Stanislas Dehaene

Le Code de la conscience

Odile Jacob
sciences

Stanislas
Dehaene

Le Code de la conscience

Odile Jacob
sciences

STANISLAS DEHAENE

Le Code de la conscience

© STANISLAS DEHAENE, 2009-2013.

Pour l'édition française :
© ODILE JACOB, OCTOBRE 2014.
15, RUE SOUFFLOT, 75005 PARIS

www.odilejacob.fr

ISBN : 978-2-7381-6904-4

Le code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5 et 3 a, d'une part, que les « copies ou reproductions strictement réservées à l'usage du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou réproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4). Cette représentation ou reproduction donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

Ce document numérique a été réalisé par Nord Compo.

À mes parents.

Et à Ann et Dan, mes parents d'Amérique.

« La conscience est la seule chose réelle au monde et de tous le plus grand mystère. »

Vladimir NABOKOV.

« Le cerveau est plus spacieux que le ciel
Car, mettez-les côte à côte,
L'un sans peine contient l'autre
Et vous de surcroît. »

Emily DICKINSON.

INTRODUCTION

La matière de la pensée

Au fond de la grotte de Lascaux, après la célèbre salle des Taureaux et sa ménagerie paléolithique, le visiteur s'engage dans un étroit couloir appelé l'Abside. C'est là, au fond d'un puits de cinq mètres de profondeur, aux côtés d'un bison blessé et d'un rhinocéros, que l'on trouve l'une des rares représentations du corps humain dans l'art pariétal (figure 1). L'homme est étendu sur le dos, les bras écartés, paumes tournées vers le haut. Près de lui, un oiseau s'est perché sur un bâton. Plus loin, une lance brisée vient d'éventrer un énorme bison dont les intestins se répandent.

Un détail ne laisse aucun doute sur le sexe de la personne représentée : son pénis est en érection. Et ce détail, selon le spécialiste français du sommeil Michel Jouvet[1], éclaire toute la signification de ce mystérieux panneau : celui-ci dépeindrait un rêveur et le contenu de son rêve – une fantastique victoire à la chasse. Comme l'ont découvert Jouvet et son équipe, le rêve survient principalement au cours d'une phase précise du sommeil, qu'ils ont nommée « sommeil paradoxal » parce qu'elle ne ressemble guère à l'image que l'on se fait du repos. Durant cette phase, en effet, le cerveau est presque aussi actif qu'à l'état de veille, et les yeux ne cessent de se mouvoir. De plus, cette phase s'accompagne systématiquement d'une forte érection chez les hommes (même si le rêve est dépourvu de tout contenu sexuel). Bien que cette curieuse propriété

de la physiologie humaine n'ait été découverte qu'au XX^e^ siècle, Jouvet fait plaisamment remarquer qu'elle ne risquait pas d'échapper à nos ancêtres. Quant à l'oiseau, il serait la métaphore naturelle de l'âme du rêveur, qui s'échappe en songe et s'envole vers des temps et des lieux lointains, aussi libre qu'une hirondelle.

Figure 1. L'esprit s'envole tandis que le corps reste inerte. Dans cette peinture rupestre de Lascaux, datée d'environ 18 000 ans, un homme est allongé, probablement endormi et plongé dans un rêve. Il présente en effet une érection caractéristique de la phase de sommeil paradoxal, au cours de laquelle les rêves sont les plus vifs. À ses côtés se trouvent un bison éventré et un oiseau. Selon Michel Jouvet, spécialiste du sommeil, il pourrait s'agir d'une des premières représentations d'un rêveur et du contenu de son rêve. Dans de nombreuses cultures, l'oiseau symbolise l'envol de l'esprit dans le sommeil ou après la mort. Le rêve est-il en partie responsable de l'intuition du dualisme, cette idée fausse selon laquelle l'esprit et le corps relèvent de deux domaines bien distincts ?

L'interprétation pourrait paraître fantaisiste si l'on ne retrouvait pas, dans l'art et le symbolisme des cultures les plus diverses, cette association archétypale du sommeil, de l'oiseau, de l'âme et de l'érection. Dans l'ancienne Égypte, un oiseau à tête humaine, souvent pourvu d'un étrange phallus en érection, symbolisait le Ba, l'âme immatérielle. Chaque personne abritait un Ba immortel qui prenait son envol après la mort. Une représentation traditionnelle du dieu Osiris, étonnamment similaire à la scène du puits de Lascaux, le montre allongé sur le dos, le pénis dressé, tandis que la chouette Isis le survole et recueille son sperme afin de donner naissance à Horus. Dans les Upanishad, les textes saints de la religion hindoue, l'âme prend aussi la forme d'une colombe qui s'envole lors de la mort et revient planer sur terre sous la forme d'un pur esprit. Des siècles plus tard, dans l'art religieux chrétien, colombes et autres oiseaux à ailes blanches symbolisent toujours l'âme, le Saint-Esprit et les anges gardiens. Du phénix égyptien, symbole de la résurrection, au Sielulintu finlandais, l'« oiseau de l'âme » qui apporte à chaque nouveau-né son âme immatérielle et la reprend au mourant, le libre vol de l'oiseau semble la métaphore universelle de l'autonomie de l'esprit.

Cette allégorie fonctionne parce qu'elle s'appuie sur une intuition répandue, selon laquelle nos pensées n'auraient rien de terrestre ou de matériel. Presque toutes les religions s'accordent sur ce point : l'étoffe de nos rêves, la substance de notre âme n'ont rien de commun avec la chair bassement matérielle qui compose nos corps. Lorsque nous rêvons, tandis que notre corps reste inerte, nos pensées vagabondent sans limite au royaume de l'imagination et de la mémoire. Quelle meilleure preuve, au moins en apparence, que l'activité mentale ne saurait se réduire au monde matériel ? Que l'esprit et le corps sont des substances distinctes ? Comment les envolées de l'esprit pourraient-elles germer des sillons du cerveau ?

Le défi de Descartes

L'idée que l'esprit et le corps appartiennent à des domaines distincts remonte à la plus haute antiquité. Théorisée par Platon dans le *Phédon* (composé au IVe siècle avant notre ère), la séparation de l'âme et du corps forme un volet primordial de la *Somme théologique* (1265-1274) de Thomas d'Aquin, texte fondateur de la doctrine chrétienne. Mais c'est au scientifique et philosophe René Descartes (1596-1650) que nous devons la formulation explicite de la thèse du dualisme : l'idée que l'âme humaine est constituée d'une substance distincte, immatérielle, qui échappe aux lois de la physique.

Il est devenu banal, en neurosciences, de ridiculiser la pensée de Descartes. Depuis la publication du best-seller d'Antonio Damasio, *L'Erreur de Descartes*[2] (1994), de nombreux livres dédiés à la conscience s'ouvrent sur une diatribe contre l'auteur du *Discours de la méthode*, accusé d'avoir, par ses hypothèses fantaisistes, retardé de plusieurs décennies l'avènement d'une neuropsychologie scientifique. La vérité, cependant, oblige à dire que Descartes était un scientifique visionnaire et, au plus profond, un réductionniste dont les analyses mécanicistes de l'esprit humain, très en avance sur leur temps, peuvent être lues comme le texte fondateur de la biologie synthétique et de la modélisation des réseaux neuronaux. Et c'est pourquoi le dualisme cartésien n'était pas la fantaisie d'un bigot – il reposait sur un argument logique, qui affirmait l'impossibilité qu'une machine puisse jamais imiter la liberté de l'esprit humain.

Le père fondateur de la psychologie moderne, William James, reconnaissait sa dette : « À Descartes revient d'avoir été le premier à avoir eu l'audace d'envisager une mécanique nerveuse qui se suffise entièrement à elle-même, et qui soit capable d'effectuer des actes complexes avec un semblant d'intelligence[3]. » De fait, dans des livres aux titres ambitieux (*Description du corps humain*, *Les Passions de l'âme*, *L'Homme*), Descartes abordait le fonctionnement du corps sous un angle

strictement mécanique. Nous sommes de subtils automates, écrivait-il intrépidement. Nos corps et nos cerveaux sont constitués d'organes étroitement intriqués qui fonctionnent comme les orgues des églises – de savants instruments de musique dans lesquels des soufflets insufflent des fluides biologiques, les « esprits animaux », à travers toute une série de réservoirs et de canalisations. Leurs mouvements composent la partition rythmique et mélodique de chacun de nos actes, sans qu'il soit nécessaire d'y postuler l'intervention de l'âme :

> « Je désire que vous considériez après cela que toutes les fonctions que j'ai attribuées à cette machine, comme la digestion des viandes, le battement du cœur et des artères, la nourriture et la croissance des membres, la respiration, la veille et le sommeil ; la réception de la lumière, des sons, des odeurs, des goûts, de la chaleur, et de telles autres qualités dans les organes des sens extérieurs ; l'impression de leurs idées dans l'organe du sens commun et de l'imagination ; la rétention ou l'empreinte de ces idées dans la mémoire ; les mouvements intérieurs des appétits et des passions ; et, enfin, les mouvements extérieurs de tous les membres, qui suivent si à propos tant des actions des objets qui se présentent aux sens que des passions et des impressions qui se rencontrent dans la mémoire, qu'ils imitent le plus parfaitement qu'il est possible ceux d'un vrai homme… Ces fonctions suivent tout naturellement en cette machine de la seule disposition de ses organes, ni plus ni moins que font les mouvements d'une horloge, ou autre automate, de celle de ses contrepoids et de ses roues[4]. »

Le cerveau hydraulique qu'imaginait Descartes n'avait aucune difficulté à déplacer sa main vers un objet. En effet, les contours de cet objet, en se projetant sur la face interne de l'œil (la rétine) activaient un jeu particulier de tuyaux, qui portait en lui le contenu de la scène visuelle.

Un mécanisme interne de décision, situé dans la glande pinéale, s'inclinait alors dans une direction précise, ce qui forçait les esprits animaux à s'écouler vers les muscles des membres et à les mettre en mouvement (figure 2). La mémoire correspondait au renforcement sélectif de certaines de ces voies d'écoulement – une étonnante préfiguration de l'idée contemporaine selon laquelle l'apprentissage repose sur la modification de la force des connexions cérébrales.

Le modèle de Descartes s'étendait même aux états de conscience. Le sommeil y était envisagé comme une baisse de pression des fluides internes. Lorsque la source des esprits animaux était abondante, ceux-ci circulaient à travers chacun de nos nerfs, et cette machine sous pression était prête à répondre à la moindre stimulation, « d'où vient que pour lors cette machine, étant disposée à obéir à toutes les actions des esprits, représente le corps d'un homme qui veille ». Lorsque la pression s'amoindrissait, que les esprits affaiblis ne gonflaient plus que quelques filets de nerfs, l'homme s'endormait.

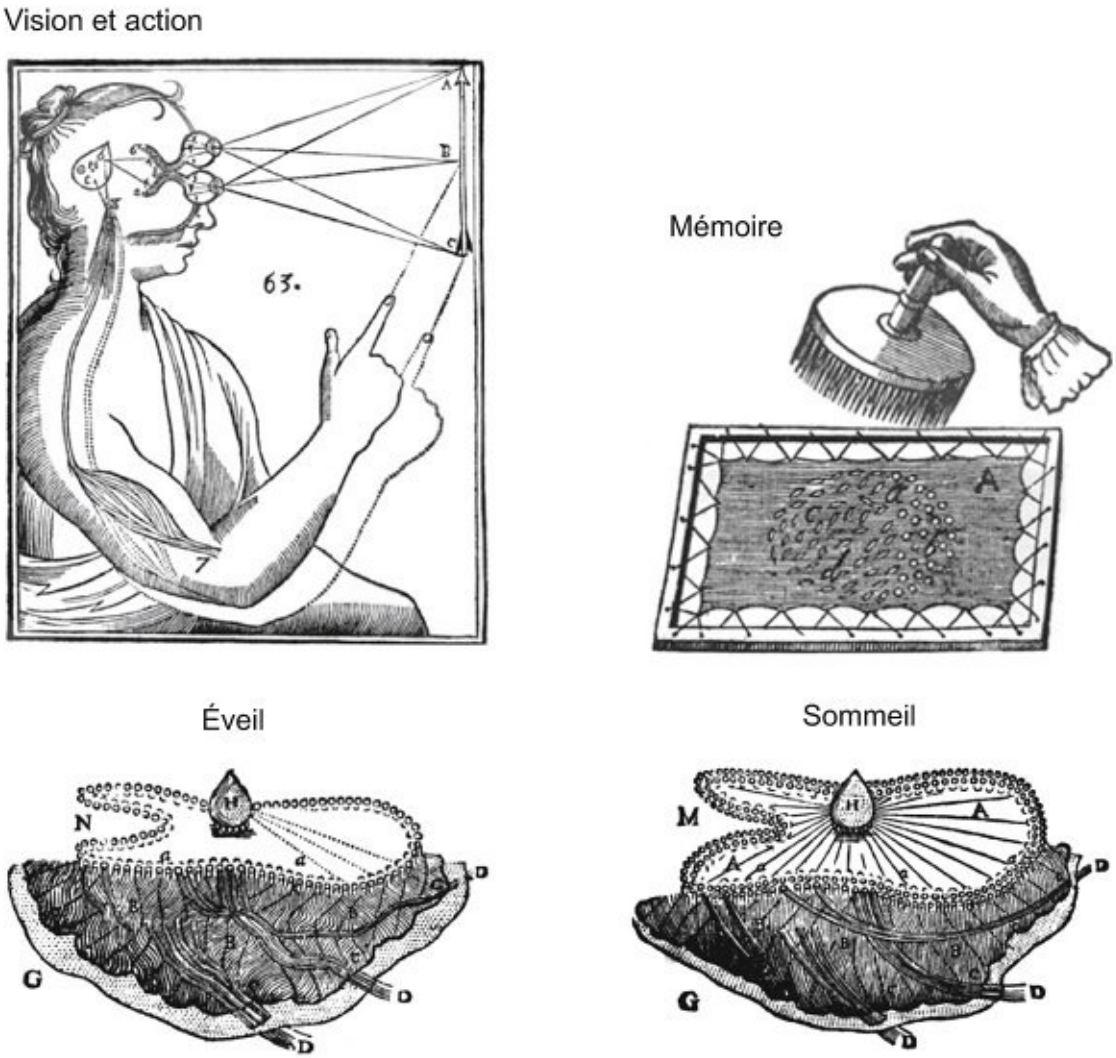


Figure 2. La théorie de René Descartes envisageait le système nerveux comme une machine, sans toutefois oser proposer une conception purement matérialiste de la pensée. Dans *L'Homme,* publié à titre posthume en 1664, Descartes expliquait comment la vision et l'action pouvaient résulter de la géométrie des connexions entre l'œil, la glande pinéale située au plus profond du cerveau et les muscles du corps. Il envisageait la mémoire comme la modulation de ces connexions, que l'on pouvait comparer au fait de trouer un tissu. Même les fluctuations de la conscience pouvaient s'expliquer par des variations de la pression des esprits animaux sur la glande pinéale : une haute pression correspondait à l'éveil et une basse pression au sommeil. En dépit de ces thèses matérialistes, Descartes soutenait que l'esprit et le corps étaient faits de substances distinctes qui interagissaient par l'entremise de la glande pinéale.

Et Descartes de conclure dans un lyrique élan matérialiste, plutôt inattendu sous la plume du père fondateur du dualisme, que pour expliquer toutes ces propriétés du corps, « il ne faut point, à leur occasion, concevoir en elle aucune autre âme végétative ni sensitive, ni aucun autre principe de mouvement et de vie, que son sang et ses esprits agités par la chaleur du feu qui brûle continuellement dans son cœur, et qui n'est point d'autre nature que tous les feux qui sont dans les corps inanimés ».

Dans ce cas, pourquoi Descartes affirmait-il l'existence d'une âme immatérielle ? Si l'on laisse de côté le fait qu'il tentait simplement de rester en vie, alors que l'Église brûlait Giordano Bruno et enfermait Galilée[5], Descartes, qui était un scientifique honnête, avait amèrement conscience que sa vision mécanique ne permettait pas de comprendre, d'une façon strictement matérialiste, les fonctions supérieures de l'esprit humain. Deux opérations mentales majeures semblaient échapper à la machinerie du corps. La première d'entre elles était la faculté de langage articulé, qui nous permet d'exprimer chacune de nos pensées. Descartes ne voyait pas comment une machine pourrait « user de paroles ou d'autres signes en les composant, comme nous faisons pour déclarer aux autres nos pensées ». Certes, un arrangement de soufflets pouvait bien émettre des cris, et même hurler « j'ai mal » si on le touchait de-ci ou de-là. Mais comment une machine aurait-elle pu arranger les mots en phrases les plus diverses, « pour répondre au sens de tout ce qui se dira en sa présence, ainsi que les hommes les plus hébétés peuvent faire » ?

La seconde disposition du cerveau humain qui posait un sérieux problème était la flexibilité de nos raisonnements. Une machine n'était, par définition, qu'un assemblage figé de rouages et de tuyaux, qui ne pouvait qu'agir de façon prédéterminée, « selon la disposition de ses organes ». Comment aurait-elle pu engendrer l'infinie variété de nos pensées ? La raison, expliquait notre philosophe, « est un instrument universel qui peut servir en toutes sortes de rencontres, [alors que] ces organes ont besoin de quelque particulière disposition pour chaque action particulière ; d'où vient qu'il est moralement impossible qu'il y en ait assez

de divers en une machine, pour la faire agir en toutes les occurrences de la vie de même façon que notre raison nous fait agir ».

Près de quatre siècles plus tard, les arguments de Descartes continuent de poser de singuliers défis aux neurosciences contemporaines. Comment la machinerie du cerveau lui permet-elle de produire des phrases infiniment variées, et d'exprimer ainsi toute la diversité de ses états mentaux ? Et comment prend-elle des décisions rationnelles d'une manière aussi flexible, qui semble infiniment adaptable à l'extraordinaire diversité de problèmes que nous rencontrons ? Toute science de la conscience se doit de répondre à ces questions.

Le dernier problème

> « L'humanité est parvenue à identifier des galaxies situées à des années-lumière d'ici, à étudier des particules plus petites qu'un atome, mais nous n'avons toujours pas élucidé le mystère de ce kilo et demi de matière qui se trouve entre nos deux oreilles. »
>
> Barack OBAMA, lors de l'annonce du programme américain BRAIN (2 avril 2013).

Grâce à Euclide, à Karl Friedrich Gauss ou à Albert Einstein, nous possédons aujourd'hui un bagage considérable de principes mathématiques qui régissent le monde de la physique. Juchés sur les épaules de géants tels qu'Isaac Newton et Edwin Hubble, nous avons compris que notre Terre n'est qu'un grain de poussière de l'une des nombreuses galaxies issues de l'explosion primordiale du Big Bang. Plus récemment, Charles Darwin, Louis Pasteur, James Watson et Francis Crick nous ont montré que la vie n'est qu'un enchaînement de milliards de réactions chimiques évoluées – un bric-à-brac de physique et de chimie.

Aujourd'hui, seule la question de l'origine de la conscience semble être restée dans les limbes. Elle constitue, sans doute, le dernier grand

problème scientifique, de ceux que tout enfant se pose spontanément : comment mon cerveau pense-t-il ? Qu'est-ce que ce moi qui dit « je pense » ? Serais-je un autre si j'étais né dans un autre lieu, un autre temps, un autre corps ? Où mon esprit part-il lorsque je rêve ? Que deviendra-t-il après ma mort ? Toutes ces pensées viennent-elles seulement de mon cerveau ? Ou bien suis-je, au moins en partie, un pur esprit fait d'une substance incorporelle ?

Ces questions lancinantes ont, de tout temps, préoccupé en vain les plus brillants esprits. Dès 1580, le grand humaniste Michel de Montaigne, dans l'un de ses *Essais*, se lamentait de ne pas trouver, chez les Anciens, la moindre réponse cohérente à la question de l'origine de l'âme et de son siège au sein du corps : « Hippocrate et Hérophile la mettent au ventricule du cerveau ; Démocrite et Aristote, par tout le corps ; Épicure, en l'estomac ; les stoïciens, autour et dedans le cœur ; Érasistrate, joignant la membrane de l'épicrâne ; Empédocle, au sang ; (…) Galien a pensé que chaque partie du corps ait son âme ; Straton l'a logée entre les deux sourcils[6]. »

Pendant la majorité du XIX^e^ et du XX^e^ siècle, la question de l'origine de la conscience a eu mauvaise réputation. Les scientifiques purs et durs la considéraient comme floue et hautement subjective, ce qui entravait toute expérimentation objective. Aborder ce sujet, c'était risquer de saborder sa carrière : au mieux, la spéculation sur la conscience était perçue comme un hobby de professeur en retraite. Dans son livre d'introduction à la *Psychologie, science de la vie mentale* (1962), George Miller, pionnier de la psychologie cognitive, alla jusqu'à proposer de bannir ce domaine de recherches : « La conscience est un mot que des millions de personnes ont usé jusqu'à la corde… nous devrions peut-être le bannir de notre vocabulaire jusqu'à ce que nous parvenions à inventer des termes plus précis pour les différents usages que le mot "conscience" ne fait qu'embrouiller aujourd'hui. »

Et ce bannissement a bel et bien existé. Dans les années 1980, au cours de ma thèse de psychologie cognitive, j'ai été surpris de découvrir

que le mot « conscience » n'était le bienvenu ni dans nos articles ni dans nos réunions de laboratoire. Bon nombre de nos recherches y touchaient pourtant de près ou de loin, puisque nous demandions par exemple à nos volontaires de rapporter ce qu'ils avaient vu, ou de former des images mentales. Cependant, le mot lui-même demeurait tabou, et une publication sérieuse se devait de l'éviter. Même lorsqu'un expérimentateur présentait des images pour une durée extrêmement brève, au seuil de la perception, il ne se préoccupait guère de savoir si les participants avaient ou non pris conscience de ces stimuli. Bien que quelques voix dissonantes s'élevassent contre ce consensus[7], la plupart des chercheurs s'accordaient à penser que le vocable « conscience » n'apportait rien de fondamental à la psychologie scientifique. Les sciences cognitives naissantes se proposaient de modéliser l'ensemble des opérations mentales en termes de traitement de l'information et de ses mécanismes neuronaux ou moléculaires. Elles n'avaient que faire d'un terme aussi éculé, mal défini et subjectif que celui de « conscience ».

Et pourtant, à la fin des années 1980, la communauté des chercheurs a changé radicalement d'attitude. Aujourd'hui, le problème de la conscience est à la pointe de la recherche en neurosciences. C'est un domaine très actif auquel sont dédiées plusieurs revues et sociétés scientifiques spécialisées. Ses avancées commencent à répondre, avec un certain succès, au défi de Descartes : comment notre cerveau génère-t-il des idées subjectives, un point de vue personnel que nous pouvons rapporter à d'autres personnes et utiliser avec flexibilité dans nos raisonnements ?

Dans ce livre, nous allons voir comment le vent a tourné.

Déchiffrer le code conscient

Au cours des vingt dernières années, les chercheurs en sciences cognitives, en neurophysiologie et en imagerie cérébrale ont combiné leurs efforts afin d'attaquer de front le problème de la conscience. Très vite, celui-ci a perdu son arrière-goût idéologique et spéculatif pour devenir une question expérimentale. Dans les chapitres qui suivent, nous allons examiner en détail la stratégie qui a permis de transformer un mystère philosophique en un simple phénomène de laboratoire. Nous verrons que trois ingrédients essentiels ont rendu cette évolution possible : la formulation d'une meilleure définition de la conscience ; la découverte que celle-ci peut être manipulée à volonté ; et un respect accru pour les phénomènes subjectifs.

Le mot « conscience », tel que nous l'utilisons dans la vie courante, revêt de multiples acceptions souvent floues et confuses. Notre première tâche consistera donc à mettre de l'ordre dans ce chaos. Nous tenterons de ramener ce vaste sujet à une question suffisamment restreinte pour être soumise à l'épreuve de l'expérience. Comme nous le verrons, la science moderne de la conscience distingue au moins trois concepts : le degré de vigilance, qui varie quasi continûment depuis la veille jusqu'au sommeil ou au coma profond ; l'attention, c'est-à-dire la focalisation de nos ressources mentales sur un objet particulier ; et, enfin, l'accès à la conscience, c'est-à-dire le fait que seule une partie de nos pensées entre dans le champ de notre conscience, devient disponible pour diverses opérations cognitives et peut être rapportée à d'autres.

Le nœud du problème consiste à comprendre comment une information accède à la conscience. Nous partirons d'une observation très simple : en général, lorsque nous sommes éveillés, et pour peu que nous y prêtions attention, n'importe quel objet mental peut devenir conscient. Vigilance et attention permettent la conscience, cependant elles n'y suffisent pas. Bien que nous soyons éveillés et attentifs, il se peut qu'un objet apparaisse sur notre rétine et soit traité par notre cortex, sans pour autant accéder à notre conscience. Le simple fait de présenter une image très brièvement, comme nous le verrons, peut suffire à la rendre

subliminale et invisible à notre conscience, sinon à nos yeux. Nous dirons qu'une image visible accède à la conscience, tandis qu'une image invisible (mais dont le contenu circule pourtant dans nos réseaux corticaux) n'y accède pas.

Au sein de la nouvelle science dont nous esquissons les contours, la notion d'« accès à la conscience » est un phénomène bien délimité, distinct de la vigilance et de l'attention. Il est facile de l'étudier en laboratoire. Nous avons découvert des dizaines de manières de faire franchir à une image la frontière mystérieuse entre le perçu et le non-perçu, le visible et l'invisible, le conscient et le non-conscient – et à suivre, avec nos instruments d'imagerie cérébrale, ce que ce pas supplémentaire change au niveau neurophysiologique.

L'accès à la conscience ouvre la porte à des formes plus complexes d'expérience subjective. Dans la vie quotidienne, nous confondons souvent la prise de conscience avec la représentation de soi – comment le cerveau crée un point de vue sur le monde, un sujet capable de dire « je », un poste d'observation subjectif d'où ce « je » envisage « les autres ». Se penchant sur elle-même, la conscience peut également devenir récursive : le « je » devient l'objet de ses propres pensées et parvient à savoir qu'il sait… ou qu'il ne sait pas. Bonne nouvelle : même ces aspects supérieurs de la conscience ne sont plus inaccessibles à l'expérimentation. Dans nos laboratoires, nous avons appris à quantifier ce que le « je » ressent, et à étudier ce qu'il parvient à rapporter, tant sur son environnement que sur lui-même. Nous parvenons même à manipuler la représentation de soi, de sorte que des personnes strictement normales éprouvent une sensation extraordinaire de « sortie du corps » alors qu'elles sont allongées dans une machine IRM.

Certains philosophes continuent de prétendre que toutes ces avancées scientifiques ne changent rien au cœur du problème. Le nœud de l'énigme, selon eux, repose dans un autre sens du mot « conscience », qu'ils appellent la « conscience phénoménale » : l'intuition, profondément ancrée en nous, que chacune de nos expériences conscientes est un

phénomène « en soi », qu'elle possède une qualité particulière, un *quale* irréductible de pensée. La douleur exquise d'une rage de dents, le vert inimitable d'une feuille fraîchement éclose sous le soleil de printemps… Selon les penseurs de cette obédience, ces qualités subjectives ne sauraient se réduire à une description scientifique ou à un état du système nerveux. Par nature, ces états internes, personnels et subjectifs échapperaient à toute tentative de description objective. Cependant, je m'inscris en faux contre cette idée. Dans ce livre, j'essaierai d'argumenter que la notion d'une conscience « phénoménale », intrinsèquement différente de celle d'accès conscient, est trompeuse et nous entraîne sur la pente glissante qui mène au dualisme.

Commençons par les choses simples, et tentons de comprendre comment une information sensorielle accède à la conscience. Une fois que nous aurons clarifié comment l'acte de perception transforme certaines des informations qui frappent notre rétine en pensées conscientes, la montagne philosophique que nous nous faisons du caractère ineffable de l'expérience subjective accouchera d'une souris… de laboratoire.

Voir ou ne pas voir

L'accès à la conscience paraît trop simple pour qu'on s'y arrête. Il nous suffit de poser les yeux sur un objet, semble-t-il, pour prendre immédiatement conscience de sa forme, de sa couleur et de son identité. Mais cette apparente simplicité n'est qu'un leurre. Derrière cette impression d'instantanéité se cache une avalanche de processus cérébraux d'une complexité redoutable. Des milliards de neurones visuels nous donnent à voir le monde, et leurs calculs prennent jusqu'à une demi-seconde. Comment analyser cette longue chaîne en étapes élémentaires ? Et surtout, est-il possible de distinguer certaines étapes purement

automatiques et inconscientes, et d'autres qui entraînent, obligatoirement, un sentiment de vision consciente ?

C'est ici qu'entre en jeu le second ingrédient essentiel de la science contemporaine de la conscience. Nous disposons aujourd'hui d'une solide approche empirique du problème, parce que les sciences cognitives des vingt dernières années ont découvert une étonnante panoplie de moyens de manipuler la conscience à volonté. Tout un éventail de conditions expérimentales nous permet de passer du visible à l'invisible, en modifiant le stimulus de façon presque insensible, voire pas du tout. Nous pouvons faire apparaître un mot si brièvement que la plupart des participants ne le voient pas. Nous avons appris à créer des images visuelles encombrées de tant d'objets que l'un d'entre eux demeure invisible, parce que ce sont toujours les autres qui gagnent la compétition interne pour l'accès à la conscience. Nous pouvons également compter sur le détournement de l'attention : comme le savent les magiciens, même un geste évident peut devenir parfaitement invisible pour peu que l'attention du spectateur soit distraite. Parfois, enfin, c'est notre cerveau qui nous joue un tour de magie : quand on lui présente des images distinctes dans chaque œil, notre système visuel balance entre les deux et nous laisse voir tantôt l'une, tantôt l'autre, mais jamais les deux à la fois.

L'image perçue, qui accède à la conscience, et l'image effacée, qui disparaît dans les limbes de l'inconscient, peuvent très bien ne différer que d'un iota. Cependant, dans le cerveau, cette différence minime fait toute la différence. Nos circuits neuronaux doivent l'amplifier, puisque, après quelques instants, nous pouvons toujours décrire la première alors que la seconde n'est plus rien à nos yeux. Comprendre précisément pourquoi et comment se produit cette divergence entre le conscient et le non-conscient, tel est l'objet précis de la nouvelle science de la conscience.

La stratégie expérimentale qui consiste à créer un contraste minimal entre la perception consciente et non consciente a fourni la clé du sanctuaire intime de l'expérience subjective[8]. Au fil des ans, nous avons découvert de nombreuses conditions expérimentales parfaitement

appariées, telles que l'une conduit à la prise de conscience et l'autre non. À chaque découverte, l'intimidant mystère de la conscience perdait de sa superbe. Il se transformait progressivement en une simple question expérimentale : parviendrions-nous à déchiffrer les mécanismes cérébraux qui séparent ces deux conditions ?

Faire du subjectif une science

La stratégie de recherche que je viens d'esquisser paraît simple. Elle repose cependant sur un point si controversé que je le considère personnellement comme le troisième ingrédient majeur de la nouvelle science de la conscience : il s'agit d'accorder foi à la subjectivité de l'observateur.

Présenter des stimuli à des volontaires ne suffit pas. Encore faut-il enregistrer, avec précision, ce que ressent chaque personne. Ce rapport introspectif est une donnée essentielle : c'est lui qui définit le phénomène même que nous souhaitons éclaircir. Si l'expérimentateur voit une image alors que le volontaire, lui, affirme qu'il n'a rien vu, c'est cette seconde réponse qui compte, toute subjective qu'elle soit : l'image doit être classée dans la catégorie des stimuli invisibles (pour cette personne, à cet instant). Ainsi la psychologie expérimentale doit-elle concevoir de nouvelles manières de mesurer l'introspection aussi précisément que possible.

Cette emphase sur l'expérience subjective est une véritable révolution en psychologie. Au début du XX^{e} siècle, les béhavioristes forcenés tels que John Broadus Watson (1878-1958) avaient vigoureusement écarté de leur science psychologique toute référence à l'introspection :

> « La psychologie telle que le béhavioriste la voit est une branche purement objective des sciences naturelles. Son but théorique est

> la prédiction et le contrôle du comportement. L'introspection ne fait pas partie de ses méthodes essentielles, et la valeur scientifique de ses données ne dépend pas de la façon dont elles se prêtent à une interprétation en termes de conscience[9]. »

Bien qu'en définitive le béhaviorisme lui-même ait été écarté, il a laissé des traces durables sur notre discipline. Tout au long du XX^e^ siècle, le recours à l'introspection est resté entaché d'une aura de suspicion. J'affirme pourtant que cette position dogmatique est tout à fait inappropriée, car elle confond deux questions bien distinctes : l'introspection en tant que méthode de recherche et l'introspection en tant que donnée brute.

En tant que méthode, l'introspection n'est guère digne de confiance[10]. De toute évidence, nous ne pouvons pas demander à un observateur de se pencher en lui-même et de nous dire comment fonctionne son esprit (si tel était le cas, notre recherche serait facile !). De plus, nous ne devons pas prendre ses réponses au pied de la lettre. Si quelqu'un me dit qu'il plane hors de son corps, que son cerveau contient une puce implantée par la CIA, ou qu'il dialogue avec Jules César en rêve, je ne prends pas ses récits pour argent comptant. Et pourtant, en un sens, même des situations aussi étranges que celles-là doivent être prises au sérieux, car si la personne ne ment pas, elles correspondent à d'authentiques états mentaux qu'il faudra bien que le psychologue explique.

La seule attitude appropriée consiste à traiter ces rapports subjectifs comme des données brutes[11]. Une personne qui prétend être sortie de son corps se sent vraiment, pendant un moment, flotter dans les airs. Tant que nous n'aurons pas expliqué d'où provient cette étrange impression, notre science de la conscience restera lettre morte. De fait, nos recherches contemporaines font souvent appel à des phénomènes subjectifs tels que les illusions visuelles, les images ambiguës, les hallucinations et tous les autres fruits de l'imaginaire normal ou pathologique. Seuls ces

événements permettent de dissocier complètement le stimulus objectif que reçoit le cerveau et la perception subjective qu'il en construit – donc de rechercher les bases cérébrales de l'un sans être contaminé par l'autre. Le psychologue qui travaille sur la conscience n'est jamais aussi heureux que lorsqu'il découvre une nouvelle illusion visuelle qui rend une image tantôt visible et tantôt invisible, ou un nouveau type de son ambigu dont l'un des aspects est parfois audible, et parfois non. Tant que nous prenons la précaution de mesurer, à chaque essai, ce que le participant déclare ressentir, l'expérience est valide : nous pouvons trier les essais en fonction de l'expérience consciente, et tenter de découvrir ce qui les distingue sur le plan cérébral.

Les signatures d'une pensée consciente

Une fois nos trois ingrédients réunis – la focalisation sur l'accès à la conscience, la manipulation d'images visibles et invisibles, et leur tri systématique en fonction de l'introspection –, l'étude de la conscience peut rentrer dans le giron de la science « normale ». À l'aide des outils modernes d'imagerie cérébrale, suivre les étapes que franchit une image non consciente dans le cerveau ne pose aucune difficulté d'ordre fondamental – ce n'est plus un problème philosophique, mais une simple question d'ingéniosité expérimentale. Comme nous le verrons, quantité d'opérations mentales se déroulent hors de notre conscience. L'étude des images subliminales permet de délimiter, comme en négatif, les contours de l'expérience consciente. L'imagerie cérébrale fournit des moyens nouveaux d'étudier jusqu'à quel point précis un stimulus peut parcourir les autoroutes de l'information du cerveau sans pour autant accéder à la conscience. En précisant où s'arrête ce traitement subliminal, ces expériences révèlent quelles configurations de l'activité cérébrale caractérisent exclusivement le traitement conscient.

Depuis près de vingt ans, mon équipe utilise tous les outils à sa disposition, depuis l'imagerie fonctionnelle par résonance magnétique (IRMf) jusqu'à l'électro-et la magnétoencéphalographie, en passant par des électrodes implantées dans la profondeur du cortex, pour tenter d'isoler des marqueurs neurophysiologiques de la pensée consciente. Aux côtés de nombreux autres chercheurs dans le monde, nous recherchons systématiquement quel type d'activité cérébrale apparaît exclusivement lorsque la personne qui se fait scanner rapporte avoir consciemment perçu un mot ou une image – ce que j'appelle une « signature de la conscience ». Et notre recherche a été couronnée de succès. D'une expérience à l'autre, nous retrouvons la même série de marqueurs physiologiques qui, tous, changent soudain d'état lorsqu'une personne prend conscience d'une image, d'un mot ou d'un son. Leur nature est remarquablement stable, et nous les avons observés en réponse à un grand nombre de stimulations visuelles, auditives, tactiles et cognitives.

La découverte empirique de marqueurs reproductibles de l'accès à la conscience, communs à tous les êtres humains, n'est qu'une étape. Il nous faut encore progresser sur le plan théorique. D'où proviennent ces signatures ? Surtout, *pourquoi* leur présence témoigne-t-elle d'un cerveau conscient ? Comment se fait-il que seul un certain type d'état cérébral s'accompagne d'une expérience consciente ? Bien qu'aucun scientifique ne soit en mesure d'affirmer qu'il connaît la réponse à ces questions, nous disposons aujourd'hui d'hypothèses fortes et qui commencent à être mises à l'épreuve. Mes collaborateurs et moi-même avons élaboré une théorie précise, l'hypothèse de l'« espace de travail neuronal global ». Nous proposons que la conscience correspond à une diffusion globale de l'information dans le cortex. Elle émerge d'un réseau de connexions corticales dont la fonction est de diffuser, à l'ensemble du cerveau, certaines informations sélectionnées en fonction de leur pertinence. Une idée est consciente lorsqu'elle est codée par une assemblée de neurones au sein de cet espace cérébral spécifique.

Le philosophe Daniel Dennett résume cette théorie en deux mots : la « célébrité cérébrale » (*fame in the brain*). La conscience diffuse à tout le cerveau les idées dignes d'être « célèbres ». Grâce à notre espace de travail global, nous gardons à l'esprit, aussi longtemps que nécessaire, toute idée qui nous frappe. Le « buzz » que la conscience crée dans notre cerveau permet à cette information d'être communiquée à toutes les aires de notre cortex, et donc d'être incorporée à tous nos plans d'action, quels qu'ils soient. Selon cette hypothèse, la prise de conscience remplit une fonction bien précise : elle sélectionne, amplifie et propage nos pensées pertinentes. Grâce à elle, l'organisme dans son ensemble, bien qu'il soit composé d'un très grand nombre de systèmes disparates, agit « comme un seul homme », d'une façon unifiée, rationnelle et en toute connaissance de cause.

Quel circuit cérébral assure la diffusion des informations conscientes ? Nous pensons que cette fonction est assurée par un type de neurones bien particulier : des cellules nerveuses géantes dont les axones longs traversent le cortex de part en part et l'interconnectent pour former un vaste réseau intégré. Des simulations informatiques de cette architecture reproduisent nos principaux résultats expérimentaux. Quand un nombre suffisant de régions cérébrales s'accordent sur l'importance d'une information sensorielle, elles se synchronisent pour former un seul état neuronal global. Un vaste ensemble de neurones s'embrase et, comme nous le verrons plus loin dans ce livre, la nature soudaine et massive de cet embrasement cortical explique que l'on parvienne à détecter de multiples signatures de la conscience dans nos signaux d'imagerie cérébrale.

Bien que de nombreuses opérations cérébrales puissent s'exécuter inconsciemment, l'accès d'une information à la conscience enrichit considérablement la nature des opérations cognitives disponibles. Tout un ensemble de fonctionnalités requiert la diffusion consciente d'une information. L'espace de travail neuronal global nous donne accès à un champ étendu d'expériences de pensée libérées des contingences du

temps présent. Grâce à lui, nous pouvons garder à l'esprit, pour une durée arbitraire, toute information que nous jugeons importante. De plus, nous pouvons la confronter avec n'importe quelle autre idée que nous avons en tête. En effet, l'espace global permet de transmettre un contenu conscient à n'importe lequel de nos « processeurs » cérébraux, ce qui confère à nos cerveaux la flexibilité même dont Descartes s'émerveillait. Dès qu'une information accède à la conscience, elle peut être intégrée à une série arbitrairement longue d'opérations mentales – elle n'est plus simplement traitée d'une façon réflexe, automatique, mais peut être contemplée aussi longtemps que nécessaire et réorientée à volonté. De plus, puisque l'une de ces interconnexions atteint les aires cérébrales du langage, nous pouvons la rapporter à d'autres.

Une propriété tout aussi fondamentale de l'espace de travail neuronal global est son autonomie. De nombreuses études récentes ont montré que le cortex est le siège d'une intense activité nerveuse spontanée. Des vagues de décharges synchronisées le traversent constamment, qui ne trouvent pas nécessairement leur origine dans les aires sensorielles, mais de l'intérieur. Cellule excitable, le neurone dispose d'une capacité à s'autoactiver spontanément et de manière partiellement aléatoire. Au sein d'un réseau de neurones interconnectés, cette excitabilité spontanée s'organise selon des formes complexes. Ainsi, à l'exact opposé de la métaphore cartésienne d'un orgue dont les tuyaux ne s'activent que si l'on en frappe le clavier, notre espace de travail neuronal global ne fonctionne pas selon un mode réflexe, entrée-sortie. En l'absence de toute stimulation, dans le noir et le silence, il est incessamment parcouru d'une sorte de ressac neuronal qui s'auto-organise pour former une séquence de configurations stables, et engendre ce que William James appelle le « flux de la conscience » – un flot ininterrompu de pensées librement déliées, sculptées seulement par les buts que nous avons en tête, et qui ne vont que rarement interroger les entrées sensorielles. René Descartes ne pouvait imaginer une machine de ce type, dans laquelle éclosent sans cesse des plans inédits, des pensées débridées qui engendrent nos

comportements. L'autonomie qui en résulte, le « libre arbitre » que l'activité neuronale spontanée confère à notre cortex, répond au défi de Descartes. Ainsi s'esquisse un modèle plausible de la conscience.

L'avenir de la conscience

Notre compréhension de la conscience reste rudimentaire. Que nous réserve l'avenir ? En fin d'ouvrage, munis de nouvelles certitudes scientifiques, nous pourrons revenir aux anciennes querelles philosophiques. Je défendrai l'idée que notre connaissance grandissante de la conscience conduira prochainement à répondre à certaines de nos plus profondes interrogations sur nous-mêmes, mais également à affronter de difficiles questions de société ou à concevoir des technologies nouvelles qui dupliquent la puissance de calcul de l'esprit humain.

Nul doute que de nombreux détails restent à comprendre – mais la science de la conscience est déjà bien plus qu'une hypothèse. Des applications médicales sont à portée de main. Dans bien des hôpitaux de par le monde, des milliers de patients dans le coma ou en état végétatif reposent dans un terrible isolement, la voix morte, le cerveau ruiné par un accident vasculaire, un accident de voiture ou une hypoxie. Reprendront-ils connaissance ? Se pourrait-il que certains soient déjà conscients, mais intégralement paralysés et incapables de le faire savoir ? Pour les aider, saurons-nous transformer l'imagerie cérébrale en un détecteur de conscience en temps réel ?

Mon laboratoire conçoit actuellement des tests innovants qui commencent à distinguer si une personne est ou n'est pas consciente. De telles signatures objectives de la conscience vont révolutionner le suivi des comas, et pourront également éclairer la difficile question de l'émergence de la conscience chez l'enfant nouveau-né ou prématuré. Bien que les découvertes scientifiques ne dictent jamais de règles de conduite, je suis

persuadé que, lorsque nous parviendrons à déterminer, d'une manière strictement objective, si un patient ou un bébé éprouvent des expériences subjectives, nos décisions éthiques en seront facilitées.

L'informatique offrira bientôt un autre débouché à la science de la conscience. Parviendrons-nous à imiter, *in silico*, les circuits du cerveau ? Notre connaissance actuelle suffira-t-elle à concevoir un ordinateur conscient ? Et dans le cas contraire, que lui manque-t-il ? À mesure que progresse la théorie de la conscience se rapproche la perspective de la dupliquer dans des puces électroniques dont l'architecture imite l'opération des neurones. La prochaine étape sera-t-elle une machine consciente de ses propres limites ? Parviendrons-nous à lui conférer un « moi » autonome, voire un sentiment de libre arbitre ?

Je vous invite à présent à un voyage aux frontières de la science, une quête qui donnera un sens nouveau à la célèbre devise grecque : « Connais-toi toi-même. »

CHAPITRE 1

Quand la conscience entre au laboratoire

Comment la conscience est-elle devenue un objet de science ? Reléguant à plus tard les obscures énigmes du libre arbitre et de la conscience de soi, nous nous sommes concentrés sur la question de l'accès à la conscience : pourquoi certaines de nos entrées sensorielles deviennent-elles des perceptions conscientes, tandis que d'autres restent confinées aux limbes de l'inconscient ? Afin de comparer la perception consciente avec le traitement non conscient des informations sensorielles, nous avons appris à rendre n'importe quelle image visible ou invisible, à volonté et sous contrôle intégral de l'expérimentateur. Nous sommes même parvenus à présenter des images au seuil de conscience, en sorte qu'elles ne soient perçues consciemment que la moitié du temps. Dans ces conditions, il est devenu essentiel de faire appel à l'introspection de chaque observateur, car elle seule permet de connaître le contenu subjectif de la conscience. Tous ces ingrédients composent un programme de recherche simple : la quête des mécanismes objectifs de la conscience subjective, des « signatures » systématiques qui marquent la transition de l'activité cérébrale de la non-conscience à la conscience.

Examinez l’illusion visuelle présentée en figure 3. Douze gros points, imprimés en gris clair, entourent une croix. À présent, fixez cette petite croix du regard. Au bout de quelques secondes, vous devriez voir certains points disparaître de la page. L’espace d’un instant, ils s’effacent de la conscience, puis redeviennent visibles. Parfois, c’est tout l’ensemble qui s’évanouit et vous laisse face à une page blanche – pour mieux revenir, un instant plus tard, dans une nuance de gris qui semble soudain plus sombre.

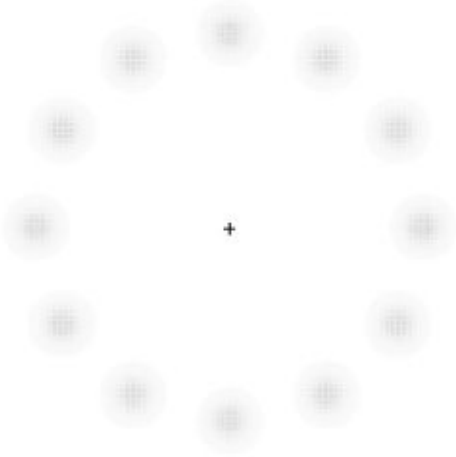


Figure 3. L'illusion de Troxler illustre l'une des nombreuses manières de manipuler la conscience subjective. Fixez la croix centrale du regard. Au bout de quelques secondes, vous verrez certains des disques gris disparaître, puis réapparaître d'une façon imprévisible. Le stimulus objectif est constant, et pourtant son interprétation subjective ne cesse de changer. Reste alors à déterminer quels aspects de l'activité cérébrale varient en fonction de la perception subjective plutôt que de la stimulation objective.

Ainsi, une image fixe peut entrer et sortir de la conscience, plus ou moins au hasard. Cette observation profonde constitue le fondement de la science moderne de la conscience. En 1990, le prix Nobel Francis Crick, aujourd'hui disparu, et le neurobiologiste Christof Koch ont avancé l'idée que ce genre d'illusion visuelle fournissait un moyen scientifique d'étudier le trajet de stimuli conscients et non conscients dans le cerveau[1].

Au moins en théorie, ce programme de recherche ne présente pas de difficulté majeure. Pendant l'illusion des douze points gris, par exemple, rien n'empêche d'enregistrer l'activité neuronale en divers endroits du cerveau pendant que les points sont visibles, et de comparer ces enregistrements avec ceux faits au moment où les points ne sont pas vus. Crick et Koch pensaient que le domaine de la vision était particulièrement mûr pour de telles recherches, car non seulement nous commençons à bien connaître les voies nerveuses qui transportent l'information de la

rétine au cortex, mais également parce que des dizaines d'illusions permettent de comparer des stimuli visibles et invisibles[2]. Toutes ces illusions partagent-elles des points communs ? Existe-t-il une forme unique d'activité cérébrale qui sous-tend tous les états conscients et fournit une signature unifiée de l'accès à la conscience ? L'identification d'une telle signature ferait faire un pas de géant à l'élucidation de cette énigme.

Crick et Koch avaient entrouvert la porte du labyrinthe. À leur suite, des dizaines de laboratoires commencèrent à explorer la perception consciente par le biais d'illusions semblables à celles que vous venez d'éprouver. Trois éléments clés rendaient excitant ce programme de recherche. Tout d'abord, les illusions ne faisaient aucunement appel à une conception alambiquée de la conscience – il s'agit juste de comprendre l'acte de voir ou de ne pas voir, ce que j'appelle l'accès à la conscience. En second lieu, des dizaines d'illusions étaient disponibles – comme nous allons bientôt le voir, les chercheurs, devenus magiciens, ont vite appris à faire disparaître de la conscience n'importe quel mot, n'importe quelle image, et même un gorille tout entier ! Enfin, ces illusions sont purement subjectives – vous seuls savez où et quand les points disparaissent de votre vision. Il n'empêche que les résultats sont reproductibles : toute personne qui fixe l'image éprouve la même expérience subjective. Nous nous accordons tous sur le fait qu'un événement bien réel, étrange et fascinant, se produit dans notre conscience. Nous devons le prendre au sérieux.

En résumé, trois ingrédients suffisent à amener la conscience à portée de la science : la concentration des recherches sur l'accès à la conscience ; la panoplie d'illusions qui permettent de manipuler la conscience à volonté ; et le fait de traiter les phénomènes subjectifs comme d'authentiques données scientifiques. Examinons-les à présent en détail.

Les facettes de la conscience

> « Conscience : le fait d'avoir des perceptions, des pensées et des sentiments. Le terme est impossible à définir, sauf en des termes qui sont eux-mêmes inintelligibles si l'on ne sait pas ce que signifie la conscience… Rien de valable n'a jamais été écrit à ce sujet. »
>
> Stuart SUTHERLAND, *International Dictionary of Psychology* (1996).

La science progresse souvent en introduisant des distinctions nouvelles qui raffinent les concepts souvent flous des langues naturelles. Un exemple classique dans l'histoire des sciences est celui de la séparation des concepts de chaleur et de température. L'intuition traite ces deux termes comme des synonymes : après tout, ajouter de la chaleur, c'est augmenter la température, n'est-ce pas ? Non – un bloc de glace, lorsqu'on le réchauffe, fond tout en restant à une température constante de zéro degrés. Un matériau peut très bien avoir une température élevée (par exemple une étincelle de feu d'artifice peut facilement atteindre quelques milliers de degrés), et contenir une chaleur si faible qu'elle ne cause aucune brûlure (parce que sa masse est très faible). Au XIX^{e} siècle, la distinction entre chaleur (la quantité d'énergie transférée) et température (l'énergie cinétique moyenne des particules qui composent un corps) a été l'une des clés du progrès fulgurant de la thermodynamique.

Le mot *conscience*, tel qu'on l'utilise dans la vie courante, ressemble au mot *chaleur* : en amalgamant plusieurs sens, il sème la confusion. Nous ne pourrons apporter de l'ordre à ce domaine qu'en les séparant. Dans ce livre, je défends l'idée que l'un d'entre eux en particulier, l'*accès à la conscience*, recouvre un concept particulièrement clair et suffisamment restreint pour faire l'objet d'expériences précises et susceptibles de jeter une certaine lumière sur l'ensemble du problème.

Qu'est-ce donc que l'accès à la conscience ? À chaque instant, notre cerveau est saturé d'innombrables stimulations sensorielles. Cependant, notre conscience ne nous donne accès qu'à une toute petite fraction d'entre elles. Chaque matin, en me rendant au travail, je passe devant les

mêmes maisons sans jamais remarquer la couleur de leur toit ou le nombre de leurs fenêtres. Assis à mon bureau, tandis que je me concentre sur l'écriture de ce livre, ma rétine est littéralement bombardée d'informations sur les objets, les photos et les œuvres d'art qui m'entourent. Au même instant, mes tympans vibrent d'une superposition de musique, de chants d'oiseaux et de bruits de voisinage. Pourtant, tant que je parviens à rester focalisé sur mon travail, toutes ces distractions restent aux confins de ma conscience : elles n'y ont pas accès.

L'accès à la conscience est à la fois extrêmement ouvert et fortement sélectif. Son répertoire *potentiel* est immense. À tout instant, en réorientant mon attention, je peux prendre conscience d'une couleur, d'un parfum, d'un son, d'un souvenir oublié, d'un sentiment, ou des multiples acceptions d'un mot familier. Si je commets une bourde, il se peut que mon esprit soit envahi d'une vive conscience de mon erreur, puis d'émotions, de regrets et de stratégies pour reprendre le dessus. À un instant donné, pourtant, le répertoire *effectif* de la conscience est minuscule. Chacun de nous ne peut guère penser qu'à une seule chose à la fois (même si cette pensée peut être constituée d'un assemblage de plusieurs éléments, comme lorsque nous réfléchissons au sens d'une phrase).

Cette contenance extrêmement réduite oblige la conscience à se retirer d'un objet avant de pouvoir accéder à un autre. Cessez un instant de lire et prêtez attention à la position de vos jambes. Peut-être ressentez-vous une pression ou une douleur à tel ou tel endroit. Cette sensation vient d'accéder à votre conscience. Une seconde plus tôt, elle n'était que *préconsciente* – accessible, certes, mais pas encore accédée, elle restait dans les limbes de l'inconscient. Cela ne signifie pas qu'elle n'était pas représentée dans votre cerveau : sans en avoir conscience, vous ajustez constamment votre posture en réponse à des signaux inconscients. Cependant, l'accès à la conscience rend cette information disponible à l'ensemble de vos facultés intellectuelles. Soudainement, elle devient accessible par la voie du langage : vous pouvez décrire ce que vous

ressentez. L'information consciente parvient également à votre mémoire, elle guide votre attention, vos intentions et vos plans d'action. C'est précisément ce passage soudain du préconscient au conscient qui fait accéder une information à la conscience et la rend disponible à mille et une opérations mentales, que j'analyserai dans les prochains chapitres. J'essaierai de clarifier ce qui se produit dans le cerveau à cet instant exact : les mécanismes cérébraux de l'accès à la conscience.

Pour ce faire, il nous faudra encore distinguer accès à la conscience et orientation de l'attention – une nuance délicate mais indispensable. En effet, comment définir ce qu'est l'attention ? Dans ses magistraux *Principes de psychologie* (1890), William James en a proposé une définition demeurée célèbre. L'attention, selon lui, est « la prise de possession par l'esprit, sous une forme claire et vive, d'un objet ou d'une suite de pensées parmi plusieurs qui semblent possibles ». Malheureusement, cette définition me semble amalgamer deux notions bien différentes et qui renvoient à des circuits cérébraux distincts : la *sélection* et l'*accès*. « La prise de possession par l'esprit » est essentiellement ce que j'ai appelé l'accès à la conscience : le fait d'amener une information au premier plan de notre pensée, de sorte qu'elle devienne un objet mental conscient que nous « gardons en tête ». Sous cet angle, l'attention se confond avec la conscience, pratiquement par définition : toutes les fois qu'un objet prend possession de notre esprit, si bien que nous puissions le décrire à d'autres (par mots ou par gestes), nous en sommes nécessairement conscients.

Cependant, la définition de l'attention selon William James recouvre également un deuxième processus : le fait d'isoler un objet de pensée parmi d'autres. C'est exactement ce que les sciences cognitives appellent aujourd'hui l'*attention sélective*. À chaque instant, notre environnement regorge de milliers de perceptions potentielles. Notre mémoire, également, fourmille de connaissances qui restent en sommeil. Afin d'éviter la saturation, notre cerveau applique une sévère procédure de sélection. Parmi d'innombrables pensées en puissance, seule « la crème de la crème » parvient à notre conscience. C'est ce filtre très complexe que

nous appelons l'attention. Notre cerveau élimine impitoyablement toutes les informations dépourvues de pertinence pour n'en retenir qu'une seule qui soit la plus saillante ou la plus adaptée à nos buts actuels. Ce stimulus est alors amplifié jusqu'à prendre le contrôle de notre comportement.

Il est évident, dès lors, que la plupart de ces opérations de sélection attentionnelle, sinon toutes, doivent se dérouler en dehors de notre conscience. Comment pourrions-nous avoir la moindre pensée si nous devions, à chaque fois, commencer par passer au crible tous les candidats potentiels ? Le filtre de l'attention agit de façon inconsciente – et c'est pourquoi l'attention est dissociable de l'accès à la conscience. Dans la vie de tous les jours, notre environnement est souvent saturé de stimulations de toutes sortes, et nous utilisons alors l'attention afin de sélectionner celles qui méritent d'accéder à notre conscience. Dans ces conditions, l'attention contrôle la voie d'accès à la conscience[3]. Au laboratoire, cependant, nous parvenons à créer des situations si dépouillées qu'il ne reste qu'une seule information sur l'écran. Dans ce cas, la sélection n'est pas indispensable, et l'attention et la conscience apparaissent dissociées[4]. Inversement, dans d'autres expériences que nous discuterons plus loin, l'attention opère à couvert, en amplifiant ou, au contraire, en filtrant certaines informations sans pour autant qu'aucune d'elles ne finisse par parvenir à la conscience. En bref, l'attention et l'accès à la conscience sont des processus bien distincts.

Un troisième concept demande à être soigneusement mis de côté : la vigilance, que l'on appelle également la « conscience intransitive » ou l'« état de conscience ». En français, l'adjectif *conscient* peut être transitif, comme lorsque nous déclarons être conscients *de* la forme d'un fruit ou de la beauté d'une femme – nous parlons alors de l'accès d'une information précise à la conscience. Mais le mot *conscient* peut également être intransitif, comme lorsque nous disons qu'un blessé est demeuré conscient. Dans ce cas, nous faisons référence à un *état* qui comprend toute une série de degrés. Dans cette acception, la conscience est une

faculté que nous perdons durant le sommeil, lorsque nous nous évanouissons, ou lorsque nous subissons une anesthésie générale.

Afin d'éviter toute confusion, les scientifiques font fréquemment référence à cet état sous le nom d'*état de veille* ou de *vigilance*. Pourtant, même ces deux termes devraient sans doute être soigneusement distingués. Le terme d'éveil, ou d'état de veille, fait principalement référence au cycle veille-sommeil, qui dépend de mécanismes sous-corticaux, tandis que la vigilance renvoie au degré d'activité des réseaux cortico-thalamiques qui sous-tendent l'état de conscience. Ces deux concepts sont bien distincts de l'accès à la conscience. L'état de veille, la vigilance et l'attention ne sont que des conditions préalables à cet accès. Ils sont nécessaires, mais pas toujours suffisants, pour nous rendre conscients d'une information spécifique. Ainsi, certains patients, à la suite d'un petit accident vasculaire dans le cortex visuel, deviennent aveugles à la couleur. Ils sont toujours éveillés et attentifs, leur vigilance est intacte, mais la destruction d'un circuit spécialisé dans la perception de la couleur les empêche de prendre conscience de cet aspect du monde extérieur. Au chapitre 6, nous rencontrerons d'autres patients, en état végétatif, qui semblent n'avoir plus aucune réaction consciente ni aucune prise de conscience de leur environnement, mais ouvrent pourtant les yeux le matin et s'endorment le soir. Leur cycle veille-sommeil est intact, mais leur cerveau, dramatiquement détérioré, ne semble plus capable de se représenter une succession d'états conscients.

Tout au long de ce livre, nous étudierons la question de l'accès : que se passe-t-il dans notre cerveau lorsque nous prenons conscience *de* quelque chose ? Au chapitre 6, cependant, nous reviendrons sur le sens intransitif du mot conscience, et sur les applications croissantes de la science de la conscience aux patients qui souffrent d'un coma, d'un état végétatif ou de troubles apparentés de l'état de conscience.

Mais le mot *conscience* possède encore bien d'autres acceptions. Bon nombre de philosophes et de scientifiques estiment que la conscience, état éminemment subjectif, ne se définit qu'en référence au soi. Comment

pourrions-nous comprendre ce qu'est la perception consciente sans commencer par clarifier ce qu'est ce « je » qui prétend percevoir ? Le neurologue Antonio Damasio définit la conscience comme « le soi dans l'acte même de connaître » (« *the self in the act of knowing* ») – une définition qui implique que l'on ne puisse pas résoudre l'énigme de la conscience si l'on ne décrypte pas ce qu'est un « soi ».

Une même intuition a conduit Gordon Gallup à concevoir son test classique de reconnaissance de soi, qui évalue si un enfant ou même un animal parvient à reconnaître son reflet dans un miroir[5]. La conscience de soi est attribuée à tout organisme qui utilise le miroir pour explorer des endroits cachés de son corps et, par exemple, s'apercevoir qu'une gommette rouge a été subrepticement placée sur son front. Les enfants parviennent à repérer la gommette à l'aide du miroir aux alentours de 18 à 24 mois. Il semblerait que les chimpanzés, les gorilles, les orangs-outans et même quelques dauphins, éléphants et autres pies parviennent à réussir ce test[6] – ce qui a conduit un groupe de scientifiques à énoncer, sans grandes nuances, dans leur déclaration de Cambridge sur la conscience du 7 juillet 2012, que « de nombreuses données indiquent que les humains ne sont pas les seuls à posséder les substrats neurologiques qui engendrent la conscience ».

Une fois de plus, pourtant, la science nous dicte de préciser les concepts. La reconnaissance dans un miroir n'atteste nullement de la conscience. Elle pourrait être accomplie par un dispositif totalement inconscient qui se contenterait de prédire à quoi devrait ressembler notre corps sous un certain angle, et qui ajusterait ses mouvements en comparant cette prédiction avec ce qu'il perçoit dans le miroir – c'est ce que je fais tous les matins, machinalement, lorsque je me rase. Même les pigeons peuvent être conditionnés à passer ce test, au prix d'un entraînement considérable qui les transforme pratiquement en automates du miroir[7]. Il se pourrait que le test du miroir ne mesure que la capacité d'un organisme à apprendre, d'une part, à quoi ressemble son corps et, d'autre part, comment utiliser un miroir pour comparer cette attente avec

la réalité. Ce n'est pas rien, certes, mais rien ne garantit qu'il s'agisse d'un test décisif de la conscience de soi[8].

Même le lien entre la perception consciente et la connaissance de soi n'a rien d'évident. Assister à un concert ou contempler un magnifique coucher de soleil peuvent très bien me mettre dans un état particulier de conscience, sans pour autant me forcer à penser que c'est *moi* qui éprouve ce plaisir-là. Dans de telles circonstances, mon corps et mon moi se font oublier, un peu comme une basse continue ou une lumière douce et constante. Ce ne sont que des cibles potentielles de mon attention, qui restent le plus souvent non conscientes, mais que je peux amener à la conscience dès que je le souhaite. Selon moi, la conscience de soi n'est guère différente de la conscience d'un son ou d'une couleur. Prendre conscience d'un aspect de moi-même n'est probablement qu'une forme particulière d'accès à la conscience, dans laquelle l'information à laquelle j'accède concerne l'une des multiples représentations mentales de mon « moi » – mon corps, mon comportement, mes sentiments ou mes pensées.

Ce qui est tout de même fascinant à propos de la conscience de soi, c'est qu'elle paraît impliquer une « boucle étrange[9] », une circularité, voire une contradiction. Quand je me penche sur moi-même, le « je » apparaît deux fois : celui qui perçoit se confond avec celui qui est perçu. Comment opère cette forme récursive d'une conscience qui s'examine elle-même ? Nous touchons ici à une nouvelle acception de la conscience, que les spécialistes de sciences cognitives appellent la *métacognition* : la capacité de réfléchir à ses propres états mentaux. Le philosophe positiviste Auguste Comte (1798-1857) la considérait comme une impossibilité logique : « L'individu pensant, écrivait-il, ne saurait se partager en deux, dont l'un raisonnerait, tandis que l'autre regarderait raisonner. L'organe observé et l'organe observateur étant, dans ce cas, identiques, comment l'observation pourrait-elle avoir lieu[10] ? »

Comte avait tort. Comme le lui fit immédiatement remarquer le philosophe britannique John Stuart Mill (1806-1873), le paradoxe s'évanouit sitôt que l'observateur et l'observé habitent des temps différents

ou font appel à des systèmes distincts. « Il aurait pu venir à l'esprit de M. Comte qu'il est possible d'étudier un fait par l'intermédiaire de la mémoire, non pas à l'instant même où nous le percevons, mais dans le moment d'après. (…) Nous réfléchissons sur ce que nous avons fait quand l'acte est passé, mais quand l'impression en est encore fraîche dans la mémoire. (…) Ce simple fait détruit l'argument entier de M. Comte[11]. »

Effectivement, un système cérébral peut fort bien évaluer la performance d'un autre. Nous le faisons très souvent, par exemple lorsque nous éprouvons le phénomène du « mot sur le bout de la langue » (nous savons que nous devrions le savoir), lorsque nous détectons une erreur dans notre raisonnement (nous savons que nous nous trompons), ou lorsque nous nous lamentons d'avoir échoué à un examen (nous savons que nous avions étudié, nous croyions avoir tout retenu, mais nous comprenons que nous nous sommes trompés). Certaines aires du cortex préfrontal surveillent le déroulement de nos plans d'action, allouent un niveau de confiance à nos décisions et détectent nos erreurs. En lien étroit avec notre mémoire à long terme et notre imagination, elles concoctent un soliloque endogène qui alimente nos réflexions sur nous-mêmes. Le mot même de « réflexion » évoque ce « miroir de l'âme » qui permet à certaines de nos aires cérébrales de « représenter » et d'évaluer l'opération des autres.

Les scientifiques aiment réduire les problèmes à leur forme la plus simple. Nous allons commencer par aborder l'acception la plus élémentaire du mot « conscience », c'est-à-dire l'accès à la conscience, ou la manière dont nous prenons conscience d'une information donnée. Reléguons à plus tard les questions épineuses du soi et de la conscience récursive. Le premier ingrédient de notre science de la conscience consiste à nous concentrer sur la question de l'accès conscient, en prenant grand soin de la distinguer des notions d'attention, d'éveil, de conscience de soi et de métacognition[12].

La méthode contrastive

Le second ingrédient qui rend possible une science de la conscience est l'existence d'une panoplie de manipulations expérimentales du contenu de la conscience. Dès les années 1990, les psychologues cognitifs se sont rendu compte qu'ils pouvaient jouer avec la conscience en contrastant, à volonté, les états conscients et non conscients du cerveau. Il leur était facile de rendre invisibles des objets, des mots et même des films entiers. Qu'arrivait-il à toutes ces images ? En cernant les contours de l'inconscient, l'expérience pouvait aussi révéler, comme en négatif, le propre de la conscience. Combinée avec les méthodes révolutionnaires de l'imagerie cérébrale, cette idée simple a conduit à une nouvelle stratégie expérimentale d'étude des mécanismes cérébraux de la conscience.

En 1989, le psychologue Bernard Baars, dans un livre important ambitieusement intitulé *Une théorie cognitive de la conscience*[13], a fait remarquer qu'il existait des dizaines d'expériences capables de révéler la nature de la conscience. Il ajoutait une observation cruciale : bon nombre de ces expériences fournissaient des « contrastes minimaux », c'est-à-dire deux conditions expérimentales minimalement différentes, mais dont une seule est perçue consciemment. De telles situations sont idéales parce qu'elles permettent aux scientifiques de traiter la conscience comme n'importe quelle autre variable expérimentale, et de la faire varier tout en gardant les autres paramètres pratiquement constants. En concentrant leurs recherches sur de tels contrastes minimaux, et en essayant de comprendre ce qu'ils changent dans le cerveau, il devenait possible de s'abstraire des étapes communes aux opérations conscientes et non conscientes, et de se focaliser exclusivement sur les événements cérébraux qui les distinguent.

Prenons l'exemple d'une personne qui apprend à taper à la machine. Dans les premières semaines, elle est lente, attentive et consciente de chacun de ses gestes. Avec l'entraînement, sa frappe s'automatise tant et si bien qu'elle finit par taper un texte machinalement, tout en parlant ou en

pensant à autre chose. Son comportement est toujours le même, mais il échappe à présent à la conscience, au point qu'elle est bien incapable d'énoncer la position des touches sur le clavier. Voilà un contraste minimal riche d'enseignements : le même comportement peut être effectué consciemment ou non. Il se trouve que ce contraste, en imagerie cérébrale, fait apparaître un vaste réseau cortical qui inclut plusieurs régions préfrontales et pariétales – celles-là mêmes qui, nous le verrons, s'activent systématiquement dès que survient l'accès à la conscience [14].

La transition inverse, de l'inconscient au conscient, est tout aussi facile à étudier. La perception visuelle offre de nombreuses opportunités de faire entrer et sortir n'importe quelle image de la conscience. L'illusion avec laquelle s'est ouvert ce chapitre en offre un excellent exemple (figure 3). Pourquoi les points gris disparaissent-ils de notre vue ? Le mécanisme n'est pas encore parfaitement compris, mais l'idée est que notre système visuel interprète la constance de l'image comme un indice que ces points ne viennent pas de l'environnement, mais d'anomalies de l'œil [15]. Lorsque nous gardons le regard fixe, chaque point apparaît comme une tache de gris strictement immobile sur notre rétine – et au bout d'un moment, notre système visuel décide de se débarrasser de ce qui n'est manifestement qu'une salissure. Notre perte de conscience des points gris résulterait ainsi d'un mécanisme évolué qui élimine les défauts de nos yeux. De fait, notre rétine foisonne d'imperfections diverses, tels que des vaisseaux sanguins qui passent devant nos photorécepteurs, que nous devons apprendre à attribuer à nos yeux plutôt qu'à notre environnement. Imaginez à quel point il serait horrible d'être constamment distrait par des volutes sanglantes qui nous barrent la vue ! Notre système visuel interprète l'immobilité parfaite d'un objet comme l'indice d'une anomalie du capteur. Il l'efface alors, comme le ferait un utilisateur du logiciel Photoshop, en remplissant la partie manquante de l'image par la texture voisine. (Au passage, ce mécanisme de remplissage explique aussi que nous ne détections pas notre « tache aveugle », l'endroit de notre rétine d'où part le nerf visuel et qui est donc dépourvu de photorécepteurs.) Dès

que nous bougeons les yeux, même très peu, les taches se déplacent sur notre rétine, ce qui permet à notre système visuel de se rendre compte qu'elles proviennent de l'environnement – et immédiatement, il les fait remonter à la conscience.

La disparition des taches immobiles n'est qu'une illusion parmi des dizaines d'autres qui permettent d'étudier la transition entre le conscient et l'inconscient. Passons à présent en revue les principaux outils qui figurent dans la panoplie du cogniticien.

Images rivales

Historiquement, l'un des premiers contrastes qui a permis l'analyse de la perception consciente est la « rivalité binoculaire », le très curieux duel qui se produit au sein de notre cerveau lorsque deux images distinctes se présentent à chacun de nos yeux.

Première observation intéressante : nous n'avons aucune conscience d'avoir deux yeux en mouvement constant. Notre cerveau nous donne à voir un monde tridimensionnel stable, mais il nous cache l'étonnante complexité des opérations par lesquelles il parvient à cette conclusion. À chaque instant, l'œil droit reçoit une image légèrement différente de celle que reçoit l'œil gauche. Cependant, nous ne voyons pas double ! La plupart du temps, nous ne remarquons pas la présence de deux images, mais nous les fusionnons en une seule scène homogène. Notre cerveau tire pourtant avantage de l'espacement de nos deux yeux, qui induit un léger déplacement horizontal d'une image par rapport à l'autre. Comme l'a découvert le scientifique anglais Charles Wheatstone en 1838, cette « disparité binoculaire » nous permet de localiser les objets en profondeur – c'est ainsi que nous éprouvons le vif sentiment de la troisième dimension.

Mais que se passerait-il, s'interrogea Wheatstone, si chaque œil recevait une image complètement différente, par exemple celle d'un visage dans l'œil droit et celle d'une maison dans l'œil gauche ? Notre cerveau continuerait-il à les fusionner ? Verrions-nous deux images en même temps ?

Pour le savoir, Wheatstone construisit un appareil qu'il nomma « stéréoscope » (durant toute la période victorienne et bien au-delà, ce gadget lança la mode des images 3D de paysages, de monuments… et de pornographie). Deux miroirs permettaient de présenter des images distinctes à chaque œil (figure 4). À sa grande surprise, Wheatstone découvrit que, dès que les images étaient incompatibles, la vision devenait instable. Au lieu de fusionner les deux images en une seule, la perception ne cessait d'alterner. Pendant quelques secondes, le visage apparaissait. Puis celui-ci se dissolvait, et en quelques instants, cédait la place à la maison ; et ainsi de suite, en une alternance incessante, entièrement concoctée par le cerveau. Comme le remarquait Wheatstone, « il ne semble pas être en mon pouvoir de déterminer l'apparition de l'une ou de l'autre image ». Le cerveau, confronté à un stimulus artificiel et dépourvu de toute plausibilité, balance spontanément entre deux interprétations : le visage et la maison. Les deux images semblent se disputer la primeur de la conscience – d'où le terme de *rivalité binoculaire*.

Pour l'expérimentateur, la rivalité binoculaire est un phénomène idéal : une pure manipulation de la conscience subjective. En effet, le stimulus est constant, et pourtant l'observateur rapporte que sa vision subjective ne cesse de changer. Au fil du temps, la même image, bien qu'elle soit fixe, change de statut : tantôt elle est visible, tantôt elle disparaît de la conscience. Que se passe-t-il alors ? En enregistrant les neurones du cortex visuel du singe macaque, les neurophysiologistes David Leopold et Nikos Logothetis ont été les premiers à mesurer le trajet d'une image vue ou pas vue au sein du cerveau[16]. Ils ont commencé par entraîner les singes à rapporter ce qu'ils voyaient, à l'aide d'un levier à deux positions (visage ou maison). Ils ont alors montré que les singes

percevaient une alternance aléatoire des deux images, exactement comme le décrit un observateur humain. Enfin, ils ont mesuré l'activité de neurones individuels, tandis que leur image préférée apparaissait puis disparaissait de la conscience. Les résultats étaient nets. À l'étape la plus précoce, au sein du cortex visuel primaire qui est le point d'entrée de la vision dans le cortex, la plupart des cellules ne faisaient que refléter le stimulus objectif : leur activité ne dépendait que des images présentes, et elle ne se modifiait pas lorsque l'animal rapportait que sa perception avait changé. Plus avant dans le système visuel, au sein des « aires visuelles supérieures » tels l'aire V4 et le cortex inféro-temporal, de plus en plus de neurones s'accordaient avec ce que percevait le singe : ils déchargeaient avec ardeur lorsque le singe déclarait percevoir leur image préférée, et beaucoup moins, voire pas du tout, lorsque cette image s'effaçait de la conscience. Leur réponse était, littéralement, le premier aperçu d'un corrélat neuronal de l'expérience consciente (figure 4).

Aujourd'hui encore, la rivalité binoculaire demeure un moyen privilégié d'accès à la machinerie neuronale qui sous-tend la conscience. Elle a inspiré des centaines d'expériences et de nombreuses variantes. Par exemple, à l'aide d'une méthode appelée « suppression continue par des flashs », il est devenu possible d'éliminer totalement l'une des deux images de la conscience. Il suffit de montrer à un œil toute une série de rectangles colorés qui changent en permanence : seul ce flux incessant est vu, et il éteint totalement ce qui est présenté à l'autre œil, même s'il s'agit d'une phrase ou d'un film[17].

Que nous apprennent ces illusions binoculaires ? Qu'une image peut être présente à nos yeux pendant un temps considérable, et pénétrer profondément dans les aires visuelles du cerveau, sans pour autant être vue. La rivalité binoculaire indique ainsi que ce qui compte pour l'accès à la conscience, ce n'est pas le traitement visuel initial (où les images sont toutes deux traitées en parallèle), mais une étape ultérieure d'où émerge une seule image gagnante. Parce que notre conscience ne peut imaginer que deux objets distincts occupent le même lieu, notre cerveau est le siège

d'une féroce compétition. À notre insu, ce ne sont pas seulement deux images qui se disputent l'avant-scène, mais des myriades d'interprétations possibles de nos entrées sensorielles qui entrent en concurrence, et dont une seule parviendra à conquérir notre esprit. La rivalité est une bonne métaphore de ce combat incessant pour l'accès à la conscience.

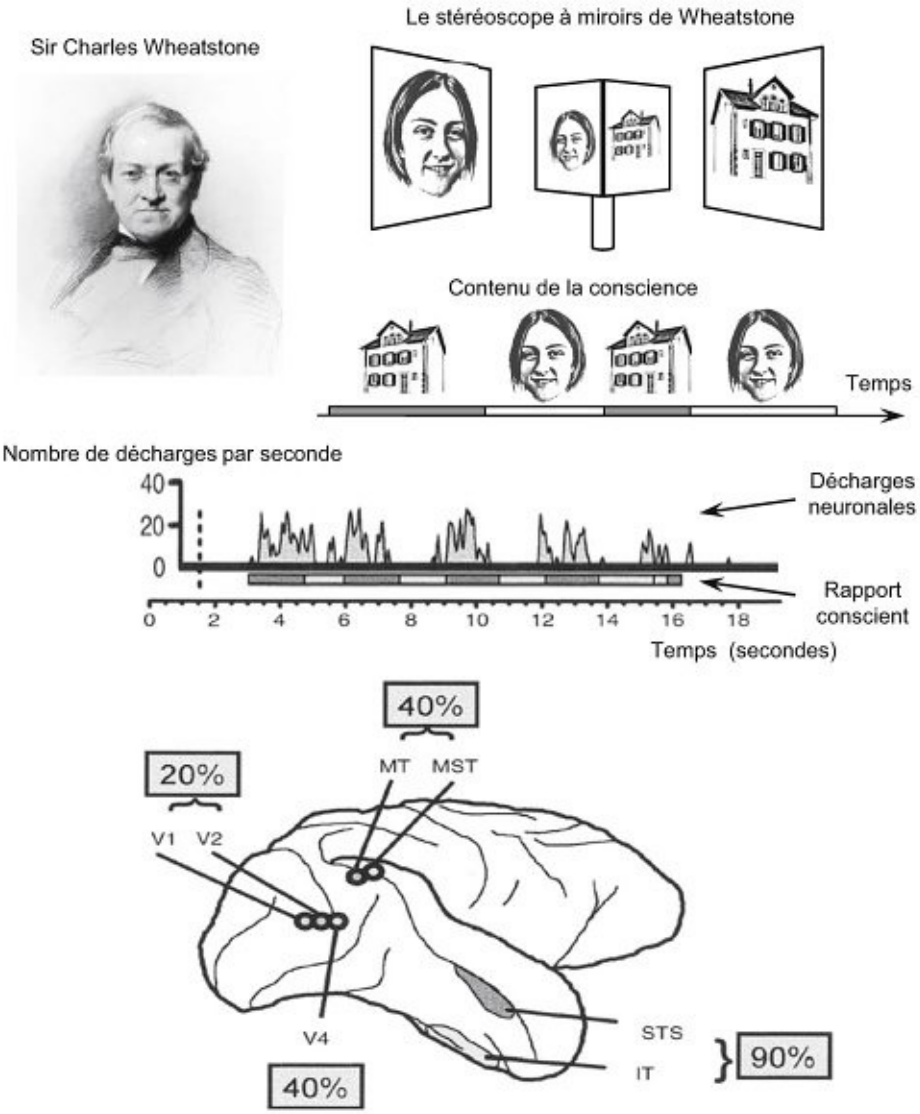


Figure 4. La rivalité binoculaire est une illusion visuelle découverte par Charles Wheatstone en 1838. Elle consiste à présenter une image différente à chaque œil (par exemple un visage et une maison). Au lieu de fusionner les deux images, notre cerveau les fait alterner : nous voyons

tantôt le visage, tantôt la maison, mais jamais les deux en même temps. Nikos Logothetis et David Leopold ont entraîné des singes à décrire ce qu'ils voyaient en bougeant une manette. Ils ont ainsi montré que les singes sont soumis à la même illusion, et ont alors enregistré l'activité des neurones de leur cerveau. Dans les aires visuelles les plus précoces (V1 et V2), la plupart des neurones répondent à la présence objective des images. Dans les aires de plus haut niveau, par contre, de nombreux neurones corrèlent avec la conscience subjective : leurs décharges indiquent quelle image est perçue à chaque instant. Le pourcentage indique la fraction de neurones qui codent pour le contenu subjectif de la conscience dans chaque région du cortex. Cette étude pionnière suggère que la perception consciente repose sur les régions associatives plutôt que sur les aires sensorielles du cortex.

Lorsque l'attention cligne

La rivalité est-elle un processus passif ? Ou bien pouvons-nous choisir l'image que nous souhaitons voir ? Lorsqu'on regarde dans un stéréoscope, on a l'impression subjective d'être passif et soumis à une alternance incessante d'images imposées. Cependant, il s'agit d'une illusion. En réalité, l'attention joue un rôle important dans la compétition corticale entre les deux images. D'abord, lorsque nous essayons de prêter attention à l'une des images, sa perception dure un peu plus longtemps[18]. Cet effet, cependant, n'est pas énorme : le combat entre les deux images démarre dans des profondeurs cérébrales qui sont au-delà de notre contrôle conscient.

Plus important est le fait que l'existence même d'un gagnant dépend de l'attention que nous lui accordons. L'arène même où se joue la compétition, c'est la conscience – et lorsque l'attention se détourne des images rivales, leur combat cesse, et leur alternance aussi[19].

Vous vous demandez peut-être comment nous le savons. Demander à quelqu'un quelles images il voit lorsqu'il n'y fait pas attention, et si ces images changent ou non, paraît contradictoire : pour le savoir, il faudrait qu'il y prête attention. À première vue, déterminer ce que l'on perçoit sans y faire attention semble une mission impossible, un peu comme essayer de

voir ses yeux bouger dans un miroir : nos yeux bougent tout le temps, mais le simple fait de les regarder dans le miroir les oblige à rester immobiles. Ainsi, espérer étudier si la rivalité binoculaire dépend de l'attention a longtemps paru aussi déraisonnable que de demander quel bruit fait la chute d'un arbre lorsque personne n'est là pour l'entendre, ou ce qu'on ressent au moment même où l'on s'endort.

La science, pourtant, accomplit souvent l'impossible. Récemment, Peng Zhang et ses collaborateurs de l'Université du Minnesota se sont rendu compte qu'ils n'avaient pas besoin de demander à l'observateur si les images alternaient toujours lorsqu'il n'y faisait plus attention[20]. Il suffisait d'identifier des marqueurs cérébraux de la rivalité, qui montreraient si, oui ou non, les images étaient encore en compétition. La recherche avait déjà montré que, pendant la rivalité, les neurones qui codent pour l'une et l'autre des images déchargent en alternance (figure 4) – pourrait-on détecter cette alternance en l'absence d'attention ? Oui, si les images sont « étiquetées » parce qu'elles scintillent à des rythmes différents. Ces deux fréquences ressortent de l'électroencéphalogramme que l'on peut aisément enregistrer à l'aide de deux électrodes placées sur la tête. Durant la rivalité, les deux fréquences s'excluent l'une l'autre : lorsqu'une oscillation augmente, l'autre diminue, en correspondance directe avec le fait que l'observateur ne perçoit que l'une des deux images à la fois. Or, dès que l'observateur cesse d'y prêter attention, ces alternances cessent, et les deux rythmes se mêlent indifféremment : l'inattention fait cesser la rivalité.

Une autre expérience confirme cette conclusion : si l'on détourne l'attention pendant une durée déterminée, puis que l'on demande à l'observateur laquelle des deux images est à présent visible, ce qu'il décrit n'est pas compatible avec la poursuite de la rivalité : tout se passe comme si les deux images avaient cessé d'alterner pendant la période de distraction[21]. Ainsi, la rivalité binoculaire dépend de l'attention : en son absence, les deux images se superposent sans entrer en compétition. La rivalité requiert la présence d'un observateur conscient et attentif.

L'attention impose ainsi une limite précise au nombre d'images que nous parvenons à traiter simultanément. Cette limite, à son tour, cause d'autres illusions de conscience. L'une d'elles, appelée « clignement attentionnel », se manifeste par une brève période d'invisibilité lorsque l'esprit est temporairement saturé d'informations[22]. La figure 5 montre les conditions typiques dans lesquelles cet effet survient. On présente une série de symboles, l'un après l'autre, à grande vitesse, sur un écran d'ordinateur. La plupart sont des chiffres, mais parfois se glisse une lettre. Il est facile de se souvenir de la première lettre qui apparaît, mais si une deuxième lettre surgit moins d'une demi-seconde après la première, elle a toutes les chances d'être manquée. L'observateur dit n'avoir vu qu'une seule lettre, la première, et il est très étonné d'apprendre qu'il y en a eu deux. Le simple fait de prêter attention à la première lettre engendre un « black-out » qui annihile toute possibilité de percevoir la seconde, un peu comme si l'observateur clignait, non pas des paupières, mais de l'attention – d'où le terme de « clignement attentionnel ».

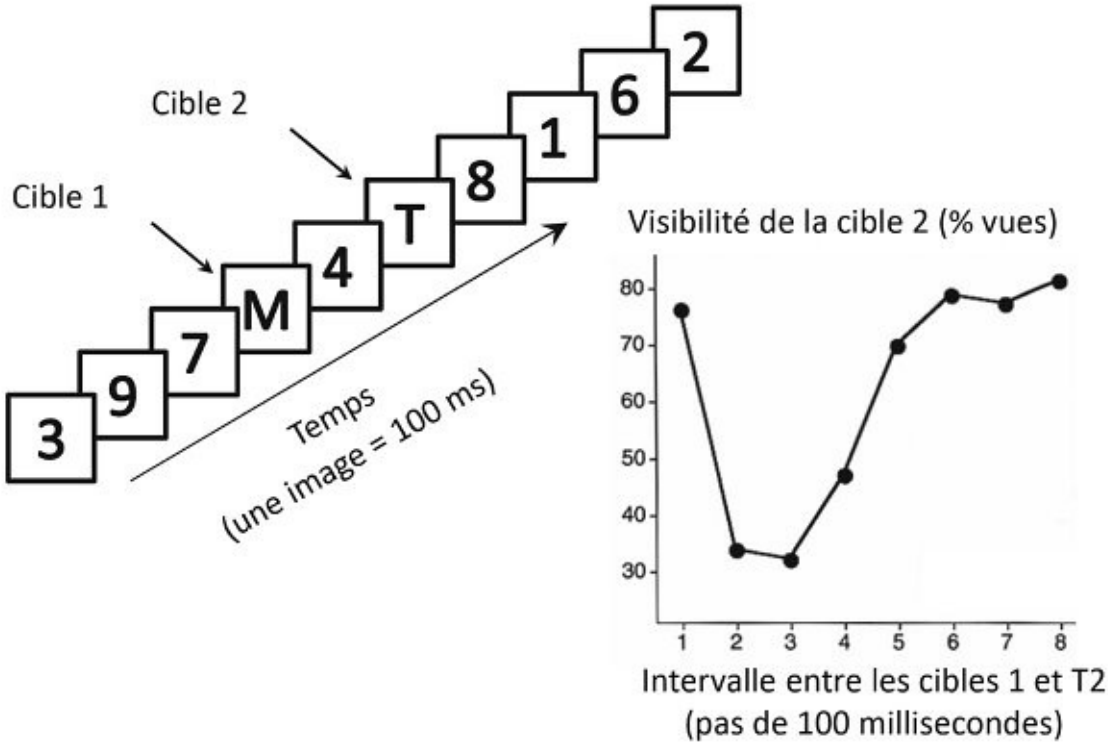


Figure 5. Le clignement attentionnel met en évidence les limites de notre conscience. Lorsque nous voyons une série de chiffres, parfois entrecoupée de lettres, nous n'éprouvons aucune difficulté à identifier la première lettre (ici un M), mais nous ratons la seconde (ici un T). Pendant que nous enregistrons la première lettre en mémoire, nos réseaux cérébraux sont occupés et ne parviennent pas à prendre conscience de la seconde cible – un peu comme si nous clignions, non pas de l'œil, mais de l'attention.

En analysant ce phénomène grâce à l'imagerie cérébrale, on constate que toutes les lettres, conscientes ou non, pénètrent profondément dans le cerveau. Elles entrent dans le système visuel, et atteignent même le point où elles sont reconnues comme des cibles : une partie du cerveau « sait » que ce sont des lettres et qu'il faudrait essayer de les retenir[23]. Pourtant, cette information n'accède pas à la conscience. Pour être perçu, il faut qu'un symbole atteigne une étape de traitement cérébral qui l'imprime dans la conscience[24]. Or cette étape agit comme un goulot

d'étranglement : à chaque instant, une seule information peut y pénétrer à la fois – tous les autres éléments de la scène visuelle doivent attendre.

Tandis que la rivalité binoculaire met en évidence la compétition entre deux images présentées au même instant, le clignement attentionnel démontre qu'une concurrence du même ordre survient entre des images présentées l'une après l'autre. Notre conscience ne parvient pas à suivre le rythme rapide de présentation des lettres sur l'écran. Nous croyons « voir » tous les symboles lorsque nous les observons passivement, mais le simple fait d'en mettre un en mémoire bloque nos ressources pendant un temps suffisant pour nous rendre aveugles aux symboles suivants. La forteresse de la conscience possède un chemin d'accès si étroit qu'il force les images à entrer en compétition. L'accès à la conscience impose un goulot d'étranglement.

Le lecteur objectera peut-être que nous parvenons parfois à retenir deux lettres successives (environ un tiers du temps à en croire les données de la figure 5). Dans la vie quotidienne, nous ne semblons pas avoir de difficulté à percevoir deux choses presque au même moment. Nous pouvons, par exemple, regarder un film sans pour autant devenir sourd au klaxon qui monte de la rue. Les psychologues parlent de « double tâche », parce qu'on demande à la personne de faire deux choses à la fois. Que se passe-t-il alors ? Le fait que nous puissions réaliser certaines doubles tâches réfute-t-il l'idée que la conscience se limite à un seul objet de pensée à la fois ? Pas du tout. Même dans cette situation, la recherche montre que notre conscience est réduite. En réalité, nous ne pensons jamais vraiment simultanément à deux idées distinctes. Lorsque nous essayons de faire deux choses à la fois, l'impression que nous avons d'être parfaitement en phase, au même instant, avec deux aspects du monde extérieur n'est qu'une illusion. En vérité, notre esprit ne les perçoit pas simultanément. L'un d'entre eux accède à la conscience en premier, tandis que l'autre doit attendre que l'espace conscient soit libéré.

Le goulot d'étranglement de notre esprit conscient entraîne alors un ralentissement considérable, que l'on appelle à juste titre la « période

réfractaire psychologique[25] ». Tandis que l'esprit se focalise consciemment sur le premier objet, il devient réfractaire à toute autre idée, et donc très lent à les traiter. Tant que dure l'analyse du premier, le second stationne dans une mémoire inconsciente. Il y demeure tant que le traitement du premier élément n'est pas terminé.

Nous n'avons aucune conscience de cette longue période d'attente inconsciente. Comment pourrait-il en être autrement ? Quand notre conscience est occupée ailleurs, nous n'avons aucune autre possibilité de prendre du recul et de nous rendre compte que le traitement d'un second élément prend du retard. Ainsi, quand l'esprit vagabonde, la perception subjective du temps se déforme[26]. Si nous nous engageons pleinement dans une première tâche et qu'on nous demande d'estimer à quel moment un deuxième stimulus apparaît, nous n'en avons pas la moindre idée, et nous souffrons d'une illusion temporelle systématique : nous confondons le moment de son apparition avec celui où nous en prenons conscience (souvent bien plus tardif). Même quand deux images apparaissent en même temps, leur simultanéité ne va pas de soi : celle à laquelle nous faisons attention en premier semble apparaître plus tôt que la seconde. En vérité, ce délai n'est dû qu'à la lenteur de notre conscience.

Le clignement attentionnel et la période réfractaire sont des phénomènes psychologiques apparentés. Dès que la conscience est occupée par une idée, les autres candidats à l'accès conscient doivent attendre dans un espace inconscient (c'est la période réfractaire), et cette attente les fragilise au point qu'à tout instant ils risquent de s'effacer (c'est le clignement attentionnel). Plus la distraction s'accroît, plus la période réfractaire s'allonge, et plus la probabilité d'un black-out total augmente[27].

Dans la plupart des expériences, le clignement attentionnel ne dure qu'une fraction de seconde. Mettre en mémoire la première lettre ne nous prend qu'un instant, pendant lequel les autres lettres deviennent invisibles à nos yeux. Qu'en va-t-il, cependant, si la première tâche se complique ? Étonnamment, il se peut que nous perdions toute conscience du monde

extérieur. Les grands lecteurs, les joueurs d'échecs ou les mathématiciens le savent bien : une activité intellectuelle intense annihile la conscience de l'environnement et du temps qui passe. Les psychologues appellent ce phénomène la « cécité attentionnelle », et il est facile de le mettre en évidence au laboratoire. Au cours d'une expérience[28], on demande aux participants de fixer le centre d'un écran d'ordinateur, tout en se concentrant sur le haut de l'écran, où va apparaître une lettre qu'ils devront mémoriser. On leur propose deux essais afin qu'ils s'entraînent, et au troisième, sans les prévenir, en même temps que la lettre du haut, on fait apparaître une image au centre de l'écran. Peu importe qu'il s'agisse d'une tache noire, d'un chiffre ou d'un mot : deux tiers des participants ne la remarquent pas, même si elle dure près d'une seconde. Ils disent n'avoir vu que la lettre en haut de l'écran et rien d'autre. Lorsqu'on leur passe une seconde fois la séquence qu'ils ont vue, ils sont stupéfaits d'avoir manqué un événement aussi évident. En bref, l'inattention engendre l'invisibilité.

Figure 6. L'inattention peut entraîner une cécité temporaire. Le simple fait de prêter attention à une chose nous empêche d'en percevoir d'autres. Dans le film classique du gorille (en haut), des étudiants en T-shirt blanc se lancent un ballon de basket, et on demande au spectateur de compter le nombre de passes. La focalisation de l'attention sur l'équipe blanche suffit à rendre invisible l'entrée en scène d'un acteur déguisé en gorille, bien qu'il se frappe la poitrine en plein

centre de l'écran ! Dans un petit film policier (en bas), pas moins de vingt et un éléments de la scène du crime changent à notre insu. Au vu de ces expériences, on peut se demander combien de « gorilles dans la brume » disparaissent ainsi de notre vie quotidienne, sans que nous en ayons la moindre conscience…

La psychologie confine parfois à la magie. Connaissez-vous l'extraordinaire démonstration du « gorille invisible » de Dan Simons et Christopher Chabris[29] (figure 6) ? Un film montre deux équipes d'étudiants, l'une vêtue de blanc, l'autre de noir, qui jouent au basket. On demande simplement aux participants de compter le nombre de passes de l'équipe blanche. Le film dure 30 secondes, et avec un peu de concentration, tout le monde parvient à compter quinze passes. C'est alors que l'expérimentateur leur demande : « Avez-vous vu le gorille ? » Un gorille ? Quel gorille ? On repasse le film, et les spectateurs constatent avec stupéfaction qu'au beau milieu, un homme déguisé en gorille entre sur le terrain, se frappe vigoureusement la poitrine, puis disparaît. L'immense majorité des gens ne le voient pas, et sont prêts à jurer qu'il n'était pas là lors du premier visionnage. Ils en sont tellement sûrs qu'ils accusent l'expérimentateur d'avoir changé de film ! Le simple fait de se concentrer sur l'équipe blanche abolit la perception d'un personnage en noir, même s'il est aussi inopiné qu'un gorille de belle taille.

L'expérience du gorille invisible a fait date. Au même moment, d'autres chercheurs en psychologie cognitive découvraient des dizaines d'autres situations dans lesquelles l'inattention induisait une cécité temporaire. Les observateurs, même attentifs, s'avéraient de pitoyables témoins. Des manipulations simplissimes pouvaient rendre invisibles les aspects les plus saillants d'une image. Kevin O'Regan et Ron Rensink ont découvert la « cécité au changement[30] », une étonnante incapacité de détecter qu'une partie d'une image a été effacée : ils faisaient alterner deux versions de la même image, identiques à une exception près, et séparées par un bref écran blanc. Les observateurs étaient prêts à jurer que les deux images étaient identiques – même quand le changement était énorme (un avion à

réaction perdait l'un de ses moteurs) ou d'une importance capitale (dans une scène de conduite automobile, la ligne en pointillé au milieu de la route devenait continue).

Dan Simons, quant à lui, a démontré que la cécité attentionnelle pouvait affecter la vie réelle. Sur le campus de Harvard, un acteur demandait son chemin à un étudiant. Leur conversation était interrompue par des déménageurs qui passaient entre eux en portant une grande planche de bois… ce qui permettait de substituer au premier acteur une autre personne dont les traits, la coiffure et le costume changeaient du tout au tout. La plupart des étudiants ne remarquaient rien.

Une étude encore plus insolite est celle menée par Peter Johansson sur la cécité à nos propres choix[31]. Dans cette expérience, on montre à un étudiant deux photos de jeunes femmes, et on lui demande de choisir celle qu'il préfère. On lui montre ensuite la photographie qu'il a choisie… sauf que l'expérimentateur, par un tour de passe-passe, a interverti les deux photos. Ainsi, le participant se retrouve avec la photo du visage qu'il n'a *pas* choisi. Non seulement la moitié des participants ne remarquent rien, mais ils se mettent à expliquer, avec force détails, en quoi ce visage est évidemment plus attrayant que l'autre !

Pour éprouver les effets de la cécité attentionnelle, connectez-vous à YouTube et cherchez *Whodunnit ?,* un petit film policier réalisé par le Département des transports londoniens[32]. On y voit un émule du commissaire Hercule Poirot interroger trois suspects et arrêter l'un d'entre eux. Rien d'anormal… jusqu'à ce qu'on rembobine le film, que la caméra recule, et que vous vous rendiez soudain compte que d'énormes anomalies vous ont échappé : sous vos yeux aveugles, en moins d'une minute, pas moins de 21 détails de la scène ont été escamotés. Cinq assistants ont échangé les meubles, remplacé un ours empaillé par une armure médiévale, aidé les acteurs à se changer et troqué les objets qu'ils tenaient dans leurs mains. Un spectateur non averti ne remarque rien !

Cette impressionnante démonstration de cécité au changement se conclut par une réflexion signée du maire de Londres : « Bien des choses

nous échappent lorsque nous n'y prêtons pas attention. Sur la route, cette erreur pourrait être fatale – faites attention aux cyclistes ! » Monsieur le maire a bien raison : des études en simulateur de vol ont montré que, distrait par une conversation avec la tour de contrôle, un pilote de ligne aguerri pouvait devenir aveugle au monde extérieur au point d'engager une procédure d'atterrissage sur une piste occupée par un autre avion…

La morale est claire : en l'absence d'attention, n'importe quel objet peut disparaître de notre conscience. Le détournement de l'attention fournit ainsi une méthode remarquable pour rendre une image inconsciente.

La perception avance masquée

Au laboratoire, l'inattention présente des difficultés pratiques. En effet, c'est un phénomène fuyant. La plupart des observateurs dont on distrait l'attention ne voient rien la première fois, mais, dès le second essai, devenus vigilants, ils remarquent souvent la manipulation, et l'effet d'inattention s'évanouit. Voilà pourquoi un magicien refuse si souvent de refaire deux fois le même tour !

De plus, même si l'inattention induit une puissante sensation d'invisibilité, les scientifiques rencontrent des difficultés à démontrer sans appel cette absence de conscience. Questionner les observateurs après chaque essai est bien sûr possible, mais cette procédure ralentit l'expérience, et surtout elle attire l'attention, ce qui gâche tout. On peut également attendre la fin de l'expérience, mais se pose alors le problème de l'oubli : après quelques dizaines de minutes, les observateurs sous-estiment souvent ce dont ils ont été conscients.

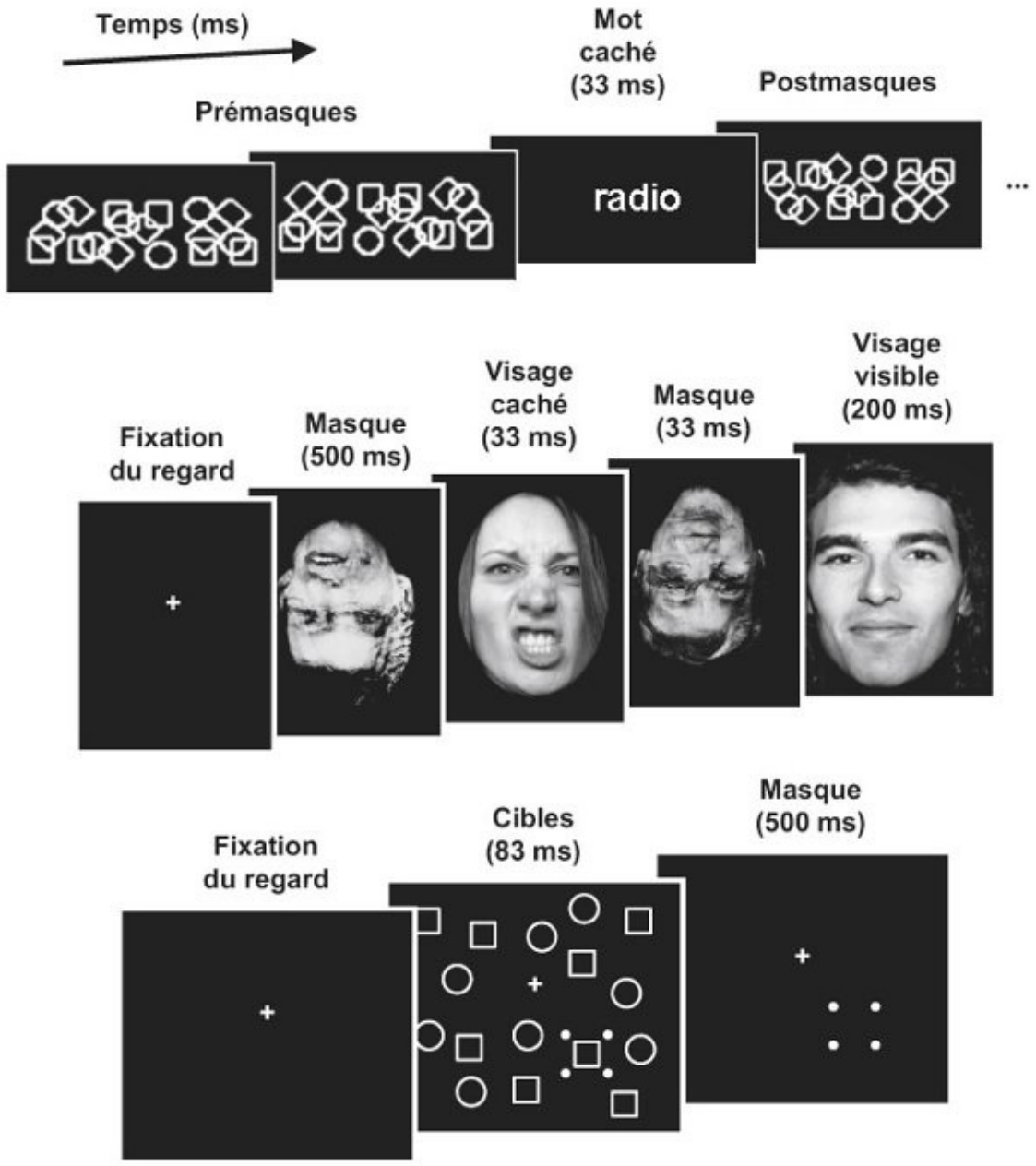


Figure 7. Le masquage peut rendre une image invisible. Cette technique consiste à présenter une image brièvement et à l'entourer de « masques » qui empêchent de la voir. Un mot demeure ainsi invisible lorsqu'il est présenté au milieu de formes géométriques (en haut). Un visage, même porteur d'une émotion forte, disparaît lorsqu'il est entouré d'images aléatoires : seuls ces masques et le visage final sont perçus consciemment (au milieu). Enfin, parmi un ensemble de formes, seule celle qui est entourée de quatre points demeure invisible. Parce qu'ils durent plus longtemps que l'écran de départ, les points masquent la forme qu'ils entourent (en bas).

