux tâches en succession rapide, puis de répondre à un questionnaire introspectif détaillé. Immédiatement après leurs réponses objectives, les sujets répondaient à pas de moins de quatre questions subjectives qui portaient sur le déroulement temporel

des événements sensoriels, moteurs et décisionnels, survenus au cours de l'essai. Sur la base de ces introspections, nous avons proposé une méthode nouvelle qui permet de reconstruire la phénoménologie d'un essai et de la comparer aux mesures objectives et aux prédictions du modèle standard de l'interférence centrale lors des tâches doubles. Les introspections étaient étroitement corrélées au déroulement réel de la tâche, à l'exception d'une distorsion subjective majeure : les sujets n'avaient pas conscience du fait que le stimulus de la seconde tâche devait parfois attendre longuement que l'étape centrale de la première tâche soit terminée avant d'être traité (effet dit de « période psychologique réfractaire »).

Ainsi, il apparaît que la perception consciente est systématiquement différée et subit des distorsions profondes pendant que les ressources centrales sont monopolisées par une autre tâche. En d'autres termes, notre introspection n'a aucun accès à l'étendue des traitements parallèles qui caractérisent le traitement non conscient de l'information dans le cerveau. Par contre, nous disposons d'une introspection véridique, détaillée et sans coût apparent sur le déroulement de l'opération mentale qui, à un instant donné, occupe notre espace de travail conscient.

### **Accumulation d'évidence et prise de décision**

Floris de Lange, lors de son stage post-doctoral, a introduit un autre paradigme qui permet d'étudier la sérialité des opérations mentales. Nos résultats antérieurs montrent que l'accumulation de données issues d'un stimulus peut se produire sans que ce stimulus soit perçu. Nous nous sommes alors demandé comment les volontaires se comporteraient dans une tâche qui demande d'accumuler séquentiellement des données issues de cinq percepts distincts. Au cours d'un enregistrement de magnéto-encéphalographie, nous leurs présentions successivement une série de flèches tournées aléatoirement vers la droite ou vers la gauche. Il pouvait y avoir entre une et cinq flèches successives, et le sujet devait répondre en appuyant sur un bouton du côté indiqué par la majorité des flèches. Par exemple la séquence G(auche)-D(droite)-G-D-D conduisait à une réponse à main droite. Une modélisation mathématique démontra que plusieurs stratégies pouvaient être adoptées pour résoudre cette tâche. Si chaque flèche est correctement perçue, alors la stratégie optimale consiste à arrêter sa décision dès que trois flèches pointant du même côté ont été vues. Cependant, en présence d'une incertitude sur l'identité des flèches, la stratégie optimale consiste à accumuler les informations sur le nombre de flèches pointant à droite et à gauche.

Les données de magnéto-encéphalographie ont validé la notion d'une accumulation continue des informations sensorielles, mais avec quelques aspects inattendus. Dans les régions pariétales et centrales du cortex, les signaux neuronaux étaient inversement reliés à la quantité totale d'information accumulée, c'est-à-dire que plus le nombre de flèches était grand, plus l'activité était faible. De plus, il n'y avait pas trace d'un raisonnement logique qui stoppe la réflexion au-delà du seuil de trois flèches. Au contraire, de façon en partie irrationnelle, la dernière flèche aperçue contraignait plus fortement le choix que les précédentes. Nous avons

récemment obtenu des données supplémentaires, non encore publiées, qui suggèrent que cette accumulation d'informations constitue un mode opératoire automatique et qui peut même se déployer de façon non consciente.

Ces résultats indiquent que, pour prendre des décisions sur la base d'informations présentées séquentiellement, le système nerveux humain dispose d'un système d'accumulation capable d'intégrer les données à travers le temps sur des durées de l'ordre de quelques secondes, et qui modifie la pondération des données sensorielles nouvelles.

#### LISTE COMPLÈTE DES PUBLICATIONS (2009-2010)

##### Articles originaux

Bekinstein T., Dehaene S., Cohen L., Naccache L., « Neural signature of the conscious processing of auditory regularities », *Proceedings of the National Academy of Sciences USA*, 106, 2009, 1672-1677.

Gaillard R., Dehaene S., Adam C., Clermenceau S., Hasboun D., Baulac M., Cohen L., Naccache L., « Converging intracranial markers of conscious access », *PLoS Biology*, 7, 2009, e1000061.

Zylberberg A., Dehaene S., Mindlin G.B., Sigman M., « Neurophysiological bases of exponential sensory decay and top-down memory retrieval: a model », *Frontiers in Computational Neurosciences*, 3, 2009, 4.

Devauchelle A.D., Oppenheim C., Rizzi L., Dehaene S., Pallier C., « Sentence syntax and content in the human temporal lobe », *Journal of Cognitive Neuroscience*, 21, 2009, 1000-12.

Kouider S., Dehaene S., « Subliminal number priming within and across the visual and auditory modalities », *Experimental Psychology*, 56, 2009, 418-33.

Pinel P., Dehaene S., « Beyond hemispheric dominance: Brain regions underlying the joint lateralization of language and arithmetic to the left hemisphere », *Journal of Cognitive Neuroscience*, 0, 2009, 1-19.

Szwed M., Cohen L., Qiao E., Dehaene S., « The role of invariant line junctions in object and visual word recognition », *Vision research*, 49, 2009, 718-725.

Dehaene S., « Origins of mathematical intuitions: the case of arithmetic », *Annals NY Acad. Sci.*, 1156, 2009, 232-59.