Certains chercheurs vont jusqu'à suggérer que, dans la cécité au changement, les participants prennent toujours conscience de l'ensemble de l'image[33]. Selon cette hypothèse, s'ils ne détectent pas les changements, c'est parce que la plupart des détails ne s'impriment pas dans leur mémoire, en sorte qu'ils ne peuvent pas comparer l'ancienne image avec la nouvelle. Dès qu'on élimine les indices de mouvement, même un très bref délai empêcherait le cerveau de comparer les deux images. Par défaut, l'observateur répondrait qu'elles ne diffèrent en rien. Il en percevrait bien tous les détails, mais sa mémoire défaillante l'empêcherait de les comparer entre eux.

Il me paraît peu probable qu'un défaut de mémoire explique tous les phénomènes d'inattention et de cécité au changement – après tout, un gorille sur un terrain de basket ou un ours empaillé sur une scène de crime devraient constituer des événements mémorables. Cependant, le doute demeure, et sur le plan scientifique, si nous voulons comprendre ce qu'est la conscience, nous avons besoin de maîtriser très précisément l'état de conscience de nos sujets. Il serait idéal de disposer d'une procédure qui rende les images absolument invisibles – quels que soient le degré d'information des participants, leurs efforts pour les discerner et le nombre de fois qu'ils les examinent.

Fort heureusement, une telle procédure existe. Les psychologues l'appellent le « masquage », le reste du monde les « images subliminales ». Une image devient subliminale lorsqu'elle est présentée en dessous du seuil de conscience (*limen* signifie « seuil » en latin), de sorte que personne ne parvienne à la voir.

Comment crée-t-on une image subliminale ? On peut tout simplement la rendre très pâle en réduisant son contraste. Malheureusement, cette solution dégrade l'image au point qu'elle n'engendre plus guère d'activité cérébrale. Une méthode plus intéressante consiste à présenter l'image pendant une durée très brève et à l'intercaler entre deux autres images. La figure 7 montre que l'on peut ainsi « masquer » le mot « radio ». On commence par présenter ce mot pendant 33 millisecondes, c'est-à-dire la

durée approximative d'une image de cinéma. En soi, cette astuce ne suffit pas à le rendre invisible – dans l'obscurité totale, même un flash d'une microseconde peut suffire à illuminer une scène et la rendre perceptible. Ce qui rend le mot « radio » invisible, c'est le fait de le « masquer » en le faisant précéder et surtout suivre de formes géométriques, présentées au même endroit sur l'écran. Si leur durée est bien choisie, on ne voit que les formes qui clignotent. Pris en sandwich entre elles, le mot devient totalement invisible.

J'ai personnellement conçu plusieurs expériences de masquage subliminal, et bien que j'aie généralement confiance dans mes capacités de programmation, il m'est arrivé d'en douter. En regardant l'écran, on jurerait que rien ne sépare les deux masques. Seule une cellule photoélectrique permet de s'assurer qu'effectivement un mot apparaît brièvement sur l'écran : sa disparition est un phénomène purement subjectif. D'ailleurs, le mot réapparaît lorsqu'on allonge sa durée de présentation.

Dans de nombreuses expériences, on observe une frontière assez nette entre « voir » et « ne pas voir » : l'image est invisible lorsqu'elle dure 40 millièmes de seconde, mais devient totalement évidente lorsque sa durée dépasse 60 millièmes de seconde. Cette observation justifie pleinement les mots *subliminal* (« en dessous du seuil ») et *supraliminal* (« au-dessus du seuil »). Métaphoriquement, une image doit franchir un seuil avant de passer la porte de la conscience, et il faut que cette porte soit ouverte ou fermée. En d'autres termes, l'accès à la conscience fonctionne en tout-ou-rien, et l'image est soit dehors, soit dedans. La taille du seuil varie d'une personne à l'autre, mais elle reste toujours aux alentours de 50 millisecondes. Lorsque la durée de l'image est juste au seuil, on la perçoit à peu près une fois sur deux. La présentation « au seuil » fournit donc des conditions idéales d'expérimentation : le stimulus reste constant, tandis que sa perception subjective, qui est le paramètre d'intérêt, varie d'un essai à l'autre.

De nombreuses variantes du masquage permettent de moduler la conscience à volonté. On peut, par exemple, effacer une image entière en l'intercalant entre deux masques composés d'un pêle-mêle de fragments d'images. Lorsque l'image masquée consiste en un visage souriant ou, au contraire, effrayé (figure 7b), elle induit la perception subliminale d'une émotion invisible – comme nous le verrons dans un instant, à l'insu de la conscience, l'émotion transparaît. Une autre variante du masquage consiste à présenter brièvement tout un ensemble de formes, dont l'une est entourée de quatre points (figure 7c). Curieusement, lorsque les points restent longtemps sur l'écran, la forme qu'ils entourent devient invisible, tandis que toutes les autres demeurent parfaitement nettes[34]. Les points encadrent un espace virtuel qui vient se substituer, dans la conscience, à la forme présentée au même endroit. C'est pourquoi on appelle cette méthode le « masquage par substitution ».

Le masquage est un merveilleux outil d'étude du sort d'un stimulus inconscient. En faisant apparaître successivement une image puis un masque, nous « injectons » dans le cerveau de l'observateur une dose fixe d'information visuelle – disons un mot. En principe, cette dose suffit à percevoir de quel mot il s'agit, puisque celui-ci devient parfaitement visible si l'on omet le masque. Cependant, la présence du masque, d'une manière ou d'une autre, efface de la conscience le mot qui précède. Le cerveau doit donc être le siège d'une étrange course : bien que le mot y pénètre en premier, le masque qui le suit semble le rattraper et finit par l'éjecter de la conscience. Il se peut que le cerveau se comporte comme un statisticien qui évalue les données disponibles avant de décider si une ou deux images étaient présentes. Lorsque la présentation du mot est très courte, une écrasante majorité des données militent en faveur de l'hypothèse que seul le masque est présent – et le cerveau en conclut qu'il n'a vu aucun mot.

La primauté du subjectif

Pouvons-nous garantir que les mots et les images, une fois masqués, échappent totalement à la conscience ? Dans les expériences les plus récentes de mon laboratoire, nous demandons tout simplement aux participants, à chaque essai, s'ils ont vu un mot ou non[35]. Cependant, certains de nos collègues critiquent cette procédure qu'ils jugent « trop subjective ». C'est un scepticisme qui me semble hors de propos : par définition, toute recherche sur la conscience doit s'intéresser aux aspects subjectifs de la perception. Quoi qu'il en soit, fort heureusement, nous disposons également d'autres moyens pour convaincre les sceptiques.

Tout d'abord, tout subjectif qu'il soit, le masquage rend généralement tous les observateurs d'accord. En deçà de 30 millisecondes, tous les participants affirment qu'ils ne voient pas le moindre mot. Seule varie la durée minimale au-delà de laquelle ils commencent à voir les mots masqués.

Plus crucialement, il est aisé de vérifier qu'au cours du masquage l'invisibilité subjective se double de conséquences objectives. Lorsqu'un sujet affirme, subjectivement, ne pas voir un mot, il ne parvient pas non plus, objectivement, à le nommer. (C'est seulement si on le force à répondre qu'il répond légèrement mieux que s'il l'avait fait au hasard – il s'agit là de l'un des marqueurs de la perception subliminale sur lesquels nous reviendrons au prochain chapitre.) Quelques secondes plus tard, l'observateur ne parvient pas à effectuer les tâches les plus simples, comme dire si un chiffre masqué est plus grand ou plus petit que 5. Dans l'une de nos expériences[36], nous présentions la même liste de 37 mots jusqu'à vingt fois chacun, tandis qu'un masquage les rendait invisibles. À l'issue de la présentation, nous demandions aux participants de choisir, parmi un vaste ensemble de mots, lesquels avaient été présentés. Ils n'y parvenaient absolument pas, ce qui suggérait que les mots masqués n'avaient laissé aucune trace en mémoire.

Toutes ces données conduisent à un point fondamental, le troisième ingrédient clé de la science de la conscience : *on peut et on doit faire confiance à l'introspection subjective*. Bien que l'invisibilité qu'induit le masquage soit un phénomène purement subjectif, il n'est pas invérifiable : l'invisibilité entraîne des conséquences bien réelles sur le traitement de l'information dans le cerveau, notamment parce qu'elle réduit considérablement les capacités de dénomination et de mémoire. Lorsqu'ils sont proches du seuil, les essais qu'un observateur déclare avoir vus s'accompagnent d'une amplification intense de l'information disponible, qui ne se traduit pas seulement par un sentiment subjectif de conscience, mais aussi par bien d'autres améliorations du traitement de l'information[37]. Quelle que soit l'information dont nous prenons conscience, nous parvenons à la nommer, à l'évaluer, à la juger ou à la mémoriser bien mieux que lorsqu'elle reste subliminale. En d'autres termes, lorsqu'on leur demande d'évaluer leur subjectivité, les observateurs humains ne se comportent pas de manière aléatoire ou fantasque : quand ils rapportent, avec honnêteté, l'impression d'avoir vu un stimulus, cet accès conscient s'accompagne d'un changement massif du traitement de l'information, qui entraîne presque toujours une amélioration des performances.

En bref, en dépit de la défiance que les béhavioristes et les cognitivistes éprouvent à son égard depuis un siècle, l'introspection est une source respectable d'informations en psychologie. Non seulement elle fournit des données valables, souvent confirmées par le comportement objectif ou par l'imagerie cérébrale, mais elle *définit* tout simplement l'essence même de la science de la conscience. Nous recherchons des explications objectives aux récits subjectifs : des signatures de la conscience, c'est-à-dire des événements neuronaux qui se déroulent systématiquement dans le cerveau d'un observateur lorsqu'il dit percevoir consciemment tel ou tel phénomène. Par définition, il n'y a que lui qui puisse nous en informer.

En 2001, dans un article qui est devenu un manifeste de ce domaine, Lionel Naccache et moi-même avions résumé ce point de la manière suivante : « Les récits subjectifs sont précisément les phénomènes que les neurosciences cognitives de la conscience se proposent d'étudier. À ce titre, ils constituent des données primaires qu'il convient de mesurer et d'enregistrer, aux côtés des autres données psychophysiologiques[38]. »

Cela dit, évitons de traiter l'introspection avec trop de naïveté : si celle-ci fournit indubitablement des données brutes au psychologue, elle ne l'informe certainement pas directement de la manière dont l'esprit fonctionne. Quand un patient atteint d'une maladie neurologique ou psychiatrique dit voir des visages flotter dans le vide devant lui, nous ne le prenons pas au mot – mais nous ne nions pas pour autant qu'il a effectivement ressenti une « vraie » illusion. Nous essayons simplement d'expliquer *pourquoi* il l'a éprouvée – par exemple parce qu'une soudaine crise d'épilepsie a activé les circuits du lobe temporal responsables de la perception des visages[39].

Notre introspection n'est pas toujours fondée[40]. Par définition, nous n'avons pas accès à nos processus inconscients – mais cela ne nous empêche pas de nous raconter des histoires à leur sujet. Par exemple, nous croyons que, lorsque nous lisons, nous reconnaissons « instantanément » les mots « comme un tout ». En réalité, au sein de notre cerveau, toute une série d'étapes sophistiquées, qui analysent les lettres et leurs combinaisons, échappe à notre conscience[41]. Deuxième exemple : lorsque nous cherchons à expliquer nos actes, nous inventons souvent toutes sortes d'explications après coup, sans nous rendre compte de nos vraies motivations inconscientes. Dans une expérience classique, on montre à des consommateurs quatre paires de bas en nylon, et on leur demande de choisir celle qui leur semble de meilleure qualité. En fait, elles sont identiques, mais cela n'empêche pas les consommateurs de montrer un biais systématique en faveur de celles situées à leur droite. Si on leur demande de justifier leur choix, pas un d'entre eux n'invoque la position dans l'espace : ils prétendent avoir repéré des différences de

qualité du tissu ! Dans cet exemple, l'introspection est de toute évidence illusoire.

Tout n'est donc pas vrai dans l'introspection… mais tout demande à être expliqué. En ce sens seulement, les béhavioristes avaient raison : en tant que méthode, l'introspection est trop fragile pour fonder une psychologie scientifique, parce qu'elle ne parvient pas à décrire notre fonctionnement mental avec exactitude ; en tant que mesure expérimentale, par contre, l'introspection constitue une plate-forme irremplaçable sur laquelle doit se fonder toute science de la conscience. Elle seule fournit une moitié cruciale de l'équation : la partie subjective, c'est-à-dire ce que les observateurs ressentent (même s'ils se leurrent sur l'origine de ce qu'ils éprouvent). Pour parvenir à une connaissance scientifique de la conscience, les neuroscientifiques doivent « simplement » identifier l'autre moitié de l'équation : quels événements objectifs, dans le cerveau, causent ces expériences subjectives ?

Parfois, comme dans le phénomène de masquage, une donnée objective vient immédiatement corroborer le récit subjectif : une personne dit qu'elle a vu un mot, et elle le prouve aussitôt en le nommant correctement. Cependant, la psychologie ne doit pas craindre d'étudier les nombreux autres cas où un sujet dit éprouver une sensation purement subjective, détachée de tout contexte réel, et qui paraît invérifiable. Même dans ces cas-là, des événements neuronaux objectifs doivent expliquer ce que la personne ressent. Dans la mesure où ce ressenti illusoire ne naît d'aucun stimulus physique, il peut même se révéler plus facile à localiser dans le cerveau, parce qu'il n'est confondu avec aucune variation des entrées sensorielles. Les chercheurs qui travaillent sur la conscience sont donc constamment à l'affût de telles situations « purement subjectives », dans lesquelles le stimulus est constant ou absent, tandis que la perception subjective varie. Ces cas idéaux transforment la conscience en variable expérimentale.

Il n'y a pas de meilleur exemple que les expériences du neurologue suisse Olaf Blanke sur la « sortie hors du corps ». Au réveil d'une

anesthésie, il arrive qu'un patient déclare avoir quitté son propre corps. Il dit avoir éprouvé l'impression irrépressible de flotter vers le plafond et de voir, de là-haut, son corps inerte se faire opérer. Faut-il prendre ce récit au sérieux ? La sortie du corps existe-t-elle « pour de vrai » ?

Afin d'évaluer les dires de leurs patients, certains pseudo-scientifiques cachent des dessins en haut des armoires des salles d'opération, en sorte que seul un patient en plein vol puisse les voir ! Cette idée est ridicule. La bonne approche consiste à se demander comment un dysfonctionnement des circuits normaux du cerveau peut engendrer le curieux sentiment de sortir de son propre corps. Comment notre cerveau adopte-t-il un point de vue sur le monde extérieur ? Comment sait-il où s'arrête le corps et où commence le reste du monde ? En étudiant de nombreux patients atteints de lésions cérébrales, Olaf Blanke a découvert que ce calcul de l'« image du corps » est l'apanage d'une région corticale située au carrefour des lobes temporaux et pariétaux de l'hémisphère droit, si bien que sa perturbation par une lésion ou par une stimulation électrique donne le sentiment d'échapper à son corps[42]. Cette région fait partie des secteurs de haut niveau du cortex, où convergent de nombreux signaux issus de la vision, des systèmes kinesthésiques et somato-sensoriels (qui cartographient les signaux du toucher, des muscles et de la motricité), et du système vestibulaire (la plate-forme inertielle située dans l'oreille interne qui évalue les mouvements de la tête). En assemblant tous ces indices, le cerveau se construit une image de la position du corps dans l'espace. Cependant, ce processus échoue parfois si les signaux sont en désaccord ou deviennent ambigus à la suite d'une lésion cérébrale. L'expérience de sortie du corps survient donc « vraiment » – c'est un authentique événement biologique, mais seulement au sein du cerveau du patient et, par voie de conséquence, de son expérience subjective. *Grosso modo*, la sortie du corps n'est qu'une forme exacerbée du malaise que nous ressentons tous lorsque notre vision ne concorde plus avec notre système vestibulaire, par exemple dans la cabine d'un bateau qui tangue.

Olaf Blanke a poursuivi ses travaux en montrant que *tout* être humain peut s'envoler hors de son corps. Il a créé une combinaison de signaux visuels et tactiles, synchronisés mais délocalisés, qui engendre systématiquement l'illusion de sortie du corps dans le cerveau d'un observateur sain de corps et d'esprit[43]. Par la robotique, il est même parvenu à recréer cette illusion dans une machine d'IRM. Dès que la personne scannée déclare éprouver cette illusion, sa région temporo-pariétale s'active – précisément là où se situaient les lésions des patients.

Nous ne savons toujours pas exactement quels calculs effectue cette région pour localiser le « point de vue subjectif » en un endroit précis de l'espace. Il n'empêche que la recherche neurologique a clairement évacué le phénomène de sortie du corps du domaine de la parapsychologie où il était confiné. La science de la conscience progresse ainsi à grands pas : même les plus insolites des sensations subjectives commencent à être rattachées à leurs origines neuronales. L'essentiel est de traiter l'introspection avec juste assez de sérieux : elle ne donne pas d'accès direct au fonctionnement du cerveau, mais constitue une donnée brute sur laquelle doit se fonder toute science de la conscience.

*

Au terme de cet aperçu de l'approche moderne de la conscience, une conclusion optimiste s'impose. Au cours des vingt dernières années sont apparus de nombreux outils expérimentaux qui permettent de manipuler la conscience à volonté. Grâce à eux, nous parvenons à faire disparaître de la conscience des mots, des images et même des films entiers – et à les faire réapparaître en un clin d'œil, soit en changeant minimalement le stimulus, soit même sans rien modifier d'objectif.

Ces outils en main, nous pouvons désormais revenir aux questions brûlantes que René Descartes aurait tant souhaité explorer. Premièrement, que devient une image invisible ? Est-elle traitée par le cortex ? Pendant combien de temps ? Jusqu'à quelle profondeur ? La réponse dépend-elle

de la manière dont on l'a fait disparaître[44] ? Deuxièmement, comment l'activité nerveuse change-t-elle lorsqu'un stimulus devient visible ? Certains événements cérébraux surviennent-ils spécifiquement au cours de la prise de conscience ? Pouvons-nous identifier de telles « signatures de la conscience » et nous appuyer sur elles pour comprendre ce qu'est réellement la conscience ?

Dans le prochain chapitre, nous examinerons la première de ces questions : les images subliminales influencent-elles profondément nos cerveaux, nos pensées et nos actes ?

CHAPITRE 2

Sonder la profondeur de l'inconscient

Jusqu'où une image invisible peut-elle cheminer dans notre cerveau ? Peut-elle traverser toute la hiérarchie des aires du cortex et influencer nos décisions ? Il est essentiel de répondre à ces questions afin d'esquisser les frontières de la pensée consciente. La psychologie et l'imagerie cérébrale montrent que notre cerveau reconnaît et catégorise inconsciemment des images invisibles, et parvient même à déchiffrer et à interpréter le sens de mots imperceptibles. À notre insu, les images subliminales déclenchent en nous les circuits de la motivation et de la récompense. Même des opérations complexes, qui relient la perception à l'action, se déroulent sans conscience : nous fonctionnons souvent « en pilotage automatique ». Dans l'ignorance de ce bouillonnement de processus inconscients, nous ne cessons de surestimer notre pouvoir de décision consciente – alors qu'en vérité notre degré de contrôle est sévèrement limité.

« Temps passé et temps futur ne laissent que peu de place à la conscience. »

T. S. ELIOT, *Burnt Norton* (1935).

En l'an 2000, au cours de la campagne présidentielle américaine, une publicité venimeuse concoctée par l'équipe de George W. Bush caricaturait le programme économique d'Al Gore en l'accompagnant du mot « RATS » en énormes majuscules (figure 8). Bien qu'elle ne fût pas strictement subliminale, l'image n'était guère visible, car elle passait pour la fin du mot *bureaucrats*. Cette insulte voilée lança un vif débat : le cerveau des spectateurs enregistrait-il le message caché ? Jusqu'où pénétrait-il ? Parvenait-il jusqu'aux circuits émotionnels du cerveau, et influençait-il les intentions de vote ?

L'élection présidentielle française, douze ans auparavant, avait été le théâtre d'un usage encore plus controversé des images subliminales. On découvrit fortuitement que le visage de François Mitterrand avait été discrètement inséré dans le générique du journal de 20 heures de la chaîne publique Antenne 2 (figure 8). Tous les jours, depuis des mois,

l'image invisible s'était imprimée sur la rétine de millions de téléspectateurs. Avait-elle biaisé leur vote ? Toute influence, même minime, aurait pu se traduire en milliers de bulletins dans l'urne.

La plus célèbre de toutes les manipulations subliminales est sans doute l'insertion d'une image portant les mots *Drink*

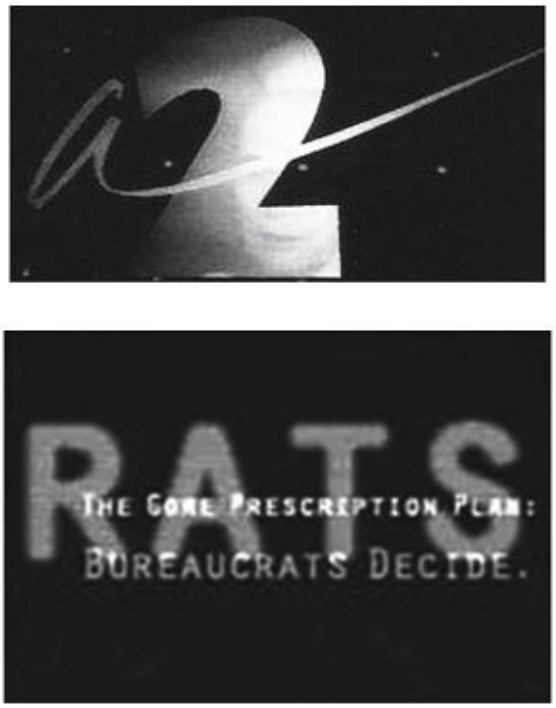


Figure 8. Des images subliminales apparaissent parfois dans les médias. Pendant la campagne présidentielle de 1988, le visage de François Mitterrand était présenté subrepticement au sein du logo du journal de 20 heures d'Antenne 2. En 2000, dans l'un des spots de campagne de George W. Bush, un plan d'Al Gore était accompagné d'une insulte, le mot « rats » (la fin du mot « *bureaucrats* »). De telles images sont-elles reconnues par notre cerveau et parviennent-elles à influencer nos décisions ?

Coca-Cola dans un film en 1957. À l'entracte, des hordes de spectateurs se seraient précipitées vers la buvette. Ce mythe fondateur n'est pourtant qu'une invention : son auteur proclamé, James Vicary, a avoué quelques années plus tard que sa prétendue expérience n'avait jamais eu lieu ! Seul

perdure le mythe, mais aussi la vraie question scientifique : une image invisible peut-elle influencer nos pensées ? Ce n'est pas seulement un problème central pour la démocratie et la manipulation de masse, mais une énigme essentielle pour la connaissance même de notre cerveau. Avons-nous besoin d'être conscients d'une image pour la traiter ? Ou bien pouvons-nous percevoir, catégoriser et décider sans conscience ?

La question est d'actualité puisque toute une panoplie de méthodes permet de présenter des informations au cerveau sans qu'il les détecte consciemment. La rivalité binoculaire, l'inattention, le masquage et bien d'autres artifices peuvent nous rendre inconscients de pratiquement n'importe quel aspect de notre environnement. En sommes-nous simplement aveugles ? Lorsque nous focalisons notre attention sur un objet, cessons-nous de percevoir tout le reste, ou continuons-nous de le percevoir sans le savoir ?

De la réponse à ces interrogations dépend l'identification des signatures de l'expérience consciente. Clarifier la profondeur des opérations non conscientes, c'est éclaircir, comme en miroir, la nature de la conscience. Si nous prouvons, par exemple, que les étapes précoces de la perception peuvent se dérouler inconsciemment, nous pouvons les rayer de notre quête du propre de la conscience. Plus nous en apprenons sur l'inconscient, mieux nous cernons, par élimination, les frontières de la conscience. Toute photographie de l'inconscient esquisse, en négatif, une image de l'esprit conscient.

Les pionniers de l'inconscient

On attribue généralement à Sigmund Freud (1856-1939) la paternité de la découverte que de nombreuses opérations mentales se déroulent hors de notre conscience. En réalité, il s'agit d'une légende que Freud lui-même a contribué à forger[1]. Comme l'a remarqué l'historien et philosophe

Marcel Gauchet, « quand Freud déclare en substance en 1925 qu'avant la psychanalyse il était de règle d'identifier psychisme et conscience, force nous est de constater que c'est rigoureusement faux[2] ».

En vérité, la découverte qu'une quantité prodigieuse d'opérations mentales échappe à la conscience précède Freud de plusieurs décennies, voire de plusieurs siècles[3]. Dès l'Antiquité, le médecin romain Galien (environ 129-200) et le philosophe Plotin (environ 204-270) avaient constaté que certaines fonctions du corps, comme la marche ou la respiration, ne demandaient aucune attention. La plupart de leur savoir clinique était un héritage d'Hippocrate (environ 460-370 avant notre ère), cet éminent observateur des maladies dont le nom reste étroitement lié à la profession médicale. Hippocrate consacra tout un traité à l'épilepsie, qu'on appelait alors la « maladie sacrée ». Il y montrait que cette maladie n'a « rien de plus divin ni de plus sacré que les autres », mais qu'elle résulte d'une affection cérébrale qui force le corps à se mouvoir indépendamment de toute volonté. Il en concluait que notre cerveau nous contrôle en permanence et, à notre insu, qu'il tisse l'étoffe même de notre vie mentale :

> « Les hommes devraient savoir que c'est du cerveau, et du cerveau seul, que proviennent les plaisirs, les joies, les rires et les jeux, ainsi que les peines, les douleurs, les chagrins et les larmes. C'est par lui surtout que nous pensons, voyons, entendons et distinguons le laid du beau, le mauvais du bon, et l'agréable du désagréable. »

Après la chute de l'Empire romain, les érudits indiens et arabes préservèrent une partie du savoir médical de l'Antiquité. Au XI^{e} siècle, le scientifique arabe que nous connaissons sous le nom d'Alhazen (Ibn al-Haytham, 965-1040) découvrit les principes essentiels de la perception visuelle. Des siècles avant Descartes, il comprit que l'œil fonctionne

comme une chambre noire qui reçoit la lumière plutôt qu'elle ne l'émet, et que diverses illusions peuvent le tromper[4]. Tout n'est donc pas sous le contrôle de la conscience, en conclut Alhazen. Il fut le premier à postuler une opération automatique d'inférence inconsciente : à notre insu, le cerveau tire des conclusions qui dépassent ce que perçoivent les sens, et nous donne parfois à voir des choses inexistantes[5]. Il faudra attendre huit siècles pour qu'en 1867, dans son traité d'*Optique physiologique*, le physicien Hermann von Helmholtz reprenne ce terme d'*inférence inconsciente* pour décrire la manière dont notre vision déduit automatiquement, à partir des entrées sensorielles, l'interprétation la plus probable de la scène qui se déroule sous les yeux.

Par-delà la perception inconsciente s'étend le vaste champ des origines de nos buts, de nos désirs, de nos tentations et de nos choix. Des siècles avant Freud, de nombreux philosophes, dont saint Augustin (354-430), Thomas d'Aquin (1225-1274), René Descartes (1596-1650), Baruch Spinoza (1632-1677) et Gottfried Wilhelm von Leibniz (1646-1716) ont noté à quel point notre ligne de conduite nous est souvent dictée par des mécanismes inaccessibles à l'introspection, depuis les réflexes jusqu'aux désirs inavoués. Spinoza recensait ainsi toute une panoplie d'instincts qui échappent au contrôle conscient : l'appétence d'un bébé pour le sein, l'emprise de la jalousie ou de la vengeance, l'attirance irrésistible de l'ivrogne pour sa bouteille, ou encore l'irrépressible flot de paroles du bavard.

Au cours du XVIII^e^ et du XIX^e^ siècle, les neurologues découvrirent l'omniprésence, dans le système nerveux, de circuits inconscients. Marshall Hall (1790-1857) fut le premier à énoncer le concept d'« arc réflexe » qui relie certaines entrées sensorielles à des mouvements spécifiques et irrépressibles. Il remarqua que certaines actions, qui trouvent leur origine dans la moelle épinière, échappent à notre contrôle volontaire. À sa suite, John Hughlings Jackson (1835-1911) souligna l'organisation hiérarchique du système nerveux, depuis le tronc cérébral jusqu'au cortex, et depuis les opérations automatiques jusqu'à celles qui

mobilisent la conscience et la volonté. En France, les psychologues et sociologues Théodule Ribot (1839-1916), Gabriel Tarde (1843-1904) et Pierre Janet (1859-1947) discernèrent l'étendue des automatismes humains, depuis la mémoire des gestes et des usages (Ribot) jusqu'à l'imitation inconsciente (Tarde), et même l'existence de désirs subconscients dont l'origine remonte à notre enfance et qui deviennent des traits de notre personnalité (Janet).

Les idées des scientifiques français étaient tellement en avance sur leur temps que, lorsque les premières théories de Freud lui valurent la gloire, Janet en revendiqua la paternité. Cependant, dès 1868, le psychiatre anglais Henry Maudsley (1835-1918) avait écrit que « la part la plus importante de nos actes mentaux, le processus essentiel sur lequel repose toute pensée, c'est l'activité mentale inconsciente[6] ». Dès 1899, le neurologue Sigmund Exner, contemporain et collègue de Freud à Vienne, écrivait que « nous ne devrions pas dire "je pense", "je ressens", mais plutôt "ça pense en moi" [*es denkt in mir*], "ça ressent en moi" [*es fühlt in mir*] » – une bonne vingtaine d'années avant que Freud ne publie sa célèbre distinction entre le ça, le moi et le surmoi (*Das Ich und das Es*, 1923).

À la fin du XIX^e^ siècle, l'ubiquité des processus inconscients était si bien acceptée que, dans ses célèbres *Principes de psychologie* (1890), le grand psychologue et philosophe américain William James affirmait : « Tous ces faits, considérés ensemble, annoncent indubitablement l'avènement d'une quête destinée à jeter une lumière nouvelle sur les abysses mêmes de notre nature… Ils prouvent de façon concluante que nous ne devons jamais interpréter le témoignage d'une personne, même la plus sincère, qu'elle n'a rien perçu comme l'assurance qu'il n'a existé aucune forme de sentiment[7]. » Tout être humain, présumait-il, « est amené à faire toutes sortes de choses saugrenues dont il n'a pas la moindre conscience ».

Parmi ce foisonnement de découvertes neurologiques et psychologiques, qui démontrent à quel point notre vie est dominée par des mécanismes inconscients, la contribution personnelle de Freud paraît

maigre. Il ne serait guère exagéré d'affirmer que, dans son œuvre, les idées solides ne sont pas les siennes, tandis que les idées qui sont les siennes ne sont pas très solides. Avec un siècle de recul, il semble particulièrement décevant que Freud n'ait jamais tenté de mettre ses théories à l'épreuve de l'expérience. La fin du XIX^e^ et le début du XX^e^ siècle ont vu la naissance de la psychologie expérimentale. De nouvelles méthodes s'épanouissaient partout, et notamment la chronométrie mentale, qui consistait à recueillir avec précision les temps de réponse et les erreurs d'un observateur humain. Bien qu'il en fût averti, Freud se contenta de proposer des modèles métaphoriques de l'esprit, sans jamais les tester sérieusement. L'un de mes écrivains favoris, Vladimir Nabokov, exécrait les méthodes de Freud, et c'est avec une exquise méchanceté qu'il les reléguait aux oubliettes : « Laissons les crédules et les vulgaires continuer à croire que toutes les infortunes mentales peuvent être guéries par une application quotidienne de vieux mythes grecs sur les parties intimes de leur individu. Peu m'importe, vraiment[8]. »

Le siège des opérations inconscientes

En dépit des grandes avancées médicales du XIX^e^ et du XX^e^ siècle, force est de constater que dans les années 1990, lorsque mes collègues et moi avons commencé à étudier la perception subliminale avec les méthodes de l'imagerie cérébrale, une énorme confusion entourait toujours la question de la nature des aires impliquées dans la reconnaissance des images invisibles. Il existait plusieurs hypothèses, incompatibles entre elles, sur la division du travail entre le conscient et l'inconscient. La plus simple était que le cortex – les minces couches de neurones qui tapissent la surface de nos deux hémisphères – était le siège de la conscience, tandis que tous les autres circuits fonctionnaient sur un mode non conscient. En effet, le cortex, partie la plus évoluée du cerveau des mammifères, abrite des

opérations avancées telles que l'attention, la planification ou la parole. Il paraissait donc naturel de supposer que l'information qui l'atteignait devait être consciente. Les opérations inconscientes, par contre, étaient supposées se dérouler au sein de groupements neuronaux plus spécialisés comme ceux de l'amygdale ou du colliculus, qui ont évolué pour remplir des fonctions précises telles que la détection des dangers ou le déplacement des yeux. Ces régions forment des circuits « sous-corticaux », ainsi nommés parce qu'ils sont situés sous la surface du cortex.

Une proposition distincte, quoique également naïve, introduisait une dichotomie entre les deux hémisphères. Le gauche, siège du langage, pouvait décrire ce qu'il faisait. Cela signifiait que lui seul était conscient – l'hémisphère droit ne l'était pas.

Une troisième hypothèse avançait que certains circuits corticaux étaient conscients, d'autres non. Plus précisément, n'importe quelle information visuelle qui passe par la voie ventrale du cortex, celle qui reconnaît l'identité des objets, des visages et des mots, serait forcément perçue consciemment. Par contre, l'information transmise à la voie visuelle dorsale, qui passe par le cortex pariétal et s'appuie sur la forme et la position des objets pour guider nos mouvements, disparaîtrait à jamais du côté obscur de la conscience.

Aucune de ces dichotomies simplistes ne tient la route. Nous savons aujourd'hui que pratiquement toutes les régions du cerveau peuvent participer, tantôt à la pensée consciente, tantôt aux opérations inconscientes. Parvenir à cette conclusion, toutefois, a demandé une série d'expériences plus raffinées les unes que les autres, qui ont progressivement étendu les frontières de l'inconscient.

Au départ, donc, la simple observation des patients atteints de lésions cérébrales avait suggéré que les opérations inconscientes se tramaient dans les tréfonds du cerveau, en deçà de la surface du cortex, et notamment dans l'amygdale. Il s'agit d'un groupe de neurones en forme d'amande, situé sous le lobe temporal, et qui colore les événements de notre vie d'une tonalité émotionnelle. L'amygdale signale tout

particulièrement la peur : dès que nous voyons une image angoissante, par exemple un serpent, elle s'active à grande vitesse, en ligne directe avec la rétine, bien avant que le cortex ne ressente quoi que ce soit[9]. Nombreuses sont les expériences qui ont démontré l'autonomie et l'extraordinaire rapidité du tri émotionnel inconscient qu'effectue l'amygdale. Dans les années 1900, le neurologue suisse Édouard Claparède fit la preuve de l'existence d'une mémoire émotionnelle inconsciente : en serrant la main d'une amnésique, il la piqua avec une aiguille, et le jour suivant, alors que son amnésie ne lui permettait pas de le reconnaître, elle refusa farouchement la main qu'il lui tendait, sans pouvoir expliquer pourquoi.

Une seconde source de données sur la perception subliminale provenait de la « vision aveugle » (*blindsight*) de certains patients dont le cortex visuel primaire, principal point d'entrée des informations visuelles dans le cortex, était détruit. La « vision aveugle » est un bizarre oxymore qui décrit pourtant à merveille la curieuse affection de ces malades qui « voient sans voir ». Les neurologues savent bien que toute lésion du cortex visuel primaire devrait rendre la personne aveugle, et de fait ces patients n'ont aucune vision *consciente* – ils affirment ne rien voir (dans la partie du champ visuel qui correspond à leur lésion cérébrale), et ils se comportent effectivement comme des aveugles. Toutefois, il est stupéfiant de constater que, lorsqu'on leur présente des objets ou des points lumineux dans leur champ aveugle, ils parviennent à indiquer leur position avec une grande précision[10]. Tels des zombies, ils tendent inconsciemment la main en direction de choses qu'ils ne voient pas – une authentique vision aveugle.

Quelles voies anatomiques intactes la vision inconsciente emprunte-t-elle chez ces patients ? Il est clair que, chez eux, la rétine transmet toujours certaines informations jusqu'à la main, en contournant la lésion qui les rend aveugles. Comme le point d'entrée dans le cortex visuel est détruit, les chercheurs ont d'abord pensé que ce comportement provenait exclusivement de circuits sous-corticaux. Le principal « suspect » était le

colliculus supérieur, un noyau situé dans le mésencéphale dont les neurones sont sensibles à la position des informations visuelles dans l'espace. Effectivement, la toute première étude d'IRM fonctionnelle de la vision aveugle a montré que les cibles invisibles déclenchaient une forte activation du colliculus supérieur[11]. Toutefois, cette étude montrait également des activations dans le cortex – et les recherches ultérieures ont effectivement confirmé que les stimuli, même s'ils n'étaient pas perçus, activaient toujours le thalamus et les aires visuelles supérieures, en trouvant le moyen de contourner la région endommagée du cortex visuel primaire[12]. De toute évidence, les circuits cérébraux qui contrôlent notre « zombie intérieur » et guident inconsciemment les mouvements de nos yeux et de nos mains débordent bien au-delà des anciennes voies sous-corticales.

Une autre patiente, décrite par le psychologue canadien Melvyn Goodale, a apporté beaucoup d'eau au moulin de l'hypothèse d'un inconscient cortical. À 34 ans, Mme D. F. a subi une grave intoxication au monoxyde de carbone[13]. Le manque d'oxygène a causé des dommages irréversibles à ses aires visuelles droites et gauches. Elle a immédiatement perdu les plus élémentaires des opérations visuelles conscientes, et contracté ce que les neurologues appellent une « agnosie visuelle ». Pour tout ce qui, de près ou de loin, touchait à la reconnaissance des formes, Mme D. F. était essentiellement aveugle – elle ne distinguait même pas un carré d'un rectangle deux ou trois fois plus long. Son handicap était tel qu'elle ne pouvait juger si une simple ligne était verticale, horizontale ou oblique. Et pourtant, ses gestes demeuraient adéquats : si on lui demandait de poster une lettre dans une rainure dont elle était incapable de percevoir l'orientation, sa main se dirigeait toujours avec la plus grande assurance. Son système moteur continuait de « voir », inconsciemment, bien mieux qu'elle ne le pouvait consciemment. Elle adaptait également l'ouverture de sa main aux objets qu'elle attrapait – et pourtant elle était incapable d'en évaluer volontairement la taille, même si

on lui demandait de l'indiquer d'un geste symbolique en écartant le pouce des autres doigts.

En résumé, chez Mme D. F., une voie visuo-motrice inconsciente demeurait intacte, en dépit de son incapacité majeure à percevoir consciemment les mêmes formes. Mel Goodale et ses collaborateurs ont soutenu que son comportement était trop sophistiqué pour s'expliquer au niveau sous-cortical, et qu'il devait impliquer la voie visuelle du cortex pariétal. À l'insu de Mme D. F., certaines informations visuelles traversaient toujours son cortex occipital en direction du lobe pariétal. Là, des circuits préservés analysaient encore les tailles, les positions, les orientations et même les formes qu'elle ne parvenait plus à voir.

Depuis ces premières observations, de nombreux autres cas de vision aveugle et d'agnosie visuelle ont été décrits. Certains patients parviennent à parcourir un couloir encombré sans jamais se heurter aux obstacles, tout en jurant être aveugles. D'autres souffrent d'une forme particulière de perte de conscience appelée « héminégligence spatiale ». Chez eux, une lésion souvent située dans l'hémisphère droit, aux confins inférieurs du lobe pariétal, empêche l'orientation de l'attention vers la moitié gauche de l'espace. Résultat : ces patients ratent systématiquement la partie gauche de ce qu'ils ont sous les yeux. Un patient se plaignait même qu'on ne le nourrissait pas suffisamment : il mangeait bien la moitié droite de l'assiette, mais ne se rendait pas compte que la moitié gauche était encore pleine.

Figure 9. L'étude des patients atteints de lésions cérébrales prouve que le cortex traite les images inconsciemment. En haut, la patiente D. F., étudiée par Goodale et Milner (1991), avait perdu toute capacité de reconnaître le moindre objet ou même de décrire l'orientation d'une fente. Pourtant, elle parvenait toujours à y glisser une lettre : la voie visuo-motrice, qui guide nos mouvements, peut se dérouler sans la moindre conscience. En bas, le patient P. S., étudié par Marshall et Halligan (1988), souffrait d'une « héminégligence » qui l'empêchait de prêter attention à la partie gauche des objets. Il ne voyait pas la moindre différence entre ces deux maisons… mais il préférait tout de même habiter dans celle du bas, ce qui suggère que son cerveau reconstruisait toujours, inconsciemment, le sens de ces dessins.

Les patients atteints d'héminégligence ne sont pas aveugles du côté gauche. Bien qu'ils souffrent d'un important problème d'appréhension de cette partie de l'espace, leur rétine et leur cortex visuel précoce fonctionnent toujours. C'est une lésion plus en aval qui, d'une manière ou d'une autre, les empêche de prêter attention à ces informations et de les percevoir consciemment. Sont-elles pour autant perdues ? Non : inconsciemment, le cortex continue de traiter les informations négligées. John Marshall et Peter Halligan en ont apporté une élégante démonstration en montrant à un patient deux maisons, dont l'une avait clairement la moitié de gauche en flammes (figure 9)[14]. Le patient niait voir la moindre différence – pour lui, les deux maisons étaient rigoureusement identiques. Cependant, lorsqu'on lui demandait de choisir celle dans laquelle il préférerait vivre, il évitait systématiquement celle qui était en feu. Son cerveau parvenait donc encore à reconnaître, dans les flammes, un danger à éviter. Quelques années plus tard, l'imagerie cérébrale démontra que, chez ces patients, un stimulus invisible parvient encore à activer les aires du cortex visuel ventral qui se spécialisent pour les maisons et les visages[15]. Même le sens d'un mot ou d'un nombre négligé chemine toujours dans le cortex d'un patient héminégligent[16].

Le côté obscur du cerveau

Initialement, toutes les données dont nous disposions étaient issues de patients atteints de lésions cérébrales souvent étendues et qui affectaient peut-être la division normale du travail entre le conscient et l'inconscient. Le cerveau normal, en l'absence de toute lésion, traite-t-il également les images inconsciemment ? Le cortex visuel peut-il entrer en activité, éventuellement jusqu'à un haut niveau, sans que cela engendre une prise de conscience ? Mon laboratoire a figuré parmi les premiers à répondre à ces questions. À l'aide de l'imagerie cérébrale, nous sommes parvenus à

démontrer que des mots invisibles atteignaient des étapes corticales très profondes.

Comme je l'ai évoqué au chapitre 1, il nous est facile de présenter une image pendant quelques dizaines de millisecondes sans qu'elle soit vue. L'astuce consiste à la masquer par d'autres formes qui apparaissent au même endroit juste avant et surtout juste après (voir la figure 7). Mais comment déterminer à quel point cette image pénètre dans le cerveau ? Pour ce faire, mes collègues et moi avons employé la méthode d'« amorçage subliminal », qui consiste à faire suivre le mot ou l'image invisible (qu'on appelle l'« amorce ») par une autre image nettement visible (la « cible »). D'un essai à l'autre, la relation entre l'amorce et la cible change : ils sont tantôt identiques (par exemple *radio* suivi de *radio*), tantôt différents (par exemple *chien* suivi de *radio*). Les participants ne se rendent même pas compte de la présence d'un mot caché. Ils focalisent toute leur attention sur le mot visible, et on leur demande de juger le plus vite possible si, par exemple, ce mot fait référence à un objet ou à un être vivant (pratiquement n'importe quelle instruction fonctionne).

Le résultat, fascinant, a été répliqué plusieurs dizaines de fois : la présentation inconsciente d'un mot accélère le traitement du même mot lorsqu'il réapparaît consciemment[17]. Du moment que les deux présentations sont séparées de moins d'une seconde, la répétition facilite la prise de décision, même si elle demeure indétectable. Les participants répondent donc plus vite, et en faisant moins d'erreurs, au mot *radio* précédé de *radio* qu'au même mot précédé de *chien*. Ce phénomène s'appelle l'« amorçage subliminal par répétition » : exactement comme on amorce une pompe en lui injectant une petite quantité d'eau, on peut amorcer le circuit de reconnaissance des mots à l'aide d'un mot invisible.

Nous savons aujourd'hui que l'information qui est injectée dans le cerveau peut atteindre un haut degré d'abstraction. Par exemple, l'effet d'amorçage continue de fonctionner même si l'amorce est en minuscules (*radio*) et la cible en majuscules (*RADIO*), ou *vice versa*. Or, sur le plan visuel, ces formes sont totalement différentes. Le *a* minuscule n'a rien à

voir avec le *A* majuscule – seule une convention culturelle lie ces deux formes à la même lettre. Pourtant, l'expérience montre que, chez les lecteurs experts, cette convention, devenue inconsciente, est parfaitement intégrée aux étapes précoces du traitement visuel : l'amorçage subliminal est tout aussi puissant quand on répète le mot à l'identique (*radio-radio*) que quand on en change la casse (*radio-RADIO*)[18]. Sans que nous en ayons conscience, notre cerveau représente l'identité abstraite des lettres et parvient à s'abstraire des variations superficielles de forme, de taille ou de police.

Restait à déterminer quelle région cérébrale réalise cette opération. Mes collègues et moi sommes parvenus à montrer que la sensibilité de l'imagerie fonctionnelle par résonance magnétique suffisait à détecter le tout petit surcroît d'activité cérébrale évoquée par un mot inconscient[19]. Avec l'IRM, nous avons obtenu une image de l'ensemble des régions cérébrales affectées par l'effet d'amorçage. Toute une partie du cortex visuel situé sur la face ventrale du cortex s'activait en réponse à un mot inconscient. Le circuit comprenait une région appelée « gyrus fusiforme » qui abrite des mécanismes avancés de reconnaissance des formes et qui est impliquée dans une étape importante de la lecture, l'extraction de la forme visuelle du mot[20]. Là, l'amorçage ne dépendait plus de la forme du mot : cette région ne s'intéressait qu'à la forme abstraite du mot, et non au fait qu'il soit écrit en majuscules ou en minuscules[21].

Avant nos expériences, certains chercheurs supposaient que toute activation du gyrus fusiforme conduisait à une prise de conscience. En effet, cette région appartient à la « voie visuelle ventrale » qui nous permet de voir la forme des objets. Seule la « voie dorsale », estimaient ces chercheurs, qui relie le cortex visuel du lobe occipital aux systèmes du lobe pariétal spécialisés pour le codage de l'espace et l'action, fonctionnait de façon inconsciente[22]. En démontrant que la voie ventrale, qui repère l'identité des mots et des images, opère parfois sans conscience, nos expériences ont fait voler en éclats cette idée simpliste[23]. En réalité, les

deux circuits implémentent des opérations susceptibles de se dérouler tantôt sur le mode subliminal, tantôt sous le contrôle de la conscience.

Tisser des liens inconscients

Année après année, les expériences d'amorçage subliminal ont dissipé bien des mythes sur la place de la conscience au sein du système visuel. L'une de ces idées aujourd'hui révoquées postulait que chaque partie de la scène visuelle pouvait être identifiée inconsciemment, mais que la conscience était indispensable à l'assemblage de l'ensemble en un tout cohérent. En l'absence d'attention consciente, les propriétés de mouvement, de couleur ou de forme, éclatées dans ces aires cérébrales différentes, flottaient indépendamment les unes des autres et ne pouvaient pas être rassemblées en un tout cohérent[24]. Puisque le système visuel les séparait dès le départ, il fallait les rassembler en un seul « fichier » pour qu'émerge enfin la perception globale d'un objet intégré. Certains chercheurs estimaient que ce processus de liage (*binding*), qui reposait sur la synchronisation d'ensembles de neurones[25] ou sur la « ré-entrée » d'informations d'une aire à l'autre[26], était l'apanage de la conscience. Nous savons aujourd'hui que cette hypothèse est fausse : certaines formes de liage visuel surviennent sans la moindre conscience. Prenons l'exemple des lettres qui composent un mot. Il faut clairement qu'elles soient rassemblées en un tout cohérent, et dans le bon ordre, pour que nous lisions un mot comme *RAIDE* sans le confondre avec *AIDER* : le déplacement d'une seule lettre fait parfois toute la différence. Or nos expériences ont démontré que l'assemblage des lettres s'effectue inconsciemment[27]. En effet, on observe un effet d'amorçage subliminal quand le mot *RAIDE* suit le mot *raide*, mais pas quand *RAIDE* suit *aider* – ce qui indique que l'amorçage est sensible non seulement à la présence des lettres, mais également à leur arrangement en un mot. D'ailleurs, les

réponses à *RAIDE* précédé d'*aider* n'étaient même pas plus rapides que lorsque *RAIDE* était précédé d'un mot comme *moulu*, sans la moindre lettre en commun. Ainsi, la perception subliminale ne se laisse pas tromper par le fait que deux mots partagent 80 % de leurs lettres. Notre inconscient est éminemment sensible, et une différence d'une seule lettre lui suffit à conclure que les deux mots sont complètement différents.

Au cours des dix dernières années, des centaines d'expériences ont répliqué le phénomène d'amorçage subliminal – non seulement avec des mots, mais également avec des visages, des photos et des dessins[28]. Nous en avons conclu que ce que nous percevons consciemment d'une scène visuelle est infiniment plus abstrait que ce qu'en reçoivent nos yeux. Nous ne voyons jamais le monde tel que notre rétine le perçoit. Ce serait d'ailleurs un bien étrange spectacle : un amas confus de points sombres ou lumineux, monstrueusement élargi en son centre (la fovéa), masqué en partie par des vaisseaux sanguins, troué d'une vaste « tache aveugle » à l'endroit où le nerf optique quitte la rétine ; et, qui plus est, une image qui changerait constamment, deux ou trois fois par seconde, dès que nous déplaçons les yeux. Ce que nous voyons, à la place, c'est une scène tridimensionnelle, corrigée de tous les défauts de la rétine, stabilisée jusqu'à éliminer toute impression de mouvement des yeux, et interprétée en fonction de notre expérience passée. Toutes ces opérations visuelles se déroulent inconsciemment, bien que leur complexité dépasse largement celle d'un logiciel comme Photoshop. Pour prendre un exemple, notre système visuel détecte les ombres et les soustrait de l'image afin de restituer la couleur et la luminosité de la scène (figure 10). En un clin d'œil, notre cerveau infère la position des sources de lumière et en déduit la forme, l'opacité, la réflectance et la luminance des objets.

Dès que nous ouvrons les yeux, notre cortex visuel se lance dans d'innombrables calculs – mais nous n'en avons cure. Comme nous n'avons aucun accès aux opérations intimes du cerveau, nous croyons qu'il ne travaille dur que lorsque nous avons conscience de travailler dur – par exemple lorsque nous faisons des calculs mentaux ou que nous jouons aux

échecs. Nous nous méprenons totalement sur la complexité des opérations qu'il met en œuvre, inconsciemment, pour créer l'impression d'un monde visuel simple et sans coutures.

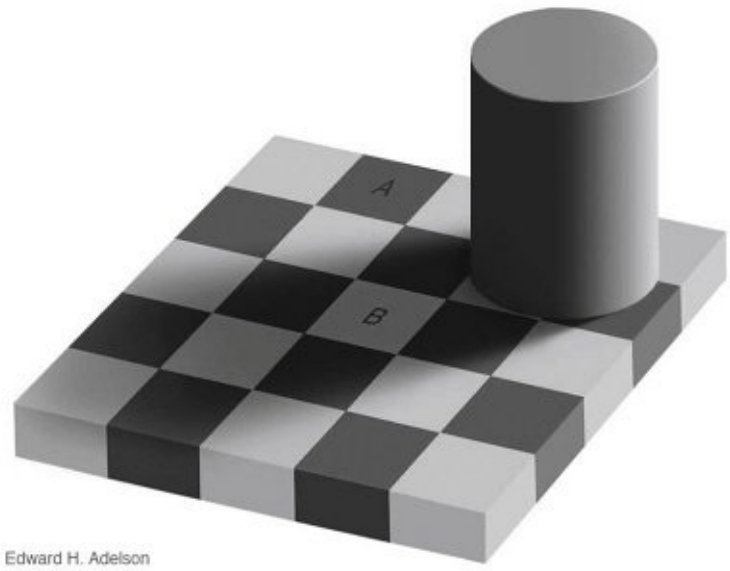


Figure 10. Notre vision exige de puissants calculs inconscients. Un simple coup d'œil suffit à reconnaître ce damier. Aucun doute : la case A est gris foncé, et la case B gris clair. Pourtant, elles ont exactement la même nuance de gris (vérifiez-le en masquant le reste de l'image). D'où provient l'illusion ? En une fraction de seconde, notre cerveau décompose cette scène, décide que la lumière vient d'en haut à droite, que le cylindre jette une ombre sur le damier, et soustrait alors cette ombre afin de nous donner à voir les « vraies » couleurs sous-jacentes. Nous n'avons aucune conscience de ces calculs : seul le résultat final entre dans notre conscience subjective.

Jouer aux échecs inconsciemment

Approfondissons l'exemple du jeu d'échecs. Lorsque le grand maître Garry Kasparov joue une partie difficile, doit-il prêter consciemment attention à la configuration de l'échiquier avant de voir que, disons, l'un des cavaliers noirs menace la reine blanche ? Ou bien peut-il se concentrer

sur le plan d'ensemble et laisser son système visuel repérer automatiquement les relations entre les pièces ?

L'intuition nous souffle que, chez un tel expert, l'analyse d'une position de l'échiquier doit s'apparenter à un réflexe. Et c'est bien ce que prouve la recherche : un seul regard suffit à un grand maître pour évaluer une partie et en mémoriser la position, parce qu'il analyse la scène en fragments cohérents[29]. Une expérience récente montre que cette décomposition se déroule hors de toute conscience : l'image d'un jeu d'échecs simplifié présentée pendant 20 millisecondes et prise en sandwich entre des masques qui la rendent invisible suffit à influencer la décision d'un joueur[30]. L'expérience ne fonctionne que si le participant est un expert, et s'il tente de résoudre un problème qui fait sens, par exemple déterminer si le roi est en échec ou non. Elle implique que le système visuel analyse l'identité des pièces (roi ou cavalier), mais également leur position, et qu'il assemble toutes ces informations en un tout cohérent (« le roi noir est mis en échec par le cavalier blanc »). De toute évidence, un tel liage perceptif ne dépend pas de la conscience.

La vision d'une voix

Tous nos exemples, jusqu'à présent, viennent de la vision. Se pourrait-il que la conscience soit la « glu » qui rassemble les différentes modalités sensorielles en un tout cohérent ? Devons-nous être conscient pour fusionner les signaux visuels et auditifs, comme lorsque nous regardons un film ? À nouveau, la réponse est non. Notre cerveau assemble, inconsciemment, des informations issues de modalités sensorielles distinctes – et seule la fusion qui en résulte accède à notre conscience. Nous devons cette conclusion à une illusion remarquable appelée l'« effet McGurk », décrite pour la première fois par Harry McGurk et John MacDonald en 1976[31]. La vidéo de cette expérience, que l'on peut trouver

sur Internet[32], montre une personne qui parle, et il paraît évident qu'elle répète la syllabe *da da da da*. Quoi de plus banal ? Sauf que si vous fermez les yeux, vous vous rendez compte que le son qui parvient vraiment à vos oreilles est la syllabe *ba* ! Pourquoi, les yeux ouverts, entendez-vous *da* ? Parce que le film montre la bouche de quelqu'un qui dit *ga* ! Quand le cerveau entend *ba* mais voit *ga*, il est confronté à un conflit qu'il résout en fusionnant les deux informations. Lorsque la vision et l'audition sont suffisamment synchronisées, une opération inconsciente harmonise les entrées en un seul percept, la syllabe *da*. Celle-ci constitue un bon compromis, dans la mesure où les syllabes *ba*, *da* et *ga* forment un continuum de production acoustique d'avant en arrière de la bouche. En bon statisticien, le cerveau calcule, inconsciemment, l'interprétation la plus probable compatible avec toutes ses entrées.

Cette illusion auditive confirme à quel point l'expérience consciente est tardive et indirecte. Si surprenant que cela puisse paraître, nous n'avons pas conscience des ondes sonores qui font vibrer nos tympans, pas plus que des photons qui entrent dans nos yeux. Notre conscience ne nous donne jamais accès à une sensation brute, mais seulement à une reconstruction experte du monde extérieur. À notre insu, notre cerveau agit comme un détective qui recueille tous les indices, évalue leur solidité et les interprète en une histoire cohérente. Mais cette enquête se mène en catimini. Subjectivement, nous n'éprouvons pas l'impression d'avoir à inférer l'identité du son *da* – nous l'entendons, un point c'est tout. Il se trouve simplement que, dans l'effet McGurk, ce que nous entendons dépend au moins autant de notre vision que de notre audition.

Quelle région du cerveau concocte ce mélange savant des modalités sensorielles ? L'imagerie cérébrale suggère qu'il s'agit du cortex frontal, et non des aires auditives ou visuelles[33]. La synthèse perceptive est distillée dans les aires cérébrales de haut niveau, communes à toutes les modalités sensorielles, avant d'être renvoyée aux régions les plus précoces. De toute évidence, une énorme quantité de calculs s'effectue à notre insu afin

d'assembler la scène qui se joue devant nos yeux et nos oreilles, et que nous prenons, bien à tort, pour la simple donnée de nos organes des sens.

Pour autant, le cerveau peut-il fusionner inconsciemment n'importe quelles sources d'information ? C'est peu probable. La vision, la reconnaissance de la parole et l'expertise pour les échecs partagent un trait commun : ce sont des activités automatiques et surentraînées, et c'est pourquoi elles parviennent à fusionner les informations pertinentes sans que nous en ayons conscience. Le neurophysiologiste Wolf Singer propose l'existence de deux types de liages perceptifs[34]. Il y a d'abord les liages de routine, ceux que nous effectuons si couramment que notre cerveau leur dédie des neurones spécifiques – il en existe probablement qui fusionnent les lettres au sein d'un mot écrit, ou les consonnes que nous entendons avec les mouvements des lèvres que nous voyons. Mais nous avons aussi parfois besoin de créer des assemblages nouveaux, des combinaisons mentales inédites, et dans ce cas, il se peut que la conscience et la synchronie jouent des rôles essentiels.

Cette vision nuancée de la manière dont le cortex fait la synthèse de nos sensations paraît plausible. Dès la naissance, le cerveau est bombardé d'informations sur le monde. Des années d'interaction avec un environnement structuré lui permettent de recueillir un monceau de statistiques sur la fréquence avec laquelle différentes sensations surviennent soit simultanément, soit en étroite succession. Avec l'expérience, l'apprentissage dédie des ensembles de neurones spécifiques aux combinaisons récurrentes qui caractérisent les objets familiers[35]. Une fois établies, ces assemblées neuronales conservent leur sélectivité, même au cours d'une anesthésie générale – une preuve manifeste que cette forme de liage ne requiert pas la conscience. La reconnaissance des mots écrits doit beaucoup à ce mécanisme d'apprentissage statistique inconscient. À l'âge adulte, le lecteur moyen a vu des millions de mots, et son cortex visuel contient vraisemblablement des milliers de neurones dédiés à des chaînes de lettres fréquentes telles que *ent, pré,* ou *tion*[36].

Chez les grands maîtres d'échecs, certains neurones se spécialisent peut-être pour les configurations classiques de l'échiquier.

Cette forme de liage perceptif, inscrite dans des circuits spécialisés, diffère de l'assemblage des mots dans une phrase. Considérez cette citation de Raymond Devos : « Il m'est arrivé de prêter l'oreille à un sourd. Il n'entendait pas mieux. » Vous souriez parce que votre cerveau réunit, pour la première fois peut-être, cette combinaison de mots à double sens. Tout assemblage nouveau nécessite probablement la conscience. De fait, l'imagerie cérébrale suggère qu'au cours d'une anesthésie générale notre cerveau perd une bonne part de sa capacité d'intégration des mots dans les phrases[37].

Sens sans conscience ?

Notre système visuel assemble inconsciemment les lettres en mots – mais en calcule-t-il le sens ? Ou bien la compréhension d'un mot requiert-elle un traitement conscient ? La simplicité de cette question est trompeuse. Deux générations de scientifiques se sont battues avec cet épineux problème tant chaque camp avait la certitude absolue d'être dans le vrai.

Comment pourrait-on comprendre un mot sans en être conscient ? Si l'on définit la conscience comme « la perception, par une personne, de ce qui traverse son propre esprit », comme le fait John Locke dans son célèbre *Essai concernant l'entendement humain* (1690), on voit mal comment l'esprit pourrait saisir le sens d'un mot sans en être aussitôt conscient. La *compréhension* est, étymologiquement, l'appréhension d'un ensemble, l'assemblage de fragments de sens dans le « sens commun ». Comment ne serait-elle pas apparentée à la *conscience*, le « rassemblement du savoir », au point de lui être virtuellement synonyme ?

Cependant, dans le camp opposé, les arguments ne manquent pas. Comment fonctionnerait une langue dont le locuteur devrait d'abord prendre conscience de chaque mot isolément ? Quand vous lisez cette phrase, est-ce que vous devez délibérément réfléchir au sens de chaque mot avant de les assembler en un tout cohérent ? Non. Votre esprit va droit à l'essentiel, au « fond » du texte. Un coup d'œil suffit à juger du rôle de chaque mot au sein de la phrase. Nous n'avons pas la moindre introspection de la manière dont le signifiant évoque un signifié.

Qui est dans le vrai ? Il aura fallu trois décennies de recherche en psychologie et en imagerie cérébrale pour trancher cette controverse. Son histoire mérite d'être contée pas à pas – une longue valse-hésitation faite de conjectures et de réfutations et qui ne converge vers la vérité que fort tardivement.

Tout commence dans les années 1950, avec la découverte de l'« effet cocktail » (*cocktail party effect*[38]). Vous participez à un bruyant vernissage. Des dizaines de conversations s'entremêlent autour de vous. Pourtant, vous parvenez à vous concentrer sur les paroles de votre interlocutrice. Votre attention agit comme un filtre qui ne laisse passer qu'une voix et élimine toutes les autres. En êtes-vous bien sûr ? Le psychologue anglais Donald Broadbent postule que l'attention agit à un niveau précoce : les conversations auxquelles nous ne prêtons pas attention sont bloquées au niveau perceptif, bien avant qu'elles n'influencent notre compréhension[39]. Mais cette hypothèse ne résiste pas à l'épreuve des faits. Imaginez que l'un des convives, situé derrière vous, mentionne votre nom au détour d'une phrase, même chuchotée. Immédiatement, vous lui accordez votre attention – ce qui implique que votre cerveau ait traité les mots au moins jusqu'au niveau du sens des noms propres, et cela en l'absence d'attention[40]. Des expériences contrôlées confirment cette impression anecdotique, et montrent même que les mots auxquels nous n'accordons aucune attention influencent la compréhension de la conversation que nous essayons de suivre[41].

L'effet cocktail et bien d'autres expériences d'attention divisée suggèrent un accès inconscient au sens des mots. Pourtant, elles n'en apportent pas une preuve irréfutable. Les volontaires ont beau jurer qu'ils ne font attention qu'à une seule personne, sans jamais se laisser distraire par les autres conversations, il est légitime d'émettre quelques doutes. Comment prouver que l'attention ne se réoriente jamais, même pour un bref instant ? Ou que quelques mots filtrent à travers les silences de la première conversation ? L'effet cocktail est un phénomène frappant de la vie quotidienne, mais il est bien difficile de le transformer en un test rigoureux.

Dans les années 1970, Anthony Marcel, psychologue à Cambridge, fit faire un pas important au domaine. Il utilisa la technique de masquage pour présenter des mots en deçà du seuil de conscience. Sa méthode permettait une invisibilité parfaite : tous les participants, à tous les essais, niaient avoir vu le moindre mot. Même lorsqu'on leur expliquait qu'un mot était caché, ils ne parvenaient toujours pas à le voir. Si on exigeait d'eux qu'ils essaient de deviner, ils ne parvenaient même pas à dire si les lettres formaient un mot ou juste une chaîne de consonnes dépourvue de sens. Et pourtant… Marcel parvint à démontrer que leur cerveau en traitait inconsciemment le sens[42]. Dans une expérience clé, on présentait un nom de couleur tel que *bleu* ou *rouge* aux participants. Ces derniers juraient ne rien voir, mais lorsqu'ils devaient ensuite sélectionner une forme de la couleur correspondante, ils répondaient un vingtième de seconde plus vite que s'ils avaient été exposés à un autre mot sans rapport avec la couleur. Ainsi, un mot invisible amorçait la couleur correspondante – ce qui semblait impliquer que son sens avait été extrait.

D'autres expériences de Marcel mirent en évidence un phénomène étonnant : le cerveau évalue inconsciemment tous les sens possibles d'un mot ambigu, même ceux qui ne sont pas pertinents dans le contexte[43]. Supposez que je vous chuchote le mot *vol*. La peur d'un larcin vous vient à l'esprit (« la propriété, c'est le vol ») –, mais à la réflexion, peut-être voulais-je évoquer le déplacement aérien (« ô temps, suspends ton vol »).

Nous ne pouvons pas réfléchir consciemment à plus d'une acception à la fois. C'est le contexte qui décide : au sein d'un film policier, le mot *vol* évoquera vraisemblablement l'hypothèse d'un cambriolage. L'expérience démontre effectivement que le contexte d'un seul mot suffit à biaiser le sens d'un autre mot : le mot « vol » amorce le mot « bijou » lorsqu'il est précédé du mot « police », mais il amorce le mot « avion » s'il est précédé du mot « plume[44] ».

Or cette adaptation au contexte semble ne survenir qu'au niveau conscient. Quand le mot-amorce est masqué au point de devenir invisible, les recherches de Marcel démontrent une activation conjointe des deux sens de ce mot. Le mot « vol », devenu invisible, amorce alors aussi bien « bijou » qu'« avion » – même dans un contexte qui privilégie fortement l'hypothèse d'un cambriolage. Les aires du langage sont d'une telle efficacité qu'elles parviennent à retrouver en mémoire, inconsciemment et simultanément, toutes les associations sémantiques d'un mot, même s'il est ambigu et même lorsque seul l'un d'entre eux convient au contexte. L'inconscient propose, le conscient choisit.

La grande bataille de la sémantique inconsciente

Les expériences d'amorçage sémantique d'Antony Marcel étaient remarquablement novatrices. Elles suggéraient fortement qu'un mot invisible était traité au niveau sémantique. Pourtant, elles n'étaient pas parfaites, et les sceptiques s'engouffrèrent dans la brèche[45]. Leurs doutes déclenchèrent une formidable bataille intellectuelle entre les défenseurs et les détracteurs de l'hypothèse d'un traitement sémantique inconscient.

Ces doutes n'étaient pas entièrement injustifiés. Après tout, l'effet d'amorçage subliminal qu'avait observé Marcel était si faible qu'il en devenait presque négligeable. Un mot masqué n'accélérait les mots reliés

que d'une quantité minuscule, de l'ordre d'un centième de seconde. Pouvait-on certifier que cet effet ne provenait pas d'une toute petite fraction d'essais où le mot aurait été vu – mais si brièvement que cet épisode ne laisserait guère de trace en mémoire ? Selon ses détracteurs, Marcel ne vérifiait pas suffisamment bien que toutes ses amorces étaient invisibles. En effet, ce n'était qu'à la fin de l'expérience que les volontaires rapportaient n'avoir vu aucun des mots. La visibilité des amorces devait être mesurée avec plus de soin, dans une expérience distincte où l'on demandait aux volontaires de deviner le mot caché ou de le catégoriser selon des critères simples. Selon les sceptiques, seule une réponse strictement aléatoire à cette tâche secondaire indiquait que les amorces étaient strictement invisibles. Encore fallait-il que ce contrôle se déroule dans des conditions strictement identiques à l'expérience principale – mêmes stimuli, même environnement lumineux, même degré d'entraînement des sujets, etc. Or les expériences de Marcel, soit ne remplissaient pas ces critères, soit, lorsqu'elles le faisaient, suggéraient que les participants avaient effectivement dû voir quelques mots.

En réponse à ces critiques, les défenseurs du traitement inconscient resserrèrent tous les boulons de leur travail – et pourtant, les résultats continuèrent à indiquer un traitement inconscient du sens des mots, des chiffres et même des images[46]. En 1996, Anthony Greenwald, psychologue à l'Université de Seattle, publia dans la revue *Science* une étude qui semblait apporter la preuve définitive d'un accès à la connotation émotionnelle d'un mot invisible. Il avait demandé à des volontaires de classifier des mots selon leur connotation émotionnelle positive ou négative. À leur insu, chaque cible visible était précédée d'une amorce invisible. Dans la moitié des essais, dits « congruents », les deux mots allaient dans le même sens (ils étaient tous deux positifs, comme « joyeux » suivi de « paix », ou tous deux négatifs, comme « viol » suivi de « mort »), tandis que dans l'autre moitié des essais, dits « incongruents », ils étaient de sens opposé (« viol » suivi de « joyeux »). Or, lorsqu'ils répondaient très vite, les volontaires faisaient moins d'erreurs dans les

essais congruents que dans les essais incongruents. Les significations des deux mots semblaient se combiner inconsciemment, ce qui facilitait la décision finale lorsqu'ils partageaient la même connotation émotionnelle et qui la ralentissait dans le cas contraire.

Les résultats de Greenwald étaient fortement reproductibles. Dans toute une série d'expériences, non seulement les participants juraient n'avoir vu aucune des amorces, mais ils étaient objectivement incapables de les nommer ou d'identifier leur contenu émotionnel. De plus, leurs performances dans ces tâches de devinette inconsciente ne prédisaient en rien la taille de leur effet d'amorçage. Ce dernier ne paraissait donc pas provenir d'un petit groupe de sujets qui auraient vu certains mots. Nous tenions enfin la démonstration irréfutable que le sens des mots pouvait s'activer sans conscience.

Irréfutable ? Voire… car quelques années plus tard, Tony Greenwald, avec son étudiant Richard Abrams, révisa complètement son interprétation des résultats[47]. Il fit remarquer que son expérience n'utilisait qu'une petite liste de mots, toujours les mêmes. Se pourrait-il que les participants finissent par associer chaque chaîne de lettres à la réponse appropriée, en court-circuitant totalement l'accès au sens ? Cette hypothèse n'était pas absurde, car les mêmes mots étaient utilisés comme amorces et comme cibles, et il fallait toujours les classer selon la même règle. Après avoir décidé plus de vingt fois que le mot « joie » évoque une émotion positive, peut-être le cerveau associe-t-il directement les lettres *j-o-i-e* avec la réponse du côté « positif », sans faire appel au sens[48].

Malheureusement, cette conjecture s'avéra correcte. Dans les expériences de Greenwald, l'amorçage était effectivement subliminal, mais il ne passait pas par le sens. Tout d'abord, Greenwald montra qu'un mélange aléatoire de lettres avait autant d'effet que le mot de départ : « ioej » et « joie » facilitaient tout autant les émotions positives. Ensuite, il manipula finement la ressemblance des mots visibles avec ceux qui servaient d'amorces. Dans l'expérience clé, les cibles comprenaient les mots *tulip* et *humor* (« tulipe » et « humour »), que les participants

anglophones jugeaient évidemment positifs. Greenwald recombina leurs lettres pour former un mot négatif, *tumor* (« tumeur »), qui n'apparaissait que dans les amorces invisibles.

Résultat : inconsciemment, ce mot *négatif* amorçait la réponse *positive*. Pourquoi ? Parce que inconsciemment le cerveau des participants classait les lettres de *tumor* au côté des mots *tulip* et *humor*, sans tenir compte du sens radicalement différent de ces mots. Voilà bien la preuve que l'amorçage ne dépend que d'une association superficielle entre les lettres et les réponses. Les recherches de Greenwald concernaient bien la perception inconsciente, mais pas le sens profond des mots. Au moins dans ces conditions, le traitement inconscient n'est guère sophistiqué : au lieu de concerner le sens des mots, il ne porte que sur la relation entre les lettres et les réponses.

Anthony Greenwald venait de réduire à néant son propre article dans *Science*.

Arithmétique inconsciente

En 1998, les preuves d'un traitement sémantique inconscient manquaient toujours. Cependant, les expériences de Greenwald n'étaient pas le mot de la fin. Elles présentaient une particularité inhabituelle : on exigeait des participants qu'ils répondent en moins de 400 millisecondes. Ce délai semblait trop court pour que le cerveau accède au sens de mots relativement rares comme « tumeur ». En un temps aussi bref, il n'était guère surprenant que le cerveau ne fasse qu'associer les lettres aux réponses. Peut-être parviendrait-il à analyser le sens d'un mot invisible si on lui laissait plus de temps ? C'est ainsi que Lionel Naccache et moi-même avons commencé une série d'expériences qui allaient finalement prouver que le sens d'un mot pouvait être activé sans conscience[49].

Pour maximiser les chances d'obtenir un effet inconscient, nous avons choisi l'une des catégories de mots les plus simples : les nombres. En effet, les nombres de 1 à 10 sont particuliers. Ce sont des mots courts, très fréquents et appris dès l'enfance. Leur sens est absolument transparent. Enfin, ils s'écrivent avec une compacité extrême : un simple chiffre. Dans notre expérience, nous présentions comme amorce subliminale un nombre, choisi parmi 1, 4, 6 et 9, précédé et suivi d'une chaîne aléatoire de lettres qui le rendait totalement invisible. Immédiatement après venait un second nombre, la cible, clairement visible.

Nous donnions à nos participants une consigne simplissime : décidez, le plus vite possible, si le nombre cible est plus petit ou plus grand que 5. Ils ignoraient tout de la présence d'un nombre caché. À la fin de l'expérience, nous prouvions que, même si on leur expliquait la présence de ces amorces invisibles, ils ne parvenaient pas à dire si elles étaient plus petites ou plus grandes que 5. Et pourtant, les amorces influençaient leur réponse. Lorsque les deux nombres étaient congruents (tous deux du même côté de 5), les participants répondaient plus vite que lorsqu'ils étaient incongruents (l'un plus petit et l'autre plus grand que 5). Ainsi, l'apparition d'un 9 invisible accélérait les réponses à 9 et à 6, mais ralentissait les réponses à 4 et à 1.

L'imagerie cérébrale nous permit ensuite de suivre cet effet à la trace dans le cortex. Elle mit en évidence une toute petite activation du cortex moteur qui commandait la main qu'il aurait fallu utiliser pour répondre au nombre caché. Tout se passait comme si le cerveau commençait par réagir à cette amorce invisible, avant de changer d'idée et de se focaliser sur la cible visible. Un vote inconscient traversait tout le cerveau, depuis la perception jusqu'au contrôle moteur (figure 11). Cet effet impliquait que le cerveau pouvait classifier, inconsciemment, le sens d'un symbole numérique.

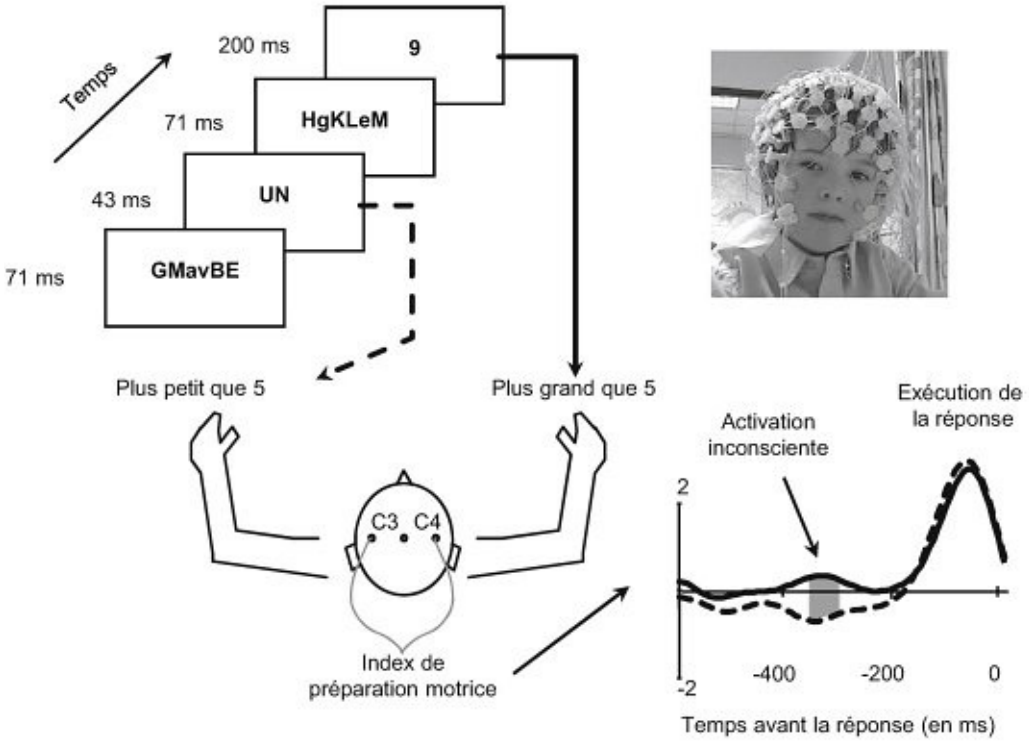


Figure 11. Le cortex moteur se prépare inconsciemment à répondre. Dans cette expérience, on demande à un volontaire de juger si des nombres sont plus petits ou plus grands que 5. Juste avant, un autre nombre est présenté inconsciemment (ici la cible est « 9 » et l'amorce inconsciente « 1 »). Bien qu'invisible, le nombre masqué parvient à activer le cortex moteur qui commande la main correspondante. Ainsi un symbole peut-il être inconsciemment identifié, classifié selon une règle arbitraire, et le résultat envoyé jusque dans les aires motrices.

Nos recherches ultérieures achevèrent de convaincre les sceptiques. Notre effet d'amorçage subliminal ne dépendait pas de la notation utilisée : le mot *quatre* amorçait le chiffre *4* tout autant que la simple répétition de *4* suivi de *4*, ce qui suggérait que l'effet provenait intégralement d'une représentation mentale du sens de ces symboles. Par la suite, nous avons démontré que l'amorçage persistait même lorsque l'amorce était un symbole *écrit* et la cible un mot *parlé*[50].

Dans l'expérience de départ, l'effet était éventuellement imputable à une association directe et « a-sémantique » entre la vision et les réponses – le même problème qui avait détruit l'interprétation de l'expérience de

Greenwald. Mais l'amorçage numérique, contrairement à l'amorçage émotionnel, résistait à la critique. En effet, même des amorces cachées qui n'apparaissaient jamais de façon visible, et qui ne pouvaient donc pas faire l'objet d'un lien sensori-moteur direct, causaient un effet substantiel[51]. L'IRM fonctionnelle montrait également que le nombre invisible activait les aires du lobe pariétal associées au sens des quantités[52]. Ces régions codent pour la quantité numérique associée à chaque nombre[53]. Chez le singe, on y trouve des neurones qui répondent sélectivement à la quantité concrète que perçoit l'animal[54] (entre 1 et 30 objets). Or, au cours de l'amorçage subliminal, leur activité diminue dès que l'on répète le même nombre (par exemple *neuf* suivi de 9). Ce phénomène classique, appelé « habituation » ou « suppression liée à la répétition », indique que les neurones ont reconnu que la même quantité revient deux fois de suite. Il semble que les neurones qui codent pour la quantité s'habituent à voir le même nombre, même quand la première présentation n'est pas perçue consciemment.

Les données devenaient de plus en plus solides : une aire de haut niveau, qui codait pour le *sens* des nombres, pouvait s'activer sans conscience. Le point final fut mis lorsque nos collègues démontrèrent que l'effet d'amorçage dépendait de la proximité numérique entre l'amorce et la cible[55]. L'effet était maximal lorsque les nombres étaient répétés à l'identique (par exemple quand l'amorce était le mot *quatre*, et la cible le chiffre *4*). Il diminuait un peu lorsque les nombres étaient consécutifs (*trois* suivi de *4*), nettement plus s'ils étaient séparés de deux unités (*deux* suivi de *4*), et ainsi de suite. Cet effet de distance numérique est la marque indubitable d'un accès au sens des nombres. Il ne peut exister que si le cerveau de l'observateur a saisi que la quantité *4* est plus proche de *3* que de *2* ou de *1* – la preuve irréfutable qu'un symbole invisible peut être compris inconsciemment.

Une combinaison inconsciente de concepts

Le dernier recours des sceptiques fut de supposer que les nombres étaient des objets mentaux d'une nature particulière. Cette toute petite catégorie de mots, estimaient-ils, nous est tellement familière qu'il n'est guère surprenant qu'on les comprenne automatiquement et sans réfléchir. Il en irait différemment d'autres catégories de mots, pensaient-ils. Mais ce dernier bastion de résistance tomba lorsqu'on découvrit des effets de congruence sémantique avec des mots masqués qui n'étaient pas des nombres[56]. Par exemple, la décision que le mot *piano* représente un objet et non un animal s'accélère lorsque celui-ci est précédé de la présentation inconsciente d'un autre nom d'objet, par exemple *chaise*, et elle est ralentie par un nom d'animal tel que *chat* – même si ces amorces subliminales ne sont jamais perçues au cours de l'expérience.

L'imagerie cérébrale a également confirmé les conclusions issues du comportement. Certains enregistrements neuronaux ont révélé l'activité sémantique induite par un mot invisible. Dans l'une de nos études, mes collègues et moi avons recueilli les signaux d'électrodes implantées directement dans les régions sous-corticales spécialisées pour les émotions[57]. Bien entendu, cet examen ne se pratique pas chez le volontaire sain, mais seulement chez certains patients épileptiques. Dans de nombreux hôpitaux de la planète, il est devenu monnaie courante d'implanter de longues électrodes à l'intérieur de la boîte crânienne, afin d'identifier les sources des crises d'épilepsie et de les exciser au cours d'une intervention chirurgicale. Entre deux crises, avec l'accord du patient, les électrodes peuvent également aider la recherche scientifique. Elles donnent un accès inespéré à l'activité neuronale d'une petite région du cerveau – parfois même aux décharges d'un neurone unique.

Dans notre cas, les électrodes pénétraient dans l'amygdale, une structure cérébrale qui répond à toutes sortes de stimuli qui évoquent la peur, depuis les serpents et les araignées jusqu'à la musique d'un film d'horreur ou au visage d'un inconnu. Même la présentation subliminale de

la photo d'un serpent ou d'un visage effrayé suffit à l'activer[58]. S'activerait-elle également en réponse à un *mot* effrayant ? Pour le savoir, nous avons présenté des mots subliminaux au contenu alarmant, tels que *viol*, *danger* ou *poison*. Pour notre plus grand plaisir, ces mots déclenchaient une réponse électrique qui n'existait pas pour des mots neutres tels que *voile*, *manger* ou *poisson*. L'amygdale « voyait » ce que voulaient dire des mots dont le patient lui-même n'avait aucune conscience.

L'effet était étonnamment lent : une demi-seconde au minimum s'écoulait avant qu'un mot invisible n'entraîne un signal émotionnel inconscient. À l'instant même où son amygdale déchargeait, le patient niait avoir vu le moindre mot, et il ne parvenait pas à le deviner. Ainsi, un mot invisible faisait son petit bonhomme de chemin dans le cerveau, était identifié et même compris, sans pour autant accéder à la conscience.

L'amygdale ne fait pas partie du cortex, mais d'autres expériences prouvent que même les régions corticales du langage répondent également aux mots inconscients. Ces expériences s'appuient sur une onde cérébrale qui signale la détection d'une anomalie sémantique. Imaginez que vous entendiez la phrase suivante : « À midi, j'ai mangé mon steak avec un couteau et une chaussette. » À l'écoute du mot saugrenu, votre cerveau émet une onde cérébrale qu'on appelle la N400 (le *N* fait référence à son signe, car le voltage est négatif au sommet de la tête, et le *400* à sa latence, de l'ordre de 400 millisecondes après l'apparition du mot).

L'onde N400 traduit une opération de haut niveau qui évalue si un mot convient dans le contexte actuel de la phrase. Son amplitude dépend du degré d'absurdité : les mots justes engendrent une toute petite N400, tandis que les mots foncièrement inappropriés sont suivis d'une onde très intense. Or cette modulation continue d'exister même pour des mots que nous ne voyons pas – qu'ils soient rendus invisibles par le masquage[59] ou par l'inattention[60]. Les réseaux de neurones du lobe temporal ne se

contentent pas d'évaluer le sens des mots, mais également leur compatibilité avec le contexte.

Dans un travail récent, Simon Van Gaal et moi-même avons montré que l'onde N400 reflète la combinaison inconsciente de plusieurs mots invisibles[61]. Notre expérience consistait à présenter deux mots rendus strictement invisibles par des masques. Leur combinaison évoquait une connotation qui allait du positif au négatif : *très heureuse*, *pas heureuse*, *pas triste* ou *très triste*. Tout de suite après, apparaissait un mot bien visible, lui aussi positif ou négatif (par exemple *guerre* ou *fête* ; toutes les combinaisons faisaient sens en hollandais, langue dans laquelle cette expérience fut réalisée). L'onde N400 émise par le mot conscient était modulée par le contexte global : non seulement le mot *guerre* engendrait une grande N400 lorsqu'il était précédé du mot *heureuse*, mais cet effet lui-même était modulé, dans un sens ou dans l'autre, par l'intensificateur *très* ou la négation *pas*. Inconsciemment, le cerveau remarquait l'incongruité d'une « très heureuse guerre ». Cette expérience constitue l'une des meilleures démonstrations que le cerveau continue de traiter, inconsciemment, la syntaxe et le sens d'une expression linguistique invisible[62].

L'aspect le plus étonnant de ces expériences, c'est que l'onde N400 ne change même pas d'amplitude, que les mots soient visibles ou invisibles. Ainsi, la conscience semble n'avoir aucun d'impact sur le traitement du sens – le cerveau effectue les mêmes calculs, jusqu'au niveau sémantique, que nous en soyons conscients ou non. Cela signifie également que les mots masqués n'évoquent pas toujours des activations cérébrales minuscules. Les réponses corticales sont parfois intenses, alors même que les stimuli qui les causent demeurent inaccessibles à la conscience.

Une conclusion s'impose : un mot invisible peut parfaitement entraîner l'activation d'un vaste circuit cérébral spécialisé pour la représentation du sens. Nuance importante, toutefois : la reconstruction des sources corticales de ces effets sémantiques montre que l'activité inconsciente reste confinée au lobe temporal, qui est le siège principal des

réseaux de la compréhension du langage[63]. Dans les prochains chapitres, au contraire, nous verrons qu'un mot conscient envahit des réseaux bien plus larges, notamment dans le lobe frontal, qui confèrent le sentiment d'avoir le mot « en tête ». En dernière analyse, les mots invisibles ne sont donc pas aussi influents que les mots conscients.

Attentif, mais inconscient

La découverte qu'un mot ou qu'un chiffre pouvaient traverser tout le cerveau, biaiser nos décisions et moduler nos aires du langage, sans pour autant accéder à la conscience, a créé un déclic chez de nombreux chercheurs. Nous avions clairement sous-estimé le pouvoir de l'inconscient. Nos intuitions n'étaient pas dignes de confiance : il était impossible de savoir à l'avance quels processus cognitifs exigeaient une perception consciente et lesquels pouvaient se dérouler inconsciemment. Cette question était strictement empirique. Il nous fallait passer au crible de l'expérience chacune de nos facultés cognitives, une à une, et décider si elles nécessitaient ou non un accès à la conscience. Seule l'expérimentation permettait de répondre – et fort heureusement, en s'appuyant sur les techniques de masquage ou de détournement de l'attention, jamais il n'avait été aussi facile d'évaluer la profondeur et les limites du traitement inconscient.

Les dix dernières années ont ainsi vu fleurir des expériences aux résultats surprenants, qui ont profondément remis en cause notre vision de la conscience. Prenez l'attention. Rien ne semble plus proche de la conscience que la capacité à faire attention. En l'absence d'attention, même des stimuli énormes peuvent demeurer invisibles – c'est le cas du gorille de Dan Simons et d'une myriade d'autres effets. Lorsque plusieurs stimuli sont en compétition, l'attention sert de porte d'entrée à l'expérience consciente[64]. Dans ces conditions au moins, la conscience

requiert l'attention. Étonnamment pourtant, l'inverse n'est pas vrai : plusieurs expériences récentes démontrent que nous pouvons faire attention à un stimulus sans pour autant en devenir conscient[65].

À la réflexion, d'ailleurs, il serait bien étrange que l'attention requière une supervision consciente. Comme l'a fait remarquer William James, l'attention assure la sélection d'« un objet parmi plusieurs qui semblent possibles ». Il serait extraordinairement inefficace que notre esprit soit constamment distrait par des dizaines, voire des centaines ou des milliers de pensées possibles, que nous devrions examiner consciemment, l'une après l'autre. Déterminer lesquelles sont les plus pertinentes et doivent faire l'objet d'une analyse approfondie ne saurait être l'apanage de la conscience : il est indispensable que cette opération se déroule à notre insu, en parallèle, et que son résultat seul devienne conscient. Et c'est exactement ce que montrent les travaux récents : toute une armée d'opérateurs inconscients s'occupe d'orienter le faisceau de l'attention et de trier les monceaux de données qui parviennent à nos sens – jusqu'à ce que l'un d'entre eux alerte la conscience de la présence d'une pépite pertinente.

Imaginez que nous vous présentions un flash lumineux au coin de l'œil, si brièvement que vous n'en ayez pas conscience. Plusieurs expériences montrent que ce flash attire tout de même votre attention : bien que vous n'ayez pas la moindre idée de ce qui a attiré votre regard, votre attention s'oriente vers cet endroit, et vous répondez plus vite aux images qui y apparaissent[66]. Inversement, une image invisible vous ralentit si son contenu n'est pas pertinent pour votre tâche actuelle. Curieusement, cet effet d'attraction automatique de l'attention fonctionne mieux quand les images distrayantes sont invisibles que quand elles sont visibles : avec un effort de votre volonté, vous arrivez à réprimer une distraction dont vous percevez consciemment la présence, tandis qu'une distraction inconsciente conserve tout son pouvoir de nuisance parce que notre cerveau ne parvient pas à la contrôler[67].

Un bruit soudain ou une image qui clignote attirent automatiquement l'attention. Ces stimulations envahissent notre esprit, quels que soient nos efforts pour les ignorer. Pourquoi ? Parce qu'elles sollicitent un système cérébral qui nous alerte des dangers potentiels. Même lorsque nous souhaitons nous consacrer exclusivement à notre déclaration d'impôts ou à notre jeu vidéo favori, notre cerveau ne peut pas se couper totalement du monde. Quel que soit notre degré de concentration, il est indispensable que certains signaux d'alerte, comme un cri ou l'appel de notre nom, puissent toujours interrompre notre train de pensées. C'est pourquoi le filtre que nous appelons l'« attention sélective » reste toujours actif : indépendamment de notre état actuel de conscience, c'est lui qui décide si un signal externe mérite une analyse approfondie. L'attention inconsciente doit toujours rester sur le qui-vive.

Les psychologues ont longtemps cru que seuls les processus automatiques et ascendants (*bottom-up*) pouvaient fonctionner inconsciemment. Leur métaphore favorite était celle d'une « dissémination de l'activation » (*spreading activation*) : un stimulus active les récepteurs sensoriels, et cette activation se répand passivement et grimpe, de proche en proche, à travers la hiérarchie des aires cérébrales. Selon cette théorie, une image subliminale contacte successivement les processeurs spécialisés dans l'identification, l'accès au sens et la décision motrice, sans que ces étapes soient influencées par la volonté consciente, l'intention ou l'attention du sujet. On estimait donc que l'amorçage subliminal évaluait les automatismes des participants, indépendamment de leurs stratégies ou de leurs attentes[68].

Imaginez donc quelle a été notre surprise lorsque nos expériences ont fait voler en éclats ce consensus établi. Nous avons prouvé que l'amorçage subliminal n'était absolument pas un processus passif, dégagé des influences de l'attention et des instructions. Au contraire, l'attention y joue un rôle déterminant[69]. Une image subliminale qui apparaît à un moment ou à un endroit inattendu ne produit pratiquement aucun effet d'amorçage. Même l'effet de répétition – l'accélération de la réponse à

radio lorsqu'il est précédé de *radio* – varie en fonction de l'attention allouée à ces stimuli. Le simple fait de prêter attention entraîne une amplification massive des ondes cérébrales évoquées par les stimuli présentés au moment et à l'endroit attendu. Étonnamment, les stimuli inconscients bénéficient de ce coup de projecteur attentionnel tout autant que les stimuli conscients. En d'autres termes, l'attention peut très bien amplifier une sensation périphérique, sans pour autant que cela suffise à la rendre accessible à la conscience.

Nos intentions conscientes affectent même l'orientation de notre attention inconsciente. Supposez que l'on vous présente un ensemble de formes, avec l'instruction de détecter les carrés sans tenir compte des ronds. Lors de certains essais, un carré apparaît à droite de l'écran, et un cercle à gauche – mais tous deux sont masqués, si bien qu'ils demeurent invisibles. Vous répondez au hasard, parce que vous n'avez pas la moindre idée consciente du côté où est apparu le carré. Et pourtant… une onde cérébrale issue du cortex pariétal, appelée la N2pc, montre que votre attention s'est inconsciemment orientée du bon côté[70]. La présence de la cible (un simple carré) attire subrepticement l'attention, même dans les essais que vous jugez invisibles et où vous finissez par répondre du mauvais côté. De même, pendant le clignement attentionnel, au sein d'une série de lettres, le symbole que l'on a arbitrairement désigné comme cible entraîne nettement plus d'activité que les autres, même s'il demeure indétectable[71]. Ces expériences prouvent que notre attention peut très bien commencer son travail de filtrage, en tenant compte de la pertinence de chaque stimulus pour nos buts actuels, sans que ce coup de projecteur ne suffise à la prise de conscience.

La valeur d'une pièce de monnaie inconsciente

Comment l'attention décide-t-elle de la pertinence d'un stimulus ? La sélection s'appuie sur un mécanisme clé : l'attribution, par le cerveau, d'une *valeur* à chacune de nos pensées potentielles. Pour survivre, tout animal doit, très vite, assigner une valence positive ou négative à chaque événement nouveau. Dois-je rester ou partir ? M'approcher ou battre en retraite ? L'estimation de la valeur des situations que nous rencontrons repose sur des circuits cérébraux spécialisés situés au sein des « ganglions de la base », des aires sous-corticales situées à la base du cerveau. Et, comme vous l'avez probablement deviné, ces circuits, une fois de plus, échappent à notre introspection. Même l'attribution d'une valeur purement symbolique et culturelle à l'argent s'effectue sans conscience.

L'une des expériences qui le prouve consiste à motiver les participants avec de l'argent… invisible (figure 12[72]). On leur demande de serrer une manette le plus fort possible, en leur expliquant que s'ils parviennent à exercer une certaine force, ils toucheront de l'argent. Au début de chaque essai, l'image d'une pièce de monnaie indique la somme en jeu : un penny ou une livre (les participants sont anglais). Cependant, cette image est parfois masquée au point de devenir invisible. Résultat : alors même qu'ils nient avoir vu la mise, les participants exercent une force plus grande lorsque celle-ci est plus élevée. De plus, avant même de commencer, leurs mains suent légèrement plus – une anticipation inconsciente de la récompense à venir. Enfin, les circuits de la récompense s'activent subrepticement. Les volontaires eux-mêmes n'en ont aucune conscience : leur comportement varie d'un essai à l'autre, sans qu'ils se rendent compte que leur motivation a été manipulée à leur insu.

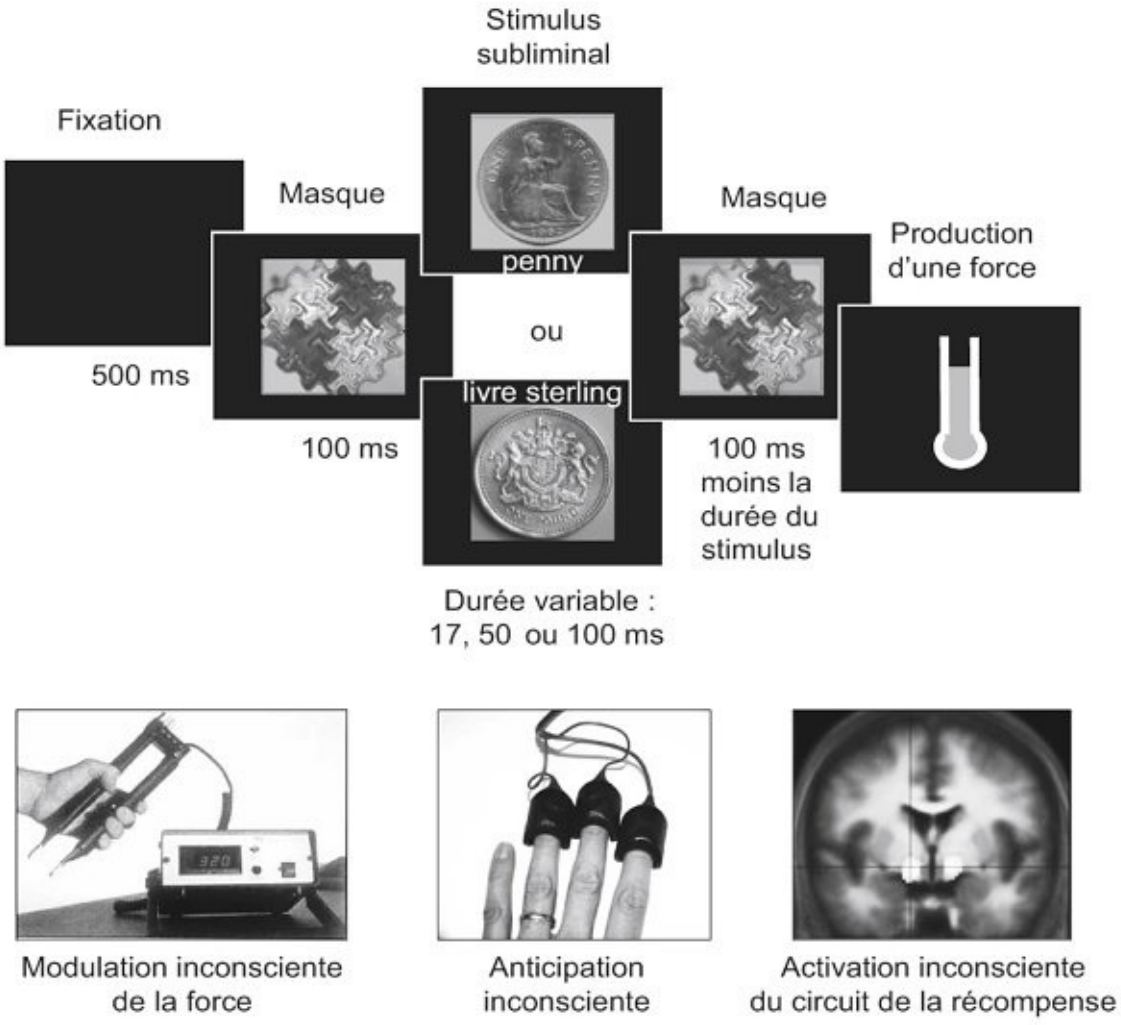


Figure 12. Nos motivations ne sont pas toujours conscientes. Dans cette expérience, on demande aux participants de serrer une pince le plus fort possible afin de gagner de l'argent. Lorsque l'enjeu est d'une livre plutôt que d'un penny, les sujets sont plus motivés et ils serrent la pince plus fort. Étonnamment, ils continuent de le faire même si la pièce de monnaie est rendue invisible par masquage. Les circuits cérébraux de la récompense sont préactivés inconsciemment, et même les mains du sujet se mettent à suer légèrement plus lorsque l'enjeu augmente. Ainsi, même une image inconsciente déclenche les circuits de la motivation, de l'émotion et de la récompense.

Dans une variante de cette expérience, les sujets ne connaissent pas la valeur des mises, et pourtant ils les apprennent… sans jamais les voir[73] ! On leur explique qu'à chaque fois qu'ils voient un « signal » ils doivent

deviner s'il vaut mieux appuyer sur un bouton ou non. Une fois qu'ils ont fait leur choix, ils apprennent s'ils ont gagné ou perdu de l'argent. À leur insu, une forme invisible, cachée dans le « signal », indique la bonne réponse. L'une des formes indique qu'il faut répondre pour recevoir une récompense, une autre qu'il faut rester immobile, et une troisième indique que chacun de ces deux choix a une chance sur deux d'être récompensé.

Après quelques minutes de jeu, les sujets acquièrent une inexplicable prescience de la bonne réponse. Ils ignorent tout des formes cachées, et pourtant leur intuition s'améliore au point qu'ils finissent par gagner une belle somme d'argent. Pourquoi ? Parce que leur système d'évaluation inconsciente apprend à attribuer, à chaque forme, l'action qui convient. L'imagerie cérébrale montre qu'une région spécifique des ganglions de la base, appelée *striatum* ventral, alloue une valeur appropriée à chaque action, en tenant compte du contexte invisible. Ainsi, les symboles invisibles acquièrent un sens : l'un devient attractif, l'autre répulsif, et leur valeur module la prise de décision.

La conclusion est claire : notre cerveau abrite toute une série de dispositifs astucieux mais inconscients, qui évaluent en permanence les entrées sensorielles, guident notre attention et orientent nos pensées. Grâce à ces étiquettes subliminales, les stimuli amorphes qui bombardent nos sens deviennent un paysage familier au sein duquel notre cerveau distingue aisément les aubaines à saisir et les risques à éviter. Seuls les événements les plus saillants et les plus pertinents attirent notre attention et acquièrent ainsi une petite chance d'accéder à la conscience. À notre insu, notre attention inconsciente évalue sans relâche les opportunités qui se présentent à l'aune de nos buts et de nos systèmes de valeur.

Mathématiques inconscientes

> « Pour bien comprendre la vie psychique, il est indispensable de cesser de surestimer la conscience. »
>
> Sigmund FREUD, *L'Interprétation des rêves* (1900).

Freud a raison : nous ne cessons de surestimer la conscience. Voici un truisme moins trivial qu'il n'y paraît : nous n'avons conscience que de nos pensées… conscientes. Parce que les opérations inconscientes nous échappent, par définition, nous exagérons l'importance de la conscience dans notre vie mentale. Aveugles au prodigieux pouvoir de l'inconscient, nous accordons trop de crédit à l'introspection consciente et sous-estimons en permanence l'importance des vraies raisons inconscientes de nos comportements. Comme l'a si bien écrit Julian Jaynes, « la conscience est une partie de notre vie mentale beaucoup plus petite que nous n'en sommes conscients, parce que nous ne pouvons pas être conscients de ce dont nous n'avons pas conscience[74] ». En paraphrasant l'esprit fantasque de la loi autoréférentielle de Douglas Hofstadter (« Un projet prend toujours plus longtemps qu'on ne le pense – même en tenant compte de la loi d'Hofstadter »), on pourrait élever cette observation au rang de loi universelle :

> Nous surestimons constamment la conscience – même lorsque nous avons conscience des limites de notre conscience.

Corollaire immédiat : nous *sous*-estimons considérablement à quel point les opérations de la vision, du langage ou de l'attention se déroulent hors de la portée de notre introspection. Se pourrait-il, dès lors, que certains calculs que nous considérons comme l'apanage de la conscience s'effectuent en réalité inconsciemment ? Prenons l'exemple des mathématiques. L'un des plus grands mathématiciens de tous les temps, Henri Poincaré, attribuait ses plus grandes découvertes à la fécondité de son inconscient :

> « À ce moment, je quittai Caen, où j'habitais alors, pour prendre part à une course géologique entreprise par l'École des mines. Les péripéties du voyage me firent vite oublier mes travaux mathématiques ; arrivés à Coutances, nous montâmes dans un omnibus pour je ne sais quelle promenade ; au moment où je mettais le pied sur le marchepied, l'idée me vint, sans que rien dans mes pensées antérieures parût m'y avoir préparé, que les transformations dont j'avais fait usage pour définir les fonctions fuchsiennes étaient identiques à celles de la géométrie non euclidienne. Je ne fis pas la vérification, je n'en aurais pas eu le temps puisque à peine dans l'omnibus, je repris la conversation commencée ; mais j'eus tout de suite une entière certitude. De retour à Caen, je vérifiais le résultat à tête reposée pour l'acquit de ma conscience. »

Et encore :

> « Je me mis alors à étudier des questions d'arithmétique sans grand résultat apparent et sans soupçonner que cela pût avoir le moindre rapport avec mes études antérieures. Dégoûté de mon insuccès, j'allais passer quelques jours au bord de la mer et je pensais à autre chose. Un jour, en me promenant sur la falaise, l'idée me vint, toujours avec le même caractère de brièveté, de soudaineté et de certitude immédiate, que les transformations arithmétiques des formes quadratiques ternaires indéfinies étaient identiques à celles de la géométrie non euclidienne. »

Ces deux anecdotes sont rapportées par Jacques Hadamard, lui-même mathématicien émérite, et qui consacra un livre fascinant à l'esprit des grands mathématiciens[75]. Hadamard proposait de décomposer le processus de la découverte mathématique en quatre étapes successives : la

préparation, l'incubation, l'illumination et la vérification. La *préparation* recouvre toute la phase d'exploration délibérée d'un problème. Cette attaque frontale, hélas, demeure souvent vaine – ce qui ne signifie pas qu'elle soit inutile, car c'est elle qui déclenche la réflexion inconsciente. La phase d'*incubation* peut alors démarrer – une période de maturation invisible durant laquelle l'esprit reste vaguement préoccupé par le problème, sans toutefois montrer le moindre signe d'une ardente cogitation. L'incubation demeure invisible mais son impact est immense. Soudain, après une nuit de sommeil, une bonne marche ou une douche, survient l'*illumination* : la solution apparaît dans toute sa splendeur. Elle est souvent juste, mais il reste encore à la soumettre au crible de la *vérification* consciente, qui en dissèque chaque détail.

La théorie d'Hadamard est séduisante, mais résiste-t-elle à l'épreuve des faits ? L'incubation inconsciente est-elle une réalité ou seulement une reconstruction imaginaire suscitée par l'euphorie de la découverte ? Pouvons-nous vraiment résoudre inconsciemment des problèmes aussi complexes ? Les sciences cognitives commencent seulement à soumettre cette question au crible de l'expérience. Le test du casino, conçu par Antoine Bechara à l'Université de l'Iowa, évalue l'intuition proto-mathématique des probabilités et de l'espérance statistique[76]. On accorde au sujet un prêt de 2 000 dollars (en fausse monnaie – les psychologues ne sont pas si riches !), et on lui montre quatre jeux de cartes, faces cachées. À chaque tour, le sujet choisit l'un de ces jeux, retourne une carte, et y découvre un gain ou une perte (par exemple « vous gagnez 200 dollars »). Pour maximiser ses gains, le joueur doit apprendre à choisir à bon escient. Or, à son insu, deux des jeux de cartes sont désavantageux : au départ, ils offrent des gains rapides, mais très vite, ils entraînent des pertes telles qu'au bout du compte le total est franchement négatif. Les deux autres jeux conduisent à des gains et des pertes plus modestes, qui conduisent lentement mais sûrement à un solde positif.

Au début, les joueurs ne peuvent que tirer les cartes au hasard, mais progressivement, ils acquièrent une sorte d'intuition consciente et,

finalement, ils parviennent à dire que deux des jeux sont néfastes. La période la plus intéressante est celle qui précède toute intuition consciente. Cette phase ressemble à la période d'incubation des mathématiciens : les participants ont déjà une certaine expérience des différents jeux de cartes, mais ils choisissent toujours au hasard et disent n'avoir pas la moindre idée de ce qu'il convient de faire. Et pourtant… Juste avant de tirer une carte dans l'une des mauvaises piles, leurs mains se mettent à suer, ce qui se traduit par une baisse mesurable de la résistance cutanée. Marqueur physiologique classique de l'activation du système nerveux sympathique, cette « réaction viscérale » implique que leur cerveau a parfaitement enregistré quels sont les choix les plus risqués.

Le signal d'alarme inconscient provient vraisemblablement du cortex préfrontal ventro-médian, une région spécialisée dans l'évaluation de nos actes. L'imagerie cérébrale montre que cette région s'active précisément lors des essais désavantageux, avec une amplitude qui prédit la réussite ultérieure[77]. Effectivement, les patients atteints de lésions de cette région ne présentent plus la moindre réponse cutanée anticipée, ils n'éprouvent d'émotion qu'*après* avoir appris leurs pertes.

Le cortex frontal ventro-médian et orbito-frontal contient toute une panoplie de processus d'évaluation qui jaugent constamment nos décisions et en anticipent la valeur future. Les recherches de Bechara suggèrent que ces calculs s'opèrent à notre insu. Chaque fois que nous avons l'impression de choisir au hasard, des pondérations inconscientes président en réalité à nos actes.

Bien qu'elle nécessite un calcul des probabilités et des espérances, l'intuition des gains et des pertes nous éloigne de la résolution de problèmes mathématiques. Cependant, une autre expérience, que l'on doit à Ap Dijksterhuis, se rapproche de la taxonomie d'Hadamard et suggère que la résolution de problèmes bénéficie elle aussi d'une période d'incubation inconsciente[78]. Ce psychologue hollandais a proposé à des étudiants de choisir parmi quatre marques de voitures, chacune étant

caractérisée par une douzaine de critères distinctifs. Au cours du test, les participants commencent par lire l'énoncé, puis on laisse à la moitié d'entre eux le temps d'y réfléchir consciemment pendant quatre minutes, tandis que l'autre moitié est distraite pendant la même durée (ils doivent résoudre des anagrammes). Enfin, tous doivent donner leur réponse. Surprise : le groupe qui a été distrait choisit la voiture optimale bien plus souvent que celui qui y a réfléchi posément (60 % de réussite contre 22 %, une différence étonnamment grande si l'on considère qu'un choix au hasard conduirait à 25 % de bonnes réponses). Ce résultat a été répliqué dans les conditions de la vie quotidienne : plusieurs semaines après un achat chez Ikea, les clients qui disent avoir beaucoup réfléchi à leur décision sont *moins* satisfaits de leur acquisition que ceux qui ont choisi sous l'impulsion du moment, sans grande réflexion consciente.

L'expérience n'est pas parfaite, car elle ne remplit pas les critères exigeants de non-conscience (la simple distraction ne garantit pas que les sujets n'ont eu aucune possibilité de réfléchir au problème). Néanmoins, elle suggère qu'il vaut mieux laisser certains problèmes aux franges de la conscience plutôt que de les attaquer de front. La sagesse populaire n'a pas complètement tort lorsqu'elle affirme que laisser reposer un problème conduit souvent à de meilleurs résultats que de suer sang et eau pour le résoudre immédiatement.

Notre inconscient peut-il résoudre n'importe quel problème ? Ou bien, plus vraisemblablement, certaines catégories d'énigmes sont-elles plus aisées à incuber inconsciemment que d'autres ? Il est intéressant de noter que les expériences de Bechara et de Dijksterhuis présentent des similitudes : toutes deux demandent de pondérer un ensemble de paramètres. Dans le cas de Bechara, il s'agit de pondérer les gains et les pertes que chaque paquet de cartes a entraînés par le passé. Chez Dijksterhuis, la voiture idéale émerge d'une pondération optimale d'une douzaine de critères. De telles quantités de données excèdent la capacité de notre mémoire de travail : elles dépassent les compétences de l'esprit conscient, dont la force est d'examiner les possibilités une par une. Voilà

sans doute pourquoi les personnes à qui on donne le temps de réfléchir, dans l'expérience de Dijksterhuis, ne s'en tirent pas bien : elles ont tendance à accorder trop d'importance à l'un ou l'autre des critères sans discerner la perspective d'ensemble. Au contraire, les processus inconscients excellent dans le traitement simultané de très nombreuses informations : ils peuvent assigner une valeur à plusieurs items en même temps, et les moyenner pour parvenir à une décision.

Plusieurs expériences démontrent que le calcul d'une somme ou d'une moyenne fait partie des opérations simples qu'un circuit neuronal élémentaire accomplit sans conscience. Même un singe macaque, une fois entraîné, parvient à prendre une décision fondée sur le montant total d'une série de quatre formes arbitraires dont il a appris la valeur individuelle, et la décharge des neurones du cortex pariétal reflète, à chaque instant, la valeur de leur somme[79]. Dans mon laboratoire, nous avons prouvé que l'addition approximative est à la portée d'un cerveau non conscient. Une première expérience affichait une série de cinq flèches, en demandant au participant de juger si la majorité était tournée vers la droite ou vers la gauche. Même quand les flèches étaient rendues invisibles par masquage, les participants, tout en croyant répondre au hasard, s'en sortaient plutôt bien, et les signaux de leur cortex pariétal indiquaient un calcul inconscient du cumul approximatif des cinq informations reçues[80]. Toutes les flèches, même celles qui étaient subjectivement invisibles, parvenaient aux centres d'évaluation, de pondération et de décision du cerveau.

Une seconde expérience consistait à afficher pas moins de huit chiffres. Quatre d'entre eux étaient visibles, et les quatre autres entièrement masqués. Nous demandions aux participants de décider si leur moyenne était plus grande ou plus petite que 5. Ils croyaient le faire uniquement sur les quatre chiffres visibles, mais en réalité, leurs réponses prenaient en compte la totalité des huit chiffres. Ainsi, lorsque la moyenne des chiffres visibles était plus grande que 5, alors que celle des chiffres cachés était plus petite que 5, les sujets montraient un biais inconscient

pour la réponse « plus petit[81] ». Le calcul de la moyenne, qu'ils essayaient de mener consciemment, s'étendait aux quatre chiffres invisibles.

Statistiques nocturnes

Il est donc clair que certaines opérations mathématiques élémentaires, comme la moyenne ou la comparaison, s'effectuent sans conscience. Mais qu'en est-il des opérations vraiment créatives, comme celles qui conduisent à l'illumination d'un Poincaré dans son omnibus ? La lumière d'une découverte nouvelle peut-elle vraiment nous frapper à tout moment, alors que nous pensons à autre chose ? Il semble bien que oui. À notre insu, notre cerveau agit constamment comme un statisticien à l'affût de la moindre régularité cachée derrière des données qui nous semblent aléatoires. Cet apprentissage statistique opère sans relâche, y compris lorsque nous dormons.

Ullrich Wagner, Jan Born et leurs collègues ont testé avec rigueur un phénomène que la plupart des scientifiques disent avoir vécu : au réveil, après une bonne nuit de sommeil, la solution d'un problème complexe émerge soudainement[82]. Afin de transposer cette anecdote au laboratoire, ils ont demandé à des volontaires de participer à un challenge d'arithmétique : il s'agissait de transformer, le plus vite possible, une séquence de sept chiffres en une autre en suivant des règles simples. On leur demandait juste de nommer le dernier des sept chiffres du résultat – mais le trouver demandait un long calcul mental. Cependant, le problème comprenait un raccourci caché. La séquence des chiffres du résultat présentait une symétrie : les trois derniers chiffres reprenaient les trois précédents dans l'ordre inverse (par exemple 4 1 4 9 9 4 1). Ainsi, le chiffre recherché, le dernier, était toujours identique au deuxième. Une fois cette régularité remarquée, les participants pouvaient s'épargner bien des efforts en se cantonnant au calcul du deuxième chiffre. Au cours de la

première heure de test, la plupart des sujets ne s'apercevaient pas de cette astuce. Étonnamment, une bonne nuit de sommeil faisait plus que doubler la probabilité de la découvrir le lendemain : nombreux étaient ceux qui se réveillaient avec la solution ! Des expériences de contrôle ont établi que le temps écoulé ne comptait pas – seul importait le sommeil. Lorsque nous dormons, notre cerveau consolide ses connaissances sous des formes plus compactes.

De nombreuses études chez l'animal montrent que les neurones de l'hippocampe et du cortex s'activent sans relâche au cours du sommeil. Leurs décharges neuronales « rejouent », à grande vitesse, les mêmes séquences d'activité que celles évoquées durant la journée précédente[83]. Un rat court dans un labyrinthe, puis s'endort : les cellules de l'hippocampe qui codent pour les lieux de l'espace se réactivent immédiatement, avec une telle précision que l'on parvient à décoder quels endroits l'animal est en train d'explorer mentalement. Souvent les décharges oniriques se déroulent plus vite que la réalité, parfois même en ordre inverse. Cette compression temporelle pourrait permettre au cerveau de traiter des informations dispersées dans le temps comme un seul épisode : accélérée, une séquence temporelle se transforme en une carte spatiale de neurones activés ou inhibés, qui permet la détection de régularités cachées, inaccessibles aux mécanismes normaux de l'apprentissage diurne. Quel qu'en soit le mécanisme ultime, il est clair que le sommeil est une période d'intense activité inconsciente qui consolide la mémoire et, parfois, dans l'obscurité de la nuit, jette soudain la lumière sur des problèmes demeurés insolubles.

La boîte à outils de l'inconscient

De telles démonstrations de laboratoire sont encore bien loin des envolées mathématiques d'un Poincaré au sein des fonctions fuchsiennes

et de la géométrie non euclidienne. Petit à petit, pourtant, la distance se réduit, à mesure que des expériences innovantes démontrent qu'un nombre croissant d'opérations cognitives élaborées s'exécutent, au moins en partie, sans la moindre conscience.

Bien des psychologues ont longtemps considéré que l'« administrateur central » de notre cerveau (*central executive*) – ce système de haut niveau qui contrôle nos actes, inhibe nos réponses instinctives, passe d'une tâche à l'autre, et repère nos erreurs – était l'apanage de l'esprit conscient. Il n'en est rien : une part au moins de ces opérations se déroulent inconsciemment et peuvent être guidées par des stimuli invisibles.

Considérons la faculté de nous contrôler et d'inhiber nos tendances automatiques. Imaginez que vous deviez effectuer une tâche répétitive telle qu'appuyer sur un bouton dès qu'une image apparaît sur l'écran – sauf que, de temps en temps, lorsque l'image représente un rond noir, vous devez absolument éviter de répondre (ce stimulus est appelé le signal « stop »). L'inhibition d'une activité routinière est typique des fonctions de l'« administrateur central ». Le psychologue hollandais Simon Van Gaal s'est demandé si elle exigeait vraiment un traitement conscient. Les sujets parviendraient-ils à refréner leurs réponses si le signal « stop » demeurait subliminal ? Étonnamment, la réponse est oui. Quand le signal « stop » est présenté si brièvement qu'il en demeure invisible, la main des participants se ralentit, et il arrive même qu'ils cessent de répondre – sans savoir pourquoi, puisque le stimulus qui déclenche cette inhibition reste inconscient[84]. Ces résultats montrent qu'*invisible* n'est pas synonyme d'*incontrôlable*. Même un panneau de « stop » inconscient peut évoquer une avalanche d'activité neuronale qui se propage jusqu'au plus profond des réseaux de l'administrateur central, ceux qui contrôlent nos actes[85].

De même, notre cerveau parvient à détecter certaines de nos erreurs sans que nous en ayons conscience. Lorsque nous bougeons les yeux, notre regard dévie parfois de la cible que nous avons volontairement choisie (par exemple parce qu'un autre objet attire notre attention). L'expérience montre que ce type d'erreur peut être détecté

inconsciemment par les réseaux spécialisés du cortex cingulaire antérieur – sans que les participants aient la moindre conscience d'avoir fait une erreur, et alors même qu'ils nient farouchement que leurs yeux se soient détachés de la cible[86].

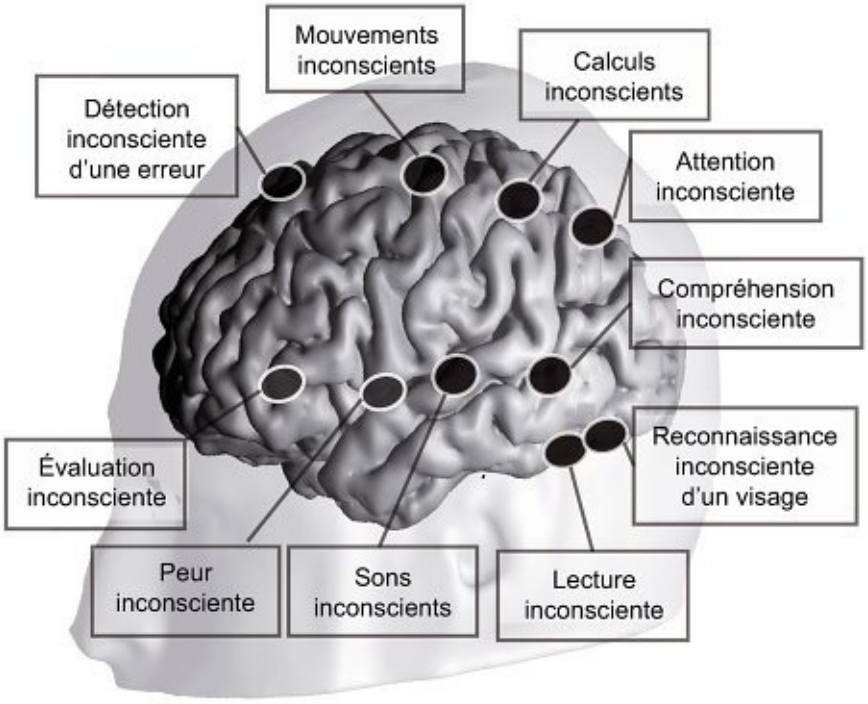


Figure 13. Un aperçu des calculs inconscients du cerveau humain. La figure ne montre qu'une petite fraction des circuits dont les sciences cognitives ont prouvé l'activation inconsciente. Nous pensons aujourd'hui que pratiquement n'importe quelle aire cérébrale peut opérer dans un mode non conscient. Par simplicité, chaque fonction a été attribuée à une région principale, mais il faut se souvenir que toutes les grandes fonctions cognitives font appel à des circuits distribués. Certains processus inconscients reposent sur des régions sous-corticales comme l'amygdale ou le striatum (en pointillé), d'autres sur des circuits corticaux. Même des fonctions culturelles de haut niveau, comme la lecture ou l'arithmétique élémentaire, peuvent se dérouler sans conscience.

Un signal inconscient peut même nous inciter à changer de tâche, et donc à reconfigurer tous les réseaux cérébraux. Demandons par exemple à un sujet d'effectuer une tâche donnée (disons, juger si deux mots riment)

et, lorsqu'il voit un indice visuel, de changer de tâche (disons, juger si deux mots sont synonymes). Lorsque l'indice est présenté en deçà du seuil de conscience, les participants ralentissent, et l'imagerie cérébrale montre qu'ils commencent à activer les réseaux de la seconde tâche[87].

En bref, la psychologie cognitive a amplement démontré non seulement que la perception subliminale est une réalité, mais également que toute une panoplie de processus mentaux peut se déclencher sans conscience (il est important de souligner qu'en général ils ne vont pas jusqu'à leur terme). La figure 13 résume les diverses régions cérébrales qui, dans l'une ou l'autre des expériences que nous avons décrites dans ce chapitre, s'activent inconsciemment. La boîte à outils de l'inconscient comprend une extraordinaire diversité d'opérations mentales, depuis la compréhension d'un mot jusqu'à l'addition de plusieurs chiffres, et depuis la détection d'erreurs jusqu'à la résolution de problèmes. Parce qu'elles travaillent rapidement et en parallèle, ces opérations surpassent souvent la réflexion consciente.

Henri Poincaré, dans *Science et méthode* (1908), avait anticipé la découverte de la supériorité du jugement inconscient, massivement parallèle, sur la pensée consciente lente et sérielle :

> « Le moi subliminal n'est nullement inférieur au moi conscient ; il n'est pas purement automatique, il est capable de discernement, il a du tact, de la délicatesse ; il sait choisir, il sait deviner. Que dis-je ? Il sait mieux deviner que le moi conscient, puisqu'il réussit là où celui-ci avait échoué. En un mot, le moi subliminal n'est-il pas supérieur au moi conscient ? »

La science contemporaine répond positivement à l'interrogation de Poincaré : oui, de bien des manières, les opérations subliminales surpassent la réflexion consciente. À tout instant, sans que nous n'en prenions conscience, notre système visuel résout des problèmes de

reconnaissance des formes qui dépassent tous nos logiciels actuels. Nous nous appuyons sur l'extraordinaire puissance de calcul de l'inconscient chaque fois que nous tentons de résoudre des problèmes mathématiques.

Toutefois, ne poussons pas cette conclusion plus loin que nécessaire. Certains psychologues vont jusqu'à affirmer que la conscience n'est qu'un mythe, un trait purement décoratif et sans aucun rôle effectif, comme la cerise sur le gâteau[88]. *Toutes* les opérations mentales qui président à nos décisions, prétendent-ils, s'accomplissent sans conscience. Selon eux, l'introspection n'est qu'un spectateur mégalomaniaque qui croit diriger ce qu'elle ne fait que subir. En cela, la conscience serait semblable à la célèbre mouche du coche de La Fontaine, qui « pense à tout moment qu'elle fait aller la machine » alors que le mérite en revient aux chevaux de trait. Nous vivrions comme les prisonniers du film *Matrix* : l'impression que nous avons de mener une vie consciente ne serait qu'une illusion, puisque « nos » décisions se prennent à notre insu, inconsciemment, par des moyens sur lesquels nous n'avons aucun contrôle.

Au prochain chapitre, je tenterai de réfuter cette théorie qui fait de la conscience une illusion et de nous des zombies impuissants. La conscience est une fonction biologique qui a émergé au cours de l'évolution parce qu'elle remplissait un rôle utile à la survie. Elle doit donc remplir une niche cognitive spécifique, et répondre à un problème que les systèmes spécialisés et parallèles de l'esprit inconscient ne pouvaient pas résoudre.

Toujours aussi pénétrant, Poincaré avait déjà remarqué qu'en dépit de leur puissance de calcul les rouages de l'inconscient ne se mettaient en route qu'à la suite d'un examen prolongé et éminemment conscient du problème initial. Il notait également qu'après l'illumination seul l'esprit conscient pouvait vérifier, pas à pas, les intuitions que l'inconscient prétendait justes. Le sculpteur Henry Moore parvint à la même conclusion :

« Bien que la part subconsciente, instinctive, non logique de l'esprit joue certainement un rôle dans le travail [de l'artiste], celui-ci dispose également d'un esprit conscient qui ne reste pas inactif. L'artiste travaille avec une concentration de toute sa personnalité, et la part consciente résout les conflits, organise les souvenirs et l'empêche d'essayer d'emprunter deux directions à la fois. »

Découvrons à présent ce domaine privé de l'esprit conscient.

CHAPITRE 3

À quoi sert la conscience ?

Pourquoi la conscience est-elle apparue au cours de l'évolution des espèces ? Existe-t-il des opérations qui ne peuvent être réalisées que par un cerveau conscient ? Ou bien la conscience n'est-elle qu'un épiphénomène, un aspect inutile de notre biologie ? Plusieurs arguments suggèrent que la conscience remplit certaines fonctions indispensables aux organismes évolués. Les représentations subliminales s'évanouissent rapidement, tandis que l'information consciente reste stable : nous pouvons y réfléchir aussi longtemps que nous le souhaitons. La conscience comprime l'information, elle la transforme en un petit jeu de symboles soigneusement sélectionnés. Ceux-ci peuvent alors être retransmis à pratiquement n'importe quel autre processus cérébral, ce qui nous permet d'enchaîner les opérations mentales, sous le contrôle de la conscience, un peu comme un ordinateur sériel. Cette fonction de « retransmission » de la conscience est essentielle. Dans l'espèce humaine, elle est considérablement augmentée par le langage, qui nous permet de diffuser nos pensées conscientes à tout un réseau social.

« Les caractéristiques de la répartition de conscience, pour autant que nous les connaissions, suggèrent qu'elle est efficace. »

William JAMES,
Principes de psychologie (1890).

Dans l'histoire de la biologie, peu de questions ont été autant disputées que celle du finalisme ou de la « téléologie » : est-il légitime de dire qu'un organe a été conçu ou a évolué « pour » une certaine fonction (une « cause finale » ou *telos* en Grec) ? Avant Darwin, le finalisme était répandu : chaque espèce, chaque organe portait l'empreinte de la main de Dieu, le grand horloger de l'Univers. Le grand anatomiste Georges Cuvier, dans son analyse des organes du corps, ne cessait de faire appel à la téléologie : les griffes étaient faites « pour » attraper des proies, les poumons « pour » respirer, les yeux « pour » voir, et ces causes finales expliquaient l'existence même de ces fonctions et de leur intégration au sein de l'organisme – chaque partie du corps avait un but.

Charles Darwin révolutionna cette idée en montrant que la sélection naturelle, plutôt que l'intelligence d'un grand horloger, façonnait la

biosphère. La thèse darwinienne n'a nul besoin d'une intention divine. Les organes évolués ne sont en rien conçus *pour* leur fonction : en définitive, ils ne font que conférer à l'organisme un avantage pour la reproduction.

Dans un étonnant renversement de la perspective, ce sont les antiévolutionnistes qui, pour contrer Darwin, se mirent à collectionner les exemples d'organes qui leur semblaient, de toute évidence, mal adaptés et désavantageux pour l'organisme. Comment expliquer que le paon arbore une queue gigantesque, certes superbe, mais qui gêne ses mouvements ? Pourquoi le mégacéros, un genre d'élan irlandais aujourd'hui disparu, portait-il des bois de près de quatre mètres d'envergure, une taille tellement disproportionnée que certains scientifiques l'ont rendue responsable de l'extinction de l'espèce ? Pour répondre à cette critique, Darwin formula l'hypothèse de la sélection sexuelle : il est dans l'intérêt des mâles, qui se disputent l'attention des femelles, de brandir des étendards élaborés, coûteux et symétriques, car ceux-ci, telles de coûteuses publicités sur papier glacé, reflètent à coup sûr l'aptitude physique et génétique de leurs possesseurs (*fitness*). Les bois du mégacéros seraient donc une réclame plus qu'une arme. La leçon était limpide : aucun organe biologique ne naît avec une étiquette qui lui attribue sa fonction, et même des assemblages bancals et bricolés par l'évolution peuvent se révéler avantageux dans la grande bataille pour la reproduction.

Au cours du XX^e^ siècle, la théorie synthétique de l'évolution a sonné la mort de la téléologie. La biologie contemporaine de l'évolution et du développement (*evo-devo*) inclut tout un arsenal de concepts qui expliquent pourquoi la biologie donne parfois l'impression d'un *intelligent design*, une conception sophistiquée… mais sans concepteur. Dans cette boîte à outils, on trouve notamment les idées suivantes :

- l'apparition spontanée de motifs : le mathématicien Alan Turing a, le premier, décrit comment certaines réactions chimiques font émerger des formes géométriques quasi régulières telles que les rayures du zèbre ou les côtes du melon[1]. Sur certains coquillages, les pigments

s'auto-organisent en des formes complexes. Parce qu'ils se forment en dessous d'une couche de matière organique opaque, il est presque certain que ces motifs ne jouent aucun rôle dans la survie de l'animal – ils ne sont que la conséquence de réactions chimiques qui ont leur propre raison d'être ;

- les relations allométriques : un accroissement de la taille globale de l'organisme, même s'il est avantageux pour la survie, exige parfois un changement proportionnel de la taille de certains de ses organes qui, lui, peut s'avérer délétère. Les bois démesurés de l'élan irlandais résultent probablement d'une modification allométrique[2] ;
- les écoinçons de Steven Jay Gould : le mot « écoinçon » (*spandrel*) est un terme technique d'architecture : c'est l'espace en forme de triangle arrondi qui se forme inévitablement à l'endroit où une arche rencontre un mur ou un pilier. Ce mot avait été choisi par le paléontologue de Harvard Steven Jay Gould, aujourd'hui disparu, pour décrire un aspect de l'organisme qui n'existe qu'en tant que sous-produit inévitable du développement de l'organisme[3]. L'exemple type est le téton sur la poitrine des hommes – conséquence inutile mais inévitable de la manière dont le génome des mammifères construit des tétines chez les femelles.

Munis de tous ces concepts, nous ne pouvons plus supposer que chacune de nos propriétés biologiques ou psychologiques joue nécessairement un rôle positif dans la survie de l'organisme. En principe, la conscience pourrait n'être qu'un trait contingent et décoratif de notre espèce, ou bien un effet indésirable de l'énorme accroissement de la taille du cerveau au sein du genre *Homo*, voire un simple écoinçon, la conséquence mécanique d'autres évolutions vitales. Ces hypothèses rejoignent le pessimisme teinté d'humour noir d'Alexandre Vialatte, pour qui « la conscience, comme l'appendice, ne sert à rien, sauf à rendre l'homme malade ». L'un des héros du film *Dans la peau de John Malkovitch*, le marionnettiste Craig Schwartz se lamente également sur l'inutilité de l'introspection : « La conscience est une terrible malédiction. Je pense. Je

ressens. Je souffre. Et tout ce que je demande, c'est de pouvoir faire mon travail. »

La conscience n'est-elle qu'un épiphénomène ? Peut-on la comparer au bruit assourdissant d'un turboréacteur – une conséquence désagréable, inutile et cependant inévitable de la machinerie cérébrale ? Le psychologue britannique Max Velmans penche clairement en faveur de cette hypothèse pessimiste. Selon lui, la plupart de nos fonctions cognitives, sinon toutes, sont indifférentes à la conscience : elles continueraient à fonctionner aussi bien si nous n'étions que de vulgaires zombies[4]. L'auteur danois Tor Nørretranders parle de l'« illusion de l'utilisateur » pour décrire l'impression que la conscience est aux commandes de nos actes et qui, selon lui, n'est qu'une vaste mystification : chacune de nos décisions, en réalité, serait issue des profondeurs de l'inconscient[5]. Bien d'autres psychologues parviennent à la même conclusion : la conscience n'est que la mouche du coche, un observateur inutile d'événements qui la dépassent[6].

Dans ce livre, cependant, j'explore une voie différente, que certains philosophes appellent une vision fonctionnaliste de la conscience. Je pense que la conscience est utile et qu'elle répond à des besoins fonctionnels spécifiques de l'organisme. La perception consciente transforme les informations sensorielles en un code interne qui, seul, permet certaines opérations distinctives. La conscience possède des caractéristiques fonctionnelles qui résultent d'une architecture élaborée. À ce titre, elle résulte vraisemblablement d'une sélection naturelle à l'échelle de plusieurs centaines de millions d'années et remplit donc probablement un rôle bien précis.

Comment déterminer ce qu'est ce rôle ? Nous ne pouvons pas rembobiner l'histoire de la vie, mais nous pouvons étudier le contraste minimal entre des images visibles et invisibles afin de préciser l'unicité des opérations conscientes. La psychologie cognitive nous permet d'évaluer quelles opérations s'effectuent sans conscience, et lesquelles exigent que l'information soit perçue consciemment. Dans ce chapitre, nous verrons

que, loin d'écarter la conscience de tout rôle fonctionnel, ces recherches indiquent qu'elle remplit certaines fonctions avec efficacité.

Statistiques inconscientes, échantillonnage conscient

Ma vision de la conscience suggère une répartition naturelle du travail. Dans les profondeurs de l'inconscient, une armée de travailleurs souterrains effectue un énorme travail de fond : tous traitent des monceaux de données. Pendant ce temps, au sommet, un petit groupe distingué d'administrateurs de haut niveau, sur la base d'un condensé de la situation, réfléchit aux décisions conscientes avec toute la pondération nécessaire.

Le chapitre 2 a mis en évidence les superpouvoirs des calculs inconscients de notre cerveau. À notre insu, tout un assortiment d'opérations cognitives, depuis la perception jusqu'à la compréhension du langage, la décision, l'action, l'évaluation et l'inhibition se déroulent sans conscience. Hors de portée de toute introspection, des myriades de processeurs inconscients œuvrent simultanément afin d'extraire l'interprétation la plus complète possible de notre environnement. Ils agissent comme de brillants statisticiens qui exploitent le moindre indice sensoriel – un minuscule mouvement, une ombre, une tache de lumière – afin de calculer la probabilité que telle propriété soit vraie dans le monde extérieur. Un peu comme Météo France combine des dizaines d'observations afin d'en déduire les risques de pluie, la perception inconsciente s'appuie sur la variété des données sensorielles afin d'en déduire la probabilité que telle couleur, telle forme, tel animal ou telle personne soit présent dans notre environnement. Notre conscience, par contre, ne nous donne qu'un aperçu très réduit de cet univers probabiliste – ce que les statisticiens appellent un « échantillon » de la distribution des

possibles. La conscience n'hésite pas à simplifier les choses : elle résume le monde à sa plus simple expression, un aperçu suffisamment condensé pour être utilisable par nos systèmes de prise de décision.

Cette division du travail, entre une armée de statisticiens indépendants et un système de décision unique, s'impose probablement à tout animal qui doit décider de la manière de se mouvoir. Personne ne peut agir sur la base d'une simple probabilité – à un moment, un processus dictatorial doit couper court à toutes les ambiguïtés et imposer sa volonté unique. *Alea jacta est* : « les dés sont jetés », comme l'aurait dit César en franchissant le Rubicon pour reprendre Rome des mains de Pompée. Toute action volontaire fait pencher la balance jusqu'à un point de non-retour : la distribution de probabilités s'effondre pour laisser place à un échantillon unique, qui permet de réfléchir à de nouvelles décisions.

La fable classique de l'âne de Buridan illustre l'importance de trancher rapidement lors d'une prise de décision complexe. Un âne assoiffé et affamé est placé entre un seau d'eau et une botte de foin. Dans l'incapacité de décider par quoi commencer, l'animal finit par mourir de faim et de soif ! Le problème paraît ridicule, pourtant nous sommes constamment confrontés à des décisions du même type : le monde ne nous offre que des opportunités brutes aux interprétations hasardeuses. La conscience résout ces dilemmes en portant à notre attention, à chaque instant, une seule parmi les milliers d'interprétations possibles du monde.

Le philosophe Charles Sanders Peirce, suivant en cela les pas pionniers du physicien Hermann von Helmholtz, a été l'un des premiers à reconnaître tout ce que notre perception consciente doit à la complexité effroyable des inférences probabilistes inconscientes :

> « Si je regarde par la fenêtre en cette charmante matinée de printemps, je vois une azalée en pleine floraison. Non, non ! Ce n'est pas cela que je vois, même si c'est la seule manière de le décrire. Ceci n'est qu'une proposition, une phrase, un fait ; mais ce

> que je perçois n'est pas une proposition, ni une phrase ou un fait, mais seulement une image que je rends en partie intelligible par le biais de cette formulation. Celle-ci est abstraite, mais ce que je vois est concret. Je fais une abduction dès que j'exprime ce que je vois sous la forme d'une phrase. En vérité, l'étoffe de notre savoir n'est qu'un entrelacs d'hypothèses pures confirmées et raffinées par l'induction. On ne pourrait pas faire la moindre avancée dans la connaissance (…) sans pratiquer l'abduction à chaque étape[7]. »

Ce que Peirce appelle l'« abduction » n'est rien d'autre que ce que les sciences cognitives contemporaines nomment l'« inférence bayésienne ». Celle-ci doit son nom au révérend Thomas Bayes (environ 1701-1761) qui a, le premier, exploré ce domaine des mathématiques. L'inférence bayésienne consiste à mener un raisonnement probabiliste en sens inverse afin de remonter aux causes cachées d'une série d'observations aléatoires. En théorie classique des probabilités, on vous décrit la situation (par exemple « quelqu'un tire trois cartes d'un jeu de 52 cartes »), et la théorie permet d'assigner une probabilité à chaque événement (par exemple, « quelle est la probabilité que les trois cartes soient des as ? »). La théorie bayésienne, elle, raisonne en sens inverse, depuis le résultat jusqu'à ses origines probables. Elle permet, par exemple, de répondre à des questions du genre : « Si quelqu'un tire trois as d'un jeu de 52 cartes, quelle est la probabilité que le jeu soit truqué et comprenne plus de quatre as ? » C'est ce qu'on appelle une inférence inverse ou un raisonnement bayésien. L'hypothèse que le cerveau effectue des inférences bayésiennes est l'un des domaines les plus actifs et les plus disputés des neurosciences contemporaines.

Notre cerveau doit effectuer des inférences inverses pour une raison très simple : toutes nos sensations sont ambiguës. Si je vous montre une assiette, vous percevez son contour comme un cercle parfait, et pourtant elle apparaît sur votre rétine sous la forme d'une ellipse qui serait

compatible avec des myriades d'autres interprétations. Une infinité de patatoïdes d'innombrables orientations peuvent se projeter sous la forme d'une assiette. Si vous percevez un cercle, c'est parce que votre cerveau, après avoir parcouru inconsciemment l'ensemble des causes possibles de cette sensation rétinienne, *décide* que le cercle est l'interprétation la plus probable. Ainsi, bien que la perception semble instantanée, elle résulte d'une inférence complexe qui écarte un nombre inimaginable d'autres explications possibles de nos entrées sensorielles.

Les neurosciences offrent de nombreux exemples de cette élimination inconsciente des hypothèses. Chaque neurone du cortex visuel, par exemple, ne perçoit qu'un petit segment du contour de l'ellipse. Cet indice est compatible avec une grande variété de formes et de mouvements. Dès que plusieurs neurones communiquent entre eux, par contre, chacun « vote » pour l'interprétation qui lui semble la plus vraisemblable, et la population entière converge vers le meilleur candidat. Le cerveau raisonne comme Sherlock Holmes : quand vous avez éliminé l'impossible, ce qui reste, pour improbable qu'il soit, doit être la vérité.

Une stricte logique régit nos circuits inconscients : ils semblent idéalement organisés pour tirer des inférences statistiques précises à partir de nos entrées sensorielles. Dans l'aire du cortex temporal moyen appelée « MT/V5 » et sensible au mouvement visuel, par exemple, chaque neurone ne perçoit le mouvement qu'à travers une étroite lucarne, le « champ récepteur ». À cette échelle, n'importe quel mouvement est ambigu. Si vous regardez un bâton par le trou d'une serrure, vous ne parvenez pas à déterminer comment il bouge, car une infinité de mouvements réels sont compatibles avec le mouvement observé (figure 14). Chaque neurone de l'aire MT/V5 est soumis à cette ambiguïté fondamentale – et pourtant nous n'en avons pas conscience. Même dans les pires circonstances, nous ne percevons qu'un mouvement particulier, jamais un ensemble de probabilités. Notre cerveau prend une décision et nous donne à voir l'interprétation qu'il juge la plus probable. En l'occurrence, c'est celle qui minimise le déplacement : nous voyons le bâton se mouvoir dans une

direction perpendiculaire à lui-même. Une armée de processeurs inconscients évalue toutes les possibilités, mais notre conscience n'en reçoit que la synthèse.

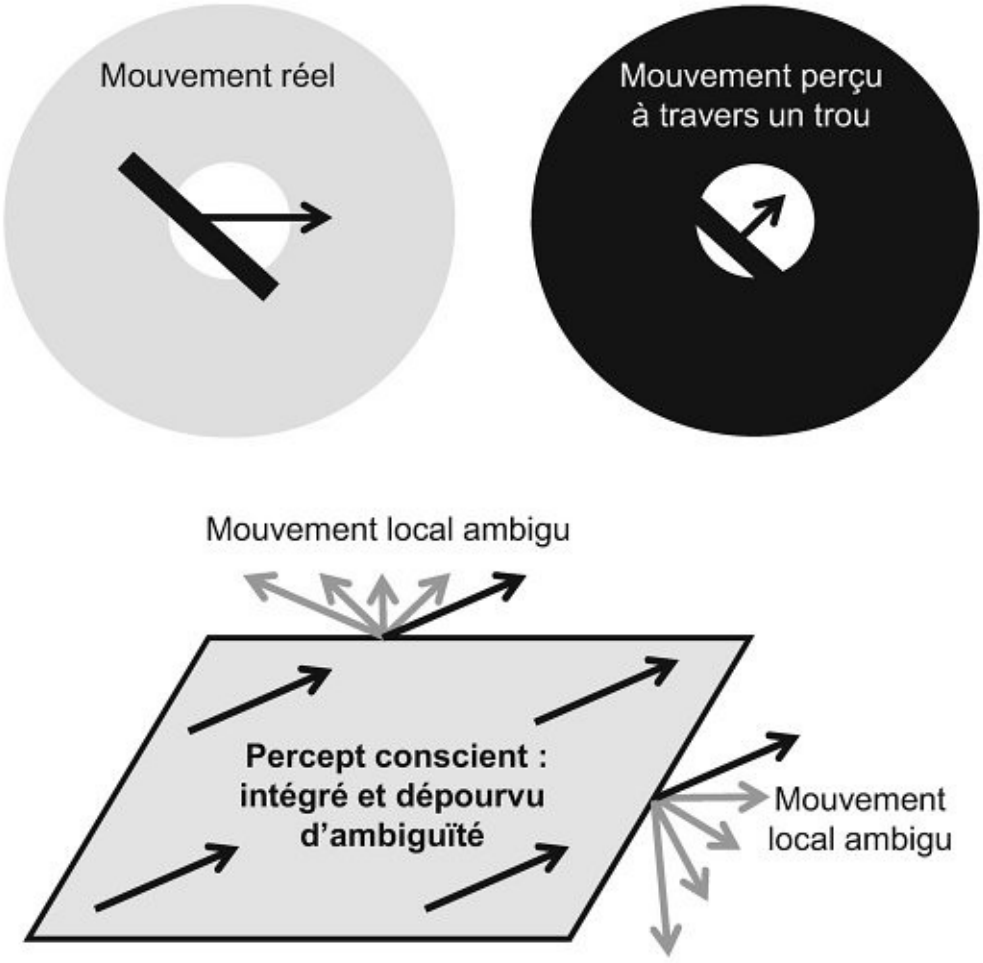


Figure 14. La conscience aide à résoudre certaines ambiguïtés. Dans l'aire MT/V5, la région du cortex sensible au mouvement, les neurones ne voient le monde qu'au travers d'une petite lucarne appelée « champ récepteur ». Cette fenêtre de vision limitée les empêche de déterminer le mouvement d'une barre : localement, chaque mouvement est ambigu. Consciemment, pourtant, notre système visuel ne nous donne à voir qu'un mouvement dépourvu d'ambiguïté. Lorsqu'un objet se déplace, le mouvement que nous percevons résulte de l'intersection des contraintes issues de chacun des côtés. Les neurones de l'aire MT/V5 codent d'abord le mouvement local, mais très vite, en échangeant des messages, ils convergent vers une interprétation globale qui coïncide avec celle que perçoit l'observateur. Cette convergence s'interrompt sous anesthésie générale : la résolution des ambiguïtés requiert la conscience.

Lorsque nous regardons une forme complexe, par exemple un rectangle en mouvement, les ambiguïtés locales persistent, mais elles peuvent à présent être résolues en combinant les informations issues des différents côtés. Une seule direction de mouvement satisfait simultanément toutes les contraintes (voir la figure 14). Notre cerveau l'infère et nous donne à voir le seul mouvement rigide qui convient. L'enregistrement des décharges neuronales montre que cela prend du temps : pendant un dixième de seconde, les neurones de l'aire MT/V5 ne « voient » que le mouvement local, et il leur faut 120 à 140 millisecondes pour changer d'avis et se mettre à coder le mouvement global[8]. La conscience, toutefois, ignore tout de ce calcul. Subjectivement, nous n'en percevons que l'aboutissement (un rectangle en mouvement cohérent), sans nous douter que nos sensations initiales étaient ambiguës et que notre cerveau a eu fort à faire pour les expliquer.

Il est fascinant de constater que, sous anesthésie générale, le processus qui permet aux neurones de l'aire MT/V5 de converger vers une interprétation cohérente s'interrompt[9]. La perte de la conscience s'accompagne d'un dysfonctionnement soudain des circuits qui rassemblent les informations sensorielles en un tout cohérent. La conscience semble nécessaire à l'échange bidirectionnel de signaux au sein de la hiérarchie du cortex, de bas en haut et de haut en bas, jusqu'à ce que tout le monde soit d'accord. En son absence, l'inférence perceptive n'aboutit plus.

La conscience joue donc un rôle proéminent dans la résolution des ambiguïtés perceptives. Pour le montrer, rien de tel que de concevoir de toutes pièces un stimulus parfaitement ambigu. Sur un écran d'ordinateur, superposons deux grilles qui bougent dans des directions différentes (figure 15). Notre cerveau n'a aucun moyen de savoir laquelle des deux grilles masque l'autre. Subjectivement, pourtant, nous ne percevons pas l'ambiguïté. Nous ne voyons jamais un mélange des deux possibilités : notre perception consciente décide et nous laisse voir l'une des deux grilles devant l'autre. Mais attendez quelques secondes, et vous verrez que

les deux interprétations alternent ! À un rythme aléatoire, la grille dominante glisse à l'arrière-plan, et les rôles s'inversent. Alexandre Pouget et ses collaborateurs ont montré que, si l'on varie les paramètres de vitesse et d'espacement des deux grilles, le temps que notre cerveau passe à nous donner à voir l'une des deux interprétations est directement proportionnel à sa probabilité, étant donné les entrées sensorielles[10]. Ce que nous voyons à un instant donné est donc, le plus souvent, l'interprétation la plus probable, mais d'autres possibilités font irruption dans notre conscience, avec une durée proportionnelle à leur vraisemblance. L'inconscient calcule toutes les probabilités, tandis que la conscience les échantillonne au hasard.

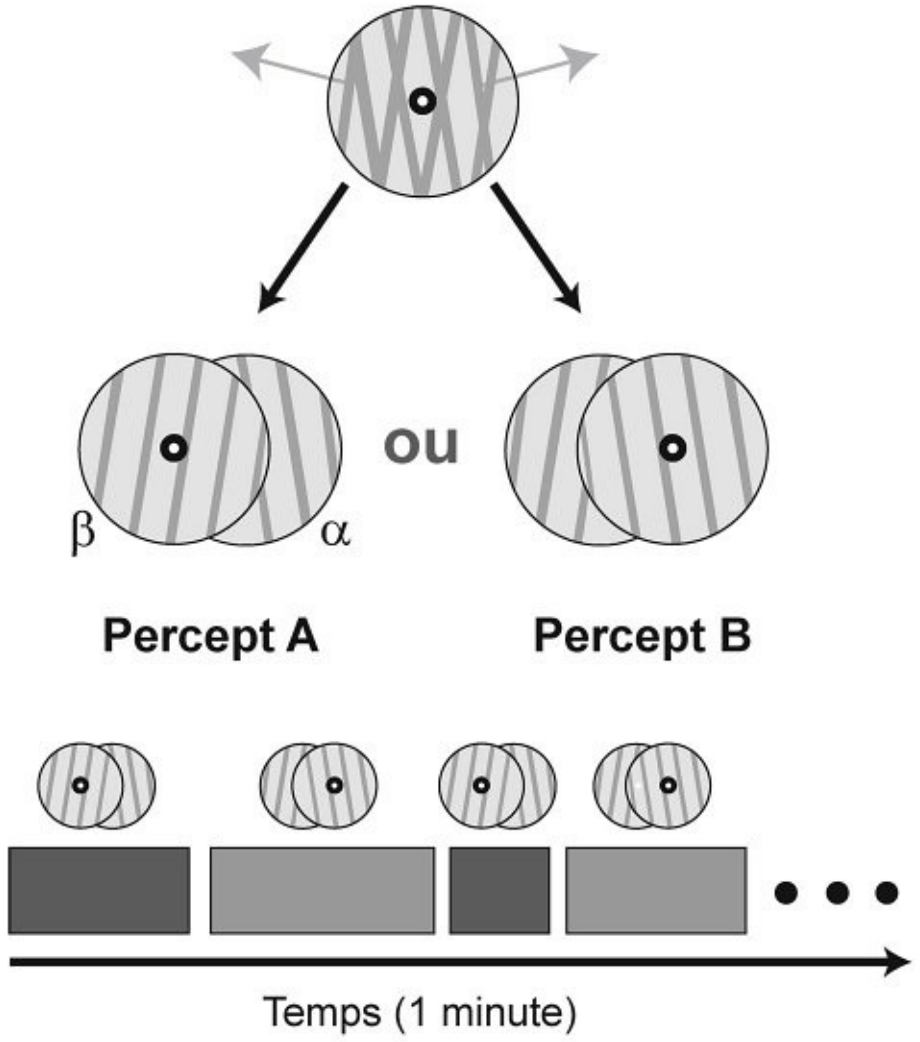


Figure 15. La conscience ne nous donne à voir qu'une seule interprétation des entrées sensorielles. Une image formée de deux grilles superposées est ambiguë : l'une ou l'autre des deux grilles pourrait être au-dessus. Cependant, à un instant donné, nous ne percevons qu'une seule de ces deux interprétations. Notre conscience alterne entre les deux possibilités, et la proportion de temps passée dans chaque état reflète directement la probabilité que cette interprétation soit la bonne. Inconsciemment, notre cerveau calcule toute la distribution de probabilités, mais notre conscience ne nous donne à voir qu'un échantillon de cette distribution.

Cette loi probabiliste démontre qu'au moment même où nous percevons l'une des interprétations d'une scène visuelle notre cerveau continue de s'interroger, inconsciemment, sur les autres possibilités et se prépare à changer d'avis à tout instant. En coulisse, un Sherlock Holmes

inconscient ne cesse de calculer le jeu des possibles. Comme le dit Peirce : « L'étoffe de notre savoir n'est qu'un entrelacs d'hypothèses pures confirmées et raffinées par l'induction. » Pourtant, consciemment, nous ne percevons qu'un seul échantillon parmi toutes ces possibilités. C'est pourquoi nous n'avons aucune conscience des calculs mathématiques qui sous-tendent notre vision : nous nous contentons d'ouvrir les yeux, et notre cerveau conscient ne nous donne qu'une seule chose à voir. Paradoxalement, le processus d'échantillonnage qui gouverne notre vision consciente nous rend aveugles à sa complexité interne.

Échantillonner le jeu des possibles semble être l'apanage de la conscience. En effet, cet échantillonnage se fige dès que nous cessons de prêter attention à ce que nous voyons. Prenons l'exemple de la vision binoculaire, que nous avons étudiée au chapitre 1 – l'instabilité qui survient quand on présente des images différentes aux deux yeux. Quand nous y faisons attention, les deux images alternent dans notre conscience. Bien que l'entrée sensorielle soit fixe et ambiguë, nous la percevons comme nette et changeante, parce que nous voyons tantôt une image, tantôt l'autre. Crucialement, dès que nous orientons notre attention ailleurs, cette rivalité cesse[11]. L'alternance d'images ne se produit qu'en présence d'une attention consciente.

Les processus inconscients sont donc, dans une certaine mesure, plus objectifs que la perception consciente. Notre armée de neurones inconscients évalue toute la distribution de probabilité des états du monde, tandis que la conscience la réduit à quelques échantillons. Les processeurs inconscients travaillent avec des probabilités continues, mais notre esprit conscient n'a accès qu'à des symboles discrets dont le contenu bascule soudainement, en tout-ou-rien. L'inconscient quantifie, la conscience discrétise.

Cette organisation ressemble étrangement à celle de la mécanique quantique. En effet, celle-ci nous dit que la réalité est faite d'une superposition de fonctions d'ondes qui gouvernent la probabilité de trouver une particule dans un certain état. Dès que nous tentons de

mesurer cette réalité, les fonctions d'onde s'effondrent et nos instruments ne détectent que des états discontinus, discrétisés, « tout-ou-rien ». On ne voit jamais d'étranges mixtures telles que le fameux chat de Schrödinger, mi-mort et mi-vivant. Selon la théorie quantique, l'acte de mesure lui-même contraint les probabilités à s'effondrer dans un état déterminé (c'est la fameuse « réduction du paquet d'onde »). Or les données que nous venons d'examiner montrent qu'un phénomène similaire survient dans le cerveau : le simple fait de prêter attention à un objet fait s'écrouler la distribution de probabilité de toutes ses interprétations possibles, et ne nous donne à voir que l'une d'entre elles. La conscience se comporte comme un instrument de mesure qui discrétise le réel et ne nous donne à voir qu'un minuscule aperçu de la vaste étendue des calculs inconscients.

Cette analogie séduisante offre-t-elle plus qu'un parallèle superficiel ? Seule la recherche nous dira si les équations mathématiques de la théorie quantique s'ajustent aux détails de la perception bistable. Sur le plan des mécanismes, il est presque certain que la biophysique des neurones ne fait appel qu'à la physique et à la chimie classiques, et que le cerveau n'est pas un ordinateur quantique (nous y reviendrons). Cependant, à un niveau plus abstrait, il se pourrait que le formalisme quantique s'applique bien au code neural, dans la mesure où le taux de décharge de chaque neurone se comporte comme l'une des coordonnées mathématiques d'un immense « espace mental » pourvu d'autant de dimensions que nous avons de neurones dans le cortex.

Ce qui est certain, c'est que notre cerveau abrite deux grandes catégories de processus : les processeurs inconscients, infatigables calculateurs prodiges de la statistique, et le traitement conscient, qui fonctionne sur la base d'un lent échantillonnage. Nous retrouvons cette division du travail dans le domaine du langage[12]. Lorsque nous entendons un mot ambigu tel que « vol », comme nous l'avons vu au chapitre 2, ses deux acceptions s'activent dans notre lexique inconscient, alors que notre esprit conscient n'en perçoit qu'une, celle qui convient le mieux au contexte[13]. Le même principe gouverne également notre attention.

Consciemment, nous ne pouvons faire attention qu'à un seul endroit à la fois, mais le mécanisme inconscient qui sélectionne les objets dignes d'intérêt évalue en réalité plusieurs possibilités en parallèle[14].

Un statisticien inconscient se cache même au sein de notre mémoire. Essayez de répondre à cette question : quel pourcentage des aéroports du monde se trouve aux États-Unis ? Allez-y, hasardez une réponse, même si cela vous semble difficile. C'est fait ? À présent, oubliez votre première réponse, et donnez-en une seconde. L'expérience montre qu'aucune de vos réponses n'est vraiment faite « au hasard » – toutes les deux sont informatives. Mieux, c'est la *moyenne* de vos deux réponses qui s'approche le plus de la vérité[15]. Une fois encore, l'accès à la conscience agit comme une main invisible qui puise dans l'urne des possibles. Nous pouvons y faire un premier tirage, un deuxième et même un troisième sans épuiser le savoir de notre mémoire inconsciente.

Tentons une analogie : la conscience ressemble au porte-parole d'une grande institution. Les grands organismes tels que le FBI, avec leurs dizaines de milliers d'employés, détiennent toujours plus d'informations que n'en expriment leurs communiqués. Comme en témoigne le triste épisode du 11 septembre 2001, lorsque le FBI doit résumer la situation au président des États-Unis, il ne lui est pas toujours facile d'extraire les éléments pertinents du monceau de connaissances qu'engrange chaque employé. Afin d'éviter de se noyer dans un océan de faits bruts, le président s'appuie sur de brèves notes de synthèse, savamment compilées par une pyramide de conseillers, et il ne laisse qu'un seul porte-parole exprimer ce « sens commun ». Cette hiérarchisation des ressources est généralement rationnelle, même si elle conduit parfois à ignorer certains indices subtils qui indiquent qu'une crise de grande ampleur se prépare.

Le cerveau lui-même se comporte comme une vaste institution dont l'effectif approche la centaine de milliards de neurones. Il lui faut donc, lui aussi, des notes de synthèse. Le rôle de la conscience semble être de simplifier la perception de l'environnement en n'en proposant qu'un

résumé pertinent, qui est transmis à toutes les autres aires impliquées dans la mémoire, la décision et l'action.

Pour être utile, cette synthèse consciente doit être stable et unifiée. Au beau milieu d'une crise nationale, il serait absurde que le FBI transmette au président des milliers de notes, chacune portant un brin de vérité, en lui laissant le soin de les réunir. De même, le cerveau ne peut se contenter d'un flux continu de données brutes et disparates. Il doit assembler ces bribes en un tout cohérent. Comme une note au président, ce résumé conscient doit fournir une interprétation lisible de l'environnement, écrite dans un « langage de l'esprit » suffisamment abstrait pour interagir avec les décideurs du plus haut niveau.

Pensées durables

> « Les améliorations que nous apportons à notre cerveau lorsque nous apprenons des langues nous permettent de revoir, de rappeler, de répéter et de repenser nos propres activités, ce qui transforme nos cerveaux en sortes de chambres à écho, au sein desquelles des processus généralement évanescents peuvent s'éterniser et devenir de vrais objets de pensée. Ceux qui persistent le plus, et qui acquièrent de l'influence à mesure qu'ils persistent, nous les appelons des pensées conscientes. »
>
> Daniel DENNETT, *Kinds of Minds* (1996).

> « La conscience est un trait d'union entre ce qui a été et ce qui sera, un pont jeté entre le passé et l'avenir. »
>
> Henri BERGSON,
> Conférence à l'Université de Birmingham (1911).

> « La mémoire et l'imagination sont toutes les deux une négation du temps. »
>
> Vladimir NABOKOV, *Intransigeances* (1973).

Il pourrait y avoir une seconde raison pour que notre conscience condense les messages sensoriels en une synthèse dépourvue d'ambiguïté : ce code compact peut se propager à travers le temps et entrer dans ce que nous appelons la mémoire de travail. Conscience et mémoire de travail semblent étroitement apparentées. On pourrait même affirmer, avec le philosophe Dan Dennett, que l'un des rôles essentiels de la conscience est de créer des pensées durables. Dès qu'une information devient consciente, elle reste présente à l'esprit tant que nous y prêtons attention. La synthèse consciente doit demeurer stable si elle veut informer nos décisions, qui prennent parfois plusieurs minutes à se forger. Cette stabilité, cet épaississement du présent caractérisent la pensée consciente.

Un mécanisme cellulaire de stabilisation de la mémoire existe chez tous les mammifères, depuis les êtres humains jusqu'aux souris en passant par les rats, les chats et les singes. Il confère manifestement bien des avantages au cours de l'évolution. Les organismes qui disposent d'une mémoire peuvent se détacher des contingences immédiates de leur environnement. Ils cessent d'être à la merci du présent, mais peuvent se rappeler le passé et se projeter dans le futur. Si un prédateur se cache derrière un rocher, se souvenir de sa présence est une question de vie ou de mort. De nombreux événements vitaux pour l'animal surviennent à des intervalles variables, en des lieux distincts, et sous les formes les plus diverses. Savoir faire la synthèse d'informations dispersées dans le temps, l'espace et les modalités sensorielles, et parvenir à réactiver ces éléments dans le futur, est une fonction essentielle de la conscience, qui a très certainement fait l'objet d'une pression de sélection positive depuis des millions d'années.

La fonction que les psychologues appellent « mémoire de travail » repose au premier chef sur le cortex préfrontal dorso-latéral et les aires qui lui sont connectées, ce qui fait de ces régions d'excellents candidats pour les mécanismes neuronaux de la conscience[16]. On les voit s'activer dès que nous retenons, même très brièvement, un élément d'information :

un numéro de téléphone, une couleur, la forme d'une image. Les neurones du cortex préfrontal forment une mémoire active : bien après qu'une image a disparu, ils continuent de décharger tout au long de la phase de mémoire, parfois jusqu'à plusieurs dizaines de secondes plus tard. Et lorsque le cortex préfrontal est endommagé par une lésion ou distrait par une autre tâche, cette mémoire de travail s'évanouit.

Les lésions du cortex préfrontal entraînent toutes sortes de déficits de planification du futur. Ces symptômes peuvent être variés, mais leur point commun est une sorte d'adhérence au présent, une incapacité de s'en détacher pour se projeter vers l'avenir. Ainsi ces patients éprouvent-ils souvent des difficultés à inhiber les actions indésirables, et ils se saisissent parfois d'outils pour les utiliser hors contexte (comportement d'utilisation) ou ne peuvent s'empêcher de copier les gestes des autres (comportement d'imitation). Leurs capacités d'inhibition consciente, de réflexion et de planification à long terme peuvent être très altérées. Dans les cas les plus sévères s'installent une apathie et d'autres symptômes qui indiquent une dégradation flagrante du contenu de la vie mentale. Les troubles en lien avec la conscience comprennent l'héminégligence (une anomalie de la prise de conscience d'une moitié de l'espace, généralement la gauche), l'aboulie (une incapacité d'initier toute action volontaire), le mutisme akinétique (un défaut de production spontanée du langage parlé, alors que la répétition est fréquemment intacte), l'anosognosie (absence de conscience d'un déficit majeur tel qu'une paralysie ou une cécité) et le déficit de mémoire « autonoétique » (l'incapacité de se souvenir de ses propres pensées). Une perturbation du cortex préfrontal peut même interférer avec des capacités aussi élémentaires que la prise de conscience d'une image présentée très brièvement[17].

En résumé, le cortex préfrontal semble jouer un rôle clé dans notre capacité de maintenir une information pendant un certain temps, d'y réfléchir et de l'intégrer dans un plan d'action. Existe-t-il une preuve plus directe que ces opérations impliquent toujours la conscience ? Robert Clark et Larry Squire ont inventé un test remarquablement simple de la

synthèse d'informations dispersées dans le temps : le conditionnement différé de la réponse réflexe de la paupière[18]. À un instant précis, une machine pneumatique envoie une bouffée d'air en direction de l'œil. La réponse est immédiate : chez le lapin comme chez l'homme, la membrane protectrice de la paupière se ferme instantanément. Maintenant, imaginez que l'on précède chaque bouffée d'air d'un bref signal sonore. Vous obtenez un conditionnement pavlovien (ainsi nommé en l'honneur du physiologiste russe Ivan Petrovich Pavlov, qui découvrit que les chiens pouvaient être conditionnés à saliver au son d'une cloche qui annonce de la nourriture). Un entraînement minime suffit pour que l'œil cligne lors de l'émission du son, avant même que ne survienne la bouffée d'air. Au bout du compte, le son seul déclenche un clignement de l'œil – on dit que le réflexe a été « conditionné ».

La fermeture de la paupière est rapide, mais est-elle inconsciente ? La réponse, curieusement, dépend de la présence ou non d'un délai temporel. Dans l'une des versions du test, appelée « conditionnement différé », le son dure jusqu'à la bouffée d'air, en sorte que les deux stimuli coïncident brièvement dans le cerveau de l'animal. L'apprentissage devient alors une simple affaire de coïncidence à détecter. Dans l'autre version, appelée « conditionnement de trace », le son ne dure qu'un bref instant et un long silence le sépare de la bouffée d'air. Bien que minimalement différente, cette version pose clairement un défi : pour parvenir à découvrir la relation temporelle systématique entre le son et la bouffée d'air, l'organisme doit garder en mémoire une trace du son passé. Pour éviter toute confusion, je propose de renommer la première version « conditionnement fondé sur la coïncidence » (le premier stimulus dure jusqu'à coïncider avec le second, sans qu'il soit besoin de le mémoriser) et la seconde, « conditionnement fondé sur la mémoire » (seule la mémoire permet de franchir le délai entre les deux stimuli).

Les résultats expérimentaux sont clairs : le conditionnement fondé sur la coïncidence se produit sans la moindre conscience, alors que pour le conditionnement fondé sur la mémoire, il semble nécessaire que

l'organisme soit conscient[19]. En fait, la détection des coïncidences ne fait même pas appel au cortex. Un lapin décérébré, privé de tout cortex, dépouillé également de ses ganglions de la base, de son système limbique, de son thalamus et de son hypothalamus, continue d'apprendre le lien entre son et bouffée d'air lorsque ces événements coïncident dans le temps. Par contre, pour le conditionnement fondé sur la mémoire, aucun apprentissage n'est possible lorsque l'hippocampe et les régions qui s'y connectent (dont le cortex préfrontal) sont détériorés. Chez l'homme, le conditionnement ne survient que dans la mesure où la personne rapporte avoir pris conscience d'un lien entre le son et la bouffée d'air. Les personnes âgées, amnésiques, ou simplement trop distraites pour prendre conscience de ce lien, ne montrent pas le moindre conditionnement – alors que ces manipulations n'ont aucun effet sur le conditionnement fondé sur la coïncidence. L'imagerie montre que les sujets qui prennent conscience du lien sont précisément ceux qui activent leur cortex préfrontal et leur hippocampe au cours de l'apprentissage.

Le paradigme de conditionnement suggère que la conscience joue un rôle essentiel : c'est elle qui nous permet de relier les événements à travers le temps plutôt que de vivre dans l'instant. Le circuit que forme le cortex préfrontal et les aires qui lui sont connectées, y compris l'hippocampe, semble avoir évolué afin de combler les intervalles de temps. La conscience nous fournit un « présent remémoré », selon la formule de Gerald Edelman[20] : grâce à elle, nous projetons notre expérience passée en direction du futur et nous lui permettons de tisser des liens avec le présent.

Le test de conditionnement fondé sur la mémoire présente un immense avantage : il est si simple qu'il peut être administré à toutes sortes d'organismes, depuis le bébé jusqu'au singe, au lapin ou à la souris. Lorsqu'une souris passe ce test, elle active massivement les régions antérieures de son cerveau, qui sont l'homologue du cortex préfrontal humain[21]. Il semble donc que le test exploite l'une des fonctions les plus

élémentaires de la conscience, une opération tellement essentielle qu'elle pourrait bien être déjà présente chez bien d'autres espèces que la nôtre.

Si l'enregistrement d'une information en mémoire de travail exige d'en être conscient, est-ce que cela implique que les représentations inconscientes ne durent pas ? Effectivement, il semble bien que l'activité cérébrale évoquée par une image subliminale ne dure qu'un bref instant[22]. La durée de vie d'un message subliminal peut être estimée en mesurant au bout de combien de temps son efficacité décroît. Le résultat est sans appel : alors qu'une image visible conserve son pouvoir d'influence pendant des dizaines de secondes, un stimulus invisible n'exerce qu'un effet ponctuel sur nos pensées. Lorsque nous rendons un mot invisible en le masquant, il continue d'activer des représentations visuelles, orthographiques, lexicales et même sémantiques, mais la durée de cette activité est brève. Au bout de 1 ou 2 secondes, l'activation inconsciente s'affaiblit tant qu'elle en devient indétectable.

Pour résumer ces observations, Lionel Naccache se permet de paraphraser Jacques Lacan : « L'inconscient est structuré comme une exponentielle décroissante[23] ! » En faisant un effort, nous parvenons à maintenir une information subliminale active pendant une période un peu plus longue, mais la qualité de cette mémoire inconsciente est si médiocre qu'après quelques secondes de délai notre capacité de rappel ne dépasse guère le niveau du hasard[24]. Seule la conscience nous donne accès à des pensées durables.

La machine de Turing humaine

Lorsque nous avons une information « en tête », qu'elle est bien protégée de l'oubli, pouvons-nous lui faire subir des opérations spécifiques, que l'inconscient ne permet pas de réaliser ? La réponse semble positive : au moins chez l'espèce humaine, l'accès d'une

information à la conscience permet de l'intégrer dans des programmes mentaux. Je postule donc que la conscience confère au cerveau humain les capacités d'un puissant ordinateur séquentiel.

Essayez par exemple de calculer mentalement le produit de 12 par 13.

Vous y êtes ?

Avez-vous ressenti toute une série d'opérations se succéder dans votre cerveau, l'une après l'autre ? Pourriez-vous décrire avec précision chacune des étapes que vous avez suivies et les résultats intermédiaires que vous avez obtenus ? En général, la réponse est oui : nous avons conscience des stratégies séquentielles que nous déployons au cours d'un calcul mental. Personnellement, je me suis d'abord souvenu que 12^2 fait 144, puis j'ai ajouté 12. D'autres personnes multiplient les chiffres l'un après l'autre, de la droite vers la gauche, selon la méthode traditionnelle qu'elles ont apprise à l'école. Quelle que soit la stratégie que nous employons, nous parvenons à la rapporter consciemment, et cette introspection subjective s'avère généralement juste, en comparaison avec les mesures objectives du temps de réaction et des mouvements oculaires[25].

En psychologie, le fait que l'introspection donne des résultats valides n'est pas une chose banale. La plupart de nos opérations mentales sont opaques : nous avons beau y réfléchir, nous n'avons pas la moindre idée des opérations qui nous permettent de reconnaître un visage, de monter un escalier ou de nommer un mot. Pourtant, le calcul à plusieurs chiffres a quelque chose de différent : il se décompose en une série d'étapes accessibles à l'introspection.

Pourquoi en va-t-il ainsi ? Selon moi, l'explication est simple. L'exécution de stratégies complexes, qui comprennent une série d'étapes et que les informaticiens appellent des « algorithmes », fait appel à la mémoire de travail pour stocker chaque résultat intermédiaire et le passer à l'étape suivante. C'est donc une fonction exclusive de la conscience.

Si le problème 12×13 vous était présenté si brièvement que vous n'en ayez pas conscience, votre cerveau pourrait-il en calculer le résultat inconsciemment ? Il est presque certain que non[26]. Ce genre de calcul est

remarquablement lent et sériel. Il demande un « routage » de l'information afin de récupérer le résultat de l'étape n et le passer à l'étape $n + 1$. Le cerveau doit donc contenir un routeur[27], un dispositif capable de recueillir l'information, où qu'elle soit initialement représentée, et de la diffuser à toutes sortes de processeurs internes. À leur tour, ces processeurs appliquent leurs calculs inconscients à ce symbole, et le cycle se répète autant de fois que nécessaire. Il en résulte une machine hybride, à la fois parallèle et sérielle, dans laquelle des étapes de calcul parallèle et inconscient alternent avec des étapes de décision consciente et de routage de l'information.

Avec mes amis physiciens Mariano Sigman et Ariel Zylberberg, nous avons commencé à explorer les propriétés de cette architecture cérébrale[28]. Elle ressemble étroitement à ce que les informaticiens appellent un « système de productions », un type de programme informatique employé en intelligence artificielle dans les années 1960. Un système de productions consiste en une base de données, qu'on appelle également « mémoire de travail », et un jeu de règles « si… alors » (par exemple, s'il y a un symbole A en mémoire de travail, le remplacer par la séquence BC). À chaque étape, le système examine si l'une des règles s'applique à l'état actuel de la mémoire de travail. Si plusieurs règles conviennent, elles entrent en compétition sous l'égide d'un système de contrôle des priorités. Enfin, la règle qui gagne est mise en œuvre. Son déclenchement, également appelé « ignition », modifie la mémoire, et on passe à l'étape suivante. Ce mécanisme engendre donc une série de cycles d'accès conscient, d'ignition et de diffusion du résultat.

Bien que les systèmes de production soient très simples, ils peuvent réaliser n'importe quel calcul : toute procédure effective peut se réaliser sous forme de règles si… alors. Leur puissance est équivalente à celle de la machine de Turing, un dispositif théorique imaginé par le mathématicien anglais Alan Turing en 1936, et qui est à l'origine de l'invention de l'ordinateur digital[29]. Ainsi, notre proposition équivaut à dire que le cerveau conscient, parce qu'il est doté d'un routeur flexible,

fonctionne comme une machine de Turing humaine qui nous permet d'exécuter mentalement n'importe quel algorithme. Ses calculs sont très lents, parce que chaque résultat intermédiaire doit être stocké en mémoire de travail avant d'être envoyé à l'étape suivante – mais sa puissance de calcul est imposante.

L'histoire de la découverte d'Alan Turing n'est pas dépourvue d'intérêt pour le psychologue. L'article de Turing répondait à un défi posé par le mathématicien David Hilbert en 1928 : peut-on trouver une procédure mécanique qui imite la réflexion d'un mathématicien ? Par de simples manipulations de symboles, une telle procédure parviendrait à décider si, oui ou non, une proposition mathématique découle logiquement des axiomes. La machine de Turing était donc conçue pour imiter « un homme en train de calculer un nombre réel » (ce sont les mots mêmes de Turing, dans son fameux article de 1936). Mais de quelles données Turing disposait-il pour simuler le cerveau du mathématicien ? Turing n'était ni un psychologue ni un spécialiste des neurosciences. La seule donnée dont il disposait, c'était son introspection consciente. C'est pourquoi, selon moi, sa machine n'imite qu'une partie des processus mentaux du mathématicien : ceux qui sont accessibles à la conscience. Les opérations sérielles et symboliques que Turing a placées au cœur de sa machine forment un modèle assez raisonnable des opérations que notre cerveau est capable d'effectuer sous le contrôle de la conscience.

Ne vous méprenez pas, je n'ai aucune intention de relancer la métaphore élimée qui identifie le cerveau à un ordinateur. Avec son architecture massivement parallèle, plastique, qui se modifie en permanence au fil de l'expérience, et qui est capable de calculer avec des distributions de probabilité plutôt que des symboles discrets, l'organisation du cerveau humain ne ressemble guère à celle d'un ordinateur d'aujourd'hui. Les neurosciences ont, depuis longtemps, rejeté cette métaphore. Et pourtant, le *comportement* du cerveau, lorsqu'il s'engage dans de longs calculs, s'approche de celui d'un système de production ou d'une machine de Turing sérielle[30]. Pour prendre un

exemple, le temps qu'il nous faut pour calculer une addition telle que 235 + 457 est la somme des durées de chaque étape élémentaire (calcul de 5 + 7 ; mémorisation de la retenue ; calcul de 3 + 5 + 1, et enfin de 2 + 4) – exactement ce qu'on prédit si les étapes se déroulent en série[31].

Le modèle de Turing est une idéalisation. Si l'on regarde le comportement humain de plus près, on découvre qu'il s'écarte de cette séquence idéale. Au lieu d'être bien distinctes dans le temps, les étapes se recouvrent partiellement, ce qui engendre des interférences indésirables entre les opérations[32]. Ainsi, au cours d'un calcul mental, il arrive que la deuxième opération démarre avant que la première ne soit complètement terminée. Jérôme Sackur et moi-même avons étudié l'exécution d'un algorithme mental simplissime : prenez un nombre n, ajoutez-lui 2 (n + 2), et décidez si le résultat est plus petit ou plus grand que 5 (n + 2 > 5 ?). Nous avons observé qu'inconsciemment, les participants commencent l'étape de comparaison trop tôt, sur le nombre n lui-même, avant même d'avoir obtenu le résultat intermédiaire n + 2[33]. Dans un ordinateur, ce genre d'erreur ne peut pas se produire : une horloge centrale et un routeur digital s'assurent que chaque bit atteigne sa destination en temps et en heure. Le cerveau, lui, n'a jamais évolué pour faire des calculs exacts. Son architecture, sélectionnée pour survivre dans un monde probabiliste, explique que nous fassions autant d'erreurs dans nos calculs. Nous « recyclons » pour l'arithmétique des réseaux destinés au calcul approximatif et au traitement du langage, en utilisant nos capacités de contrôle conscient pour les faire échanger des informations d'une manière lente et sérielle[34].

Si l'une des fonctions de la conscience est d'être la *lingua franca* du cerveau, le « sens commun » qui permet aux processeurs cérébraux spécialisés d'échanger des informations avec une grande flexibilité, il s'ensuit une prédiction simple : une opération routinière devrait pouvoir s'exécuter sans conscience, mais il devrait être impossible d'enchaîner inconsciemment plusieurs opérations. Dans le domaine du calcul mental, par exemple, notre cerveau pourrait calculer 3 + 2 inconsciemment, mais

pas $(3 + 2)^2$, $(3 + 2) - 1$, ou $\frac{3+2}{5}$. L'exécution d'une série de calculs devrait toujours exiger un effort conscient[35].

Jérôme Sackur et moi avons mis cette idée à l'épreuve. Nous avons masqué un chiffre *n* en sorte que les participants ne parvenaient à le voir que la moitié du temps, et nous leur avons demandé d'exécuter toutes sortes d'opérations. Ils devaient tantôt nommer le chiffre, tantôt lui ajouter 2 (la tâche « $n + 2$ »), tantôt le comparer avec 5 (tâche « $n > 5$? »). Une quatrième instruction, enfin, consistait à ajouter 2, puis à comparer le résultat avec 5 (la tâche « $n + 2 > 5$? »).

Dans les trois premières tâches, nos participants montrèrent des performances qui dépassaient le niveau du hasard. Même s'ils juraient n'avoir absolument rien vu, nous les forcions à répondre, et ils étaient surpris d'apprendre l'étendue de leurs performances inconscientes. Ils pouvaient nommer un chiffre invisible : leur réponse était correcte la moitié du temps, alors que s'ils se fiaient au hasard, leur score n'aurait pas dû excéder 25 % (étant donné qu'il y avait quatre chiffres possibles). Ils parvenaient même à ajouter 2 ou à comparer avec 5. Toutes ces opérations font appel à des routines familières, et comme nous l'avons vu au chapitre 2, bien des données montrent qu'elles s'exécutent, au moins en partie, sans que nous en ayons conscience. Crucialement, pourtant, la dernière tâche (« $n + 2 > 5$? ») était différente : lors des essais inconscients, les participants y répondaient au hasard. Ce résultat est étrange puisque, s'ils avaient pensé à nommer le chiffre *n*, et à utiliser leur réponse comme le point de départ de leur calcul, ils auraient obtenu d'excellents résultats ! De toute évidence, leur cerveau contenait des informations subliminales, puisqu'ils parvenaient à nommer le chiffre la moitié du temps. Pourtant, en l'absence de conscience, ces informations ne parvenaient pas à passer d'une étape à l'autre.

Au chapitre 2, nous avons vu que le cerveau peut additionner, inconsciemment, les données issues de plusieurs flèches[36], d'une série de chiffres[37], ou même d'indications sur l'achat d'une voiture[38]. Est-ce

contradictoire ? Non – parce que l'addition de plusieurs indices fait partie des opérations élémentaires du cerveau. Lorsqu'un accumulateur neuronal s'ouvre, n'importe quelle information, consciente ou inconsciente, peut le biaiser dans un sens ou dans un autre. La seule chose que notre système de décision inconsciente ne fait pas bien, c'est de parvenir à une réponse claire et nette, qui puisse être transmise aux étapes suivantes du calcul. Notre accumulateur central laisse entrer les informations inconscientes, mais cela ne lui suffit pas à atteindre le seuil requis pour émettre une décision sans appel et passer à la suite. C'est pourquoi, dès que la tâche demande une série d'étapes, notre inconscient reste bloqué à l'étape d'accumulation nécessaire à la première décision et ne passe jamais à la seconde.

Autre conséquence plus générale : nous ne pouvons pas raisonner logiquement à partir d'une intuition inconsciente. Les informations subliminales n'entrent pas dans nos réflexions stratégiques. Cette conclusion peut sembler évidente, mais il n'en est rien. Les stratégies auxquelles nous réfléchissons ne sont, après tout, qu'un exemple parmi d'autres de processus cérébraux – il n'y a donc rien de trivial à constater que ce type de processus ne peut pas s'exécuter sans conscience. De plus, cette observation est riche de conséquences. Vous souvenez-vous de la tâche d'addition de flèches, dans laquelle on voit une série de cinq flèches tournées vers la droite ou vers la gauche, et pour laquelle il s'agit de dire vers où pointe la majorité d'entre elles ? N'importe quel esprit conscient découvre rapidement une stratégie gagnante : dès qu'on a vu trois flèches pointer du même côté, les dés sont jetés et rien ne pourra plus changer le résultat. Les participants exploitent ces stratégies afin de répondre plus vite. Cependant, une fois de plus, ils n'y parviennent que si les flèches sont conscientes, pas si elles sont subliminales. Quand les flèches sont masquées au point de devenir pratiquement invisibles, le cerveau ne peut que les additionner – il ne parvient pas à appliquer la stratégie des trois flèches et à passer à la suite.

Dans leur ensemble, toutes ces expériences révèlent un nouveau rôle clé de la conscience. Nous avons besoin d'être conscients pour réfléchir rationnellement. L'extraordinaire puissance de calcul de l'inconscient aboutit certes à des intuitions profondes, mais seul l'esprit conscient peut suivre une stratégie rationnelle, pas après pas. En jouant le rôle de « routeur », capable de faire circuler l'information à travers toute une série d'étapes, la conscience nous donne accès à un mode opératoire inédit – une machine de Turing humaine.

Le premier réseau social

> « La conscience n'est en somme qu'un réseau de communications d'homme à homme, ce n'est que comme telle qu'elle a été forcée de se développer : l'homme solitaire et bête de proie aurait pu s'en passer. »
>
> Friedrich NIETZSCHE, *Le Gai Savoir* (1882).

Chez l'*Homo sapiens*, l'information consciente ne voyage pas seulement à l'intérieur d'une tête. Par la magie du langage, elle peut également sauter d'une tête à l'autre. Dans l'évolution récente de l'humanité, le partage de l'information au sein d'un réseau social pourrait bien avoir été l'une des fonctions essentielles de la conscience. L'« homme bête de proie » de Nietzsche possédait certainement une mémoire de travail consciente depuis des dizaines de millions d'années – mais ce n'est qu'avec le genre *Homo* qu'est apparue la capacité de communiquer cet état de conscience. Grâce au langage autant qu'aux gestes, la synthèse consciente qui émerge dans l'esprit d'une personne peut être partagée avec d'autres. Cette transmission sociale des symboles conscients ouvre la voie à de nouveaux modes de « calcul en société ». Elle donne à notre espèce accès à des algorithmes sociaux « multicœurs » (comme disent les informaticiens), qui ne reposent pas sur le savoir d'un seul, mais sur la

confrontation des points de vue, des expertises et des sources de connaissances.

Ce n'est certainement pas un accident si la capacité de formuler une pensée en mots est considérée comme le critère essentiel de conscience. Nous estimons qu'une personne est consciente d'une information dès lors qu'elle peut l'exprimer, au moins en partie, sous forme de mots (en supposant, évidemment, qu'elle ne soit pas paralysée, aphasique ou trop jeune pour savoir parler). Chez l'homme, l'« interpréteur verbal » qui nous permet d'exprimer le contenu de nos pensées fait partie des processus cognitifs centraux que nous ne parvenons à déployer que lorsque nous sommes conscients[39].

Je ne prétends pas, bien sûr, que nous soyons toujours capables d'exposer notre état d'esprit avec une précision digne des plus grandes pages de Proust. Notre conscience outrepasse nos talents linguistiques : nous percevons bien plus que nous ne saurions décrire. Un chef-d'œuvre du Caravage, un splendide coucher de soleil sur le Grand Canyon ou l'expression furtive d'un visage de bébé échappent à toute description verbale exhaustive – ce qui contribue sans doute grandement à la fascination qu'ils exercent. Néanmoins, et pratiquement par définition, tout ce dont nous prenons conscience peut être formulé en mots. Au minimum, nous parvenons à l'indiquer aux autres par le biais d'un arsenal de pointeurs que les linguistes appellent des pronoms (« lui »), des déictiques (« ça, là ») ou des explétifs (« ouah, tu as vu ? ») – ou tout simplement en pointant du doigt plutôt que du verbe. Le langage nous permet de couler nos pensées conscientes dans des catégories et des structures syntaxiques qui structurent nos pensées autant qu'elles en assurent le partage avec d'autres.

La nécessité de diffuser nos informations à d'autres explique pourquoi notre cerveau juge avantageux de s'abstraire des détails de nos sensations et d'en faire une « note de synthèse » consciente. Le canal de communication qu'offrent les mots et les gestes est étonnamment réduit – il ne dépasse pas 40 à 60 bits par seconde[40], soit 300 fois plus lent que les

modems et les fax qui ont révolutionné nos bureaux dans les années 1990 (et qui sont aujourd'hui relégués au musée). Notre cerveau doit donc comprimer l'information en un tout petit jeu de symboles qui seront envoyés dans le réseau social. Il serait d'ailleurs sans intérêt que je transmette à d'autres une image mentale précise de ce que je perçois. Ce qui compte pour les autres, ce n'est pas de recevoir un portrait fidèle de ce que je vois, mais un résumé des aspects qui ont de bonnes chances d'être utiles de leur point de vue : une représentation multisensorielle, stable, invariante et durable de l'environnement que nous partageons. Chez l'homme au moins, la conscience semble condenser l'information perçue sous la forme d'un résumé de ce qui pourrait intéresser les autres.

Le lecteur pourrait objecter que le langage ne sert, le plus souvent, qu'à des choses triviales, comme de cancaner sur les frasques de telle actrice de Hollywood. Selon l'anthropologue d'Oxford Robin Dunbar, près des deux tiers de nos conversations portent sur des commérages. Il a même proposé une théorie de l'évolution du langage dite « épouillage et commérage » (*grooming and gossip*), selon laquelle le langage n'est apparu que pour créer du lien social, pas vraiment pour communiquer[41].

Pouvons-nous prouver que nos conversations dépassent le niveau d'un tabloïd ? Qu'elles transmettent aux autres la synthèse exacte dont ils ont besoin pour prendre des décisions ? Le psychologue iranien Bahador Bahrami a réalisé une expérience ingénieuse[42]. Il a demandé à deux personnes d'effectuer la même tâche. Elles voyaient deux images successives et on leur demandait de décider lequel de ces deux écrans contenait une image cible, présentée au seuil de perception consciente. À chaque essai, chacun des participants devait d'abord donner sa propre réponse, en toute autonomie. L'ordinateur affichait ensuite ces choix, et s'ils étaient en désaccord, il demandait aux sujets d'en discuter jusqu'à ce qu'ils parviennent à résoudre ce conflit.

Un aspect de cette expérience était particulièrement intelligent : au bout du compte, chaque paire de sujets se comportait comme un seul homme. Tous deux ne voyaient qu'un seul stimulus et ne donnent qu'une

seule réponse, dont la précision pouvait être évaluée avec les bonnes vieilles méthodes de la psychophysique, comme s'ils ne faisaient vraiment qu'un. Or les résultats étaient clairs : dès que les participants étaient de niveau comparable, les faire discuter augmentait leurs performances. Le groupe faisait systématiquement mieux que le meilleur de deux individus – ce qui faisait dire aux auteurs que « deux têtes valent mieux qu'une ».

L'un des avantages de l'expérience de Bahrami est que cette situation de communication est suffisamment simple pour être modélisée mathématiquement. En supposant que chaque personne perçoit le monde avec une incertitude qui lui est propre, il est aisé de calculer comment leurs sensations devraient se combiner : pour maximiser les chances de réussite, les sensations que reçoit chaque joueur doivent être divisées par la précision de chacun, puis additionnées en une seule « sensation de groupe ». Cette règle optimale pour la décision multicerveaux est, en fait, exactement la même que celle qui gouverne la fusion d'indices multisensoriels au sein d'un même cerveau. On peut la remplacer par un raccourci simple : la plupart du temps, les gens n'ont pas besoin de communiquer toutes les nuances de ce qu'ils ont vu (ce qui serait impossible), mais seulement une réponse catégorique, tout ou rien (ici, écran 1 ou écran 2), accompagnée d'un jugement de confiance en soi.

Il se trouve que les paires de participants qui passent le test avec succès sont précisément celles qui adoptent cette stratégie. Elles échangent leur degré de confiance en utilisant des mots tels que *sûr et certain, aucune idée,* ou *j'ai cru voir quelque chose*. Au fil des essais, certains couples adoptent même une échelle numérique afin d'exprimer au mieux leur degré de certitude. Dès qu'ils mettent en commun leurs doutes et leurs croyances, la performance de leur équipe s'élève jusqu'à l'optimum prédit par la théorie. Le tout dépasse la somme des parties.

Les expériences de Bahrami permettent de spéculer sur l'origine de la conscience de soi dans l'évolution. En nous laissant partager avec d'autres nos doutes et nos certitudes, la conscience réflexive, celle qui évalue en permanence nos propres jugements, nous permet de prendre,

collectivement, de meilleures décisions. Conscience et confiance en soi sont étroitement liées. Pour qu'elles soient utiles à d'autres, nos pensées conscientes doivent être marquées au sceau de la confiance. En tant qu'*Homo sapiens sapiens*, non seulement nous savons que nous savons, mais nous parvenons à estimer le niveau de certitude de chacune de nos idées et à le comparer avec celui des autres. Nous jaugeons en permanence la fiabilité de nos sources, en gardant à l'esprit qui a dit quoi à qui, et si cela s'est avéré juste ou faux (c'est pourquoi le commérage occupe une place centrale dans nos conversations). Ces évolutions, pratiquement uniques à l'espèce humaine, font de l'autoévaluation de la certitude un ingrédient essentiel de l'algorithme qui régit nos décisions collectives.

La théorie du révérend Bayes nous dit qu'il convient d'appliquer exactement les mêmes règles de décision à nos pensées et aux informations que nous donnent les autres. Chaque source d'information, qu'elle soit interne ou externe, doit être pondérée, aussi précisément que possible, par son degré d'incertitude, avant que toutes ces informations ne soient combinées entre elles. Avant l'hominisation, le cortex préfrontal de tous les primates fournissait déjà un espace de mémoire dans lequel les informations passées et présentes, pondérées par leur fiabilité, pouvaient être rassemblées afin de guider la prise de décision. À partir de là, un pas majeur de l'évolution a été franchi lorsque cet espace de travail mental s'est ouvert aux informations issues de l'esprit des autres. Le développement de cette interface sociale nous a permis de récolter les bénéfices d'un algorithme de prise de décision collective : en comparant notre savoir à celui des autres, nous affinons notre jugement.

L'imagerie cérébrale commence à élucider les réseaux neuronaux du partage social. Dès que nous tentons de lire l'esprit des autres, les régions les plus antérieures du cortex préfrontal, dans le pôle frontal et le long de la ligne médiane du cerveau (le cortex préfrontal ventro-médian) entrent en activité. On observe également l'activation de la jonction des lobes temporaux et pariétaux, ainsi que de la face ventrale du cortex pariétal (le

précuneus). Toutes ces régions forment un réseau à grande échelle, interconnecté par des faisceaux de fibres nerveuses à longue distance. Ce réseau s'active également dans l'état de repos : dès que nous avons quelques secondes de libre, notre cerveau retourne spontanément à ce « mode par défaut » où s'activent les réseaux sociaux du cortex[43].

Plus surprenant, quoique en accord avec l'origine de la conscience humaine dans le partage social, ce réseau s'active lorsque nous réfléchissons à nous-mêmes – par exemple lorsque nous utilisons notre introspection pour réfléchir au degré de confiance que nous avons dans nos propres décisions[44] – tout autant que lorsque nous songeons aux pensées des autres[45]. Le pôle frontal et le cortex préfrontal ventro-médian, en particulier, montrent des réponses étonnamment similaires lorsque nous portons des jugements sur nous-mêmes et sur les autres[46] – à tel point que penser à l'un facilite le fait de penser à l'autre[47]. Ainsi, ce réseau cortical dispose de toutes les caractéristiques nécessaires à l'évaluation de nos connaissances et leur confrontation à celles d'autres.

En résumé, notre cerveau comprend un jeu de circuits neuronaux de la conscience de soi et des autres. Nous utilisons les mêmes bases de données pour nous représenter nous-mêmes et pour accumuler des informations sur les autres. Ces réseaux cérébraux spécialisés construisent une image interne de nous-mêmes au même titre que les autres personnes de notre connaissance. Chacun d'entre nous se figure « soi-même comme un autre », selon le mot du philosophe français Paul Ricœur[48].

Si cette théorie est vraie, on est en droit de parler d'une véritable construction du « soi », indirecte et fragmentaire. Nous passons notre vie à observer notre propre comportement et celui des autres. Notre cerveau tire de ces observations des inférences statistiques qui élèvent notre esprit – au sens le plus littéral du terme. Notre conception de l'âme n'est qu'une construction de notre cerveau, un modèle interne qui rend à peu près compte des notes de nos observations[49]. À force de passer notre vie avec nous-mêmes, nous nous construisons une image cérébrale de notre

caractère, de nos connaissances et de notre confiance en nous-mêmes, forcément un peu plus raffinée que celle que nous avons des autres.

Notre représentation du « moi » n'est pas seulement informée de notre comportement, mais aussi de certaines de ses causes cachées (intentions, sentiments, croyances, etc.)[50]. L'introspection rend certaines de nos motivations et de nos stratégies conscientes, donc transparentes à nous-mêmes, alors que nous n'avons aucun moyen aussi direct de déchiffrer les desseins des autres. Pour autant, nous n'avons jamais vraiment accès aux tréfonds de notre être. Nous restons dans l'ignorance des vraies raisons, inconscientes, de nos comportements, et cela nous empêche de les prédire, dès que nous sortons de la routine des activités quotidiennes. La devise grecque : « Connais-toi toi-même », lorsqu'elle s'applique aux détails de nos actes, n'est qu'un idéal inaccessible. Notre moi n'est qu'une base de données imparfaites, formatée selon les mêmes règles que notre connaissance des autres, donc pétrie de zones d'ombre, de malentendus et d'aveuglement.

Ces limites de la condition humaine n'ont évidemment pas échappé aux romanciers. Dans son roman introspectif *Pensées secrètes,* l'écrivain anglais David Lodge invente un dialogue entre Ralph, scientifique matérialiste et spécialiste de l'intelligence artificielle, et Helen, une romancière professeur de littérature anglaise portée à la philosophie. On les surprend à cogiter sur la conscience de soi, légèrement ivres, tout en flirtant dans un jacuzzi[51] :

> « Je suppose qu'il doit y avoir un thermostat. Est-ce que ça signifie qu'un jacuzzi est conscient ?
> – Pas de soi, en tout cas. Il ne sait pas qu'il est en train de passer un bon moment – pas comme vous et moi.
> – Je croyais que le soi n'existait pas.
> – Ce n'est certainement pas une chose réelle et tangible, mais bien sûr qu'il existe un soi. Nous l'inventons sans cesse. Exactement

> comme vous inventez vos histoires.
> – Vous voulez dire que nos vies ne sont que des fictions ?
> – En quelque sorte. C'est quelque chose que nous faisons dès que nous avons du temps de cerveau disponible. Nous inventons des histoires sur nous-même. »

Nous leurrer nous-mêmes… tel est sans doute le prix que nous devons payer pour ce don qu'est la conscience humaine : la capacité de communiquer à d'autres, de façon rudimentaire, notre état d'esprit, avec juste assez d'introspection sur nos propres limites pour nous permettre de prendre, collectivement, des décisions avantageuses pour tous. Tout imparfaite qu'elle soit, c'est bien cette capacité humaine d'introspection et de partage social qui a engendré les alphabets, les cathédrales, les avions à réaction et le homard thermidor. Pour la première fois sans doute dans l'évolution, cette compétence nous a permis d'inventer des mondes fictifs. Nous avons appris à détourner l'algorithme de prise de décision sociale à notre avantage par la simulation, la contrefaçon, la flatterie, le mensonge, l'argumentation, la réfutation et la moquerie. Vladimir Nabokov, dans ses *Cours de littérature*[52] (1980), l'avait bien vu :

> « La littérature n'est pas née le jour où un jeune garçon criant "au loup, au loup !" a jailli d'une vallée néanderthalienne, un grand loup gris sur ses talons : la littérature est née le jour où un jeune garçon a crié "au loup, au loup !" alors qu'il n'y avait aucun loup derrière lui. »

La conscience est le simulateur de réalité virtuelle où se confrontent, dans un format commun, nos représentations mentales et celles des autres. Mais comment notre cerveau construit-il ce monde mental ?

CHAPITRE 4

Les signatures de la pensée consciente

L'imagerie cérébrale a conduit à une percée dans le domaine de la conscience. Elle a mis en évidence le trajet cérébral d'une information qui accède à la conscience ou qui, au contraire, demeure inconsciente. La comparaison de ces deux états a révélé ce que j'appelle des « signatures de la conscience » : des marqueurs fiables du fait qu'un stimulus a été consciemment perçu. Dans ce chapitre, nous ne découvrirons pas moins de quatre signatures de la conscience. En premier lieu, alors qu'un stimulus subliminal s'émousse à mesure qu'il se propage dans le cortex, l'activité cérébrale évoquée s'amplifie soudainement lors de l'accès à la conscience. Elle envahit de nombreuses régions supplémentaires et entraîne l'embrasement ou l'« ignition » de circuits pariétaux et frontaux (signature 1). L'électroencéphalogramme montre alors une onde tardive appelée P3 ou P300 (signature 2). Cet événement survient au moins un tiers de seconde après le stimulus : notre conscience est très en retard sur le monde extérieur. À l'aide d'électrodes placées dans le cortex, deux autres signatures ont été détectées : une explosion tardive et soudaine d'ondes de haute fréquence (signature 3) et une synchronisation des signaux que s'échangent les aires corticales les plus éloignées les unes des autres (signature 4). Tous ces événements constituent autant de marqueurs cérébraux d'une pensée consciente.

> « Et ainsi ce fut elle qui la première me donna l'idée qu'une personne n'est pas, comme j'avais cru, claire et immobile devant nous avec ses qualités, ses défauts, ses projets, ses intentions à notre égard […] mais est une ombre où nous ne pouvons jamais pénétrer, pour laquelle il n'existe pas de connaissance directe. »
>
> Marcel PROUST,
> *Le Côté de Guermantes* (1921).

La métaphore de Proust renouvelle un cliché qui remonte à l'Antiquité : notre esprit est comme une citadelle intérieure. Retiré derrière d'épaisses murailles, protégé du regard inquisiteur des autres, chacun de nous est libre de penser sans crainte d'être confondu. Notre conscience est un sanctuaire impénétrable où notre esprit peut vagabonder librement, tout en laissant croire à notre entourage que nous lui prêtons attention. « 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 », écrit Julian Jaynes. Comment la science pourrait-elle jamais conquérir ce bastion intérieur ?

Et pourtant… En l'espace de quelques décennies, le crâne est devenu transparent. Dès 1900, le chercheur japonais Seiji Ogawa et ses collègues ont inventé l'imagerie fonctionnelle par résonance magnétique (IRMf), une technique puissante et sans danger qui permet de visualiser l'activité de l'ensemble du cerveau[1]. L'IRM fonctionnelle repose sur le couplage qui relie les neurones aux vaisseaux sanguins. Dès qu'un circuit neuronal entre en activité, les cellules gliales qui entourent ces neurones détectent un surcroît d'activité synaptique. Pour contrebalancer rapidement cette dépense énergétique, elles déclenchent la dilatation des artères voisines. 2 ou 3 secondes plus tard, un sang riche en oxygène afflue dans les tissus. L'oxygène y est transporté par les molécules d'hémoglobine que contiennent les globules rouges du sang.

Le grand exploit de l'IRM fonctionnelle est de détecter, à distance, les propriétés de l'hémoglobine. Lorsqu'elle ne porte pas d'atomes d'oxygène, l'hémoglobine se comporte comme un petit aimant, alors qu'une fois oxygénée elle devient transparente au champ magnétique. Les machines d'IRM sont des aimants ultrapuissants dont les paramètres sont ajustés afin de détecter les petites fluctuations du champ magnétique en chaque point du cerveau. Ainsi, indirectement, en mesurant l'aimantation locale, on évalue l'oxygénation du sang qui elle-même reflète l'activité récente des réseaux neuronaux.

L'IRM fonctionnelle visualise aisément l'ensemble du cerveau, à l'échelle du millimètre, jusqu'à plusieurs fois par seconde. Cependant, elle ne permet pas de suivre l'activité neuronale en temps réel. Nous devons pour cela faire appel à d'autres techniques qui détectent les minuscules courants électriques au niveau des synapses – sans jamais ouvrir le crâne. L'électroencéphalographie (EEG), la bonne vieille mesure des ondes cérébrales inventée dès les années 1930, est devenue une technique raffinée, qui combine les signaux de 256 électrodes pour fournir un enregistrement digital de haute qualité de l'ensemble de l'activité électrique du cerveau. Dans les années 1960, une technologie nouvelle a vu le jour : la magnétoencéphalographie (MEG), l'enregistrement

ultraprécis des minuscules ondes magnétiques qu'émettent les mouvements des ions dans le cortex. L'EEG et la MEG ne sont pas compliquées à mettre en œuvre : il suffit de poser des électrodes sur la surface de la tête, ou bien de l'entourer d'une couronne de capteurs de champ magnétique.

Avec l'IRM, l'EEG et la MEG, il devient possible de suivre toute la séquence d'activations cérébrales qui sous-tend la vision, depuis la rétine jusqu'au cortex frontal. Combinés avec les méthodes de la psychologie cognitive, ces outils ouvrent des perspectives nouvelles sur le cerveau conscient. Comme nous l'avons vu au chapitre 1, une panoplie de stimuli permet de contraster le traitement conscient et non conscient. Nous savons rendre n'importe quelle image invisible en la masquant ou en détournant l'attention des participants. Nous pouvons même la présenter au seuil de conscience, en sorte qu'elle ne soit vue que la moitié du temps et que son degré de conscience subjective soit la seule chose qui varie. Dans les meilleures expériences, tout est identique : le stimulus, la tâche, le taux de réussite… Seule la conscience y est manipulée : tantôt le sujet dit voir l'image, tantôt il rapporte ne rien voir.

Dans ces conditions idéales, il ne reste plus qu'à aller voir dans le cerveau ! Quelle différence la conscience introduit-elle ? Certains circuits s'activent-ils uniquement lors des essais conscients ? La prise de conscience s'accompagne-t-elle d'événements qui lui sont propres, tels que des ondes ou des oscillations particulières ? De tels marqueurs, s'ils existent, mériteraient d'être appelés « signatures » de la conscience. Leur présence au sein de l'activité cérébrale, telle une signature au bas d'un chèque, certifierait l'identité de leur émetteur : un sujet conscient.

Grâce à l'imagerie cérébrale, le mystère des mécanismes neuronaux de la conscience commence à se dissiper. Dans ce chapitre, nous allons découvrir plusieurs signatures de la conscience.

L'avalanche consciente

En l'an 2000, la scientifique israélienne Kalanit Grill-Spector, qui travaillait alors à l'Institut Weizmann, a réalisé une expérience très simple[2]. Elle a présenté des images très brièvement, pendant une durée qui variait entre un cinquantième et un huitième de seconde. Chaque image était suivie d'un méli-mélo de photos. Ainsi masquées, certaines images demeuraient détectables, tandis que d'autres tombaient en dessous du seuil de perception consciente et devenaient invisibles. Les rapports subjectifs des participants traçaient une élégante fonction à seuil : en dessous de 50 millisecondes de présentation, la plupart des images étaient invisibles, tandis qu'au-delà de 100 millisecondes elles apparaissaient en pleine conscience.

Grâce à l'IRM fonctionnelle, Kalanit Grill-Spector effectua ensuite une série de mesures du cortex visuel. Elle découvrit une nette dissociation. Les aires visuelles précoces s'activaient tout le temps, que l'image soit consciente ou non. Toutes les images allumaient ainsi le cortex visuel primaire et les régions voisines, quel que soit le degré de masquage. Par contre, dans les aires visuelles supérieures telles que le gyrus fusiforme et la région occipito-temporale latérale, l'activité cérébrale était étroitement liée au degré de conscience. Ces régions sont impliquées dans la catégorisation et l'identification des images (visages, objets, mots, lieux, etc.). Il semblait que, dès qu'une image atteignait ce niveau de traitement cérébral, elle avait de fortes chances de devenir consciente.

Au même moment, mes collègues et moi menions des expériences similaires sur la perception des mots masqués[3]. Nous présentions des mots en dessous ou au-dessus du seuil de perception consciente, tandis que notre IRM produisait des images de l'activité du cerveau tout entier. Les résultats étaient clairs : même les aires visuelles de haut niveau situées dans le gyrus fusiforme pouvaient s'activer inconsciemment. Nos mesures montraient que même une image subliminale subissait un traitement visuel approfondi qui impliquait certaines aires du cortex

pariétal et temporal. Le cerveau pouvait par exemple reconnaître que *piano* et *PIANO* représentent le même mot, ou que le chiffre *3* et le mot *trois* évoquent la même quantité, sans pour autant que le sujet en prenne conscience[4].