Knops A., Thirion B., Hubbard E.M., Michel V., Dehaene S., « Recruitment of an area involved in eye movements during mental arithmetic », *Science*, 324, 2009, 1583-1585.

Knops A., Viarouge A., Dehaene S., « Dynamic representations underlying symbolic and nonsymbolic calculation: Evidence from the operational momentum effect », *Attention, Perception, & Psychophysics*, 71, 2009, 803-821.

Reuter F., Del Cul A., Malikova I., Naccache L., Confort-Gouny S., Cohen L., Cherif A.A., Cozzone P.J., Pelletier J., Ranjeva J.P., Dehaene S., Audoin B., « White matter damage impairs access to consciousness in multiple sclerosis », *NeuroImage*, 44, 2009, 590-599.

Sackur J., Dehaene S., « The cognitive architecture for chaining of two mental operations », *Cognition*, 111, 2009, 187-211.

Del Cul A., Dehaene S., Reyes P., Bravo E., Slachevsky A., « Causal role of prefrontal cortex in the threshold for access to consciousness », *Brain*, 132, 2009, 2531-40.

Nieder A., Dehaene S., « Representation of number in the brain », *Annual Rev. Neurosci.*, 32, 2009, 185-208.

Forget J., Buiatti M. et Dehaene S., « Temporal Integration in Visual Word Recognition », *J. Cogn. Neurosci.*, 0, 2009, 1-15.

Ranzini M., Dehaene S., Piazza M., Hubbard E.M., « Neural mechanisms of attentional shifts due to irrelevant spatial and numerical cues », *Neuropsychologia*, 47, 2009, 2615-24.

Hubbard E., Ranzini M., Piazza M., Dehaene S., « What Information is Critical to Elicit Interference in Number-Form Synaesthesia? », *Cortex*, 45, 2009, 1200-16.

Eger E., Michel V., Thirion B., Amadon A., Dehaene S., Kleinschmidt A., « Deciphering Cortical Number Coding from Human Brain Activity Patterns », *Current Biology*, 19, 2009, 1608-15.

Dehaene-Lambertz G., Montavont A., Jobert A., Alliroi L., Dubois J., Hertz-Pannier L., Dehaene S., « Language or Music, Mother or Mozart? Structural and environmental influences on infants' language networks », *Brain and Language*, 26 oct 2009.

Dehaene S., Dehaene-Lambertz G., « Neuro-imagerie cognitive: phylogénèse et ontogénèse » *Bull. Acad. Natl. Méd.*, 193(4), 2009, 883-89.

Dehaene S., Nakamura K., Jobert A., Kuroki C., Ogawa S., Cohen L., « Why do children make mirror errors in reading? Neural correlates of mirror invariance in the visual word form area », *NeuroImage*, 49, 2010, 1837-48.

Qiao E., Vinckier F., Szwed M., Naccache L., Valabrègue R., Dehaene S., Cohen L., « Unconsciously deciphering handwriting: Subliminal invariance for handwritten words in the visual word form area », *NeuroImage*, 49(2), 2010, 1786-99.

Kouider S., de Gardelle V., Dehaene S., Dupoux E., Pallier C., « Cerebral bases of subliminal speech priming », *Neuroimage*, 49(1), 2010, 922-929.

Marti S., Sackur J., Sigman M., Dehaene S., « Mapping the introspection's blind spot: Reconstruction of dual-task phenomenology using quantified introspection », *Cognition*, 115, 2010, 303-13.

Rasanen P., Salminen J., Wilson A., Aunio P., Dehaene S., « Computer-assisted intervention for children with low numeracy skills », *Cognitive Development*, 24, 2009, 450-72.

De Lange F., Jensen O., Dehaene S., « Accumulation of Evidence during Sequential Decision Making: the Importance of Top-down Factors », *Journal of Neuroscience*, 30, 2010, 731-738.

Rusconi E., Pinel P., Dehaene S., Kleinschmidt A., « The enigma of Gerstmann's syndrome revisited: a telling tale of the vicissitudes of neuropsychology », *Brain*, 133(Pt 2), 2010, 320-32.

Berteletti I., Lucangeli D., Piazza M., Dehaene S. and Zorzi M., « Numerical Estimation in Preschoolers », *Developmental Psychology*, 2009, sous presse.

Viarouge A., Hubbard E.M., Dehaene S., Sackur J., « Number Line Compression and the Illusory Perception of Random Numbers », *Exp. Psychol.*, 2010, sous presse.

## Livres

Dehaene S., *Reading in the brain*, Penguin Viking, novembre 2009.

## Chapitres de livre

Hubbard E.M., Piazza M., Pinel P., Dehaene S., « Numerical and spatial intuitions: A role for posterior parietal cortex? » in Tommasi L., Nadel L. et Peterson M.A. (éds.), *Cognitive Biology: Evolutionary and Developmental Perspectives on Mind, Brain and Behavior*. Cambridge, MA, MIT Press, 2009, 221-246.

**Revue, commentaires, diffusion des connaissances**

Dehaene S., Izard V., Pica P., & Spelke E., « Response to Comment “Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures” », *Science*, 323, 38c, 2009.

Dehaene S., « The case for a notation-independent representation of number », *Behav Brain Sci.* 32(3-4) : 333-5, discussion 356-73, 2009.

**HONNEURS ET DISTINCTIONS**

Stanislas Dehaene a été élu :

- membre étranger de la *National Academy of Sciences* des États-Unis ;
- membre de l'*Academia Europaea* ;
- *corresponding fellow* de la *British Academy* ;
- membre de l'*American Philosophical Society*.