Néanmoins, nos images étaient en accord avec celles de Kalanit Grill-Spector sur un point : lorsqu'une image franchissait le seuil de perception consciente, l'activité des aires visuelles supérieures montrait un changement massif : elle s'amplifiait soudainement. Dans l'aire de la forme visuelle des mots, que j'ai appelée la « boîte aux lettres du cerveau[5] », elle était multipliée par douze ! De plus, tout un jeu de régions supplémentaires, qui ne répondait pas aux mots subliminaux, s'activait dès que la conscience était présente. Ces régions appartenaient aux secteurs latéraux des lobes pariétaux et frontaux, ainsi qu'au gyrus cingulaire antérieur situé sur la ligne médiane qui sépare les deux hémisphères (figure 16).

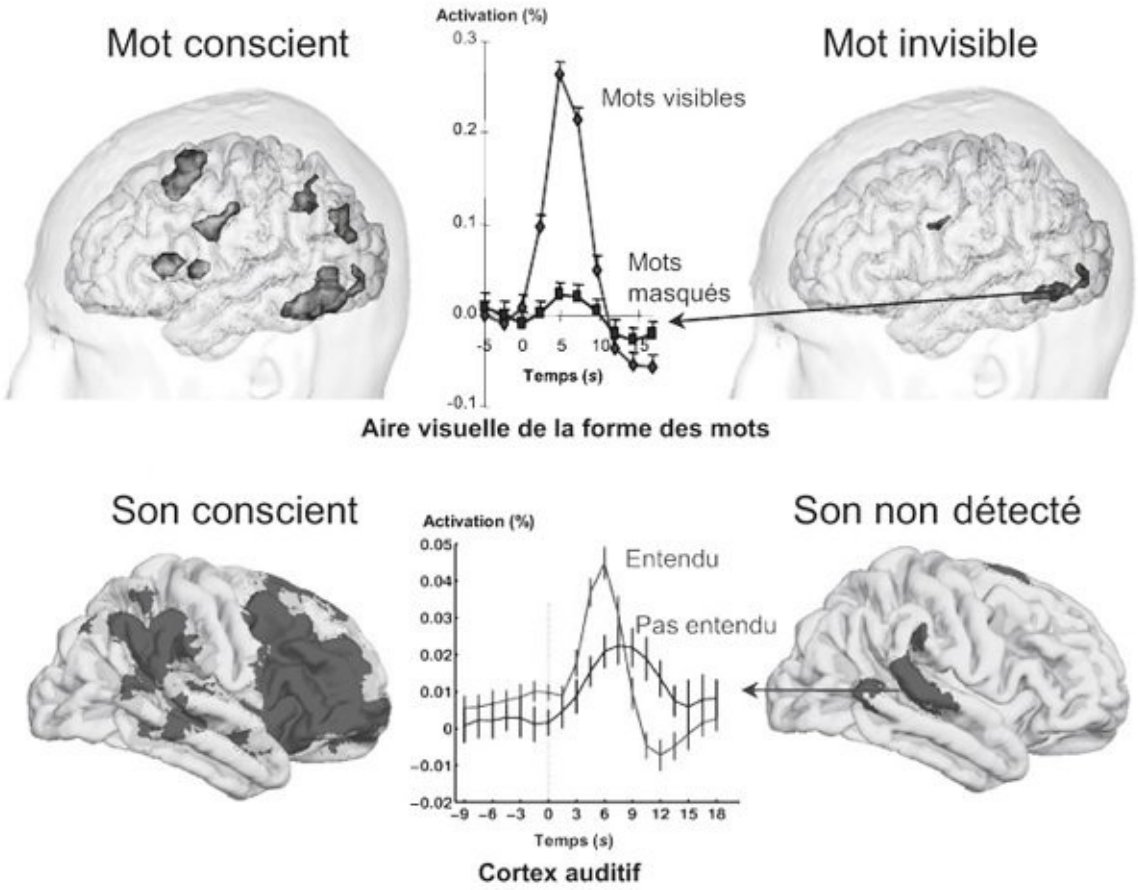


Figure 16. La première signature de l'accès à la conscience est un intense embrasement de nombreuses régions du cortex, particulièrement les aires pariétales et préfrontales bilatérales. Un mot rendu subliminal par masquage (en haut) active certains circuits spécialisés dans la lecture, mais le même mot, s'il est vu, entraîne une énorme amplification de cette activité, qui envahit alors les lobes pariétaux et préfrontaux. De même, les aires auditives peuvent être activées par un son non conscient (en bas), mais le même son, une fois détecté, envahit de vastes zones du cortex préfrontal et pariétal inférieur.

En mesurant l'amplitude de l'activité évoquée, nous avons découvert que le facteur d'amplification, qui distingue le traitement conscient du traitement non conscient d'un mot, varie d'une aire visuelle à l'autre. À son point d'entrée dans le cortex, dans l'aire visuelle primaire, un mot invisible entraîne une activation vigoureuse et facile à détecter. Cependant, à mesure qu'il progresse dans le cortex, le masquage lui fait

perdre de sa force[6]. La perception subliminale ressemble ainsi à une vague qui, au large, impressionne par son ampleur, mais déçoit ensuite lorsqu'elle meurt sur le rivage en vous léchant à peine les pieds. Selon cette métaphore, la perception consciente serait un tsunami, une onde impossible à arrêter qui s'amplifie et laisse des traces irréversibles loin dans les terres. Je préfère encore la comparaison avec une avalanche, parce que l'activation cérébrale consciente semble gagner en force à mesure qu'elle progresse, exactement comme une toute petite boule de neige, en roulant, s'agrandit jusqu'à causer un glissement de terrain.

Pour clarifier cette idée, je voudrais vous rappeler que, dans notre expérience, nous présentions chaque mot pendant une durée très brève : 43 millisecondes seulement. Un tel stimulus n'injectait donc qu'une quantité très limitée d'activité dans la rétine. Et pourtant, l'activation ne cessait de progresser dans le cerveau et, dans les essais conscients, elle s'amplifiait jusqu'à entraîner une énorme vague d'activité dans des régions très distantes du cortex visuel. Leur activation devenait également corrélée : la vague entrait et sortait simultanément dans toutes ces régions, comme si elles échangeaient des messages qui se renforçaient les uns les autres jusqu'à former une avalanche impossible à stopper. La synchronisation était nettement plus grande dans les essais conscients que dans les essais non conscients, ce qui suggérait déjà que la corrélation de l'activité cérébrale entre plusieurs régions est un facteur important dans la prise de conscience[7].

Ces expériences simples ont donc révélé une première signature de la conscience : l'amplification de l'activité des aires sensorielles, qui gagne en force et envahit de nombreuses régions des lobes pariétaux et frontaux. Cette signature caractéristique a souvent été retrouvée par la suite, y compris dans d'autres modalités sensorielles. Imaginez que vous soyez allongé au beau milieu d'une IRM, dans un bruit considérable. De temps à autre, par le biais d'écouteurs, un son vous est présenté. Sans que vous le sachiez, son niveau est finement ajusté afin que vous ne parveniez à le détecter que la moitié du temps. Voilà une nouvelle manière de contraster

la perception consciente et non consciente, cette fois dans la modalité auditive. Et le résultat est toujours aussi net : les sons non conscients activent seulement les régions situées autour du cortex auditif primaire, tandis que lors des essais conscients, une avalanche d'activité se déclenche dans les mêmes régions et gagne les régions préfrontales et pariétales inférieures[8] (figure 16).

Troisième exemple : l'action motrice. Supposez qu'on vous demande de bouger le doigt dès que voyez une cible… sauf si celle-ci est précédée d'un signe qui signifie « stop », auquel cas vous devez absolument éviter tout mouvement[9]. Cette tâche évalue votre faculté d'inhibition : vous devez mobiliser vos capacités de contrôle conscient afin d'inhiber la tendance naturelle à répondre dès que la cible apparaît. Maintenant, imaginez que le signe « stop » vous soit présenté en dessous du seuil de perception consciente. Parviendrez-vous à suivre un ordre subliminal dont vous n'avez pas conscience ? Cela peut paraître paradoxal, mais le cerveau se montre à la hauteur de ce défi. Même dans les essais subliminaux, les réponses des participants se ralentissent légèrement dès que le signe « stop » apparaît, ce qui indique que l'inhibition peut se déployer en l'absence de conscience (comme nous l'avons vu au chapitre 2). L'imagerie cérébrale montre que l'inhibition inconsciente repose sur deux régions spécialisées dans le contrôle de la motricité (l'aire motrice présupplémentaire et l'insula antérieure). Ce qui est le plus remarquable, toutefois, c'est qu'une fois de plus la perception consciente induit un changement massif : dès que le signe « stop » cesse d'être invisible, l'activation est multipliée par deux dans ces régions, et elle envahit également un vaste réseau d'aires pariétales et préfrontales (figure 17). À présent, ce circuit doit vous être familier : son activité soudaine constitue une signature systématique de l'accès à la conscience[10].

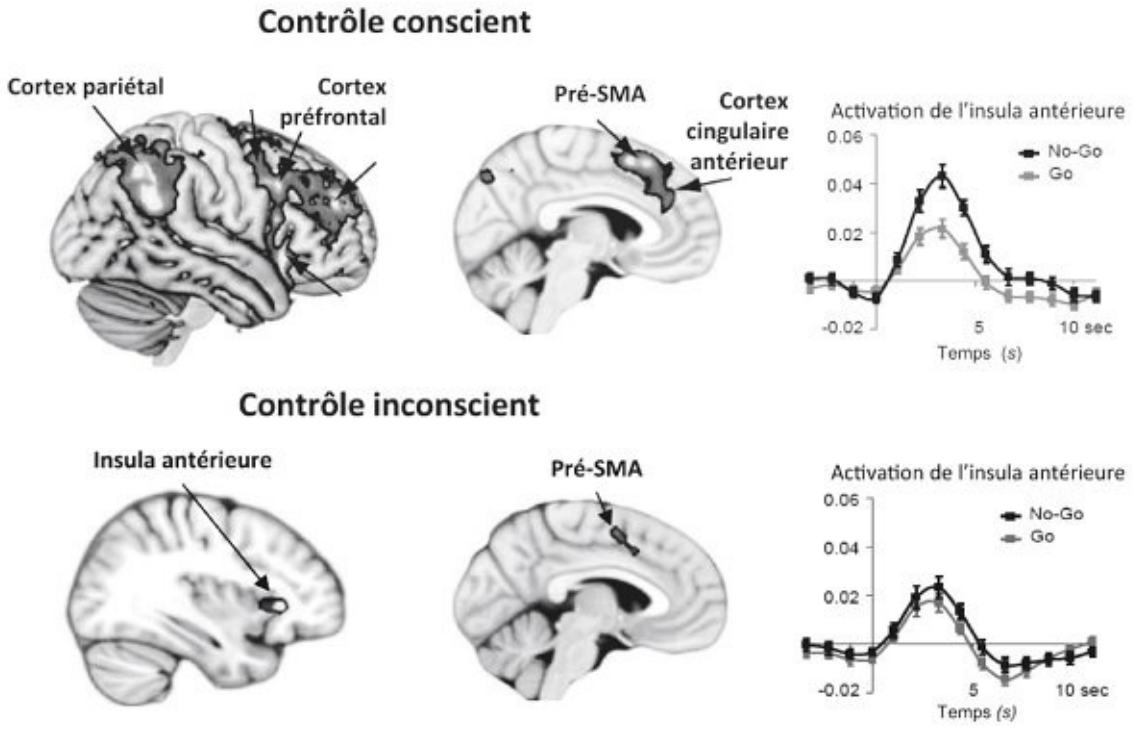


Figure 17. Le contrôle conscient ou non conscient de l'action repose sur des circuits cérébraux en partie distincts. Un signal « stop », même invisible, active quelques circuits spécialisés de l'insula antérieure et de l'aire motrice présupplémentaire (pré-SMA) qui supervisent nos gestes (colonne de droite). Le même signal, dès qu'il devient visible, active un nombre bien plus grand de régions pariétales et préfrontales associées au contrôle volontaire de l'action.

Chronométrer l'accès à la conscience

Bien que l'IRM fonctionnelle soit un merveilleux outil de localisation de l'activité cérébrale, elle ne répond qu'à une partie de la question : elle nous dit *où*, mais elle ne nous dit pas *quand*. Elle ne peut pas déterminer à quelle vitesse ni dans quel ordre les différentes aires cérébrales s'allument lorsque nous prenons conscience d'un stimulus. Pour chronométrer l'avalanche consciente, nous devons faire appel aux méthodes plus raffinées de l'électro-et de la magnétoencéphalographie (EEG et MEG). Quelques électrodes collées sur la peau, ou quelques capteurs

magnétiques disposés autour de la tête, suffisent à suivre l'activité cérébrale à la milliseconde près.

En 1995, ma thésarde Claire Sergent et moi avons conçu une expérience d'EEG qui a révélé, pour la première fois, l'enchaînement des étapes qui conduisent à la conscience[11]. Nous avons suivi le trajet cortical d'images rigoureusement identiques qui étaient tantôt invisibles, tantôt perçues consciemment (figure 18). Pour ce faire, nous avons utilisé le phénomène du clignement attentionnel – le fait que, lorsque nous sommes distraits, nous ne parvenons pas à voir les stimuli qui sont pourtant sous nos yeux. Nous avons demandé à des volontaires de détecter des mots, tout en les distrayant brièvement en précédant chaque mot d'une autre série de lettres qu'ils devaient mémoriser. Enregistrer ces lettres en mémoire leur demandait de se concentrer pendant une fraction de seconde, et cela suffisait à leur faire régulièrement rater le mot. Pour savoir précisément quand survenaient ces défaillances de l'accès à la conscience, nous leur demandions, après chaque essai, de rapporter ce qu'ils avaient vu à l'aide d'un curseur. Ils pouvaient le déplacer continûment afin d'indiquer qu'ils n'avaient rien vu du tout, quelques lettres, une part importante du mot, le mot tout entier… ou n'importe quel autre état intermédiaire.

Sergent et moi avons ajusté tous les paramètres jusqu'à ce que les mêmes mots soient tantôt vus, tantôt pas vus. Une fois ce réglage effectué, chaque participant disait avoir parfaitement vu le mot dans une moitié des essais et n'avoir rien vu du tout dans l'autre. Leur introspection variait donc en tout-ou-rien : soient ils percevaient le mot entier, soit ils le rataient entièrement. Il était rare qu'ils disent l'avoir vu partiellement[12]. C'est un point fondamental : l'accès à la conscience fonctionne en tout-ou-rien et bascule dans l'état « on » ou dans l'état « off ».

Au même moment, nos enregistrements d'EEG montraient que l'activité cérébrale elle aussi changeait brutalement d'état. Initialement, au sein du système visuel, les mots visibles et invisibles ne montraient aucune différence. Tous deux, comme n'importe quel stimulus visuel, évoquaient

dans le cortex occipital des ondes cérébrales appelées P100 et N170 (cette terminologie indique que la première est positive et survient environ 100 millisecondes après le mot, tandis que la seconde est négative et atteint son pic vers 170 millisecondes). Les deux ondes sont le reflet de la progression des informations visuelles à travers une hiérarchie de régions impliquées dans la reconnaissance des mots – une progression qui semble totalement indépendante de la conscience. L'activité était tout aussi intense, que le mot soit vu ou qu'il demeure invisible. De toute évidence, le mot entrait dans le cortex visuel avec force, même quand son possesseur, une fraction de seconde plus tard, affirmait qu'il n'y avait rien à voir !

Quelques centièmes de seconde plus tard, toutefois, cette organisation était bouleversée. Soudainement, entre 200 et 300 millisecondes après l'apparition du mot sur l'écran, l'activité cérébrale diminuait pour les essais inconscients alors que, lors des essais conscients, elle continuait sa progression vers l'avant du cerveau. Vers 400 millisecondes, cette différence était devenue gigantesque : seuls les mots conscients entraînaient une intense activation des lobes frontaux, du cortex cingulaire antérieur et de certaines régions pariétales. Au bout d'une demi-seconde, l'activité retournait vers l'arrière du cerveau, y compris le cortex visuel primaire. De nombreux autres chercheurs ont observé cette onde rétrograde, mais personne ne sait vraiment ce qu'elle signifie – peut-être une trace de mémoire durable de la représentation visuelle devenue consciente[13].

Puisque notre stimulus était *exactement* le même dans les essais visibles et invisibles, la brutalité de la transition entre le non-conscient et le conscient était étonnante. En moins d'un dixième de seconde, entre 200 et 300 millisecondes, le cerveau passait d'une absence de différence à un énorme effet tout-ou-rien. On aurait dit que tous les mots débutaient avec la même quantité d'activation visuelle, mais que cette vague, uniquement lors des essais conscients, prenait de l'ampleur et franchissait soudain la digue des réseaux frontaux et pariétaux, en inondant d'un coup les vastes

champs des cortex supérieurs. Lors des essais invisibles, par contre, la même vague restait confinée au sein des aires postérieures du cerveau, épargnant les systèmes associés aux opérations conscientes et, donc, laissant la personne dans l'ignorance totale de ce qui s'était réellement passé.

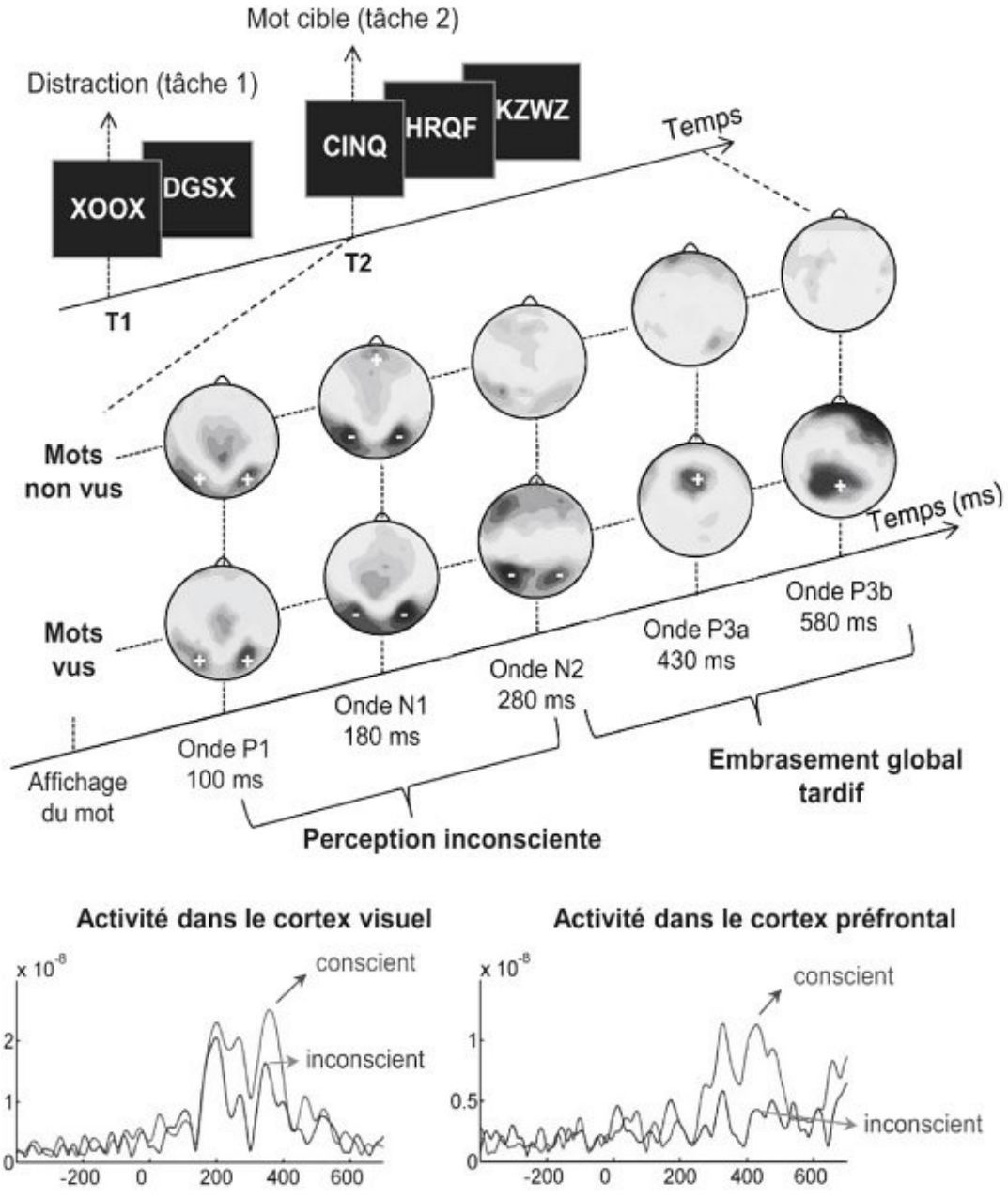


Figure 18. Une intense onde positive, l'onde P3b, fournit une deuxième signature de la perception consciente. Dans l'expérience illustrée ici, on présente des mots pendant le « clignement attentionnel », alors que le sujet est distrait par une autre tâche. Environ 50 % des mots sont ainsi rendus invisibles. L'enregistrement de l'électroencéphalogramme permet de suivre le trajet cérébral d'un même mot selon qu'il est vu ou pas vu. La divergence survient soudainement vers 300 millisecondes : seuls les mots consciemment perçus sont amplifiés et entraînent l'embrasement du cortex préfrontal et d'autres aires corticales associatives, ainsi

qu'un retour vers les aires visuelles. Cet embrasement global se traduit par une mesure de voltage positive au sommet et à l'arrière de la tête – l'onde P3b.

L'embrasement du cortex caractérisait donc les essais conscients. Cependant, lors des essais non conscients, l'activité cérébrale ne s'évanouissait pas immédiatement. Pendant près d'une demi-seconde, des ondes inconscientes parcouraient le lobe temporal gauche, dans des régions associées à la représentation du sens des mots. Au chapitre 2, nous avons vu que, durant le clignement attentionnel, même des mots invisibles continuent d'accéder au sens[14]. Ce processus d'interprétation inconsciente reste toutefois confiné au lobe temporal. Seul un débordement de l'activité dans les lobes pariétaux et frontaux signale une perception consciente.

L'avalanche de la conscience se traduit par une signature que l'on recueille aisément sur les électrodes situées au sommet du crâne. Lors de la prise de conscience d'une information nouvelle, une onde de grande amplitude balaie cette région. Elle démarre vers 270 millisecondes et atteint son pic entre 350 et 500 millisecondes après l'arrivée du mot. Cet événement lent et massif s'appelle l'onde P3[15] (parce que c'est le troisième pic positif majeur qui suit le stimulus) ou encore l'onde P300 (parce qu'elle démarre vers 300 millisecondes). Elle ne fait que quelques microvolts, un million de fois moins qu'une pile AA, mais c'est déjà beaucoup pour le cerveau, et sa mesure ne pose aucune difficulté aux amplificateurs modernes. L'onde P3 constitue la deuxième signature de la conscience. Une grande variété d'expériences a montré qu'elle surgit dès qu'une information accède à la conscience[16].

En examinant nos enregistrements de plus près, nous avons découvert que l'évolution de l'onde P3 pouvait expliquer *pourquoi* nos participants rataient si souvent le mot cible. Dans notre expérience, il y avait en fait *deux* ondes P3 : l'une était évoquée par la série de lettres à mémoriser, qui servait de distraction et qui était toujours perçue consciemment ; l'autre

était suscitée par le mot, mais seulement lorsqu'il était vu. Ce qui était fascinant, c'est qu'il existait un compromis entre ces deux phénomènes. Lorsque la première onde P3 était ample et durable, la seconde avait bien plus de chances de disparaître complètement – et c'était précisément lors de ces essais que le sujet disait n'avoir rien vu. Autrement dit, la conscience opérait comme un va-et-vient : tant que le cerveau était occupé par les lettres, comme le montrait l'allongement de la première onde P3, il ne pouvait en même temps prêter attention au mot. L'évaluation de l'un semblait bloquer la conscience de l'autre.

René Descartes aurait été ravi de ces résultats, lui qui écrivait dans son traité de *L'Homme* que « les idées s'empêchent l'une l'autre, (…) d'où vient qu'on ne peut être fort attentif à plusieurs choses en même temps ». Descartes avait vu juste : notre cerveau conscient ne peut pas connaître deux embrasements en même temps, il ne nous laisse percevoir qu'un seul aspect de réalité à la fois. Dès lors que les régions pariétales et préfrontales se synchronisent pour traiter un premier stimulus, elles ne peuvent pas se réengager immédiatement en faveur d'un second. Le simple fait de se concentrer sur un objet de pensée nous empêche d'en percevoir un second[17]. C'est la « période réfractaire » que nous avons rencontrée au chapitre 1 : avant qu'un objet ne pénètre notre conscience, il doit d'abord attendre que celle-ci se libère.

La conscience en retard sur le monde

Une conséquence importante de ces découvertes est que notre conscience est en retard sur les événements. Non seulement nous ne percevons qu'une toute petite fraction des signaux qui bombardent nos sens, mais quand nous y parvenons, c'est avec un délai important. À ce titre, notre cerveau ressemble à un astronome qui recueille la lumière des étoiles : parce que la vitesse de la lumière n'est pas infinie, la nouvelle de

l'explosion d'une supernova ne lui parvient que des millions d'années après qu'elle a eu lieu. De même, parce que notre cerveau accumule les données avec lenteur, ce que nous jugeons comme le « temps présent » de la conscience reste à la traîne de la réalité physique. La prétendue « vitesse de la pensée » n'a rien d'extraordinaire : notre cerveau est tellement plus lent qu'un ordinateur qu'il lui faut au moins un tiers de seconde pour réagir consciemment. Bien souvent, cette durée s'allonge lorsque l'entrée sensorielle est si faible que le cerveau doit accumuler de nombreuses données avant de franchir le seuil de la prise de conscience[18] (cette situation est semblable à celle de l'astronome qui prolonge l'exposition pendant plusieurs minutes afin de photographier la lumière des étoiles les plus faibles). Comme nous venons de le voir, la prise de conscience se ralentit parfois bien plus encore lorsque notre esprit est occupé ailleurs. C'est pourquoi il ne faut pas téléphoner en conduisant – même un réflexe aussi élémentaire que d'appuyer sur la pédale de frein lorsque vous voyez s'allumer les feux de la voiture devant vous est ralenti lorsque vous êtes distrait[19].

Nous sommes tous aveugles aux limites de notre attention et n'avons donc aucune conscience que notre perception subjective est en retard sur le monde extérieur. La plupart du temps, d'ailleurs, ce délai est sans conséquence. Nous pouvons parfaitement apprécier un coucher de soleil ou un concert symphonique sans nous rendre compte que les couleurs que nous voyons ou la musique que nous entendons remontent déjà à une fraction de seconde. Lorsque nous nous contentons d'écouter, quelle importance de savoir exactement quand les sons qui nous parviennent ont été émis ? Même lorsque nous devons agir, le monde change souvent suffisamment lentement pour que nos gestes ne soient pas complètement hors de propos. C'est seulement lorsque nous devons agir « en temps réel » que nous prenons conscience de l'extraordinaire lenteur de notre pensée. N'importe quel pianiste qui interprète un *allegro* sait bien qu'il serait vain de tenter de contrôler chacun de ses doigts véloces qui volent sur le clavier – le contrôle conscient, avec ses gros sabots nonchalants, est

bien trop lent pour entrer dans cette danse. Pour vous rendre compte du retard de votre conscience, essayez de photographier un événement bref et imprévisible, par exemple un lézard qui sort sa langue : au moment où votre doigt appuie sur le déclencheur, le moment que vous souhaitiez immortaliser est passé depuis longtemps.

Fort heureusement, notre cerveau contient également des mécanismes sophistiqués qui compensent les délais de notre conscience. Tout d'abord, nous fonctionnons souvent en « pilote automatique », sur un mode inconscient. Comme René Descartes l'avait noté, un doigt qui se brûle se retire à la vitesse de l'éclair, bien avant que nous ne prenions conscience de la douleur. Nos yeux et nos mains réagissent souvent à temps, parce qu'ils sont guidés par des boucles neuronales qui échappent à la conscience. Ces circuits moteurs peuvent certainement être *modulés* par nos intentions conscientes, comme lorsque nous approchons lentement le doigt de la flamme d'une bougie. Ensuite, pourtant, l'action elle-même se déroule souvent inconsciemment. Ainsi nos doigts s'ajustent-ils presque instantanément à un changement de position de leur cible, bien avant que nous ne détections consciemment que quelque chose a changé[20].

L'anticipation est le second mécanisme qui nous permet d'atténuer les conséquences de notre propre lenteur. Pratiquement tous nos systèmes sensoriels et moteurs abritent des mécanismes d'apprentissage qui anticipent sur le monde extérieur et nous donnent à percevoir les événements à un moment plus proche de la réalité. Évidemment, la conséquence néfaste de cet état de choses, c'est que lorsqu'un événement inattendu survient, nous nous méprenons sur le moment où il apparaît. Ainsi, par rapport à un point qui bouge régulièrement, un flash lumineux inattendu semble en retard par rapport à sa position réelle[21]. Cette illusion temporelle, appelée *flash lag illusion* en anglais, montre que nous percevons plus vite un stimulus quand il est prévisible que quand il est imprévisible. Elle témoigne des longs chemins tortueux qu'emprunte l'information avant d'atteindre la citadelle de notre conscience.

C’est seulement lorsque nos attentes sont mises en échec que nous prenons vivement conscience de l’énorme délai avec lequel nous réagissons consciemment. Bousculez accidentellement un verre de lait, et vous aurez une conscience aiguë de ce phénomène : pendant un instant, votre conscience tente désespérément de reprendre le contrôle sur vos mains, mais elle tâtonne dans le vide, en constant décalage avec les événements en cours. Simple spectateur de cet accident, vous ne pouvez que constater, avec retard, ses déplorables atermoiements.

La perception d’une erreur motrice se produit en deux temps, exactement comme celle d’un mot ou d’une image : une prompte évaluation inconsciente suivie d’un lent embrasement conscient. Votre corps agit d’abord par réflexe, et la conscience survient bien plus tard. Supposez que l’on vous demande de bouger les yeux du côté opposé à une cible : si une lumière apparaît à droite, vous devez regarder à gauche, et *vice versa*. Cette instruction engendre un grand nombre d’erreurs : très souvent, votre regard est d’abord attiré du côté de la cible, et ce n’est que plus tard qu’il se détourne, sous le contrôle de la conscience. Ce qui est fascinant, c’est que l’erreur demeure souvent inconsciente. Vous avez le sentiment que vos yeux se sont directement éloignés de la lumière, alors que ce n’est pas le cas. L’électroencéphalographie permet de mesurer la réponse du cerveau aux erreurs conscientes et inconscientes[22]. Pendant un cinquième de seconde, le cortex réagit pratiquement à l’identique : un pilote automatique, situé dans le gyrus cingulaire, détecte que le plan d’action ne se déroule pas selon les instructions, et il décharge violemment pour signaler cette erreur – même lorsque celle-ci demeure inconsciente[23]. Comme d’autres réponses sensorielles, cette décharge initiale reflète un traitement purement inconscient et indétectable. Ce n’est que si la réponse cérébrale se prolonge, jusqu’à devenir intense et positive au sommet de la tête, que la personne prend conscience de son erreur. L’onde cérébrale correspondante s’appelle « positivité liée à l’erreur » (PE). Elle est pratiquement identique à l’onde P3 qui accompagne la perception des événements sensoriels. Actions et

sensations semblent ainsi être perçues de façon très semblable. Une fois de plus, l'onde P3 signale l'accès à la conscience – et cette signature survient toujours tardivement, bien après l'événement qui l'a déclenchée[24].

Un pur moment de conscience

Les chercheurs sont des sceptiques professionnels et ils n'ont pas manqué de mettre en doute l'hypothèse que l'onde P3 signale l'accès à la conscience. L'embrasement des réseaux préfrontaux et pariétaux, accompagné de son onde P3, ne pourrait-il pas s'expliquer autrement ? Au cours de la dernière décennie, de nombreuses expériences ont été conduites afin d'éliminer, une à une, toutes les autres possibilités. Bien que le débat reste ouvert, certaines expériences ingénieuses sont parvenues à isoler l'accès conscient des autres événements sensoriels et moteurs, afin de « distiller » un pur moment de conscience[25]. Voyons comment elles s'y sont prises.

L'une des difficultés est que la conscience est riche de conséquences. Dès que nous prenons conscience d'un fait, des myriades de possibilités s'ouvrent à nous. Nous pouvons le rapporter à d'autres, en mots ou en gestes. Nous pouvons le stocker en mémoire et le rappeler ensuite. Nous savons l'évaluer et l'utiliser pour guider nos actions. Tous ces processus deviennent disponibles au moment même où l'information accède à notre conscience – en sorte que nous risquons de les confondre avec l'accès conscient lui-même. L'activité cérébrale que nous mesurons dans les essais conscients est-elle vraiment spécifique à la prise de conscience ou bien reflète-t-elle seulement les conséquences de celle-ci ?

Pour résoudre ce dilemme, mes collègues et moi avons tenté d'apparier, autant que faire se peut, les essais conscients et non conscients. C'est à dessein que, dans nos expériences, les participants devaient agir de

façon presque identique dans les deux cas. Dans notre étude du clignement attentionnel, par exemple, ils devaient toujours mémoriser les lettres, puis décider s'ils avaient vu un mot ou non[26]. Il est raisonnable de penser que ces décisions sont d'égale difficulté : décider que l'on n'a *pas* vu un mot est probablement au moins aussi difficile que de décider qu'on l'a vu. De plus, les participants répondaient « vu » ou « pas vu » à l'aide des mêmes mouvements d'un curseur. Aucun de ces facteurs ne semble donc en mesure d'expliquer pourquoi l'onde P3 et l'activité pariétale et préfrontale sont tellement plus intenses dans les essais vus que dans ceux demeurés invisibles.

Le diable est dans les détails, cependant, et l'avocat du diable pourrait affirmer que la vue soudaine d'un mot évoque une série d'opérations mentales à un moment bien précis, tandis que le fait de « ne pas voir » est dilué dans le temps. En effet, l'absence de conscience n'a pas de point de départ bien défini : ce n'est qu'à la fin d'un essai que l'on peut conclure avec certitude que l'on n'a rien vu. Cette dilution temporelle pourrait-elle expliquer les différences d'activité cérébrale ?

Hakwan Lau et Richard Passingham ont très astucieusement rejeté cette possibilité[27]. Ils ont exploité le phénomène de la « vision aveugle ». Comme nous l'avons vu au chapitre 2, une image subliminale, affichée pendant une durée très brève, bien qu'elle soit invisible, peut induire des activations corticales qui atteignent le cortex moteur. Ainsi, une personne peut très bien répondre à une cible tout en niant l'avoir vue – d'où le terme de « vision aveugle ». Lau et Passingham ont utilisé cet effet pour égaliser les performances motrices dans les essais conscients et non conscients : leurs participants faisaient *exactement* les mêmes réponses dans les deux cas. Même avec ce contrôle très subtil, le surcroît de conscience était associé à une activation plus intense du cortex préfrontal gauche. Ces résultats ont été obtenus chez des participants normaux, mais le patient GY, atteint de « vision aveugle » à la suite d'une lésion de l'aire visuelle primaire, a montré des effets similaires. Chez lui, c'est tout le

réseau habituel, pariétal et préfrontal, qui s'active sélectivement lors des essais conscients[28].

Merveilleux, dit l'avocat du diable, toujours sceptique ; vous avez égalisé les réponses, mais à présent ce sont les stimuli qui diffèrent dans les essais conscients et non conscients. Pourriez-vous égaliser aussi bien les stimuli que les réponses, afin que *tout* soit identique – sauf, bien sûr, le sentiment subjectif de conscience ? Ce n'est qu'à ce prix que je serai vraiment convaincu que vous détenez une signature de la conscience.

Ce défi paraît impossible à relever… Et pourtant, durant sa thèse, le psychologue israélien Moti Salti, sous la direction de Dominique Lamy, y est parvenu et a confirmé que l'onde P3 est bien une signature de l'accès conscient[29]. Son stratagème a consisté à combiner la « vision aveugle » et le tri des essais en fonction de leur visibilité subjective. À chaque essai, un stimulus apparaissait brièvement à l'une des quatre positions possibles, et chaque participant devait immédiatement répondre à deux questions : 1) Où était situé le stimulus ? 2) L'avez-vous vu, ou bien avez-vous répondu au hasard ? Les réponses permettaient de trier les essais. Lors des essais « conscients et corrects », la personne disait avoir vu le stimulus et avait bien identifié sa position. Cependant, à cause du phénomène de vision aveugle, il y avait également un grand nombre d'essais « non conscients et corrects » où la personne niait avoir vu quoi que ce soit et pourtant avait répondu correctement à la première question.

Voilà donc la comparaison idéale : stimulus identique, réponse identique, mais conscience différente. Les enregistrements EEG montrèrent que toute la partie initiale de l'activation, jusqu'à 250 millisecondes après l'apparition de la cible, était strictement identique. Les deux types d'essais ne différaient que sur un point : l'onde P3, à partir de 270 millisecondes, croissait jusqu'à devenir nettement plus grande dans les essais conscients que dans les essais non conscients. Non seulement son amplitude, mais également sa topographie étaient bien distinctes : tandis que les stimuli invisibles évoquaient une petite onde positive au niveau du cortex pariétal, ce qui reflétait probablement la

chaîne d'opérations « aveugles » qui permettaient de répondre avec justesse, seule la perception consciente conduisait à une expansion de cette activation dans les lobes frontaux gauche et droit.

Se faisant lui-même l'avocat du diable, Moti Salti s'est demandé si ses résultats non conscients pouvaient s'expliquer par un mélange de deux types d'essais : certains sans onde P3, où le sujet répondait au hasard, et d'autres avec une onde P3 de taille normale. Ses analyses éliminèrent cette possibilité. La petite onde P3 qui survenait lors des essais non conscients était trop petite, trop courte et trop postérieure pour correspondre à celle observée lors des essais conscients. Elle indiquait seulement que, lors des essais où le sujet répondait « pas vu », l'activité cérébrale commençait par s'amplifier, mais faisait ensuite long feu sans déclencher une onde P3 globale. Seule une P3 massive, qui se déploie dans le cortex préfrontal des deux hémisphères, constitue la signature authentique d'une perception consciente.

L'embrasement de la conscience

Lorsque nous prenons conscience d'une information inattendue, l'activité du cerveau s'embrase soudainement dans de nombreuses régions synchronisées du cerveau. Mes collègues et moi-même avons nommé cette propriété l'« embrasement global[30] » (*global ignition* en anglais). Le terme d'embrasement nous a été soufflé par la lecture du neurophysiologiste canadien Donald Hebb qui, en 1949, a été le premier à analyser le comportement collectif de grands ensembles de neurones[31]. Hebb expliquait, en des termes très intuitifs, comment un réseau de neurones qui s'excitent l'un l'autre peut vite s'embraser et devenir un vif foyer d'activité globale. Imaginez la fin d'un concert : les premiers applaudissements fusent et, en quelques dizaines de secondes, toute la salle se retrouve debout à battre des mains. Quelques spectateurs qui se

lèvent pour applaudir à tout rompre suffisent souvent à communiquer leur enthousiasme à toute une salle. De même, les grands neurones pyramidaux des couches supérieures du cortex propagent l'excitation à une vaste audience de neurones récepteurs. L'embrasement global, selon moi, survient lorsque cette excitation va *crescendo* jusqu'à excéder un seuil au-delà duquel elle se renforce elle-même. Certains neurones en excitent d'autres qui, à leur tour, leur rendent la pareille[32]. Résultat : une explosion soudaine d'activité, dans laquelle tous les neurones qui sont suffisamment interconnectés s'autoactivent mutuellement pour former une « assemblée réverbérante », selon le terme de Hebb.

Un tel phénomène collectif n'est pas rare : il constitue ce que les physiciens appellent une « transition de phase », ou les mathématiciens une « bifurcation » : un changement d'état brutal au point d'apparaître pratiquement discontinu. La glace qui gèle en fournit une illustration parfaite. Dès le tout début de nos réflexions sur la conscience, mes collègues et moi avions remarqué que le concept de transition de phase capturait bon nombre des propriétés de la perception consciente[33]. Comme l'eau qui gèle, la conscience possède un seuil : un stimulus bref va rester subliminal, tandis qu'un autre à peine plus long sera pleinement visible. La plupart des systèmes physiques autoamplifiés possèdent un point de non-retour au-delà duquel le changement survient d'un seul coup, en fonction de la présence de petites impuretés ou de fluctuations aléatoires. Selon nous, il en va de même pour le cerveau.

Comment pourrions-nous prouver qu'un message conscient entraîne une transition de phase à l'échelle de tout le cerveau, qui « gèle » transitoirement les aires cérébrales dans un état cohérent d'activité ? Avec mon étudiant Antoine Del Cul, nous avons conçu une expérience simple[34]. Nous avons fait varier un seul paramètre physique, exactement comme un physicien ferait décroître progressivement la température d'un liquide. Nous avons ensuite mesuré si la perception subjective, ainsi que les marqueurs objectifs de l'activité cérébrale, se comportaient de manière

discontinue et changeaient brutalement, comme s'ils subissaient une transition de phase.

Notre expérience consistait à afficher un chiffre pendant une durée très brève (16 millisecondes), puis un blanc, et enfin un masque formé de lettres aléatoires. Nous faisions varier la durée du blanc par paliers de 16 millisecondes. Que rapportaient les observateurs ? Leur perception traçait une courbe en tout-ou-rien. Lorsque le délai était long, ils voyaient parfaitement le chiffre, mais celui-ci disparaissait soudainement lorsque l'on raccourcissait l'intervalle de blanc. La perception subjective n'était pas linéaire : à mesure que la durée augmentait, les participants ne voyaient pas simplement un peu mieux, mais leur visibilité s'améliorait d'un seul coup. Les deux états, « vu » et « pas vu », étaient séparés par un seuil, qui se situait aux environs de 50 millisecondes de délai entre le chiffre et le masque[35].

Ce résultat en main, nous nous sommes tournés vers l'EEG afin d'étudier quels événements cérébraux montraient également une réponse en tout-ou-rien. La réponse ne nous a pas surpris : c'était notre fameuse onde P3. Tous les autres événements précédents soit ne variaient pas avec le délai de masquage, soit le faisaient d'une manière continue qui ne ressemblait guère à la perception subjective de nos observateurs.

La réponse initiale du cortex, par exemple, qui se traduisait par des ondes P1 et N1, demeurait pratiquement inchangée, quel que soit le degré de masquage du chiffre. Cela n'avait rien de surprenant : après tout, sur la rétine, c'était exactement le même chiffre qui était présenté pour la même durée. Rien d'étonnant, donc, à ce que les premières étapes de traitement visuel soient essentiellement constantes, que l'on finisse par voir le chiffre ou non.

Les ondes suivantes, issues des aires visuelles des deux hémisphères, se comportaient de façon continue : leur taille croissait lentement, en proportion directe de la durée qui séparait le chiffre du masque. Ainsi, le chiffre, bien qu'il ne s'affichât que pendant un instant, continuait de progresser dans la hiérarchie des aires cérébrales, jusqu'au moment où le

masque venait l'interrompre. Le strict parallélisme entre la durée des stimuli et la taille des réponses cérébrales observées ne correspondait pas à la perception subjective des observateurs, qui suivait une courbe non linéaire. Cela signifiait donc que ces activations cérébrales ne codaient pas encore pour la perception consciente. À ce stade, l'activité évoquée par le chiffre continuait d'ailleurs d'être très forte, même dans les essais où les participants ne parvenaient pas à le voir.

Vers 270 millisecondes après l'arrivée du chiffre, toutefois, les enregistrements changeaient (figure 19). Les ondes cérébrales divergeaient soudain. Une avalanche d'activité se déclenchait, forte et vive, lorsque le participant disait voir le chiffre. La taille de ce changement d'état cérébral était sans commune mesure avec le petit allongement du stimulus. Ces résultats constituaient une preuve directe que l'accès à la conscience ressemble à une transition de phase dans la dynamique des réseaux cérébraux.

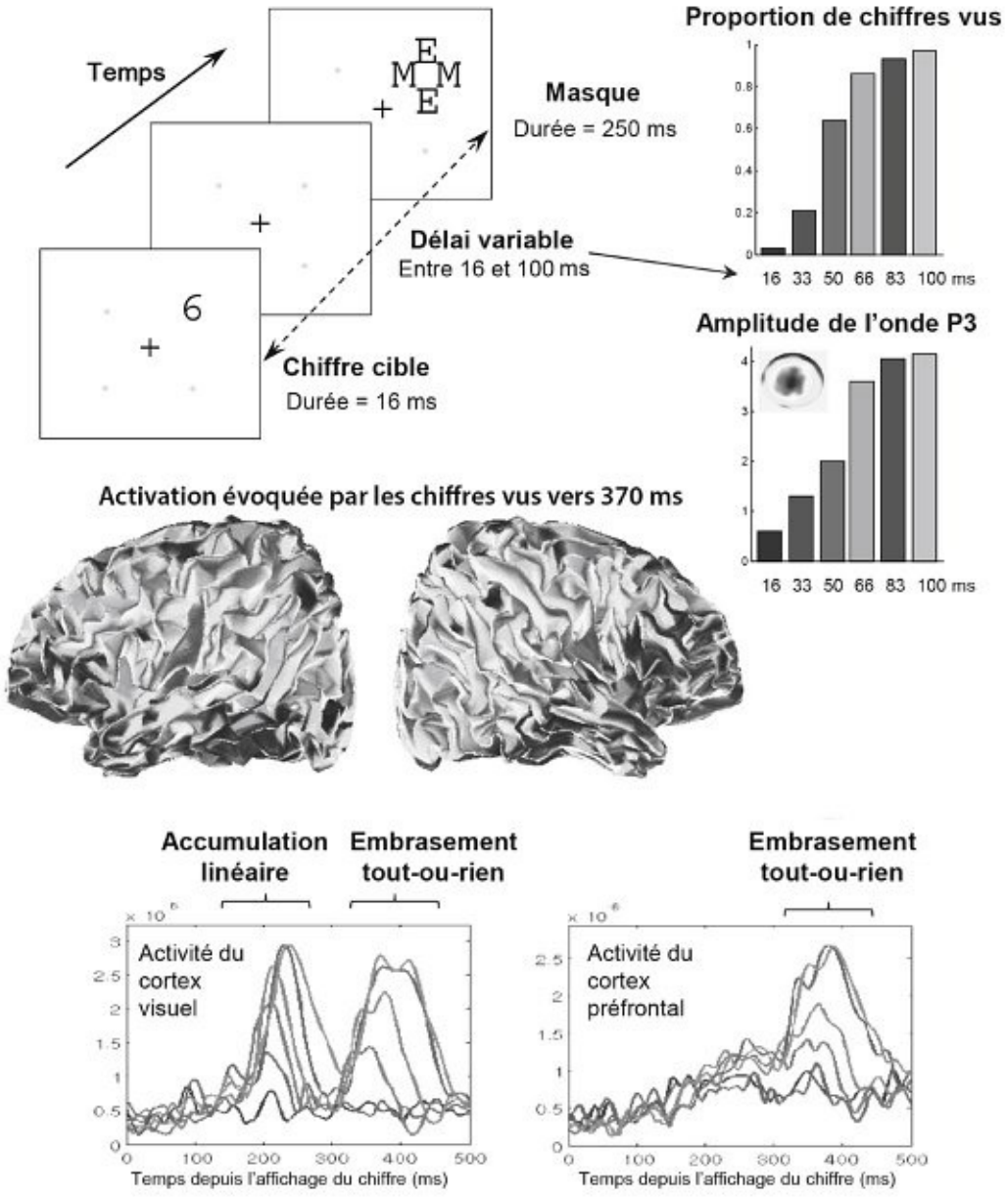


Figure 19. Un changement brutal de l'activité cérébrale tardive reflète l'accès d'une information à la conscience. Dans cette expérience, on présente brièvement un chiffre, puis, après un délai variable, on le masque avec des lettres. L'activité du cortex visuel augmente continûment avec le délai. La perception consciente, par contre, est discontinue : dès que le délai dépasse 50 millisecondes, le chiffre devient visible. Une fois de plus, l'onde P3b reflète la prise de conscience. Environ 300 millisecondes après l'apparition du chiffre, de nombreuses régions du cortex, dont les aires préfrontales, s'embrasent brutalement uniquement quand l'observateur rapporte avoir vu le chiffre.

Une fois de plus, la divergence consciente prenait la forme d'une onde P3 – une augmentation massive du voltage à la surface de la tête. Elle prenait sa source dans l'activation simultanée d'un vaste réseau de régions corticales situées dans les lobes préfrontaux, pariétaux et occipitaux gauches et droits. Comme nous ne présentions le chiffre que d'un côté de l'écran, il était frappant de voir l'embrasement cortical envahir les deux hémisphères. Clairement, la perception consciente amplifie massivement le minuscule filet d'activité qui émerge initialement de la rétine. L'avalanche d'opérations mentales culmine au moment où de nombreuses régions du cerveau, distribuées dans les deux hémisphères, se mettent à décharger en synchronie, ce qui indique que le petit stimulus est devenu un grand état de conscience.

Au plus profond du cerveau conscient

Les expériences que nous avons examinées jusqu'ici ne s'approchent guère des neurones. L'IRM fonctionnelle et les potentiels évoqués ne donnent qu'un aperçu très partiel de l'activité neuronale sous-jacente. Récemment, cependant, l'exploration de la conscience a pris un nouveau tour. Chez certains patients épileptiques, des électrodes placées directement dans le cerveau ont jeté une lumière nouvelle sur les réponses du cortex. Dès que cette méthode est apparue, mon équipe l'a employée afin de suivre, pas à pas, le trajet d'un mot invisible ou visible à travers le cortex[36]. Nos observations et celles d'autres équipes soutiennent pleinement le concept d'une avalanche suivie d'un embrasement du cortex[37].

Dans une étude, nous avons combiné les données de dix patients afin de disposer de suffisamment d'électrodes tout au long du trajet et de suffisamment d'essais pour les trier en fonction du degré de perception consciente[38] (figure 20). Comme d'habitude, l'activité était très semblable

au départ, puis elle divergeait rapidement selon que le mot était vu ou non. Après environ 300 millisecondes, la différence devenait énorme. Dans les essais non conscients, l'activité décroissait si vite qu'elle était pratiquement indétectable dans le lobe frontal. Au contraire, si le sujet disait avoir pris conscience du mot, l'activité ne cessait de s'amplifier. En un tiers de seconde, le cerveau passait d'une toute petite différence à un gigantesque embrasement.

Nos électrodes nous permettaient également d'évaluer jusqu'où le message conscient était diffusé. Les électrodes que nous enregistrions n'avaient été placées que dans un but clinique : mesurer les crises d'épilepsie. Leur position dans le cortex n'était donc pas influencée par les objectifs de notre étude. Et pourtant, près de 70 % d'entre elles étaient modulées par les mots conscients – contre seulement 25 % pour les mots inconscients. Conclusion : le message conscient s'étend pratiquement à l'ensemble du cortex, alors que l'information inconsciente reste confinée à certains circuits spécialisés.

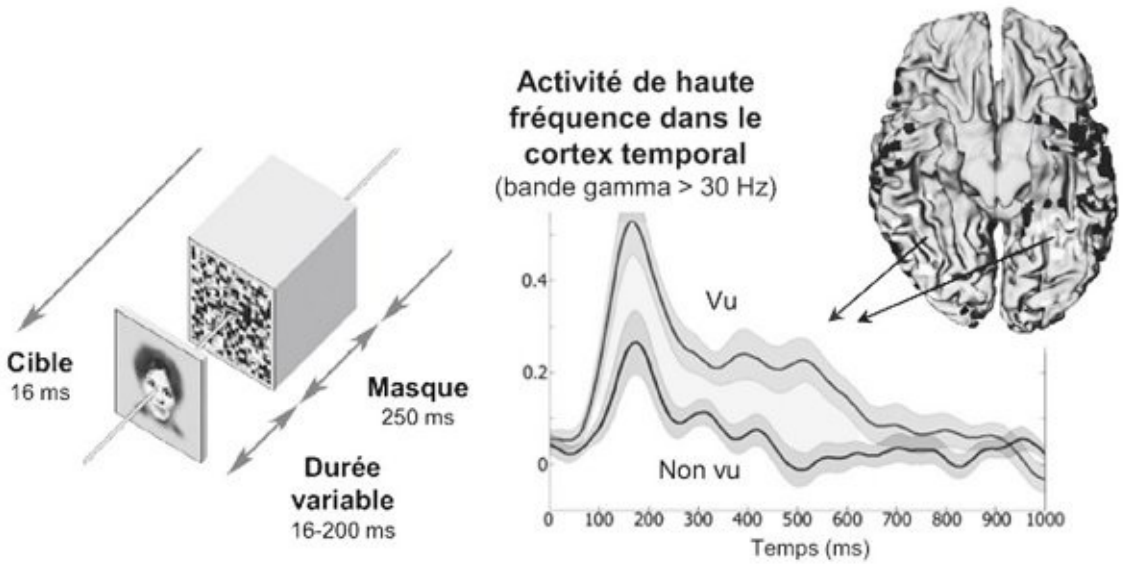


Figure 20. Une longue bouffée d'activité de haute fréquence accompagne la perception d'une image et constitue une troisième signature de la conscience. Chez certains patients épileptiques,

des électrodes placées à la surface du cortex enregistrent l'avalanche d'activité évoquée par une image. Lorsque les sujets ne parviennent pas à la voir, seule une brève impulsion de haute fréquence parcourt le cortex visuel ventral. Lorsque l'image est reconnue, par contre, l'avalanche s'autoamplifie jusqu'à causer un embrasement global. La perception consciente se caractérise par une bouffée prolongée d'activité électrique de haute fréquence, qui reflète l'intense activation des circuits neuronaux sous-jacents.

Les enregistrements intracrâniens ouvrent une fenêtre exceptionnelle sur la dynamique de l'activité corticale. Les électrophysiologistes distinguent de nombreux rythmes au sein du cerveau. En état de veille, le cortex émet quantité de fluctuations électriques qui peuvent être grossièrement triées en fonction de leur fréquence dominante. À chaque bande de fréquence sa lettre grecque : le bestiaire des oscillations cérébrales comprend ainsi la bande alpha (8 à 13 hertz), bêta (13 à 30 hertz), et gamma (30 hertz et au-delà). Quand un stimulus pénètre dans le cerveau, il perturbe ces rythmes endogènes, les déplace ou impose des fréquences qui lui sont propres. L'analyse de ces fluctuations rythmiques nous a permis de découvrir de nouvelles signatures de la conscience.

Dès que le mot apparaissait, qu'il soit ou non conscient, une onde d'activité gamma se propageait dans le cerveau. Le cerveau émettait des signaux électriques de haute fréquence dont on pense qu'ils sont le reflet direct de l'activité neuronale. Cette bouffée d'activité durait 200 millisecondes, puis elle s'écroulait pour les mots invisibles, tandis qu'elle restait soutenue pour les mots visibles. Une fois de plus, à partir de 300 millisecondes, on observait une différence massive, un phénomène de « tout-ou-rien ». Par la suite, Rafi Malach et ses collègues de l'Institut Weizmann ont décrit un phénomène très semblable[39] (figure 20). Un accroissement soudain de la puissance des ondes gamma, à partir de 300 millisecondes et au-delà, constitue donc une troisième signature de la perception consciente.

Ces observations jettent une lumière nouvelle sur une hypothèse ancienne : l'idée que le cerveau conscient émet des ondes autour de 40

hertz. Dès les années 1990, le prix Nobel Francis Crick, aujourd'hui décédé, et son élève Christof Koch spéculèrent que la conscience serait portée par une oscillation autour de 40 cycles par seconde (soit une onde toutes les 25 millisecondes), qui serait la fréquence de circulation des informations entre le cortex et le thalamus. Nous savons à présent que cette hypothèse était fausse : même un stimulus invisible induit des ondes de haute fréquence, non seulement à 40 hertz, mais aussi dans toute la bande gamma[40]. Pourquoi serions-nous surpris, d'ailleurs, qu'une forte activation cérébrale accompagne aussi bien le traitement conscient que les opérations non conscientes ? Les fréquences gamma envahissent pratiquement n'importe quel groupe de neurones corticaux activés, dès que l'inhibition sculpte les décharges neuronales en bouffées rythmiques[41]. Ce que montrent nos expériences, c'est que cette activité s'amplifie grandement au cours de l'embrasement conscient. C'est donc l'amplification tardive de l'activité gamma, plutôt que sa présence pure et simple, qui constitue une signature solide de l'accès à la conscience.

Le réseau neuronal global

Si le cerveau émet des oscillations synchrones, c'est peut-être parce que la synchronie facilite la transmission des informations[42]. Au cœur des vastes forêts du cortex, parcourues de millions de décharges aléatoires, une petite assemblée de neurones actifs pourrait aisément se perdre. Si ces neurones crient à l'unisson, par contre, leur voix se propage plus loin. Il se pourrait donc que les neurones excitateurs orchestrent leurs décharges afin d'être mieux entendus à distance. La synchronie ouvrirait ainsi une voie de communication entre deux assemblées de neurones[43]. Les neurones qui oscilleraient en rythme partageraient périodiquement des périodes privilégiées de communication pendant lesquelles ils seraient mieux aptes à échanger des messages.

Si nous parvenons à mesurer une synchronisation des signaux électriques macroscopiques émis par le cerveau, cela signifie que des milliers de neurones échangent des informations synchrones. Il se pourrait donc que ce soit la diffusion de ces échanges à l'ensemble du cortex qui signale un moment de conscience. En effet, cette synchronie globale permettrait la formation d'une vaste assemblée de neurones.

À l'appui de cette hypothèse, de nombreuses équipes ont montré que la synchronisation massive des signaux électromagnétiques du cortex constitue une quatrième signature de la perception consciente[44]. Comme toujours, cet effet émerge tardivement : environ un tiers de seconde après qu'une image est apparue, de nombreuses électrodes oscillent en rythme – mais seulement si l'image est visible (figure 21). Les images subliminales n'induisent qu'une forme temporaire de synchronie, restreinte aux aires postérieures où s'effectue un traitement non conscient. La perception consciente, par contre, implique une communication corticale à longue distance, un échange bidirectionnel de signaux tellement intense qu'on le décrit comme un « réseau cérébral global[45] » (*brain web*). La bande de fréquence qui permet aux aires cérébrales de communiquer entre elles varie selon les études, mais elle concerne plutôt les fréquences basses (soit la bande bêta, 13 à 30 hertz, soit la bande thêta, 3 à 8 hertz). On présume que ces fréquences lentes permettent de compenser les délais de transmission assez élevés qui séparent des régions corticales distantes de plusieurs centimètres.

Nous ne comprenons toujours pas bien comment des millions de décharges neuronales, réparties dans le temps et dans l'espace, parviennent à créer un code neural qui représente n'importe quel état mental. Un nombre croissant d'indices suggère que la décomposition en fréquences est une technique utile mais insuffisante pour comprendre le cerveau. La plupart du temps, les signaux cérébraux n'oscillent pas à une fréquence très précise. Ils fluctuent sans cesse, dans des gammes de fréquences étendues et qui changent à tout instant, tout en parvenant, on ne sait comment, à rester synchrones d'une aire à l'autre. De plus, les

bandes de fréquences sont « enchâssées » les unes dans les autres : les bouffées de haute fréquence surviennent à des instants précis de l'onde de basse fréquence[46]. Pour comprendre ces motifs complexes, nous allons avoir besoin de nouveaux outils mathématiques.

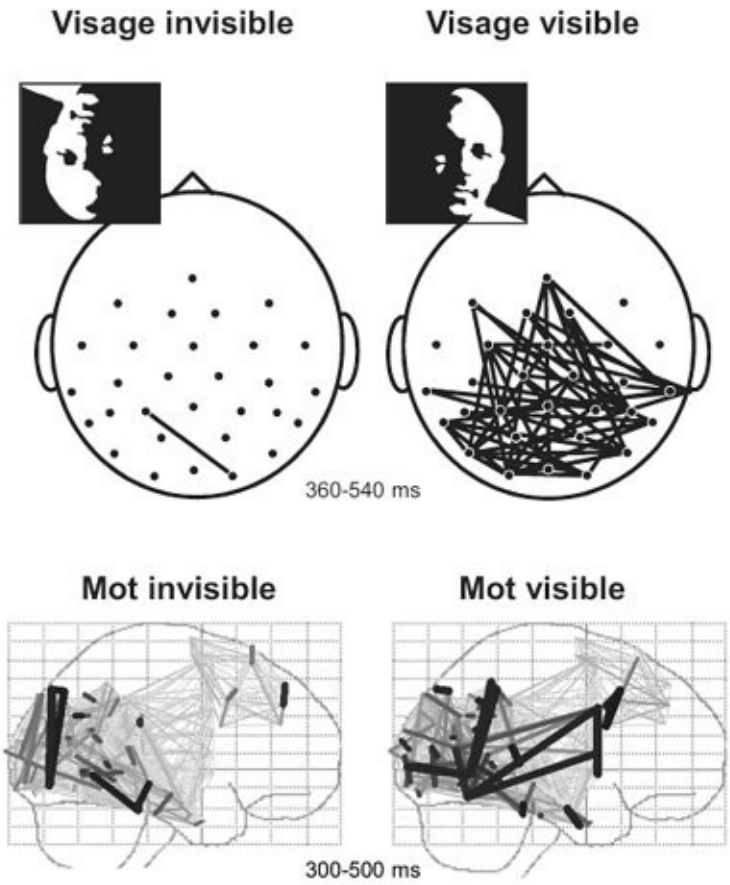


Figure 21. La synchronisation de nombreuses aires cérébrales distantes, sous la forme d'un réseau global, fournit une quatrième signature de la conscience. Un tiers de seconde après avoir vu un visage (en haut), l'activité électrique du cerveau se synchronise (chaque ligne représente une paire d'électrodes). Les oscillations de haute fréquence dans la bande « gamma » (au-delà de 30 hertz) fluctuent en synchronie, ce qui suggère que les régions cérébrales sous-jacentes échangent des messages à haut débit. De même, pendant la perception consciente d'un mot (en bas), de multiples relations de causalité bidirectionnelle s'établissent entre des aires cérébrales distantes, particulièrement avec le lobe frontal. Seule une faible synchronisation locale apparaît lorsque les participants ne parviennent pas à voir le visage ou le mot.

L'un des outils intéressants s'appelle la « causalité de Granger ». Inventée en 1969 par l'économiste britannique Clive Granger, cette méthode permet de déterminer si deux séries de mesures (par exemple deux indicateurs économiques) sont reliées entre elles à tel point qu'on puisse dire que l'une « cause » l'autre. Récemment, la méthode a été étendue aux signaux cérébraux. En effet, dans le cerveau, la question de la causalité représente un vrai défi. Tant de connexions s'y entrecroisent qu'on ne parvient guère à savoir qui cause quoi. L'activité neuronale progresse-t-elle de bas en haut, depuis les organes des sens jusqu'aux aires corticales supérieures ? Ou bien circule-t-elle en sens inverse, du haut vers le bas, dans la mesure où les aires de haut niveau envoient des signaux prédictifs aux régions sensorielles afin de les interroger et de les préparer à percevoir ? Sur le plan anatomique, le cortex comprend des projections ascendantes aussi bien que descendantes. La grande majorité des faisceaux de connexions à longue distance sont bidirectionnels. Étrangement, les connexions de bas en haut, qui transmettent les données sensorielles, sont bien *moins* nombreuses que les projections de haut en bas. Nul ne connaît la raison de cette organisation contre-intuitive. Se pourrait-il qu'elle joue un rôle dans la perception consciente ?

Grâce à la causalité de Granger, nous avons clarifié cette question. La méthode consiste à comparer deux signaux afin d'évaluer si l'un des signaux précède et prédit systématiquement l'autre. Pour savoir si le signal A joue un rôle causal dans les fluctuations du signal B, on regarde si le passé récent de A prédit l'état présent de B, mieux qu'on ne pourrait le faire en examinant le seul signal B. Selon cette définition, rien n'interdit l'existence de relations causales dans les deux sens : deux régions cérébrales peuvent s'influencer mutuellement.

Quand nous l'avons appliquée à nos enregistrements intracrâniens, la causalité de Granger a immédiatement clarifié la dynamique de l'embrasement conscient[47]. Nous avons observé un accroissement spectaculaire de la causalité *bidirectionnelle* dans presque tout le cerveau, spécifiquement lors des essais conscients. Une fois de plus, cette

« explosion causale » émergeait après 300 millisecondes de traitement des données visuelles. À cet instant, la grande majorité des signaux cérébraux s'étaient fondus en un seul état intégré, un gigantesque enchevêtrement de relations réciproques où la majorité des connexions propagent les informations vers l'avant du cerveau, du cortex visuel au lobe frontal, tandis qu'une forte minorité opère en sens inverse.

La propagation de l'onde vers l'avant n'est pas compliquée à comprendre : il faut bien que l'information sensorielle qui définit l'objet perçu quitte la rétine et grimpe dans la hiérarchie des aires corticales, depuis le cortex visuel primaire jusqu'aux représentations plus abstraites de son identité et de son sens. L'onde qui parcourt le cerveau en sens inverse, par contre, n'est pas encore bien comprise. Ce pourrait être un signal attentionnel qui amplifie l'activité sensorielle. Il se pourrait également que le cerveau renvoie des messages aux aires sensorielles afin de leur confirmer que toutes leurs sensations ont bien été interprétées, et de vérifier s'il reste certains détails à expliquer. L'interprétation la plus générale est que, lors de la prise de conscience, l'activité cérébrale converge vers un « attracteur global » – un ensemble de régions coactives qui se soutiennent l'une l'autre et qui produisent collectivement un état de réverbération stable.

Rien de tout cela ne survenait lors des essais non conscients. Le réseau cérébral global ne s'embrasait pas. On ne voyait qu'une période transitoire de relations causales restreintes aux aires visuelles situées dans la partie postérieure du cerveau – et celle-ci ne durait guère au-delà de 300 millisecondes. Étonnamment, cette activité non consciente comprenait une majorité de signaux descendants. Tout se passait comme si les régions antérieures du cerveau interrogeaient désespérément les aires sensorielles. Lorsque leurs questions restaient sans réponse, le cerveau concluait à l'absence d'une perception consciente.

Le point de non-retour et ses précurseurs

Résumons brièvement toutes ces découvertes. L'accès à la conscience correspond à une onde d'activité neuronale qui progresse jusqu'à un point de non-retour. Chaque stimulation consciente déclenche une avalanche de décharges neuronales qui s'amplifie elle-même et finit par embraser des régions corticales distantes les unes des autres pour former un état cohérent. Cet état de conscience démarre environ un tiers de seconde après que le stimulus est apparu sur la rétine. Le cortex préfrontal reçoit des informations ascendantes en provenance des organes des sens, mais il dialogue également avec elle en retour, en revoyant de nombreux messages descendants. Il en résulte un réseau cérébral global formé de multiples régions synchrones, dont les différentes facettes expliquent les quatre signatures de la conscience : une activité distribuée, particulièrement dans les aires pariétales et préfrontales ; une onde P3 ; une amplification de l'activité dans la bande gamma ; et une synchronisation massive à longue distance.

La métaphore de l'avalanche, avec son point de non-retour au-delà duquel le phénomène ne cesse de s'amplifier, clarifie le déroulement temporel de la conscience. Mes propres données, comme celles de nombreux autres collègues, suggèrent que la formation d'un état de conscience prend au moins un tiers de seconde. Pourtant d'autres laboratoires ont également observé des différences bien plus précoces entre les essais conscients et non conscients – parfois un dixième de seconde seulement après l'apparition de l'image[48]. Est-ce qu'ils se trompent ? Pas forcément. Avec une méthode sensible, on détecte fréquemment de petits signes avant-coureurs de l'avalanche, qui précèdent l'embrasement cortical. Mais est-ce que ces signaux signifient que la conscience est déjà là ? Je ne crois pas. D'abord, ils ne sont pas toujours présents – il existe toute une série d'excellentes expériences, qui utilisent exactement le même stimulus dans les essais vus et pas vus, et où le seul corrélat de la perception consciente est un embrasement tardif[49].

Ensuite, la forme de ces signes avant-coureurs ne correspond pas à ce que rapportent les sujets eux-mêmes. Pendant le masquage, par exemple, les événements précoces croissent en proportion directe de la durée du stimulus, sans montrer l'effet de seuil qui caractérise pourtant la conscience subjective. Enfin, les événements précoces ne montrent guère qu'une petite amplification lors des essais conscients, qui surnage à peine au-dessus d'une énorme activation inconsciente[50]. À nouveau, cela ne concorde pas avec la subjectivité : si cette activité neuronale précoce correspondait déjà à l'accès conscient, elle ne devrait pas rester aussi grande alors que le sujet affirme n'avoir rien vu du tout.

Mais alors, comment expliquer que l'activité visuelle précoce prédise la prise de conscience ? Pour le comprendre, il faut se souvenir que le cerveau est parcouru d'incessantes fluctuations aléatoires. Ce « bruit » neuronal s'ajoute à l'activité sensorielle et fait varier les chances de franchir le seuil – exactement comme une petite boule de neige peut déclencher une avalanche si le terrain est favorable, ou le battement d'une aile de papillon, un gigantesque ouragan. Les avalanches, les tremblements de terre et les ouragans sont des phénomènes probabilistes qui dépendent de fluctuations aléatoires autant que de causes extérieures. Il en va de même pour l'embrasement cortical : lorsque le stimulus est faible, le phénomène cesse d'être déterministe et varie d'un essai à l'autre. Les fluctuations spontanées de l'activité cérébrale s'alignent parfois avec le stimulus sensoriel, auquel cas ses chances d'être perçu augmentent. D'autres fois, l'activité spontanée entre en compétition avec l'image entrante et diminue la probabilité qu'elle soit détectée. Une fois que l'on moyenne des dizaines d'essais avec ou sans conscience, toutes ces fluctuations aléatoires ressortent du bruit de fond. Toutes choses égales par ailleurs, l'activité cérébrale précoce paraît un peu plus grande dans les essais vus que dans les essais invisibles. En conclure que le cerveau est déjà conscient serait une erreur aussi grossière que d'appeler « avalanche » la moindre boule de neige qui commence à dévaler une pente !

Certaines expériences détectent même des signaux qui prédisent (partiellement) la conscience avant même que le stimulus visuel ne soit présenté[51]. Cela paraît de plus en plus étrange. Comment le cerveau pourrait-il contenir une représentation de l'image plusieurs secondes avant qu'elle n'apparaisse ? S'agit-il d'un cas de précognition ? Pas du tout. Les différences qui précèdent l'apparition de l'image ne font que refléter les conditions qui, *en moyenne*, facilitent le déclenchement de l'avalanche consciente. En effet, parmi les fluctuations de l'activité cérébrale, certaines facilitent la prise de conscience, tandis que d'autres nous empêchent de nous concentrer sur l'image à percevoir. L'imagerie cérébrale est devenue tellement sensible qu'elle détecte ces petits signaux qui, avant même que le stimulus ne survienne, reflètent l'état de réceptivité du cortex. Le simple fait de trier rétrospectivement les essais en fonction du résultat final (« vu » ou « pas vu ») fait apparaître des prédicteurs précoces de la prise de conscience. Ces marqueurs ne sont pas constitutifs de l'état de conscience – celle-ci ne survient que bien plus tard, lorsque les biais et les données se combinent en un vaste embrasement global.

Toutes ces observations conduisent à la même conclusion : il faut absolument distinguer les simples *corrélats* de la conscience des authentiques *signatures* d'un accès conscient. La recherche des mécanismes cérébraux de l'expérience consciente est parfois décrite comme une quête des « corrélats neuronaux de la conscience », mais ce terme est inapproprié. Corrélation n'est pas causalité. Bien trop de signaux cérébraux corrèlent avec la perception consciente – y compris, comme nous venons de le voir, les fluctuations qui précèdent le stimulus. Nous ne recherchons pas une simple relation statistique entre l'activité cérébrale et la perception consciente, mais un code neural de la conscience : une signature systématiquement présente quand survient l'accès conscient, absente quand il échoue, et qui code l'ensemble de l'expérience subjective que ressent l'observateur.

Le décodage d'une pensée consciente

Écoutons de nouveau la plaidoirie de l'avocat du diable. L'embrasement global n'est-il pas qu'un signal d'alerte qui se déclenche dès que nous prenons conscience de quelque chose ? Qu'est-ce qui prouve que cette décharge d'activité neuronale détient tous les détails de notre pensée consciente ? Se pourrait-il qu'il ne s'agisse que d'une bouffée d'excitation globale, sans rapport avec le contenu de notre expérience subjective ?

La question n'est pas seulement théorique. En effet, de nombreux noyaux cérébraux semblent jouer un rôle générique : ils déchargent dès qu'un événement requiert notre attention. Le locus cœruleus, par exemple, est un groupe de neurones du tronc cérébral qui libèrent un neurotransmetteur bien particulier, la noradrénaline, dès que survient un événement suffisamment important pour demander toute notre attention. Une bouffée de noradrénaline pourrait donc accompagner chaque prise de conscience, et certains ont suggéré que c'est exactement cela qu'indique l'onde P3[52]. Dans ce cas, la décharge des neurones du locus cœruleus n'aurait pas de lien direct avec la conscience. Elle ne serait qu'un signal non spécifique, essentiel à notre vigilance, mais dépourvu de toutes les distinctions plus fines qui tissent la trame de notre vie mentale[53]. Faire d'un tel signal cérébral la signature de la conscience serait comme de confondre le coup de sonnette du livreur de journaux avec le texte du journal lui-même.

Mais comment séparer le bon grain de l'ivraie, la conscience des signaux qui l'accompagnent ? Sur le principe, les choses sont claires. Nous sommes à la recherche d'un code neural qui, une fois décrypté, fournisse une image exacte du contenu de la conscience de la personne scannée – ni plus ni moins[54]. Ce code conscient doit contenir tous les détails de l'expérience subjective. Il doit être insensible à ce que la personne ne voit pas, même si ce sont des aspects saillants de l'image. Inversement, il doit coder les moindres aspects de la perception, y compris les illusions ou les hallucinations. Il doit également préserver notre sentiment de

ressemblance. Si nous jugeons, par exemple, un losange et un carré comme des formes distinctes (■ et ♦), ces formes doivent être représentées dans le cerveau par des codes différents.

Le code conscient doit également être invariant : dès que notre perception du monde est stable, il ne doit plus bouger. Ce critère contraint énormément la recherche des signatures de la conscience. Il exclut presque à coup sûr les aires sensorielles de bas niveau. En effet, lorsque nous nous déplaçons dans notre salon, l'image des murs et des meubles ne cesse de changer dans le cortex, et pourtant nous n'en avons pas conscience et pensons que la pièce est stable. Notre conscience reste fixe alors même que le mouvement est omniprésent dans nos aires sensorielles. Trois ou quatre fois par seconde, notre regard se déplace et aussitôt, sur notre rétine comme dans nos aires visuelles, toute l'image du monde extérieur glisse dans le sens opposé. Il est heureux que nous n'ayons pas conscience de ce roulis nauséeux ! De même, quand nous suivons des yeux un oiseau en vol, nous n'avons pas conscience que tout le reste du monde défile dans l'autre sens. Grâce aux capteurs inertiels de l'oreille interne et aux prédictions qu'envoient nos aires motrices, nous parvenons à nous abstraire du déplacement de notre corps et à percevoir notre environnement comme une entité stable. Ce n'est que lorsque ces signaux prédictifs sont perturbés (par exemple si vous tapotez légèrement de la main sur le côté de votre œil) que le monde semble se déplacer.

Le glissement visuel qu'engendre notre mouvement propre n'est qu'un des nombreux indices que notre cerveau efface de notre contenu conscient. Bien d'autres propriétés font de notre conscience un monde à part, lent et serein, à cent lieues des caprices que subissent nos organes des sens. Lorsque nous regardons un vieil écran de télévision muni d'un tube cathodique, l'image clignote cinquante ou soixante fois par seconde, et l'enregistrement des neurones du cortex visuel montre qu'ils clignotent à la même fréquence[55]. Heureusement, ce rythme obsédant ne va pas plus loin : l'information temporelle fine qui envahit nos aires visuelles est filtrée avant d'atteindre notre conscience. De même, l'expérience montre

qu'une grille formée de fines barres parallèles noires et blanches peut très bien être représentée dans notre cortex sans que nous ne la voyions[56].

Cependant, notre conscience n'est pas simplement aveugle : c'est un observateur actif qui ne cesse de transformer et d'améliorer l'image rétinienne. Dans nos aires visuelles, les informations en provenance du centre de la rétine, appelé la *fovéa*, sont massivement amplifiées : bien plus de neurones s'intéressent au centre qu'à la périphérie du champ visuel. Pourtant, nous n'avons pas l'impression de voir le monde à travers une loupe, et les visages et les mots ne grossissent pas soudain lorsque nous décidons de les regarder. La conscience assemble une image stable à partir d'informations fragmentaires.

Voulez-vous un dernier exemple du décalage énorme qui sépare les données sensorielles du contenu de la conscience ? Prenez la couleur. En dehors de la fovéa, nos rétines ne contiennent que très peu de cônes sensibles à la couleur – et pourtant la totalité de la scène visuelle nous paraît colorée. Nous n'avons pas l'impression de voir des couleurs vives entourées d'un anneau noir et blanc, et nous ne nous émerveillons pas de voir des couleurs apparaître chaque fois que nous posons le regard quelque part. Les couleurs que nous croyons « voir » dans le monde extérieur ne sont qu'une reconstruction cérébrale. Chacune de nos rétines contient même un énorme trou, la tache aveugle, à l'endroit où le nerf optique s'écarte de l'œil – et pourtant il n'y a pas de trou de la taille d'une grosse pièce de monnaie dans l'image que nous avons du monde !

Tous ces arguments prouvent que les réponses visuelles précoces ne peuvent pas détenir le secret de la conscience. Notre cerveau effectue un traitement considérable de l'image avant de nous donner à voir un monde stable. C'est pourquoi les signatures de la conscience sont si tardives : il faut un tiers de seconde, au minimum, avant que notre cortex ne résolve le puzzle de la perception et n'assemble les données en un tout cohérent.

Si cette hypothèse est vraie, alors l'activité cérébrale tardive devrait contenir un enregistrement complet de toute notre expérience subjective – le code exhaustif de nos pensées. Si nous savions le décrypter, nous

aurions accès à l'ensemble de la vie mentale d'une personne. Tout ce qu'elle voit, pense ou ressent consciemment devrait s'y trouver inscrit.

Ce rêve n'appartient déjà plus à la science-fiction. En enregistrant quelques neurones dans le cerveau humain, Quian Quiroga et ses collègues israéliens Itzhak Fried et Rafi Malach ont ouvert la voie d'un décodage de la perception consciente[57]. Ils ont découvert des neurones qui ne répondent qu'à certaines images, à certains lieux ou à certaines personnes – et ce, uniquement lorsque ces stimuli sont perçus consciemment. Leurs résultats réfutent donc définitivement l'hypothèse d'une excitation non spécifique du cortex. Au cours de l'embrasement conscient, le cerveau n'est pas simplement excité globalement. Au contraire, à chaque image que nous voyons correspond l'entrée en activité d'un petit jeu de neurones bien spécifiques, dont les contours définissent le contenu de l'expérience subjective.

Comment est-il devenu possible d'enregistrer des neurones uniques dans le cortex humain ? J'ai expliqué plus haut comment les neurochirurgiens étudient l'épilepsie en plaçant des électrodes à l'intérieur de la boîte crânienne. Les électrodes traditionnelles sont de gros diamètre et elles enregistrent l'activité moyenne de plusieurs milliers de neurones. Cependant, en s'appuyant sur des études antérieures[58], Itzhak Fried a conçu un jeu d'électrodes tellement fines qu'elles mesurent l'activité d'un seul neurone[59]. Chez l'homme comme chez l'animal, les neurones corticaux échangent des signaux électriques tout-ou-rien qu'on appelle des « décharges » (*spikes* en anglais), car ils se traduisent par des changements brutaux du potentiel électrique que l'on observe aisément sur un oscilloscope. Les neurones excitateurs émettent typiquement quelques décharges par seconde, et chacune d'elles se propage le long de l'axone en direction d'autres neurones, proches ou lointains. Grâce à Itzhak Fried et ses collègues, on peut aujourd'hui enregistrer, pendant plusieurs heures ou même plusieurs jours, *toutes* les décharges qu'un neurone émet alors que le patient, parfaitement éveillé, vit une vie normale.

Quand Fried et ses collaborateurs ont placé leurs électrodes dans le lobe temporal antérieur humain et qu'ils ont bombardé le patient de centaines de photos de visages, de lieux, d'objets ou de mots, ils ont fait une découverte remarquable : chaque neurone ne répond qu'à un tout petit nombre d'entre elles. Le plus souvent, un neurone ne décharge qu'à la présentation d'une seule image. L'un des neurones, par exemple, ne déchargeait qu'aux photos de Bill Clinton – aucune autre personne ne le faisait vibrer[60] ! Au fil des années, on a découvert des neurones qui répondent à toutes sortes d'images, depuis les membres de la famille du patient jusqu'aux lieux célèbres comme l'Opéra de Sydney ou la Maison Blanche et à des célébrités comme Jennifer Aniston ou Homer Simpson. Mieux encore, le mot seul suffit souvent à les activer. Ainsi, le même neurone décharge-t-il en réponse aux mots « tour Eiffel » et à la vue d'une photo de ce monument.

Il est étonnant qu'en insérant une électrode au petit bonheur dans le cerveau, et en enregistrant le premier neurone venu, on trouve des cellules qui répondent à Bill Clinton. Cela ne peut signifier qu'une seule chose : ces neurones ne sont pas uniques. Chaque image que voit le patient doit faire décharger quelques millions de cellules nerveuses dans son cerveau, sinon nous n'aurions aucune chance de tomber pile sur la bonne cellule.

Pris dans leur ensemble, ces neurones forment un code interne des lieux, des personnes et des autres concepts qui nous sont familiers. Chaque image évoque une sorte de carte qui comprend des neurones actifs et d'autres inactifs, et dont la topographie est propre à l'objet reconnu. Ce code est si précis qu'en examinant une centaine de neurones et en regardant lesquels déchargent et lesquels restent silencieux on peut entraîner un ordinateur à deviner, avec une très haute précision, quelle image la personne était en train de voir[61].

Ces neurones sont donc à la fois spécifiques et invariants. Leurs décharges ne représentent ni un signal global, tel que la surprise ou la vigilance, ni les détails d'une image particulière, mais une abstraction du

concept sous-jacent (« la tour Eiffel ») – exactement le genre de code stable que nous prédisions pour les pensées conscientes. Reste une question clé : ces neurones s'activent-ils uniquement lorsque la personne prend conscience de l'image ? La réponse est oui. Dans la région temporale antérieure, de nombreux neurones ne déchargent pratiquement *que* lorsque leur image préférée est vue consciemment. La même image, lorsqu'elle est présentée trop brièvement pour être identifiée, n'engendre aucune activité[62]. Si, par exemple, le patient voit des images masquées et rapporte, à chaque essai, s'il a reconnu l'image ou non, la majorité des cellules suivent étroitement la perception consciente. Bien que les photographies soient exactement les mêmes dans les essais conscients et non conscients, c'est l'expérience subjective qui détermine les décharges neuronales dans cette région du cerveau.

La figure 22 montre une cellule qui décharge à n'importe quelle image du World Trade Center, mais seulement lors des essais conscients. Lorsque le patient déclare n'avoir rien vu, le neurone reste silencieux. Au seuil de conscience, lorsque l'image est présentée si brièvement que le patient ne la voit que la moitié du temps, seuls les essais vus s'accompagnent de décharges neuronales. L'activité du neurone est si tranchée que l'on peut séparer les essais conscients et non conscients, simplement en regardant si le nombre de décharges dépasse un certain seuil. Un état d'esprit subjectif (« j'ai vu le World Trade Center ») peut donc être déduit de l'activité objective du cerveau.

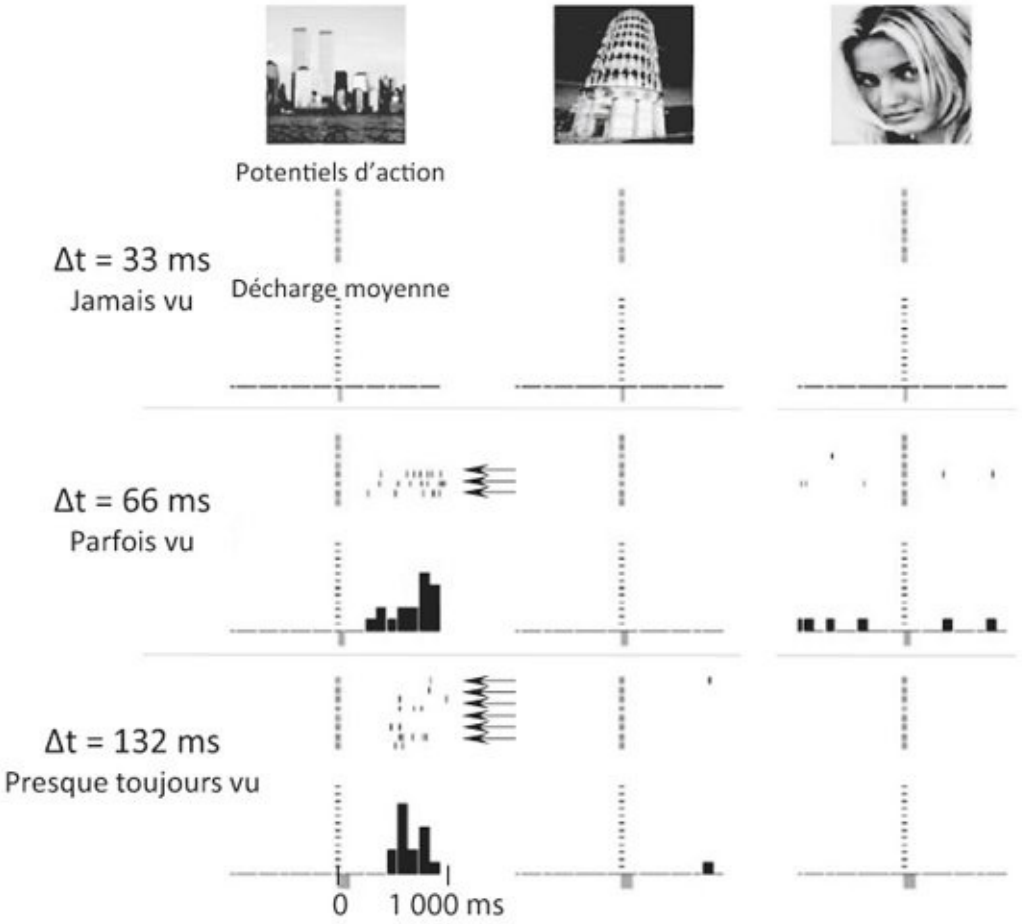


Figure 22. L'activité d'un seul neurone suffit parfois à identifier le contenu de la conscience. Dans cet exemple, un neurone de la région temporale antérieure décharge à la vue d'une image du World Trade Center. Lorsqu'on augmente la durée de présentation, la probabilité de percevoir l'image augmente. Cependant, le neurone ne décharge que lorsque le sujet prend conscience de l'image (essais marqués d'une flèche). Il est également très sélectif et ne répond guère à la vue d'autres images (tour de Pise, visage, etc.). Ainsi, ses potentiels d'action indiquent la prise de conscience d'une image spécifique. Des millions de neurones de ce type, qui déchargent tous au même instant, codent pour chacune de nos pensées.

Si les cellules du cortex temporal antérieur reflètent la perception consciente, alors leurs décharges ne doivent pas dépendre de la manière dont la conscience est manipulée. Effectivement, Fried et ses collaborateurs sont parvenus à répliquer leurs résultats en utilisant, non plus le masquage, mais la rivalité binoculaire. Ayant découvert un

neurone sensible au visage de Bill Clinton, ils ont présenté l'image du président américain à un seul œil. Dès que le sujet en prenait conscience, la cellule se mettait à décharger vigoureusement. Cependant, elle cessait immédiatement de répondre lorsqu'une image rivale (un damier) était présentée à l'œil gauche, ce qui effaçait toute conscience de Bill Clinton[63]. Le visage de l'ancien président s'affichait toujours sur la rétine de l'œil droit, mais en raison de la rivalité binoculaire, les régions visuelles de haut niveau, où se concocte la prise de conscience, ne parvenaient plus à relayer cette information.

Ces expériences mettent en évidence les mécanismes neuronaux de l'embrasement cortical. Un tiers de seconde après qu'une image a été vue, certains neurones du cortex temporal antérieur se mettent à décharger avec vigueur. Cet embrasement soudain n'est pas juste un signal non spécifique d'attention ou d'excitation globale, puisque chaque concept active des neurones différents. Nous sommes tout simplement en train de visualiser le codage neuronal d'une information consciente. La distribution des cellules actives et inactives compose un code interne qui reflète fidèlement le contenu de la perception subjective.

Ce code conscient est stable et reproductible : ce sont toujours les mêmes neurones qui déchargent dès que le patient pense à Bill Clinton. Il suffit, pour les activer, d'*imaginer* le visage du président : la plupart des neurones du cortex temporal antérieur répondent avec la même sélectivité aux images réelles et aux images mentales[64]. La mémoire suffit également à les réactiver. Une cellule, qui déchargeait chaque fois que le patient visionnait une vidéo de la famille Simpson, s'activait à nouveau, dans le noir absolu, dès que le patient se remémorait cette vidéo.

Bien que chaque neurone de cette région reflète un fragment de nos pensées, on aurait tort d'en conclure qu'à l'inverse chacune de nos pensées est codée par une seule cellule. L'information consciente est très certainement distribuée dans des myriades de neurones. Imaginez plusieurs millions de neurones, répartis dans l'ensemble des aires associatives du cortex et qui, chacun, codent pour un fragment de la scène

actuellement visible. Leurs décharges synchrones engendrent des potentiels électriques macroscopiques suffisamment importants pour être mesurés par des électrodes disposées à la surface du cerveau ou même du crâne. La décharge d'un seul neurone n'est pas détectable à distance. C'est seulement parce que la perception consciente mobilise de vastes assemblées de neurones que nous pouvons, dans une certaine mesure, déterminer si une personne a vu un visage ou une maison en s'appuyant sur la topographie des potentiels électriques émis par son cortex visuel[65]. De la même manière, la position et même le nombre d'objets qu'une personne a vus et garde en mémoire peuvent être décodés à partir des ondes qui parcourent le cortex pariétal[66].

Comme ce code conscient reste longtemps stable, même l'IRM, technique imprécise qui fusionne les signaux de millions de neurones, parvient partiellement à le déchiffrer. Dans une expérience récente, lorsqu'une personne voyait un visage ou une maison, l'organisation spatiale de l'activation mesurée par IRM dans les régions antérieures du cortex temporal ventral suffisait à déterminer ce qu'elle avait perçu[67]. Cette configuration codante demeurait stable d'essai en essai, alors que ce n'était pas le cas lors des essais inconscients.

Imaginez à présent que vous soyez réduit à une taille submillimétrique et que vous soyez envoyé dans le cortex. Vous êtes alors entouré de milliers de décharges neuronales. Comment pouvez-vous reconnaître lesquelles de ces décharges codent pour une pensée consciente ? D'après les résultats actuels, trois propriétés distinctives pourraient vous y aider : la *stabilité* à travers le temps, la *reproductibilité* d'essai en essai et l'*invariance* pour des changements superficiels qui conservent le même contenu conceptuel. Ces critères ne sont pas seulement remplis dans le cortex temporal antérieur, mais également dans la région cingulaire postérieure, une autre région de haut niveau, intégrative, située sur la face interne des deux hémisphères. L'activité neuronale évoquée par un objet visuel y reste stable tant que l'objet ne bouge pas, même si les yeux bougent[68]. De plus, les neurones de cette région codent pour la position

des objets dans le monde extérieur, indépendamment de la direction du regard : lorsque nous explorons une scène du regard, leur activité ne varie pas. Cette propriété est loin d'être évidente, car lorsque nous bougeons les yeux, toute l'image se déplace sur le cortex visuel primaire. Sans qu'on sache comment, lorsque l'image parvient au cortex cingulaire, elle a été stabilisée.

La région du cortex cingulaire postérieur où se trouvent ces neurones invariants est étroitement interconnectée avec un site appelé le gyrus parahippocampique, situé juste à côté de l'hippocampe, où l'on a découvert d'autres neurones extraordinaires : les « cellules de lieu[69] ». Ces neurones ne déchargent que lorsque l'animal occupe un certain lieu – par exemple, le coin nord-ouest d'une pièce qui lui est familière. Les cellules de lieu sont elles aussi invariantes aux indices sensoriels : elles continuent même de décharger dans le noir absolu, du moment que l'animal « sait » où il est. On a même démontré qu'elles codent pour l'endroit où l'animal *croit* être. Si on fait croire à un rat qu'il vient d'être « téléporté » ailleurs, simplement en changeant brutalement la couleur des murs, du plancher et du plafond de la cage pour qu'elle ressemble à un autre endroit familier, les neurones de l'hippocampe vacillent brièvement entre les deux interprétations, puis leur activité se stabilise et finit par représenter le nouveau lieu illusoire[70]. Dans cette aire, le décodage des signaux est tellement avancé qu'il est devenu possible, à partir de l'état d'activité de quelques dizaines de neurones, de deviner où se trouve l'animal (ou plutôt, où il croit être) – et ce, même pendant le sommeil, où l'animal ne fait qu'imaginer qu'il se promène dans son environnement habituel. Il n'est pas invraisemblable de penser que, d'ici quelques années, certains codes neuronaux encore plus abstraits, qui cryptent l'étoffe même de nos pensées, pourront être décodés dans le cerveau humain.

En un mot, le mystère de l'expérience subjective est aujourd'hui éventé. Au cours de la perception consciente, les neurophysiologistes n'ont aucune difficulté à enregistrer des décharges neuronales spécifiques d'une image ou d'un concept, et ce dans plusieurs régions du cerveau. Ces

cellules déchargent si et seulement si la personne déclare avoir vu l'image – peu importe que cette perception soit authentique ou imaginaire. Chaque scène visuelle consciente est codée par un état d'activité neuronale reproductible et qui reste stable pendant au moins une demi-seconde, tant que la personne la voit.

Induire une hallucination

Notre quête des signatures de la conscience touche-t-elle à sa fin ? Il est trop tôt pour crier victoire, car un dernier critère doit encore être rempli, celui de la causalité. Pour mériter le label de « signature de la conscience », un signal cérébral doit non seulement apparaître chaque fois que la personne prend conscience de l'information correspondante, mais nous devons également prouver qu'il est la *cause* de la prise de conscience de cette information.

La prédiction est simple : si nous parvenions à recréer l'état d'activité cérébrale qui code pour un certain contenu de conscience, l'observateur devrait ressentir l'état mental correspondant. Par exemple, si un stimulateur, tout droit sorti du film *Matrix*, parvenait à recréer dans leurs moindres détails les décharges neuronales qui circulaient dans notre cerveau la dernière fois que nous avons vu un beau coucher de soleil, nous devrions le percevoir avec la même clarté – une hallucination *totale*, absolument impossible à distinguer de l'impression que nous avions éprouvée à l'époque.

Figure 23. La stimulation magnétique transcrânienne permet de modifier l'activité cérébrale et le contenu de la conscience. Un siècle après les travaux pionniers de Thompson (1910, en haut)

et Magnusson et Stevens (1911, au milieu), la technique est devenue beaucoup plus simple à utiliser (en bas). L'application d'un champ magnétique transitoire déclenche un courant électrique dans le cortex qui vient perturber la perception en cours ou, au contraire, engendrer une hallucination lumineuse. Cette technique démontre ainsi l'existence d'un lien causal entre l'activité cérébrale et l'expérience consciente.

Recréer ainsi un état cérébral peut sembler impossible. Pourtant, nous le faisons chaque nuit. Lorsque nous rêvons, tandis que notre corps repose dans le noir, notre esprit s'envole, simplement parce que notre cerveau est traversé de décharges neuronales organisées qui évoquent des contenus mentaux bien précis. L'enregistrement des neurones d'un rat qui dort montre que le cortex et l'hippocampe « rejouent » les séquences de décharges neuronales qui, quelques heures auparavant, codaient pour les événements de la journée précédente[71]. Chez l'homme, de la même façon, l'activité des aires du cortex au cours du sommeil prédit le contenu du rêve[72]. Si, par exemple, l'activité se concentre sur les aires spécialisées dans la reconnaissance des visages, et que l'on réveille alors le dormeur, il est probable que celui-ci rapporte la présence d'autres personnes dans son rêve.

Ces découvertes fascinantes démontrent l'étroite correspondance entre les états neuronaux et les états mentaux – mais elles ne prouvent toujours pas l'existence d'une relation causale. Prouver qu'une activité neuronale cause une certaine pensée est l'un des problèmes les plus difficiles auxquels sont confrontées les neurosciences contemporaines. Pratiquement toutes les méthodes d'imagerie cérébrale démontrent la corrélation, pas la causalité. Elles ne font que révéler un lien systématique entre l'activation cérébrale et les pensées de la personne scannée, alors qu'il faudrait stimuler le cerveau et observer une modification correspondante du contenu mental pour démontrer une authentique relation de cause à effet. À l'heure actuelle, deux techniques permettent cependant d'aller plus loin et de stimuler le cerveau humain sans danger et de façon réversible.

Chez n'importe quelle personne volontaire, nous pouvons activer le cerveau à partir de l'extérieur, à l'aide d'une technique appelée « stimulation magnétique transcrânienne » (TMS). Inventée dès le début du XX^e^ siècle[73], puis modernisée grâce aux progrès de l'électronique[74], cette méthode est désormais très répandue dans les laboratoires (figure 23). Une batterie d'accumulateurs décharge un violent courant électrique dans une boucle de fil placée à la surface de la tête. Ce courant engendre un champ magnétique qui traverse le crâne et induit, à son tour, un courant en un point bien précis du cortex sous-jacent. De strictes règles de sécurité assurent l'innocuité de cette technique. Seul un clic nettement audible et, de temps à autre, une désagréable contraction musculaire témoignent de la décharge. De cette manière, pratiquement n'importe quelle région du cortex peut être stimulée à l'instant désiré.

Si l'on souhaite une meilleure précision spatiale, l'alternative consiste à stimuler directement les neurones à l'aide d'électrodes disposées à la surface du cortex ou insérées dans le cerveau. Bien évidemment, cette option n'est disponible que chez de rares patients atteints d'épilepsie, de la maladie de Parkinson ou d'autres troubles moteurs que les médecins explorent de plus en plus fréquemment à l'aide de ces fines électrodes. Avec l'accord du patient, de petits courants, éventuellement synchronisés avec des stimuli visuels ou auditifs, peuvent ainsi être envoyés. On peut même stimuler les neurones au cours d'une intervention chirurgicale. Puisque le cerveau est dépourvu de récepteurs sensibles à la douleur, la stimulation électrique est totalement indolore. Elle joue un rôle essentiel dans la panoplie du neurochirurgien, car c'est elle qui lui permet de détecter les aires cérébrales du langage, que son bistouri doit épargner. Leur stimulation interrompt en effet la production de la parole. Certaines expériences sont troublantes : allongé sur la table d'opération, le crâne à moitié ouvert, mais parfaitement éveillé, un patient raconte à haute voix ce qu'il ressent, tandis qu'une électrode, profondément implantée dans son cerveau, lui injecte un courant électrique…

Les résultats de ces travaux sont fondamentaux. En effet, qu'ils soient menés chez l'homme ou chez le singe, ils établissent une relation causale directe entre l'activité neuronale et la perception consciente. La simple stimulation d'un circuit nerveux, en l'absence de tout événement objectif, suffit à induire un sentiment subjectif dont le contenu varie en fonction de l'endroit stimulé. Ainsi, la stimulation magnétique du cortex visuel, dans le noir absolu, engendre une impression de lumière, que l'on appelle techniquement un phosphène : dès l'application du courant électrique, un petit halo de lumière apparaît et se déplace en fonction de l'endroit stimulé. Que l'on place à présent la boucle de stimulation au-dessus de l'aire cérébrale MT/V5, qui répond au mouvement, et la perception change : l'observateur dont on triture ainsi le cerveau dit voir un mouvement furtif. Induire une impression de couleur ne pose guère plus de difficulté.

Les enregistrements neuronaux ont établi depuis belle lurette que chaque paramètre de la scène visuelle se projette en un point donné du cortex. Différents secteurs du cortex occipital abritent des mosaïques de neurones sensibles à la forme, au mouvement ou à la couleur. La stimulation cérébrale démontre que cette relation entre la perception et les décharges neuronales est causale. Une décharge focale, en l'absence de toute sensation réelle, suffit à évoquer un embryon de conscience doté d'attributs bien précis de couleur ou d'illumination.

Avec des électrodes intracrâniennes, l'effet peut devenir plus spécifique[75]. Stimuler une électrode située en regard de l'aire des visages, dans le cortex visuel ventral, suffit à induire la perception subjective d'un visage. Déplacer le courant en direction des régions antérieures du lobe temporal éveille des souvenirs complexes extraits de la mémoire du patient. L'un d'entre eux sentait une odeur de pain grillé, l'autre voyait et entendait un grand orchestre symphonique. D'autres encore éprouvaient des sensations remarquablement vives et riches, proches de l'état de rêve : ils se voyaient en train d'accoucher, de visionner un film d'horreur ou de revivre un épisode de leur enfance. Wilder Penfield, neurochirurgien

canadien et pionnier de cette technique, en a conclu que nos microcircuits corticaux contiennent, à l'état latent, un enregistrement complet des épisodes majeurs et mineurs de nos vies, prêts à resurgir lors d'une stimulation cérébrale.

Une exploration systématique suggère que chaque région corticale abrite des connaissances particulières. Prenez l'insula, un pli de cortex profondément enterré sous une partie des lobes temporaux et frontaux. Stimulez cette région, et vous évoquerez toute une panoplie d'effets désagréables, depuis l'impression d'étouffer, de se brûler ou d'être piqué, jusqu'à des bouffées de chaleur, des nausées ou une sensation de chute sans fin[76]. Enfoncez l'électrode quelques centimètres plus profondément, en direction du noyau sous-thalamique, et la même impulsion électrique déclenchera instantanément un état dépressif caractéristique, avec son cortège d'idées noires, de pleurs et de sanglots, sa voix monocorde et sa posture corporelle maussade. Titillez le lobe pariétal, et vous évoquerez des vertiges ou même une « sortie du corps », l'étrange impression de léviter jusqu'au plafond et de voir son corps de là-haut[77].

Si vous doutiez encore que chacune de vos pensées provient de l'activité de votre cerveau, ces exemples devraient vous convaincre. La stimulation cérébrale semble pouvoir évoquer pratiquement n'importe quelle sensation consciente, depuis l'orgasme jusqu'au sentiment de déjà-vu. Toutefois, ces faits ne permettent pas de tirer de conclusions fermes sur les causes de la conscience. En effet, l'activité neuronale démarre à l'endroit stimulé, mais elle se propage aussitôt à d'autres circuits, ce qui nous empêche de déterminer précisément lequel de ces effets est responsable de la sensation consciente. Des données récentes suggèrent que l'activité de départ n'est pas consciente : ce n'est que si l'activation s'étend à des régions distantes des lobes pariétaux et préfrontaux que survient un sentiment subjectif de conscience.

Considérez l'étonnante dissociation récemment décrite par le neuroscientifique français Michel Desmurget[78]. Lorsqu'il stimulait le cortex prémoteur à un niveau suffisamment bas, au cours d'une

intervention chirurgicale, la main du patient bougeait, mais celui-ci niait avoir perçu quoi que ce soit (il ne pouvait pas voir ses mains). Inversement, lorsque Desmurget stimulait le cortex pariétal inférieur, le patient disait ressentir une envie irrépressible de bouger, et lorsque le courant augmentait, il jurait avoir bougé la main – alors qu'en réalité son corps était resté immobile.

Ces résultats sont riches d'implications : tous les circuits ne sont pas d'égale importance pour la prise de conscience. Les circuits sensoriels et moteurs périphériques peuvent s'activer sans engendrer de sensation consciente. Par contre, les aires supérieures du cortex temporal, pariétal et préfrontal sont étroitement associées à l'expérience consciente, dans la mesure où leur stimulation peut induire des hallucinations sans aucune réalité objective.

L'étape suivante consiste à créer des stimulations minimalement différentes, mais tantôt perçues et tantôt non, et d'examiner en quoi leurs effets diffèrent. Comme tant de scientifiques avant eux, Paul Taylor, Vincent Walsh et Martin Eimer ont utilisé la stimulation magnétique pour induire des phosphènes visuels – des lumières hallucinées, suscitées uniquement par l'activité du cortex[79]. Cependant, ils ont également ajusté l'intensité du courant injecté jusqu'à ce que le sujet dise ne voir le point lumineux qu'une fois sur deux. Ils ont alors suivi l'activité cérébrale évoquée par ces stimulations au seuil, en enregistrant l'EEG milliseconde par milliseconde après la décharge initiale.

Les résultats se sont avérés… lumineux. Après la stimulation, l'activité ne restait pas confinée à son point de départ, mais elle se propageait au reste du cerveau. Cette propagation, initialement, n'avait pas de rapport avec la conscience, puisque pendant 160 millisecondes elle se déroulait à l'identique dans les essais visibles et invisibles. C'est seulement après cette période que notre fameuse onde P3 apparaissait à la surface de la tête, avec une intensité bien plus grande dans les essais perçus que dans les essais non perçus. Sa topographie était strictement normale, et seul son point de départ, dès 200 millisecondes, survenait un peu plus tôt que

d'habitude – sans doute parce que l'activation directe du cortex, contrairement à la stimulation visuelle habituelle, court-circuite les étapes initiales de la vision rétinienne et accélère d'autant l'accès à la conscience.

La stimulation cérébrale démontre donc une relation causale entre l'activité électrique du cortex et l'expérience subjective. Même dans le noir, les yeux fermés, une impulsion électrique qui induit une décharge du cortex visuel peut provoquer une impression consciente. Mais l'expérience de Taylor et de ses collègues indique que cette relation n'est pas directe : ce n'est pas l'activité neuronale locale qui suscite la conscience, mais sa propagation à des régions distantes du cortex. Une fois de plus, c'est la partie tardive du train de décharges, et sa diffusion au sein d'un réseau cortical global, qui induit une perception consciente. L'activité neuronale doit se répandre dans le cortex, et peut-être même revenir aux aires sensorielles de départ, avant que nous ne ressentions l'impression de « voir ».

Détruire la conscience

Nous parvenons à créer une perception consciente, mais pouvons-nous également la détruire ? Si c'est bien l'activation tardive du réseau cérébral global qui cause chacune de nos pensées conscientes, alors l'empêcher de s'activer devrait éradiquer toute conscience. Conceptuellement, l'expérience est simple. Présentez d'abord un stimulus visible, bien au-dessus du seuil minimal de conscience, puis utilisez une impulsion de courant électrique pour perturber le réseau à longue distance qui sous-tend la conscience : le sujet devrait jurer qu'aucun stimulus n'était présent. Imaginez à présent que l'impulsion ne soit pas simplement destructrice, mais parvienne à remplacer un état global d'activité neuronale par un autre. Le sujet devrait dire qu'il a pris conscience, non de la réalité, mais de l'information que l'on a induite en lui.

Bien que ces idées puissent sembler issues d'un livre de science-fiction, plusieurs variantes de cette expérience de pensée ont déjà été menées avec succès. L'une d'entre elles utilisait un double stimulateur afin d'induire des courants dans deux régions du cerveau à des moments arbitraires. La recette était simple : exciter d'abord l'aire MT/V5, et vérifier que cette impulsion seule engendrait une impression de mouvement ; puis appliquer une seconde bouffée de courant dans le cortex visuel primaire. La seconde impulsion éradiquait la perception consciente que le premier stimulus, seul, parvenait à susciter. Ainsi, l'activation initiale ne suffisait pas à rendre l'information consciente : il fallait encore que cette activité revienne au cortex visuel primaire avant d'être perçue[80]. La conscience réside dans les boucles du cortex : c'est la réverbération de l'activité neuronale dans les méandres des connexions corticales qui cause chacune de nos sensations conscientes.

Plus fascinant encore, la combinaison d'une stimulation corticale avec la présentation d'images réelles engendre de nouvelles illusions. Ainsi, la stimulation du cortex visuel, un cinquième de seconde après la brève apparition d'une photo sur la rétine, réveille sa perception consciente : le participant dit voir l'image deux fois, ce qui confirme qu'une trace électrique inconsciente se promenait encore dans son cortex près de 200 millisecondes plus tard[81]. L'effet est particulièrement net lorsqu'on demande à la personne de conserver l'image en mémoire : quand nous gardons une image à l'esprit, notre cerveau la maintient littéralement en tête dans le taux de décharge de nos neurones visuels, à un niveau subconscient, prête à resurgir à la moindre impulsion[82].

L'étendue du réseau cérébral global qui engendre une impression consciente fait encore l'objet d'un débat. Selon le neurophysiologiste hollandais Viktor Lamme, dès que deux régions forment une boucle, en sorte que A projette vers B et que B, à son tour, réponde à A, cette réverbération suffit à induire une prise de conscience[83]. Elle crée un « traitement récurrent » des données, qui réinjecte sans cesse l'information dans le circuit. « Nous pourrions même définir la conscience comme un

traitement récurrent », écrit Lamme[84]. Pour lui, chaque boucle cérébrale engendre un fragment de conscience. Personnellement, je doute que cette idée soit juste. Notre cortex est plein de boucles : la communication bidirectionnelle y abonde à toutes les échelles, depuis les microcircuits locaux de l'ordre du millimètre jusqu'aux autoroutes cérébrales longues de plusieurs centimètres. Il serait très surprenant que chacune de ces boucles, si petite soit-elle, suffise à évoquer un brin de conscience (si tant est que cette expression ait un sens[85]). Il me paraît plus plausible que la réverbération de l'activité soit une condition nécessaire mais pas suffisante de l'émergence de la conscience. Seules les boucles à longue distance qui amènent l'information aux régions pariétales et préfrontales créeraient un code conscient.

Quel serait alors le rôle des boucles locales ? Elles sont vraisemblablement indispensables aux opérations visuelles *inconscientes* qui assemblent les fragments d'une scène visuelle en un tout cohérent[86]. Avec leurs petits champs récepteurs, les neurones visuels ne peuvent pas appréhender immédiatement les propriétés globales d'une image, telle que la présence d'une grande ombre (souvenez-vous de l'illusion de l'ombre sur l'échiquier, présentée en figure 10). De nombreux échanges et itérations sont nécessaires pour que ces propriétés globales soient reconnues[87].

Certains scientifiques prétendent que ce sont les petites boucles qui induisent la conscience, parce que leur activité disparaît pendant l'anesthésie[88]. Pourtant, ces données ne permettent guère de conclure : la réverbération de l'activité semble être l'une des premières choses à disparaître lorsque le cerveau baigne dans les anesthésiques, mais ce pourrait être la conséquence plutôt que la cause de la perte de conscience.

Les expériences d'interférence induite par la stimulation cérébrale racontent une autre histoire. Si l'on interfère avec les boucles locales au cortex visuel, environ 60 millisecondes après l'apparition d'une image, cette impulsion affecte aussi bien la prise de conscience que le traitement *inconscient*[89]. En effet, une telle stimulation affecte la « vision aveugle »,

c'est-à-dire la capacité d'émettre une réponse appropriée à des stimuli subliminaux, aussi bien que la vision consciente. Cette observation confirme que les premières étapes du traitement cortical, où l'information circule dans des boucles locales, ne sont pas exclusivement associées à la perception consciente. Elles réalisent des opérations inconscientes qui ne font que préparer le cerveau à la prise de conscience qui surviendra bien plus tard.

Si ces déductions sont justes, l'accès à la conscience émerge de l'activité synchronisée de multiples aires pariétales et préfrontales. Contrecarrer ces régions devrait donc avoir un impact majeur. Effectivement, toute une série d'expériences confirme que la sidération des régions frontales ou pariétales cause une cécité transitoire. Pratiquement toutes les illusions qui permettent de rendre une image invisible, comme le masquage ou l'inattention, sont renforcées lorsque l'on perturbe en même temps certaines régions pariétales droite ou gauche[90]. Ainsi, une tache de couleur pâle mais bien visible disparaît totalement de la conscience lorsqu'on interfère avec l'activité des aires pariétales[91].

Plus étonnante encore est l'expérience réalisée par Hakwan Lau et son équipe à l'Université d'Oxford, et qui consistait à sidérer transitoirement le cortex préfrontal[92]. Pas moins de 600 impulsions bombardaient chaque aire préfrontale dorso-latérale, d'abord dans l'hémisphère gauche, puis à droite. Les impulsions survenaient sous la forme de « bouffées thêta », c'est-à-dire cinq fois par seconde, une cadence spécialement conçue pour annihiler le « rythme thêta », l'une des fréquences privilégiées par laquelle les aires corticales se communiquent des informations à distance. L'impact de ce bombardement bilatéral ressemblait à une lobotomie virtuelle : pendant près de vingt minutes, les lobes frontaux étaient inhibés, ce qui laissait suffisamment de temps pour en mesurer les effets sur le comportement.

Objectivement, rien n'avait changé : les participants, bien qu'assommés de décharges, continuaient de répondre aussi bien qu'avant. On leur avait demandé d'identifier si une forme, présentée au seuil de

conscience, était un carré ou un losange, et leur performance restait bonne. Subjectivement, par contre, c'était une autre histoire. Pendant de longues minutes, ils perdaient toute confiance dans leur jugement. Ils devenaient incapables d'évaluer leur propre compétence, et ils avaient l'impression subjective que leur vision n'était plus fiable. Ils se comportaient un peu comme le fameux « zombie » dont discutent les philosophes : leur perception et leurs actions étaient normales, mais ils n'avaient plus conscience de leurs performances.

Avant la stimulation, les participants parvenaient à émettre des jugements subjectifs de visibilité qui correspondaient étroitement à leurs performances objectives. Comme vous et moi, quand ils avaient l'impression d'avoir vu le stimulus, ils l'identifiaient presque toujours à la perfection, alors que quand ils disaient n'avoir rien vu, leurs réponses étaient proches du hasard. Pendant la « lobotomie virtuelle » du cortex préfrontal, par contre, cette corrélation disparaissait : les réponses subjectives se dissociaient du comportement.

Ce comportement correspond à la définition de la *vision aveugle* – une dissociation entre la perception consciente et le comportement objectif. Un tel état survient normalement à la suite d'une lésion cérébrale majeure, mais il peut être transitoirement simulé en sidérant les deux lobes frontaux. Une conclusion s'impose : ces régions jouent un rôle majeur dans les boucles corticales qui engendrent la conscience.

Une chose qui pense

> « Mais qu'est-ce donc que je suis ? Une chose qui pense. Qu'est-ce qu'une chose qui pense ? C'est une chose qui doute, qui entend, qui conçoit, qui affirme, qui nie, qui veut, qui ne veut pas, qui imagine aussi, et qui sent. »
>
> René DESCARTES, *Méditation seconde* (1641).

Toutes ces observations expérimentales sur la conscience, et elles sont nombreuses, convergent vers une conclusion réductionniste. Chacun de nos états de conscience, depuis l'écoute d'un orchestre jusqu'à l'odeur du pain grillé, provient d'une seule source : l'activation des vastes circuits cérébraux du cortex. Au cours de la perception consciente, des groupes de neurones se mettent à décharger de manière coordonnée, d'abord au sein des aires spécialisées, puis dans la quasi-totalité du cortex. Ces décharges finissent par envahir les lobes pariétaux et frontaux, tout en restant synchrones avec les aires sensorielles de départ. C'est à cet instant, quand le réseau global s'embrase, que s'allume l'étincelle de la conscience.

Au cours de ce chapitre, nous avons découvert pas moins de quatre signatures fiables de la conscience – des marqueurs physiologiques qui indiquent qu'une personne a pris conscience d'un objet de pensée. En premier lieu, une sensation devient consciente quand elle conduit à l'embrasement soudain de circuits pariétaux et préfrontaux. En second, dans l'électroencéphalogramme, l'accès à la conscience se traduit par l'émergence d'une onde lente appelée onde P3, qui survient près d'un tiers de seconde après l'apparition d'une image sur la rétine. Troisièmement, l'embrasement conscient s'accompagne d'une bouffée d'oscillations de haute fréquence. Quatrièmement, enfin, de nombreuses régions distantes du cortex se mettent à échanger des messages synchronisés, dans les deux sens, jusqu'à former un réseau cérébral global.

Il se pourrait que l'un ou l'autre de ces événements ne soit qu'un épiphénomène de la conscience, un peu comme le sifflet de la locomotive : il l'accompagne, mais ne lui apporte rien. En neurosciences, établir une relation de causalité reste difficile. Pourtant, quelques expériences pionnières ont commencé à montrer que la perturbation des aires corticales de haut niveau peut altérer la perception consciente tout en laissant intact le traitement non conscient. D'autres recherches, en stimulant le cerveau, sont parvenues à induire des hallucinations de lumière ou de mouvement. Bien qu'elles demeurent trop rudimentaires pour peindre le tableau détaillé d'un état de conscience, ces études ne

laissent aucun doute sur le fond : l'entrée en activité de certains neurones peut causer une pensée ou, tout aussi facilement, en détruire une.

La plupart des neuroscientifiques accordent foi à l'idée du « cerveau dans une cuve » introduite par le philosophe contemporain Hilary Putnam et popularisée par le film *Matrix*. Selon cette expérience de pensée, rien n'empêcherait un cerveau détaché du corps et placé dans une cuve de continuer à ressentir des sensations conscientes. En stimulant les neurones adéquats, et en faisant se taire les autres, on devrait pouvoir recréer, à tout moment, n'importe quelle hallucination. Les avalanches neuronales doivent déclencher des symphonies mentales.

La technologie actuelle reste toutefois bien loin de la fiction des frères Wachowski. Nous ne savons pas encore contrôler les milliards de neurones nécessaires pour peindre, sur la surface du cortex, l'équivalent neuronal d'une rue bondée de Chicago ou d'un coucher de soleil aux Bahamas. Ce rêve est-il définitivement hors d'atteinte ? Je ne parierai pas là-dessus. Dans les mains des bio-ingénieurs d'aujourd'hui, motivés par l'objectif de redonner aux patients aveugles, paralysés ou parkinsoniens une vie normale, les neurotechnologies progressent rapidement.

Il existe déjà, par exemple, des puces en silicone dotées de milliers d'électrodes que l'on applique sur le cortex des animaux et qui augmentent massivement la bande passante des signaux échangés entre le cerveau et un ordinateur. Plus fascinant encore, les avancées spectaculaires de l'optogénétique permettent de piloter des neurones par la lumière plutôt que par le courant électrique. Au cœur de cette technique se trouve la découverte, chez certaines algues et bactéries, de molécules sensibles à la lumière. Ces « opsines » convertissent les photons en signaux électriques, qui sont la monnaie d'échange des neurones. Les gènes qui codent pour les opsines sont connus, et leurs propriétés peuvent être modifiées par génie génétique. Il suffit désormais d'injecter dans le cerveau un virus qui porte ces gènes et de restreindre leur expression à un sous-groupe de neurones pour ajouter des photorécepteurs à l'endroit souhaité du cerveau. Au plus profond du cortex, dans des régions

obscures et habituellement insensibles à la lumière, un laser déclenche soudain une avalanche de décharges avec une précision inférieure au millième de seconde.

À l'aide de l'optogénétique, les neuroscientifiques peuvent aujourd'hui activer ou inhiber sélectivement n'importe quel circuit cérébral[93]. La technique a même été utilisée pour réveiller une souris endormie, en stimulant son hypothalamus[94]. Bientôt, elle permettra d'induire des états d'activité plus différenciés et, pourquoi pas, d'évoquer ainsi une perception consciente. Les dix prochaines années verront certainement des avancées majeures dans la compréhension du code neuronal qui fonde notre vie mentale.

CHAPITRE 5

Théoriser la conscience

Que signifient les signatures de la conscience que nous avons découvertes ? Pourquoi ces mesures objectives reflètent-elles l'introspection subjective ? La théorie de l'« espace de travail neuronal global », fruit de quinze ans de réflexion avec Jean-Pierre Changeux, Lionel Naccache et d'autres membres de mon laboratoire, tente de répondre à ces questions. L'hypothèse de départ est simple : la conscience n'est rien d'autre que le partage global d'une information. Le cerveau humain possède des réseaux de connexions à longue distance, particulièrement au sein du cortex préfrontal, dont le rôle est de sélectionner les informations les plus pertinentes et de les diffuser à l'ensemble du cerveau. La conscience est un dispositif évolué qui nous permet de maintenir l'information en ligne. Une fois qu'une information est devenue consciente, elle peut être redirigée vers n'importe quelle autre région du cerveau, en fonction de nos objectifs, et donc être nommée, évaluée, mémorisée ou incorporée à nos plans d'action. La simulation informatique de réseaux neuronaux montre que cette hypothèse rend bien compte des signatures neurophysiologiques observées dans nos expériences. Elle explique également pourquoi une si grande part de nos connaissances reste inaccessible à la conscience.

« Je considérerai les actions et les appétits humains comme s'il était question de lignes, de surfaces et de solides. »

Baruch SPINOZA, *Éthique* (1677).

La découverte de signatures de la conscience est une avancée importante. Cependant, l'observation que certaines ondes cérébrales et autres décharges neuronales sont systématiquement liées à la prise de conscience n'explique toujours pas ce qu'est réellement la conscience ni pourquoi elle survient. Pourquoi une activité tardive du cortex préfrontal, un embrasement cortical ou une synchronisation à longue distance devraient-ils se traduire par un changement d'état mental ? Comment ces événements biologiques, si complexes soient-ils, engendrent-ils une pensée ? Pourquoi les neurones de l'aire V4, quand ils déchargent, induisent-ils une sensation de couleur, et les neurones de MT/V5 un sentiment de mouvement ? Bien que les neurosciences aient identifié, empiriquement, de nombreuses correspondances entre l'activité cérébrale et la vie mentale, un gouffre conceptuel semble toujours séparer les états du cerveau de ceux de l'esprit.

En l'absence d'une théorie explicite, la quête des corrélats neuronaux de la conscience risque d'être aussi vaine que la proposition de Descartes qui faisait de la glande pinéale le siège de l'âme. Cette hypothèse ne convainc guère, car elle se contente de perpétuer une division que la science de la conscience est censée dissiper : l'intuition que le neural et le mental appartiennent à des domaines entièrement disjoints. La simple observation d'une corrélation systématique entre ces deux domaines ne saurait suffire. Ce qu'il nous faut, c'est un cadre théorique qui englobe les deux phénomènes, une série de lois qui expliquent, pas à pas, comment les états d'activité du cerveau produisent des états mentaux.

Les énigmes qui confondent les neurosciences contemporaines ne sont pourtant pas très différentes de celles que les physiciens ont fini par résoudre. Comment les propriétés de la matière émergent-elles d'interactions entre atomes ? D'où provient la solidité d'une table, si elle n'est faite que de vide peuplé de rares atomes de carbone, d'oxygène et d'hydrogène ? Qu'est-ce qu'un liquide ? Un solide ? Un cristal ? Un gaz ? Une flamme ? Comment leurs caractéristiques résultent-elles d'un tissu d'atomes en perpétuel mouvement ? Si la physique a aujourd'hui réponse à toutes ces questions, cette compréhension profonde a exigé d'importants efforts théoriques. Pour comprendre les propriétés « émergentes » de la matière, il ne suffit pas d'en décrire, de plus en plus précisément, les éléments constitutifs. Il faut également rassembler tous ces faits en une théorie synthétique. La théorie cinétique des gaz, édifiée par James Clerk Maxwell et Ludwig Boltzmann au XIX^{e} siècle, a établi comment les variables macroscopiques de pression et de température émergent du mouvement des atomes d'un gaz. Ce n'était que la première d'une longue série de modèles mathématiques de la matière – une chaîne réductionniste qui rend à présent compte de substances exotiques aussi diverses que les colles, les bulles de savon, l'eau qui percole à travers le marc de notre café ou le plasma qui anime notre soleil.

Les neurosciences ont besoin d'un intense effort théorique du même ordre afin de réduire l'écart qui sépare la matière du cerveau de l'étoffe

des pensées qu'il engendre. Aucun appareil d'imagerie cérébrale ne visualisera jamais les décharges des cent mille millions de neurones que contient le cerveau humain au moment où il accède à la conscience – et même s'il le pouvait, nous n'y comprendrions toujours rien. Seule une théorie mathématique peut expliquer comment le mental se réduit au neural. Les neurosciences attendent encore leur Maxwell ou leur Boltzmann qui formulera les lois de mise en liaison de ces deux domaines. Cela ne sera pas chose facile, car la « matière condensée » du cerveau est l'objet le plus complexe de l'univers. Rien à voir, en effet, avec la structure d'un gaz ou d'un cristal : dans le cerveau humain s'emboîte, comme une série de poupées russes, toute une hiérarchie de niveaux d'organisation. La pensée émerge d'une architecture sophistiquée de routines mentales, un assemblage de processeurs élémentaires interconnectés en circuits distribués dans plusieurs régions du cerveau, eux-mêmes formés de dizaines de types de neurones. Chaque neurone, avec ses dizaines de milliers de synapses, est à lui seul un univers de molécules en interaction permanente, qui donnera sans doute du travail aux modélisateurs pour quelques siècles.

En dépit de ces difficultés, au cours des quinze dernières années, mes collègues Jean-Pierre Changeux, Lionel Naccache et moi-même avons lancé quelques pistes de mise en relation psycho-neurale. Nous avons ébauché un embryon de théorie de la conscience, que nous avons appelée la théorie de l'« espace de travail neuronal global ». Dans ce chapitre, j'espère vous convaincre que, même si ses lois mathématiques précises nous échappent encore, nous commençons à entrevoir à quoi pourrait ressembler une théorie intégrative de la conscience, comment celle-ci surgit de l'activité coordonnée de réseaux neuronaux, et pourquoi elle engendre les signatures que nous avons identifiées dans nos expériences.

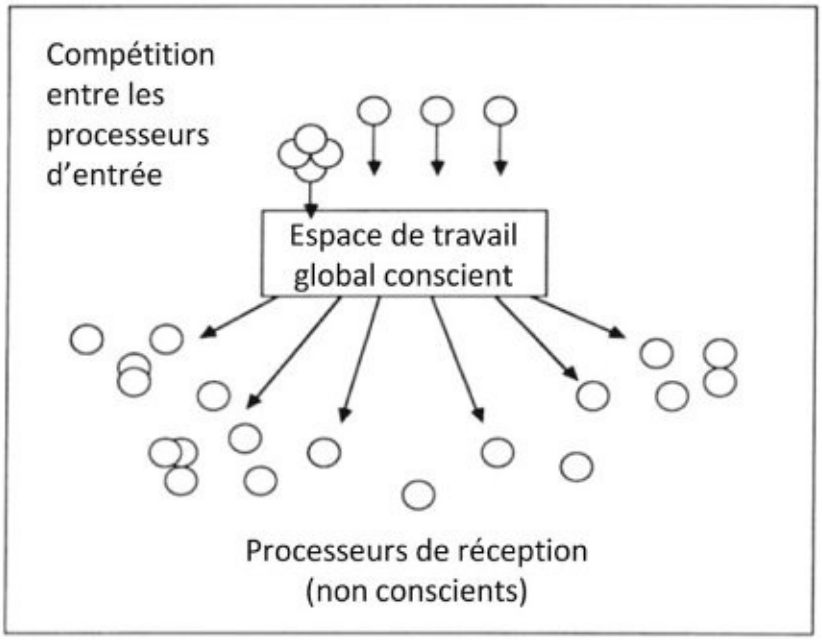
La conscience est une diffusion globale

À quoi sert la conscience ? Quelle sorte de calculs permet-elle ? Quelle est sa raison d'être, son rôle dans le traitement cérébral de l'information ? Ma proposition tient en peu de mots[1]. Lorsque nous prenons conscience d'une information, celle-ci entre dans un système de stockage qui la maintient en ligne et la rend disponible au reste du cerveau. Parmi les millions de représentations mentales inconscientes qui, à tout instant, traversent nos circuits cérébraux, l'une d'entre elles est sélectionnée pour sa pertinence par rapport à nos buts actuels. La conscience est le dispositif qui la stocke et la rend disponible à tous les systèmes de décision de haut niveau. Nous possédons un poste d'aiguillage, une architecture cérébrale qui a évolué afin d'extraire l'information pertinente et de l'expédier aux systèmes adéquats. Le psychologue Bernard Baars l'appelle « espace de travail global » : un système interne, découplé du monde extérieur, au sein duquel nous sommes libres de créer nos propres images mentales et de les transmettre à n'importe quel processeur cérébral spécialisé (figure 24).

D'après cette théorie, la conscience n'est rien d'autre que la diffusion globale d'une information à l'échelle de tout le cerveau. Tout ce dont nous prenons conscience, nous pouvons le garder à l'esprit longtemps après qu'il a disparu de nos organes sensoriels. Une fois l'information acheminée vers l'espace de travail, elle y reste stable, indépendamment du moment et du lieu où nous l'avions initialement perçue. Nous pouvons alors l'utiliser de mille manières, et en particulier l'expédier aux aires du langage, donc la nommer : voilà pourquoi la capacité de formuler en mots ce que nous avons perçu est un excellent critère de conscience. Nous pouvons également la stocker dans notre mémoire à long terme ou l'intégrer à nos plans d'action, quels qu'ils soient. La dissémination flexible de l'information caractérise l'état conscient.

Le concept d'espace de travail mental est issu d'une synthèse de plusieurs idées théoriques. Dès 1870, le philosophe français Hippolyte Taine introduisit la métaphore du « théâtre de la conscience[2] ». L'esprit conscient, expliquait-il, ressemble à un théâtre si petit qu'un seul acteur s'y produit :

Baars, 1989



Dehaene et Changeux, 1998

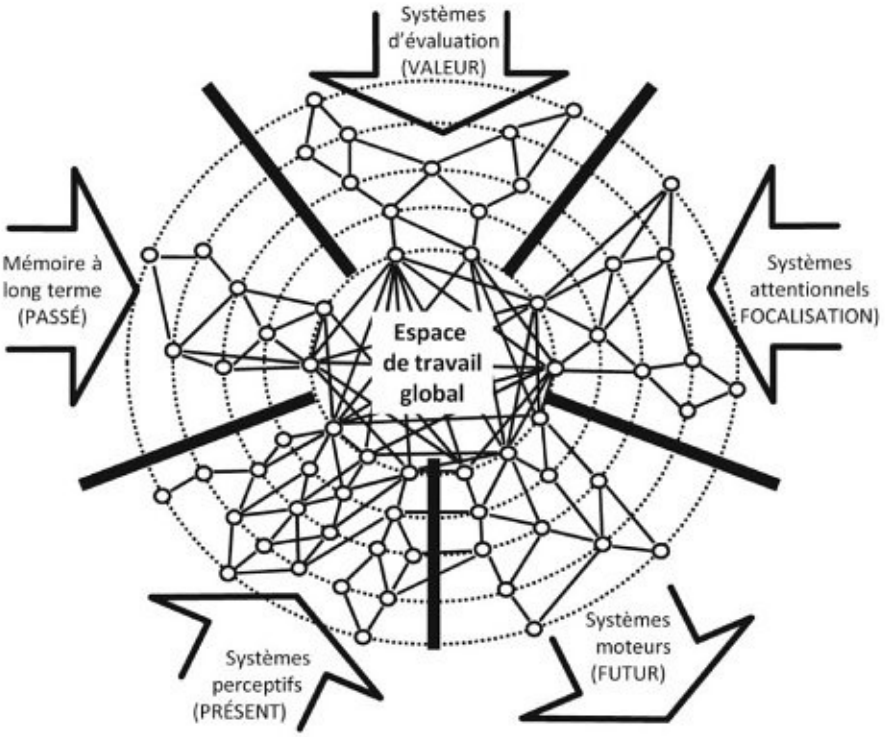


Figure 24. La théorie de l'espace de travail neuronal global propose que ce que nous appelons « conscience » correspond à la diffusion globale de l'information. Le cerveau contient des dizaines de processeurs spécialisés (représentés par des cercles). Un système de communication à longue distance leur permet d'échanger leurs informations. À chaque instant, l'espace de travail sélectionne un jeu de processeurs, forme une représentation cohérente de l'information qu'ils contiennent, la maintient en ligne et la transmet à n'importe quel autre processeur. Dès qu'une information accède à l'espace de travail, elle devient consciente.

> « 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. Au-delà de ces groupes, dans les coulisses et l'arrière-fond lointain, se trouve une multitude de formes obscures qu'un appel soudain amène parfois sur la scène ou même jusque sous les feux de la rampe, et des évolutions inconnues s'opèrent incessamment dans cette fourmilière d'acteurs de tous ordres pour fournir les coryphées qui tour à tour, comme en une lanterne magique, viennent défiler devant nos yeux. »

Des décennies avant Freud, la métaphore de Taine impliquait déjà que, dans l'arrière-scène de la conscience, notre esprit abrite une extraordinaire diversité de processeurs inconscients. Quelle énorme équipe pour un *one-man-show* ! À tout instant, le contenu de la conscience émerge de myriades d'opérations invisibles. En coulisse se produit un ballet de techniciens condamnés à un perpétuel anonymat.

Le philosophe Daniel Dennett a souligné les dangers de l'allégorie du théâtre conscient, car elle peut nous entraîner à commettre une erreur majeure : le sophisme de l'*homunculus*[3]. Si la conscience est un théâtre, qui en sont les spectateurs ? Faut-il imaginer de petits bonshommes

(*homunculi*) aux commandes du cerveau ? Ont-ils, eux aussi, de tout petits cerveaux, pourvus à leur tour d'un théâtre de poche ? Et, si oui, qui assiste à leur représentation ? Raisonnement par l'absurde ! La science de la conscience doit absolument résister à cette métaphore circulaire, digne seulement d'un dessin animé de Walt Disney. Aucun personnage ne tient les rênes de notre cerveau, n'en consulte les écrans ni n'en commande les gestes. Il n'existe aucun « moi » qui contemple le théâtre de la conscience : c'est le théâtre lui-même qui constitue le mécanisme du « moi ». La métaphore du théâtre n'est pas intrinsèquement nocive, mais il faut écarter l'idée d'une audience intelligente, et la remplacer par des opérations explicites de nature algorithmique. Comme Dennett le dit si bien, « on élimine l'*homunculus* en organisant des armées d'idiots pour faire le travail[4] ».

La version de l'espace de travail qu'a introduite Bernard Baars se passe de tout *homunculus*. L'audience n'est pas constituée d'êtres humains dans toute leur complexité, mais d'un ensemble de processeurs inconscients qui reçoivent le message conscient et qui le prennent en compte, chacun selon ses compétences. Une intelligence collective émerge de l'échange global de messages sélectionnés pour leur pertinence. Cette idée n'est pas neuve : elle remonte aux débuts de l'intelligence artificielle, quand les chercheurs proposaient que les sous-programmes échangent des informations par le biais d'un « tableau noir », une structure de données semblable au « presse-papiers » des ordinateurs actuels. L'espace de travail est le presse-papiers de la conscience.

L'étroite scène qu'imaginait Taine, trop petite pour laisser entrer plus d'un acteur à la fois, fait écho à une autre idée ancienne, celle d'une limite de la conscience. Juste après la Seconde Guerre mondiale, le psychologue britannique Donald Broadbent a étudié cette limite en empruntant les concepts de la théorie de l'information qui commençait à poindre[5]. En étudiant les pilotes de combat, il s'est rendu compte que, même avec beaucoup d'entraînement, ils ne parvenaient pas à faire attention à deux dialogues à la fois, un dans chaque oreille. Selon lui, la perception

consciente ressemble donc à un « canal de communication à capacité limitée » – un goulot d'étranglement qui ne peut traiter qu'une information à la fois.

La découverte ultérieure du clignement de l'attention et de la période psychologique réfractaire, que nous avons décrits au chapitre 2, renforce cette conclusion : dès que notre attention est attirée vers un premier objet de pensée, nous devenons aveugles à tous les autres. La psychologie cognitive moderne fait ainsi de l'accès à la conscience un « goulot d'étranglement central[6] », une « seconde étape de traitement[7] », ou un salon VIP où seuls quelques heureux élus sont invités.

Une autre métaphore a vu le jour dans les années 1960 et 1970. Celle-ci compare la conscience à un « système de supervision centrale », un puissant conseil d'administration qui contrôle les flux d'information dans le cerveau[8]. Comme l'avait déjà écrit William James dans ses *Principes de psychologie*, la conscience ressemble à « un organe ajouté afin de diriger un système nerveux devenu trop complexe pour se réguler lui-même[9] ». Au sens littéral, cette proposition fleure le dualisme : la conscience n'est pas un observateur externe au système nerveux, elle doit au contraire être considérée comme l'un de ses principaux acteurs. Mais ce que veut dire James, c'est que le système nerveux « se régule lui-même » par le biais d'une organisation hiérarchique. Les circuits d'ordre supérieur du cortex préfrontal, d'évolution récente, prennent le dessus sur les systèmes de bas niveau qu'abritent les aires postérieures du cortex et les noyaux sous-corticaux – souvent en les inhibant[10].

Les neuropsychologues Michael Posner et Tim Shallice ont proposé qu'une information devient consciente dès qu'elle accède à ce système régulateur de haut niveau. Nous savons à présent que cette idée n'est pas entièrement juste : au chapitre 3, nous avons vu que même un stimulus subliminal, qui reste invisible, peut déclencher certaines des opérations inhibitrices et régulatrices attribuées à l'« administrateur central[11] ». Inversement, par contre, toute information qui parvient à l'espace de travail conscient devient immédiatement susceptible de réguler, d'une

façon radicale, toutes nos pensées. L'attention exécutive n'est qu'un des nombreux systèmes que l'espace de travail global irrigue d'informations. Dès que nous prenons conscience d'une information, celle-ci devient capable d'influencer nos décisions et nos actions volontaires, en nous donnant le sentiment qu'elles sont sous le contrôle de notre volonté. La parole, la mémoire à long terme, l'attention et l'intention sont autant de systèmes qui forment ce cercle intérieur de « cadres supérieurs du cerveau » qui partagent les informations conscientes. Grâce à l'espace de travail, tout ce dont nous prenons conscience peut être arbitrairement transmis à n'importe lequel de ces processus et devenir ainsi le sujet d'une phrase, l'embryon d'un souvenir, le centre de l'attention ou le cœur de notre prochaine action volontaire (figure 24).

Au-delà de la modularité

Selon l'hypothèse de Bernard Baars, les opérations conscientes coïncident avec celles que permet l'espace de travail, qui assure la diffusion flexible des informations pertinentes à une grande diversité de circuits cérébraux. En principe, rien n'empêche de reproduire ces fonctions dans un ordinateur. En pratique, ces opérations n'ont rien d'évident. Nous ne savons pas encore comment le cerveau les réalise ni comment une machine pourrait les accomplir.

Nos logiciels sont généralement conçus d'une façon strictement modulaire : chaque programme reçoit certaines informations en entrée et les transforme selon des règles précises. Cette modularité a des conséquences fâcheuses. Il se peut, par exemple, qu'un traitement de texte contienne une information précieuse (l'adresse d'un ami), alors que l'ordinateur dans son ensemble ne dispose d'aucune possibilité d'examiner cette information, de décider de sa pertinence pour l'ensemble de l'organisme, et de la rendre accessible à d'autres programmes. C'est

pourquoi nos ordinateurs demeurent désespérément bornés : ils accomplissent à la perfection les tâches pour lesquelles ils ont été programmés, mais ce qui est connu d'un programme, quel que soit son degré d'« intelligence », ne peut pas être partagé avec d'autres. Seul un mécanisme rudimentaire, le presse-papiers, permet quelques échanges – mais seulement sous le contrôle d'un *deus ex machina,* l'utilisateur humain.

Notre cortex, à la différence de l'ordinateur, semble surmonter ce problème de modularité. Il adopte simultanément une division du travail en processeurs spécialisés, et un dispositif flexible d'aiguillage et de diffusion des informations. De nombreux secteurs du cortex se dédient à un processus particulier. Il existe, par exemple, des régions entières qui ne contiennent que des neurones sensibles aux visages[12]. Certaines aires pariétales et motrices ne s'intéressent qu'à certains gestes, ou aux parties du corps qui permettent de les accomplir. D'autres régions, plus abstraites, se spécialisent pour la connaissance des nombres, des animaux, des objets ou des verbes. Si la théorie de l'espace de travail est juste, la conscience a évolué afin d'atténuer les conséquences néfastes de cette modularité. Grâce à l'espace de travail neuronal global, nos modules cérébraux peuvent partager librement certaines informations. Leur disponibilité globale est précisément ce que nous appelons « conscience[13] ».

Les avantages évolutifs que confère cette organisation sont évidents. La modularité est utile parce que chaque domaine de connaissance nécessite des microprogrammes dédiés : la programmation d'un geste exige des calculs différents de la reconnaissance d'un paysage ou de la récupération d'un souvenir en mémoire. Prendre une bonne décision, par contre, nécessite souvent de briser cette modularité en recueillant un maximum d'informations les plus diverses. Imaginez un éléphant assoiffé, seul dans la savane. Sa survie dépend de la découverte d'un point d'eau. Sa décision de se mettre en route vers telle ou telle destination lointaine doit se fonder sur un usage optimal de toutes les informations disponibles : cartographie de l'espace environnant, reconnaissance des

sentiers et des arbres, souvenir des succès et des échecs antérieurs… Un choix aussi vital que celui d'entamer un épuisant périple sous le soleil africain doit s'appuyer sur une solide base de données. La conscience aurait ainsi évolué, il y a des millions d'années, afin d'extraire et de diffuser au reste du cerveau un maximum d'informations pertinentes initialement confinées dans des circuits cérébraux spécialisés [14].

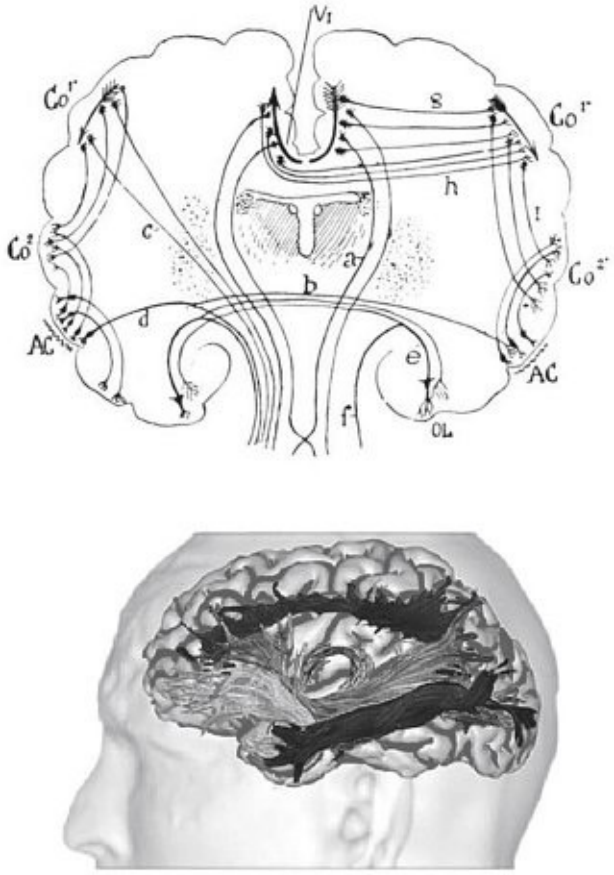


Figure 25. Les connexions neuronales à longue distance jouent un rôle essentiel dans la théorie de l'espace de travail global. Le célèbre neuroanatomiste Ramón y Cajal, dès la fin du XIX^e^ siècle, avait remarqué que les grands neurones pyramidaux du cortex humain envoient leur axone vers des régions très distantes (en haut). Nous savons aujourd'hui que certaines de ces projections à longue distance propagent les informations sensorielles à tout le réseau d'aires pariétales, temporales et préfrontales qui participent à l'espace de travail neuronal global (en bas). La lésion de ces connexions peut entraîner une héminégligence, c'est-à-dire une atteinte sélective de la prise de conscience d'une moitié de l'espace.

Un réseau de communication évolué

Selon cet argument évolutionniste, qui dit conscience dit connectivité. Le partage global des informations exige une architecture neuronale spécifique qui relie les nombreuses aires corticales, souvent distantes les unes des autres, en un tout cohérent. Nos cerveaux contiennent-ils une telle architecture ? Dès le XIX^{e} siècle, le neuroanatomiste espagnol Santiago Ramón y Cajal découvre au microscope un aspect particulier du tissu cérébral. À la différence de la mosaïque de cellules imbriquées qui forme notre peau, le cerveau est formé d'un réseau assez lâche de cellules extraordinairement étirées, les neurones. Chaque neurone est doté d'un prolongement, l'axone, qui peut mesurer plusieurs mètres de long. Afin de commander les muscles, chaque neurone du cortex moteur peut envoyer son axone jusque dans la moelle épinière, à plusieurs dizaines de centimètres de là. Plus fascinant encore, Cajal observe une grande densité de neurones à axones longs dans tout le cortex, la fine écorce qui recouvre toute la surface de nos deux hémisphères (figure 25). Il n'est pas rare qu'un axone quitte le cortex frontal pour se promener dans la substance blanche et atteindre l'arrière du cerveau ou l'autre hémisphère. Les axones longs se regroupent en grands faisceaux de câbles de quelques millimètres de diamètre et jusqu'à plusieurs centimètres de long. L'IRM permet aujourd'hui de démêler cet écheveau de fibres dans le cerveau humain vivant.

Toutes les régions corticales ne sont pas interconnectées de façon identique. Les aires sensorielles telles que le cortex visuel primaire (une aire appelée « V1 ») n'établissent qu'un petit nombre de contacts, principalement avec leurs voisins les plus proches. Les connexions des aires visuelles forment une hiérarchie : l'aire V1 parle principalement à l'aire V2, qui à son tour communique avec les aires V3 et V4, et ainsi de suite. Ainsi, les opérations visuelles précoces sont modulaires : chaque neurone visuel ne reçoit qu'une petite fraction des informations en

provenance de la rétine et la traite en toute autonomie, sans avoir « conscience » de l'ensemble.

Dans les régions associatives supérieures du cortex, par contre, les neurones ne privilégient plus seulement leurs voisins géographiques. La modularité est abandonnée au profit d'une diffusion plus éclectique de l'information. Les neurones présentent de grands axones avec de nombreux branchements. Ceux-ci sont particulièrement abondants dans le cortex préfrontal, la partie antérieure du cerveau. Ils le connectent avec de nombreuses autres régions situées dans le lobe pariétal inférieur, le lobe temporal moyen et antérieur, et les régions cingulaires antérieures et postérieures situées de part et d'autre de la ligne médiane qui sépare les deux hémisphères. Ces régions agissent comme des plaques tournantes, des postes d'aiguillage, des *hubs* d'un vaste réseau d'interconnexions[15]. Toutes leurs liaisons sont réciproques : si la région A projette vers la région B, alors il est pratiquement certain que la région B lui répond en retour (figure 25). De plus, les connexions à longue distance forment souvent des triangles : si A parle simultanément à B et à C, il y a fort à parier que ces derniers sont également reliés[16].

Ce « club » de régions interconnectées communique aussi avec d'autres circuits, notamment les noyaux centro-latéraux et intralaminaires du thalamus (impliqués dans l'attention, la vigilance et la synchronisation), les ganglions de la base (cruciaux pour la prise de décision et l'action) et l'hippocampe (essentiel pour la remémoration des épisodes de notre vie). Les voies qui relient le cortex et le thalamus sont particulièrement essentielles. Le thalamus est formé d'un paquet de noyaux qui se connectent en boucle avec une ou plusieurs régions spécifiques du cortex. Chaque voie neuronale directe entre deux régions du cortex est doublée d'une voie sous-corticale parallèle qui passe par un relais dans le thalamus[17]. Les entrées en provenance du thalamus jouent un rôle fondamental dans l'excitation du cortex : elles le maintiennent dans un état « haut » (*up state*), proche du seuil d'activation spontanée, où les décharges neuronales sont faciles à évoquer[18]. Comme nous le verrons,

la réduction des entrées thalamiques joue un rôle important chez les patients en coma ou en état végétatif : sans l'appui du thalamus, le cerveau peut perdre conscience.

L'espace de travail global repose donc sur un réseau extraordinairement dense de connexions corticales et sous-corticales – une organisation décentralisée, sans point de rencontre unique. Au sommet de la hiérarchie corticale, un club d'élite, constitué d'administrateurs chevronnés répartis dans différents territoires du cortex, travaille en coordination étroite, par le biais d'un incessant échange de messages. Une coïncidence est frappante : le réseau anatomique d'aires corticales reliées par des connexions à longue distance, et qui implique principalement les régions préfrontales et pariétales, correspond précisément à celui que j'ai décrit au chapitre 4, et dont l'activation soudaine constitue notre première signature de la conscience. Nous comprenons à présent pourquoi ces régions s'embrasent dès qu'une information parvient à la conscience : ce sont précisément elles qui possèdent la connectivité anatomique requise pour diffuser l'information consciente à l'ensemble du cortex.

Les neurones corticaux qui participent aux réseaux à longue distance sont bien adaptés à cette tâche (figure 26). Ce sont des cellules pyramidales géantes, dont le vaste corps cellulaire abrite toute la machinerie moléculaire nécessaire à la maintenance d'un immense axone. Souvenez-vous que c'est dans le noyau de la cellule que se situe la molécule d'ADN qui contient le code génétique – et pourtant, les sous-produits de la transcription génétique doivent s'acheminer jusqu'aux synapses situées à plusieurs centimètres de là. Les grandes cellules nerveuses capables d'une telle prouesse ont tendance à se concentrer dans certaines couches du cortex (les couches II et III). Elles sont particulièrement responsables des connexions du corps calleux, qui distribuent l'information aux deux hémisphères.

Dès les années 1920, le neuroanatomiste autrichien Constantin von Economo avait observé que les couches de cellules nerveuses n'ont pas

partout la même taille. Elles sont bien plus épaisses dans les régions préfrontales et cingulaires, ainsi que dans les aires associatives des lobes pariétaux et temporaux – exactement celles qui s'activent au cours de la prise de conscience !

Plus récemment, Guy Elston, de la région du Queensland en Australie, et Javier DeFelipe, en Espagne, ont montré que les neurones géants de l'espace de travail possèdent également des dendrites immenses. Ces vastes « antennes » les rendent particulièrement aptes à la réception de messages issus des régions distantes[19]. En effet, les neurones pyramidaux collectent les informations par le biais d'un fouillis de petites arborescences (c'est le sens du mot « dendrite » en grec). À l'endroit où atterrit une synapse en provenance d'une autre cellule, la dendrite produit une protubérance microscopique en forme de champignon qu'on appelle « épine dendritique ». Plusieurs milliers d'épines recouvrent toute la surface de l'arbre dendritique. En accord parfait avec l'hypothèse de l'espace de travail, Elston et DeFelipe ont montré que, dans le cortex préfrontal, les dendrites sont plus grandes et les épines plus nombreuses que dans les régions postérieures du cerveau – une adaptation à la « vie en réseau », à la communication corticale à longue distance (figure 26).

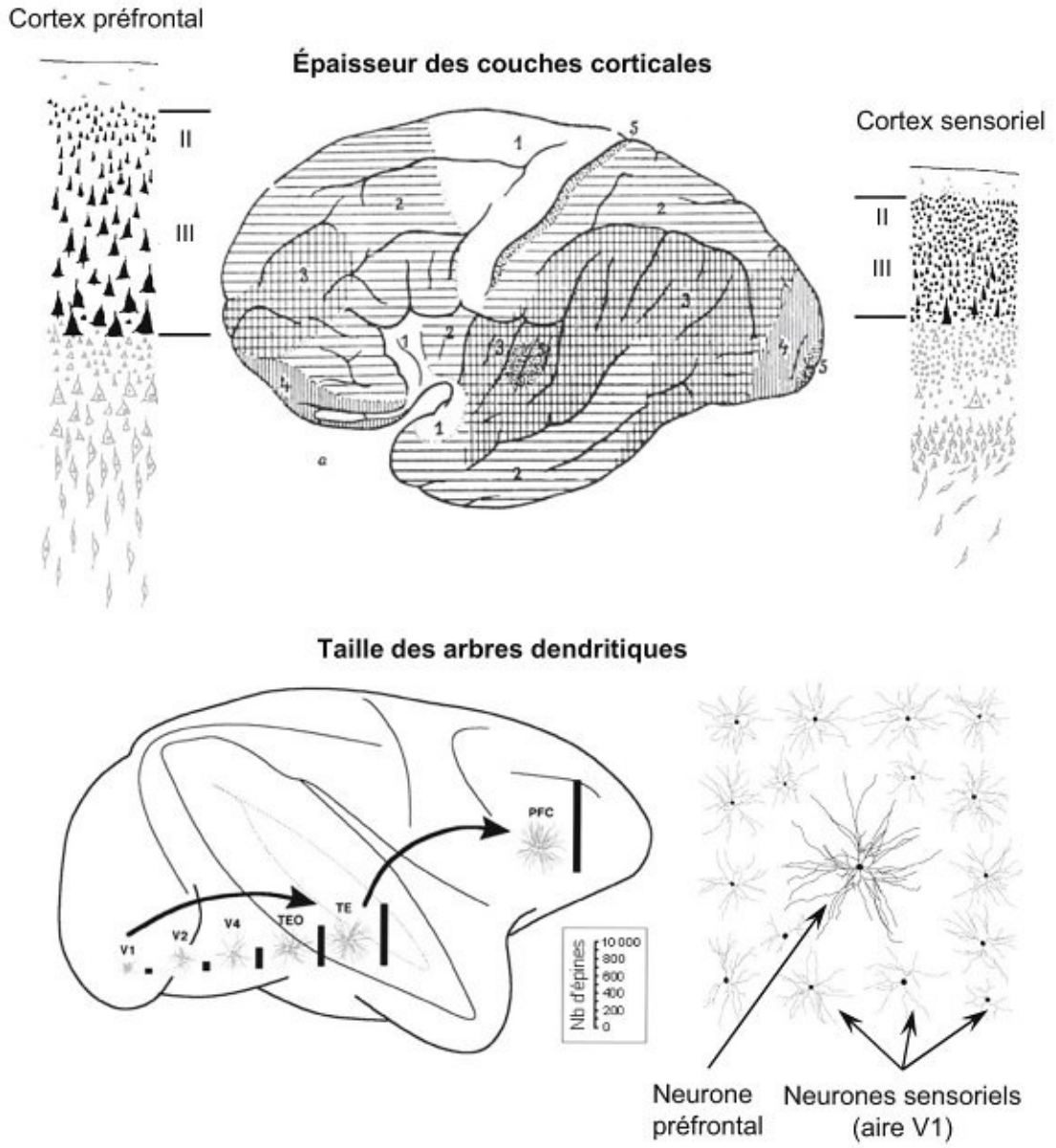


Figure 26. Les grands neurones pyramidaux sont bien adaptés à la diffusion globale des informations conscientes. Notre cortex est formé de six couches de cellules, et les couches II et III contiennent une grande densité de neurones pyramidaux à axones longs, capables de propager des messages à grande distance. Ces couches sont nettement plus épaisses dans les aires associatives, telles que le cortex préfrontal, que dans les aires sensorielles (en haut). Les aires pourvues d'épaisses couches II et III coïncident avec celles activées au cours de l'accès à la conscience. Les grands neurones pyramidaux sont également bien adaptés à la réception de messages globaux. Leurs arbres dendritiques (en bas), qui reçoivent les entrées en provenance d'autres régions, sont nettement plus touffus dans le cortex préfrontal qu'ailleurs. Ces adaptations

à la communication à longue distance sont plus prononcées dans le cerveau humain que chez toutes les autres espèces de primates.

Toutes ces caractéristiques anatomiques de l'espace de travail sont particulièrement développées dans le cerveau humain[20]. Par rapport aux autres espèces de primates, les neurones du cortex préfrontal humain possèdent des dendrites qui ont plus de branchements et contiennent plus d'épines. La jungle des dendrites est contrôlée par une famille de gènes qui ont muté dans la lignée humaine[21]. Cette famille inclut le fameux gène FoxP2, qui comprend deux mutations ponctuelles uniques au genre *Homo*[22], qui module les aires du langage[23], et dont les anomalies engendrent des troubles importants de l'articulation et du langage[24]. La famille FoxP2 comprend plusieurs autres gènes impliqués dans la construction des neurones, des dendrites, des axones et des synapses. Exploit de la génétique contemporaine, il est devenu possible de créer des souris mutantes dont le gène FoxP2 a été pourvu des deux mutations spécifiques de l'espèce humaine. Le cerveau de ces souris « humanisées » contient des neurones pyramidaux aux dendrites bien plus grandes que la normale, et leur faculté d'apprentissage est considérablement augmentée[25] (bien qu'elles ne parlent toujours pas !).

Grâce à FoxP2 et à la famille de gènes qui lui est associée, chaque neurone du cortex préfrontal humain contient environ quinze mille épines dendritiques, parfois plus. Cela veut dire qu'il échange des informations avec un nombre équivalent d'autres neurones, situés pour la plupart dans des régions lointaines du cortex et du thalamus. Cette organisation anatomique ressemble à s'y méprendre à une adaptation évolutive qui répond au grand défi de la communication cérébrale : collecter les informations n'importe où dans le cerveau et, une fois leur pertinence reconnue, les diffuser à des milliers d'autres sites.

Supposez que l'on puisse suivre, une par une, toutes les connexions nerveuses qui servent à reconnaître un visage – un peu comme le FBI

trace un appel téléphonique à travers toute une série de relais. Que verrait-on ? Au tout départ, des connexions courtes, au sein même de la rétine, nettoient et compriment l'image. Celle-ci est ensuite envoyée, *via* le gros câble du nerf optique, jusqu'au thalamus, puis au cortex visuel primaire qui se niche dans le lobe occipital. De là, par un réseau local de courtes fibres en U, l'information progresse de proche en proche vers le gyrus fusiforme de l'hémisphère droit, où les chercheurs ont découvert des zones entièrement dédiées aux visages. Mais ce foisonnement de transmissions demeure inconscient. Que se passe-t-il ensuite ? L'anatomiste suisse Stéphanie Clarke a fait une découverte étonnante[26] : à ce niveau, des neurones à axone long prennent le relais et envoient ces informations visuelles dans pratiquement tous les recoins du cortex. À partir du lobe temporal inférieur, des connexions massives et directes, en une seule synapse, projettent vers toutes les aires associatives, y compris dans l'autre hémisphère. Ces projections se concentrent sur la région de Broca (le cortex frontal inférieur) et celle de Wernicke (le cortex temporal associatif). Toutes deux sont des points névralgiques du circuit du langage : dès que l'information leur parvient, elles peuvent mettre un nom sur le visage.

Comme ces régions elles-mêmes participent au réseau bien plus large de l'espace de travail global, l'information peut à présent être disséminée à tout le cercle d'administrateurs de haut vol. Ils la font circuler entre eux, en sorte qu'elle y réverbère et ne s'éteint pas, même si le visage lui-même disparaît de la rétine. Selon ma théorie, c'est à cet instant que l'information devient consciente.

Sculpter une pensée consciente

Essayez d'estimer le nombre énorme de pensées conscientes qui pourraient traverser votre cerveau : tous les visages, les objets, les scènes

que vous reconnaissez ; toutes les émotions que vous ayez jamais ressenties ; chaque connaissance, vraie ou fausse, que vous avez ou que vous pourriez avoir, en géographie, en histoire, en mathématiques, ou tout simplement que vous auriez lue dans un magazine ; la prononciation et le sens de tous les mots que vous connaissez ou que vous pourriez connaître, dans n'importe quelle langue… La liste paraît infinie, et pourtant chacun de ses éléments pourrait, dans la prochaine minute, devenir l'objet de votre réflexion consciente. Comment une assemblée de neurones parvient-elle à coder une telle diversité d'états mentaux ? Quel est le code neural de la conscience, et comment nous donne-t-il accès à un nombre pratiquement infini de pensées ?

Le neuroscientifique Giulio Tononi souligne que la taille énorme de notre répertoire d'idées contraint fortement le code neuronal de la conscience[27]. La *différenciation* est sa caractéristique principale : la combinatoire des neurones actifs et inactifs, au sein de l'espace de travail global, doit pouvoir former des milliards de configurations. Chacun de nos états mentaux doit être alloué à l'une de ces configurations, et celle-ci doit être bien différenciée des autres pour éviter toute confusion. Il faut donc qu'il y ait des frontières nettes : soit c'est un oiseau, soit c'est un avion, soit c'est Superman, mais pas les trois à la fois. Un esprit lucide, avec des myriades de pensées différentes, implique un cerveau doté de myriades d'états stables et bien différenciés.

Dans son livre *L'Organisation du comportement* (1949), Donald Hebb avait déjà proposé une théorie visionnaire de la manière dont le cerveau code les pensées. Il y introduisait le concept d'« assemblées de cellules » – des groupes de neurones reliés entre eux par des synapses excitatrices, et qui ont donc tendance à rester actif longtemps après que la stimulation a disparu. « Toute stimulation qui se répète fréquemment, affirmait-il, conduira au développement progressif d'une assemblée de cellules, une structure diffuse comportant des cellules du cortex et du diencéphale (et également, peut-être, les ganglions de la base), capable d'agir brièvement comme un système fermé[28]. »

Selon Hebb, tous les neurones qui appartiennent à une assemblée se soutiennent l'un l'autre en se transmettant des décharges excitatrices. Ils forment donc une « colline » d'activité bien délimitée dans l'espace des neurones. Et parce que de nombreuses assemblées locales peuvent s'activer indépendamment les unes des autres en différents endroits du cerveau, il en résulte un code combinatoire capable de représenter des milliards de possibilités. N'importe quel objet visuel peut être représenté par une combinaison de fragments de couleurs et de formes. Les enregistrements neuronaux réalisés dans le cortex visuel du singe vont dans ce sens : l'image d'un extincteur, par exemple, semble codée par une combinaison de parcelles de cortex activées, chacune contenant quelques centaines de neurones qui représentent une petite partie de l'image[29] (la poignée, le réservoir, le tuyau, etc.).

En 1959, John Selfridge, pionnier de l'intelligence artificielle, introduisit une autre métaphore utile : le « pandémonium[30] ». Il envisageait le cerveau comme une hiérarchie de « démons » spécialisés qui tentent de proposer une interprétation de l'image entrante. Trois décennies de neurophysiologie plus tard, l'idée tient toujours, comme le montre la découverte spectaculaire de neurones qui ne répondent qu'à certaines lignes, à une couleur, à la forme des yeux, à un visage entier ou même aux présidents des États-Unis et aux stars de Hollywood. Dans le modèle de Selfridge, les démons excités par l'image se mettent à crier leur interprétation favorite, avec une ardeur proportionnelle au soutien qu'ils reçoivent des entrées sensorielles. Des ondes de cris se propagent à travers une hiérarchie de démons de plus en plus abstraits, qui correspondent aux neurones capables de répondre à des aspects de plus en plus élevés de l'image. Trois démons, par exemple, qui détectent la présence des yeux, du nez et de la chevelure, conspirent pour en exciter un quatrième qui code pour la présence d'un visage. En écoutant les démons qui se font le plus entendre, un système de décision parvient à se forger une opinion du contenu de l'image – une perception consciente.

Une amélioration importante est venue compléter le modèle du pandémonium. Dans l'idée originale de Selfridge, l'excitation ne se propageait que de bas en haut, selon une stricte hiérarchie : les démons les plus proches de l'entrée sensorielle braillaient en direction de leurs supérieurs hiérarchiques, mais dans le sens inverse, un démon de haut rang ne s'abaissait jamais à répondre à ses collègues de niveau inférieur ni même à un autre démon du même niveau. Dans la réalité, pourtant, les assemblées de neurones communiquent dans les deux sens. Les boucles et les projections bidirectionnelles abondent dans le cortex[31]. Même les neurones individuels se parlent mutuellement : si un neurone α se connecte au neurone β, alors il est probable que β, à son tour, se connecte à α[32]. À chaque niveau, les neurones interconnectés se soutiennent mutuellement, et le haut de la hiérarchie communique à ses subordonnés, si bien que les messages se propagent au moins autant vers le bas que vers le haut.

La simulation informatique et la modélisation mathématique de réseaux « connexionnistes » réalistes, qui comprennent beaucoup de boucles, font apparaître une propriété utile. Lorsqu'un sous-groupe de neurones s'excite, l'assemblée tout entière s'auto-organise. Elle tombe dans un « attracteur » : tous les neurones adoptent rapidement une configuration reproductible d'activité qui reste stable pendant une longue durée[33]. Comme l'avait prédit Hebb, les neurones interconnectés forment des assemblées cellulaires stables.

Autre avantage : ces réseaux convergent souvent vers un consensus raisonnable. Lorsque le réseau inclut des connexions réciproques, et pas seulement unidirectionnelles comme dans le pandémonium de Selfridge, les neurones ne se contentent pas de s'invectiver : ils convergent progressivement vers un accord satisfaisant pour tous, une interprétation unifiée de la scène perçue. Les neurones qui reçoivent le plus d'activation se soutiennent l'un l'autre et inhibent progressivement tous les autres. Ce simple mécanisme parvient à restaurer les parties manquantes de l'image. Quelques itérations suffisent à en éliminer le bruit et les incertitudes. La

représentation neuronale devient stable, cohérente, résistante aux interférences, et bien différenciée des configurations d'activité qui codent pour d'autres images. Francis Crick et Christof Koch l'intitulent une « coalition neuronale gagnante », et ils y voient le code idéal d'une représentation consciente[34].

Le mot « coalition » souligne un autre aspect essentiel du code neuronal conscient : toutes ses composantes doivent être fortement soudées, intégrées les unes aux autres[35]. Chacune de nos pensées conscientes est constituée d'une seule pièce. Lorsque nous contemplons *La Joconde*, notre conscience ne nous donne jamais à voir une sorte de Picasso éviscéré dont les mains, les yeux et le sourire magique tireraient à hue et à dia. Nous intégrons tous ces fragments de la scène en un tout cohérent. Au départ, pourtant, ils sont éclatés dans des assemblées de neurones épars et distants de plusieurs centimètres à la surface du cortex. Comment les attachons-nous ensemble ?

Une solution consiste à profiter des plaques tournantes que forment les aires associatives du cortex. Ces centres névralgiques, que le neurologue Antonio Damasio appelle des « zones de convergence[36] », prédominent dans le cortex préfrontal, mais également dans certaines régions antérieures du lobe temporal, dans le lobule pariétal inférieur, et une région de la ligne médiane appelée le précuneus. Toutes reçoivent et renvoient un grand nombre de connexions en direction des autres secteurs du cortex, ce qui permet aux neurones qui s'y trouvent d'intégrer les informations à travers l'espace et le temps. Ainsi mis en contact, les différents modules sensoriels peuvent s'accorder sur une interprétation unifiée et cohérente (« une séduisante Italienne ») qui sera alors renvoyée aux aires sensorielles d'origine. Il en résulte une représentation intégrée. En diffusant des messages globaux depuis le cortex préfrontal et les autres aires associatives supérieures en direction des régions sensorielles, les neurones pourvus d'axones longs créent les conditions nécessaires à l'émergence d'un état unifié de conscience, à la fois différencié et intégré.

Ce dialogue permanent entre les aires cérébrales, qui inclut un retour vers les aires d'origine, a été nommé « réentrée » par le prix Nobel Gerald Edelman[37]. La modélisation de réseaux de neurones montre que la réentrée permet des calculs sophistiqués qui convergent vers l'interprétation statistique la plus probable d'une scène visuelle[38]. Chaque groupe de neurones agit comme un statisticien expert, qui collabore avec ses collègues afin d'expliquer les données sensorielles dans leurs moindres détails[39]. Un expert des ombres, par exemple, décide qu'il peut expliquer une zone sombre de l'image – mais seulement si la lumière vient d'en haut à gauche. Un expert de l'éclairage lui répond que c'est possible et que cette hypothèse pourrait expliquer pourquoi le haut des objets semble plus clair. Un troisième expert entre en lice et déclare qu'une fois ces effets pris en compte le reste de l'image ressemble à un visage. Ce dialogue se poursuit jusqu'à ce que le moindre recoin de l'image ait été expliqué.

La forme d'une idée

Assemblées de cellules, pandémonium, coalitions en compétition, attracteurs, zones de convergence, réentrée… Chacune de ces hypothèses théoriques semble détenir une part de vérité. Ma théorie de l'espace de travail neuronal en propose une synthèse[40]. Selon elle, un état mental conscient est codé par la décharge simultanée, stable pendant quelques dixièmes de seconde, d'un sous-ensemble de neurones actifs de l'espace de travail. Ces neurones sont distribués dans de nombreuses aires cérébrales, et codent tous pour différentes facettes de la même représentation mentale. Prendre conscience de *La Joconde*, c'est coactiver quelques millions de neurones qui représentent chacun un fragment d'objet, de sens ou de souvenir.

Au cours de l'accès à la conscience, grâce à leurs axones longs, tous ces neurones échangent des messages réciproques, en parallèle, afin de

tenter de parvenir à une interprétation cohérente et synchrone des données. La perception consciente s'achève dès qu'ils convergent vers un état relativement stable. La « superassemblée » de cellules qui abrite le contenu de la conscience est donc éparpillée dans le cerveau – mais ses fragments s'ajustent comme les pièces d'un puzzle parce que les assemblées locales se synchronisent grâce aux messages descendants propagés par les neurones à axones longs.

La synchronie neuronale semble être un ingrédient important du fonctionnement cortical. Un nombre croissant d'observations indique que les neurones distants forment des superassemblées géantes en synchronisant leurs décharges avec les oscillations électriques qui se déroulent en toile de fond[41]. Selon cette image, le réseau global qui code chacune de nos pensées ressemble à un essaim de lucioles qui harmonisent leurs flashs lumineux en fonction du rythme de l'ensemble du groupe. En l'absence de conscience, des assemblées neuronales de taille moyenne parviennent toujours à se synchroniser localement, par exemple lorsque nous codons le sens d'un mot invisible au sein des aires du langage du lobe temporal gauche. Cependant, tant que le cortex préfrontal n'entre pas dans cette danse, ce message ne peut pas être partagé globalement et reste donc confiné aux profondeurs de l'inconscient.

Imaginez les seize milliards de neurones de votre cortex. Chacun d'eux s'intéresse à un petit nombre de stimuli. Leur diversité est étonnante : dans le seul cortex visuel, on trouve des neurones qui répondent aux visages, aux mains, aux objets, à la perspective, aux formes, aux lignes, aux courbes, aux couleurs, à la troisième dimension… Chaque cellule ne transmet que quelques bits d'information sur la scène observée. Collectivement, pourtant, elles parviennent à coder un immense répertoire de pensées. Le modèle de l'espace de travail propose qu'à un instant donné la dynamique du réseau réduit cet espace potentiellement infini en y sélectionnant un seul objet de pensée qui occupe le cœur de notre conscience. À cet instant, toutes les cellules pertinentes s'activent en

synchronie partielle, sous l'égide d'un sous-ensemble de neurones du cortex préfrontal.

Il est crucial de bien comprendre que, dans ce type de code, les neurones silencieux, ceux qui ne déchargent *pas*, participent au codage global. Implicitement, leur mutisme indique aux autres que leur trait favori n'est pas présent ou n'est pas pertinent dans la scène actuelle. Un contenu conscient se définit au moins autant par ses neurones silencieux que par ses neurones actifs.

En dernière analyse, la perception consciente ressemble à une sculpture. L'artiste fait émerger sa vision d'un grand bloc de marbre brut dont il élimine, miette à miette, une grande partie de la matière. De même, la conscience s'appuie sur un immense espace de travail comprenant plusieurs centaines de millions de neurones, dont les décharges fluctuent aléatoirement autour d'un niveau par défaut, et que la perception d'une image particulière « sculpte » en faisant taire un grand nombre d'entre eux. Ne reste alors qu'une petite fraction de cellules actives dont les contours délimitent, littéralement, les traits de la pensée consciente.

Ce paysage de neurones actifs et inactifs peut expliquer notre deuxième signature de la conscience : l'onde P3 que j'ai décrite au chapitre 4, une grande déviation positive du voltage mesuré au sommet du crâne. Au cours de la perception consciente, un sous-ensemble de neurones de l'espace de travail s'active et définit le contenu conscient actuel, tandis que les autres sont inhibés. Les neurones actifs propagent alors leur message au reste du cortex, en envoyant des potentiels d'action le long de leur axone. En de nombreux endroits du cortex, cependant, ces signaux doivent contacter des neurones inhibiteurs afin de faire taire des groupes entiers de neurones. « Chut, semblent-ils dire à ces neurones, les propriétés que vous préférez ne sont pas pertinentes actuellement. » Une idée consciente est codée par des petites assemblées neuronales actives, entourées de vastes populations de neurones inhibés.

À présent, réfléchissons à la géométrie des courants qui traversent ces neurones. Dans les cellules pyramidales actives, les courants synaptiques se propagent depuis les dendrites superficielles jusqu'au corps cellulaire. Comme tous ces neurones sont alignés comme à la parade, parallèlement les uns aux autres, leurs courants électriques s'additionnent et, à la surface du cortex, ils engendrent une onde *négative* lente au-dessus des régions qui codent pour la représentation consciente[42]. Les neurones inhibés, cependant, dominent le paysage cortical – et leur activité s'additionne pour donner, à l'inverse, un potentiel électrique *positif*. S'il y a nettement plus de neurones inhibés qu'activés, le résultat net qui apparaît à la surface de la tête est une grande onde positive – la fameuse P3 que nous détectons que survient un accès à la conscience[43]. Voilà une explication plausible de notre deuxième signature de la conscience.

Cette théorie explique pourquoi l'onde P3 est aussi forte, générique et reproductible : elle indique essentiellement tout ce que la pensée actuelle *n'est pas*. Ce sont les petites négativités focales qui définissent le contenu de la conscience, et celles-ci sont masquées par la positivité diffuse. Dans certains cas, nous parvenons tout de même à les détecter. Edward Vogel et ses collègues de l'Université d'Oregon ont publié de magnifiques exemples de voltages négatifs au-dessus des aires pariétales du cortex, qui reflètent fidèlement le contenu de la mémoire de travail spatiale[44]. Dès que nous mémorisons un ensemble d'objets, ces potentiels négatifs indiquent exactement combien d'objets nous avons vus, ainsi que leur répartition dans l'espace (si les objets sont à droite, les ondes négatives apparaissent au-dessus du lobe pariétal gauche, et *vice versa*). Ces voltages restent stables tant que nous gardons un ensemble d'objets en tête ; ils s'accroissent lorsque nous y ajoutons de nouveaux objets ; ils saturent précisément au moment où notre mémoire déborde ; ils s'écroulent au moment précis où nous perdons la mémoire ; et leur niveau reflète le nombre d'objets dont nous nous souvenons. En bref, dans le travail d'Edward Vogel, les ondes négatives reflètent, à tout instant, le contenu de la conscience spatiale, exactement comme le prédit la théorie.

Simuler l'ignition consciente

> « La science de la réalité ne se contente plus du *comment* phénoménologique ; elle cherche le *pourquoi* mathématique. »
> Gaston BACHELARD, *La Formation de l'esprit scientifique* (1938).

L'accès à la conscience cisèle une pensée en sculptant des assemblées de neurones actives et inactives dans l'espace de travail neuronal global. Cette métaphore suffit peut-être à nourrir notre intuition de ce qu'est la conscience, mais au bout du compte, il faudra bien la remplacer par une théorie mathématique plus précise de la manière dont un réseau de neurones engendre les signatures neurophysiologiques que nous observons dans nos enregistrements.

Bien que nous n'en soyons qu'aux balbutiements, Jean-Pierre Changeux et moi-même avons commencé à créer des simulations informatiques qui reproduisent les propriétés les plus basiques de l'accès à la conscience[45]. Notre but était d'examiner comment des neurones se comportent lorsqu'ils sont connectés selon les préceptes de notre théorie (figure 27). Afin de recréer, dans l'ordinateur, la dynamique d'une petite coalition de neurones, nous sommes partis d'un modèle simplifié du neurone, où celui-ci se contente de faire la somme de ses entrées et, dès que ce total dépasse un seuil, de déclencher un potentiel d'action. Chaque neurone possède des synapses réalistes, dont les équations mathématiques reproduisent les principaux types de récepteurs excitateurs et inhibiteurs présents dans le cortex.

Nous avons connecté ces neurones virtuels pour former des colonnes corticales, qui imitaient grossièrement la subdivision du cortex en couches cellulaires. Le concept de la « colonne » vient du fait que les neurones corticaux qui se situent l'un au-dessus de l'autre, perpendiculairement à la surface du cortex, ont tendance à être interconnectés, à répondre aux mêmes stimuli, et à être issus de la division d'une seule et même cellule progénitrice au cours du développement. Notre modèle respectait cette

disposition : au sein d'une colonne simulée, les neurones avaient tendance à se soutenir mutuellement et à réagir aux mêmes entrées.

Nous avons également simulé un morceau de thalamus – cette structure sous-corticale qui comprend de nombreux noyaux interconnectés avec le cortex. Nous l'avons relié au reste du modèle, avec une connectivité et des délais temporels réalistes qui tenaient compte de la distance que les décharges neuronales devaient parcourir. Le tout formait un modèle approximatif de l'unité fondamentale de calcul dans le cerveau des primates : la colonne thalamocorticale. Nous nous sommes assurés que le modèle reproduisait quelques données physiologiques fondamentales : en l'absence de toute entrée, les neurones virtuels déchargeaient spontanément et produisaient même un électroencéphalogramme assez réaliste.

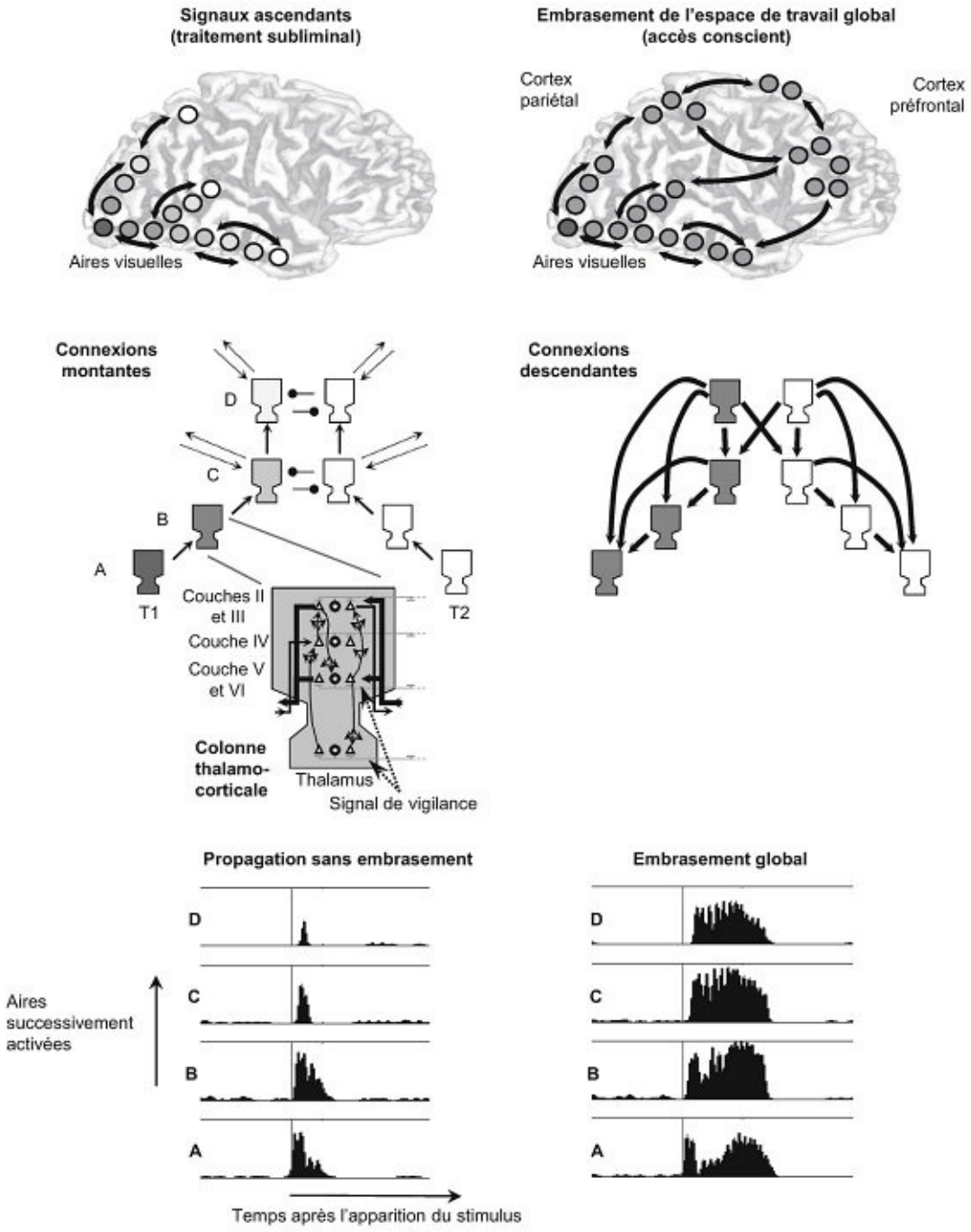


Figure 27. Une simulation informatique reproduit les principales signatures de la perception consciente et non consciente. Nous avons simulé sur ordinateur un petit sous-ensemble des aires visuelles et associatives qui participent au traitement subliminal et à l'accès conscient (en haut). Quatre aires étaient interconnectées par des connexions ascendantes et descendantes (au milieu). Chacune comprenait des neurones organisés en couches et connectées avec le thalamus. Lorsqu'on stimulait brièvement l'une des aires sensorielles (A), l'activation se propageait de bas en haut avant

de s'éteindre, simulant ainsi l'activité transitoire observée lors de la perception subliminale. Une stimulation plus longue entraînait un embrasement global : les connexions descendantes amplifiaient l'information montante et induisaient une seconde vague d'activité nettement plus durable, dont les propriétés reproduisaient celles de la perception consciente.

Nous avons ensuite connecté plusieurs de ces colonnes pour former des réseaux à longue distance. Notre simulation comprenait une hiérarchie de quatre régions cérébrales, chacune dotée de deux colonnes qui codaient respectivement pour un son et une lumière. Notre réseau ne pouvait donc distinguer que deux états du monde – une simplification radicale qui était malheureusement indispensable pour limiter les ressources informatiques nécessaires à la simulation. Nous supposions que les propriétés physiologiques du modèle ne seraient pas trop affectées par cette réduction massive de sa complexité[46].

En périphérie, les étapes les plus précoces de la perception opéraient en parallèle : les neurones qui codaient pour le son et pour la lumière pouvaient s'activer simultanément, sans interférences. Aux niveaux supérieurs, par contre, ils s'inhibaient mutuellement, en sorte que ces régions ne pouvaient représenter qu'un seul état à la fois – une seule « pensée ».

Comme dans un vrai cerveau, les connexions « montantes » s'établissaient de proche en proche : l'aire primaire recevait les stimulations extérieures, puis transmettait ses décharges à l'aire secondaire, qui à son tour les communiquait à une troisième et enfin à une quatrième région. Les projections « descendantes », en retour, reconnectaient le réseau avec lui-même, en permettant aux aires de haut niveau d'envoyer leur soutien aux aires sensorielles qui les avaient initialement excitées. L'ensemble formait un tout petit espace de travail global : un enchevêtrement de connexions montantes et descendantes doté de plusieurs niveaux enchâssés : le neurone, la colonne, l'aire corticale et le réseau à longue distance qui intègre l'ensemble en un tout cohérent.

Après toute cette programmation informatique, quel plaisir de faire enfin tourner la simulation complète et de voir s'exciter tous ces neurones virtuels ! Pour simuler la perception, nous injections un petit courant dans les neurones du thalamus – imitant ainsi ce qui se passe lorsqu'un rai de lumière parvient aux photorécepteurs de la rétine et, après un prétraitement, excite les neurones du corps genouillé latéral du thalamus. Nous laissions ensuite les milliers d'équations de la simulation se dérouler librement. Comme nous l'espérions, bien que simplifiée à l'extrême, notre maquette reproduisait bon nombre de propriétés physiologiques qui avaient été observées expérimentalement, et dont les origines devenaient enfin transparentes.

La première de ces propriétés était un embrasement soudain : lorsque nous présentions une brève impulsion en entrée, l'activité neuronale se propageait tout le long de la hiérarchie corticale dans un ordre fixe, depuis l'aire primaire jusqu'à la deuxième, puis la troisième, enfin la quatrième. La propagation de cette onde imitait la séquence bien connue d'entrée en activité de la hiérarchie des aires visuelles. Ensuite, toutes les colonnes qui codaient pour le même stimulus se mettaient à s'embraser. Grâce aux connexions descendantes, les neurones échangeaient des signaux qui se renforçaient mutuellement et, très vite, l'activité s'emballait jusqu'à atteindre un niveau très élevé, sans commune mesure avec la force de l'entrée initiale. En parallèle, l'autre perception était inhibée. L'activation soutenue pouvait durer quelques centaines de millisecondes – bien plus que la durée initiale de stimulation : même une impulsion très brève, pourvu qu'elle soit intense, pouvait induire un état de réverbération neuronale durable. Ces simulations cernaient donc l'essence de la façon dont le cerveau se forme une représentation stable d'une image présentée pendant un court instant.

La dynamique du modèle reproduisait les propriétés que nous avions observées dans nos enregistrements. La plupart des neurones virtuels montraient une augmentation tardive et soudaine de leurs courants synaptiques. L'excitation se propageait d'abord de bas en haut, puis elle

revenait à son point de départ, imitant ainsi l'amplification tardive que nous avions constatée dans les aires sensorielles pendant l'accès à la conscience. Dans la simulation, l'embrasement global entraînait aussi une réverbération de l'activation neuronale au sein des nombreuses boucles du réseau : à l'intérieur de chaque colonne corticale, mais aussi entre le cortex et le thalamus, et d'une région corticale à l'autre. Il s'ensuivait une augmentation massive des oscillations dans une vaste bande de fréquences, avec un pic dans la bande gamma (30 hertz et au-delà). Enfin, pendant l'embrasement, les potentiels d'action se synchronisaient entre eux et avec le potentiel de membrane des cellules de toutes les assemblées qui codaient pour la représentation en cours. En un mot, la simulation reproduisait nos quatre signatures expérimentales de la conscience.

Simuler aide à formuler de nouvelles hypothèses mathématiques. Dans le modèle, l'accès à la conscience correspondait à ce que la physique théorique appelle une « transition de phase » – le passage soudain d'un état à un autre. Comme je l'ai expliqué au chapitre 4, l'eau qui gèle illustre bien ce concept : tout à coup, les molécules de H_2O s'assemblent pour former une structure rigide dotée de caractéristiques émergentes nouvelles. Au cours d'une transition de phase, les propriétés physiques du système changent souvent brutalement, d'une façon discontinue. Dans nos simulations, de même, l'état des neurones passait soudainement d'une activité spontanée assez basse à une phase très différente de décharges rapides et synchronisées.

Il est facile de comprendre, sans faire appel aux mathématiques, d'où provient cette discontinuité. Comme les neurones qui codent pour un même objet de pensée sont connectés par des boucles excitatrices, ils finissent par amplifier leur propre activité. Le système qui en résulte ne possède que deux états stables. Soit les neurones restent à un niveau d'activité faible, soit leur excitation dépasse un seuil critique et ils s'emballent d'un seul coup. L'avalanche d'activité s'autoamplifie et certains groupes de neurones voient leur activité devenir frénétique. La frontière entre ces deux états est mince. Elle rend imprévisible le sort d'un stimulus

d'intensité intermédiaire. En fonction des conditions initiales et du hasard propre à chaque simulation, la même stimulation sensorielle peut soit s'évanouir rapidement, soit déclencher une avalanche qui modifie l'ensemble du réseau.

Cet aspect de nos simulations rend compte d'un concept de la psychologie qui date d'au moins cent cinquante ans : l'idée que la conscience possède un seuil qui sépare les images subliminales et supraliminales. Les opérations subliminales correspondent à la propagation d'une activité neuronale d'une aire à l'autre, sans déclencher d'embrasement global. L'accès à la conscience, lui, correspond à la transition soudaine vers un état plus élevé d'activité synchronisée.

Personne ne doute, cependant, que le cerveau soit immensément plus complexe qu'une boule de neige ! Modéliser correctement les transitions de phase qui surviennent dans la dynamique des réseaux de neurones réels nous occupera encore de nombreuses années[47]. Même nos simulations simplifiées contenaient déjà deux transitions imbriquées. L'une, que je viens de décrire, correspondait à l'embrasement global. Cependant, le seuil de cet embrasement était lui-même sous le contrôle d'une seconde transition de phase, qui modélisait l'état de vigilance du réseau. Chaque neurone pyramidal de la simulation recevait un signal de vigilance, un petit courant synaptique qui résumait, sous une forme grandement simplifiée, l'influence de neuromodulateurs tels que l'acétylcholine, la noradrénaline et la sérotonine. Ces molécules, issues de divers noyaux du tronc cérébral, du mésencéphale et de l'hypothalamus, ont pour effet d'activer le cortex. Par ce biais, notre modèle rendait compte des changements de l'*état de conscience* – le passage d'un cerveau inconscient, incapable de représenter le moindre contenu de façon explicite, à un cerveau conscient.

Dans nos simulations, lorsque le signal de vigilance était faible, l'activité neuronale spontanée diminuait massivement, et avec elle disparaissait la capacité d'embrasement : même une entrée sensorielle très forte, qui activait vigoureusement le thalamus et les aires sensorielles

primaires et secondaires, faisait long feu et se dissipait sans franchir le seuil d'embrasement global. Dans cet état, notre réseau se comportait donc comme un cerveau anesthésié ou endormi[48]. Il répondait encore aux stimulations, mais seulement de façon superficielle – l'activation sensorielle ne grimpait pratiquement jamais jusqu'aux aires corticales de niveau supérieur. À mesure que nous augmentions le paramètre de vigilance, un électroencéphalogramme structuré émergeait, et le modèle recouvrait sa capacité d'embrasement. Le seuil d'accès à la conscience variait en proportion directe de la vigilance : plus le niveau d'éveil était bas, moins le réseau avait de chances de détecter un stimulus de faible intensité.

Pas de repos pour le cerveau

> « Je vous le dis : il faut porter encore en soi un chaos, pour pouvoir mettre au monde une étoile dansante. Je vous le dis : vous portez en vous un chaos. »
>
> Friedrich NIETZSCHE,
> *Ainsi parlait Zarathoustra* (1883-1885).

Un autre phénomène fascinant émergeait de nos simulations : l'activité neuronale spontanée. Nous n'avions pas besoin de stimuler le réseau pour qu'il reste actif. Même en l'absence de toute entrée sensorielle, les neurones déchargeaient spontanément, en réponse à des fluctuations aléatoires de la transmission synaptique – et cette activité chaotique s'auto-organisait pour former des états d'activité cohérents et reconnaissables.

Lorsque le niveau de vigilance était élevé, un flux permanent de configurations complexes se déroulait sur nos écrans d'ordinateur. Parmi elles, nous détections parfois un embrasement global – déclenché en l'absence de tout stimulus extérieur. Toute une série de colonnes corticales

codant pour le même objet s'activaient spontanément pendant un bref instant, puis s'évanouissaient. Une fraction de seconde plus tard, une autre assemblée de neurones la remplaçait. Sans qu'on le lui demande, le réseau s'auto-organisait pour former une séquence stochastique d'embrasements successifs, proches de ceux observés au cours de la perception de stimuli externes. La seule différence importante était que l'activité spontanée avait tendance à démarrer dans les aires corticales du haut de la hiérarchie, au sein de l'espace de travail global, et à se propager de haut en bas, en sens inverse de la perception.

Cette propriété n'est pas qu'une pure invention du modèle. Non seulement le cerveau présente également des périodes d'activité spontanée, intense et bien organisée, mais cette activité spontanée est omniprésente dans le système nerveux. Quiconque a déjà vu un EEG sait que les deux hémisphères du cortex émettent en permanence des ondes électriques aléatoires de haute fréquence. Que la personne soit éveillée ou endormie, cette excitation spontanée est tellement intense qu'elle domine le signal d'EEG. En comparaison, l'activité évoquée par un stimulus sensoriel est à peine détectable, et il faut moyenner beaucoup d'essais pour la voir apparaître. L'activité cérébrale évoquée ne rend compte que d'une toute petite fraction de l'énergie consommée par le cerveau, sans doute moins de 5 %. Le système nerveux se comporte essentiellement comme un système autonome qui engendre la pensée à son propre rythme. Allongé en silence dans le noir, tandis que nous nous reposons « sans penser à rien », notre cerveau produit constamment des configurations changeantes d'activité neuronale.

C'est chez l'animal que l'organisation complexe de l'activité spontanée a été observée pour la première fois. En utilisant des colorants sensibles au voltage, qui transforment les ondes électriques cérébrales en configurations visibles d'ombre et de lumière, Amiram Grinvald et ses collègues de l'Institut Weizmann ont obtenu des enregistrements de longue durée de l'activité d'une zone de cortex[49]. Bien que l'animal ait été anesthésié, l'activité de son cerveau composait des figures complexes.

Dans le noir, en l'absence de toute stimulation, un neurone visuel se mettait à décharger. Il n'était pas seul : l'imagerie montrait qu'au même instant toute une assemblée de neurones s'était spontanément mise en branle.

Le même phénomène survient également dans le cerveau humain[50]. Les images d'activité cérébrale « au repos », alors qu'on demande à la personne de ne penser à rien, montrent que, loin de se taire, le cortex est parcouru d'incessantes fluctuations d'activité. De vastes réseaux d'aires corticales, distribuées dans les deux hémisphères, s'activent spontanément dans tous les cerveaux. Certains correspondent aux réseaux évoqués par un stimulus externe. Il existe, par exemple, un circuit des aires du langage qui s'active dès que nous entendons une phrase dans notre langue, mais qui décharge également spontanément lorsque nous nous reposons dans le noir – en plein accord avec l'idée d'un « langage intérieur ».

L'interprétation de l'activité cérébrale au repos reste débattue. Certains estiment que les décharges neuronales aléatoires ne font que suivre le réseau de connexions anatomiques cortico-corticales – que pourraient-elles faire d'autre ? Effectivement, l'activité spontanée continue de former des réseaux complexes alors même que la personne scannée est endormie, anesthésiée, dans le coma ou en état végétatif[51]. Cependant, l'histoire ne s'arrête pas là : une partie de l'activité semble refléter les pensées de la personne. Ainsi, l'un des réseaux de repos, qu'on appelle le « mode par défaut », apparaît dès que la personne scannée dispose d'un peu de temps pour réfléchir à sa vie, pour évoquer des souvenirs autobiographiques ou pour comparer ses pensées à celles des autres[52]. Si on laisse un individu se reposer dans le scanner, son cerveau bascule fréquemment dans ce « mode par défaut ». Lorsqu'on lui demande alors : « À quoi étiez-vous en train de penser ? », il répond souvent qu'il était plongé dans des réflexions et des souvenirs intimes – plus souvent que si on l'interrompt à un autre moment[53]. Ainsi, le réseau qui s'active spontanément prédit, au moins partiellement, l'état mental de la personne scannée.

En un mot, nos ruminations mentales résultent d'incessantes décharges neuronales spontanées. De plus, ce flux de pensées endogènes entre en compétition avec la conscience du monde extérieur. Pendant les périodes d'intense activité cérébrale « par défaut », l'onde P3 évoquée par un stimulus extérieur disparaît[54]. Les états internes de conscience interfèrent avec la prise de conscience d'événements externes. L'activité spontanée envahit l'espace de travail global et, lorsqu'elle nous absorbe, elle peut bloquer la conscience d'autres stimuli pendant un temps considérable. Nous avons rencontré une variante de ce phénomène au chapitre 1 sous le terme de « cécité inattentionnelle ».

Mes collègues et moi-même avons eu le plaisir de constater que ce phénomène survenait spontanément dans notre simulation informatique[55]. Des périodes d'embrasement spontané survenaient sous nos yeux, particulièrement lorsque le paramètre de vigilance était élevé. Lorsque nous stimulions le réseau dans ces moments-là, même si l'entrée sensorielle était nettement au-dessus du seuil, sa progression au sein des aires sensorielles était bloquée, et elle ne parvenait pas à déclencher d'embrasement global, parce que l'activité interne entrait en compétition avec la stimulation externe. Notre simulation imitait ainsi la cécité inattentionnelle et le clignement de l'attention, deux phénomènes psychologiques classiques qui illustrent l'incapacité de notre espace de travail à prêter attention à deux choses à la fois.

L'activité spontanée expliquait également pourquoi le même stimulus était tantôt perçu, tantôt non. Tout dépendait de la forme exacte des fluctuations neuronales juste avant la stimulation. Dans notre simulation, comme dans les données réelles, les variations stochastiques de l'activité cérébrale biaisaient la perception d'un stimulus faible[56].

Darwin dans le cerveau

L'activité spontanée est l'un des aspects fréquemment négligés du modèle de l'espace de travail neuronal global. Il s'agit pourtant d'un de ses traits les plus importants et les plus originaux. Trop de neuroscientifiques adhèrent encore, implicitement, à la métaphore obsolète de l'arc réflexe[57]. Cette idée, qui remonte à René Descartes, Charles Sherrington et Ivan Pavlov, réduit le cerveau à un dispositif d'entrée-sortie qui se contente de transformer les entrées sensorielles en commandes musculaires, comme dans le fameux schéma de Descartes où l'œil commande la main (figure 2). Or nous savons aujourd'hui à quel point cette vision est erronée. L'autonomie est la première propriété du système nerveux. L'activité neuronale intrinsèque domine sur l'excitation externe. Notre cerveau ne s'abandonne jamais passivement à son environnement, mais il engendre en permanence ses propres configurations d'activité. Au cours du développement, les configurations utiles sont préservées, tandis que les autres disparaissent avec l'apprentissage[58]. Cet algorithme joyeusement créatif, qui teste et qui rejette, se manifeste tout particulièrement dans le comportement exploratoire des jeunes enfants. Il soumet chacune de nos pensées à un processus de sélection darwinienne.

Cette idée n'avait pas échappé à William James, qui en faisait un point important de sa théorie. « Pourquoi ne pas dire que, comme la moelle épinière est une machine pourvue d'un petit nombre de réflexes, les hémisphères cérébraux sont une machine qui en comprend beaucoup, ce qui fait toute la différence ? » Parce que, se répondait-il à lui-même, les circuits les plus évolués du cerveau opèrent comme « un organe dont l'état naturel est un équilibre instable », qui permet « à son possesseur d'adapter sa conduite aux plus minuscules altérations du contexte ».

Notre conscience n'est donc pas soumise à son environnement : elle lui impose un sens, un contexte, un but. L'origine de cette faculté réside dans la capacité des cellules nerveuses à s'autoexciter : très tôt dans l'évolution, les neurones ont acquis la possibilité de déclencher spontanément des potentiels d'action. Filtrée et amplifiée par des circuits

bouclés sur eux-mêmes, cette excitabilité se transforme en comportements exploratoires, dirigés vers des buts que l'organisme se fixe de lui-même. Tout animal explore son environnement d'une manière semi-aléatoire, en s'appuyant sur une hiérarchie de « générateurs centraux du mouvement » (*central pattern generators*) – des réseaux de neurones dont l'activité spontanée engendre les rythmes de la marche, de la nage ou de la mastication.

Chez les primates et probablement chez bien d'autres espèces, une exploration similaire se produit à l'intérieur même du cerveau, d'une façon cognitive plutôt que motrice. En produisant spontanément des états changeants d'activité neuronale, en l'absence de toute stimulation, l'espace de travail global nous permet d'engendrer sans cesse de nouveaux plans, de les tester et de les modifier à volonté s'ils ne remplissent pas nos attentes.

Un processus darwinien de variation suivi de sélection se produit dans l'espace de travail global[59]. L'activité spontanée opère comme un « générateur de diversité » dont les configurations sont perpétuellement sculptées par les systèmes cérébraux d'anticipation de la récompense. Les réseaux de neurones conçus selon ce principe peuvent s'avérer très puissants. Dans nos simulations informatiques, Jean-Pierre Changeux et moi-même avons montré qu'ils peuvent résoudre des problèmes complexes tels que le casse-tête classique de la tour de Londres[60]. La logique de l'apprentissage par sélection, combinée avec des règles classiques de modification des poids synaptiques, conduit à une architecture robuste, capable d'apprendre de ses erreurs et d'extraire les règles abstraites qui sous-tendent un problème[61].

En anglais, le *generator of diversity* s'abrège en GOD ! N'allez pas croire pour autant que la notion d'activité spontanée soit une propriété mystérieuse. Elle n'implique aucune étincelle divine, aucune action de l'esprit sur la matière, ni aucun retour au dualisme. L'excitabilité est une propriété physique des cellules nerveuses. Dans chaque neurone, le potentiel de membrane fluctue sans cesse, notamment à cause de la

libération stochastique des vésicules de neurotransmetteur au niveau de la synapse. En dernière analyse, ces fluctuations résultent du bruit thermique qui secoue en permanence nos molécules. On aurait pu penser que l'évolution minimiserait l'impact de ce bruit (comme le font les concepteurs de puces électroniques lorsqu'ils choisissent des niveaux de voltage très différents pour coder les valeurs binaires 0 et 1, en sorte que le bruit thermique ne puisse pas en modifier la valeur). Mais ce n'est absolument pas le cas dans le cerveau : non seulement les neurones tolèrent le bruit, mais ils l'amplifient, probablement parce que ces fluctuations aléatoires facilitent la résolution de certains problèmes (de nombreux algorithmes tels que les « méthodes de Monte-Carlo par chaînes de Markov » ou le « recuit simulé » exigent une bonne source de bruit).

Dès que les fluctuations du potentiel de membrane excèdent un certain seuil, le neurone émet un potentiel d'action. Nos simulations montrent que ces décharges aléatoires, sculptées par les faisceaux de connexions qui relient les neurones, font émerger des configurations complexes d'activité. Ce qui commence par un « bruit » sans signification se termine par une avalanche structurée qui engendre nos pensées secrètes. Il n'y a pas de plus belle leçon d'humilité que de prendre conscience que le flux de notre conscience, tout ce flot d'images et de mots qui surgit en nous et forme la texture même de notre vie mentale, provient des décharges aléatoires de nos neurones corticaux, sculptées par les trillions de synapses que des années de maturation cérébrale et d'éducation ont inscrites en nous.

Un catalogue de l'inconscient

Ces dernières années, la théorie de l'espace de travail neuronal global a pris de l'importance. Pour bon nombre de chercheurs, elle est devenue le principal prisme d'interprétation des observations empiriques. L'un de ses

succès a consisté à proposer une classification systématique des processus inconscients. De la même manière que le naturaliste suédois Carl von Linné, au XVIIIe siècle, avait proposé une première « taxonomie » du vivant (un système de classification des plantes et des animaux), nous commençons à envisager une taxonomie de l'inconscient.

Souvenez-vous du message principal du chapitre 2 : la plupart des calculs qu'effectue notre cerveau sont inconscients. Nous n'avons aucune conscience de la majorité des processus qui s'y déroulent, depuis la respiration jusqu'au contrôle de notre posture, de la vision de bas niveau jusqu'aux mouvements fins de nos doigts, ou des statistiques de l'orthographe jusqu'aux règles grammaticales de notre langue maternelle. À notre insu, une étonnante profusion de processeurs inconscients façonne nos pensées et nos actes.

La théorie de l'espace de travail global met un peu d'ordre dans cette jungle[62]. Elle catalogue les capacités inconscientes en fonction des mécanismes cérébraux mis en jeu (figure 28). Commençons par la cécité inattentionnelle. Un stimulus est présenté bien au-dessus du seuil normal de perception consciente – et pourtant nous ne le voyons pas, parce que notre esprit est occupé ailleurs. J'écris ces mots dans la maison de mes beaux-parents, une gentilhommière du XVIIe siècle où trône une belle horloge comtoise. Son pendule oscille sous mes yeux, et j'entends aisément son tic-tac – mais dès que je me concentre sur mon texte, ce mouvement et ce bruit s'évanouissent : l'inattention bloque la prise de conscience.

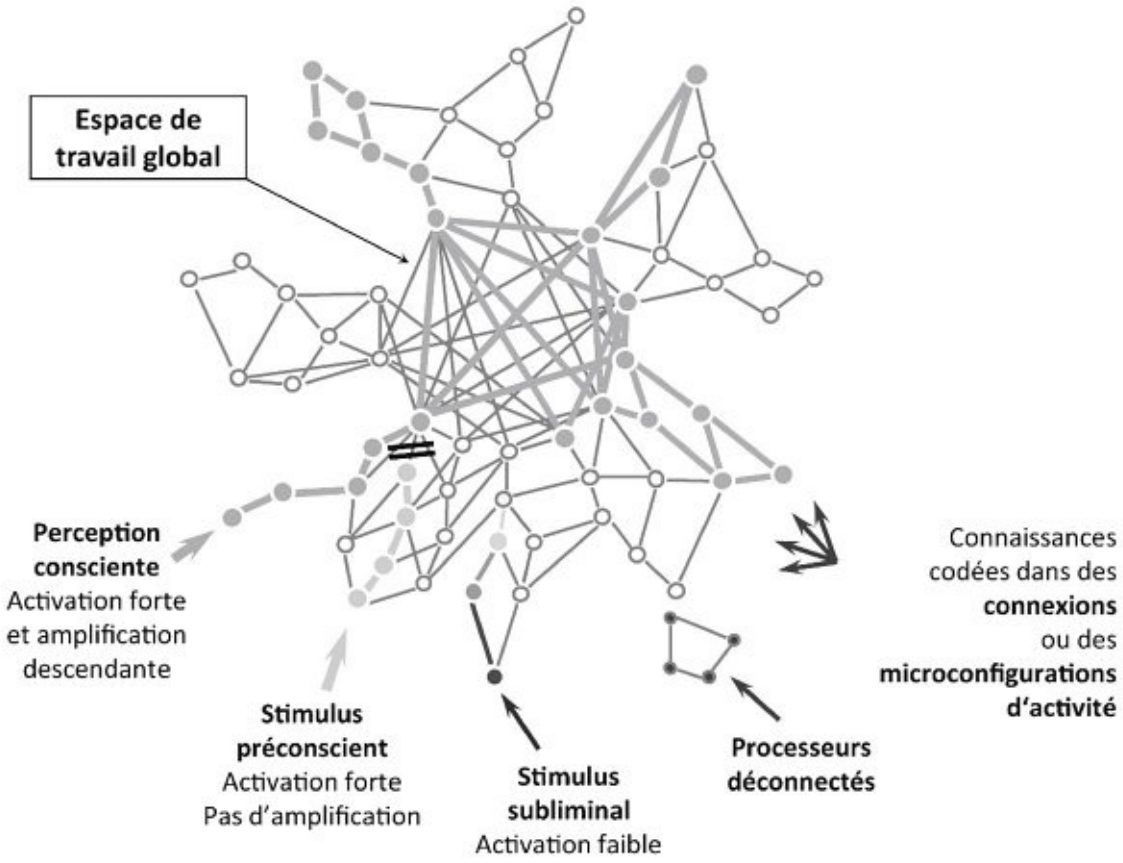


Figure 28. De multiples raisons peuvent expliquer qu'une information demeure inconsciente. À chaque instant, une seule pensée consciente embrase l'espace de travail. D'autres représentations n'accèdent pas à la conscience, soit parce qu'elles ne sont pas amplifiées par l'attention (*préconscient*) ou parce qu'elles sont trop faibles pour entraîner une avalanche d'activité globale (*subliminal*). Nous n'avons également aucune conscience des informations qui sont codées par des neurones déconnectés de l'espace de travail. Enfin, quantité d'informations demeurent inconscientes parce qu'elles sont cryptées dans les connexions et dans les microconfigurations de l'activité cérébrale.

Dans notre catalogue de l'inconscient, mes collègues et moi-même avons proposé d'appeler « préconsciente » toute information de cette nature[63]. Le préconscient est du conscient en puissance : il s'agit d'une information codée par une assemblée de neurones actifs, qui pourrait devenir consciente à tout moment, si seulement on lui prêtait attention – mais ce n'est pas le cas. Le terme *préconscient* est dû à Sigmund Freud

qui, dans son *Abrégé de psychanalyse*, observait que « maints processus deviennent facilement conscients, puis cessent de l'être pour ensuite le redevenir sans difficulté. (…) De cette partie d'inconscient, qui tantôt reste inconsciente, tantôt devient consciente, nous dirons qu'elle est "capable de devenir consciente" et nous lui donnerons de préférence le nom de préconscient ».

Les simulations de l'espace de travail neuronal global fournissent un mécanisme neuronal plausible du préconscient[64]. Lorsqu'un stimulus pénètre dans la simulation, son activité se propage et entraîne un embrasement global. À son tour, cette représentation consciente engendre une frange d'inhibition qui empêche d'autres informations d'entrer dans la conscience au même instant. Cette compétition centrale est inévitable. J'ai déjà souligné qu'une représentation consciente est définie au moins autant par ce qu'elle n'est pas que par ce qu'elle est. Selon notre hypothèse, de nombreux neurones de l'espace de travail doivent être réduits au silence afin de délimiter les contours de l'objet conscient et de cerner ce qu'il n'est pas. Cette inhibition diffuse engendre un goulot d'étranglement au sein des aires supérieures de la hiérarchie corticale. Elle nous empêche de voir simultanément deux images, ou de faire deux choses en même temps. Cependant, elle ne bloque pas l'activité des aires sensorielles périphériques – celles-ci s'allument, pratiquement au même niveau que d'habitude, même quand l'espace de travail est déjà occupé par un premier objet de pensée. L'information préconsciente est donc tenue à l'écart de l'espace de travail conscient, dans une mémoire tampon sensorielle. Là, elle tombe progressivement dans l'oubli, à mesure que l'activité neuronale revient à son niveau de base – sauf si nous décidons de lui prêter attention. Pendant quelques instants, l'information préconsciente évanescente peut encore être sauvée de l'oubli et portée à notre conscience, auquel cas nous la percevons de façon rétrospective[65].

L'état préconscient diffère nettement d'une seconde forme d'inconscient que nous avons appelée l'état *subliminal*. Présentez une image brièvement et faites-la suivre par une autre forme, en sorte que, par

effet de masquage, elle demeure invisible. Dans ce cas, à la différence du préconscient, l'attention ne change rien – quels que soient nos efforts, nous ne parvenons pas à percevoir l'image cachée. Prise en sandwich entre d'autres formes, une image subliminale nous échappe définitivement. Elle induit pourtant une activité détectable dans les aires visuelles, sémantiques et motrices du cortex, mais cette activation est trop brève pour déclencher un embrasement global. Là encore, nos simulations rendent compte de cet état. Dans l'ordinateur également, une impulsion trop courte ne parvient pas à embraser l'ensemble du réseau. Pourquoi ? Parce que le temps que l'activation progresse jusqu'aux aires supérieures, puis en redescende pour amplifier l'entrée, celle-ci a déjà disparu, remplacée par le masque[66]. L'état *subliminal* est une pure invention du psychologue qui, en présentant des stimuli trop faibles, trop courts ou trop encombrés, joue des tours au cerveau et entrave l'embrasement global. Dans notre taxonomie, nous proposons de réserver le terme « subliminal » aux situations où l'onde d'activité sensorielle meurt avant d'avoir pu créer une avalanche au sein de l'espace de travail, quels que soient nos efforts pour l'amplifier en y prêtant attention.

La distinction entre le préconscient et le subliminal n'épuise pas encore notre catalogue de connaissances inconscientes. Prenez l'exemple de la respiration. À chaque minute de votre vie, d'harmonieuses oscillations des potentiels d'action, issues du plus profond de votre tronc cérébral et transmises aux muscles de votre poitrine, contrôlent le processus de ventilation qui vous maintient en vie. D'ingénieuses boucles de rétroaction adaptent leur rythme en fonction de la concentration d'oxygène et de gaz carbonique dans votre sang. Toute cette activité neuronale échappe à la conscience. Pour quelle raison ? Elle est intense et durable, donc elle n'est pas subliminale ; et l'attention ne l'aide en rien, donc elle n'est pas préconsciente. Dans notre taxonomie, cette situation constitue une troisième catégorie de représentation inconsciente : l'*activité déconnectée*.

Enfermées dans les circuits du tronc cérébral, les décharges neuronales qui contrôlent la respiration ne peuvent pas être transmises à l'espace de travail formé par les aires pariétales et préfrontales. Ces informations respiratoires sont séquestrées à vie dans les oubliettes d'un circuit qui ne communique pas avec le reste du cortex. C'est pourquoi elles demeurent inconscientes. Nombreuses sont les informations qui demeurent ainsi en retrait du cortex, faute de connexions qui leur permettraient d'accéder à la conscience. La seule manière d'y accéder est de les recoder dans une autre modalité sensorielle. Nous ne prenons conscience de notre respiration que de manière indirecte, lorsque nous prêtons attention aux mouvements de notre poitrine : ce sont les signaux musculaires qui nous permettent de la percevoir.

Bien que nous ayons le sentiment d'être aux commandes de notre corps, des centaines de signaux neuronaux parcourent en permanence nos circuits cérébraux sans jamais percer le seuil de la conscience, car ils sont déconnectés des régions corticales de l'espace de travail. Chez certains patients, à la suite d'un accident vasculaire, la situation empire. Une lésion des connexions du cortex peut rendre inaccessibles des informations dont une personne saine a pleinement conscience. Par exemple, la déconnexion d'une partie du corps calleux, ce vaste faisceau de fibres qui relie les deux hémisphères, peut faire perdre au patient toute conscience de ses propres plans d'action. Il va jusqu'à renier les mouvements de sa propre main gauche, en affirmant qu'elle est devenue incontrôlable. En réalité, la commande de la main gauche provient des aires motrices de l'hémisphère droit, alors que les commentaires verbaux sont issus de l'hémisphère gauche. Lorsque ces deux circuits sont déconnectés, tout se passe comme si le patient abritait deux espaces de travail dans la même tête, chacun devenu partiellement inconscient de ce que concocte l'autre.

Une quatrième possibilité de rendre une information inconsciente, selon la théorie, consiste à la *diluer* au sein d'une configuration complexe de décharges neuronales. Prenons l'exemple d'une image à la Buren,

constituée exclusivement de rayures noires et blanches. Si les rayures sont trop finement espacées, ou si elles clignotent trop vite (plus de cinquante fois par seconde), vous cessez de les percevoir. L'expérience montre pourtant que, là où vous ne voyez qu'un gris uniforme, votre cortex visuel continue de « voir » la grille : les neurones de l'aire visuelle primaire continuent d'en détecter l'orientation[67]. Pourquoi ce code neural ne peut-il pas être amené à la conscience ? Probablement parce qu'il est formé d'un embrouillamini de décharges, une configuration spatio-temporelle trop complexe pour être décodée par les neurones des aires corticales supérieures. Avant de devenir consciente, une information sensorielle doit être recodée sous une forme explicite par une assemblée compacte de neurones. Les régions antérieures du cortex visuel doivent dédier des neurones spécifiques aux entrées sensorielles qui font sens, pour que cette activité elle-même puisse être amplifiée et mise en relation avec les aires pariétales et préfrontales. Si l'information reste diluée dans de subtiles altérations des décharges de millions de neurones, cela lui interdit tout accès à la conscience.

Chaque mot que nous entendons entame ainsi sa carrière dans l'aire auditive visuelle primaire (V1) où il est codé de façon inconsciente sous la forme d'une série absurdement complexe de millions de décharges neuronales. Cette configuration spatio-temporelle d'activité contient une quantité presque infinie d'informations sur la personne, sa taille, son message, ses émotions, l'endroit où elle se trouve… si seulement nous pouvions les décrypter – mais c'est impossible. Nous ne prenons conscience de cette information latente que lorsque nos aires cérébrales de haut niveau parviennent à la catégoriser dans des cases qui font sens. Rendre le message neuronal explicite, telle est la fonction essentielle de la pyramide de neurones qui, de proche en proche, extrait des informations de plus en plus abstraites de l'entrée sensorielle. Avec de l'entraînement, nous parvenons à prendre conscience de sons faibles ou d'autres aspects de nos sensations, parce qu'à tous les niveaux les neurones réorientent leurs préférences afin d'amplifier ces messages sensoriels[68]. Avant

l'apprentissage, le message neuronal était déjà présent dans nos aires sensorielles, mais seulement d'une manière implicite, sous la forme d'une configuration de décharges neuronales diluée, donc inaccessible à la conscience. Après l'apprentissage, il est devenu explicite et accessible à la conscience.

Ce phénomène a une conséquence fascinante : le cerveau contient des signaux dont même son propriétaire ignore l'existence[69]. L'imagerie cérébrale commence à savoir décoder ces cryptogrammes neuronaux. L'armée américaine mène un programme de recherches qui consiste à présenter à un observateur entraîné des photos satellites au rythme ultrarapide de 10 images par seconde. On recherche alors, dans ses ondes cérébrales, la marque inconsciente de la présence d'un avion ennemi. Notre inconscient regorge d'informations insoupçonnées qui attendent d'être déchiffrées. À l'avenir, en amplifiant ces microcodes neuronaux implicites que notre conscience ne parvient pas à déceler, le décodage cérébral assisté par ordinateur pourrait nous conférer une forme rigoureuse de perception extrasensorielle – un sens plus aiguisé de ce qui nous entoure.

Enfin, il existe encore une cinquième catégorie de connaissances inconscientes : les *connexions latentes*. Selon la théorie de l'espace de travail neuronal, nous ne pouvons prendre conscience d'une information que si elle est codée sous la forme d'une configuration d'activité neuronale, une assemblée active de cellules. Cependant, des quantités bien plus importantes d'informations sont stockées dans les connexions qui relient les neurones entre eux. Avant même notre naissance, nos neurones s'adaptent aux statistiques du monde extérieur en modifiant leurs synapses. Les centaines de milliers de milliards de synapses qui parsèment notre cortex gardent une trace latente de toute notre vie. Des millions de synapses se font et se défont chaque jour, particulièrement au cours des premières années de vie, quand notre cerveau s'adapte à son environnement.

Chaque synapse n'enregistre qu'un minuscule fragment de savoir statistique : quelle est la probabilité que mon neurone présynaptique décharge juste avant mon neurone postsynaptique ? Cumulées, toutes ces informations finissent par poser les fondements d'une riche intuition du monde. Dans les aires visuelles, nos connexions corticales accumulent des statistiques sur les combinaisons de traits et de couleurs qui composent des images[70]. Dans les régions auditives et motrices s'inscrit une intuition inconsciente de la musique : des années de pratique du piano y induisent des changements détectables de la densité de matière grise, sans doute liés à la densité de synapses, à la taille des dendrites, ainsi qu'au grossissement des axones et des cellules gliales qui leur viennent en aide[71]. Dans les connexions de l'hippocampe (une structure qui s'enroule sur elle-même en dessous du lobe temporal) s'accumulent les souvenirs des événements forts de nos vies : où, quand, comment ils sont survenus.

Une fois codées par un foisonnement d'épines dendritiques, ces mémoires synaptiques peuvent demeurer latentes pendant des années. Nous ne pouvons pas y puiser directement, parce que leur format diffère radicalement des configurations d'activité neuronale qui représentent nos pensées conscientes. Pour qu'un souvenir remonte à la conscience, nous devons d'abord le convertir de l'état latent à la forme active. Au cours de l'accès à la mémoire, nos synapses permettent la reconstitution d'une assemblée neuronale proche de celle d'origine – et ce n'est qu'alors que nous prenons conscience de cette réminiscence. Un souvenir conscient n'est que la reconstruction approximative d'une configuration de décharges neuronales qui a traversé notre cerveau par le passé. L'imagerie cérébrale confirme que les souvenirs latents doivent être convertis en configurations explicites d'activité qui envahissent le cortex préfrontal et les régions associées, au moment précis où nous reprenons conscience d'un épisode passé de nos vies[72]. Cette réactivation d'un réseau d'aires corticales distantes correspond parfaitement aux prédictions de la théorie de l'espace de travail.

La distinction entre connexions latentes et assemblées neuronales actives explique pourquoi nous n'avons aucune conscience des règles de grammaire qui nous permettent de parler et de comprendre notre langue maternelle. Dans la phrase : « Jean croit qu'*il* est intelligent », le pronom *il* peut-il faire référence à Jean ? Oui. Et dans : « *Il* croit que Jean est intelligent » ? Non. Et dans : « La vitesse avec laquelle *il* a résolu le problème a fait plaisir à Jean. » Oui – bien que le mot « Jean » vienne après le pronom. Nous connaissons la réponse sans avoir la moindre idée des règles que nous appliquons, et que les linguistes s'évertuent toujours à découvrir. Nos aires du langage sont câblées pour traiter les mots et les phrases, mais le plan de ce câblage reste inaccessible à la conscience. La théorie explique pourquoi : cette connaissance, stockée dans les connexions, n'est tout simplement pas au bon format.

La grammaire diffère de l'arithmétique. Lorsque nous multiplions 24 par 31, nous sommes conscients de nos processus mentaux. Notre introspection parvient à suivre, pas à pas, chacune des opérations intermédiaires, leur nature, leur ordre, et même les erreurs que nous commettons. Le contraste est vif avec la production d'une phrase, situation dans laquelle nous restons paradoxalement muets sur la nature des processus internes qui nous ont permis de l'exprimer. Les problèmes que résout notre processeur grammatical ne sont pas moins complexes que ceux de l'arithmétique, mais nous n'avons pas la moindre idée de la manière dont nous les débrouillons. Pourquoi cette différence ? Un calcul arithmétique s'effectue pas à pas sous le contrôle des aires de l'espace de travail global (le cortex pariétal, préfrontal et cingulaire). Cette séquence est codée explicitement par des assemblées de neurones du cortex préfrontal : certains neurones y représentent les intentions, d'autres les plans d'action, les étapes individuelles, leur nombre, leur ordre et même la détection et la correction d'une erreur[73]. Ainsi, en arithmétique, le cortex préfrontal, élément clé de l'espace de travail conscient, contient un code neural de toutes les étapes. Les calculs grammaticaux sont bien différents : on pense qu'ils reposent sur des circuits spécialisés du sillon

temporal supérieur et du gyrus frontal inférieur gauche (la fameuse « aire de Broca »), en évitant les réseaux du traitement conscient et volontaire qui passent par le cortex préfrontal dorso-latéral[74]. Au cours de l'anesthésie, une partie des aires du langage continue de traiter les phrases de façon autonome, sans conscience[75]. Nous ne savons pas encore comment les neurones représentent les règles de grammaire – mais on peut prédire que leur code neural s'écarte radicalement de celui de l'arithmétique.

Les états subjectifs de la matière

En résumé, la théorie de l'espace de travail neuronal global rend compte d'un grand nombre d'observations sur la conscience et ses bases cérébrales. Elle explique pourquoi nous ne prenons conscience que d'une infime partie des connaissances stockées dans notre cerveau. Avant de devenir consciente, une information doit d'abord être codée par une assemblée de neurones active, à la fois dans les aires sensorielles de haut niveau, et dans les régions du cortex préfrontal et pariétal qui constituent l'espace de travail. L'embrasement synchrone de toutes ces régions rend compte des signatures de la conscience identifiées par l'imagerie cérébrale.

Les simulations informatiques de mon laboratoire reproduisent certains aspects de l'accès à la conscience, mais elles sont encore loin d'imiter le cerveau en détail – et la simulation est donc loin d'être consciente ! Sur le principe, je n'ai aucun doute qu'un programme d'ordinateur puisse un jour reproduire un état de conscience. Mais cette simulation devrait posséder des milliards d'états neuronaux différenciés. Elle ne devrait pas juste faire circuler l'activité neuronale, mais l'utiliser pour réaliser des inférences statistiques, par exemple calculer la

probabilité de la présence d'un visage ou les chances qu'un geste atteigne son but.

Nous commençons à comprendre comment un réseau de neurones pourrait réaliser ce genre de calcul[76]. Certains neurones spécialisés prennent des décisions perceptives en accumulant les données issues des organes sensoriels[77]. Au cours de l'embrasement conscient, une fraction de ces neurones converge vers une interprétation unifiée de la scène. Le franchissement du seuil d'embrasement décide, d'une façon irrévocable, de ce qui a été perçu et de la conduite à tenir.

Imaginez-vous l'espace de travail global comme une vaste arène où de multiples régions cérébrales, tels les démons du pandémonium de Selfridge, se disputent le devant de la scène. Leurs règles de fonctionnement les forcent à rechercher un consensus, une interprétation unique et cohérente des messages qu'ils reçoivent. Leurs connexions à longue distance leur permettent de confronter les signaux issus de toute l'arène, et d'accumuler ainsi les connaissances à une échelle globale. Ils en débattent, parfois pendant quelques centaines de millisecondes, jusqu'à ce qu'une réponse cohérente émerge qui satisfasse les buts actuels de l'organisme.

La machine tout entière n'est pas complètement soumise à ses entrées externes. L'autonomie est son point fort. Elle engendre ses propres buts, grâce à son activité spontanée, et ces configurations neuronales émergentes façonnent à leur tour, de façon descendante, le reste de l'activité du cerveau. Elles incitent les autres régions à récupérer des souvenirs dans la mémoire à long terme, à en former des images mentales et à les transformer selon des règles intuitives, linguistiques ou logiques. Un flux constant d'activité neuronale circule dans l'espace de travail et se transforme en information qui passe au crible de millions de processeurs inconscients et parallèles. Chaque résultat cohérent, chaque embrasement conscient nous fait avancer d'un pas dans cet algorithme mental qui ne s'arrête jamais – le « flux de la conscience » cher à William James.

Il serait fascinant de parvenir à simuler une telle machine statistique, à la fois sérielle et massivement parallèle. En Europe, un réseau de laboratoires, regroupé au sein du « Projet du cerveau humain » (Human Brain Project), propose purement et simplement de reproduire, dans un superordinateur, la géométrie et les propriétés moléculaires de tous les neurones et toutes les connexions d'un cerveau humain. Des puces « neuromorphiques » spécialisées dans la simulation de la physique des réseaux de neurones devraient accélérer les calculs et permettre de simuler quelques centaines de millions de neurones dans le moindre détail[78]. Toutefois, simuler suffira-t-il à comprendre ? Parallèlement à ces efforts de simulation, il est indispensable de développer la théorie mathématique des calculs neuronaux. Ce n'est qu'à ce prix que l'on parviendra à une théorie intégrée de l'activité cérébrale objective et de l'expérience consciente subjective.

CHAPITRE 6

L'épreuve de vérité

Toute théorie de la conscience doit affronter l'épreuve de vérité : la clinique. Chaque année, des milliers de patients plongent dans le coma. Bon nombre d'entre eux sombrent dans un état mental parfois difficile à déterminer. Sont-ils conscients ? La nouvelle science de la conscience peut-elle les aider ? Le rêve d'un « conscientimètre », un dispositif qui serait capable d'évaluer le degré de conscience résiduelle de n'importe quel patient, n'est pas hors de portée. Des analyses mathématiques sophistiquées des signaux cérébraux indiquent que certains patients sont toujours conscients, mais incapables de communiquer. Des traitements innovants apparaissent. La stimulation des régions profondes du cerveau parvient parfois à accélérer le recouvrement de la conscience. Les interfaces cerveau-machine peuvent également permettre aux patients conscients, mais totalement paralysés, de communiquer leurs pensées. À l'avenir, ces neurotechnologies vont radicalement modifier la prise en charge des pathologies de la conscience.

« Ne demande pas, lecteur, quelle fut alors mon épouvante : je ne la peindrai pas dans ces chants ; mes expressions seraient impuissantes. Je ne mourus pas, et je ne restai pas vivant : si tu as quelque génie, pense à ce que je devins dans cet état où j'étais hors de la vie et de la mort. »

DANTE ALIGHIERI, *L'Enfer*,
chant XXXIV (env. 1307-1321).

Chaque année, un nombre épouvantable d'accidents de la route, de suicides manqués, de noyades, d'accidents vasculaires ou d'intoxications au monoxyde de carbone rendent invalides des dizaines d'adultes et d'enfants. Quadriplégiques, incapables de se mouvoir et parfois de parler, ils semblent avoir perdu l'étincelle même de la vie. Et pourtant… au sein de leur cortex subsiste parfois un esprit vivace. Dans *Le Comte de Monte-Cristo,* Alexandre Dumas a peint une image terriblement réaliste de cette persistance de la conscience, enterrée vivante dans le caveau d'un corps inerte :

> « M. Noirtier, immobile comme un cadavre, regardait avec des yeux intelligents et vifs ses enfants (…). La vue et l'ouïe étaient les deux seuls sens qui animassent encore, comme deux étincelles, cette matière humaine déjà aux trois quarts façonnée pour la tombe ; encore, de ces deux sens, un seul pouvait-il révéler au dehors la vie intérieure qui animait la statue : et le regard qui dénonçait cette vie intérieure était semblable à une de ces lumières lointaines qui, durant la nuit, apprennent au voyageur perdu dans un désert qu'il y a encore un être existant qui veille dans ce silence et cette obscurité. (…) C'était un cadavre avec des yeux vivants, et rien n'était plus effrayant parfois que ce visage de marbre au haut duquel s'allumait une colère ou luisait une joie. »

Noirtier est un personnage de roman – sans doute la toute première description littéraire du syndrome d'enfermement (*locked-in syndrome*). Sa pathologie, par contre, est bien réelle. Jean-Dominique Bauby, l'éditeur du magazine *Elle,* n'avait que 43 ans quand sa vie prit un tour dramatique : « Jusqu'alors, explique-t-il, je n'avais jamais entendu parler du tronc cérébral. Ce jour-là, j'ai découvert de plein fouet cette pièce maîtresse de notre ordinateur de bord, passage obligé entre le cerveau et les terminaisons nerveuses, quand un accident cardiovasculaire a mis ledit tronc hors circuit. »

Ce jour-là, le 8 décembre 1995, Bauby plongeait dans un coma qui allait durer vingt et un jours. Il en ressortit pour se retrouver allongé sur un lit d'hôpital, presque entièrement paralysé, ne parvenant plus à effectuer que quelques mouvements de l'œil et de la tête. Il survécut quinze mois, que cet homme hors du commun consacra à imaginer, mémoriser, dicter et publier un livre. Témoignage exceptionnel de la vie mentale d'un patient *locked-in*, *Le Scaphandre et le Papillon* devint instantanément un best-seller mondial. Embastillé dans un corps inanimé, tel un Noirtier contemporain, Jean-Dominique Bauby dicta son livre lettre

à lettre, en clignant de la paupière gauche tandis qu'un assistant récitait les lettres E, S, A, R, I, N, T, U, L, O, M. Deux cent mille clins d'œil racontent l'histoire émouvante d'un bel esprit qu'un accident vasculaire cérébral a condamné à la prison à vie. Une pneumonie finit par l'emporter, trois jours seulement après la parution de son livre.

Avec des mots sobres, souvent pince-sans-rire, l'ancien éditeur à succès décrit son enfer quotidien, un tissu de frustration, d'isolement, d'incommunicabilité et, parfois, de pur désespoir. Bien qu'il soit le prisonnier d'un corps sans vie qu'il compare à un scaphandre de plongée, sa prose élégante et concise s'élève avec la grâce d'un papillon – métaphore d'une pensée qui reste libre de voleter où elle veut. Il n'existe pas de meilleure preuve de l'autonomie de la conscience que l'écriture alerte de Jean-Dominique Bauby :

> « Pour le plaisir, j'ai recours à la mémoire vive des goûts et des odeurs, un inépuisable réservoir de sensations. Il y avait l'art d'accommoder les restes. Je cultive celui de mitonner les souvenirs. (…) Le bourguignon est onctueux, le bœuf en gelée translucide, et la tarte à l'abricot a la pointe d'acidité nécessaire. Selon mon humeur, je m'offre une douzaine d'escargots, une choucroute garnie et une bouteille de gewurztraminer "Cuvée vendanges tardives" à la teinte dorée, ou je déguste un simple œuf à la coque accompagné de mouillettes au beurre salé. Quel régal ! Le jaune d'œuf m'envahit le palais et la gorge en longues coulées tièdes. »

À cette lecture, il ne fait aucun doute qu'un immense répertoire d'états mentaux, depuis la vision jusqu'au toucher, de l'odeur la plus savoureuse à l'émotion la plus odieuse, peut subsister au sein d'un corps aussi inerte qu'un cadavre. Mais le plus atroce est que, chez bon nombre de patients semblables à Bauby, la présence d'une intense vie mentale n'est pas

toujours détectée[1]. Selon une enquête récente de l'Association française du *locked-in syndrome* (fondée par Bauby lui-même), c'est le plus souvent un membre de la famille, et non le personnel soignant, qui décèle les premiers signes de conscience[2]. Pire, après un accident cérébral, il faut attendre deux mois et demi en moyenne avant que le diagnostic soit correctement établi. Certains patients patientent parfois plusieurs années avant que leurs efforts désespérés ne soient enfin détectés. En effet, de leur corps brisé jaillit parfois un tic ou un rictus réflexe qui rend plus difficile encore la détection de leurs mouvements volontaires des yeux ou des paupières. Même dans les meilleurs hôpitaux, près de 40 % des patients initialement jugés « végétatifs » et dépourvus de toute réaction s'avèrent, après un examen approfondi, présenter des signes de conscience minimale[3].

Ces patients représentent un défi pour les neurosciences. Une théorie de la conscience, si elle est effective, devrait expliquer pourquoi certains patients perdent conscience et d'autres non. Surtout, elle devrait leur apporter de l'aide. Si les signatures de la conscience sont tellement faciles à détecter, il est urgent de les mettre au service de ceux qui en ont le plus besoin : les patients non communicants, pour qui la détection d'une trace de conscience peut devenir, littéralement, une question de vie ou de mort. Dans les services de soins intensifs, partout dans le monde, la moitié des décès résultent d'une décision médicale de cesser la réanimation[4]. On ne peut que se demander combien de Noirtier et de Bauby meurent parce que l'équipe médicale n'a pas su détecter leur conscience résiduelle ou entrevoir qu'ils sortiraient prochainement de leur coma et recouvriraient une vie mentale digne d'être vécue.

L'avenir, cependant, s'annonce prometteur. Les neurologues et les spécialistes de l'imagerie cérébrale font des progrès considérables dans l'identification des états de conscience. Ce champ, en pleine évolution, s'oriente à présent vers des méthodes simples, robustes et peu coûteuses, telles que l'électroencéphalographie (EEG), afin de détecter une conscience résiduelle et de rendre à ces patients une partie de leurs

facultés de communication. Penchons-nous à présent sur ce nouveau domaine à la frontière de la science, de la médecine et de la haute technologie.

Perdre conscience

Commençons par passer en revue les différents types de troubles neurologiques de la conscience et de la communication[5] (figure 29). Nous débuterons par le plus familier : le *coma* (du grec κωμα, « sommeil profond »). C'est l'état dans lequel plongent initialement la plupart des patients. Après une atteinte cérébrale, le coma dure généralement de quelques minutes à quelques heures. Ses causes peuvent être des plus diverses : un traumatisme crânien bien sûr (le plus souvent dû à un accident de la route), mais aussi un accident vasculaire cérébral (la rupture ou l'obstruction d'une artère du cerveau), une anoxie (l'interruption de l'alimentation du cerveau en oxygène, elle-même due à un arrêt cardiaque, une intoxication au monoxyde de carbone ou une noyade), ou encore un empoisonnement (un grave excès d'alcool, par exemple). À l'examen clinique, le coma se définit par une perte de connaissance prolongée qu'aucune stimulation ne parvient à dissiper. Le patient reste sans aucune réaction, les yeux fermés. Rien ne le réveille, et il ne présente aucun signe de conscience de soi ni de son environnement. Le terme de coma ne s'applique que si cet état se prolonge au-delà d'une heure (ce qui le distingue de la simple syncope, de la commotion cérébrale ou de la stupeur, qui sont toujours de courte durée).

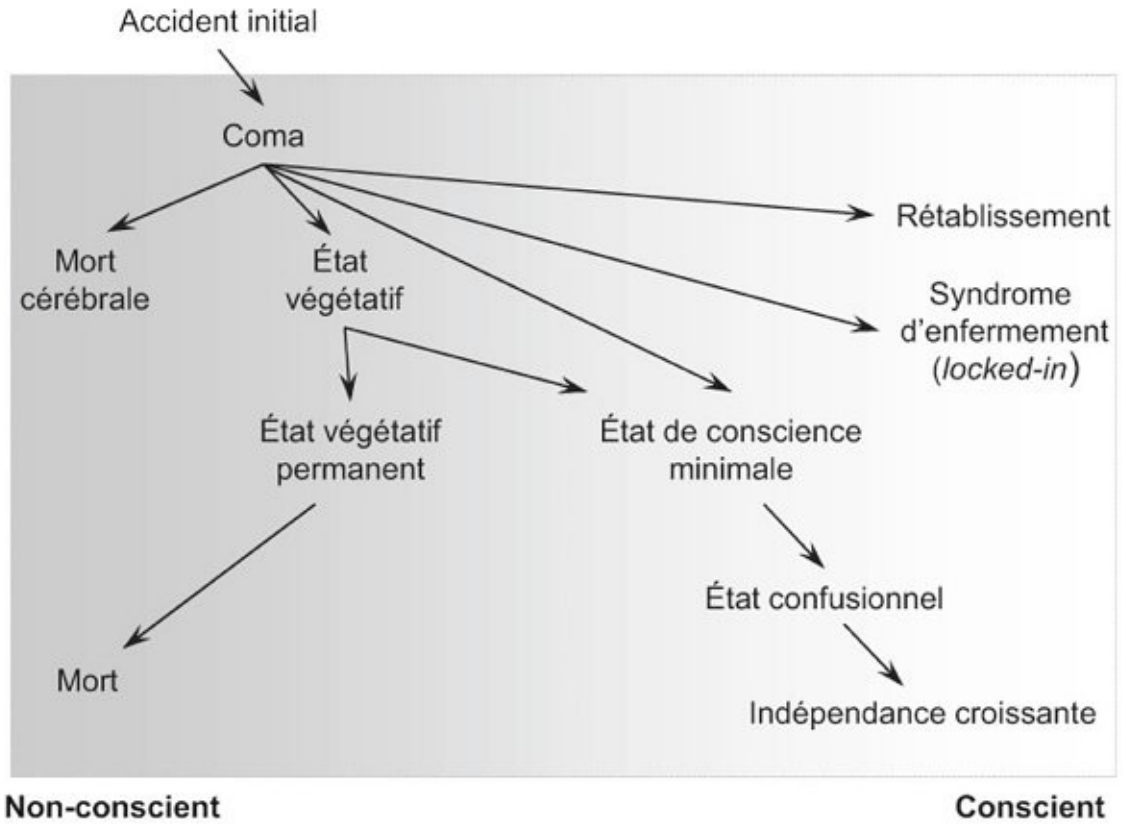


Figure 29. Une lésion cérébrale peut entraîner des troubles de la conscience et de la communication. Les principales catégories de patients sont ici ordonnées de la gauche vers la droite, en fonction de leur degré de conscience résiduelle. Les flèches indiquent l'évolution des patients au fil du temps. L'un des principaux objectifs de la recherche actuelle est de comprendre la différence entre l'état végétatif, dans lequel on n'observe aucun signe de conscience, et l'état de conscience minimale, dans lequel il existe des indices reproductibles de compréhension et d'action volontaire.

Il ne faut surtout pas confondre le coma avec la mort cérébrale. Cette dernière est un état bien différent qui se caractérise par une absence totale de réflexes du tronc cérébral, accompagnée d'un électroencéphalogramme plat et d'une absence de respiration autonome. Chez les patients en état de mort cérébrale, l'imagerie en caméra à positons et d'autres mesures telles que l'échographie à effet Doppler mettent en évidence l'annihilation totale du métabolisme cortical et de la

perfusion sanguine du cerveau. Une fois exclue une hypothermie ou l'effet de drogues ou de toxiques, le diagnostic de mort cérébrale peut être posé au bout de six heures à un jour. Les neurones corticaux et thalamiques dégénèrent très vite. Leurs tissus se délitent, effaçant à tout jamais les traces de mémoire qui constituent l'essence même de la personne. C'est pourquoi l'état de mort cérébral est irréversible : aucune technologie ne pourra jamais ravauder des chaînes moléculaires désagrégées. La plupart des pays, y compris le Vatican[6], identifient donc la mort cérébrale à la mort tout court.

En quoi le coma est-il différent ? Et comment un neurologue distingue-t-il le coma de la mort cérébrale ? Tout d'abord, le corps d'un patient dans le coma continue de présenter certaines réponses organisées. De nombreux réflexes demeurent intacts. Ainsi, la plupart des patients comateux conservent souvent un réflexe nauséeux lorsqu'on stimule le fond de leur gorge. De même, leur pupille continue de se contracter en réponse à une lumière intense. Ces réponses indiquent qu'une partie des circuits inconscients du cerveau, situés dans les profondeurs du tronc cérébral, sont toujours intacts.

L'électroencéphalogramme des patients dans le coma n'a également rien à voir avec une ligne plate. Il continue de fluctuer à un rythme ralenti, en produisant des ondes lentes qui ressemblent à celles que l'on observe au cours du sommeil ou de l'anesthésie. De nombreux neurones corticaux et thalamiques demeurent donc intacts et actifs, mais le réseau dans son ensemble ne retrouve pas son état normal d'activité rapide. Dans de rares cas, on observe des rythmes de haute fréquence, dans les bandes thêta et alpha, mais avec une régularité excessive, comme si de vastes pans du cerveau, au lieu de présenter l'activité fluctuante qui caractérise l'état normal du réseau thalamocortical, étaient saturés d'immenses ondes synchrones[7]. Mon collègue le neurologue Andreas Kleinschmidt aime parler du rythme alpha comme des « essuie-glaces du cerveau » : dans le cerveau sain et conscient, les ondes alpha servent à endormir sélectivement certaines régions, par exemple les aires visuelles lorsque

nous prêtons attention à un son[8]. Dans certains comas, comme dans l'anesthésie au propofol[9] (le sédatif qui a tué Michael Jackson), un rythme alpha géant envahit le cortex frontal et efface jusqu'à la possibilité même d'un état conscient. Mais puisque les neurones sont encore actifs, il se peut qu'ils reviennent un jour à leur rythme normal.

Les patients comateux possèdent donc un cortex actif et capable d'engendrer des signaux d'EEG, mais qui ne parvient plus à émerger de son « sommeil profond ». Fort heureusement, le coma dure rarement longtemps. Après quelques jours, voire quelques semaines, si l'on parvient à éviter infections et autres complications, la grande majorité des patients finissent par récupérer. Le premier signe est généralement le retour d'un cycle de veille et de sommeil. La plupart des patients sortent ensuite du coma et recouvrent progressivement la conscience, la volonté et la capacité de communiquer.

Dans certains cas, hélas, le patient reste dans un étrange état d'éveil sans conscience[10]. Il ouvre les yeux chaque matin, mais durant ces phases de veille, il reste muet, inerte et apparemment inconscient de ce qui l'entoure, perdu dans les limbes d'un enfer semblable à celui de Dante, « hors de la vie et de la mort ». La préservation du cycle veille-sommeil, en l'absence de signes de conscience, est la marque de ce que les neurologues appellent l'« état végétatif », également connu sous le nom d'« éveil sans réponse ». Cet état peut durer des années, car le patient respire de lui-même et peut être alimenté par une sonde ou par la famille. Les Français ont en mémoire la polémique qui a entouré le cas de Vincent Lambert ; Les Américains, eux, se souviennent du cas emblématique de Terri Schiavo, une jeune femme qui passa quinze ans en état végétatif tandis que sa famille se déchirait dans une bataille d'avocats qui vit intervenir l'État de Floride et même le président George W. Bush ; elle finit par mourir en mars 2005 lorsqu'une décision de justice autorisa enfin l'arrêt de son alimentation artificielle.

Que signifie exactement l'adjectif *végétatif* ? Ce mot est regrettable, car il associe le patient à un « légume », une sinistre comparaison qui n'a

hélas pas totalement disparu. Pourtant, ce sont deux neurologues américains, Bryan Jennett et Fred Plum, qui ont formé l'adjectif *végétatif* à partir du verbe *végéter* qui signifie en anglais « vivre une vie purement matérielle, dépourvue d'activité intellectuelle ou de relations sociales[11] ». Effectivement, demeurent intactes la plupart des fonctions vitales qui dépendent du système nerveux autonome, telles que la régulation du rythme cardiaque, de la tension artérielle et de la température corporelle. Le patient n'est pas immobile, et il effectue parfois de grands mouvements spectaculaires du corps ou du regard. Son visage se fend parfois d'un immense sourire, de pleurs ou des froncements des sourcils, sans cause apparente. De tels comportements, on l'imagine, sèment la confusion chez les proches (dans le cas de Terri Schiavo, ils persuadèrent ses parents qu'elle était encore consciente). Mais les neurologues savent que ces réactions, strictement involontaires, peuvent survenir à la manière d'un réflexe : les circuits de la moelle épinière et du tronc cérébral suffisent à les expliquer. Jamais, par contre, le patient ne répond à des commandes verbales. Jamais non plus il ne s'exprime, bien qu'il émette parfois quelques grognements.

Lorsque cet état se prolonge au-delà d'un mois, les médecins parlent d'un « état végétatif persistant », et après trois à douze mois, selon qu'il s'agisse d'une anoxie ou d'un traumatisme crânien, ils posent un diagnostic d'« état végétatif permanent ». Ces termes sont cependant débattus, car ils semblent impliquer qu'aucune récupération n'est possible et peuvent dès lors conduire à une décision prématurée de fin de vie. Plusieurs chercheurs et cliniciens militent en faveur de l'expression « éveil sans réponse », un terme strictement factuel qui ne présage ni du présent ni de l'avenir. La vérité, comme nous allons le voir, est que cet état est un amalgame de pathologies encore mal comprises, et qui inclut même quelques patients parfaitement conscients mais totalement incapables de le manifester à leur entourage.

Chez certains patients atteints de lésions massives, le niveau de conscience présente parfois d'importantes fluctuations dans le temps.

Pendant quelques heures, il arrive que le patient regagne un certain contrôle sur ses actions. On ne parle plus alors d'état végétatif, mais d'état de « conscience minimale ». Cette terminologie a été introduite en 2005, lors d'une conférence de consensus en neurologie, afin d'attirer l'attention sur cette catégorie particulière de patients qui, de façon rare et intermittente, montrent des signes indubitables de compréhension et de volonté de répondre[12]. Les patients en état de conscience minimale parviennent parfois à cligner des yeux en réponse à une question, ou à suivre du regard leur image dans un miroir. Une certaine forme de communication peut alors s'établir : de nombreux patients parviennent à répondre par oui ou par non, en hochant la tête ou même par oral. À la différence d'un patient végétatif, qui rit, pleure ou crie sans raison, une personne minimalement consciente exprime souvent des émotions appropriées au contexte.

Un seul indice ne suffit pas : le diagnostic d'état de conscience minimale exige que ces signes soient reproduits plusieurs fois. Paradoxalement, cependant, cet état se caractérise par des fluctuations incessantes qui ne permettent pas au patient d'exercer sa volonté de façon régulière. Certains jours, on n'observe aucun indice de conscience. D'autres fois, c'est le matin seulement que l'on détecte d'indubitables signes qui témoignent que la personne est « présente ». Il est parfois bien difficile d'établir avec certitude si elle a cligné, ri ou pleuré au moment adéquat. Afin d'améliorer la fiabilité du diagnostic, le neuropsychologue Joseph Giacino a conçu l'Échelle de récupération du coma, une série de tests comportementaux qui sont appliqués avec rigueur au chevet du patient[13]. Ces épreuves évaluent des fonctions cognitives élémentaires telles que la reconnaissance et la manipulation des objets, l'orientation du regard, spontanément ou à la demande, ou encore la réaction à un bruit inattendu. L'équipe médicale doit s'entraîner à administrer la totalité de cette batterie en restant à l'affût de la moindre réponse, même si elle est lente et approximative. Il est recommandé de répéter l'évaluation à différents moments de la journée.

À l'aide de cette échelle, un clinicien averti parvient à distinguer avec fiabilité l'état végétatif de l'état de conscience minimale[14]. Or c'est une information essentielle, non seulement pour les décisions de fin de vie, mais également pour anticiper sur la récupération éventuelle d'une certaine autonomie. Statistiquement parlant, un patient en état de conscience minimale a de bien meilleures chances de recouvrer la conscience qu'un patient resté pendant des années en état végétatif (bien qu'il reste hélas encore difficile de prédire l'avenir d'un patient donné). La récupération est souvent d'une lenteur frustrante. Cependant, de semaine en semaine, les comportements volontaires deviennent de plus en plus fréquents et reproductibles. Dans quelques cas spectaculaires, on assiste à une récupération soudaine, en quelques jours seulement. Dès que les patients redeviennent capables de communiquer de façon stable, ils cessent d'être qualifiés de « minimalement conscients ».

À quoi ressemble la vie mentale des patients en état de conscience minimale ? Est-elle pleine de souvenirs du passé, de projections dans le futur, et surtout d'une vive conscience du moment présent, riche de son lot d'envies, mais aussi de souffrance et de désespoir ? Ou s'apparente-t-elle plutôt à une sorte de brouillard permanent d'où émerge parfois, à la suite d'un immense effort de la volonté, une vague pensée volontaire ? Nous n'en savons rien, mais les fluctuations considérables du comportement de ces patients suggèrent que, dans bien des cas au moins, la seconde option ne doit pas être loin de la réalité. Gageons qu'elle ressemble probablement à l'état de torpeur et de confusion que nous éprouvons lorsque nous émergeons d'une anesthésie générale, d'un knock-out ou d'une cuite sévère.

De ce point de vue, l'état de conscience minimale diffère sans doute beaucoup du dernier type de pathologie qu'il nous reste à évoquer : le syndrome d'enfermement (*locked-in syndrome*) qu'a vécu Jean-Dominique Bauby. Cet état résulte d'une lésion bien délimitée, le plus souvent au niveau de la protubérance du tronc cérébral. Avec une précision quasi chirurgicale, cette lésion vient déconnecter le cortex de ses voies de sortie

en direction de la moelle épinière. Résultat, une paralysie presque complète, mais qui, parce que le cortex et le thalamus sont préservés, laisse souvent la conscience parfaitement intacte. Le patient sort donc du coma pour se retrouver enfermé dans un corps inerte, totalement incapable de parler ou de se mouvoir. Même le regard demeure généralement figé. Seuls de petits mouvements verticaux de l'œil et des clignements de la paupière, commandés par des circuits neuronaux distincts, fréquemment épargnés, offrent une modeste voie de communication avec le monde extérieur.

Dans *Thérèse Raquin* (1867) d'Émile Zola, Mme Raquin est une vieille femme quadriplégique atteinte d'un syndrome d'enfermement. En fin observateur, Zola note comment les yeux seuls ouvrent une fenêtre sur l'âme de cette malheureuse :

> « On eût dit le masque dissous d'une morte, au milieu duquel on aurait mis deux yeux vivants ; ces yeux seuls bougeaient, roulant rapidement dans leur orbite ; les joues, la bouche étaient comme pétrifiées, elles gardaient une immobilité qui épouvantait. (…) Ses yeux prenaient chaque jour une douceur, une clarté plus pénétrantes. Elle en était arrivée à se servir de ses yeux comme d'une main, comme d'une bouche, pour demander et remercier. Elle suppléait ainsi, d'une façon étrange et charmante, aux organes qui lui faisaient défaut. Ses regards étaient beaux d'une beauté céleste, au milieu de sa face dont les chairs pendaient molles et grimaçantes. »

En dépit de leur paralysie, les patients *locked-in* conservent effectivement toute leur lucidité. Ils ont une conscience très vive non seulement de leur déficit, mais aussi de leurs capacités résiduelles et de l'affection que leur témoigne leur entourage. C'est pourquoi, une fois leur état détecté et leurs douleurs soulagées, leur vie peut s'avérer pleinement

satisfaisante. Même leur moral peut être au beau fixe : une enquête récente révèle que la majorité d'entre eux, une fois dissipée l'horreur des premiers mois, juge sa qualité de vie comparable à celle de la population bien portante[15].

Preuves vivantes que le cortex et le thalamus suffisent à engendrer une pensée autonome, les cerveaux *locked-in* connaissent encore, en esprit, toute la gamme des expériences de la vie. Dans le roman de Zola, Mme Raquin savoure une douce vengeance lorsqu'elle voit sa nièce et son amant, qu'elle déteste depuis qu'ils ont tué son fils, se suicider sous ses yeux pénétrants. Dans *Le Comte de Monte-Cristo* de Dumas, Noirtier, en dépit de sa paralysie, parvient à révéler à sa petite-fille Valentine qu'elle s'apprête à épouser le fils de l'homme qu'il a lui-même tué des années plus tôt.

Dans la vraie vie, la destinée des patients *locked-in*, pour être moins mouvementée, n'en est pas moins singulière. À l'aide d'ordinateurs spécialisés dans le suivi du regard, certains patients répondent à leur courrier électronique ou dirigent des associations. Certains, tel l'écrivain Philippe Vigand, publient quatre livres et font un enfant. Contrairement aux patients comateux, végétatifs ou minimalement conscients, leur syndrome n'est pas une pathologie de la conscience, mais seulement de la motricité et de la communication. Encore faut-il que leur conscience soit bien détectée…

Cortico ergo sum

Mort cérébrale, coma, état végétatif, conscience minimale et syndrome d'enfermement : en 2006, la nosologie des troubles neurologiques de la vigilance et de la communication semblait bien établie, quand un article de la prestigieuse revue *Science* vint bousculer ce consensus. Le neuroscientifique britannique Adrian Owen y décrivait un patient qui

présentait toutes les manifestations cliniques d'un état végétatif, mais dont l'activité cérébrale indiquait un haut degré de conscience[16]. Cet article suggérait donc l'existence d'un état plus épouvantable encore que le *locked-in* : une personne parfaitement consciente, mais dépourvue de *tout* moyen de le manifester à l'extérieur, pas même un battement de la paupière. Tout en détruisant nos certitudes cliniques, cette recherche était également porteuse d'un message d'espoir : l'imagerie cérébrale était devenue suffisamment sensible pour détecter la présence d'une conscience résiduelle et même, comme nous allons le voir, de restaurer une forme de communication avec l'entourage.

La patiente décrite dans l'article de *Science* était une jeune femme de 23 ans qui, lors d'un accident de la route, avait subi des lésions bilatérales des lobes frontaux. Cinq mois plus tard, en dépit d'une préservation du cycle veille-sommeil, elle demeurait immobile et sans réponse – la définition même d'un état végétatif. Même une équipe expérimentée de neurologues ne parvenait pas à détecter le moindre signe d'une conscience résiduelle : aucune forme de communication, aucun contrôle volontaire, aucun geste intentionnel.

À l'imagerie cérébrale, pourtant, une découverte attendait les chercheurs. Dans le cadre d'un protocole de recherche, ils firent passer à leur patiente une IRM fonctionnelle. Quelle ne fut pas leur surprise d'observer que, lorsqu'on lui faisait écouter des phrases, toutes les aires corticales du langage s'activaient. Non seulement les régions temporales supérieures et moyennes, qui abritent les circuits de l'audition et de la compréhension des mots, s'allumaient normalement, mais on observait même une activation de la région frontale inférieure gauche, l'aire de Broca, impliquée dans la syntaxe des phrases. Cette activation augmentait même lorsque la compréhension était rendue plus difficile parce que la phrase contenait de nombreux mots ambigus (par exemple « Une petite *pièce* servait de *remise* »).

Cette intense activité corticale suggérait que le cerveau de la patiente continuait à traiter les mots et les phrases. Cependant, comprenait-elle ce

qu'on lui disait ? L'activation des aires du langage ne suffisait pas à prouver que la jeune femme parvenait à prendre conscience du sens des phrases. En effet, de nombreuses études ont montré qu'une grande part de ce réseau continue de fonctionner pendant le sommeil ou l'anesthésie[17].

Pour en avoir le cœur net, Owen présenta à sa patiente des phrases qui lui demandaient de réaliser certaines activités mentales. On lui disait tantôt « imaginez que vous jouez au tennis », tantôt « imaginez que vous visitez les pièces de votre appartement », ou encore « reposez-vous ». Les instructions lui demandaient de démarrer et de cesser ces activités à des moments bien précis : les mots « tennis » ou « appartement » indiquaient le début d'une période de 30 secondes d'imagination mentale, tandis que les mots « reposez-vous » imposaient 30 secondes de repos.

Owen n'avait aucun moyen de savoir si sa patiente, muette et inerte, comprenait un traître mot de ces instructions complexes. Cependant, l'IRM fonctionnelle apporta une réponse claire : son cerveau suivait fidèlement les ordres. Dès qu'on lui demandait d'imaginer une partie de tennis, son aire motrice supplémentaire s'activait pendant 30 secondes, puis elle s'éteignait, exactement comme on le lui avait demandé. Et lorsqu'elle visitait son appartement par la pensée, un autre réseau d'aires cérébrales s'allumait, qui comprenait diverses régions impliquées dans la représentation de l'espace (le gyrus parahippocampique, la région pariétale postérieure et le cortex prémoteur). L'imagerie cérébrale montrait qu'elle activait exactement les mêmes régions que des volontaires en bonne santé à qui on donnait les mêmes consignes.

Était-elle nécessairement consciente ? Quelques scientifiques se firent les avocats du diable[18]. Selon eux, il n'était pas impossible que ces régions s'activent de façon inconsciente, sans que la patiente comprenne les instructions. L'écoute du mot *tennis* suffisait peut-être à activer automatiquement les aires motrices, simplement parce que le sens de ce mot implique une action manuelle. De même, il se pouvait que le mot *appartement* évoque, inconsciemment, les circuits neuronaux de codage

de l'espace. Le cortex s'activait donc peut-être sans que cela n'indique nécessairement un état conscient. D'autres critiques, se plaçant sur un plan plus philosophique, rejetaient en bloc l'idée que l'imagerie cérébrale puisse jamais prouver ou réfuter la présence d'une conscience résiduelle. Pour le neurologue Allan Roper, « ni les médecins ni la société ne sont prêts à accepter le syllogisme "j'ai une activité cérébrale, donc je suis[19]" ».

Nous avons vu au chapitre 4 que cette dernière critique ne résiste pas à l'analyse. L'imagerie cérébrale a considérablement progressé, au point que l'identification d'un état de conscience subjective à partir d'images objectives de l'activité cérébrale a cessé de poser un problème insurmontable. Les autres critiques se sont évanouies lorsque Owen a réalisé une élégante expérience de contrôle. Il a scanné des volontaires en bonne santé, qui se contentaient d'entendre les mots *tennis*, *appartement* et *relaxation* sans avoir reçu auparavant la moindre instruction sur ce qu'ils devaient faire[20]. Sans surprise, ces mots ne causaient qu'une activation modeste, indistincte, et bien différente des signaux intenses et durables que présentait la patiente. Cette observation réfutait l'interprétation des avocats du diable. Lorsqu'elle activait ses aires prémotrices, pariétales et parahippocampiques, au bon moment et pendant 30 secondes, la patiente ne se contentait pas de répondre inconsciemment à l'écoute d'un mot. De toute évidence, elle *réfléchissait* à la tâche qu'on lui avait demandé d'exécuter.

Owen et ses collègues soulignèrent qu'il n'était guère probable que l'écoute d'un seul mot déclenche une activité inconsciente qui dure plusieurs dizaines de secondes – à moins que la patiente n'interprète effectivement ce mot comme une consigne, l'indication qu'il fallait commencer à exécuter les instructions demandées. Effectivement, dans le cadre du modèle de l'espace de travail neuronal global, un simple mot, s'il demeure inconscient, ne peut que déclencher une activation transitoire qui retombe rapidement à zéro. Inversement, l'observation d'une activation soutenue d'aires préfrontales et pariétales, pendant 30 secondes, indique presque à coup sûr la présence d'une représentation

consciente en mémoire de travail. On peut certes critiquer la tâche choisie par Owen et ses collègues pour son côté arbitraire et sans lien direct avec les travaux sur la conscience, mais force est de reconnaître que leur choix était intelligent et pragmatique : jouer au tennis mentalement ou se promener en pensée sont des activités relativement simples et motivantes, mais qui font très certainement appel à la conscience.

Libérer le papillon de la conscience

Si les premiers travaux du groupe d'Owen laissaient peut-être planer quelques doutes sur l'existence d'une conscience résiduelle chez certains patients en état végétatif, leur second article, publié dans une autre revue de premier plan (le *New England Journal of Medicine*[21]) les dissipait totalement, puisqu'elle démontrait l'existence de réponses volontaires. L'expérience était étonnamment simple. Les chercheurs commencèrent par répliquer leurs résultats sur l'imagination mentale. Sur 54 patients atteints de troubles de la conscience, cinq présentèrent des activations cérébrales typiques d'une personne normale qui imagine jouer au tennis ou visiter son appartement. Quatre d'entre eux étaient pourtant en état végétatif apparent.

L'un d'entre eux fut alors convié à une seconde séance d'IRM. Avant chaque scan, on lui posait à présent une question, par exemple : « Avez-vous des frères et sœurs ? » Il ne pouvait ni bouger ni parler – mais très astucieusement, les chercheurs lui demandèrent seulement une réponse mentale. « Pour répondre "oui", lui dirent-ils, imaginez que vous jouez au tennis dans votre tête ; et si vous voulez répondre "non", imaginez que vous visitez votre appartement. Commencez seulement quand vous entendrez "allez-y", et arrêtez dès que vous entendrez "reposez-vous". »

Cette astuce fonctionna à merveille (figure 30). Pour cinq des six questions posées, l'un des deux réseaux préalablement identifiés s'activa

très nettement (pour la sixième question, aucune activation des deux réseaux ne fut observée, ce qui fut interprété comme une absence de réponse). Les chercheurs ne connaissaient pas les réponses aux questions – mais quand ils comparèrent les activations cérébrales du patient avec les réponses données par la famille, ils se réjouirent de voir que chacune de ses cinq réponses était bonne !

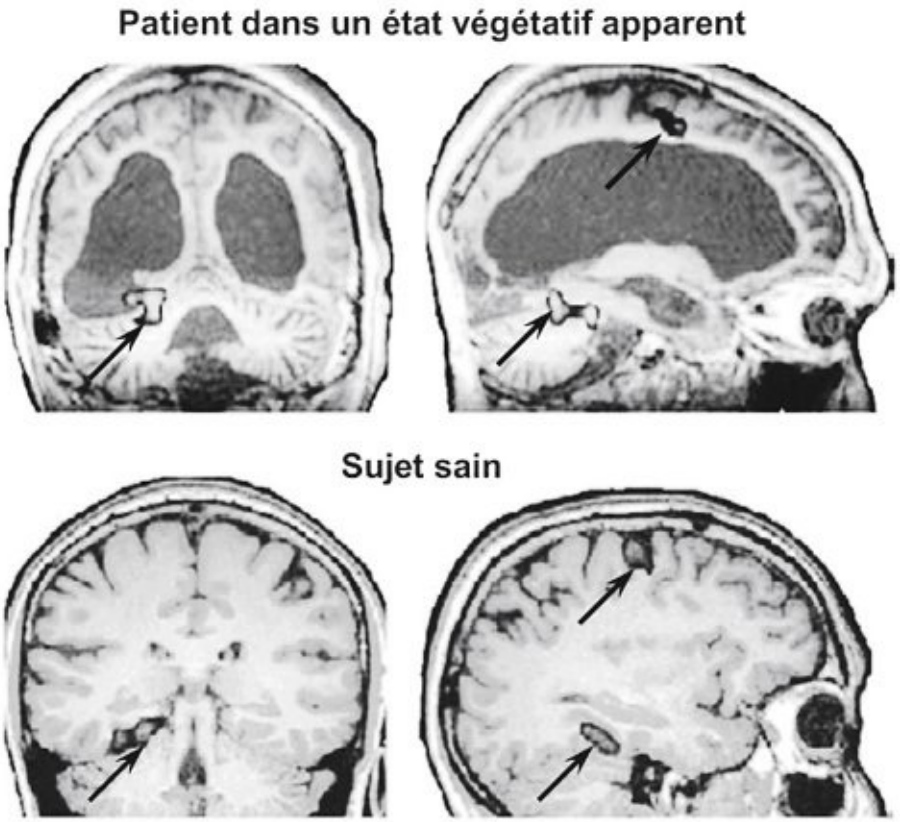


Figure 30. Certains patients en état végétatif apparent présentent une activité cérébrale presque normale qui suggère qu'ils sont encore conscients. Le patient présenté ici ne pouvait ni bouger ni parler, mais il parvenait toujours à répondre aux questions qu'on lui posait par le biais de son activité cérébrale. Pour répondre « non », on lui demandait d'imaginer qu'il se promenait dans son appartement, et pour répondre « oui », qu'il jouait au tennis. Lorsque les chercheurs lui demandèrent si son père s'appelait Thomas, ils virent s'activer les réseaux cérébraux de la navigation dans l'espace, exactement comme chez un sujet sain, donnant ainsi la bonne réponse : « Non. » Comme le patient ne présentait aucun signe de communication ou d'intention motrice, il était considéré en état végétatif. Ses lésions massives sont clairement visibles.

Arrêtons-nous un instant sur les implications de cette étonnante observation. Dans le cerveau du patient, toute une série de processus mentaux avaient dû s'enchaîner. Le patient avait forcément compris la question elle-même, récupéré la bonne réponse dans sa mémoire, et l'avait gardée en tête pendant quelques minutes avant que le scan ne commence. Ainsi, sa compréhension du langage de même que ses mémoires à long terme et à court terme étaient préservées. De plus, il avait suivi fidèlement les instructions, qui établissaient un lien arbitraire entre la réponse « oui » et l'imagerie mentale du tennis, et la réponse « non » et la visite de son appartement. Ainsi, le patient parvenait toujours, avec une grande flexibilité, à rediriger les informations en sa possession vers n'importe lequel de ses processus mentaux – un résultat qui, à lui seul, laisse penser que son espace de travail neuronal global était intact. Enfin, le patient appliquait les instructions au bon moment et n'éprouvait pas de difficulté à changer de réponse d'un scan à l'autre. Les circuits cérébraux de l'« administrateur central », responsables de l'attention et du changement de tâche, devaient donc être préservés. Certes, les données restaient maigres, et n'importe quel statisticien aurait préféré que le patient ait répondu à vingt questions plutôt qu'à cinq. Malgré tout, une conclusion semblait inévitable : ce patient disposait toujours du plein exercice de sa conscience et de sa volonté.

Cette expérience faisait ainsi voler en éclats les catégories standard de la neurologie clinique. Elle nous forçait à regarder la réalité en face : certains patients ne sont végétatifs qu'en apparence. Tempête dans un crâne : à l'intérieur du scaphandre plombé de ces corps inertes, le papillon de la conscience peut très bien continuer de battre librement des ailes, même si l'examen clinique le plus approfondi ne parvient pas à en détecter la moindre trace.

Dès la publication des travaux d'Owen et de Monti, la nouvelle se répandit dans les médias. Elle fut malheureusement souvent mal interprétée. Certains journalistes allèrent jusqu'à affirmer que les personnes dans le coma sont toujours conscientes. Pas du tout ! Rappelez-

vous que l'étude ne portait absolument pas sur le coma, mais seulement sur des personnes en état végétatif ou en état de conscience minimale. Qui plus est, seule une petite fraction, de l'ordre de 10 à 20 %, parvenait à répondre au test, ce qui signifiait que ce syndrome d'enfermement total avec une conscience intacte était relativement rare.

Nous ignorons quelle proportion exacte de patients est concernée, parce que le test d'imagerie cérébrale est asymétrique. S'il donne une réponse positive, on est pratiquement certain que le patient est conscient ; mais à l'inverse, on peut très bien demeurer conscient tout en échouant à ce test, et ce pour toutes sortes de raisons : surdité, troubles du langage, baisse de vigilance, incapacité de maintenir son attention, etc. De ce point de vue, il est notable que les patients qui parvenaient encore à répondre avaient tous subi un traumatisme crânien. Les autres, dont la perte de conscience était due à un accident vasculaire ou à un manque d'oxygène, ne présentaient aucune capacité résiduelle, peut-être par ce que leur cerveau, tel celui de Terri Schiavo, avait souffert de dommages diffus et irréversibles des neurones corticaux. En bref, le « miracle » de la découverte d'une conscience intacte chez un patient en état végétatif ne concernait qu'un tout petit nombre de cas. Il serait clairement irrationnel de l'utiliser pour défendre sans vergogne le « droit à la vie » et la poursuite des soins médicaux à tous les patients dans le coma, sans limite de temps.

Plus curieux encore, trente des trente et un patients en état de conscience minimale échouèrent à ce test. Par définition, tous ces patients présentaient parfois des signes manifestes de conscience et de volonté de répondre – et pourtant, par quelque terrible ironie, tous sauf un ratèrent leur seule chance de le prouver une bonne fois pour toutes au cours de l'examen d'imagerie cérébrale. Qui sait pourquoi ? Peut-être avaient-ils été testés au mauvais moment, quand leur vigilance était faible. Peut-être étaient-ils incapables de se concentrer dans l'environnement inhabituel et bruyant de l'IRM. Peut-être leurs compétences cognitives étaient-elles trop faibles pour suivre les instructions. Quoi qu'il en soit, nous devons en tirer

deux conclusions. Premièrement, le diagnostic clinique de « conscience minimale » ne garantit pas que ces patients jouissent encore de toutes leurs facultés mentales. Deuxièmement, le test d'Owen sous-estime probablement la conscience – le fait qu'un patient échoue ne signifie pas qu'il n'est pas conscient.

Pour toutes ces raisons, il n'y aura probablement jamais de test miracle qui prouve, une fois pour toutes, la présence ou l'absence de conscience. Dans l'état actuel des connaissances, l'éthique médicale exige de faire passer toute une batterie d'épreuves, et de voir si certaines d'entre elles permettent encore d'établir une forme de communication avec la personne dont l'esprit bat peut-être encore. Dans un monde idéal, il faudrait que ces tests soient plus simples que d'imaginer de jouer au tennis pendant 30 secondes. Il faudrait aussi qu'ils soient répétés plusieurs fois, afin de ne pas passer à côté d'un patient *locked-in* dont la vigilance fluctuerait d'heure en heure. Malheureusement, l'IRM fonctionnelle ne répond pas à ces critères, car elle exige un équipement complexe et si coûteux que les patients ne passent guère qu'une ou deux IRM au grand maximum. Comme le dit Adrian Owen lui-même, « il est affreux d'ouvrir une fenêtre de communication avec un patient, et de ne pas pouvoir ensuite enchaîner en lui proposant immédiatement une méthode de routine pour communiquer avec sa famille[22] ». Même le second patient d'Owen, qui montrait des signes évidents de communication volontaire, ne fut testé qu'une seule fois en IRM, avant d'être impitoyablement renvoyé à la prison de son syndrome d'enfermement total.

Il est urgent de dépasser cet état de fait. Dans ce but, de nombreuses équipes de recherche commencent à développer des interfaces cerveau-machine fondées sur une méthode bien plus simple : l'électroencéphalographie (EEG), une technique de routine, peu coûteuse, disponible dans tous les services, et qui ne nécessite qu'un modeste amplificateur pour détecter les signaux électriques à la surface de la tête[23].

Malheureusement, jouer au tennis ou se promener mentalement sont des états trop complexes pour être suivis par EEG. Dans une étude récente, les chercheurs ont testé une instruction bien plus simple : « Dès que vous entendez un bip, imaginez que vous serrez le poing droit, puis que vous détendez votre main ; concentrez-vous sur la sensation musculaire que vous ressentiriez si vous le faisiez pour de vrai[24]. » À d'autres moments, les patients devaient imaginer qu'ils remuaient les orteils. Pendant ce temps, une machine d'EEG enregistrait les oscillations en provenance des aires motrices du cortex. Pour chaque patient, un algorithme informatisé tentait d'apprendre à trier les essais : main ou pied ? Et chez trois patients sur seize en état végétatif, cela semblait fonctionner – mais hélas, la technique demeurait trop rudimentaire pour affirmer avec certitude que ce résultat n'était pas dû au hasard[25] (le test n'a d'ailleurs fonctionné que sur neuf des douze volontaires sains, sans lésion cérébrale et clairement conscients). Une autre équipe, dirigée par Nicholas Schiff à New York, a réalisé un test qui demandait à cinq volontaires et à trois patients d'imaginer qu'ils nageaient ou qu'ils visitaient leur appartement[26]. Là encore, ce test semble donner de bons résultats, mais le nombre de patients est encore insuffisant pour en tirer des conclusions définitives.

Nous n'en sommes donc qu'aux balbutiements. Cependant, il ne fait guère de doute que l'EEG représente la meilleure piste actuelle pour rétablir la communication avec les patients conscients mais paralysés[27]. De brillants ingénieurs, fascinés par l'extraordinaire défi technologique que représente la mise en connexion d'un ordinateur avec un cerveau humain, ne cessent de progresser. La plupart des dispositifs de communication homme-machine reposent encore sur l'orientation du regard et de l'attention visuelle, ce qui n'est pas idéal pour les patients. Cependant, le décodage de l'attention auditive et de la motricité progresse également à grands pas. L'industrie du jeu vidéo s'y met également en proposant des casques d'EEG légers et sans fil. Il se peut même qu'on implante des électrodes directement dans le cortex de certains patients

paralysés. À l'aide d'un dispositif de ce type, une patiente quadriplégique est parvenue récemment à contrôler un bras robotisé par la seule pensée[28]. Certaines équipes estiment que, si les électrodes étaient implantées dans les aires du langage, on pourrait transformer l'intention de parler en une synthèse vocale[29].

De vastes perspectives de recherche s'ouvrent donc. À l'avenir, non seulement elles aideront les patients *locked-in* à communiquer avec leur entourage, mais elles fourniront des moyens nouveaux de détection de la conscience. Dans les services cliniques les plus avancés, tel celui dirigé par Steven Laureys à Liège, en Belgique, les interfaces cerveau-machine font d'ores et déjà partie de la batterie de tests qui sont systématiquement déployés dès qu'un nouveau patient est admis à l'hôpital. D'ici une vingtaine d'années, nous trouverons tous parfaitement banal qu'un patient quadriplégique et *locked-in* conduise son fauteuil roulant par la seule force de la pensée.

La détection consciente de la nouveauté

J'ai beaucoup d'admiration pour les recherches pionnières d'Adrian Owen. Pourtant, sur le plan théorique, elles me laissent sur ma faim. Je n'ai guère de doute qu'il faille être conscient pour réussir ses tests, mais ils ont été conçus de façon purement intuitive, sans être guidés par aucune théorie spécifique de la conscience. Puisqu'ils font appel au langage, à la mémoire et à l'imagination, il existe cent raisons de ne pas les réussir, tout en étant parfaitement conscient. Serait-il possible de concevoir une mesure plus simple de la conscience ? Comme nous l'avons vu, les avancées de l'imagerie cérébrale ont permis d'identifier des signatures de la conscience. Ne pourraient-elles pas servir à détecter si une personne est ou n'est pas consciente ?

En 2008, lors d'un déjeuner mémorable à Orsay, au sud de Paris, mes collègues Tristan Bekinschtein, Lionel Naccache, Mariano Sigman et moi-même avons tenté d'imaginer un test vraiment élémentaire. Nous avons tout de suite décidé que ce test s'appuierait sur l'EEG – la plus simple et la moins coûteuse des méthodes d'imagerie. Nous avons ensuite jugé qu'il faudrait utiliser la modalité auditive, parce que la plupart des patients n'ont pas de difficultés majeures d'audition, alors que la vision leur pose souvent de grandes difficultés. Cette décision n'avait rien d'évident, parce que les signatures de la conscience dont nous disposions chez le volontaire sain provenaient, pour la plupart, de tests visuels. Cependant, nous étions confiants que les signatures de l'accès à la conscience que nous avions découvertes s'étendent également à l'audition.

Nous avons décidé de nous appuyer sur la signature la plus claire que nous avions systématiquement détectée dans toutes nos expériences : l'onde P3, massive, qui indique l'embrasement synchronisé d'un réseau global d'aires corticales. Il n'y a rien de plus facile que d'évoquer une onde P3 auditive. Imaginez que vous écoutiez un *adagio* quand, soudain, le téléphone sonne. L'irruption de ce son incongru déclenche une grande onde P3, qui traduit la réorientation de l'attention et la prise de conscience d'un événement inattendu[30].

Notre expérience consisterait donc à présenter une série de sons réguliers, *bip bip bip bip*… qui serait tout à coup interrompue par un intrus, disons le son *boup*. Quand une personne est éveillée et attentive, cet intrus engendre effectivement un événement de type P3, qui peut servir de marqueur de conscience. Au milieu de l'expérience, on inverse les deux sons : *boup* devient le son régulier, et *bip* l'intrus. Ainsi parvient-on à prouver que l'onde P3 provient uniquement de la nouveauté du son dans le contexte actuel.

Il nous restait à contourner une difficulté. Les sons inattendus ne déclenchent pas seulement une onde P3, mais aussi toute une série de réponses corticales plus précoces qui traduisent une détection inconsciente de la nouveauté. À peine un dixième de seconde après le

début du son, le cortex auditif accroît déjà sa réponse aux sons intrus par rapport aux sons habituels. Il émet une « réponse de discordance » (*mismatch response*), également appelée « onde négative de discordance » (*mismatch negativity* ou MMN), car elle se manifeste par un voltage négatif au sommet de la tête[31]. Cette réponse à la nouveauté n'a rien à voir avec la conscience : elle survient même quand l'auditeur pense à autre chose, lit un livre, regarde un film, s'endort ou tombe dans le coma. Notre système nerveux contient un système inconscient de détection de la nouveauté, qui compare simplement le son présent avec la prédiction issue des sons précédents. C'est une fonction omniprésente dans le cortex : chaque bout de cortex abrite probablement un réseau de neurones qui prédit et qui compare cette prédiction avec la réalité[32]. Ces opérations sont automatiques : seul leur résultat attire l'attention et peut, le cas échéant, devenir conscient.

Le test de l'intrus n'est donc pas suffisant : même le cerveau d'un patient comateux peut sursauter à l'écoute d'un son nouveau. L'onde MMN n'est pas une signature de la conscience. Sa présence montre simplement que le cortex auditif demeure suffisamment intact pour détecter la nouveauté, pas que le patient est conscient[33]. Elle appartient à ce vaste catalogue d'opérations sensorielles précoces, non dépourvues d'une certaine sophistication, mais qui opèrent de façon strictement inconsciente. Nous avions besoin de l'écarter pour ne détecter que la suite de l'activité corticale : l'avalanche tardive de décharges neuronales qui indexe l'entrée d'une information dans la conscience.

Pour ce faire, nous avons inventé une nouvelle astuce : l'opposition systématique de deux formes de nouveauté auditive. Imaginez que vous entendiez une séquence de cinq sons, dont le dernier est différent : *bip bip bip bip boup*. Dès que votre cerveau détecte l'anomalie finale, il engendre deux ondes successives, la MMN et la P3. Maintenant, répétez cette séquence plusieurs fois. Votre cerveau s'habitue rapidement à l'idée d'entendre quatre *bips* suivis d'un *boup* – au niveau conscient, la surprise disparaît. Ce qui est curieux, c'est que l'onde MMN reste présente. Elle

reflète un mécanisme primitif de détection de la nouveauté qui est incapable de mémoriser la séquence globale et s'en tient à une prédiction de courte vue : la plupart des *bips* sont suivis d'un autre *bip* – ce qui s'avère faux pour le dernier *boup*.

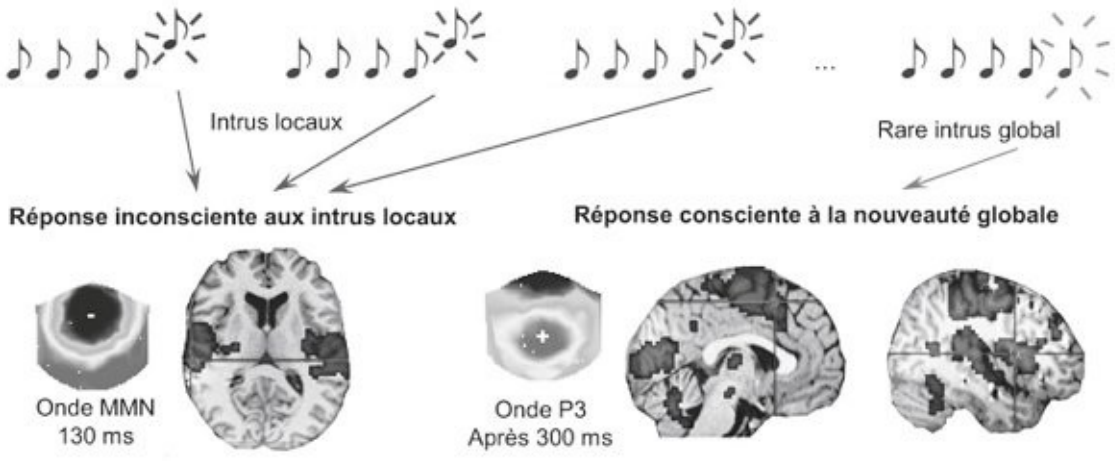


Figure 31. Le test local-global permet de détecter une conscience résiduelle chez certains patients. Il consiste à répéter plusieurs fois une séquence de cinq sons. Lorsque le dernier son diffère des précédents, les aires auditives émettent une réponse automatique à la nouveauté (*mismatch response*). Cette première réponse est inconsciente et persiste dans le sommeil ou le coma. Au niveau conscient, cependant, le cerveau mémorise la séquence globale. Après une période d'adaptation, c'est désormais l'*absence* de la nouveauté finale – une séquence de cinq sons identiques – qui déclenche une seconde réaction corticale. Cette réponse d'ordre supérieur présente toutes les signatures de la conscience (onde P3, activité synchrone des aires pariétales et préfrontales). Elle semble n'être présente que chez les patients conscients.

L'onde P3, elle, reflète un calcul de plus haut niveau qui suit étroitement le contenu de la conscience. Dès que la personne retient la séquence de cinq sons et cesse d'être surprise par l'arrivée du dernier *boup*, la P3 disparaît. Le cerveau conscient s'attend à ce que le dernier son soit différent. Une fois que cette attente est mise en place, on peut s'amuser à la transgresser en présentant, sans prévenir, une séquence de cinq sons identiques : *bip bip bip bip bip*. Et soudain, l'onde P3 est de

retour ! Remarquez à quel point ce phénomène est étrange. Le cerveau attribue l'étiquette « nouveau » à une séquence de cinq sons identiques, parfaitement régulière et banale. Il ne le fait que parce que cette séquence est rare et qu'il s'attendait à autre chose : c'est la différence entre la prédiction (je m'attends à un *boup* final) et la réalité perçue (cinq *bips*) qui crée la nouveauté. Mais, pour cela, il faut que le cerveau soit conscient, attentif, et fasse des prédictions à l'échelle de toute la séquence. La prédiction locale (la plupart des *bips* sont suivis d'un autre *bip*) est inconsciente, la prédiction globale (le dernier doit être un *boup*) est consciente.

Nous avons atteint notre but : nous savons comment évoquer une onde P3, signature d'un traitement conscient, sans qu'elle soit précédée de réponses non conscientes plus précoces. Nous pouvons même l'amplifier en donnant au patient l'instruction de compter les séquences déviantes. Le fait de compter accroît considérablement l'amplitude de l'onde P3, au point de la rendre facilement détectable (figure 31). Dès qu'on la voit, on est pratiquement certain que le patient suit fidèlement les instructions.

D'un point de vue empirique, ce test « local-global » fonctionne bien. Nous n'avons aucune difficulté à détecter la réponse P3 globale chez chaque individu : même un court enregistrement d'EEG suffit. L'onde P3 semble n'être présente que lorsque la personne est attentive et consciente de la règle[34]. Dès qu'on la distrait en lui demandant d'accomplir en même temps une tâche visuelle, l'onde P3 disparaît. Si on ne lui demande pas de compter, mais seulement de ne rien faire de particulier, seules les personnes qui détectent consciemment la séquence globale montrent une P3. Chez ceux qui, à la fin de l'expérience, disent n'avoir rien remarqué de particulier, cette signature de la conscience est absente.

L'examen des aires cérébrales activées par les séquences intruses confirme qu'elles engendrent un embrasement cortical global. Tout le réseau de l'espace de travail neuronal global s'active vigoureusement. L'activité cérébrale ne reste pas confinée aux aires auditives, mais elle envahit un vaste réseau cortical qui comprend le cortex préfrontal

bilatéral, le cortex cingulaire antérieur, les aires pariétales et même certaines régions occipitales. Ainsi, l'information qu'une anomalie vient d'être détectée est effectivement diffusée à une grande partie du cortex – la marque que cette information est devenue consciente.

Reste la question clé : ce test fonctionne-t-il en clinique ? Le cerveau d'un patient paralysé mais conscient continue-t-il de répondre à la nouveauté auditive globale ? Nos premiers essais sur un petit nombre de patients ont été couronnés de succès[35]. Chez quatre patients en état végétatif, la réponse à la nouveauté globale brillait par son absence, tandis que chez trois patients minimalement conscients sur quatre, elle était bien présente (et ces trois patients recouvrèrent ensuite une conscience stable).

Mon collègue le neurologue Lionel Naccache a ensuite commencé à appliquer ce test en routine à l'hôpital de la Salpêtrière à Paris, avec des résultats très encourageants[36]. Dès qu'on voyait une P3 globale, il semblait bien que le patient était conscient. Sur vingt-trois patients en état végétatif apparent, seuls deux ont montré une onde P3, et ces deux-là sont passés en état de conscience minimale dans les jours qui ont suivi, ce qui suggère qu'ils étaient peut-être déjà conscients, tout comme les patients d'Adrian Owen.

Dans l'unité de soins intensifs, le test local-global fournit parfois des informations vitales, dans tous les sens du terme. Ainsi, à la suite d'un terrible accident de voiture, un jeune homme se trouvait dans le coma depuis trois semaines, restait sans réponse et souffrait de tant de complications que l'équipe médicale hésitait à poursuivre les soins. Cependant, son cerveau répondait toujours fortement aux intrus globaux. Peut-être était-il dans une sorte d'état *locked-in*, enfermé en lui-même et incapable de manifester, même d'un seul geste, la préservation de sa conscience ? Lionel persuada les médecins qu'il y avait encore une possibilité d'évolution positive… et de fait, quelques jours plus tard, le patient recouvra toute sa lucidité. Son état s'améliora à tel point qu'il vit aujourd'hui une vie presque normale.

La théorie de l'espace de travail global explique pourquoi notre test fonctionne. Pour détecter la séquence qui se répète, les patients doivent enregistrer une séquence de cinq sons en mémoire de travail. Ils doivent ensuite la comparer à la séquence suivante, qui survient plus d'une seconde plus tard. Comme je l'ai expliqué au chapitre 3, garder une information en tête pendant quelques secondes est l'apanage de la conscience. Notre test fait appel à cette fonction d'au moins deux manières : le cerveau doit rassembler les cinq notes successives d'une séquence en un seul motif, et il doit également comparer plusieurs motifs successifs.

Le test exige également une réflexion de haut niveau. En effet, les opérations qui permettent de décider qu'une séquence monotone de cinq *bips* est, en fait, rare et digne d'intérêt ne sont pas triviales. Lorsqu'il entend la séquence fréquente, *bip bip bip bip boup*, le cerveau s'habitue à l'écoute de l'intrus final. Ce son engendre toujours une réaction automatique de nouveauté dans les aires auditives (la MMN), mais une prédiction de second ordre annule ce signal d'alarme sans intérêt, car parfaitement prévisible[37]. Dans les rares cas où le cerveau reçoit cinq *bips* identiques, ce système de second niveau manifeste sa surprise. Pour lui, il est surprenant que la séquence se termine sans surprise. Le test fonctionne parce qu'il court-circuite le premier niveau de détection de la nouveauté, qui est inconscient, et évalue sélectivement l'intégrité de la seconde étape, qui est étroitement liée à l'ignition globale, au cortex préfrontal, donc à la conscience.

Sonder le cortex

Notre équipe de recherche a accumulé suffisamment de succès avec le test local-global pour penser qu'il mesure effectivement un traitement conscient. Cependant, cette épreuve conserve une imperfection majeure.

Nous avons encore trop de faux négatifs – des patients qui sont sortis du coma et dont personne ne doute qu'ils sont conscients, mais chez qui le test échoue. Notre test n'est donc pas assez sensible. Nous l'avons amélioré à l'aide d'un algorithme d'apprentissage assisté par ordinateur[38]. Cet outil, un peu comme le moteur de recherche Google, explore tout le cerveau à la recherche d'une réponse quelconque à la nouveauté globale, même si elle ne prend pas la forme d'une onde P3. En effet, les lésions massives qu'ont subies les patients peuvent modifier la topographie de cette signature de la conscience. Cette astuce améliore effectivement le taux de réussite, mais sans que nous ne parvenions à déceler une réaction chez la moitié des patients en état de conscience minimale.

Les statisticiens décrivent cette situation en deux mots : bonne spécificité, mais faible sensibilité. Notre test, comme celui d'Owen, est asymétrique. Lorsqu'il donne une réponse positive, nous sommes pratiquement certains que le patient est conscient (c'est la « spécificité »), mais lorsqu'il donne une réponse négative, nous ne pouvons pas en conclure que le patient n'est pas conscient (le test manque de « sensibilité »). De multiples raisons peuvent l'expliquer. Les enregistrements EEG sont souvent bruités : tout le monde sait qu'il est difficile d'enregistrer des signaux de qualité dans un lit d'hôpital, entouré d'équipements électroniques, et chez un patient incapable de conserver son regard et son corps immobiles. Il est également probable que certains patients soient conscients mais dans l'incapacité de comprendre le test. Leurs lésions sont si vastes qu'ils ne parviennent plus à compter les sons rares, voire à les détecter ou même à focaliser leur attention sur l'audition pendant quelques secondes.

Bien entendu, ces patients, conscients, continuent d'avoir une vie mentale. Si notre théorie est juste, leur espace de travail neuronal continue de diffuser des informations à l'ensemble du cortex par le biais de connexions à longue distance. Mais comment pourrions-nous le détecter ? Marcello Massimini, de l'Université de Milan, a imaginé une solution originale[39]. Alors que tous les tests de mon laboratoire se fondent

sur la propagation d'un signal sensoriel au sein du cortex, il a proposé d'utiliser la stimulation magnétique. Son idée était d'envoyer une décharge électrique directement dans le cortex. Comme le *ping* d'un sonar sonde les abysses de l'océan, ce stimulus intense se propagerait dans les profondeurs du cortex et du thalamus, et la force et la durée de cet écho évalueraient l'intégrité des régions traversées. Si l'activité se diffusait jusqu'aux régions distantes du cortex, et si l'écho s'y réverbérait pendant un long moment, cela signifierait que le patient est conscient. Le plus remarquable, c'est que le patient n'aurait pas spécialement besoin de prêter attention à ce *ping*. Une impulsion pourrait mesurer l'état des voies de communication corticale à longue distance sans que le patient en soit nécessairement conscient.

Pour mettre en pratique son idée, Massimini combina deux technologies : la TMS et l'EEG. La stimulation magnétique transcrânienne, ou TMS en anglais, comme je l'ai expliqué au chapitre 4, utilise l'induction magnétique pour stimuler le cortex, en faisant décharger un courant dans une bobine placée tout près de la tête. L'EEG, comme le lecteur s'en souvient sûrement, est la bonne vieille méthode d'enregistrement des ondes cérébrales. Massimini décida d'envoyer d'abord une impulsion de TMS dans le cortex, puis d'enregistrer les ondes EEG évoquées par la propagation de cette impulsion. Cela nécessitait des amplificateurs spéciaux, capables d'absorber sans broncher les intenses courants induits par la TMS et, quelques millisecondes plus tard, de mesurer fidèlement les petits courants résiduels que cette stimulation avait provoqués au sein des circuits neuronaux.

Les résultats de Massimini se sont avérés très encourageants. Il a commencé par tester sa technique chez l'homme normal au cours de la veille, du sommeil et de l'anesthésie générale. Dès que le sujet perdait conscience, quelle qu'en soit la cause, l'impulsion TMS n'induisait plus qu'une brève activation focale qui ne durait guère que 200 millisecondes environ. Par contre, dès que le patient était conscient (même s'il rêvait), la même impulsion causait toute une séquence longue et complexe d'activité

cérébrale. L'endroit où la stimulation était appliquée importait peu : quel que soit le point de départ, la durée de propagation et la complexité des ondes évoquées fournissaient d'excellents indices du degré de conscience[40]. Cette observation concorde avec les résultats de mon équipe : dans notre cas, le stimulus est sensoriel, mais c'est toujours sa diffusion globale, au-delà d'une durée de 300 millisecondes, qui indique un état conscient.

Massimini a ensuite testé son stimulateur chez cinq patients en état végétatif, cinq autres minimalement conscients et deux patients *locked-in*[41]. Ce sont de tout petits effectifs pour une étude clinique, mais le résultat était net : chez *tous* les patients conscients, le stimulus magnétique déclenchait une réponse complexe et durable, tandis que ce n'était le cas chez *aucun* des patients végétatifs. Massimini a suivi cinq autres patients en état végétatif pendant plusieurs mois, jusqu'à ce que trois d'entre eux récupèrent certaines capacités de communication et répondent donc au diagnostic de « conscience minimale ». C'est précisément chez ces trois patients que les signaux cérébraux ont augmenté en complexité. De plus, en plein accord avec le modèle de l'espace de travail neuronal global, la progression des signaux en direction des aires préfrontales et pariétales a fourni un excellent indicateur de leur degré de conscience.

Détecter l'existence d'une pensée autonome

L'avenir dira si le test du *ping* magnétique de Massimini est aussi efficace qu'il y paraît. S'il fonctionne chez *chaque* patient, comme les résultats le suggèrent, il ne fait guère de doute qu'il deviendra un test clinique de référence. Cependant, il exige tout de même un grand déploiement de technologie. Tous les hôpitaux n'ont pas d'appareil d'EEG

multiélectrodes capable d'absorber les chocs électriques causés par une énorme impulsion magnétique.

En théorie, il existe une solution plus simple. Si l'hypothèse de l'espace de travail global est juste, alors, même en l'absence de toute stimulation, un cerveau conscient doit présenter des signes détectables de communication à longue distance. La conscience consiste en un flux incessant d'informations. Un flot ininterrompu d'activité cérébrale doit donc parcourir les aires pariétales et préfrontales, et ses fluctuations doivent se synchroniser avec celles d'autres régions cérébrales distantes, qui varient selon le contenu actuel de la conscience. Cette communication à longue distance doit s'accompagner d'une augmentation de la puissance des ondes émises, particulièrement dans les fréquences moyennes (bêta) et hautes (gamma). La diffusion globale des messages conscients consomme beaucoup d'énergie. Ne pourrions-nous pas la détecter directement ?

En fait, nous savons depuis des années que le métabolisme global du cerveau, tel qu'on peut le mesurer avec une caméra à positrons (PET), diminue au cours de la perte de conscience. Une caméra PET est un détecteur sophistiqué de rayons gamma, qui permet de mesurer la consommation de glucose (source d'énergie chimique) en n'importe quel point du corps. L'astuce consiste à injecter dans le sang une molécule précurseur du glucose, marquée avec un traceur radioactif, et d'utiliser la caméra pour détecter où se produisent les émissions de photons gamma. Ces émissions indiquent avec précision en quels endroits du corps le glucose est consommé, et en quelles quantités. Or, dès que le volontaire est anesthésié ou qu'il tombe en sommeil profond, on observe une réduction massive, de l'ordre de 50 %, du glucose consommé dans tout le cortex. Une diminution du même ordre de la consommation énergétique du cortex survient chez les patients en coma ou en état végétatif. Dès les années 1990, l'équipe de Steven Laureys, à Liège, a produit des images remarquables des anomalies du métabolisme cérébral dans l'état végétatif[42] (figure 32).

Cette diminution, qui affecte aussi bien la consommation d'oxygène que celle de glucose, varie d'une région à l'autre du cerveau. La perte de la conscience semble être spécifiquement associée à une réduction d'activité des régions préfrontales et pariétales bilatérales, ainsi que du cortex cingulaire et du précuneus. Ce sont exactement les régions de l'espace de travail global, celles qui sont les plus riches en connexions corticales à longue distance – un indice supplémentaire que ces aires jouent un rôle essentiel dans l'expérience consciente. À l'inverse, certaines régions restreintes du cortex sensoriel ou moteur peuvent très bien demeurer intactes et conserver une consommation énergétique normale chez des patients dépourvus de toute réponse volontaire[43]. Ainsi, certains patients en état végétatif qui ont encore des mouvements automatiques présentent parfois une activité normale dans les aires motrices. Depuis plus de vingt ans, l'un de ces patients émet parfois un mot à haute voix, inconsciemment semble-t-il, en tout cas sans aucune pertinence dans la conversation. Le métabolisme et l'activation de son cerveau ne montrent que quelques îlots de cortex restés intacts dans les aires du langage de l'hémisphère gauche. De toute évidence, cette activité isolée et sporadique ne suffit pas à évoquer un état de conscience.

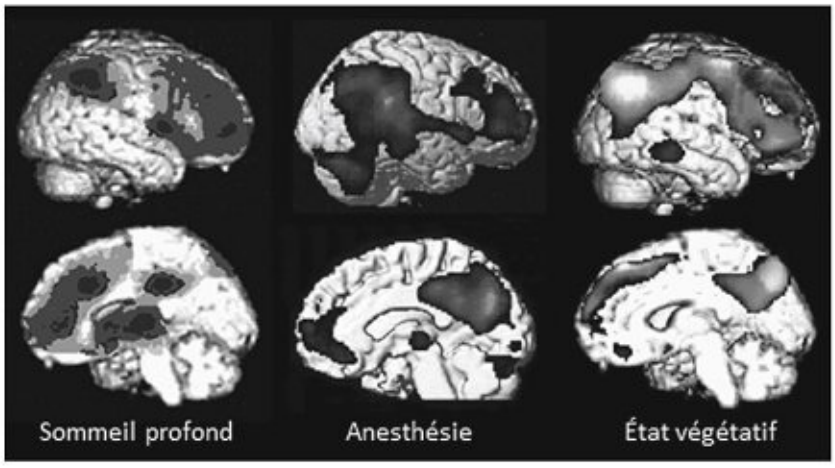


Figure 32. Une forte réduction du métabolisme des régions pariétales et frontales accompagne la perte de conscience au cours du sommeil, de l'anesthésie ou de l'état végétatif. D'autres régions réduisent également leur activité, mais la diminution de la consommation énergétique est particulièrement prononcée dans les aires corticales qui participent à l'espace de travail neuronal global.

Hélas, le métabolisme cérébral ne donne pas toujours des indications aussi précises sur la présence ou l'absence de conscience. Certains patients conservent une consommation énergétique quasi normale dans le cortex. Il est probable que leur lésion n'affecte que les connexions ascendantes en provenance du diencéphale, tandis que le cortex lui-même est épargné. À l'inverse, et plus crucialement, de nombreux patients passent de l'état végétatif à la « conscience minimale » sans pour autant que leur métabolisme ne revienne à la normale. Une comparaison de la consommation énergétique avant et après la récupération de la conscience montre certes une augmentation sélective dans les aires corticales de l'espace de travail global, mais le gain est faible et variable. Si le métabolisme ne remonte guère, c'est sans doute parce que le cortex reste gravement endommagé. Même une IRM anatomique de haute résolution ne joue guère qu'un rôle indicatif[44] : elle ne parvient pas à fournir un indicateur infaillible de conscience. Ni l'anatomie ni le métabolisme ne permettent d'évaluer avec certitude si la circulation globale d'informations, qui sous-tend l'état conscient, continue de se produire ou non.

Afin de construire un « conscientimètre », un détecteur de conscience résiduelle, mes collègues Jean-Rémi King, Jacobo Sitt, Lionel Naccache et moi nous sommes à nouveau tournés vers l'EEG, avec l'idée d'en faire un marqueur de la communication corticale[45]. En quelques années, l'équipe de Lionel Naccache a collecté près de 200 enregistrements de haute densité, à l'aide de 256 électrodes, chez des dizaines de patients végétatifs, minimalement conscients ou conscients. Pour tirer de ces données une mesure quantitative de l'échange d'informations à longue

distance, Jacobo Sitt, qui est à la fois psychiatre, informaticien et physicien de talent, a eu une idée brillante[46]. Il a conçu une nouvelle mesure mathématique, que nous avons appelée « information symbolique mutuelle pondérée », et qui évalue combien d'informations sont échangées entre deux régions du cerveau.

Lorsque nous l'avons appliquée aux données des patients, cette mesure a nettement séparé les patients en état végétatif de tous les autres (figure 33). Par rapport au cerveau conscient, le cerveau en état végétatif montre une réduction dramatique des échanges d'information. Cette diminution du trafic cortical est particulièrement évidente dès que les électrodes sont séparées d'au moins 6 centimètres : une fois de plus, la diffusion des informations à longue distance s'avère être l'apanage du cerveau conscient. En utilisant une autre mesure, cette fois sensible à la direction des échanges, nous avons également vu que le dialogue cérébral était bidirectionnel : dans le cerveau conscient, les aires spécialisées situées dans la partie postérieure du cerveau parlent aux aires généralistes des régions pariétales et préfrontales, qui en retour leur renvoient des signaux rétrogrades.

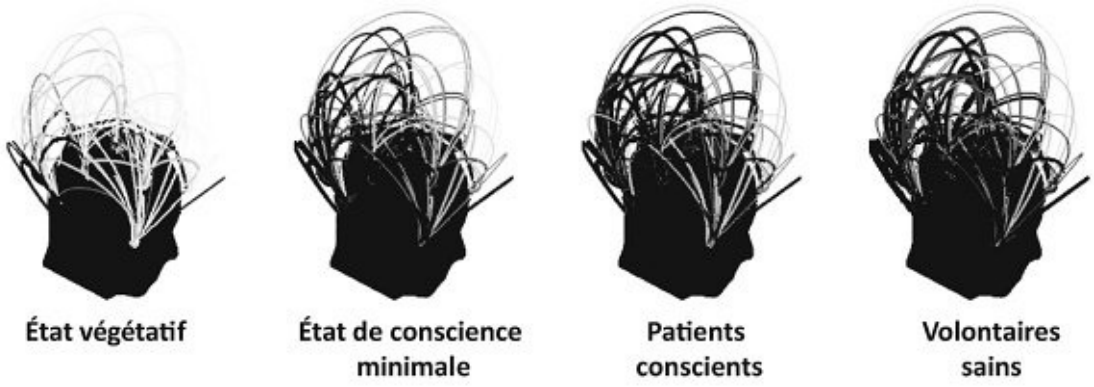


Figure 33. La quantité d'information échangée à travers le cortex fournit un excellent indice de conscience chez les patients atteints de lésions cérébrales. Pour créer cette image, notre équipe a enregistré un électroencéphalogramme (EEG) de haute résolution à 256 électrodes chez près de 200 patients. Pour chaque paire d'électrodes, symbolisée par un arc de cercle, nous avons

calculé un index mathématique de la quantité d'information partagée par les aires sous-jacentes. Chez les patients en état végétatif, le partage d'information est considérablement réduit par rapport à d'autres patients conscients ou à des sujets sains. Cette observation corrobore l'une des hypothèses clés du modèle de l'espace de travail, selon laquelle l'échange d'informations à longue distance est une des fonctions essentielles de la conscience.

La conscience se reflète encore dans l'EEG de bien d'autres manières[47]. En évaluant la puissance émise dans différentes bandes de fréquence, nous avons observé, sans surprise, que la perte de la conscience conduit à la disparition des fluctuations de haute fréquence, au bénéfice d'ondes lentes de basse fréquence typiques du sommeil ou de l'anesthésie[48]. L'évaluation de la synchronie montre également que, dans l'état conscient, les régions corticales tendent à synchroniser leurs signaux de haute fréquence, nettement plus que dans l'état non conscient.

Chacune de ces mesures mathématiques jette donc une certaine lumière sur l'état de conscience. Afin de les combiner, Jean-Rémi King a programmé un algorithme qui découvre, par apprentissage, quelles pondérations accorder à chacune d'entre elles afin de prédire au mieux les observations cliniques. Vingt minutes d'EEG, sans qu'il soit besoin de poser la moindre question au patient, suffisent à poser un bon diagnostic. Le programme fait peu d'erreurs. Il est très rare qu'il juge parfaitement conscient un patient en état végétatif. La plupart des erreurs consistent à étiqueter comme « minimalement conscient » un patient en état végétatif, ou inversement. Or ces erreurs n'en sont peut-être pas. Au cours des vingt minutes d'enregistrement, il est fort possible qu'un patient minimalement conscient se soit endormi (auquel cas la répétition de l'examen devrait améliorer le diagnostic). À l'inverse, quand le programme affirme qu'un patient jugé végétatif par les cliniciens est en réalité en état de conscience minimale, qui faut-il croire ? Se pourrait-il que ce patient soit conscient, mais dans cet état paradoxal de *locked-in* complet, incapable de réaliser la moindre action volontaire ? L'examen du devenir des patients, plusieurs mois après l'enregistrement EEG, donna des résultats fascinants. Pour

deux tiers d'entre eux, l'ordinateur et la clinique s'accordaient sur le diagnostic d'état végétatif – et dans ce cas, seuls 20 % d'entre eux recouvraient un minimum de conscience. Par contre, pour le dernier tiers, notre système informatique détectait des signes de conscience là où le clinicien n'en voyait aucun – et dans ce cas, la probabilité de recouvrer la conscience dans les mois suivants grimpait à 50 % !

Les implications de ces résultats sont considérables. À l'aide de mesures automatisées de l'activité cérébrale, nous commençons à détecter des traces de conscience longtemps avant qu'elles ne se manifestent dans le comportement. Nos signatures de la conscience ont cessé d'appartenir à la théorie pure : elles sont en train de devenir plus sensibles que l'examen clinique traditionnel. La nouvelle science de la conscience commence à porter ses fruits.

Vers des interventions cliniques

> « Ne peux-tu donc soigner un esprit malade,
> Arracher de la mémoire un chagrin enraciné,
> Effacer les soucis gravés dans le cerveau,
> Et, par la vertu de quelque bienfaisant antidote d'oubli,
> Nettoyer le sein encombré de cette matière pernicieuse
> Qui pèse sur le cœur ? »
>
> William SHAKESPEARE, *Macbeth* (1606).

Détecter un frémissement de conscience n'est qu'un commencement. Ce que les patients et leurs familles attendent vraiment, c'est une réponse à l'interrogation shakespearienne : « Ne peux-tu donc soigner un esprit malade ? » Pouvons-nous aider les patients comateux ou en état végétatif à recouvrer une pleine conscience ? Certains récupèrent leurs facultés mentales des années après l'accident. Pourrions-nous accélérer ce retour à la normale ?

Quand les familles anéanties soulèvent ces questions, la communauté médicale donne généralement une réponse pessimiste. Dès qu'une année s'est écoulée, le patient est dit en « état végétatif permanent ». En filigrane, cette étiquette clinique suggère que, quels que soient les efforts de l'entourage, il est improbable que les choses changent. Et, dans bien des cas, telle est la triste vérité.

En 2007, pourtant, Nicholas Schiff et Joseph Giacino ont publié un article spectaculaire dans la célèbre revue *Nature*, qui suggérait que ce consensus devait être réexaminé[49]. Pour la première fois, ils démontraient qu'un traitement pouvait, peu à peu, ramener un patient minimalement conscient à un niveau de conscience stable. Leur stratégie thérapeutique consistait à insérer de longues électrodes dans les profondeurs du cerveau, afin de stimuler un site clé : le bien nommé « thalamus central » et les noyaux intralaminaires qui l'entourent.

Grâce aux travaux pionniers de Giuseppe Moruzzi et Horace Magoun dans les années 1940, ces régions sont bien connues : ce sont les principaux nœuds d'un réseau ascendant qui régule le niveau de vigilance du cortex[50]. Les noyaux du thalamus central sont denses en neurones dits « de projection » qui envoient leurs axones de façon diffuse en direction du cortex, particulièrement vers les lobes frontaux. Leurs connexions ciblent tout particulièrement les neurones pyramidaux qui occupent les couches supérieures du cortex – précisément celles qui établissent en priorité les connexions corticales à longue distance qui forment l'espace de travail neuronal global. Chez l'animal, l'activation du thalamus central module le niveau d'activité du cortex, accroît l'activité motrice et stimule l'apprentissage[51].

Dans un cerveau intact, l'activité du thalamus central est elle-même modulée, en sens inverse, par les aires du cortex préfrontal et cingulaire. Cette boucle de feedback nous permet probablement d'ajuster notre niveau d'excitation corticale en fonction du contexte : toute situation qui captive notre attention embrase ce circuit et galvanise nos facultés conscientes[52]. Lorsqu'un des éléments de cette boucle est lésé, par contre,

cela peut entraîner une réduction globale de l'activité neurale qui y circule, et diminuer d'un coup le niveau de vigilance. C'est pourquoi Schiff et Giacino ont émis l'hypothèse que la stimulation du thalamus central pourrait peut-être « réveiller » le cortex et restaurer, de l'extérieur, le niveau de vigilance que le cerveau du patient était devenu incapable d'atteindre de lui-même.

Comme nous l'avons vu, la vigilance et la conscience sont deux choses distinctes. Chez les patients en état végétatif, la vigilance demeure souvent intacte : ils s'éveillent le matin, ouvrent les yeux, mais cela ne suffit pas à replacer le cortex dans un état conscient. La plupart des patients en état végétatif persistant ne tirent guère de bénéfice de la stimulation thalamique. Les médecins implantèrent un stimulateur dans le cerveau de Terri Schiavo, mais son état de conscience demeura strictement inchangé, probablement parce que son cortex et ses connexions à longue distance avaient été dramatiquement touchés lors de son accident de voiture. Dans les rares cas de réussite, on ne peut pas exclure l'hypothèse d'une récupération spontanée.

Conscients de ces modestes résultats antérieurs, Schiff et Giacino tentèrent de maximiser leurs chances de succès. D'abord, ils ciblèrent le noyau latéral du thalamus central, dont les boucles contactent directement le cortex préfrontal. Ensuite, ils sélectionnèrent un patient qui avait déjà pratiquement atteint le seuil de la conscience. Souvenez-vous que c'est Joseph Giacino lui-même qui avait aidé à cerner l'« état de conscience minimale », cette catégorie de patients qui montrent des signes occasionnels de traitement conscient et de communication volontaire, sans parvenir à émettre ces réponses d'une façon systématique et reproductible. L'équipe de Schiff identifia un patient de ce type dont le cortex semblait particulièrement épargné. Bien qu'il soit dans un état stable de conscience minimale depuis des années, ses deux hémisphères s'activaient toujours lorsqu'on lui faisait entendre des phrases. Son métabolisme cortical au repos, par contre, était terriblement réduit. La

stimulation thalamique lui fournirait-elle l'impulsion qui lui permettrait de rebasculer dans un état de conscience stable ?

Schiff et Giacino procédèrent pas à pas. Avant d'implanter leurs électrodes, ils étudièrent le patient minutieusement pendant plusieurs mois. Ils le testèrent très régulièrement, avec la batterie d'évaluation de la récupération du coma, jusqu'à obtenir une idée très précise de ses capacités et de leurs fluctuations. Les résultats n'étaient ni blancs ni noirs : le patient parvenait parfois à faire quelques gestes volontaires, voire à marmonner un mot, mais ces comportements restaient erratiques. C'était, sans aucun doute, un patient en état de conscience minimale avec une très grande marge de progression.

Forts de ces observations, Schiff et Giacino passèrent à l'étape suivante : la mise en place des électrodes. Au cours d'une intervention chirurgicale délicate, ils guidèrent deux longs fils à travers le cortex, jusqu'au thalamus central. Quarante-huit heures plus tard, ils excitèrent ces électrodes. Le résultat ne se fit pas attendre : le patient, qui était en état de conscience minimale depuis six ans, ouvrit immédiatement les yeux, son rythme cardiaque s'accéléra, et il se tourna spontanément vers les personnes qui l'entouraient et qui lui parlaient. Cependant, lorsqu'on lui demanda de nommer divers objets, ses paroles restèrent « inintelligibles et limitées à des marmonnements incompréhensibles[53] ». Dès qu'on éteignit le stimulateur, tous ces comportements s'évanouirent.

Afin d'écarter toute influence de l'intervention chirurgicale, les chercheurs laissèrent s'écouler deux mois sans aucune stimulation. Il n'y eut aucune amélioration spontanée. Ensuite, un mois sur deux, en double aveugle, ils alternèrent la présence et l'absence de stimulation thalamique. Il s'ensuivit une rémission spectaculaire des symptômes. Tous les indicateurs d'éveil, de communication, de contrôle volontaire et de dénomination passaient au vert lorsque le stimulateur était allumé. De plus, et c'est un point essentiel, ces mesures ne baissaient guère pendant les mois où le stimulateur était éteint – le patient ne revenait pas en arrière. On assistait donc à une récupération lente, mais irréversible. Six

mois plus tard, le patient parvenait à se nourrir en portant une tasse à sa bouche. Sa famille rapportait d'immenses progrès dans sa capacité d'interagir avec les autres. Tout en restant très handicapé, il participait désormais activement à toutes les décisions qui le concernaient, et parvenait même à discuter des traitements qu'on lui proposait.

Ce succès thérapeutique apporte de grands espoirs. En ramenant l'activité corticale à un niveau proche de la normale, la stimulation cérébrale profonde peut permettre au cerveau de recouvrer une partie de son autonomie consciente.

Même chez les patients dont l'état semble stable depuis des décennies, le cerveau reste plastique et une récupération spectaculaire survient parfois spontanément. Les dossiers médicaux regorgent d'observations étonnantes. Un patient, figé en conscience minimale pendant dix-neuf ans, recouvra soudainement la parole et la mémoire. Les images de son cerveau, réalisées à l'aide de l'IRM de diffusion, suggéraient que les connexions à longue distance de son cortex avaient lentement repoussé[54]. Chez un autre patient, c'est la communication entre le lobe frontal et le thalamus, fortement diminuée lorsqu'il était dans l'état végétatif, qui revint à la normale lors de sa guérison spontanée.

Il ne faut certes pas s'attendre à ce qu'une telle récupération survienne à tous les coups – mais alors, comment expliquer que certains patients guérissent et d'autres non ? Clairement, si un nombre considérable de neurones préfrontaux sont morts, aucune stimulation ne les fera revenir. Parfois, par contre, ces neurones sont toujours présents, mais leurs connexions sont détériorées. Dans d'autres cas, c'est la dynamique autoentretenue de l'activité neuronale qui semble en cause : les connexions sont toujours présentes, mais l'information qui y circule ne suffit plus à maintenir un état soutenu d'activité cérébrale, et le cerveau s'éteint de lui-même. Tant que le circuit reste entier, il existe toujours un espoir de le rallumer et d'observer alors une récupération rapide et spectaculaire.

Par quels moyens pourrions-nous basculer l'interrupteur cortical en position « on » ? Les molécules qui modulent les circuits cérébraux de la dopamine sont de puissants agents pharmacologiques. Les neurones à dopamine envoient de nombreuses connexions modulatrices en direction du cortex préfrontal et des ganglions de la base qui contrôlent nos actions volontaires – les stimuler peut donc faciliter la récupération. De fait, trois patients en état végétatif persistant recouvrèrent soudainement la conscience dès qu'on leur administra une molécule appelée lévodopa, un précurseur chimique de la dopamine que l'on donne d'habitude aux patients atteints de la maladie de Parkinson[55]. L'amantadine est un autre stimulant du circuit de la dopamine qui, dans certaines études cliniques bien contrôlées, s'est avéré efficace pour accélérer la sortie de l'état végétatif ou de conscience minimale[56].

D'autres médicaments ont des effets plus étranges encore. Le zolpidem, vendu sous le nom de Stilnox, est un somnifère qui, dans certains cas paradoxaux, ravive la conscience et l'action volontaire. Un patient était totalement léthargique, muet et immobile depuis des mois, figé dans un tableau clinique appelé « mutisme akinétique ». Pour l'aider à s'endormir, on lui donna un cachet de Zolpidem – et il s'éveilla d'un coup, se mit à bouger et à parler[57]. Autre cas étrange : une femme qui avait subi un grave accident vasculaire de l'hémisphère gauche, était devenue si aphasique qu'elle ne parvenait pas à prononcer plus d'une syllabe de temps à autre. Elle se vit prescrire du Zolpidem afin de lutter contre ses insomnies. La première fois qu'elle en prit, elle se remit immédiatement à parler pendant quelques heures ! Elle pouvait répondre à des questions, compter, et même nommer des objets. Elle finit par s'endormir, et le lendemain son aphasie était revenue. Le phénomène se produisit chaque soir, dès que sa famille lui donnait son somnifère[58]. Non seulement il ne la faisait pas dormir, mais il réveillait ses circuits linguistiques endormis.

Nous commençons à peine à entrevoir une explication de ces phénomènes. Ils semblent provenir des nombreuses connexions qui relient l'espace de travail cortical au thalamus et à deux des ganglions de la base

(le striatum et le pallidum). Grâce à ces circuits, le cortex peut s'exciter lui-même : l'activation se propage en boucle, depuis le cortex frontal vers le striatum, puis le pallidum, le thalamus, et enfin le retour au cortex. Deux de ces étapes reposent sur des connexions inhibitrices plutôt qu'excitatrices : le striatum inhibe le pallidum, et le pallidum, à son tour, inhibe le thalamus. Lorsque le cerveau cesse d'être oxygéné, les cellules inhibitrices du striatum semblent être les premières à en souffrir. Résultat, le pallidum n'est plus suffisamment inhibé, son activité s'accroît et finit par éteindre celle du thalamus et du cortex, ce qui les empêche d'atteindre le niveau d'activation suffisant pour entretenir un état conscient.

Cependant, la plupart de ces circuits demeurent intacts. Ils sont simplement inhibés, et peuvent toujours se rallumer si l'on brise ce cercle vicieux, par exemple en implantant une électrode dans le thalamus qui contrecarre l'excès d'inhibition des neurones thalamiques. Une autre possibilité consiste à injecter de la dopamine ou de l'amantadine, qui ont pour effet d'exciter le cortex soit directement, soit par l'intermédiaire des neurones du striatum. Enfin, un médicament tel que le Zolpidem peut servir à inhiber l'inhibition : en se fixant sur les nombreux récepteurs qui tapissent les synapses inhibitrices du pallidum, il force ces cellules inhibitrices à se taire, ce qui, par contrecoup, libère le cortex et le thalamus. Tous ces mécanismes, bien qu'ils soient encore hypothétiques, expliquent pourquoi ces interventions finissent par avoir le même effet : elles permettent à l'activité corticale de remonter à un niveau normal[59].

Ces astuces ne peuvent fonctionner que si le cortex lui-même n'est pas trop endommagé. Lorsque le cortex préfrontal semble intact à l'IRM anatomique, alors que son métabolisme est fortement réduit, c'est un signe favorable : peut-être est-il simplement désactivé et prêt à sortir de sa léthargie. Une fois stimulé, il récupérera lentement ses capacités d'autorégulation. En fonctionnement normal, la plupart des synapses du cerveau sont plastiques et peuvent augmenter leur efficacité. Grâce à cette plasticité, les connexions de l'espace de travail neuronal global peuvent

regagner en force et redevenir capables de promouvoir un état de conscience stable et prolongé.

Même si le cortex est endommagé, des solutions futuristes verront peut-être prochainement le jour. Si l'hypothèse de l'espace de travail est juste, la conscience n'est rien d'autre que la circulation flexible d'informations au sein d'un vaste réseau de neurones. Serait-il tiré par les cheveux d'imaginer que certaines de ces régions et de leurs connexions soient remplacées un jour par des boucles externes ? Les interfaces cerveau-machine, qui s'appuient sur des puces implantées directement dans le cortex, offriront bientôt l'espoir d'une restauration de la connectivité à longue distance. Nous pourrons recueillir des décharges neuronales spontanées dans le cortex préfrontal ou prémoteur, et les rejouer à d'autres régions auditives ou visuelles – soit directement, sous la forme de petites décharges électriques, soit, plus simplement encore, en les recodant sous la forme de signaux visuels ou auditifs. La substitution sensorielle est déjà utilisée chez les aveugles de naissance, afin de leur faire « voir » une image en la transcrivant en sons[60]. En suivant le même principe, peut-être pourrait-on reconnecter le cerveau avec lui-même et restaurer ainsi la densité critique de communications internes nécessaires au maintien d'un état conscient.

L'avenir dira si cette idée tient la route. Quoi qu'il en soit, il ne fait aucun doute que, dans la prochaine décennie, l'intérêt renouvelé pour le coma et les états perturbés de conscience, fondé sur de solides bases théoriques et empiriques, conduira à des changements importants en médecine. Dans le domaine des troubles de la conscience, une révolution thérapeutique est en route.

CHAPITRE 7

L'avenir de la conscience

La science de la conscience comporte encore de nombreuses zones d'ombre. Quand apparaît la conscience chez le bébé ? Un singe, un chien ou un dauphin sont-ils conscients de leur environnement ou d'eux-mêmes ? Comment aborder l'énigme de la conscience de soi, cette capacité paradoxale de penser à nos propres pensées ? Le cerveau humain présente-t-il, dans ce domaine, des compétences singulières ? Contient-il des circuits neuronaux propres à notre espèce, dont le dysfonctionnement pourrait expliquer l'origine de maladies telles que la schizophrénie ? Et si nous parvenons à comprendre ces réseaux cérébraux, pourrons-nous les répliquer dans un ordinateur et donner naissance à une conscience artificielle ?

« Au fond, je suis assez contrariée que la science vienne fourrer son nez dans cette affaire, mon affaire à moi. Ne s'est-elle pas déjà approprié une part suffisante de la réalité ? Doit-elle aussi avoir des prétentions sur l'essence intangible, invisible de la personne humaine ? »

David LODGE, *Pensées secrètes* (2001).

« En fait, plus votre science est grande, plus est profonde votre conscience du mystère. »

Vladimir NABOKOV,
Intransigeances (1973).

La boîte noire de la conscience est désormais entrouverte. Grâce à toute une panoplie d'astuces expérimentales, nous avons appris, à volonté, à rendre des images visibles ou invisibles. Par l'imagerie cérébrale, nous avons pu suivre dans le cerveau l'activité qu'elles évoquent et y déceler le propre des états conscients. Clarifier comment le cortex manipule des images conscientes et inconscientes s'est avéré plus facile qu'il n'y paraissait au premier abord. Certes, nous sommes loin d'avoir

tout compris, mais nous avons identifié de nombreuses signatures électrophysiologiques qui témoignent de la présence d'un embrasement conscient. Ces signatures de la conscience se sont révélées suffisamment solides pour être utilisées en clinique, afin d'évaluer le degré de conscience résiduelle des patients atteints de lésions cérébrales.

Indubitablement, nous n'en sommes qu'aux balbutiements de la science de la conscience. De nombreuses énigmes fondamentales nous échappent encore. Dans ce chapitre de conclusion, je voudrais décrire comment je vois l'avenir de la recherche sur la conscience, quels sont les problèmes majeurs qui donneront du fil à retordre aux chercheurs dans les années à venir et comment les aborder.

Certaines de ces questions sont purement empiriques, et elles ont déjà reçu un début de réponse. Par exemple, quand la conscience émerge-t-elle, aussi bien dans l'évolution qu'au cours du développement de l'enfant ? Un nouveau-né est-il déjà conscient ? *Quid* d'un prématuré ou d'un fœtus ? Les singes, les souris et les oiseaux possèdent-ils un espace de travail conscient semblable au nôtre ?

D'autres questions confinent à la philosophie – et pourtant, je suis persuadé qu'elles finiront par recevoir une réponse empirique, dès lors que nous découvrirons un angle d'attaque expérimentale. Par exemple, qu'est-ce que la conscience de soi ? Il existe très certainement, dans le cerveau humain, un système qui permet de retourner le projecteur de la conscience sur lui-même et d'analyser ses propres pensées. Cette architecture neuronale, si tant est que nous puissions l'identifier, est-elle propre à l'espèce humaine ? Qu'est-ce qui rend le cerveau humain si puissant, mais également si vulnérable aux maladies psychiatriques telles que la schizophrénie ? Les réponses à toutes ces questions nous permettront-elles, un jour, de construire une conscience artificielle – un robot doué de sensations conscientes ? Ressentirait-il des émotions proches des nôtres ? Éprouverait-il un sentiment de libre arbitre ?

Nul ne peut aujourd'hui donner des réponses fermes à ces énigmes, et je ne prétendrai pas les résoudre. Mais je voudrais montrer comment les

neurosciences cognitives peuvent les aborder.

Un bébé est-il conscient ?

Prenez la question de l'origine de la conscience dans l'enfance. Un bébé est-il conscient ? Que dire d'un nouveau-né ? d'un prématuré ? d'un fœtus dans le ventre de sa mère ? Il ne fait guère de doute qu'un minimum d'organisation cérébrale doit être présent pour qu'émergent des pensées conscientes – mais combien exactement ?

Pendant des décennies, une vive controverse a opposé les rationalistes et les défenseurs du caractère sacré de la personne humaine. Les provocations abondent dans les deux camps. Ainsi le philosophe Michael Tooley, de l'Université du Colorado, écrit-il sans ménagements que « les nouveau-nés humains ne sont ni des personnes, ni même des quasi-personnes, et leur destruction n'a donc rien d'intrinsèquement mauvais[1] ». Tooley va jusqu'à dire qu'avant l'âge de 3 mois l'infanticide peut se justifier sur le plan moral parce qu'un nouveau-né « ne possède aucun concept de la permanence de soi, pas plus qu'un chaton qui vient de naître », et qu'en conséquence il n'a « aucun droit à la vie[2] ». En écho à ces sinistres propos, le professeur Peter Singer, qui enseigne pourtant la bioéthique à l'Université de Princeton, soutient que « la vie, dans le sens moralement fort du terme, ne commence que lorsqu'il existe une conscience de sa propre existence à travers le temps […]. Le fait qu'un être soit un être humain, qu'il appartienne à l'espèce *Homo sapiens*, n'est pas pertinent dans le débat s'il est bien ou mal de le tuer. Ce qui compte, ce sont des caractéristiques telles que la rationalité, l'autonomie et la conscience de soi. Or les bébés ne possèdent aucune de ces caractéristiques. Les tuer n'a donc rien de comparable au meurtre d'êtres humains normaux ou d'autres êtres dotés d'une conscience de soi[3] ».

Ces affirmations péremptoires sont ridicules. Elles s'opposent violemment aux intuitions morales que nous partageons tous, et selon lesquelles tous les êtres humains, depuis les prix Nobel jusqu'aux enfants handicapés, ont les mêmes droits à la vie. Elles violent également toutes nos intuitions sur la conscience – posez la question à n'importe quelle mère qui vient d'échanger son premier regard avec son nouveau-né. Mais le plus choquant, à mon sens, c'est que Tooley et Singer prononcent leurs terrifiants oukases sans se préoccuper le moins du monde des données expérimentales. Comment savent-ils que les bébés sont dépourvus de toute conscience ? S'appuient-ils sur la littérature scientifique ? Absolument pas : leurs affirmations ne reposent que sur des *a priori* détachés de toute réalité expérimentale – et souvent complètement faux. Ainsi Singer écrit-il que, « à maints égards, rien d'important ne distingue [les patients comateux et en état végétatif] des bébés démunis. Ils ne sont pas conscients d'eux-mêmes, ni rationnels, ni autonomes. (…) leur vie n'a aucune valeur intrinsèque. Leur voyage dans cette vie est arrivé à son terme ». Au chapitre précédent, nous avons vu à quel point cette thèse est fausse. Au contraire, l'imagerie cérébrale démontre régulièrement l'existence d'une conscience résiduelle chez certains adultes en état végétatif apparent. L'arrogance de ce point de vue, qui bafoue allégrement tout ce que nous savons de la complexité de la vie et de la conscience, me répugne. Le cerveau mérite une meilleure philosophie !

La voie que je propose est simple : nous devons poursuivre les recherches afin de trouver comment poser ces questions de façon empirique. Chez le bébé, de nouveaux travaux fondés sur le comportement, l'anatomie et même l'imagerie cérébrale commencent à jeter une certaine lumière sur le développement de la conscience. Les signatures de la conscience, que nous avons validées chez l'adulte, peuvent et doivent être recherchées chez l'enfant aux différentes étapes de son développement.

Bien sûr, cette stratégie reste imparfaite, car elle se fonde sur une analogie. Nous espérons détecter, dans l'enfance, l'apparition des

marqueurs objectifs qui définissent la conscience chez l'adulte. Si nous parvenons à les mettre en évidence, nous en conclurons qu'à cet âge l'enfant possède déjà un état mental conscient, un point de vue subjectif sur le monde extérieur. Soyons clairs : les choses pourraient être plus complexes. Les marqueurs de la conscience pourraient être différents selon l'âge des enfants. Qui plus est, il se peut que l'expérience ne nous donne pas une réponse dépourvue d'ambiguïté. L'espace de travail neuronal qui fonctionne comme un tout chez l'adulte pourrait, dans la petite enfance, être fragmenté en éléments distincts qui se développent à leur propre rythme. Dans ce cas, les différentes signatures de la conscience pourraient être en désaccord : certaines suggéreraient que le bébé est conscient, d'autres qu'il ne l'est pas. En dépit de toutes ces difficultés potentielles, seule la méthode expérimentale peut nous permettre d'avancer dans ce débat. Toute connaissance scientifique objective, même partielle, vaut mieux que les exhortations *a priori*, qu'elles proviennent de religieux ou de philosophes.

De quelles données disposons-nous donc sur l'existence d'un espace de travail conscient chez le bébé ? Et d'abord, que dit l'anatomie ? Au siècle dernier, les pédiatres ont noté l'immaturité du cortex des bébés, empli de neurones maigrichons aux dendrites chétives et à l'axone dépourvu de toute isolation électrique. C'est pourquoi on a longtemps pensé qu'à la naissance les opérations mentales du bébé étaient pratiquement inexistantes. Seuls quelques îlots de matière grise, dans les aires visuelles, auditives et motrices, semblaient suffisamment mûrs pour conférer à l'enfant quelques sensations primitives et quelques réflexes élémentaires. Les entrées sensorielles se mélangeaient pour engendrer « une immense confusion bourdonnante », selon les mots fameux de William James. On estimait que les aires du cortex préfrontal impliquées dans le raisonnement, très immatures, ne s'activaient pas avant la fin de la première année de vie. Cette lobotomie virtuelle expliquait pourquoi le bébé échouait systématiquement dans certains tests comportementaux qui exigeaient de planifier et de contrôler ses gestes, comme le célèbre test

« A-non-B » de Jean Piaget[4]. Il semblait également évident, pour la plupart des pédiatres, que les nouveau-nés ne ressentaient pas la douleur. Dans ce cas, pourquoi les anesthésier ? On pratiquait couramment des injections et même des interventions chirurgicales sans se préoccuper d'une éventuelle conscience du bébé.

Les avancées récentes des sciences cognitives et de l'imagerie cérébrale ont bouleversé cette vision de l'enfant. En confondant l'immaturité avec l'absence de fonction, nos prédécesseurs commettaient une erreur importante. Au sein même de l'utérus, dès six mois et demi de grossesse, le cortex de l'enfant commence à se former et à se plisser. Dès la naissance, même les régions corticales les plus distantes sont déjà reliées par des faisceaux de connexions à longue distance[5]. Bien qu'elles ne soient pas encore recouvertes de myéline, l'isolant des axones, ces connexions sont opérationnelles et transmettent déjà des informations, à une vitesse nettement plus lente que chez l'adulte. Le nouveau-né présente déjà une activité cérébrale spontanée qui s'organise en réseaux fonctionnels partiellement semblables à ceux du cerveau adulte[6].

L'un des domaines les mieux étudiés est celui du langage. Les bébés sont attirés par la parole comme des aimants. L'apprentissage du langage commence probablement *in utero*, car même des nouveau-nés de quelques heures distinguent déjà des phrases dans leur langue maternelle de celles prononcées dans une langue étrangère[7]. L'acquisition du langage est si rapide qu'une longue lignée de scientifiques prestigieux, de Darwin à Chomsky et Pinker, a postulé qu'elle repose sur des mécanismes spécialisés propres au cerveau humain. Mon épouse, Ghislaine Dehaene-Lambertz, et moi-même avons directement mis à l'épreuve cette idée : nous avons été les premiers à utiliser l'IRM pour visualiser l'activité du cerveau d'un bébé qui écoute sa langue maternelle[8]. Confortablement emmitouflés, allongés sur un matelas en mousse, les oreilles protégées du bruit de l'IRM par un puissant casque antibruit, les bébés, âgés de 2 mois, écoutaient tranquillement des histoires pour enfants tandis que nous

prenions une image de l'activité de tout leur cerveau toutes les 3 secondes.

À notre stupéfaction, l'activité cérébrale ne se limitait absolument pas aux aires auditives primaires. Au contraire, tout un réseau de régions corticales s'allumait (figure 34). Leur activité traçait très précisément les contours des aires classiques du langage chez l'adulte. Dès cet âge, les informations linguistiques sont traitées par les aires temporales et frontales de l'hémisphère gauche, tandis que d'autres stimuli tout aussi complexes comme la musique de Mozart sont envoyés aux régions symétriques de l'hémisphère droit[9]. Même la célèbre aire de Broca, dans le cortex préfrontal inférieur gauche, s'active déjà en réponse aux phrases entendues. Le silence total du cortex frontal des bébés n'est donc qu'un mythe. En fait, l'étude anatomique de la région de Broca a démontré que ce site mûrit rapidement et qu'il est l'une des régions les mieux connectées du cortex frontal du bébé[10].

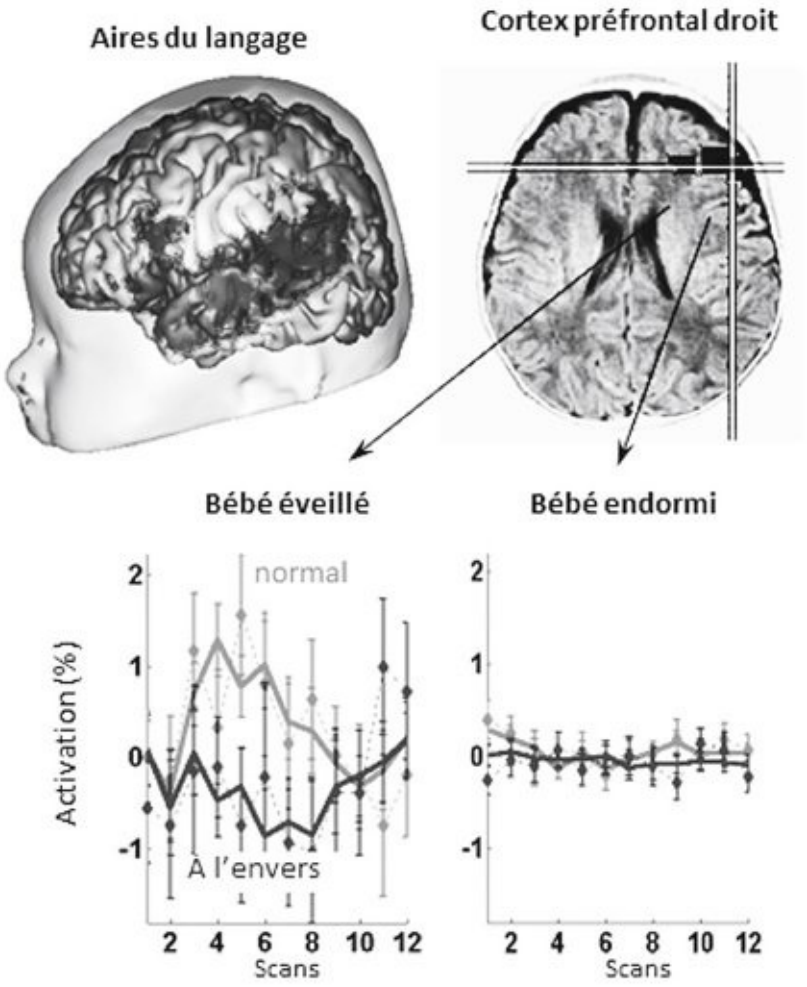


Figure 34. Le cortex préfrontal est déjà actif chez le bébé éveillé. Nous avons enregistré, par IRM fonctionnelle, l'activité cérébrale évoquée par des phrases dans la langue maternelle chez le bébé de 2 mois. L'écoute de parole active déjà le réseau des aires du langage, y compris la région frontale inférieure gauche connue sous le nom d'« aire de Broca ». Lorsqu'on joue la même bande à l'envers, ce qui détruit la plupart des indices linguistiques, l'activité cérébrale est réduite. Seuls les bébés éveillés activent leur cortex préfrontal droit – une activation qui semble associée à l'accès à la conscience, dans la mesure où elle disparaît lorsque les enfants s'endorment.

En mesurant la vitesse de cette activation, nous avons confirmé que le réseau du langage s'active déjà chez le bébé, mais à une vitesse nettement inférieure à celle de l'adulte, particulièrement dans le cortex préfrontal[11]. Cette lenteur empêche-t-elle la prise de conscience ? Les bébés traitent-ils la parole comme des « zombies », de la même manière qu'un cerveau

comateux continue de détecter inconsciemment les sons nouveaux ? Le simple fait qu'un enfant de 2 mois, exposé à des phrases, active le même réseau d'aires cérébrales qu'un adulte ne permet pas de conclure, parce que nous savons que la quasi-totalité de ce réseau peut s'activer en l'absence de conscience, par exemple au cours d'une anesthésie générale[12]. Toutefois, notre expérience montrait également que les bébés possèdent une forme rudimentaire de mémoire de travail. Lorsque nous leur répétions la même phrase après un délai de 14 secondes, leur région de Broca s'activait plus fortement la seconde fois que la première[13]. Ainsi, ils se souvenaient de ce qu'ils avaient entendu. Leur cerveau présentait l'une des caractéristiques de la conscience, la capacité de maintenir une information en mémoire de travail pendant quelques secondes.

Tout aussi important est le fait que les réponses cérébrales variaient selon que les bébés étaient endormis ou éveillés. Le cortex auditif s'allumait toujours, mais l'activité ne se propageait jusqu'au cortex préfrontal dorso-latéral que chez les bébés éveillés. Chez ceux qui s'étaient endormis dans l'IRM, l'activité de cette région ne déviait pas de zéro (figure 34). Le cortex préfrontal, nœud essentiel du réseau de l'espace de travail global chez l'adulte, semblait donc, là encore, contribuer au traitement conscient uniquement chez le bébé éveillé.

Plus récemment, nous avons obtenu une preuve directe que les bébés de quelques mois sont déjà conscients à l'aide du test local-global, que j'ai décrit au chapitre 6, et qui évalue la conscience résiduelle chez les patients en état végétatif. Souvenez-vous, ce test consiste à écouter une série de sons *bip bip bip bip boup*, tandis que nous enregistrons les ondes cérébrales à l'aide de l'EEG. Occasionnellement, une séquence viole la règle établie, par exemple parce qu'elle se termine par un cinquième *bip* au lieu du *boup* habituel. Lorsque ce stimulus nouveau évoque une onde P3, qui envahit le cortex préfrontal et les autres régions de l'espace de travail neuronal global, l'auditeur a de très grandes chances d'être conscient.

L'immense avantage de ce test est qu'il ne requiert aucune éducation ni aucune instruction. Il est donc idéal pour évaluer la conscience du petit enfant ou, comme nous le verrons plus loin, celle d'autres espèces animales. Même un bébé peut écouter une série de sons et, si son cerveau est suffisamment organisé, en extraire les régularités. Enregistrer l'EEG dès la naissance ne pose pas de difficultés particulières. Il faut juste éviter que le bébé ne s'ennuie et cesse de prêter attention aux stimuli. Pour rechercher une signature de la conscience chez le bébé, Ghislaine, qui est neuropédiatre et spécialiste de la cognition infantile, a donc adapté notre test local-global aux tout-petits. Elle l'a transformé en un spectacle multimédia, où des visages attrayants prononcent une série de voyelles, par exemple *a, a, a, i*. Les visages en mouvement fascinent les bébés, et une fois leur attention captivée, nous avons eu le plaisir de constater qu'à 2 mois leur cerveau émet déjà une réponse à la nouveauté globale – une signature de la conscience[14].

La plupart des parents ne seront guère surpris d'apprendre que leur bébé de 2 mois est déjà conscient. Cependant, notre test montre également que leur conscience diffère de celle de l'adulte sur un point important : leurs réponses cérébrales sont bien plus lentes. Chaque étape de traitement semble prendre nettement plus de temps. Nos bébés avaient besoin de 300 millisecondes pour détecter le changement de voyelle, alors que, chez l'adulte, cette réponse locale (*mismatch negativity*) ne prend que 100 à 150 millisecondes. Surtout, il faut à un bébé plus d'une seconde pour que le cortex préfrontal réagisse à la nouveauté globale – près de trois à quatre fois plus de temps qu'un adulte. Ainsi, nous pensons que le cerveau du bébé, dès les premières semaines de vie, abrite déjà un espace de travail neuronal global, mais que celui-ci fonctionne au ralenti.

Mon collègue Sid Kouider est parvenu à la même conclusion en utilisant la vision plutôt que l'audition. Il s'est intéressé au traitement des visages, un domaine pour lequel même les nouveau-nés présentent des compétences innées[15]. Les bébés adorent les visages et se tournent spontanément vers eux dès la naissance. Sid Kouider s'est appuyé sur ce

tropisme naturel pour étudier si les bébés sont déjà sensibles au masquage visuel et présentent un seuil d'accès à la conscience comparable à celui de l'adulte. Il a adapté aux bébés de 5 mois l'expérience de masquage que nous avions menée chez l'adulte[16]. Un visage attrayant apparaissait sur l'écran, pendant une durée courte et variable. Il était immédiatement suivi d'une image aléatoire, assez laide, qui servait de masque (figure 35). La question était simple : le bébé parvenait-il à voir le visage ? Prenait-il conscience de sa présence ?

Rappelez-vous que lorsqu'on présente à un adulte une image ainsi masquée, il ne la voit que si sa durée dépasse un vingtième de seconde. Les bébés, évidemment, ne peuvent pas dire s'ils voient le visage, mais leurs yeux, comme ceux d'un patient *locked-in*, racontent la même histoire. Lorsque le visage est présenté pendant une durée très courte, les bébés n'orientent pas leur regard, ce qui suggère qu'ils ne l'ont pas vu. C'est seulement au-delà d'un certain seuil qu'ils finissent par tourner les yeux et la tête. Comme les adultes, ils souffrent du masquage et ne perçoivent pas les images « subliminales », c'est-à-dire présentées en dessous du seuil. Les mesures indiquent que ce seuil est deux à trois fois plus élevé chez le bébé que chez l'adulte. Un bébé de 5 mois ne détecte le visage que s'il est présenté pendant au moins 100 millisecondes, alors que chez l'adulte, le seuil est de 40 ou 50 millisecondes. Il est fascinant de voir que le seuil s'abaisse brutalement pour atteindre la valeur adulte aux alentours de 10 à 12 mois, quand apparaissent de nombreux comportements qui dépendent de la maturation du cortex préfrontal[17].

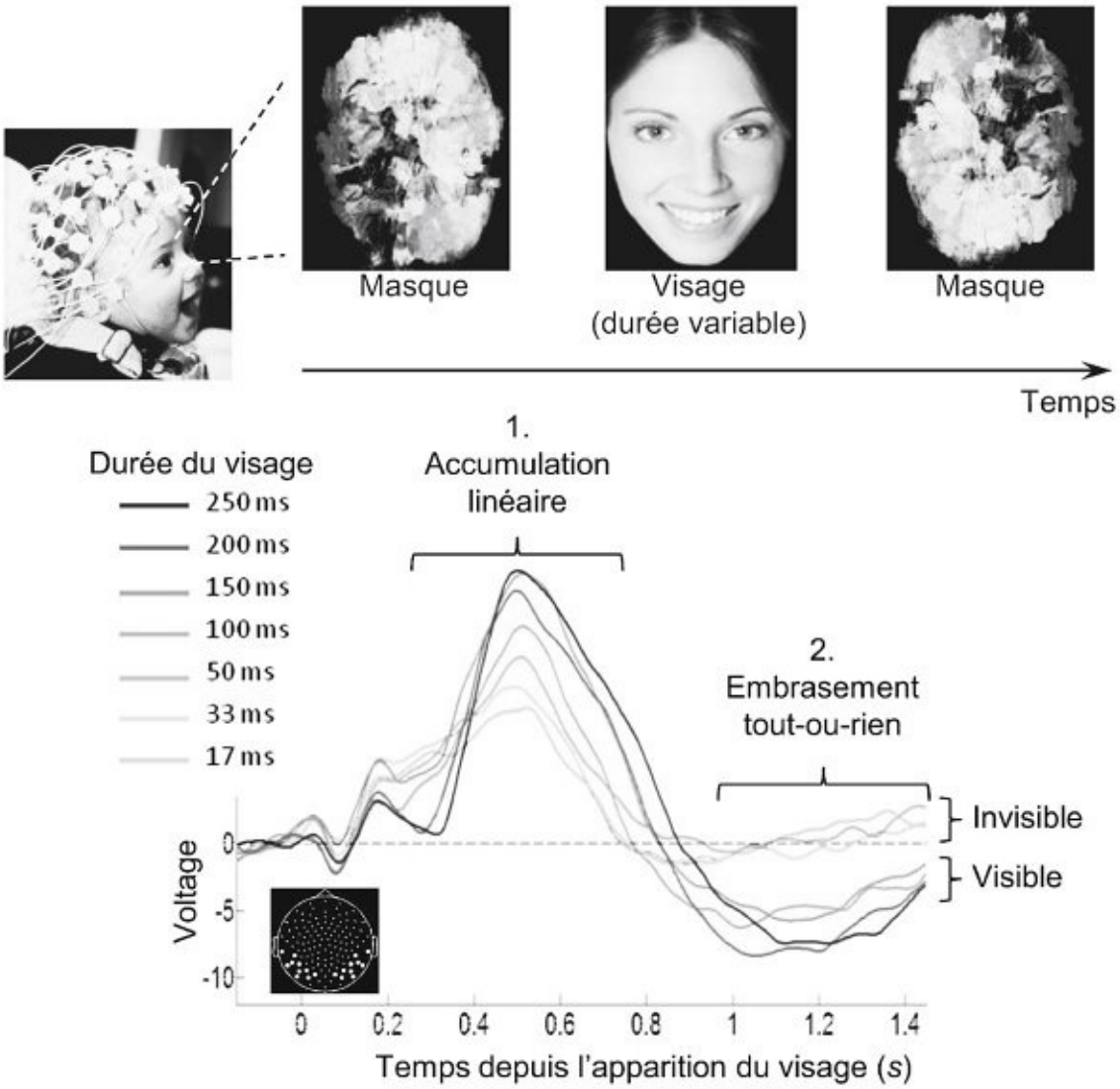


Figure 35. Les bébés présentent déjà certaines signatures de la perception consciente, bien qu'ils traitent les informations à vitesse réduite. Dans cette expérience, un visage s'affiche brièvement. Il est entouré d'images qui le rendent visible ou invisible. Plusieurs étapes de traitement sont visibles : d'abord une accumulation linéaire des informations sensorielles, puis un embrasement non linéaire qui semble coïncider avec l'accès à la conscience. Le bébé présente donc les mêmes étapes de traitement que l'adulte, mais il lui faut plus d'une seconde pour déclencher un embrasement global, ce qui est trois fois plus lent que dans le cerveau d'un adulte.

Une fois démontrée l'existence d'un seuil de conscience, Sid Kouider, Ghislaine Dehaene-Lambertz et moi avons étudié les réponses cérébrales

des bébés aux visages masqués. Nous avons observé exactement la même séquence que chez l'adulte : une étape perceptive linéaire suivie d'un embrasement global non linéaire (figure 35). Au cours de la première phase, l'activité des régions visuelles du cerveau croît en proportion directe de la durée de présentation du visage : que les images soient présentées en dessous ou au-dessus du seuil, le cerveau de l'enfant accumule simplement les données disponibles. Dans une seconde phase, seules les images présentées au-dessus du seuil déclenchent une onde négative lente et de grande amplitude, qui envahit les régions antérieures. Tant dans sa forme que dans sa localisation, cette activité tardive ressemble à l'onde P3 que l'on voit chez l'adulte. De toute évidence, lorsque les données sensorielles sont suffisantes, le cerveau du bébé déclenche une avalanche d'activation qui se propage dans le cortex préfrontal. Cet embrasement global se produit bien plus lentement chez le bébé que chez l'adulte, mais son organisation est essentiellement la même, ce qui laisse supposer que le bébé prend déjà conscience des images. Contrairement à l'adulte, il ne peut pas décrire ce qu'il voit – mais son cerveau présente déjà une signature claire de l'accès à la conscience.

Dès qu'une image ou un son nouveau attire l'attention du bébé, une onde lente et négative apparaît sur les électrodes situées dans les régions frontales[18]. D'autres chercheurs ont remarqué à quel point elle ressemble à l'onde P3 émise par le cerveau adulte[19], qui accompagne la prise de conscience de n'importe quelle information nouvelle. Une négativité frontale survient, par exemple, lorsqu'un bébé prête attention à un son différent des autres[20] – mais cette réponse disparaît sélectivement lorsque le bébé s'endort[21]. D'expérience en expérience, cette onde frontale ne cesse de se comporter comme un marqueur d'un traitement conscient.

Une conclusion s'impose : la prise de conscience fonctionne déjà chez le bébé, mais à une vitesse trois ou quatre fois plus lente que chez l'adulte. D'où provient cette lenteur ? Souvenez-vous que dans la première année de vie, le cerveau est immature. Tous les grands faisceaux de connexions à longue distance qui forment l'espace de travail neuronal global sont

présents dès la naissance[22], mais leur isolation électrique fait défaut. Les couches de myéline, cette membrane de lipides qui entoure les axones, ne se développent que très lentement, tout au long de l'enfance et même de l'adolescence. En isolant les axones, elles accroissent la vitesse et la fidélité de la transmission de l'influx nerveux à travers le cortex. Câblé mais encore mal isolé, le cerveau conscient du bébé intègre les informations au ralenti. Peut-être pouvons-nous comparer l'indolence du bébé à la torpeur d'un patient qui vient de sortir du coma ou d'une anesthésie générale : il leur faut parfois 1 ou 2 secondes pour qu'émerge un sourire. Leurs pensées sont diffuses, affaiblies et engourdies – mais ce sont des pensées conscientes.

Jusqu'à présent, les plus jeunes enfants que nous avons testés étaient âgés de 2 mois. Nous ne savons donc pas à quel moment précis apparaît la conscience. Un nouveau-né est-il déjà conscient ? Doit-il attendre plusieurs semaines avant que son architecture corticale globale ne se mette à fonctionner ? Nous ne le savons pas, mais je ne serai pas surpris de découvrir que la conscience est présente dès la naissance. De nombreuses connexions anatomiques à longue distance parcourent déjà tout le cortex du nouveau-né. Ne sous-estimons pas leur puissance de calcul. Quelques heures après la naissance, un bébé présente déjà des compétences remarquables, telles que la capacité de discriminer des ensembles de quatre ou de douze objets sur la base de leur nombre[23]. Dès lors, pourquoi ne serait-il pas déjà conscient ?

Le pédiatre suédois Hugo Langercrantz et le neurobiologiste français Jean-Pierre Changeux ont proposé une hypothèse passionnante : la naissance coïnciderait avec la première prise de conscience[24]. Dans l'utérus, le fœtus serait sous anesthésie générale, baigné dans un flot permanent de drogues qui comprennent notamment la prégnanolone, une hormone stéroïde dont l'effet est anesthésique, et la prostaglandine D2, un somnifère sécrété par le placenta. La naissance coïncide avec un afflux massif d'hormones de stress et de neurotransmetteurs aux effets stimulants comme les catécholamines. Dans les heures qui suivent, le

nouveau-né est dans un état d'éveil maximal, les yeux grands ouverts. Éprouve-t-il alors sa première expérience consciente ? La pharmacologie nous souffle que l'accouchement pourrait bien être un moment plus important encore que nous ne le pensions : la naissance d'un esprit conscient.

Une conscience chez l'animal ?

> « Celui qui comprendra le babouin contribuerait davantage à la métaphysique que Locke. »
>
> Charles DARWIN, *Carnets* (1838).

Les questions que nous posons aux bébés, il est tout aussi légitime de les adresser à nos cousins du règne animal. Dépourvu de langage, comme le nouveau-né, l'animal ne peut rapporter ses pensées conscientes – mais cela n'implique pas qu'il en soit dépourvu. Notre planète est peuplée d'une extraordinaire diversité d'espèces animales : prédateurs patients et attentifs (guépard, aigle, murène), habiles navigateurs (oie, éléphant), joueurs malicieux (chat, loutre), puissants esprits capables de résoudre des problèmes nouveaux (pie, poulpe), génies de l'imitation vocale (perroquet gris du Gabon), champions de la vie en société (loup, chauve-souris)… Je serais bien surpris qu'aucun d'entre eux ne ressente la moindre expérience consciente. Ma théorie propose que l'architecture de l'espace neuronal conscient facilite l'échange d'informations dans le cerveau. Essentielle à la survie, cette fonction pourrait donc avoir émergé depuis bien longtemps dans l'évolution, et peut-être plus d'une fois.

Sur le plan anatomique, un réseau de connexions à longue distance dense, reliant le cortex préfrontal aux autres aires associatives, est présent chez le singe macaque, et peut-être chez tous les mammifères. Même la souris dispose d'un tout petit bout de cortex préfrontal et cingulaire qui

s'active lorsqu'elle garde une information en mémoire pendant un bref instant[25]. Certains oiseaux, particulièrement ceux qui imitent et qui vocalisent, pourraient posséder des circuits neuronaux analogues et qui remplissent le même rôle[26].

L'attribution d'une conscience animale ne se fonde pas seulement sur l'anatomie. Bien qu'ils ne parlent pas, les singes peuvent être entraînés à décrire ce qu'ils voient en appuyant sur les touches d'un clavier d'ordinateur. De nombreuses expériences de ce type suggèrent qu'ils éprouvent des sensations subjectives semblables aux nôtres. On peut par exemple les entraîner à appuyer sur une touche lorsqu'une lumière s'allume, et sur une autre touche lorsque l'écran reste noir. Cette action motrice peut alors servir de « rapport conscient », un geste équivalent à la phrase « je crois que j'ai vu une lumière » ou « je n'ai rien vu ». On peut également entraîner un singe à classifier des images selon qu'il s'agit ou non de visages. Une fois entraîné avec des images normales, l'animal peut être testé avec des images masquées, les mêmes qui servent à évaluer la prise de conscience chez l'homme.

Toutes ces études convergent : les singes perçoivent les illusions visuelles de la même manière que nous. Lorsqu'on leur montre deux images différentes à chaque œil, ils « disent » faire l'expérience de la rivalité binoculaire. En effet, bien que le stimulus soit fixe, ils appuient en alternance sur les touches pour indiquer qu'ils perçoivent tantôt l'une et tantôt l'autre. Leur perception consciente des images ne cesse d'alterner, au même rythme que la nôtre[27].

Les singes sont également sensibles au masquage visuel. Lorsqu'on leur présente une image subliminale, très brève et suivie d'un masque, les singes macaques rapportent qu'ils n'ont rien vu – bien que leur cortex visuel, comme le nôtre, ait été parcouru d'une série de décharges neuronales[28]. Comme nous, ils possèdent un seuil bien précis au-delà duquel l'image devient visible.

Enfin, lorsque leur cortex visuel primaire est lésé, les singes aussi présentent un syndrome de « vision aveugle ». Ils parviennent toujours à

pointer du doigt ou à orienter leur regard vers un point lumineux qui apparaît à l'endroit de leur champ visuel qui correspond à la lésion. Pourtant, si on les entraîne à rapporter la présence ou l'absence de lumière, ils « décrivent » les stimuli qui apparaissent à cet endroit en utilisant la réponse « pas de lumière ». Exactement comme les humains qui souffrent de vision aveugle, ils s'orientent inconsciemment, mais leur capacité de détection consciente semble avoir disparu[29].

Il ne fait guère de doute que les singes macaques parviennent à utiliser leur espace de travail pour réfléchir au passé. Ils passent sans difficulté un test de réponse différée, qui exige de retenir une information en mémoire de travail pendant quelques secondes. Leur mémoire à court terme repose, comme la nôtre, sur une augmentation des décharges des neurones du cortex préfrontal et pariétal[30]. En fait, quand ils regardent un film, leur cortex préfrontal s'active *plus* que le nôtre[31]. Notre espèce présente sans doute des capacités supérieures d'inhibition et d'attention sélective qui nous permettent, lorsque nous regardons un film insipide, de détacher nos pensées de l'instant présent et de laisser notre esprit vagabonder librement[32]. Mais les singes aussi possèdent un réseau neuronal « par défaut » qui s'active spontanément lorsque l'animal est au repos, sans avoir de tâche particulière à accomplir[33]. Les régions en question sont exactement les mêmes que celles qui s'activent dans notre cerveau lorsque nous pratiquons l'introspection, que nous imaginons des situations du passé ou que nous laissons nos pensées flotter librement[34].

Que dit notre test décisif de la conscience auditive, le test local-global que nous avons mis au point pour détecter une éventuelle conscience résiduelle chez les patients en état végétatif ? Mes collègues Béchir Jarraya, Lynn Uhrig et moi-même avons testé si les singes parviennent à détecter la nouveauté de la séquence *bip bip bip bip* lorsqu'elle est présentée occasionnellement au milieu de nombreuses séquences *bip bip bip boup*. La réponse est positive : l'IRM fonctionnelle montre clairement qu'un réseau d'aires préfrontales et pariétales s'allume en réponse à la nouveauté globale[35]. Cette réponse disparaît presque totalement lorsque

le singe est sous anesthésie générale. Une fois de plus, le cerveau du singe présente une signature de la conscience.

Des recherches pilotes, menées par le chercheur français Karim Benchenane, suggèrent que même la souris pourrait réussir ce test. Dans les années à venir, à mesure que nous testons systématiquement une grande diversité d'espèces animales, je ne serais pas surpris si nous découvrions que tous les mammifères, et sans doute certaines espèces d'oiseaux et de poissons, présentaient les traces d'une évolution convergente et de l'apparition d'un équivalent fonctionnel de l'espace de travail neuronal global.

Une conscience de soi chez le singe ?

Dans ce livre, je me suis concentré sur l'aspect le plus élémentaire de la conscience : l'accès d'une information sensorielle à une représentation mentale stable qui permet de la manipuler et de la rapporter. Cette compétence est si fondamentale qu'elle existe chez le singe et, sans aucun doute, chez bien d'autres espèces animales. Mais qu'en est-il d'autres acceptions du mot « conscience » ? L'espace de travail humain possède-t-il des propriétés spécifiques qui le distinguent de celui des autres animaux ?

Un préjugé répandu fait de la conscience de soi l'apanage de l'espèce humaine. Ne sommes pas des *sapiens sapiens* – l'espèce qui sait qu'elle sait ? La capacité de réfléchir à notre propre existence n'est-elle pas le propre de l'homme ? Dans son recueil d'entretiens *Intransigeances* (1973), Vladimir Nabokov, brillant romancier doublé d'un entomologiste distingué, l'énonce clairement :

> « Être conscient d'être conscient d'être… Si je sais non seulement que je suis, mais également que je sais que je le sais, alors j'appartiens à l'espèce humaine. Tout le reste en découle – le

> fleuron de la pensée, la poésie, une vision de l'univers. À cet égard, l'écart entre le singe et l'homme est incommensurablement plus vaste que celui qui sépare l'amibe du singe. »

Et pourtant, Nabokov se trompe. « Connais-toi toi-même », le célèbre adage inscrit dans le *pronaos* du temple d'Apollon à Delphes, n'est pas le propre de l'homme. La recherche des dix dernières années a révélé que de nombreux animaux sont doués d'introspection. Même dans des tâches de second ordre, qui exigent de réfléchir à leurs propres erreurs, leurs succès et leurs échecs, leurs connaissances et leurs limites, ils ne sont pas aussi incompétents qu'on le pense.

Ce domaine de compétence s'appelle la « métacognition » – la cognition d'ordre supérieur, ou la représentation mentale de ses propres représentations mentales. Donald Rumsfeld, le secrétaire d'État à la Défense de George W. Bush, en a cerné les contours lorsque, dans un discours resté célèbre, il a distingué trois types de métaconnaissances : « Il y a des choses connues connues : ce sont les choses que nous savons que nous savons. Il y a des inconnues connues, c'est-à-dire des choses que nous savons que nous ignorons. Mais il y a aussi des inconnues inconnues : ce sont les choses que nous ne savons pas que nous ignorons. »

La métacognition est donc la représentation de nos propres connaissances et de leurs limites, l'attribution d'un degré de crédibilité à nos pensées. Et ainsi définie, nous avons de bonnes raisons de penser que les singes, les dauphins, et même les rats et les pigeons, en possèdent les rudiments. Comment savoir si les animaux savent ce qu'ils savent ? Prenez l'exemple de Natua, un dauphin en semi-liberté dans sa piscine de corail au Centre de recherches sur les dauphins de Marathon, en Floride[36]. Natua a été entraîné à classifier des sons sous-marins selon leur hauteur : il appuie sur une manette située sur le mur de gauche de la piscine lorsque le son est grave, et sur une autre, située sur le mur de droite,

lorsqu'il est aigu. La frontière entre l'aigu et le grave a été placée arbitrairement à la fréquence de 2 100 hertz. Lorsque le son est suffisamment loin de cette référence, le dauphin se dirige très vite du bon côté. Lorsque la fréquence du son est proche de 2 100 hertz, par contre, Natua prend son temps. Il secoue la tête avant de nager lentement d'un côté ou de l'autre, en se trompant souvent.

Ces hésitations suffisent-elles à prouver que l'animal « sait » qu'il a du mal à décider ? Non. En soi, l'accroissement de la difficulté lorsque la distance se réduit n'a rien d'exceptionnel. Chez l'homme comme chez de nombreux autres animaux, le temps de décision et le taux d'erreurs augmentent lorsque la distance à discriminer diminue. Par contre, il existe également chez l'homme un sentiment de second ordre : une perte de confiance en soi. Dès que le son est proche de la frontière, nous prenons conscience que la tâche devient difficile. Nous nous rendons compte que nous pourrions nous tromper. Si nous en avons la possibilité, nous refusons de répondre, en disant que nous n'avons aucune idée de la bonne réponse. C'est une métaconnaissance : *je sais que je ne sais pas, ou que je ne suis pas sûr de moi.*

Natua possède-t-il de telles connaissances ? Pourrait-il « dire » qu'il n'est pas certain de connaître la bonne réponse ? Éprouve-t-il un sentiment de confiance dans ses propres décisions ? Pour répondre à ces questions, J. David Smith, de l'Université de l'État de New York, a inventé une astuce expérimentale : la manette d'échappatoire. Une troisième manette est introduite dans le bassin. Petit à petit, Natua découvre que, chaque fois qu'il appuie dessus, le son en cours est remplacé par un son franchement grave (1 200 hertz), donc facile, et qui lui rapporte une récompense plus petite que d'habitude. Ainsi, quand cette manette est présente, l'animal peut échapper à la tâche en cours. Par contre, il ne peut pas la choisir à chaque essai : la manette d'échappatoire doit être utilisée avec modération, sinon l'expérience s'arrête pendant un long moment, et l'obtention de la récompense est donc considérablement retardée.

Nous pouvons à présent juger de l'élégance du résultat qu'a obtenu David Smith : dès qu'il a compris la tâche, le dauphin Natua a décidé, spontanément, de n'appuyer sur la manette d'échappatoire que lors des essais difficiles. Il ne la presse donc que si le son présenté est proche de 2 100 hertz – précisément les essais au cours desquels il risque de se tromper. Tout se passe comme s'il utilisait cette troisième manette comme un « commentaire » de second niveau sur ses performances. En la choisissant, il nous « dit » qu'il lui paraît trop difficile de répondre au son, et qu'il préfère recevoir un essai plus facile, quitte à ce que la récompense soit moindre. Un dauphin est donc suffisamment intelligent pour détecter quand il manque de confiance en lui. Comme Rumsfeld, il sait qu'il ne sait pas.

Certains chercheurs contestent cette interprétation mentaliste. Ils font remarquer que l'on peut décrire le comportement de l'animal en des termes béhavioristes : le dauphin se contente d'exécuter un comportement moteur auquel il a été entraîné et qui maximise sa récompense moyenne. Des milliers d'essais d'apprentissage par renforcement ont permis à l'animal de découvrir quels sons prédisent les moments où il est plus avantageux d'appuyer sur la troisième manette – rien de plus qu'un apprentissage conditionné.

Bien que cette critique s'applique à bon nombre d'expériences passées, de nouvelles données chez le singe, le rat et le pigeon la réfutent et font pencher la balance du côté d'une authentique compétence métacognitive. En effet, les animaux optent pour la touche d'échappatoire plus intelligemment que ne le prédirait un simple conditionnement sensori-moteur[37]. Ainsi, lorsqu'on introduit la troisième touche *après* que l'animal a fait son choix, mais *avant* qu'on lui ait dit s'il était juste ou faux, les animaux continuent de savoir quels essais sont difficiles pour eux. On le voit parce que leurs performances sont effectivement pires quand ils appuient sur la touche d'échappatoire que quand ils restent fidèles à leur choix de départ, y compris quand le stimulus est exactement le même dans les deux cas. Ils évaluent donc bien leur propre état mental et

parviennent à rejeter uniquement les essais où, pour une raison ou pour une autre, ils étaient distraits et n'ont pas perçu le stimulus avec la même acuité que les autres. On dirait vraiment qu'ils s'interrogent sur leur degré de confiance en eux et n'appuient sur la touche d'échappatoire que s'ils se sentent peu sûrs d'eux[38].

À quel point cette conscience de soi est-elle abstraite ? Chez le singe, une expérience récente montre que la compétence n'est pas liée à une tâche précise : les macaques généralisent spontanément l'utilisation de la touche d'échappatoire à des situations nouvelles. Une fois qu'ils ont appris ce qu'elle signifie dans le contexte d'une tâche sensorielle, ils l'utilisent immédiatement à bon escient dans un autre contexte, celui d'une tâche de mémoire. Dès qu'ils ont appris à « dire » « je n'ai pas bien perçu », ils généralisent spontanément à « je ne me souviens pas bien[39] ».

Ces animaux possèdent donc un certain degré de connaissance de soi, mais se pourrait-il que tout cela soit inconscient ? Souvenez-vous qu'au chapitre 2 nous avons vu que la plupart de nos comportements sont issus de mécanismes inconscients. Même l'autoévaluation pourrait être inconsciente. Lorsque je frappe sur la mauvaise touche sur mon clavier, ou que mes yeux sont distraits de leur cible, mon cerveau détecte automatiquement ces erreurs et les corrige parfois sans que j'en aie la moindre conscience[40]. Cependant, plusieurs arguments suggèrent que la métacognition des singes ne reflète pas simplement ce genre de mécanisme subliminal. Les animaux utilisent la touche d'échappatoire d'une façon flexible, intelligente, qui généralise à des situations nouvelles – pas du tout dans le style d'un traitement inconscient. Leur décision prend parfois plusieurs secondes, durée qui dépasse l'empan temporel typique des activations inconscientes. Elle implique l'utilisation d'une réponse arbitraire, la troisième manette. Au niveau neurophysiologique, elle implique une lente accumulation de données dans les régions préfrontales et pariétales[41]. En extrapolant ce que nous savons du cerveau humain, il ne semble guère probable qu'un comportement aussi lent et aussi complexe, d'ordre supérieur, puisse se dérouler sans conscience.

Si cette conclusion s'avère juste (et il faudra certainement d'autres travaux pour la valider), alors le comportement animal porte le sceau d'une pensée consciente et réflexive. Nous ne sommes pas les seuls à savoir que nous savons, et l'épithète *sapiens sapiens* ne devrait pas être l'apanage du genre *Homo*. D'autres espèces que la nôtre possèdent une capacité d'introspection et une certaine forme de conscience de soi.

Le propre de l'homme ?

Bien que les singes possèdent un espace de travail global et s'en servent pour réfléchir à eux-mêmes ainsi qu'à leur environnement, il est probable que l'introspection soit une faculté particulièrement développée dans l'espèce humaine. Qu'est-ce qui distingue notre espèce : la taille du cerveau ? Le langage articulé ? Une plasticité cérébrale supérieure ? La faculté de modifier le cerveau par l'éducation ?

Répondre à ces questions est l'une des missions les plus excitantes des neurosciences cognitives. Je ne prétends pas détenir la réponse, mais seulement une hypothèse : bien que nous partagions la plupart, sinon la totalité de nos processeurs cérébraux avec d'autres espèces animales, le cerveau humain pourrait bien s'avérer unique dans sa capacité de *combiner* ces « modules » dans des représentations composites, pour former un véritable « langage de la pensée ». René Descartes avait raison sur un point : seul l'*Homo sapiens* peut « user de paroles ou d'autres signes *en les composant*, comme nous faisons pour déclarer aux autres nos pensées ». Cette capacité combinatoire pourrait bien être l'ingrédient crucial qui démultiplie nos facultés mentales. L'unicité de l'espèce humaine est peut-être à rechercher dans la façon particulière que nous avons de formuler nos idées à l'aide de structures symboliques enchâssées les unes dans les autres. C'est la *récursivité* – la propriété qu'ont nos pensées de s'emboîter comme des poupées russes, en sorte que chaque pensée, même

complexe, peut elle-même devenir le point de départ d'une autre pensée de niveau supérieur : *partir, partir vite, vouloir partir vite, ne pas vouloir partir vite, prétendre ne pas vouloir partir vite, etc.*

Selon cet argument, et en accord avec les idées de Noam Chomsky, le langage aurait évolué en tant que système interne de représentation mentale, avant de jouer un rôle dans la communication avec les autres. L'avantage principal qu'il confère est la possibilité de *penser* des idées nouvelles (qui ensuite peuvent bien sûr être expliquées à d'autres). Notre espèce semble posséder un don particulier pour les symboles : nous savons les allouer à n'importe quelle représentation mentale et les agencer selon des combinaisons entièrement nouvelles. L'espace de travail neuronal global de l'espèce humaine pourrait être unique, dans le monde animal, en raison de sa capacité à formuler des pensées complexes, formées de façon récursive, telles que « plus grand que Tom », « à gauche de la porte verte » ou « ne pas donner à Jean ». Chacun de ces exemples combine plusieurs concepts élémentaires qui appartiennent à des domaines complètement différents : la taille (grand), la personne (Tom, Jean), l'espace (à gauche), la couleur (vert), l'objet (la porte), la logique (pas) ou l'action (donner). Bien que chacun d'eux soit codé, au moins au départ, dans des aires cérébrales distinctes, le cerveau humain parvient à les assembler à volonté – et non seulement à les associer entre eux, comme le font indubitablement d'autres animaux, mais à les *composer* à l'aide d'une syntaxe qui distingue, par exemple, « la mère de mon amie » de « l'amie de ma mère ».

De nombreuses compétences propres à l'espèce humaine semblent reposer sur ce système de composition mentale, depuis l'invention d'outils complexes jusqu'à la création des mathématiques. Dans le domaine de la conscience, il pourrait expliquer le développement remarquable de la conscience de soi. Les humains possèdent un sens incroyablement raffiné des choses de l'esprit – ce que les psychologues appellent une « théorie de l'esprit », un vaste ensemble de règles intuitives qui nous permettent d'imaginer ce que pensent les autres et d'en tirer des conséquences

logiques. Toutes les langues humaines possèdent un vaste éventail de mots pour parler de l'esprit. Parmi les dix verbes les plus fréquents de la langue française, six font référence à des opérations mentales ou à des états mentaux (*dire, pouvoir, vouloir, savoir, voir, devoir*). Nous parlons des autres aussi bien que de nous-mêmes, en employant des constructions presque identiques à base de pronoms (« *je* » est le mot le plus fréquent du français parlé, « *tu* » le sixième, et « *il* » le dixième). Le langage de la pensée nous permet de formuler ce que nous savons dans le même format que ce que les autres savent (« je sais que X, mais tu crois que Y »). Cette capacité de « mentalisation » est présente très précocement : même un bébé de 7 mois semble déjà confronter ses connaissances à celles des autres[42].

Réfléchir aux pensées, qu'il s'agisse des siennes ou de celles des autres, est vraiment la marque de fabrique de notre espèce. Les compétences d'un enfant de deux ans et demi pour le raisonnement social surpassent déjà celles d'un chimpanzé adulte ou de n'importe quel autre primate[43]. Une particularité essentielle de ce mécanisme est sa récursivité : nos pensées s'enchâssent les unes dans les autres à l'infini, comme lorsque *je crois qu'il pense que je ne sais pas qu'il ment*. De telles pensées possèdent une syntaxe analogue à celle du langage, qui leur permet d'exprimer avec une extraordinaire compacité des états complexes du monde physique et mental. Ces représentations semblent hors de portée des autres primates[44]. Leur métacognition n'implique, au mieux, que deux niveaux de représentation : une pensée et un degré de croyance[45]. Un singe peut, par exemple, penser : « La pomme est là, j'en suis sûr », ou bien : « La pomme est là, mais le mâle dominant l'ignore », mais pas : « La pomme est là, mais il pense que c'est une pierre », encore moins : « Il ne sait pas que je sais où est la pomme. » Seul le cerveau humain semble être capable de manipuler mentalement des croyances qu'il sait être fausses et une infinité d'autres pensées enchâssées les unes dans les autres.

Unique parmi les primates, l'espace de travail neuronal humain pourrait présenter des adaptations qui le rendent apte à la manipulation

de pensées et de croyances complexes. Les rares données disponibles vont dans ce sens. Comme nous l'avons vu au chapitre 5, le cortex préfrontal, ce pivot de l'espace conscient, occupe une place importante dans le cerveau de tous les primates – mais chez l'homme, il a connu une nouvelle expansion inédite[46]. Parmi tous les primates, les neurones du cortex préfrontal humain sont ceux qui possèdent les plus grands arbres dendritiques[47]. En augmentant les capacités de connexion, cette particularité anatomique pourrait faciliter le recueil et l'intégration des informations en provenance de processeurs cérébraux très distants, ce qui expliquerait notre remarquable capacité d'introspection et de réflexion autonome, détachée du monde extérieur.

Certaines régions, dans la partie la plus antérieure du lobe frontal, s'activent systématiquement lorsque nous déployons nos talents pour le raisonnement social et l'introspection[48]. L'une d'elles s'appelle le cortex fronto-polaire, ou encore l'aire 10 de Brodmann. Or elle est plus vaste chez *Homo sapiens* que chez n'importe quel autre grand singe (les experts débattent encore de son existence chez le singe macaque). La substance blanche sous-jacente, qui regroupe les faisceaux de connexion à longue distance, est également disproportionnellement plus grande dans l'espèce humaine, même si l'on tient compte des différences de taille du cerveau[49]. Toutes ces observations font du cortex préfrontal antérieur un candidat de choix pour expliquer nos capacités d'introspection.

Une autre région singulière est l'aire de Broca, cette partie du cortex frontal inférieur gauche qui intervient dans le langage. Les neurones de la couche 3, qui envoient des projections à longue distance, y sont bien plus espacés chez l'homme que chez les autres grands singes – là encore, c'est une adaptation qui facilite les interconnexions[50]. Dans cette aire, ainsi que dans le cortex cingulaire antérieur, une région essentielle pour le contrôle de soi, Constantin von Economo a découvert des neurones géants qui, selon certains, seraient spécifiques au cerveau de l'homme et des grands singes comme les chimpanzés et les bonobos, et tout simplement absents chez le macaque ou chez toute autre espèce de primate[51]. Avec leur corps

cellulaire géant et leur axone long, ces cellules contribuent sans doute de façon importante à la diffusion globale des messages conscients dans le cerveau humain.

Toutes ces adaptations relèvent de la même tendance évolutive. Au cours de l'hominisation, les réseaux du cortex préfrontal se sont densifiés au-delà de ce que prédit un accroissement uniforme du cerveau. Les circuits de l'espace de travail ont connu une expansion disproportionnée. Cependant, la taille n'est qu'un paramètre grossier. Nous ne sommes pas seulement des primates avec un gros cerveau ! Je ne serais pas surpris si, dans les années à venir, les neuroscientifiques découvraient l'existence de microcircuits propres au cerveau humain, qui lui donnent accès aux opérations récursives. Nos cousins les primates possèdent certainement une vie mentale et la capacité de prendre conscience de leur environnement, mais notre vie intérieure est immensément plus riche, en raison de notre faculté de composer nos pensées.

En résumé, si cette hypothèse est correcte, la conscience humaine est le résultat de deux évolutions successives. Elle a initialement évolué pour répondre aux nécessités de la communication interne : le cortex préfrontal et ses connexions à longue distance brisent la modularité des circuits neuronaux locaux et diffusent l'information à l'ensemble du cerveau. Cette fonction existe chez tous les primates. Une seconde évolution, propre à l'espèce humaine, aurait ensuite fait émerger un « langage de la pensée » qui nous permet de formuler des pensées complexes et de les partager avec d'autres.

Des maladies de la conscience ?

Les deux évolutions successives de la conscience humaine doivent reposer sur des mécanismes biologiques et génétiques qui leur sont propres. Dès lors, une question se pose : certaines maladies pourraient-

elles sélectivement affecter la machinerie de la conscience ? Certaines anomalies génétiques ou certaines lésions cérébrales pourraient-elles inverser la tendance évolutive et causer des déficits sélectifs de l'espace de travail neuronal global ?

Il est probable que les connexions à longue distance sur lesquelles repose la conscience soient fragiles. Comparés à d'autres cellules du corps, les neurones sont des cellules énormes et difformes, dont l'axone peut atteindre plusieurs dizaines de centimètres. Assurer la maintenance d'un appendice aussi gigantesque, dont la taille dépasse d'un facteur mille celle du corps cellulaire, pose des problèmes particuliers d'expression des gènes et de transport des molécules. En effet, la transcription et la traduction de l'ADN se produisent au niveau du noyau de la cellule, et pourtant il faut bien que les protéines qui en résultent soient transportées jusqu'aux synapses qui se trouvent parfois à des distances considérables. Ce problème de logistique mobilise de nombreux systèmes biologiques. Il n'est donc pas déraisonnable de penser que le système de connexions à longue distance de l'espace de travail est la cible de maladies spécifiques.

Jean-Pierre Changeux et moi-même faisons l'hypothèse qu'un déficit de ce type pourrait peut-être rendre compte de ce mystérieux amalgame de symptômes psychiatriques qu'on appelle « schizophrénie[52] ». La schizophrénie est une maladie fréquente, qui touche environ 0,7 % de la population adulte. C'est une pathologie mentale dévastatrice, dans laquelle certains adolescents ou de jeunes adultes finissent par perdre le contact avec la réalité. Ils ont des idées délirantes et des hallucinations (ce qu'on appelle les « symptômes positifs »). Dans le même temps, leurs capacités intellectuelles et émotionnelles s'émoussent, et ils présentent souvent une désorganisation de la pensée et de la parole ainsi que des comportements répétitifs (les « symptômes négatifs »).

Les psychiatres continuent de se demander s'il existe un principe général qui sous-tend toutes ces manifestations de la schizophrénie. Il est frappant de constater qu'elles concernent presque toujours des fonctions associées, dans l'espèce humaine, à l'espace de travail global : la théorie

de l'esprit, la formation des croyances, l'introspection et la métacognition, mais aussi l'accès conscient à des informations perceptives élémentaires[53].

Sur le plan clinique, les patients souffrant de schizophrénie présentent un étonnant excès de confiance dans leurs idées délirantes. Leur métacognition et leur théorie de l'esprit sont parfois tellement affectées qu'ils ne parviennent plus à distinguer leurs propres pensées, leurs connaissances, leurs souvenirs et même leurs actions de celles des autres. La schizophrénie altère radicalement l'intégration consciente de la pensée en un tout cohérent. Elle perturbe par exemple l'accès à la mémoire consciente : quelques minutes à peine après avoir vu une liste de mots ou d'images, les patients s'en souviennent très mal. Leur pathologie affecte tout particulièrement les connaissances métacognitives : où, quand, comment ai-je appris tel ou tel fait ? En suis-je certain ? Il est remarquable qu'en dépit de ce déficit majeur leur mémoire implicite et inconsciente reste intacte[54].

Mes collègues et moi nous sommes demandé s'il n'existerait pas, dans la schizophrénie, un déficit fondamental d'accès à la conscience, quelle que soit l'origine de l'information. Effectivement, nous avons découvert que la prise de conscience d'une image masquée est perturbée, alors que les opérations subliminales sont préservées. Lorsqu'un mot est suivi immédiatement par un masque, et que sa durée est suffisamment courte, il disparaît de notre conscience et devient « subliminal ». Or la durée minimale de présentation en deçà de laquelle un mot devient invisible est fortement altérée dans la schizophrénie[55]. Le seuil d'accès à la conscience est trop élevé : là où un sujet sain voit parfaitement le mot, les patients souffrant de schizophrénie disent ne rien voir, et il leur faut une quantité de données sensorielles nettement plus grande avant qu'ils n'éprouvent un sentiment subjectif de vision consciente. Pourtant, leurs capacités de traitement inconscient sont intactes. Un chiffre invisible, présenté pendant une durée de 29 millisecondes seulement, continue de causer un amorçage subliminal, exactement comme dans le cerveau normal. Le fait qu'une mesure aussi subtile soit préservée suggère que la schizophrénie

épargne la cascade de traitements corticaux qui s'enchaînent « de bas en haut », depuis les aires visuelles jusqu'à l'attribution d'un sens. Le problème semble se situer au niveau des opérations « de haut en bas », qui permettent l'intégration globale et consciente des informations en un tout cohérent. Les opérations subliminales sont normales, tandis que l'accès à la conscience présente des anomalies évidentes.

Mes collègues et moi avons été fascinés de retrouver cette même dissociation dans une tout autre maladie : la sclérose en plaques, une altération sélective des connexions qui forment la substance blanche du cerveau[56]. Au tout début de la maladie, avant qu'on observe d'autres déficits majeurs, les patients présentent déjà une altération de la prise de conscience de mots et de chiffres masqués, alors que leur cerveau continue de les traiter inconsciemment. D'un patient à l'autre, la gravité du déficit d'accès à la conscience varie en fonction de la sévérité des anomalies des faisceaux de connexions à longue distance qui relient le cortex préfrontal aux régions postérieures du cortex visuel[57]. Ce résultat est essentiel, d'abord parce qu'il confirme qu'une maladie des faisceaux de la substance blanche peut sélectivement altérer l'accès conscient, et ensuite parce qu'une petite proportion de patients atteints de sclérose en plaques présente des troubles psychiatriques apparentés à la schizophrénie, ce qui suggère que la perte de connexions à longue distance peut effectivement jouer un rôle important dans cette maladie mentale.

L'imagerie cérébrale de patients qui souffrent de schizophrénie prouve également que leur cerveau a perdu sa capacité d'embrasement conscient. Les traitements visuels et attentionnels précoces sont largement intacts, tandis qu'on constate une réduction forte de l'activation tardive et synchronisée qui engendre l'onde P3 à la surface du crâne et signale l'accès à la conscience[58]. Une autre signature de la conscience, l'apparition soudaine d'un « réseau cérébral global » formé par la synchronisation de régions corticales distantes, dans la bande de fréquences « bêta » (entre 13 et 30 hertz), présente également de très fortes anomalies[59].

Sur le plan anatomique, il existe des preuves encore plus directes d'une altération des réseaux de l'espace de travail neuronal global dans la schizophrénie. L'IRM de diffusion révèle des anomalies majeures des faisceaux constitués d'axones longs qui interconnectent les aires corticales les plus distantes. Les fibres du corps calleux, qui relient les deux hémisphères, sont particulièrement affectées, tout autant que les connexions du cortex préfrontal avec les autres régions du cortex, de l'hippocampe et du thalamus[60]. Il en résulte de graves perturbations de l'échange d'informations entre les aires cérébrales, même au repos. L'examen de l'activité spontanée du cerveau montre que le cortex préfrontal perd son statut privilégié de pivot ou de centre névralgique autour duquel gravite le reste du réseau cortical ; l'activité cérébrale se désorganise et cesse de former un tout intégré[61].

Au niveau microscopique, les neurones pyramidaux géants que l'on trouve dans les couches 2 et 3 du cortex préfrontal dorso-latéral, avec leurs dendrites immenses capables de recevoir des milliers de connexions synaptiques, voient leur taille considérablement réduite chez les patients atteints de schizophrénie. Ils ont moins d'épines dendritiques, ces petites échardes qui tapissent les dendrites, qui servent de réceptacle aux synapses excitatrices et qui sont exceptionnellement denses dans le cerveau humain. Cette diminution des connexions semble bien jouer un rôle décisif dans la maladie. En effet, bon nombre des gènes qui sont perturbés dans la schizophrénie interviennent dans deux grandes voies de transmission nerveuse, qui reposent respectivement sur le récepteur D2 de la dopamine et sur le récepteur NMDA du glutamate, et qui jouent un rôle essentiel dans la transmission synaptique et la plasticité au sein du cortex préfrontal[62].

Encore plus intéressant peut-être, même un adulte qui ne souffre d'aucune pathologie peut éprouver une psychose temporaire proche de la schizophrénie s'il prend des drogues telles que la kétamine ou la phéncyclidine (plus connue sous le nom de PCP ou, en anglais, d'*angel dust*). Ces molécules bloquent la transmission entre les neurones,

spécifiquement au niveau des synapses NMDA du glutamate, dont on sait qu'elles sont particulièrement impliquées dans les connexions à longue distance entre les aires du cortex[63]. Dans mes simulations informatiques de l'espace de travail neuronal global, les synapses NMDA jouent un rôle fondamental dans l'embrasement conscient : ce sont elles qui forment les boucles « descendantes » permettant aux aires de haut niveau de réactiver, en retour, les aires sensorielles qui les sollicitent. Retirez les récepteurs NMDA de la simulation, et l'embrasement global disparaît[64]. D'autres simulations montrent que le récepteur NMDA est essentiel à l'accumulation progressive de statistiques qui conduit à la prise d'une décision réfléchie[65].

Une réduction globale de la connectivité « descendante », depuis les aires supérieures en direction des processeurs de plus bas niveau, pourrait donc expliquer une bonne part des symptômes négatifs de la schizophrénie. Elle laisserait intacte la transmission « montante » des informations en provenance des organes des sens, mais elle affecterait sélectivement leur diffusion et leur intégration au sein d'un réseau global. Cela rendrait parfaitement compte de la préservation des opérations subliminales précoces dans la schizophrénie. Le déficit ne se manifesterait qu'au niveau de l'embrasement conscient, où il perturberait la diffusion globale des messages neuronaux et affecterait l'ensemble des capacités de supervision consciente, de mémoire de travail, de prise de décision et d'allocation volontaire de l'attention.

Mais que dire des symptômes positifs, les étranges hallucinations et les idées délirantes qui caractérisent la schizophrénie ? Deux spécialistes de neurosciences cognitives, Paul Fletcher et Chris Frith, en ont proposé une explication précise, toujours fondée sur l'hypothèse d'une propagation anormale des signaux au sein des circuits du cortex[66]. Comme nous l'avons vu au chapitre 2, le cerveau raisonne à la manière de Sherlock Holmes : comme le célèbre détective, il utilise la logique et le calcul des probabilités afin de tirer un maximum de conclusions des entrées qu'il reçoit, que celles-ci soient perceptives ou sociales. Or l'inférence

statistique, dans un réseau de neurones, exige un échange bidirectionnel d'informations[67] : les aires sensorielles envoient les données qu'elles reçoivent vers le haut de la hiérarchie corticale, et les aires supérieures, en retour, leur répondent avec des prédictions qui tentent d'anticiper sur la suite. Tous ces échanges participent d'un algorithme d'apprentissage qui vise à expliquer, le mieux possible, les informations reçues par les organes des sens. L'apprentissage ne cesse que lorsque les représentations mentales de haut niveau sont si précises qu'elles rendent compte de tous les détails des entrées sensorielles. À ce stade, le « signal d'erreur », c'est-à-dire la différence entre ce que le cerveau prédit et ce qu'il reçoit, devient négligeable. En conséquence, toute surprise disparaît. Le signal d'entrée perd son intérêt et ne déclenche plus d'apprentissage.

Imaginez à présent que, dans la schizophrénie, les signaux neuronaux descendants, ceux qui transmettent les prédictions, soient réduits – soit parce que certaines connexions manquent à l'appel, soit parce que les récepteurs NMDA fonctionnent mal. Selon Fletcher et Frith, cela aurait un effet dramatique sur les mécanismes de raisonnement statistique du cerveau. Les entrées sensorielles ne seraient jamais expliquées de façon satisfaisante. Il resterait toujours des signaux d'erreur, qui déclencheraient sans cesse de nouvelles interprétations. Les patients auraient systématiquement l'impression paranoïaque que le sens profond des choses leur échappe, qu'une infinité de causes se cachent derrière les phénomènes les plus triviaux de la vie quotidienne dont eux seuls perçoivent la vraie complexité. En conséquence, ils ne cesseraient jamais de concocter des explications de plus en plus délirantes de ce qu'ils ressentent.

Examinons, par exemple, comment un cerveau schizophrène interpréterait ses propres actions. Normalement, lorsque nous faisons un geste, un mécanisme prédictif en anticipe les conséquences et en atténue les effets. Grâce à lui, nous n'éprouvons aucune surprise lorsque nous saisissons, pour la cent millième fois peut-être, notre tasse de café : sa position, son poids et même son toucher brûlant sont prévisibles, et avant

même que nous ayons commencé à agir, nos aires motrices envoient des prédictions aux aires sensorielles afin de les informer qu'elles ne vont pas tarder à éprouver chacune de ces sensations. Cette anticipation fonctionne si bien que nous n'avons généralement pas conscience des conséquences sensorielles de nos gestes. Ce n'est que lorsque notre prédiction s'avère fausse, par exemple parce que la tasse nous glisse des mains, que nous prenons soudain conscience de la situation.

À présent, imaginez que, parce que votre cerveau souffre de schizophrénie, vous viviez dans un monde où ces prédictions restent systématiquement en deçà de la réalité. Même votre tasse de café ne paraît pas normale. Lorsque vous la prenez en main, son toucher, son poids et sa température ne ressemblent pas à ce que vous attendiez et vous mettent en alerte : quelque chose ou quelqu'un s'amuse à vous jouer des tours. Même le fait de parler vous semble bizarre : vous entendez votre propre voix, et elle n'est jamais conforme à vos attentes. Ces bizarreries ne cessent d'attirer votre attention et vous empêchent de vous concentrer. Vous commencez à penser que quelqu'un vous espionne, vous manipule et vous empoisonne l'existence au point de modifier vos propres paroles. Il n'y a qu'un petit pas à franchir pour conclure que vous entendez des voix qui cherchent à contrôler votre vie. Vous êtes sans cesse en quête d'une explication rationnelle de tous ces événements mystérieux qui perturbent votre existence et que personne d'autre que vous ne semble remarquer. Voilà un tableau plutôt juste des symptômes positifs de la schizophrénie.

En résumé, s'il existe une maladie des réseaux de connexions à longue distance qui constituent l'espace de travail conscient dans le cerveau humain, la schizophrénie pourrait bien en être l'une des manifestations les plus évidentes. Il ne s'agit évidemment pas de dire que les patients qui souffrent de schizophrénie sont des zombies dépourvus de toute conscience subjective. Constatons seulement que, dans cette maladie, les mécanismes de la conscience sont bien plus altérés que les autres processus mentaux automatiques et inconscients. Les maladies

neurologiques tendent à respecter les subdivisions du système nerveux, et la schizophrénie pourrait concerner plus particulièrement les mécanismes biologiques responsables de la maintenance des grands faisceaux de connexions corticales, particulièrement ceux qui transmettent les informations du haut de la hiérarchie vers le bas.

Dans la schizophrénie, cette atteinte n'est pas totale, sinon le patient perdrait tout simplement conscience. Une pathologie aussi sévère existe-t-elle ? En 2007, des neurologues de l'Université de Pennsylvanie ont découvert une stupéfiante maladie jusqu'alors inconnue : l'encéphalite à anticorps ciblés contre les récepteurs NMDA[68]. Les patients présentaient initialement des symptômes divers qui semblaient bénins. Beaucoup étaient des jeunes femmes atteintes d'un cancer des ovaires, mais d'autres se plaignaient simplement de maux de tête, de fièvre ou d'un syndrome grippal. Soudain, leur maladie prenait un tour dramatique. Ils commençaient par présenter « des symptômes psychiatriques prononcés : anxiété, agitation, comportements bizarres, pensées délirantes et paranoïdes, hallucinations visuelles et auditives » – autrement dit une crise de schizophrénie aiguë, acquise et d'évolution rapide. Trois semaines plus tard, leur conscience se mettait à décliner. Leur EEG montrait des ondes cérébrales lentes, comparables à celle d'une personne endormie ou dans le coma. Ils ne bougeaient pratiquement plus, cessaient de répondre aux commandes verbales, et parfois même de respirer. Plusieurs d'entre eux moururent en quelques mois. D'autres recouvrèrent lentement une santé et une vie normales, y compris sur le plan psychiatrique ; ils affirmaient alors n'avoir aucun souvenir de leur épisode de perte de conscience.

Que s'était-il passé ? L'enquête montra que tous ces patients souffraient d'une grave maladie auto-immune. Leur système immunitaire, au lieu d'attaquer les intrus habituels que sont les virus et les bactéries, se retournait contre le corps du patient lui-même. Il y détruisait avec un soin diabolique une molécule bien précise : le récepteur NMDA du glutamate. Comme nous venons de le voir, cet ingrédient fondamental de la synapse

joue un rôle essentiel dans la transmission des informations corticales descendantes, du haut de la hiérarchie des aires cérébrales vers le bas. Quand on exposait, dans une boîte de Petri, une culture de cellules nerveuses au sérum physiologique des patients, toutes les synapses NMDA s'évanouissaient littéralement en quelques heures – mais le récepteur réapparaissait dès qu'on cessait d'injecter ce liquide empoisonné.

Il est fascinant que l'éradication d'une seule molécule suffise à engendrer la totalité des symptômes psychiatriques de la schizophrénie, suivie d'une perte de conscience. Peut-être vient-on de découvrir la première pathologie spécifique des réseaux qui sous-tendent l'expérience consciente – et il est frappant que cette maladie commence par une psychose proche de la schizophrénie. Les deux pathologies semblent donc étroitement liées. À l'avenir, l'encéphalite à anticorps ciblés contre les récepteurs NMDA pourrait servir de maladie modèle dont les mécanismes moléculaires, bien délimités, éclairent l'origine des psychoses et les relient aux mécanismes de la conscience.

Des machines conscientes ?

Dès lors que nous commençons à comprendre la fonction de la conscience, son architecture corticale, ses bases moléculaires et même ses pathologies, serait-il envisageable de la simuler sur ordinateur ? Et cette machine deviendrait-elle consciente ? Non seulement cette éventualité n'a rien d'impossible, mais elle constitue un passionnant programme de recherche pour les années à venir – un grand défi pour les sciences de l'informatique. Il va sans dire que nous sommes encore loin de savoir comment construire une telle machine, mais le simple fait que nous commencions à énumérer les propriétés qu'elle devrait posséder montre à quel point la science de la conscience a progressé ces dernières années.

Au chapitre 5, j'ai esquissé les contours d'une simulation informatique de l'accès à la conscience. Cette simulation fournit les premiers ingrédients d'une nouvelle architecture logicielle. Comme la plupart des ordinateurs actuels, une machine consciente devrait contenir de nombreux programmes spécialisés qui tournent en parallèle, un peu comme les « applications » de votre téléphone portable. Elle posséderait donc toute une panoplie de systèmes de reconnaissance des visages, de détection du mouvement, de cartographie spatiale, de production et de compréhension de la parole ou de commande motrice. Il faudrait que certains au moins de ces programmes reçoivent leurs entrées et leurs sorties non pas du monde extérieur, mais de l'intérieur, c'est-à-dire d'autres programmes. Ainsi le système aurait-il accès à une certaine forme d'introspection ou de connaissance de soi. Il y aurait, par exemple, un programme spécialisé dans la détection des erreurs, dont le rôle serait de prédire à quel moment le système dévie de son but actuel. Les ordinateurs d'aujourd'hui possèdent déjà les rudiments de cette idée, puisque leur système d'exploitation comprend une batterie de logiciels d'évaluation de l'intégrité de la mémoire, de l'espace disque, de la capacité de la batterie, ou de l'existence de conflits internes.

Cependant, le modèle de l'espace de travail pointe vers trois fonctions essentielles dont les machines contemporaines sont dépourvues : la communication flexible, la plasticité et l'autonomie. Dans une machine consciente, il faudrait avant tout que les programmes puissent communiquer entre eux avec une grande flexibilité. À chaque instant, le système choisirait la sortie de l'un des programmes spécialisés et en ferait le centre d'intérêt de tout l'organisme. L'information ainsi sélectionnée entrerait dans un espace spécifique, un système à capacité limitée, incapable de traiter plus d'une information à la fois, mais qui aurait l'immense avantage de pouvoir la diffuser à n'importe quel autre programme. Les ordinateurs actuels interdisent généralement ce type d'échanges : chaque application s'exécute dans une zone de mémoire qui lui est propre, et ses sorties s'affichent dans une fenêtre dédiée. Les

programmes ne disposent d'aucune possibilité de s'informer l'un l'autre. Seule exception : le presse-papiers, qui permet un copier-coller d'une application à l'autre – mais son fonctionnement reste rudimentaire et sous le contrôle de l'utilisateur plutôt que du système lui-même. L'architecture envisagée permettrait d'augmenter massivement la flexibilité des échanges d'information, en fournissant une sorte de presse-papiers universel et autonome : l'espace de travail global.

Comment les programmes utiliseraient-ils les informations issues de cet espace global ? Le deuxième ingrédient est un puissant algorithme d'apprentissage. Les applications cesseraient ainsi d'être des programmes figés. Chacune serait dotée d'une capacité de découvrir comment utiliser au mieux les informations qu'elle reçoit. Chaque morceau de logiciel s'améliorerait constamment, à l'aide d'une règle d'apprentissage inspirée de la biologie et capable de discerner les nombreuses relations statistiques qui gouvernent les informations en circulation. Ainsi le système s'ajusterait-il à son environnement ainsi qu'aux particularités de sa propre architecture. Si une panne affectait l'un de ses composants, par exemple, il ne cesserait pas de fonctionner, mais réapprendrait à utiliser au mieux les éléments restants. Il découvrirait de lui-même lesquelles de ses entrées sont les plus pertinentes et méritent d'entrer dans l'espace de travail global, et comment les combiner pour réaliser des calculs utiles à l'organisme.

Cet aspect nous conduit au troisième ingrédient essentiel : l'autonomie. Même en l'absence de toute interaction avec un utilisateur, l'ordinateur pourrait utiliser sa propre échelle de valeurs afin de décider quelles données, présentes ou passées, méritent un examen approfondi au sein de l'espace de travail. Cette décision pourrait être probabiliste : le système laisserait alors des « pensées » plus ou moins aléatoires entrer dans l'espace de travail. Elles y seraient maintenues ou rejetées en fonction de leur adéquation aux objectifs actuels de l'organisme. Ainsi, même en l'absence de données nouvelles, un flux partiellement stochastique de représentations internes se succéderait en mémoire.

Le comportement de cet organisme simulé évoquerait initialement celui d'un enfant auquel nous attribuerions spontanément un certain degré de conscience. Sans aucune intervention extérieure, il se fixerait ses propres objectifs, explorerait son environnement et découvrirait progressivement des régularités dans ses entrées ainsi que dans son organisation interne. À chaque instant, des dizaines de programmes se dérouleraient en parallèle, mais l'espace de travail ne se consacrerait qu'à une seule représentation, qui correspondrait au contenu de la conscience.

Ces idées restent vagues, et il faudra beaucoup d'efforts pour les concrétiser. Mais, au moins en principe, aucune raison fondamentale ne s'oppose à l'idée d'une conscience artificielle. Pourtant, de nombreux penseurs contestent cette conclusion. Examinons brièvement leurs arguments. Certains estiment que la conscience ne saurait se réduire au seul traitement de l'information. Manipulez l'information autant que vous le souhaitez, disent-ils, jamais il n'en sortira une expérience consciente subjective. Le philosophe Ned Block, de l'Université de New York, concède que le mécanisme de l'espace de travail global pourrait expliquer l'accès à la conscience, mais il l'estime fondamentalement incapable d'élucider ce que sont les *qualia* – les quanta d'expérience subjective, les états d'âme à l'état brut, « ce que cela fait » de ressentir une rage de dents ou la beauté d'un coucher de soleil[69].

David Chalmers, philosophe à l'Université d'Arizona, juge que, même si la théorie de l'espace de travail parvenait à rendre compte des opérations conscientes et inconscientes, jamais elle n'expliquerait l'énigme de la subjectivité : le ressenti intime que nous éprouvons au plus profond de notre « moi[70] ». Chalmers est célèbre pour sa distinction, dans le domaine de la conscience, entre ce qu'il appelle les problèmes faciles et *le* problème difficile. Les problèmes faciles, selon lui, consistent à expliquer les nombreuses fonctions du cerveau : comment reconnaît-il un visage, un mot ou un paysage ? Comment extrait-il des informations des organes des sens et les utilise-t-il pour guider nos gestes ? Comment produisons-nous des paroles pour décrire ce que nous ressentons ? « Bien que toutes ces

questions soient reliées à la conscience, écrit Chalmers, elles portent toutes sur les mécanismes objectifs du système cognitif, et par conséquent, nous avons toutes les raisons de penser que la poursuite des efforts en psychologie cognitive et en neurosciences finira par y répondre[71]. »

Le problème difficile, lui, est d'une tout autre nature. Il s'agit de « savoir comment les processus physiques du cerveau engendrent l'expérience subjective, le ressenti propre à chaque personne. Lorsque nous voyons, par exemple, nous ressentons des sensations visuelles telles qu'un bleu intense. Ou bien pensez au son ineffable d'un hautbois dans le lointain, au calvaire d'une douleur aiguë, à la chaleur d'un instant de bonheur, ou à la qualité méditative d'un moment où nous nous perdons dans nos pensées. […] Ce sont ces phénomènes-là qui constituent le réel mystère de l'esprit ».

Mon sentiment est que Chalmers s'est trompé dans ses étiquettes : ce sont les problèmes supposés faciles qui présentent un vrai défi scientifique, tandis que le problème difficile ne nous paraît ardu que parce qu'il fait appel à des intuitions imprécises. Une fois revisité à l'aune des neurosciences cognitives et de l'informatique, le problème difficile de Chalmers s'évaporera sans laisser de traces. Ce ne sera pas la première fois que la science remettra en cause nos intuitions les plus sûres (pensez au lever de soleil, qui est en fait une rotation de la Terre dans le sens opposé). Dans quelques décennies, la notion même de *qualia*, ces quanta d'expérience pure, dépourvus de tout rôle dans le traitement de l'information, sera considérée comme une idée étrange de l'ère préscientifique. Ce débat aura la même issue que celui du vitalisme – l'idée erronée du XIX^e^ siècle selon laquelle, quelle que soit la quantité de détails accumulés sur la biochimie des organismes, nous ne rendrons jamais compte de ce qu'est le vivant. La biologie moléculaire moderne a fait voler en éclats cette croyance et révélé la beauté et la complexité de l'automate autoreproducteur qu'abrite chacune de nos cellules. De la même manière, la science de la conscience va grignoter, petit à petit, chacun des aspects du problème difficile de la conscience jusqu'à ce qu'il

parte en fumée. Les modèles actuels de la perception visuelle expliquent déjà non seulement pourquoi notre cerveau est affecté par certaines illusions visuelles, mais également pourquoi n'importe quelle machine rationnelle, qui résout efficacement des problèmes de reconnaissance visuelle en présence d'incertitudes, devra pâtir des mêmes illusions[72]. La science de la conscience rend déjà compte d'une fraction importante de nos expériences subjectives, et je ne vois pas de limites évidentes à cette approche scientifique classique.

Un autre argument philosophique prétend que, quels que soient nos efforts de simulation du cerveau, il manquera toujours à nos logiciels une propriété fondamentale de l'esprit humain : le libre arbitre. L'existence même d'une machine pourvue d'un libre arbitre serait une contradiction dans les termes, parce que les machines sont toujours déterministes : leur comportement résulte de leur organisation interne et de leur état initial. Il se peut que leurs actions ne soient pas prévisibles, à cause du chaos et de l'imprécision de nos mesures, mais elles ne peuvent en aucun cas dévier de la chaîne causale que dicte leur composition physique. Ce déterminisme, selon ces philosophes, ne laisse aucune place à la liberté individuelle. Comme l'a écrit, dès le I^er^ siècle avant notre ère, le poète et philosophe Lucrèce,

> « si tous les mouvements sont enchaînés dans la nature, si toujours d'un premier naît un second suivant un ordre rigoureux ; si, par leur déclinaison, les atomes ne provoquent pas un mouvement qui rompe les lois de la fatalité et qui empêche que les causes ne se succèdent à l'infini ; d'où vient donc cette liberté accordée sur terre aux êtres vivants, d'où vient, dis-je, cette libre faculté arrachée au destin, qui nous fait aller partout où la volonté nous mène[73] ? »

De talentueux scientifiques contemporains jugent ce problème tellement insurmontable qu'ils cherchent la solution dans de nouvelles lois de la physique. Seule la mécanique quantique, estiment-ils, introduit une indispensable pincée de liberté dans la physique de nos neurones. John Eccles (1903-1997), qui reçut le prix Nobel en 1963 pour ses découvertes sur les mécanismes chimiques de la transmission synaptique, faisait partie de ces neurosceptiques. Selon lui, le problème central des neurosciences était de découvrir « comment le soi contrôle son cerveau » – titre d'un de ses nombreux livres sur le sujet[74]. La formulation même de cette question est dualiste, puisqu'elle sépare le soi du cerveau. La solution proposée ne l'est pas moins : sans le moindre commencement de preuve expérimentale, Eccles postule que l'âme immatérielle agit sur la matière du cerveau en modifiant la probabilité des événements quantiques au niveau de chaque synapse.

Un autre brillant scientifique contemporain, le physicien sir Roger Penrose, estime également que seule la mécanique quantique peut rendre compte de la conscience et du libre arbitre[75]. Penrose, aux côtés de l'anesthésiste Stuart Hameroff, défend l'idée que le cerveau est un ordinateur quantique. Selon eux, au cours de son évolution, le cerveau humain serait parvenu à exploiter une propriété fondamentale des systèmes quantiques : la superposition d'états quantiques. Cela lui permettrait d'explorer un nombre gigantesque d'hypothèses en un temps limité et expliquerait, on ne sait trop comment, la faculté des mathématiciens à déjouer les complexités du théorème de Gödel.

Ces propositions baroques ne reposent toutefois pas sur la moindre donnée de neurobiologie ou de sciences cognitives. Je ne nie pas qu'il faille expliquer l'intuition du libre arbitre, cette impression que nous partageons tous que notre esprit prend ses décisions en toute liberté et sans contrainte. Cependant, la physique quantique, version moderne de la « déclinaison des atomes » de Lucrèce, n'est d'aucun secours. La plupart des physiciens s'accordent à penser que le milieu humide et chaud dans lequel s'ébattent nos neurones est incompatible avec le calcul quantique,

car celui-ci nécessite des températures extrêmement basses afin d'éviter la dissipation ultrarapide de la cohérence quantique. L'échelle de la femtoseconde (10^{-15} seconde) qui régit la décohérence des états quantiques est également sans commune mesure avec le temps qu'il nous faut pour prendre conscience d'un aspect du monde extérieur (de l'ordre d'un tiers de seconde).

Enfin, même si on découvrait des phénomènes quantiques dans le cerveau, leur caractère strictement imprévisible ne permettrait pas d'expliquer notre conception du libre arbitre. Comme l'a bien montré le philosophe Dan Dennett, l'aléatoire pur ne confère à nos cerveaux « aucune forme valable de liberté » (« *any kind of freedom worth having*[76] »). Souhaitons-nous vraiment que nos corps soient secoués de mouvements aléatoires et incontrôlables engendrés au niveau subatomique, qui rapprocheraient nos décisions des convulsions et des tics des patients souffrants du syndrome de Gilles de La Tourette, la fameuse « danse de Saint-Guy » ? Rien n'est plus éloigné de notre conception de la liberté. La maladie de Tourette ne rend pas libre, bien au contraire. Jamais un coup de dés n'engendrera d'esprit libre.

Lorsque nous parlons du « libre arbitre », nous pensons à une forme beaucoup plus intéressante de liberté. Notre croyance en un libre arbitre résulte d'une observation élémentaire : dans des circonstances normales, nous prenons nos décisions en toute indépendance, en nous laissant seulement guider par nos idées, nos croyances et notre expérience passée, et en contrôlant nos pulsions indésirables. Nous exerçons notre libre arbitre chaque fois que nous avons la possibilité d'examiner les choix qui s'offrent à nous, d'y réfléchir posément et d'opter pour celui qui nous paraît le meilleur. Une part de hasard entre dans nos choix volontaires, mais elle n'en constitue pas un élément indispensable. La plupart du temps, nos actions volontaires n'ont rien d'aléatoire : elles résultent d'un examen attentif des options disponibles, suivi du choix délibéré de celle qui emporte notre préférence.

Cette conception du libre arbitre n'a nul besoin de la physique quantique – elle pourrait être simulée par un ordinateur standard. Elle exige simplement un espace de travail qui recueille les informations en provenance des sens et de la mémoire, en fasse la synthèse, évalue les conséquences de chaque option, y consacre autant de temps que nécessaire et utilise cette réflexion pour guider notre choix. Voilà ce que nous appelons une décision volontaire, délibérée, prise « en toute conscience ».

En bref, l'intuition du libre arbitre doit être décomposée : elle recouvre, d'une part, l'idée que nos décisions sont fondamentalement indéterminées, non contraintes par la physique (une idée fausse) ; et d'autre part, celle que nous les prenons en toute autonomie (une idée respectable). Nos états cérébraux sont nécessairement déterminés par des causes physiques, car rien de ce qui est matériel n'échappe aux lois de la nature. Mais cela n'exclut pas que nos décisions soient réellement *libres*, si l'on entend par là qu'elles s'appuient sur une délibération consciente, autonome, qui ne rencontre aucun obstacle et qui dispose du temps suffisant pour évaluer le pour et le contre avant de s'engager. Quand toutes ces conditions sont remplies, nous avons raison de dire que nous avons exercé notre libre arbitre et pris une décision volontaire – même si celle-ci est toujours, en dernière analyse, déterminée par nos gènes, notre histoire et les fonctions de valeur qui sont inscrites dans nos circuits neuronaux. Les fluctuations de l'activité spontanée de ces réseaux rendent nos décisions imprévisibles, y compris à nos propres yeux. Cependant, ce caractère imprévisible ne devrait pas être retenu comme l'un des critères essentiels du libre arbitre, ni ne devrait être confondu avec l'indétermination fondamentale de la physique quantique. Ce qui compte pour qu'une décision soit libre, c'est l'autonomie de la délibération.

Une machine pourvue d'un libre arbitre n'est absolument pas une contradiction dans les termes, juste une définition de ce que nous sommes. Je n'ai aucune difficulté à imaginer un dispositif artificiel qui soit capable de décider, par lui seul, de la conduite à tenir. Peut-être est-ce

déjà le cas de la Google Car, la voiture autonome qui se conduit seule, sans chauffeur, dans les rues de Californie. Même si notre architecture cérébrale était totalement déterministe, comme l'est sans doute le logiciel de Google, il ne serait pas illégitime de lui attribuer une forme de libre arbitre. Lorsque l'architecture d'un réseau de neurones permet une délibération autonome, nous sommes en droit de lui attribuer une certaine forme de liberté – et lorsque nous aurons pleinement identifié comment cette architecture fonctionne, nous pourrons la reproduire dans une machine.

En bref, ni la notion de *qualia* ni celle de libre arbitre ne constituent un obstacle philosophique sérieux à l'idée d'une machine consciente. Parvenus au terme de notre voyage au sein du cerveau conscient, nous comprenons avec quelle prudence il nous faut traiter nos intuitions de ce qu'une machine neuronale immensément complexe peut faire. La puissance de calcul que permet un réseau de 16 milliards de neurones corticaux dépasse notre imagination actuelle. Ce réseau fluctue sans cesse d'une façon partiellement autonome, créant ainsi un monde interne de représentations subjectives. Même lorsqu'il est confronté à des entrées sensorielles strictement identiques, il ne réagit pas de la même manière en fonction de son humeur, de ses buts et de ses souvenirs. Le code neural varie également d'un cerveau à l'autre. Bien que nous partagions tous le même attirail de neurones codant pour la couleur, la forme ou le mouvement, les détails de leur organisation résultent d'un long développement qui sculpte différemment chacun de nos cerveaux, en amplifiant ou en éliminant certaines de nos synapses.

L'architecture neuronale qui résulte de ce foisonnement de règles génétiques, d'expériences nouvelles et de rencontres fortuites fait de chacun de nous des individus uniques. Elle nous donne accès à un nombre immense d'états d'activités qui correspondent à autant d'états de conscience possibles, un vaste univers de représentations mentales reliées à l'environnement, mais jamais entièrement déterminées par lui. Chacune de nos expériences subjectives (douleur, beauté, désir, regret, etc.)

correspond à un attracteur stable de ce paysage neuronal. Ce sont des états nécessairement subjectifs, parce que la dynamique du cerveau oblige toute entrée sensorielle à s'insérer dans la trame de nos souvenirs, ce qui lui confère une dimension personnelle.

De notre cerveau émerge un « présent remémoré[77] » : un code neural individuel de l'événement présent, épaissi de souvenirs et d'anticipations, et qui projette constamment sur le monde extérieur le point de vue subjectif d'une personnalité singulière – un espace interne dont surgit la conscience.

En ce moment même, cette délicate machine est à l'œuvre dans votre cerveau. Tandis que vous refermez ce livre et méditez sur votre propre existence, l'embrasement d'une assemblée de neurones constitue la matière de votre pensée.

Notes

INTRODUCTION
La matière de la pensée

1. Jouvet, 1999, 169-71.

2. Damasio, 1994.

3. James, 1890, chap. 5.

4. Les citations de Descartes sont issues de *L'Homme*, écrit autour de 1632-1633, et publié pour la première fois en 1662.

5. Il ne fait guère de doute que Descartes craignait le conflit avec l'Église. Il n'avait que 4 ans en 1600, lorsque Giordano Bruno fut envoyé au bûcher, et 37 ans en 1633, quand Galilée échappa de justesse au même sort. Descartes fit tout pour que son grand œuvre, *Le Monde*, qui contenait un traité particulièrement subversif sur *L'Homme*, ne paraisse pas de son vivant. Celui-ci ne fut publié qu'en 1664, quatorze ans après sa mort en 1650. Seules quelques allusions à son contenu transparaissent dans le *Discours de la méthode* (1637) et les *Passions de l'âme* (1649). Sans doute cette prudence n'était-elle pas exagérée, puisque, dès 1663, le Saint-Siège mit ces œuvres à l'index des livres frappés d'interdit. L'insistance de Descartes sur l'immatérialité de l'âme était donc peut-être, au moins en partie, une façade destinée à protéger sa vie – même si sa correspondance suggère que l'autonomie de la pensée humaine lui posait également de réelles difficultés théoriques.

6. Michel de Montaigne, *Essais*, 2 : 12.

7. Voir par exemple Posner et Snyder, 1975/2004 ; Shallice, 1979 ; Shallice, 1972 ; Marcel, 1983 ; Libet, Alberts, Wright et Feinstein, 1967 ; Bisiach, Luzzatti et Perani, 1979 ; Weiskrantz, 1986 ; Frith, 1979 ; Weiskrantz, 1997.

8. Baars, 1989.

9. Watson, 1913.

10. Nisbett et Wilson, 1977 ; Johansson, Hall, Sikstrom et Olsson, 2005.

11. Le philosophe américain Daniel Dennett appelle cette approche l'« hétérophénoménologie » (Dennett, 1991).

CHAPITRE 1

Quand la conscience entre au laboratoire

1. Crick et Koch, 1990a ; Crick et Koch, 1990b. De nombreux autres psychologues et neuroscientifiques ont également proposé une approche empirique et réductionniste de la conscience (voir par exemple : Churchland, 1986 ; Changeux, 1983 ; Baars, 1989 ; Weiskrantz, 1986 ; Posner, 1975/2004 ; Shallice, 1972). Cependant, les articles de Crick et Koch, avec leur approche très concrète et focalisée sur la vision, ont joué un rôle essentiel dans l'attrait que les expérimentateurs ont soudain éprouvé pour ce domaine.

2. Kim et Blake, 2005.

3. Posner, 1994.

4. Wyart, Dehaene et Tallon-Baudry, 2012 ; Wyart et Tallon-Baudry, 2008.

5. Gallup, 1970.

6. Plotnik, de Waal et Reiss, 2006 ; Prior, Schwarz et Gunturkun, 2008 ; Reiss et Marino, 2001.

7. Epstein, Lanza et Skinner, 1981.

8. Pour une discussion approfondie du test de reconnaissance de soi dans un miroir, voir Suddendorf et Butler, 2013.

9. Hofstadter, 2007.

10. Comte, vol. 1, p. 31-32.

11. John Stuart Mill, *Auguste Comte et le positivisme* (1865), p. 68-69.

12. Certains scientifiques anglophones distinguent *awareness* et *consciousness*. *Awareness* ferait référence à une forme élémentaire de conscience perceptive dans laquelle l'information serait d'origine strictement sensorielle – ce que j'appelle l'« accès d'une information sensorielle à la conscience ». La plupart des dictionnaires, cependant, ne s'accordent pas sur cet usage et considèrent les mots *awareness* et *consciousness* comme des synonymes. Ma théorie ne fait aucun distinguo entre ces deux termes, mais distingue l'accès à la conscience (*conscious access*), l'éveil (*wakefulness*), la vigilance (*vigilance*), la conscience de soi (*self-consciousness*) et la métacognition (*metacognition*).

13. Baars, 1989.

14. Schneider et Shiffrin, 1977 ; Shiffrin et Schneider, 1977 ; Posner, 1975/2004 ; Raichle, Fiesz, Videen et MacLeod, 1994 ; Chein et Schneider, 2005.

15. New et Scholl, 2008 ; Ramachandran et Gregory, 1991.

16. Leopold et Logothetis, 1996 ; Logothetis, Leopold et Sheinberg, 1996 ; Leopold et Logothetis, 1999. Ces travaux pionniers ont été répliqués à l'aide de la technique de « suppression continue par des flashs » (*continuous flash suppression*) qui permet de contrôler beaucoup plus finement le

moment où l'image disparaît de la conscience (voir par exemple : Maier, Wilke, Aura, Zhu, Ye et Leopold, 2008 ; Wilke, Logothetis et Leopold, 2006 ; Fries, Schroder, Roelfsema, Singer et Engel, 2002). Plusieurs chercheurs ont également utilisé l'imagerie cérébrale afin d'examiner le parcours d'une image visible ou invisible dans le cerveau (voir par exemple : Srinivasan, Russell, Edelman et Tononi, 1999 ; Lumer, Friston et Rees, 1998 ; Haynes, Deichmann et Rees, 2005 ; Haynes, Driver et Rees, 2005).

17. Wilke, Logothetis et Leopold, 2003 ; Tsuchiya et Koch, 2005.

18. Chong, Tadin et Blake, 2005 ; Chong et Blake, 2006.

19. Zhang, Jamison, Engel, He et He, 2011 ; Brascamp et Blake, 2012.

20. Zhang, Jamison, Engel, He et He, 2011.

21. Brascamp et Blake, 2012.

22. Raymond, Shapiro et Arnell, 1992.

23. Marti, Sigman et Dehaene, 2012.

24. Chun et Potter, 1995.

25. Telford, 1931 ; Pashler, 1984 ; Pashler, 1994 ; Sigman et Dehaene, 2005.

26. Marti, Sackur, Sigman et Dehaene, 2010 ; Dehaene, Pegado, Braga, Ventural, Nunes Filho, Jobert, Dehaene-Lambertz *et al*., 2010 ; Corallo, Sackur, Dehaene et Sigman, 2008.

27. Marti, Sigman et Dehaene, 2012 ; Wong, 2002 ; Jolicœur, 1999.

28. Mack et Rock, 1998.

29. Simons et Chabris, 1999. Le film du gorille se trouve sur http://www.youtube.com/watch?v=vJG698U2Mvo.

30. Rensink, O'Regan et Clark, 1997. Pour quelques travaux récents sur les bases cérébrales de la cécité au changement, voir Beck, Rees, Frith et Lavie, 2001 ; Landman, Spekreijse et Lamme, 2003 ; Simons et Ambinder, 2005 ; Beck, Muggleton, Walsh et Lavie, 2006 ; Reddy, Quiroga, Wilken, Koch et Fried, 2006.

31. Johansson, Hall, Sikstrom et Olsson, 2005.

32. Le film se trouve sur http://www.youtube.com/watch?v=ubNF9QNEQLA.

33. Ces expériences sont discutées par Simons et Ambinder, 2005 ; Landman, Spekreijse et Lamme, 2003 ; Block, 2007.

34. Woodman et Luck, 2003 ; Giesbrecht et Di Lollo, 1998 ; Di Lollo, Enns et Rensink, 2000.

35. Del Cul, Dehaene et Leboyer, 2006 ; Gaillard, Del Cul, Naccache, Vinckier, Cohen et Dehaene, 2006 ; Del Cul, Baillet et Dehaene, 2007 ; Del Cul, Dehaene, Reyes, Bravo et Slachevsky, 2009 ; Sergent et Dehaene, 2004.

36. Dehaene, Naccache, Cohen, Le Bihan, Mangin, Poline et Rivière, 2001.

37. Del Cul, Dehaene, Reyes, Bravo et Slachevsky, 2009 ; Charles, Van Opstal, Marti et Dehaene, 2013.

38. Dehaene et Naccache, 2001.

39. Ffytche, Hoard, Brammer, David, Woodruff et Williams, 1998.

40. Kruger et Dunning, 1999 ; Johansson, Hall, Sikstrom et Olsson, 2005 ; Nisbett et Wilson, 1977.

41. Dehaene, 2009 ; Dehaene, Naccache, Cohen, Le Bihan, Mangin, Poline et Rivière, 2001.

42. Blanke, Landis, Spinelli et Seeck, 2004 ; Blanke, Ortigue, Landis et Seeck, 2002.

43. Lenggenhager, Mouthon et Blanke, 2009 ; Lenggenhager, Tadi, Metzinger et Blanke, 2007. Voir également Ehrsson, 2007. Un précurseur célèbre de ces expériences de sortie du corps est l'illusion de la main en caoutchouc découverte par Botvinick et Cohen, 1998 ; voir aussi Ehrsson, Spence et Passingham, 2004.

44. Toutes les méthodes qui permettent de rendre une image inconsciente ne bloquent pas le traitement de l'information au même stade. Par exemple, la rivalité binoculaire semble interférer avec la vision à un niveau plus précoce que ne le fait le masquage (Almeida, Mahon, Nakayama et Caramazza, 2008 ; Breitmeyer, Koc, Ogmen et Ziegler, 2008). Il est donc essentiel de comparer plusieurs méthodes d'invisibilité afin de déterminer les conditions nécessaires et suffisantes d'un accès à la conscience.

CHAPITRE 2
Sonder la profondeur de l'inconscient

1. Pour un historique détaillé du concept d'inconscient, voir Ellenberger, 1970.

2. Gauchet, 1992.

3. Pour un exposé lucide, détaillé et accessible de l'histoire des neurosciences, voir Finger, 2001.

4. Howard, 1996.

5. *Ibid*.

6. Maudsley, 1868.

7. James, 1890, p. 211 et 208. Voir Ellenberger, 1970 ; et Weinberger, 2000.

8. Vladimir Nabokov, *Intransigeances* (1973 ; Julliard, 1985 pour l'édition française, traduction de Vladimir Sikorsky), p. 77.

9. Ledoux, 1996.

10. Weiskrantz, 1997.

11. Sahraie, Weiskrantz, Barbur, Simmons, Williams et Brammer, 1997. Voir également Morris, DeGelder, Weiskrantz et Dolan, 2001.

12. Morland, Le, Carroll, Hoffmann et Pambakian, 2004 ; Schmid, Mrowka, Turchi, Sauinders, Wilke, Peters, Ye et Leopold, 2010 ; Schmid, Panagiotaropoulos, Augath, Logothetis et Smirnakis, 2009 ; Goebel, Muckli, Zanella, Singer et Stoerig, 2001.

13. Goodale, Milner, Jakobson et Carey, 1991 ; Milner et Goodale, 1995.

14. Marshall et Halligan, 1988.

15. Driver et Vuilleumier, 2001 ; Vuilleumier, Sagiv, Hazeltine, Poldrack, Swick, Rafal et Gabrieli, 2001.

16. Sackur, Naccache, Pradat-Diehl, Azouvi, Mazevet, Katz, Cohen et Dehaene, 2008 ; McGlinchey-Berroth, Milberg, Verfaellie, Alexander et Kilduff, 1993.

17. Marcel, 1983 ; Forster, 1998 ; Forster et Davis, 1984. Pour une revue historique des recherches sur la perception subliminale, voir Kouider et Dehaene, 2007.

18. Bowers, Vigliocco et Haan, 1998 ; Forster et Davis, 1984.

19. Dehaene, Naccache, LeClec'H, Koechlin, Mueller, Dehaene-Lambertz, Van de Moortele et Le Bihan, 1998 ; Dehaene, Naccache, Cohen, Le Bihan, Mangin, Poline et Rivière, 2001.

20. Dehaene, 2009.

21. Dehaene et Naccache, 2001 ; Dehaene, Naccache, Cohen, Le Bihan, Mangin, Poline et Rivière, 2001 ; Dehaene, Jobert, Naccache, Ciuciu, Poline, Le Bihan et Cohen, 2004.

22. Goodale, Milner, Jakobson et Carey, 1991 ; Milner, 1995.

23. Kanwisher, 2001.

24. Treisman et Gelade, 1980 ; Kahneman et Treisman, 1984 ; Treisman et Souther, 1986.

25. Crick, 2003 ; Singer, 1998.

26. Finkel et Edelman, 1989 ; Edelman, 1989.

27. Dehaene, Jobert, Naccache, Ciuciu, Poline, Le Bihan et Cohen, 2004.

28. Voir par exemple Henson, Mouchlianitis, Matthews et Kouider, 2008 ; Kouider, Eger, Dolan et Henson, 2009 ; Dell'Acqua et Grainger, 1999.

29. de Groot et Gobet, 1996 ; Gobet et Simon, 1998.

30. Kiesel, Kunde, Pohl, Berner et Hoffmann, 2009.

31. McGurk et MacDonald, 1976.

32. On trouvera une démonstration de l'illusion de McGurk sur http://www.youtube.com/watch?v=jtsfidRq2tw.

33. Hasson, Skiper, Nusbaum et Small, 2007.

34. Singer, 1998.

35. Tsunoda, Yamane, Nishizaki et Tanifuji, 2001 ; Baker, Behrmann et Olson, 2002 ; Brincat et Connor, 2004.

36. Dehaene, 2009 ; Dehaene, Pegado, Braga, Ventural, Nunes Filho, Jobert, Dehaene-Lambertz *et al.*, 2010.

37. Davis, Coleman, Absalom, Rodd, Johnsrude, Matta, Owen et Menon, 2007.

38. Une expérience pionnière fut réalisée par Sidis (1898). Il montra qu'une lettre placée si loin de l'observateur qu'il affirme ne pas la voir continue d'être nommée bien mieux qu'au hasard.

39. Broadbent, 1962.

40. Moray, 1959.

41. Lewis, 1970.

42. Marcel, 1983.

43. Marcel, 1980.

44. Schvaneveldt et Meyer, 1976.

45. Holender, 1986 ; Holender et Duscherer, 2004.

46. Dell'Acqua et Grainger, 1999 ; Dehaene, Naccache, LeClec'H, Koechlin, Mueller, Dehaene-Lambertz, Van de Moortele et Le Bihan, 1998 ; Naccache et Dehaene, 2001b ; Merikle, 1992 ; Merikle et Joordens, 1997.

47. Abrams et Greenwald, 2000.

48. En principe, il se pourrait même que cette association aille des lettres à la réponse motrice (*j-o-i-e* → appuyer à droite). Anthony Greenwald parvint à rejeter cette interprétation en montrant que lorsqu'on échange les mains associées respectivement aux deux réponses (émotion « positive » ou « négative »), le mot « joie » continue d'amorcer la réponse « positive », bien que celle-ci soit désormais produite avec l'autre main. Voir Abrams, Klinger et Greenwald, 2002.

49. Dehaene, Naccache, LeClec'H, Koechlin, Mueller, Dehaene-Lambertz, Van de Moortele et Le Bihan, 1998 ; Naccache et Dehaene, 2001a, Naccache et Dehaene, 2001b ; Greenwald, Abrams, Naccache et Dehaene, 2003 ; Kouider et Dehaene, 2009.

50. Kouider et Dehaene, 2009.

51. Naccache et Dehaene, 2001b ; Greenwald, Abrams, Naccache et Dehaene, 2003.

52. Naccache et Dehaene, 2001a.

53. Dehaene, 2011.

54. Nieder et Miller, 2004 ; Piazza, Izard, Pinel, Le Bihan et Dehaene, 2004 ; Piazza, Pinel, Le Bihan et Dehaene, 2007 ; Nieder et Dehaene, 2009.

55. den Heyer et Briand, 1986 ; Koechlin, Naccache, Block et Dehaene, 1999 ; Reynvoet et Brysbaert, 1999 ; Reynvoet, Brysbaert et Fias, 2002 ; Reynvoet et Brysbaert, 2004 ; Reynvoet, Gevers et Caessens, 2005.

56. Van den Bussche et Reynvoet, 2007 ; Van den Bussche, Notebaert et Reynvoet, 2009.

57. Naccache, Gaillard, Adam, Hasbou, Clémenceau, Baulac, Dehaene et Cohen, 2005.

58. Morris, Ohman et Dolan, 1999 ; Morris, Ohman et Dolan, 1998.

59. Kiefer et Spitzer, 2000 ; Kiefer, 2002 ; Kiefer et Brendel, 2006.

60. Vogel, Luck et Shapiro, 1998 ; Luck, Vogel et Shapiro, 1996.

61. Van Gaal, Naccache, Meeuwese, Van Loon, Cohen et Dehaene, 2013.

62. Pour une démonstration de traitement inconscient de la syntaxe, voir Batterink et Neville, 2013.

63. Sergent, Baillet et Dehaene, 2005.

64. Cohen, Cavanagh, Chun et Nakayama, 2012 ; Posner et Rothbart, 1998 ; Posner, 1994.

65. Pour une revue des dissociations entre l'attention et la conscience, voir Koch et Tsuchiya, 2007.

66. McCormick, 1997.

67. Bressan et Pizzighello, 2008 ; Tsushima, Seitz et Watanabe, 2008 ; Tsushima, Sasaki et Watanabe, 2006.

68. Posner et Snyder, 1975.

69. Naccache, Blandin et Dehaene, 2002 ; voir également Lachter, Forster et Ruthruff, 2004 ; Kentridge, Nijboer et Heywood, 2008 ; Kiefer et Brendel, 2006.

70. Woodman et Luck, 2003.

71. Marti, Sigman et Dehaene, 2012.

72. Pessiglione, Schmidt, Draganski, Kalisch, Lau, Dolan et Frith, 2007.

73. Pessiglione, Petrovic, Daunizeau, Palminteri, Dolan et Frith, 2008.

74. Jaynes, 1976, p. 23.

75. Hadamard, 1945.

76. Bechara, Damasio, Tranel et Damasio, 1997. Les résultats ont été remis en question par Maia et McClelland, 2004, puis clarifiés par Persaud, Davidson, Maniscalco, Mobbs, Passingham, Cowey et Lau, 2011.

77. Lawrence, Jollant, O'Daly, Zelaya et Phillips, 2009.

78. Dijksterhuis, Bos, Nordgren et Van Baaren, 2006.

79. Yang et Shadlen, 2007.

80. de Lange, Van Gaal, Lamme et Dehaene, 2011.

81. Van Opstal, de Lange et Dehaene, 2011.

82. Wagner, Gais, Haider, Verleger et Born, 2004.

83. Ji et Wilson, 2007 ; Louie et Wilson, 2001.

84. Van Gaal, Ridderinkof, Fahrenfort, Scholte et Lamme, 2008.

85. Van Gaal, Ridderinkhof, Scholte et Lamme, 2010.

86. Nieuwenhuis, Ridderinkhof, Blom, Band et Kok, 2001.

87. Lau et Passingham, 2007 ; voir également Reuss, Kiesel, Kunde et Hommel, 2011.

88. Lau et Rosenthal, 2011 ; Rosenthal, 2008 ; Bargh et Morsella, 2008 ; Velmans, 1991.

CHAPITRE 3

À quoi sert la conscience ?

1. Turing, 1952.

2. Gould, 1974.

3. Gould et Lewontin, 1979.

4. Velmans, 1991.

5. Nørretranders, 1999.

6. Lau et Rosenthal, 2011 ; Velmans, 1991 ; Wegner, 2003. Benjamin Libet porte un jugement plus nuancé, en arguant que la conscience ne joue aucun rôle dans l'initiation de l'action, mais peut lui imposer un « veto » tardif (voir Libet, 2004 ; Libet, Gleason, Wright et Pearl, 1983).

7. Peirce, 1901.

8. Pack et Born, 2001.

9. Pack, Berezovskii et Born, 2001.

10. Moreno-Bote, Knill et Pouget, 2011.

11. Nous en avons discuté au chapitre 1. Voir Brascamp et Blake, 2012 ; Zhang, Jamison, Engel, He et He, 2011.

12. Norris, 2009 ; Norris, 2006.

13. Schvaneveldt et Meyer, 1976.

14. Vul, Hanus et Kanwisher, 2009 ; Vul, Nieuwenstein et Kanwisher, 2008.

15. Vul et Pashler, 2008.

16. Fuster, 1973 ; Fuster, 2008 ; Funahashi, Bruce et Goldman-Rakic, 1989 ; Goldman-Rakic, 1995.

17. Rounis, Maniscalco, Rothwell, Passingham et Lau, 2010 ; Del Cul, Dehaene, Reyes, Bravo et Slachevsky, 2009.

18. Clark, Manns et Squire, 2002 ; Clark et Squire, 1998.

19. Carter, O'Doherty, Seymoure, Koch et Dolan, 2006. Voir également Carter, Hofstotter, Tsuchiya et Koch, 2003. La valeur du test de conditionnement fondé sur la mémoire reste débattue, car certains patients en état végétatif semblent capables de réussir ce test, sans qu'on sache si cela signifie qu'ils sont conscients ou, au contraire, que le test n'exige pas d'être conscient (Bekinschtein, Shalom, Forcato, Herrera, Coleman, Manes et Sigman, 2009 ; Bekinschtein, Peeters, Shalom et Sigman, 2011).

20. Edelman, 1989.

21. Han, O'Tuathaigh, Van Trigt, Quinn, Fanselow, Mongeau, Koch et Anderson, 2003.

22. Mattler, 2005 ; Greenwald, Draine et Abrams, 1996 ; Dupoux, de Gardelle et Kouider, 2008.

23. Naccache, 2006b.

24. Soto, Matyla et Silvanto, 2011.

25. Siegler, 1987 ; Siegler, 1988 ; Siegler, 1989 ; Siegler et Jenkins, 1989.

26. Un article récent et controversé prétend qu'il est possible de résoudre inconsciemment des soustractions complexes telles que : 9 – 4 – 3, même lorsqu'elles sont présentées à un seul œil et rendues invisibles par *continuous flash suppression,* c'est-à-dire en présentant une succession de formes à l'autre œil (Sklar, Levy, Goldstein, Mandel, Maril et Hassin, 2012). La conception de cette étude, cependant, n'exclut pas que les réponses s'appuient sur un calcul partiel (par exemple 9 – 4 seulement). Même si des recherches ultérieures confirmaient qu'il était possible de combiner inconsciemment plusieurs nombres au sein d'un calcul, ce calcul pourrait s'effectuer de façon très différente suivant qu'il soit conscient ou inconscient. Nous savons que des opérations sophistiquées, telles que le calcul de la moyenne de huit chiffres, peuvent se dérouler inconsciemment (De Lange, Van Gaal, Lamme et Dehaene, 2011 ; Van Opstal, de Lange et Dehaene, 2011). Toutefois, l'exécution mentale d'un algorithme, dans un mode lent, sériel, flexible et contrôlé semble être l'apanage de la conscience.

27. Zylberberg, Fernandez Slezak, Roelfsema, Dehaene et Sigman, 2010.

28. Zylberberg, Dehaene, Roelfsema et Sigman, 2011 ; Zylberberg, Fernandez Slezak, Roelfsema, Dehaene et Sigman, 2010 ; Zylberberg, Dehaene, Mindlin et Sigman, 2009 ; Dehaene et Sigman, 2012. Voir aussi Shanahan et Baars, 2005.

29. Turing, 1936.

30. Anderson, 1983 ; Anderson et Lebiere, 1998.

31. Ashcraft et Stazyk, 1981 ; Widaman, Geary, Cormier et Little, 1989.

32. Tombu et Jolicœur, 2003 ; Logan et Schulkind, 2000 ; Moro, Tolboom, Khayat et Roelfsema, 2010.

33. Sackur et Dehaene, 2009.

34. Dehaene et Cohen, 2007 ; Dehaene, 2009.

35. Les calculateurs prodiges semblent contredire cette prédiction. Cependant, il me semble que nous en savons encore trop peu sur leurs stratégies, au point que nous ignorons si elles s'exécutent consciemment ou inconsciemment. Leurs calculs leur demandent souvent plusieurs secondes d'attention, pendant lesquelles il ne faut surtout pas les distraire. Il est vrai qu'ils sont souvent incapables de décrire leurs algorithmes, ou refusent de le faire, mais cela ne signifie pas nécessairement que ceux-ci s'exécutent inconsciemment. Certains calculateurs disent voir l'image du calendrier qu'ils manipulent mentalement (voir Howe et Smith, 1988).

36. de Lange, Van Gaal, Lamme et Dehaene, 2011.

37. Van Opstal, de Lange et Dehaene, 2011.

38. Dijksterhuis, Bos, Nordgren et Van Baaren, 2006.

39. Levelt, 1989.

40. Reed et Durlach, 1998.

41. Dunbar, 1996.

42. Bahrami, Olsen, Latham, Roepstorff, Rees et Frith, 2010.

43. Buckner, Andrews-Hanna et Schacter, 2008.

44. Yokoyama, Miura, Watanabe, Takemoto, Uchida, Sugiura, Horie *et al.*, 2010 ; Kikyo, Ohki et Miyashita, 2002 ; voir aussi Rounis, Maniscalco, Rothwell, Passingham et Lau, 2010 ; Del Cul, Dehaene, Reyes, Bravo et Slachevsky, 2009 ; Fleming, Weil, Nagy, Dolan et Rees, 2010.

45. Saxe et Powell, 2006 ; Perner et Aichhorn, 2008.

46. Ochsner, Knierim, Ludlow, Hanelin, Ramachandran, Glover et Mackey, 2004 ; Vogeley, Bussfeld, Newen, Herrmann, Happe, Falkai, Maier *et al.*, 2001.

47. Jenkins, Macrae et Mitchell, 2008.

48. Ricœur, 1990.

49. Frith, 2007.

50. Marti, Sackur, Sigman et Dehaene, 2010 ; Corallo, Sackur, Dehaene et Sigman, 2008.

51. David Lodge, *Pensées secrètes*, traduction de Suzanne V. Mayoux, révisée par l'auteur, Payot & Rivages, 2001.

52. Vladimir Nabokov, *Littératures 1* (1980), traduction d'Hélène Pasquier, Robert Laffont, « Bouquins », 2009 pour l'édition française, p. 40.

CHAPITRE 4

Les signatures de la pensée consciente

1. Ogawa, Lee, Kay et Tank, 1990.

2. Grill-Spector, Kushnir, Hendler et Malach, 2000.

3. Dehaene, Naccache, Cohen, Le Bihan, Mangin, Poline et Rivière, 2001.

4. Naccache et Dehaene, 2001a.

5. Dehaene, 2009.

6. Dehaene, Naccache, Cohen, Le Bihan, Mangin, Poline et Rivière, 2001. Nikos Logothesis et ses collègues sont parvenus à des conclusions similaires en enregistrant des neurones uniques chez le singe éveillé ; voir Leopold et Logothetis, 1996 ; Logothetis, Leopold et Sheinberg, 1996 ; Logothetis, 1998.

7. Dehaene, Naccache, Cohen, Le Bihan, Mangin, Poline et Rivière, 2001. Voir aussi Rodriguez, George, Lachaux, Martinerie, Renault et Varela, 1999 ; Varela, Lachaux, Rodriguez et Martinerie, 2001, pour une suggestion de la même nature, sans toutefois contraster de stimuli conscients et non conscients.

8. Sadaghiani, Hesselmann et Kleinschmidt, 2009.

9. Van Gaal, Ridderinkhof, Scholte et Lamme, 2010.

10. Pour d'autres exemples d'activité préfrontale et pariétale au cours d'un effort conscient, voir par exemple : Marois, Yi et Chun, 2004 ; Kouider, Dehaene, Jobert et Le Bihan, 2007 ; Stephan, Thaut, Wunderlich, Schicks, Tian, Tellmann, Schmitz et *al.*, 2002 ; McIntosh, Rajah et Lobaugh, 1999 ; Petersen, Van Mier, Fiez et Raichle, 1998.

11. Sergent, Baillet et Dehaene, 2005.

12. *Ibid.* ; Sergent et Dehaene, 2004.

13. Williams, Baker, Op de Beeck, Shim, Dang, Triantafyllou et Kanwisher, 2008 ; Roelfsema, Lamme et Spekreijse, 1998 ; Roelfsema, Khayat et Spekreijse, 2003 ; Supèr, 2001a ; Supèr, 2001b ; Haynes, Driver et Rees, 2005 ; voir aussi Williams, Visser, Cunnington et Mattingley, 2008.

14. Luck, Vogel et Shapiro, 1996.

15. Les neuroscientifiques distinguent l'onde P3a, qui est engendrée automatiquement dans les aires mésiales du lobe frontal dès qu'un événement inattendu survient, et l'onde P3b, qui reflète l'activation simultanée de multiples aires du cortex. L'onde P3a continue d'être évoquée dans des conditions non conscientes, alors que l'onde P3b est associée aux états conscients. C'est de cette dernière que nous parlons dans le texte.

16. Voir par exemple Lamy, Salti et Bar-Haim, 2009 ; Del Cul, Baillet et Dehaene, 2007 ; Donchin et Coles, 1988 ; Bekinschtein, Dehaene, Rouhaut, Tadel, Cohen et Naccache, 2009 ; Picton, 1992 ; Melloni, Molina, Pena, Torres, Singer et Rodriguez, 2007. Pour une revue, voir Dehaene, 2011.

17. Marti, Sackur, Sigman et Dehaene, 2010 ; Sigman et Dehaene, 2008 ; Marti, Sigman et Dehaene, 2012.

18. Dehaene, 2008.

19. Levy et Boer, 2006 ; Strayer, Drews et Johnston, 2003.

20. Pisella, Grea, Tilikete, Vighetto, Desmurget, Rode, Boisson et Rossetti, 2000.

21. Les mécanismes exacts de cet effet restent discutés. Pour un aperçu de ce fascinant débat, voir Kanai, Carlson, Verstraten et Walsh, 2009 ; Eagleman et Sejnowski, 2007 ; Krekelberg et Lappe, 2001 ; Eagleman et Sejnowski, 2000.

22. Nieuwenhuis, Ridderinkhof, Blom, Band et Kok, 2001.

23. Dehaene, Posner et Tucker, 1994 ; Gehring, Goss, Coles, Meyer et Donchin, 1993.

24. L'idée que la conscience survient longtemps après coup a été introduite par le psychologue californien Benjamin Libet (voir Libet, 1991 ; Libet, Gleason, Wright et Pearl, 1983 ; Libet, Wright, Feinstein et Pearl, 1979 ; Libet, Alberts, Wright et Feinstein, 1967 ; Libet, Alberts, Wright, Delattre, Levin et Feinstein, 1964). Ses expériences étaient très en avance sur leur temps (sans jeu de mots). Dès 1967, il avait déjà remarqué qu'un stimulus non conscient continuait d'évoquer des potentiels évoqués précoces, et que seules les réponses cérébrales tardives étaient un bon marqueur de la conscience (Libet, Alberts, Wright et Feinstein, 1967 ; voir aussi Libet, 1965 ; Schiller et Chorover, 1966). Malheureusement, ses interprétations théoriques allaient bien au-delà des données dont il disposait. Il n'hésitait pas à faire appel à des mécanismes farfelus tels que des « champs mentaux » immatériels et une inversion du temps (voir Libet, 2004). En conséquence, son travail est resté controversé, et ce n'est que récemment que des mécanismes neurophysiologiques précis ont été proposés pour certaines de ses observations (voir par exemple Schurger, Sitt et Dehaene, 2012).

25. Aru, Bachmann, Singer et Melloni, 2012.

26. Sergent, Baillet et Dehaene, 2005.

27. Lau et Passingham, 2006.

28. Persaud, Davidson, Maniscalco, Mobbs, Passingham, Cowey et Lau, 2011.

29. Lamy, Salti et Bar-Haim, 2009.

30. Dehaene et Naccache, 2001.

31. Hebb, 1949.

32. Dehaene, Sergent et Changeux, 2003.

33. Dehaene et Naccache, 2001.

34. Del Cul, Baillet et Dehaene, 2007.

35. *Ibid*. ; Del Cul, Dehaene et Leboyer, 2006. Nous avons fait des observations semblables dans d'autres paradigmes expérimentaux (Sergent, Baillet et Dehaene, 2005 ; Sergent et Dehaene, 2004). Le caractère « tout-ou-rien » de la perception consciente continue d'être débattu (voir par exemple Overgaard, Rote, Mouridsen et Ramsoy, 2006). Une partie de la confusion vient du fait que les critiques ne distinguent pas suffisamment notre hypothèse d'un *accès* tout-ou-rien à un contenu donné (un chiffre, par exemple, qui peut être vu ou pas vu), du fait que le *contenu* de la conscience lui-même peut changer graduellement (on voit un trait, puis une lettre, puis le mot tout entier). Voir Kouider, de Gardelle, Sackur et Dupoux, 2010 ; Kouider et Dupoux, 2004.

36. Gaillard, Dehaene, Adam, Clemenceau, Hasboun, Baulac, Cohen et Naccache, 2009 ; Gaillard, Del Cul, Naccache, Vinckier, Cohen et Dehaene, 2006 ; Gaillard, Naccache, Pinel, Clemenceau, Volle, Hasboun, Dupont *et al.*, 2006.

37. Fisch, Privman, Ramot, Harel, Nir, Kipervasser, Andelman *et al.*, 2009 ; Quiroga, Mukamel, Isham, Malach et Fried, 2008 ; Kreiman, Fried et Koch, 2002.

38. Gaillard, Dehaene, Adam, Clemenceau, Hasboun, Baulac, Cohen et Naccache, 2009.

39. Fisch, Privman, Ramot, Harel, Nir, Kipervasser, Andelman *et al.*, 2009.

40. Gaillard, Dehaene, Adam, Clemenceau, Hasboun, Baulac, Cohen et Naccache, 2009 ; Fisch, Privman, Ramot, Harel, Nir, Kipervasser, Andelman *et al.*, 2009 ; Aru, Axmacher, Do Lam, Fell, Elger, Singer et Melloni, 2012.

41. Whittingstall et Logothetis, 2009 ; Fries, Nikolic et Singer, 2007 ; Cardin, Carlen, Meletis, Knoblich, Zhang, Deisseroth, Tsai et Moore, 2009 ; Buzsaki, 2006.

42. Fries, 2005.

43. Womelsdorf, Schoffelen, Oostenveld, Singer, Desimone, Engel et Fries, 2007 ; Fries, 2005 ; Varela, Lachaux, Rodriguez et Martinerie, 2001.

44. Rodriguez, George, Lachaux, Martinerie, Renault et Varela, 1999 ; Gaillard, Dehaene, Adam, Clemenceau, Hasboun, Baulac, Cohen et Naccache, 2009 ; Gross, Schmitz, Schnitzler, Kessler, Shapiro, Hommel et Schnitzler, 2004 ; Melloni, Molina, Pena, Torres, Singer et Rodriguez, 2007.

45. Varela, Lachaux, Rodriguez et Martinerie, 2001.

46. He, Snyder, Zempel, Smyth et Raichle, 2008 ; He, Zempel, Snyder et Raichle, 2010 ; Canolty, Edwards, Dalal, Soltani, Nagarajan, Kirsch, Berger *et al.*, 2006.

47. Gaillard, Dehaene, Adam, Clemenceau, Hasboun, Baulac, Cohen et Naccache, 2009.

48. Pins et Ffytche, 2003 ; Palva, Linkenkaer-Hansen, Naatanen et Palva, 2005 ; Fahrenfort, Scholte et Lamme, 2007 ; Railo et Koivisto, 2009 ; Koivisto, Lahteenmaki, Sorensen, Vangkilde, Overgaard et Revonsuo, 2008.

49. Van Aalderen-Smeets, Oosstenveld et Schwarzbach, 2006 ; Lamy, Salti et Bar-Haim, 2009.

50. Wyart, Dehaene et Tallon-Baudry, 2012.

51. Palva, Linkenkaer-Hansen, Naatanen et Palva, 2005 ; Wyart et Tallon-Baudry, 2009 ; Boly, Balteau, Schnakers, Degueldre, Moonen, Luxen, Phillips *et al.*, 2007 ; Supèr, Van der Togt, Spekreijse et Lamme, 2003 ; Sadaghiani, Hesselmann, Friston et Kleinschmidt, 2010.

52. Nieuwenhuis, Gilzenrat, Holmes et Cohen, 2005.

53. Certaines lésions du tronc cérébral, voisines du locus cœruleus, peuvent effectivement induire un coma ; voir Parvizi et Damasio, 2003.

54. Haynes, 2009.

55. Shady, MacLeod et Fisher, 2004 ; Krolak-Salmon, Henaff, Tallon-Baudry, Yvert, Guenot, Vighetto, Mauguiere et Bertrand, 2003.

56. MacLeod et He, 1993 ; He et MacLeod, 2001.

57. Quiroga, Kreiman, Koch et Fried, 2008 ; Quiroga, Mukamel, Isham, Malach et Fried, 2008.

58. Wyler, Ojemann et Ward, 1982 ; Heit, Smith et Halgren, 1988.

59. Fried, MacDonald et Wilson, 1997.

60. Quiroga, Kreiman, Koch et Fried, 2008 ; Quiroga, Mukamel, Isham, Malach et Fried, 2008 ; Quiroga, Reddy, Kreiman, Koch et Fried, 2005 ; Kreiman, Fried et Koch, 2002 ; Kreiman, 2000a ; Kreiman, 2000b.

61. Quiroga, Reddy, Kreiman, Koch et Fried, 2007.

62. Quiroga, Mukamel, Isham, Malach et Fried, 2008.

63. Kreiman, Fried et Koch, 2002. Cette recherche s'appuie sur les travaux pionniers de Nikos Logothetis et David Leopold chez le singe macaque (Leopold et Logothetis, 1996 ; Logothetis, Leopold et Sheinberg, 1996 ; Leopold et Logothetis, 1999).

64. Kreiman, Koch et Fried, 2000b.

65. Fisch, Privman, Ramot, Harel, Nir, Kipervasser, Andelman et *al.*, 2009.

66. Vogel, McCollough et Machizawa, 2005 ; Vogel et Machizawa, 2004.

67. Schurger, Pereira, Treisman et Cohen, 2009.

68. Dean et Platt, 2006.

69. Derdikman et Moser, 2010.

70. Jezek, Henriksen, Treves, Moser et Moser, 2011.

71. Peyrache, Khamassi, Benchenane, Wiener et Battaglia, 2009 ; Ji et Wilson, 2007 ; Louie et Wilson, 2001.

72. Horikawa, Tamaki, Miyawaki et Kamitani, 2013.

73. Thompson, 1910 ; Magnusson et Stevens, 1911.

74. Barker, Jalinous et Freeston, 1985 ; Pascual-Leone, Walsh et Rothwell, 2000 ; Hallett, 2000.

75. Selimbeyoglu et Parvizi, 2010 ; Parvizi, Jacques, Foster, Withoft, Rangarajan, Weiner et Grill-Spector, 2012.

76. Selimbeyoglu et Parvizi, 2010.

77. Blanke, Ortigue, Landis et Seeck, 2002.

78. Desmurget, Reilly, Richard, Szathmari, Mottolese et Sirigu, 2009.

79. Taylor, Walsh et Eimer, 2010.

80. Silvanto, Lavie et Walsh, 2005 ; Silvanto, Cowey, Lavie et Walsh, 2005.

81. Halelamien, Wu et Shimojo, 2007.

82. Silvanto et Cattaneo, 2010.

83. Lamme et Roelfsema, 2000.

84. Lamme, 2006.

85. Semir Zeki défend l'hypothèse d'une « désunion de la conscience » et spécule que chaque région du cerveau code pour un fragment distinct de « microconscience » (Zeki, 2003).

86. Edelman, 1987 ; Sporns, Tononi et Edelman, 1991.

87. Lamme et Roelfsema, 2000 ; Roelfsema, 2005.

88. Lamme, Zipser et Spekreijse, 1998 ; Pack et Born, 2001.

89. Koivisto, Railo et Salminen-Vaparanta, 2010 ; Koivisto, Mantyla et Silvanto, 2010.

90. Sur la cécité au changement, voir Beck, Muggleton, Walsh et Lavie, 2006. Sur la rivalité binoculaire : Carmel, Walsh, Lavie et Rees, 2010. Sur la cécité attentionnelle : Babiloni, Vecchio, Rossi, De Capua, Bartalini, Ulivelli et Rossini, 2007. Sur le clignement de l'attention : Kihara, Ikeda, Matsuyoshi, Hirose, Mima, Fukuyama et Osaka, 2010.

91. Kanai, Muggleton et Walsh, 2008.

92. Rounis, Maniscalco, Rothwell, Passingham et Lau, 2010. Mon opinion est que cette expérience pose d'importantes difficultés éthiques. Contrairement à la stimulation magnétique transcrânienne avec une impulsion unique, qui semble sans danger, la stimulation cérébrale intense, répétée et bilatérale qu'ont utilisée Rounis, Maniscalco, Rothwell, Passingham et Lau (2010) devrait être évitée en dehors d'un contexte thérapeutique. On prétend parfois que les effets de cette stimulation répétitive disparaissent dans l'heure qui suit, et pourtant certains psychiatres l'utilisent pour traiter la dépression, sur de longues durées, afin d'induire une rémission qui dure plusieurs semaines. Cette stimulation s'accompagne de changements détectables de l'anatomie cérébrale (May, Hajak, Ganssbauer, Steffens, Langguth, Kleinjung et Eichhammer, 2007). Dans l'état actuel des connaissances, je ne prêterais pas mon cerveau à ce genre d'expériences.

93. Carlen, Meletis, Siegle, Cardin, Futai, Vierling-Claassen, Ruhlmann *et al.*, 2011 ; Cardin, Carlen, Meletis, Knoblich, Zhang, Deisseroth, Tsai et Moore, 2009.

94. Adamantidis, Zhang, Aravanis, Deisseroth et de Lecea, 2007.

CHAPITRE 5

Théoriser la conscience

1. Dehaene, Kerszberg et Changeux, 1998 ; Dehaene, Changeux, Naccache, Sackur et Sergent, 2006 ; Dehaene et Naccache, 2001. Notre théorie descend en ligne direct de l'hypothèse de l'espace de travail global (*global workspace*) présentée en 1989 par Bernard Baars dans un livre important intitulé *A Cognitive Theory of Consciousness*. Mes collègues et moi avons donné corps à cette idée, sur le plan neuronal, en proposant que les connexions corticales à longue distance jouent un rôle fondamental dans l'espace de travail global (Dehaene, Kerszberg et Changeux, 1998).

2. Taine, 1870.

3. Dennett, 1991.

4. Dennett, 1978.

5. Broadbent, 1958.

6. Pashler, 1994.

7. Chun et Potter, 1995.

8. Shallice, 1972 ; Shallice, 1979 ; Posner et Snyder, 1975 ; Posner et Rothbart, 1998.

9. James, 1890.

10. Cette organisation hiérarchique, proposée dès le XIX^e^ siècle par le neurologue britannique John Hughling Jackson, fait à présent partie de la doxa en neurologie.

11. Van Gaal, Ridderinkhof, Fahrenfort, Scholte et Lamme, 2008 ; Van Gaal, Ridderinkhof, Scholte et Lamme, 2010.

12. Tsao, Freiwald, Tootell et Livingstone, 2006.

13. Dehaene et Naccache, 2001.

14. Denton, Shade, Zamarippa, Egan, Blair-West, McKinley, Lancaster et Fox, 1999.

15. Hagmann, Cammoun, Gigandet, Meuli, Honey, Wedeen et Sporns, 2008 ; Parvizi, Van Hoesen, Buckwalter et Damasio, 2006.

16. Goldman-Rakic, 1988.

17. Sherman, 2012.

18. Rigas et Castro-Alamancos, 2007.

19. Elston, 2003 ; Elston, 2000.

20. Elston, Benavides-Piccione et DeFelipe, 2001.

21. Konopka, Wexler, Rosen, Mukamel, Osborn, Chen, Lu *et al.*, 2012.

22. Enard, Przeworski, Fisher, Lai, Wiebe, Kitano, Monaco et Paabo, 2002.

23. Pinel, Fauchereau, Moreno, Barbot, Lathrop, Zelenika, Le Bihan *et al.*, 2012.

24. Lai, Fisher, Hurst, Vargha-Khadem et Monaco, 2001.

25. Enard, Gehre, Hammerschmidt, Holter, Blass, Somel, Bruckner *et al.*, 2009 ; Vernes, Oliver, Spiteri, Lockstone, Puliyadi, Taylor, Ho *et al.*, 2011.

26. Di Virgilio et Clarke, 1997.

27. Tononi et Edelman, 1998.

28. Hebb, 1949.

29. Tsunoda, Yamane, Nishizaki et Tanifuji, 2001.

30. Selfridge, 1959.

31. Felleman et Van Essen, 1991 ; Salin et Bullier, 1995.

32. Perin, Berger et Markram, 2011.

33. Hopfield, 1982 ; Ackley, Hinton et Sejnowski, 1985 ; Amit, 1989.

34. Crick, 2003 ; Koch et Crick, 2001.

35. Tononi, 2008. Giulio Tononi a proposé un formalisme mathématique qui définit précisément les concepts de différenciation et d'intégration, et conduit à une mesure quantitative de l'intégration informationnelle appellée Φ. Pour qu'un système soit conscient, selon Tononi, il faut et il suffit qu'il ait une valeur élevée de Φ : « La conscience *est* l'information intégrée. » Cette théorie me paraît problématique car elle conduit inévitablement au panpsychisme – l'idée que tout système connecté, qu'il s'agisse d'une colonie de bactéries ou d'une galaxie, possède un certain degré de conscience. Elle n'explique absolument pas pourquoi de nombreuses fonctions complexes et intégrées, telles que la compréhension des mots, peuvent se dérouler sans la moindre conscience dans le cerveau humain.

36. Meyer et Damasio, 2009 ; Damasio, 1989.

37. Edelman, 1987.

38. Friston, 2005 ; Kersten, Mamassian et Yuille, 2004.

39. Beck, Ma, Kiani, Hanks, Churchland, Roitman, Shadlen *et al.*, 2008.

40. Dehaene, Kerszberg et Changeux, 1998 ; Dehaene, Changeux, Naccache, Sackur et Sergent, 2006 ; Dehaene et Naccache, 2001 ; Dehaene, 2011.

41. Fries, 2005 ; Womelsdorf, Schoffelen, Oostenveld, Singer, Desimone, Engel et Fries, 2007 ; Buschman et Miller, 2007 ; Engel et Singer, 2001.

42. He et Raichle, 2009.

43. Rockstroh, Müller, Cohen et Elbert, 1992.

44. Vogel, McCollough et Machizawa, 2005 ; Vogel et Machizawa, 2004.

45. Dehaene et Changeux, 2005 ; Dehaene, Sergent et Changeux, 2003 ; Dehaene, Kerszberg et Changeux, 1998. Nos simulations se sont inspirées d'un modèle antérieur (Lumer, Edelman et Tononi, 1997a ; Lumer, Edelman et Tononi, 1997b), qui était cependant limité au cortex visuel précoce. Des simulations beaucoup plus réalistes de ces idées ont été ultérieurement conduites par Ariel Zylberberg et Mariano Sigman de l'Université de Buenos Aires (Zylberberg, Fernandez Slezak, Roelfsema, Dehaene et Sigman, 2010 ; Zylberberg, Dehaene, Mindlin et Sigman, 2009). Nancy Kopell et ses collègues de l'Université de Boston ont également développé des modèles neurophysiologiques détaillés de la dynamique corticale, capables de simuler les états de sommeil et d'anesthésie (Ching, Cimenser, Purdon, Brown et Kopell, 2010 ; McCarthy, Brown et Kopell, 2008).

46. Ariel Zylberberg a ensuite étendu ces simulations. Voir Zylberberg, Fernandez Slezak, Roelfsema, Dehaene et Sigman, 2010 ; Zylberberg, Dehaene, Mindlin et Sigman, 2009.

47. La littérature scientifique contient plusieurs propositions précises sur les transitions de phase qui pourraient correspondre à la conscience, la vigilance et l'accès à la conscience. Voir Steyn-Ross, Steyn-Ross et Sleigh, 2004 ; Breshears, Roland, Sharma, Gaona, Freudenburg, Tempelhoff, Avidan et Leuthardt, 2010 ; Jordan, Stockmanns, Kochs, Pilge et Schneider, 2008 ; Ching, Cimenser, Purdon, Brown et Kopell, 2010 ; Dehaene et Changeux, 2005.

48. Portas, Krakow, Allen, Josephs, Armony et Frith, 2000 ; Davis, Coleman, Absalom, Rodd, Johnsrude, Matta, Owen et Menon, 2007 ; Supp, Siegel, Hipp et Engel, 2011.

49. Tsodyks, Kenet, Grinvald et Arieli, 1999 ; Kenet, Bibitchkov, Tsodyks, Grinvald et Arieli, 2003.

50. He, Snyder, Zempel, Smyth et Raichle, 2008 ; Raichle, MacLeod, Snyder, Powers, Gusnard et Shulman, 2001 ; Raichle, 2010 ; Greicius, Krasnow, Reiss et Menon, 2003.

51. He, Snyder, Zempel, Smyth et Raichle, 2008 ; Boly, Tshibanda, Vanhaudenhuyse, Noirhomme, Schnakers, Ledoux, Boveroux *et al.*, 2009 ; Greicius, Kiviniemi, Tervonen, Vainionpaa, Alahuhta, Reiss et Menon, 2008 ; Vincent, Patel, Fox, Snyder, Baker, Van Essen, Zempel *et al.*, 2007.

52. Buckner, Andrews-Hanna et Schacter, 2008.

53. Mason, Norton, Van Horn, Wegner, Grafton et Macrae, 2007 ; Christoff, Gordon, Smallwood, Smith et Schooler, 2009.

54. Smallwood, Beach, Schooler et Handy, 2008.

55. Dehaene, 2005.

56. Sadaghiani, Hesselmann, Friston et Kleinschmidt, 2010.

57. Raichle, 2010.

58. Berkes, Orban, Lengyel et Fiser, 2011.

59. Changeux, Heidmann et Patte, 1984 ; Changeux et Danchin, 1976 ; Edelman, 1987 ; Changeux et Dehaene, 1989.

60. Dehaene et Changeux, 1997 ; Dehaene, Kerszberg et Changeux, 1998 ; Dehaene et Changeux, 1991.

61. Rougier, Noelle, Braver, Cohen et O'Reilly, 2005.

62. Dehaene, Changeux, Naccache, Sackur et Sergent, 2006.

63. *Ibid*.

64. Sergent, Baillet et Dehaene, 2005 ; Dehaene, Sergent et Changeux, 2003 ; Zylberberg, Fernandez Slezak, Roelfsema, Dehaene et Sigman, 2010 ; Zylberberg, Dehaene, Mindlin et Sigman, 2009.

65. Sergent, Wyart, Babo-Rebelo, Cohen, Naccache et Tallon-Baudry, 2012 ; Marti, Sigman et Dehaene, 2012.

66. Voir également Enns et Di Lollo, 2000 ; Di Lollo, Enns et Rensink, 2000.

67. Shady, MacLeod et Fisher, 2004 ; He et MacLeod, 2001.

68. Gilbert, Sigman et Crist, 2001.

69. Haynes, 2005a ; Haynes, 2005b ; Haynes, Sakai, Rees, Gilbert, Frith et Passingham, 2007.

70. Stettler, Das, Bennett, Gilbert, 2002.

71. Gaser et Schlaug, 2003 ; Bengtsson, Nagy, Skare, Forsman, Forssberg et Ullen, 2005.

72. Buckner et Koutstaal, 1998 ; Buckner, Andrews-Hanna et Schacter, 2008.

73. Sigala, Kusunoki, Nimmo-Smith, Gaffan et Duncan, 2008 ; Saga, Iba, Tanji, Hoshi, 2011 ; Shima, Isoda, Mushiake et Tanji, 2007 ; Fujii et Graybiel, 2003 ; pour une revue, voir Dehaene et Sigman, 2012.

74. Tyler et Marslen-Wilson, 2008 ; Griffiths, Marslen-Wilson, Stamatakis et Tyler, 2012 ; Pallier, Devauchelle et Dehaene, 2011 ; Saur, Schelter, Schnell, Kratochvil, Kupper, Kellmeyer, Kummerer *et al.,* 2010 ; Fedorenko, Duncan et Kanwisher, 2012.

75. Davis, Coleman, Absalom, Rodd, Johnsrude, Matta, Owen et Menon, 2007.

76. Beck, Ma, Kiani, Hanks, Churchland, Roitman, Shadlen *et al.*, 2008 ; Friston, 2005 ; Deneve, Latham et Pouget, 2001.

77. Yang et Shadlen, 2007.

78. Izhikevich et Edelman, 2008.

CHAPITRE 6

L'épreuve de vérité

1. Laureys, 2005.

2. Leon-Carrion, Van Eeckhout, Dominguez-Morales Mdel et Perez-Santamaria, 2002.

3. Schnakers, Vanhaudenhuyse, Giacino, Ventura, Boly, Majerus, Moonen et Laureys, 2009.

4. Smedira, Evans, Grais, Cohen, Lo, Cooke, Schecter *et al.*, 1990.

5. Laureys, Owen et Schiff, 2004.

6. Académie pontificale des sciences, 2008.

7. Alving, Moller, Sindrup et Nielsen, 1979 ; Grindal, Suter et Martinez, 1977 ; Westmoreland, Klass, Sharbrough et Reagan, 1975.

8. Hanslmayr, Gross, Klimesch et Shapiro, 2011 ; Capotosto, Babiloni, Romani et Corbetta, 2009.

9. Supp, Siegel, Hipp et Engel, 2011.

10. Jennett et Plum, 1972.

11. Jennett, 2002.

12. Giacino, 2005.

13. Giacino, Kezmarsky, DeLuca et Cicerone, 1991. Les neurologues utilisent aujourd'hui l'*Échelle révisée de récupération du coma* (*Coma Recovery Scale Revised* ou *CRS-R*), introduite par Giacino, Kalmar et Whyte, 2004. Cette batterie de tests ne cesse d'être discutée et améliorée. Voir par exemple Schnakers, Vanhaudenhuyse, Giacino, Ventura, Boly, Majerus, Moonen et Laureys, 2009.

14. Giacino, Kalmar et Whyte, 2004 ; Schnakers, Vanhaudenhuyse, Giacino, Ventura, Boly, Majerus, Moonen et Laureys, 2009.

15. Bruno, Bernheim, Ledoux, Pellas, Demertzi et Laureys, 2011. Voir aussi Laureys, 2005.

16. Owen, Coleman, Boly, Davis, Laureys et Pickard, 2006. Dans la mesure où cette patiente présentait toujours quelques réponses fluctuantes aux stimuli extérieurs, certains neurologues soutiennent qu'elle était en état de conscience minimale. Quoi qu'il en soit, le contraste est frappant entre la pauvreté de son comportement volontaire et l'étendue de ses activations cérébrales.

17. Voir par exemple Davis, Coleman, Absalom, Rodd, Johnsrude, Matta, Owen et Menon, 2007 ; Portas, Krakow, Allen, Josephs, Armony et Frith, 2000.

18. Naccache, 2006a ; Nachev et Husain, 2007 ; Greenberg, 2007.

19. Ropper, 2010. La phrase d'origine contient un jeu de mots qui n'est guère traduisible : « *Physicians and society are not ready for "I have brain activation, therefore I am". That would seriously put Descartes before the horse*. »

20. Owen, Coleman, Boly, Davis, Laureys, Jolles et Pickard, 2007.

21. Monti, Vanhaudenhuyse, Coleman, Boly, Pickard, Tshibanda, Owen et Laureys, 2010.

22. Cyranoski, 2012.

23. Le pionnier du décodage de l'EEG et des interfaces cerveau-machine est l'Allemand Neils Birbaumer, de l'Université de Tübingen. Pour une revue de ses travaux, voir Birbaumer, Mirguialday et Cohen, 2008.

24. Cruse, Chennu, Chatelle, Bekinschtein, Fernandez-Espejo, Pickard, Laureys et Owen, 2011.

25. Goldfine, Victor, Conte, Bardin et Schiff, 2012.

26. Goldfine, Victor, Conte, Bardin et Schiff, 2011.

27. Chatelle, Chennu, Noirhomme, Cruse, Owen et Laureys, 2012.

28. Hochberg, Bacher, Jarosiewicz, Masse, Simeral, Vogel, Haddadin *et al.*, 2012.

29. Brumberg, Nieto-Castanon, Kennedy et Guenther, 2010.

30. Squires, Squires et Hillyard, 1975 ; Squires, Wickens, Squires et Donchin, 1976.

31. Naatanen, Paavilainen, Rinne et Alho, 2007.

32. Wacongne, Changeux et Dehaene, 2012.

33. Bien que la réponse négative à la nouveauté (MMN) ne soit pas un marqueur de conscience, elle est très utile en clinique. En effet, les patients comateux dont le cerveau émet toujours une MMN ont de bien plus grandes chances de sortir du coma que les autres (Fischer, Luaute, Adeleine et Morlet, 2004 ; Kane, Curry, Butler et Cummins, 1993 ; Naccache, Puybasset, Gaillard, Serve et Willer, 2005).

34. Bekinschtein, Dehaene, Rohaut, Tadel, Cohen et Naccache, 2009.

35. *Ibid*.

36. Faugeras, Rohaut, Weiss, Bekinschtein, Galanaud, Puybasset, Bolgert *et al.*, 2012 ; Faugeras, Rohaut, Weiss, Bekinschtein, Galanaud, Puybasset, Bolgert *et al.*, 2011.

37. Friston, 2005 ; Wacongne, Labyt, Van Wassenhove, Bekinschtein, Naccache et Dehaene, 2011.

38. King, Faugeras, Gramfort, Schurger, El Karoui, Sitt, Wacongne *et al.*, 2013. Pour une approche similaire, voir également, Tzovara, Rossetti, Spierer, Grivel, Murray, Oddo et De Lucia, 2012.

39. Massimini, Ferrarelli, Huber, Esser, Singh et Tononi, 2005 ; Massimini, Boly, Casali, Rosanova et Tononi, 2009 ; Ferrarelli, Massimini, Sarasso, Casali, Riedner, Angelini, Tononi et Pearce, 2010.

40. Casali, Gosseries, Rosanova, Boly, Sarasso, Casali, Casarotto *et al.*, 2013.

41. Rosanova, Gosseries, Casarotto, Boly, Casali, Bruno, Mariotti *et al.*, 2012.

42. Laureys, 2005 ; Laureys, Lemaire, Maquet, Phillips et Franck, 1999.

43. Schiff, Ribary, Moreno, Beattie, Kronberg, Blasberg, Giacino *et al.*, 2002 ; Schiff, Ribary, Plum et Llinas, 1999.

44. Galanaud, Perlbarg, Gupta, Stevens, Sanchez, Tollard, de Champfleur *et al.*, 2012 ; Tshibanda, Vanhaudenhuyse, Galanaud, Boly, Laureys et Puybasset, 2009 ; Galanaud, Naccache et Puybasset, 2007.

45. King, Faugeras, Gramfort, Schurger, El Karoui, Sitt, Wacongne *et al.*, 2012.

46. Notre mesure, appelée « information symbolique mutuelle pondérée » (*weighted symbolic mutual information*) s'inspire de l'« entropie de transfert symbolique » (*symbolic transfer entropy*). Voir Staniek et Lehnertz, 2008.

47. Sitt, King, El Karoui, Rohaut, Faugeras, Gramfort, Cohen *et al.*, 2014.

48. La pondération des fréquences hautes et basses intervient dans le calcul de l'indice bispectral, une mesure commerciale qui prétend mesurer la profondeur de l'anesthésie. Pour un examen critique, voir par exemple Miller, Sleigh, Barnard et Steyn-Ross, 2004 ; Schnakers, Ledoux, Majerus, Damas, Damas, Lambermont, Lamy *et al.*, 2008.

49. Schiff, Giacino, Kalmar, Victor, Baker, Gerber, Fritz *et al.*, 2007. L'originalité de cette recherche a été mise en doute par Staunton (2008), parce que la stimulation cérébrale profonde a souvent été utilisée, dès les années 1960, chez les patients en coma ou en état végétatif (voir par exemple

Tsubokawa, Yamamoto, Katayama, Hirayama, Maejima et Moriya, 1990). Pour une réponse, voir Schiff, Giacino, Kalmar, Victor, Baker, Gerber, Fritz *et al.*, 2008.

50. Moruzzi et Magoun, 1949.

51. Shirvalkar, Seth, Schiff et Herrera, 2006.

52. Giacino, Fins, Machado et Schiff, 2012.

53. Schiff, Giacino, Kalmar, Victor, Baker, Gerber, Fritz *et al.*, 2007.

54. Voss, Uluc, Dyke, Watts, Kobylarz, McCandliss, Heier *et al.*, 2006. Voir aussi Sidaros, Engberg, Sidaros, Liptrot, Herning, Petersen, Paulson *et al.,* 2008.

55. Matsuda, Matsumura, Komatsu, Yanaka et Nose, 2003.

56. Giacino, Fins, Machado et Schiff, 2012.

57. Brefel-Courbon, Payoux, Ory, Sommet, Slaoui, Raboyeau, Lemesle *et al.*, 2007.

58. Cohen, Chaaban et Habert, 2004.

59. Schiff, 2010.

60. Striem-Amit, Cohen, Dehaene et Amedi, 2012.

CHAPITRE 7

L'avenir de la conscience

1. Tooley, 1983.

2. Tooley, 1972.

3. Singer, 1993.

4. Diamond et Doar, 1989 ; Diamond et Gilbert, 1989 ; Diamond et Goldman-Rakic, 1989.

5. Dubois, Dehaene-Lambertz, Perrin, Mangin, Cointepas, Duchesnay, Le Bihan et Hertz-Pannier, 2007 ; Jessica Dubois, recherche en cours au laboratoire Unicog, centre NeuroSpin, Gif-sur-Yvette, France.

6. Fransson, Skiold, Horsch, Nordell, Blennow, Lagercrantz et Aden, 2007 ; Doria, Beckmann, Arichi, Merchant, Groppo, Turkheimer, Counsell *et al.*, 2010 ; Lagercrantz et Changeux, 2010.

7. Mehler, Jusczyk, Lambertz, Halsted, Bertoncini et Amiel-Tison, 1988.

8. Dehaene-Lambertz, Dehaene et Hertz-Pannier, 2002 ; Dehaene-Lambertz, Hertz-Pannier et Dubois, 2006 ; Dehaene-Lambertz, Hertz-Pannier, Dubois, Meriaux, Roche, Sigman et Dehaene, 2006 ; Dehaene-Lambertz, Montavont, Jobert, Allirol, Dubois, Hertz-Pannier et Dehaene, 2009.

9. Dehaene-Lambertz, Montavont, Jobert, Allirol, Dubois, Hertz-Pannier et Dehaene, 2009.

10. Leroy, Glasel, Dubois, Hertz-Pannier, Thirion, Mangin et Dehaene-Lambertz, 2011.

11. Dehaene-Lambertz, Hertz-Pannier, Dubois, Meriaux, Roche, Sigman et Dehaene, 2006.

12. Davis, Coleman, Absalom, Rodd, Johnsrude, Matta, Owen et Menon, 2007.

13. Dehaene-Lambertz, Hertz-Pannier, Dubois, Meriaux, Roche, Sigman et Dehaene, 2006.

14. Basirat, Dehaene et Dehaene-Lambertz, 2012.

15. Johnson, Dziurawiec, Ellis et Morton, 1991.

16. Pour les expériences chez le bébé, voir : Gelskov et Kouider, 2010 ; Kouider, Stahlhut, Gelskov, Barbosa, Dutat, de Gardelle, Christophe *et al.*, 2013. L'expérience chez l'adulte, que j'ai décrite au chapitre 4, a été réalisée par Del Cul, Baillet et Dehaene, 2007.

17. Diamond et Doar, 1989.

18. de Haan et Nelson, 1999 ; Csibra, Kushnerenko et Grossman, 2008.

19. Nelson, Thomas, de Haan et Wewerka, 1998.

20. Dehaene-Lambertz et Dehaene, 1994.

21. Friederici, Friedrich et Weber, 2002.

22. Dubois, Dehaene-Lambertz, Perrin, Mangin, Cointepas, Duchesnay, Le Bihan et Hertz-Pannier, 2007.

23. Izard, Sann, Spelke et Streri, 2009.

24. Lagercrantz et Changeux, 2009.

25. Han, O'Tuathaigh, Van Trigt, Quinn, Fanselow, Mongeau, Koch et Anderson, 2003 ; Dos Santos Coura et Granon, 2012.

26. Bolhuis et Gahr, 2006.

27. Leopold et Logothetis, 1996.

28. Kovacs, Vogels et Orban, 1995 ; Macknik et Haglund, 1999.

29. Cowey et Stoerig, 1995.

30. Fuster, 2008.

31. Denys, Vanduffel, Fize, Nelissen, Sawamura, Georgieva, Vogels *et al.*, 2004.

32. Hasson, Nir, Levy, Fuhrmann et Malach, 2004.

33. Hayden, Smith et Platt, 2009.

34. Buckner, Andrews-Hanna et Schacter, 2008.

35. Mes collègues et moi poursuivons l'exploration du paradigme local-global chez le singe (en collaboration avec Lynn Uhrig et Bechir Jarraya) et chez la souris (avec Karim Benchenane et Catherine Wacongne).

36. Smith, Schull, Strote, McGee, Egnor et Erb, 1995.

37. Terrace et Son, 2009.

38. Hampton, 2001 ; Kornell, Son et Terrace, 2007 ; Kiani et Shadlen, 2009.

39. Kornell, Son et Terrace, 2007.

40. Nieuwenhuis, Ridderinkhof, Blom, Band et Kok, 2001 ; Logan et Crump, 2010 ; Charles, Van Opstal, Marti et Dehaene, 2013.

41. Kiani et Shadlen, 2009 ; Fleming, Weil, Nagy, Dolan et Rees, 2010. Un noyau spécifique du thalamus, le pulvinar, étroitement connecté aux régions préfrontales et pariétales, semble également jouer un rôle essentiel dans les jugements métacognitifs (Komura, Nikkuni, Hirashima, Uetake et Miyamoto, 2013).

42. Meltzoff et Brooks, 2008 ; Kovacs, Teglas et Endress, 2010.

43. Herrmann, Call, Hernandez-Lloreda, Hare et Tomasello, 2007.

44. Marticorena, Ruiz, Mukerji, Goddu et Santos, 2011.

45. Marticorena, Ruiz, Mukerji, Goddu et Santos, 2011.

46. Fuster, 2008.

47. Elston, Benavides-Piccione et DeFelipe, 2001 ; Elston, 2003.

48. Ochsner, Knierim, Ludlow, Hanelin, Ramachandran, Glover et Mackey, 2004 ; Saxe et Powell, 2006 ; Fleming, Weil, Nagy, Dolan et Rees, 2010.

49. Schoenemann, Sheehan et Glotzer, 2005.

50. Schenker, Buxhoeveden, Blackmon, Amunts, Zilles et Semendeferi, 2008 ; Schenker, Hopkins, Spocter, Garrison, Stimpson, Erwin, Hof et Sherwood, 2009.

51. Nimchinsky, Gilissen, Allman, Perl, Erwin et Hof, 1999 ; Allman, Hakeem et Watson, 2002 ; Allman, Watson, Tetreault et Hakeem, 2005.

52. Dehaene et Changeux, 2011.

53. Frith, 1979 ; Frith, 1996 ; Stephan, Friston et Frith, 2009.

54. Huron, Danion, Giacomoni, Grange, Robert et Rizzo, 1995 ; Danion, Meulemans, Kauffmann-Muller et Vermaat, 2001 ; Danion, Cuervo, Piolino, Huron, Riutort, Peretti et Eustache, 2005.

55. Dehaene, Artiges, Naccache, Martelli, Viard, Schurhoff, Recasens *et al.*, 2003 ; Del Cul, Dehaene et Leboyer, 2006. Notre travail s'est focalisé sur la dissocation entre le traitement subliminal et l'accès à la conscience. Pour une revue des travaux antérieurs sur les anomalies du masquage dans la schizophrénie, voir McClure, 2001.

56. Reuter, Del Cul, Audoin, Malikova, Naccache, Ranjeva, Lyon-Caen *et al.*, 2007.

57. Reuter, Del Cul, Malikova, Naccache, Confort-Gouny, Cohen, Cherif *et al.*, 2009.

58. Luck, Fuller, Braun, Robinson, Summerfelt et Gold, 2006 ; Luck, Kappenman, Fuller, Robinson, Summerfelt et Gold, 2009 ; Antoine Del Cul, Stanislas Dehaene, Marion Leboyer *et al.*, expériences non publiées.

59. Uhlhaas, Linden, Singer, Haenschel, Lindner, Maurer et Rodriguez, 2006 ; Uhlhaas et Singer, 2010.

60. Kubicki, Park, Westin, Nestor, Mulkern, Maier, Niznikiewicz *et al.*, 2005 ; Karlsgodt, Sun, Jimenez, Lutkenhoff, Willhite, Van Erp et Cannon, 2008 ; Knochel, Oertel-Knochel, Schonmeyer, Rotarska-Jagiela, Van de Ven, Prvulovic, Haenschel *et al.*, 2012.

61. Bassett, Bullmore, Verchinski, Mattay, Weinberger et Meyer-Lindenberg, 2008 ; Liu, Liang, Zhou, He, Hao, Song, Yu *et al.*, 2008 ; Bassett, Bullmore, Meyer-Lindenberg, Apud, Weinberger et Coppola, 2009 ; Lynall, Bassett, Kerwin, McKenna, Kitzbichler, Muller et Bullmore, 2010.

62. Ross, Margolis, Reading, Pletnikov et Coyle, 2006 ; Dickman et Davis, 2009 ; Tang, Yang, Chen, Lu, Ji, Roche et Lu, 2009 ; Shao, Shuai, Wang, Feng, Lu, Li, Zhao *et al.*, 2011.

63. Self, Kooijmans, Supèr, Lamme et Roelfsema, 2010.

64. Dehaene, Sergent et Changeux, 2003 ; Dehaene et Changeux, 2005.

65. Wong et Wang, 2006.

66. Fletcher et Frith, 2009 ; voir aussi Stephan, Friston et Frith, 2009.

67. Friston, 2005.

68. Dalmau, Tuzun, Wu, Masjuan, Rossi, Voloschin, Baehring *et al.*, 2007 ; Dalmau, Gleichman, Hughes, Rossi, Peng, Lai, Dessain *et al.*, 2008.

69. Block, 2001 ; Block, 2007.

70. Chalmers,1996.

71. Chalmers, 1995, p. 81.

72. Weiss, Simoncelli et Adelson, 2002.

73. Lucrèce, *De rerum natura* (*De la nature des choses*), livre 2.

74. Eccles, 1994.

75. Penrose et Hameroff, 1998.

76. Dennett, 1984.

77. Edelman, 1989.

Bibliographie

ABRAMS, R. L. et A. G. GREENWALD, 2000, « Parts Outweigh the Whole (Word) in Unconscious Analysis of Meaning », *Psychological Science*, 11 (2) : 118-124.

ABRAMS, R. L., M. R. KLINGER et A. G. GREENWALD, 2002, « Subliminal Words Activate Semantic Categories (Not Automated Motor Responses) », *Psychonomic Bulletin and Review*, 9 (1) : 100-106.

ACADÉMIE PONTIFICALE DES SCIENCES, 2008, *Why the Concept of Death Is Valid as a Definition of Brain Death. Statement by the Pontifical Academy of Sciences and Responses to Objections*, http://www.pas.va/content/accademia/en/publications/extraseries/br

ACKLEY, D. H., G. E. HINTON, T. SEJNOWSKI, G. E. HINTON et T. J. SEJNOWSKI, 1985, « A Learning Algorithm for Boltzmann Machines », *Cognitive Science*, 9 : 147-169.

ADAMANTIDIS, A. R., F. ZHANG, A. M. ARAVANIS, K. DEISSEROTH et L. DE LECEA, 2007, « Neural Substrates of Awakening Probed with Optogenetic Control of Hypocretin Neurons », *Nature*, 450 (7168) : 420-424.

ALLMAN, J., A. HAKEEM et K. WATSON, 2002, « Two Phylogenetic Specializations in the Human Brain », *Neuroscientist*, 8 (4) : 335-346.

ALLMAN, J. M., K. K. WATSON, N. A. TETREAULT et A. Y. HAKEEM, 2005, « Intuition and Autism : A Possible Role for Von Economo Neurons », *Trends in Cognitive Sciences*, 9 (8) : 367-373.

ALMEIDA, J., B. Z. MAHON, K. NAKAYAMA et A. CARAMAZZA, 2008, « Unconscious Processing Dissociates Along Categorical Lines », *Proceedings of the National Academy of Sciences*, 105 (39) : 15214-15218.

ALVING, J., M. MOLLER, E. SINDRUP et B. L. NIELSEN, 1979, « "Alpha Pattern Coma" Following Cerebral Anoxia », *Electroencephalography and Clinical Neurophysiology*, 47 (1) : 95-101.

AMIT, D., 1989, *Modeling Brain Function : The World of Attractor Neural Networks,* New York, Cambridge University Press.

ANDERSON, J. R., 1983, *The Architecture of Cognition*, Cambridge, Mass., Harvard University Press.

ANDERSON, J. R et C. LEBIERE, 1998, *The Atomic Components of Thought*, Mahwah, NJ, Lawrence Erlbaum.

ARU, J., N. AXMACHER, A. T. DO LAM, J. FELL, C. E. ELGER, W. SINGER et L. MELLONI, 2012, « Local Category-Specific Gamma Band Responses in the Visual Cortex Do Not Reflect Conscious Perception », *Journal of Neuroscience*, 32 (43) : 14909-14914.

ARU, J., T. BACHMANN, W. SINGER et L. MELLONI, 2012, « Distilling the Neural Correlates of Consciousness », *Neurosci. Biobehav. Rev.*, 36 (2), 737-746.

ASHCRAFT, M. H. et E. H. STAZYK, 1981, « Mental Addition : A Test of Three Verification Models », *Memory and Cognition*, 9, 185-196.

BAARS, B. J., 1989, *A Cognitive Theory of Consciousness*, Cambridge, R.-U., Cambridge University Press.

BABILONI, C., F. VECCHIO, S. ROSSI, A. DE CAPUA, S. BARTALINI, M. ULIVELLI et P. M. ROSSINI, 2007, « Human Ventral Parietal Cortex Plays a Functional Role on Visuospatial Attention and Primary Consciousness : A Repetitive Transcranial Magnetic Stimulation Study », *Cerebral Cortex*, 17 (6) : 1486-1492.

BAHRAMI, B., K. OLSEN, P. E. LATHAM, A. ROEPSTORFF, G. REES et C. D. FRITH, 2010, « Optimally Interacting Minds », *Science*, 329 (5995) : 1081-

1085.

Baker, C., M. Behrmann et C. Olson, 2002, « Impact of Learning on Representation of Parts and Wholes in Monkey Inferotemporal Cortex », *Nature Neuroscience*, 5 (11) : 1210-1216.

Bargh, J. A. et E. Morsella, 2008, « The Unconscious Mind », *Perspectives on Psychological Science*, 3 (1) : 73-79.

Barker, A. T., R. Jalinous et I. L. Freeston, 1985, « Non-Invasive Magnetic Stimulation of Human Motor Cortex », *The Lancet*, 1 (8437) : 1106-1107.

Basirat, A., S. Dehaene et G. Dehaene-Lambertz, 2012, « A Hierarchy of Cortical Responses to Sequence Violations in Two-Month-Old Infants » *Cognition,* 132 (2) : 137-150.

Bassett, D. S., E. Bullmore, B. A. Verchinski, V. S. Mattay, D. R. Weinberger et A. Meyer-Lindenberg, 2008, « Hierarchical Organization of Human Cortical Networks in Health and Schizophrenia », *Journal of Neuroscience*, 28 (37) : 9239-9248.

Bassett, D. S., E. T. Bullmore, A. Meyer-Lindenberg, J. A. Apud, D. R. Weinberger et R. Coppola, 2009, « Cognitive Fitness of Cost-Efficient Brain Functional Networks », *Proceedings of the National Academy of Sciences*, 106 (28) : 11747-11752.

Batterink, L. et H. J. Neville, 2013, « The Human Brain Processes Syntax in the Absence of Conscious Awareness », *Journal of Neuroscience*, 33 (19) : 8528-8533.

Bechara, A., H. Damasio, D. Tranel et A. R. Damasio, 1997, « Deciding Advantageously Before Knowing the Advantageous Strategy », *Science*, 275 (5304) : 1293-1295.

Beck, D. M., N. Muggleton, V. Walsh et N. Lavie, 2006, « Right Parietal Cortex Plays a Critical Role in Change Blindness », *Cerebral Cortex*, 16 (5) : 712-717.

Beck, D. M., G. Rees, C. D. Frith et N. Lavie, 2001, « Neural Correlates of Change Detection and Change Blindness », *Nature Neuroscience*, 4,

645-650.

Beck, J. M., W. J. Ma, R. Kiani, T. Hanks, A. K. Churchland, J. Roitman, M. N. Shadlen *et al.*, 2008, « Probabilistic Population Codes for Bayesian Decision Making », *Neuron*, 60 (6) : 1142-1152.

Bekinschtein, T. A., S. Dehaene, B. Rohaut, F. Tadel, L. Cohen et L. Naccache, 2009, « Neural Signature of the Conscious Processing of Auditory Regularities », *Proceedings of the National Academy of Sciences* 106 (5) : 1672-1677.

Bekinschtein, T. A., M. Peeters, D. Shalom et M. Sigman, 2011, « Sea Slugs, Subliminal Pictures, and Vegetative State Patients : Boundaries of Consciousness in Classical Conditioning », *Frontiers in Psychology*, 2, 337.

Bekinschtein, T. A., D. E. Shalom, C. Forcato, M. Herrera, M. R. Coleman, F. F. Manes et M. Sigman, 2009, « Classical Conditioning in the Vegetative and Minimally Conscious State », *Nature Neuroscience*, 12 (10) : 1343-1349.

Bengtsson, S. L., Z. Nagy, S. Skare, L. Forsman, H. Forssberg et F. Ullen, 2005, « Extensive Piano Practicing Has Regionally Specific Effects on White Matter Development », *Nature Neuroscience*, 8 (9) : 1148-1150.

Berkes, P., G. Orban, M. Lengyel et J. Fiser, 2011, « Spontaneous Cortical Activity Reveals Hallmarks of an Optimal Internal Model of the Environment », *Science*, 331 (6013) : 83-87.

Birbaumer, N., A. R. Murguialday et L. Cohen, 2008, « Brain-computer Interface in Paralysis », *Current Opinion in Neurology*, 21 (6) : 634-638.

Bisiach, E., C. Luzzatti et D. Perani, 1979, « Unilateral Neglect Representational Schema and Consciousness », *Brain*, 102, 609-618.

Blanke, O., T. Landis, L. Spinelli et M. Seeck, 2004, « Out-of-body Experience and Autoscopy of Neurological Origin », *Brain*, 127 (Pt 2) : 243-258.

BLANKE, O., S. ORTIGUE, T. LANDIS et M. SEECK, 2002, « Stimulating Illusory Own-Body Perceptions », *Nature*, 419 (6904), 269-270.

BLOCK, N., 2001, « Paradox and Cross Purposes in Recent Work on Consciousness », *Cognition*, 79 (1-2) : 197-219.

–, 2007, « Consciousness, Accessibility, and the Mesh Between Psychology and Neuroscience », *Behavioral and Brain Sciences*, 30 (5-6) : 481-499 ; discussion : 499-548.

BOLHUIS, J. J. et M. GAHR, 2006, « Neural Mechanisms of Birdsong Memory », *Nature Reviews Neuroscience*, 7 (5) : 347-357.

BOLY, M., E. BALTEAU, C. SCHNAKERS, C. DEGUELDRE, G. MOONEN, A. LUXEN, C. PHILLIPS *et al.*, 2007, « Baseline Brain Activity Fluctuations Predict Somatosensory Perception in Humans », *Proceedings of the National Academy of Sciences*, 104 (29) : 12187-12192.

BOLY, M., L. TSHIBANDA, A. VANHAUDENHUYSE, Q. NOIRHOMME, C. SCHNAKERS, D. LEDOUX, P. BOVEROUX *et al.*, 2009, « Functional Connectivity in the Default Network During Resting State Is Preserved in a Vegetative But Not in a Brain Dead Patient », *Human Brain Mapping*.

Botvinick, M. et J. Cohen, 1998, « Rubber Hands “Feel” Touch That Eyes See », *Nature*, 391 (6669) : 756.

BOWERS, J. S., G. VIGLIOCCO et R. HAAN, 1998, « Orthographic, Phonological, and Articulatory Contributions to Masked Letter and Word Priming », *Journal of Experimental Psychology : Human Perception and Performance*, 24 (6) : 1705-1719.

BRASCAMP, J. W. et R. BLAKE, 2012, « Inattention Abolishes Binocular Rivalry : Perceptual Evidence », *Psychological Science*, 23 (10) : 1159-1167.

BREFEL-COURBON, C., P. PAYOUX, F. ORY, A. SOMMET, T. SLAOUI, G. RABOYEAU, B. LEMESLE *et al.*, 2007, « Clinical and Imaging Evidence of Zolpidem Effect in Hypoxic Encephalopathy », *Annals of Neurology*, 62 (1) : 102-105.

Breitmeyer, B. G., A. Koc, H. Ogmen et R. Ziegler, 2008, « Functional Hierarchies of Nonconscious Visual Processing », *Vision Research*, 48 (14) : 1509-1513.

Breshears, J. D., J. L. Roland, M. Sharma, C. M. Gaona, Z. V. Freudenburg, R. Tempelhoff, M. S. Avidan et E. C. Leuthardt, 2010, « Stable and Dynamic Cortical Electrophysiology of Induction and Emergence with Propofol Anesthesia », *Proceedings of the National Academy of Sciences*, 107 (49) : 21170-21175.

Bressan, P. et S. Pizzighello, 2008, « The Attentional Cost of Inattentional Blindness », *Cognition*, 106 (1) : 370-383.

Brincat, S. L. et C. E. Connor, 2004, « Underlying Principles of Visual Shape Selectivity in Posterior Inferotemporal Cortex », *Nature Neuroscience*, 7 (8) : 880-886.

Broadbent, D. E., 1958, *Perception and Communication*, Londres, Pergamon.

–, 1962, « Attention and the Perception of Speech », *Scientific American*, 206 (4) : 143-151.

Brumberg, J. S., A. Nieto-Castanon, P. R. Kennedy et F. H. Guenther, 2010, « Brain-Computer Interfaces for Speech Communication », *Speech Communication*, 52 (4) : 367-379.

Bruno, M. A., J. L. Bernheim, D. Ledoux, F. Pellas, A. Demertzi et S. Laureys, 2011, « A Survey on Self-assessed Well-being in a Cohort of Chronic Locked-in Syndrome Patients : Happy Majority, Miserable Minority », *BMJ Open*, 1 (1) : e000039.

Buckner, R. L., J. R. Andrews-Hanna et D. L. Schacter, 2008. « The Brain's Default Network : Anatomy, Function, and Relevance to Disease », *Annals of the New York Academy of Sciences*, 1124 : 1-38.

Buckner, R. L. et W. Koutstaal, 1998, « Functional Neuroimaging Studies of Encoding, Priming, and Explicit Memory Retrieval », *Proceedings of the National Academy of Sciences*, 95 (3) : 891-898.

BUSCHMAN, T. J. et E. K. MILLER, 2007, « Top-down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices », *Science*, 315 (5820) : 1860-1862.

BUZSAKI, G., 2006, *Rhythms of the Brain*, New York, Oxford University Press.

CANOLTY, R. T., E. EDWARDS, S. S. DALAL, M. SOLTANI, S. S. NAGARAJAN, H. E. KIRSCH, M. S. BERGER *et al.*, 2006, « High Gamma Power Is Phase-Locked to Theta Oscillations in Human Neocortex », *Science*, 313 (5793) : 1626-1628.

CAPOTOSTO, P., C. BABILONI, G. L. ROMANI et M. CORBETTA, 2009, « Frontoparietal Cortex Controls Spatial Attention Through Modulation of Anticipatory Alpha Rhythms », *Journal of Neuroscience* 29 (18) : 5863-5872.

CARDIN, J. A., M. CARLEN, K. MELETIS, U. KNOBLICH, F. ZHANG, K. DEISSEROTH, L. H. TSAI et C. I. MOORE, 2009, « Driving Fast-spiking Cells Induces Gamma Rhythm and Controls Sensory Responses », *Nature*, 459 (7247) : 663-667.

CARLEN, M., K. MELETIS, J. H. SIEGLE, J. A. CARDIN, K. FUTAI, D. VIERLING-CLAASSEN, C. RUHLMANN *et al.*, 2011, « A Critical Role for NMDA Receptors in Parvalbumin Interneurons for Gamma Rhythm Induction and Behavior », *Molecular Psychiatry*, 17 (5) : 537-548.

CARMEL, D., V. WALSH, N. LAVIE et G. REES, 2010, « Right Parietal TMS Shortens Dominance Durations in Binocular Rivalry », *Current Biology*, 20 (18) : R799-800.

CARTER, R. M., C. HOFSTOTTER, N. TSUCHIYA et C. KOCH, 2003, « Working Memory and Fear Conditioning », *Proceedings of the National Academy of Sciences*, 100 (3) : 1399-1404.

CARTER, R. M., J. P. O'DOHERTY, B. SEYMOUR, C. KOCH et R. J. DOLAN, 2006, « Contingency Awareness in Human Aversive Conditioning Involves the Middle Frontal Gyrus », *NeuroImage*, 29 (3) : 1007-1012.

CASALI, A., O. GOSSERIES, M. ROSANOVA, M. BOLY, S. SARASSO, K. R. CASALI, S. CASAROTTO *et al.*, 2013, « A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior », *Science Translational Medicine,* 5 (198) : 198.

CHALMERS, D., 1996, *The Conscious Mind*, New York, Oxford University Press.

CHALMERS, D. J., 1995, « The Puzzle of Conscious Experience », *Scientific American*, 273 (6) : 80-86.

CHANGEUX, J.-P., 1983, *L'Homme neuronal*, Paris, Fayard.

CHANGEUX, J.-P. et A. DANCHIN, 1976, « Selective Stabilization of Developing Synapses as a Mechanism for the Specification of Neuronal Networks », *Nature*, 264 : 705-712.

CHANGEUX, J.-P. et S. DEHAENE, 1989, « Neuronal Models of Cognitive Functions », *Cognition*, 33 (1-2) : 63-109.

CHANGEUX, J.-P., T. HEIDMANN et P. PATTE, 1984, « Learning by Selection », *in* P. Marler et H. S. Terrace (éd.), *The Biology of Learning,* Berlin, Springer : 115-139.

CHARLES, L., F. VAN OPSTAL, S. MARTI et S. DEHAENE, 2013, « Distinct Brain Mechanisms for Conscious Versus Subliminal Error Detection », *NeuroImage,* 73 : 80-94.

CHATELLE, C., S. CHENNU, Q. NOIRHOMME, D. CRUSE, A. M. OWEN et S. LAUREYS, 2012, « Brain-computer Interfacing in Disorders of Consciousness », *Brain Injury*, 26 (12) : 1510-1522.

CHEIN, J. M. et W. SCHNEIDER, 2005, « Neuroimaging Studies of Practice-Related Change : fMRI and Meta-analytic Evidence of a Domain-general Control Network for Learning », *Brain Research : Cognitive Brain Research*, 25 (3) : 607-623.

CHING, S., A. CIMENSER, P. L. PURDON, E. N. BROWN et N. J. KOPELL, 2010, « Thalamocortical Model for a Propofol-induced Alpha-rhythm Associated with Loss of Consciousness », *Proceedings of the National Academy of Sciences*, 107 (52) : 22665-22670.

CHONG, S. C. et R. BLAKE, 2006, « Exogenous Attention and Endogenous Attention Influence Initial Dominance in Binocular Rivalry », *Vision Research*, 46 (11) : 1794-1803.

CHONG, S. C., D. TADIN et R. BLAKE, 2005, « Endogenous Attention Prolongs Dominance Durations in Binocular Rivalry », *Journal of Vision*, 5 (11) : 1004-1012.

CHRISTOFF, K., A. M. GORDON, J. SMALLWOOD, R. SMITH et J. W. SCHOOLER, 2009, « Experience Sampling During Fmri Reveals Default Network and Executive System Contributions to Mind Wandering », *Proceedings of the National Academy of Sciences*, 106 (21) : 8719-8724.

CHUN, M. M. et M. C. POTTER, 1995, « A Two-stage Model for Multiple Target Detection in Rapid Serial Visual Presentation », *Journal of Experimental Psychology : Human Perception and Performance*, 21 (1) : 109-127.

CHURCHLAND, P. S., 1986, *Neurophilosophy : Toward a Unified Understanding of the Mind/Brain*, Cambridge, Mass., MIT Press.

CLARK, R. E., J. R. MANNS et L. R. SQUIRE, 2002, « Classical Conditioning, Awareness, and Brain Systems », *Trends in Cognitive Sciences*, 6 (12) : 524-531.

CLARK, R. E. et L. R. SQUIRE, 1998, « Classical Conditioning and Brain Systems : The Role of Awareness », *Science*, 280 (5360) : 77-81.

COHEN, L., B. CHAABAN et M. O. HABERT, 2004, « Transient Improvement of Aphasia with Zolpidem », *New England Journal of Medicine*, 350 (9) : 949-950.

Cohen, M. A., P. Cavanagh, M. M. Chun et K. Nakayama, 2012, « The Attentional Requirements of Consciousness », *Trends in Cognitive Sciences*, 16 (8) : 411-417.

COMTE, A., 1830-1842, *Cours de philosophie positive*, Paris, Bachelier.

CORALLO, G., J. SACKUR, S. DEHAENE et M. SIGMAN, 2008, « Limits on Introspection : Distorted Subjective Time During the Dual-task Bottleneck », *Psychological Science*, 19 (11) : 1110-1117.

COWEY, A. et P. STOERIG, 1995, « Blindsight in Monkeys », *Nature*, 373 (6511) : 247-249.

CRICK, F. et C. KOCH, 1990a, « Some Reflections on Visual Awareness », *Cold Spring Harbor Symposia on Quantitative Biology*, 55 : 953-962.

–, 1990b, « Toward a Neurobiological Theory of Consciousness », *Seminars in Neuroscience*, 2 : 263-275.

–, 2003, « A Framework for Consciousness », *Nature Neuroscience*, 6 (2) : 119-126.

CRUSE, D., S. CHENNU, C. CHATELLE, T. A. BEKINSCHTEIN, D. FERNANDEZ-ESPEJO, J. D. PICKARD, S. LAUREYS et A. M. OWEN, 2011, « Bedside Detection of Awareness in the Vegetative State : A Cohort Study », *The Lancet*, 378 (9809) : 2088-2094.

CSIBRA, G., E. KUSHNERENKO et T. GROSSMAN, 2008, « Electrophysiological Methods in Studying Infant Cognitive Development », *in* C. A. Nelson et M. Luciana (éd.), *Handbook of Developmental Cognitive Neuroscience*, Cambridge, Mass., MIT Press, 2e éd.

CYRANOSKI, D., 2012, « Neuroscience : The Mind Reader », *Nature*, 486 (7402) : 178-180.

DALMAU, J., A. J. GLEICHMAN, E. G. HUGHES, J. E. ROSSI, X. PENG, M. LAI, S. K. DESSAIN *et al.*, 2008, « Anti-NMDA-receptor Encephalitis : Case Series and Analysis of the Effects of Antibodies », *The Lancet Neurology*, 7 (12) : 1091-1098.

DALMAU, J., E. TUZUN, H. Y. WU, J. MASJUAN, J. E. ROSSI, A. VOLOSCHIN, J. M. BAEHRING *et al.*, 2007, « Paraneoplastic Anti-N-methyl-D-aspartate Receptor Encephalitis Associated with Ovarian Teratoma », *Annals of Neurology*, 61 (1) : 25-36.

DAMASIO, A. R., 1989, « The Brain Binds Entities and Events by Multiregional Activation from Convergence Zones », *Neural Computation*, 1 : 123-132.

–, 1997, *L'Erreur de Descartes. La raison des émotions*, Paris, Odile Jacob.

DANION, J. M., C. CUERVO, P. PIOLINO, C. HURON, M. RIUTORT, C. S. PERETTI et F. EUSTACHE, 2005, « Conscious Recollection in Autobiographical Memory : An Investigation in Schizophrenia », *Consciousness and Cognition*, 14 (3) : 535-547.

DANION, J. M., T. MEULEMANS, F. KAUFFMANN-MULLER et H. VERMAAT, 2001, « Intact Implicit Learning in Schizophrenia », *American Journal of Psychiatry*, 158 (6) : 944-948.

DAVIS, M. H., M. R. COLEMAN, A. R. ABSALOM, J. M. RODD, I. S. JOHNSRUDE, B. F. Matta, A. M. OWEN et D. K. MENON, 2007, « Dissociating Speech Perception and Comprehension at Reduced Levels of Awareness », *Proceedings of the National Academy of Sciences*, 104 (41) : 16032-16037.

DE GROOT, A. D. et F. GOBET, 1996, *Perception and Memory in Chess*, Assen, Pays-Bas, Van Gorcum.

DE HAAN, M. et C. A. NELSON, 1999, « Brain Activity Differentiates Face and Object Processing in 6-Month-Old Infants », *Developmental Psychology*, 35 (4) : 1113-1121.

DE LANGE, F. P., S. VAN GAAL, V. A. LAMME et S. DEHAENE, 2011, « How Awareness Changes the Relative Weights of Evidence During Human Decision-Making », *PLOS Biology*, 9 (11) : e1001203.

DEAN, H. L. et M. L. PLATT, 2006, « Allocentric Spatial Referencing of Neuronal Activity in Macaque Posterior Cingulate Cortex », *Journal of Neuroscience* 26 (4) : 1117-1127.

DEHAENE, S., 2008, « Conscious and Nonconscious Processes : Distinct Forms of Evidence Accumulation ? », *in* C. Engel et W. Singer (éd.), *Better Than Conscious ? Decision Making, the Human Mind, and Implications for Institutions. Strüngmann Forum Report*, Cambridge, Mass., MIT Press.

–, 2009, *Les Neurones de la lecture*, Paris, Odile Jacob.

–, 2011, *La Bosse des maths, quinze ans après*, Paris, Odile Jacob.

DEHAENE, S., E. ARTIGES, L. NACCACHE, C. MARTELLI, A. VIARD, F. SCHURHOFF, C. RECASENS *et al.*, 2003, « Conscious and Subliminal Conflicts in Normal Subjects and Patients with Schizophrenia : The Role of the Anterior Cingulate », *Proceedings of the National Academy of Sciences*, 100 (23) : 13722-13727.

DEHAENE, S. et J.-P. CHANGEUX, 1991, « The Wisconsin Card Sorting Test : Theoretical Analysis and Modelling in a Neuronal Network », *Cerebral Cortex*, 1 : 62-79.

–, 1997, « A Hierarchical Neuronal Network for Planning Behavior », *Proceedings of the National Academy of Sciences*, 94 (24) : 13293-13298.

–, 2005, « Ongoing Spontaneous Activity Controls Access to Consciousness : A Neuronal Model for Inattentional Blindness », *PLOS Biology* 3 (5) : e141.

–, 2011, « Experimental and Theoretical Approaches to Conscious Processing », *Neuron* 70 (2) : 200-227.

DEHAENE, S., J.-P. CHANGEUX, L. NACCACHE, SACKUR, J. et C. SERGENT, 2006, « Conscious, Preconscious, and Subliminal Processing : A Testable Taxonomy », *Trends in Cognitive Sciences*, 10 (5) : 204-211.

DEHAENE, S. et L. COHEN, 2007, « Cultural Recycling of Cortical Maps », *Neuron*, 56 (2) : 384-398.

DEHAENE, S., A. JOBERT, L. NACCACHE, P. CIUCIU, J. B. POLINE, D. LE BIHAN et L. COHEN, 2004, « Letter Binding and Invariant Recognition of Masked Words : Behavioral and Neuroimaging Evidence », *Psychological Science*, 15 (5) : 307-313.

DEHAENE, S., M. KERSZBERG et J.-P. CHANGEUX, 1998a, « A Neuronal Model of a Global Workspace in Effortful Cognitive Tasks », *Proceedings of the National Academy of Sciences*, 95 (24) : 14529-14534.

DEHAENE, S. et L. NACCACHE, 2001, « Towards a Cognitive Neuroscience of Consciousness : Basic Evidence and a Workspace Framework », *Cognition*, 79 (1-2) : 1-37.

Dehaene, S., L. Naccache, L. Cohen, D. Le Bihan, J. F. Mangin, J. B. Poline et D. Rivière, 2001, « Cerebral Mechanisms of Word Masking and Unconscious Repetition Priming », *Nature Neuroscience*, 4 (7) : 752-758.

Dehaene, S., L. Naccache, G. Le Clec'H, E. Koechlin, M. Mueller, G. Dehaene-Lambertz, P. F. Van de Moortele et D. Le Bihan, 1998, « Imaging Unconscious Semantic Priming », *Nature*, 395 (6702) : 597-600.

Dehaene, S., F. Pegado, L. W. Braga, P. Ventura, G. Nunes Filho, A. Jobert, G. Dehaene-Lambertz *et al.*, 2010, « How Learning to Read Changes the Cortical Networks for Vision and Language », *Science* 330 (6009) : 1359-1364.

Dehaene, S., M. I. Posner et D. M. Tucker, 1994, « Localization of a Neural System for Error Detection and Compensation », *Psychological Science*, 5 : 303-305.

Dehaene, S., C. Sergent et J.-P. Changeux, 2003, « A Neuronal Network Model Linking Subjective Reports and Objective Physiological Data During Conscious Perception », *Proceedings of the National Academy of Sciences*, 100 : 8520-8525.

Dehaene, S. et M. Sigman, 2012, « From a Single Decision to a Multi-step Algorithm », *Current Opinion in Neurobiology*, 22 (6) : 937-945.

Dehaene-Lambertz, G. et S. Dehaene, 1994, « Speed and Cerebral Correlates of Syllable Discrimination in Infants », *Nature*, 370 : 292-295.

Dehaene-Lambertz, G., S. Dehaene et L. Hertz-Pannier, 2002, « Functional Neuroimaging of Speech Perception in Infants », *Science*, 298 (5600) : 2013-2015.

Dehaene-Lambertz, G., L. Hertz-Pannier et J. Dubois, 2006a, « Nature and Nurture in Language Acquisition : Anatomical and Functional Brain-imaging Studies in Infants », *Trends in Neurosciences*, 29 (7) : 367-373.

DEHAENE-LAMBERTZ, G., L. HERTZ-PANNIER, J. DUBOIS, S. MERIAUX, A. ROCHE, M. SIGMAN et S. DEHAENE, 2006b, « Functional Organization of Perisylvian Activation During Presentation of Sentences in Preverbal Infants », *Proceedings of the National Academy of Sciences* 103 (38) : 14240-14245.

DEHAENE-LAMBERTZ, G., A. MONTAVONT, A. JOBERT, L. ALLIROL, J. DUBOIS, L. HERTZ-PANNIER et S. DEHAENE, 2009, « Language or Music, Mother or Mozart ? Structural and Environmental Influences on Infants' Language Networks », *Brain Language*, 114 (2) : 53-65.

DEL CUL, A., S. BAILLET et S. DEHAENE, 2007, « Brain Dynamics Underlying the Nonlinear Threshold for Access to Consciousness », *PLOS Biology*, 5 (10) : e260.

DEL CUL, A., S. DEHAENE et M. LEBOYER, 2006, « Preserved Subliminal Processing and Impaired Conscious Access in Schizophrenia », *Archives of General Psychiatry*, 63 (12) : 1313-1323.

DEL CUL, A., S. DEHAENE, P. REYES, E. BRAVO et A. SLACHEVSKY, 2009, « Causal Role of Prefrontal Cortex in the Threshold for Access to Consciousness », *Brain*, 132 (9) : 2531-2540.

DELL'ACQUA, R. et J. GRAINGER, 1999, « Unconscious Semantic Priming from Pictures », *Cognition*, 73 (1) : B1-B15.

DEN HEYER, K. et K. BRIAND, 1986, « Priming Single Digit Numbers : Automatic Spreading Activation Dissipates as a Function of Semantic Distance », *American Journal of Psychology*, 99 (3) : 315-340.

DENEVE, S., P. E. LATHAM et A. POUGET, 2001, « Efficient Computation and Cue Integration with Noisy Population Codes », *Nature Neuroscience*, 4 (8) : 826-831.

DENNETT, D., 1978, *Brainstorms*, Cambridge, Mass., MIT Press.

–, 1984, *Elbow Room : The Varieties of Free Will Worth Wanting*, Cambridge, Mass., MIT Press.

–, 1993, *La Conscience expliquée*, Paris, Odile Jacob.

DENTON, D., R. SHADE, F. ZAMARIPPA, G. EGAN, J. BLAIR-WEST, M. MCKINLEY, J. LANCASTER et P. FOX, 1999, « Neuroimaging of Genesis and Satiation of Thirst and an Interoceptor-driven Theory of Origins of Primary Consciousness », *Proceedings of the National Academy of Sciences*, 96 (9) : 5304-5309.

DENYS, K., W. VANDUFFEL, D. FIZE, K. NELISSEN, H. SAWAMURA, S. GEORGIEVA, R. VOGELS *et al.*, 2004, « Visual Activation in Prefrontal Cortex Is Stronger in Monkeys Than in Humans », *Journal of Cognitive Neuroscience*, 16 (9) : 1505-1516.

DERDIKMAN, D. et E. I. MOSER, 2010, « A Manifold of Spatial Maps in the Brain », *Trends in Cognitive Sciences*, 14 (12) : 561-569.

DESCARTES, R., 1987, *Œuvres et lettres*, Paris, Gallimard, « Bibliothèque de la Pléiade ».

DESMURGET, M., K. T. REILLY, N. RICHARD, A. SZATHMARI, C. MOTTOLESE et A. SIRIGU, 2009, « Movement Intention After Parietal Cortex Stimulation in Humans », *Science*, 324 (5928) : 811-813.

DI LOLLO, V., J. T. ENNS et R. A. RENSINK, 2000, « Competition for Consciousness Among Visual Events : The Psychophysics of Reentrant Visual Processes », *Journal of Experimental Psychology : General*, 129 (4) : 481-507.

DI VIRGILIO, G. et S. CLARKE, 1997, « Direct Interhemispheric Visual Input to Human Speech Areas », *Human Brain Mapping*, 5 (5) : 347-354.

DIAMOND, A. et B. DOAR, 1989, « The Performance of Human Infants on a Measure of Frontal Cortex Function the Delayed Response Task », *Developmental Psychobiology*, 22 (3) : 271-294.

DIAMOND, A. et J. GILBERT, 1989, « Development as Progressive Inhibitory Control of Action Retrieval of a Contiguous Object », *Cognitive Development*, 4 (3) : 223-250.

DIAMOND, A. et P. S. GOLDMAN-RAKIC, 1989, « Comparison of Human Infants and Rhesus Monkeys on Piaget's A-not-B Task : Evidence for

Dependence on Dorsolateral Prefrontal Cortex », *Experimental Brain Research*, 74 (1) : 24-40.

DICKMAN, D. K. et G. W. DAVIS, 2009, « The Schizophrenia Susceptibility Gene Dysbindin Controls Synaptic Homeostasis », *Science*, 326 (5956) : 1127-1130.

DIJKSTERHUIS, A., M. W. BOS, L. F. NORDGREN et R. B. VAN BAAREN, 2006, « On Making the Right Choice : The Deliberation-without-attention Effect », *Science*, 311 (5763) : 1005-1007.

DONCHIN, E. et M. G. H. COLES, 1988, « Is the P300 Component a Manifestation of Context Updating ? », *Behavioral and Brain Sciences*, 11 (3) : 357-427.

DORIA, V., C. F. BECKMANN, T. ARICHI, N. MERCHANT, M. GROPPO, F. E. TURKHEIMER, S. J. COUNSELL *et al.*, 2010, « Emergence of Resting State Networks in the Preterm Human Brain », *Proceedings of the National Academy of Sciences*, 107 (46) : 20015-20020.

DOS SANTOS COURA, R. et S. GRANON, 2012, « Prefrontal Neuromodulation by Nicotinic Receptors for Cognitive Processes », *Psychopharmacology (Berlin)*, 221 (1) : 1-18.

DRIVER, J. et P. VUILLEUMIER, 2001, « Perceptual Awareness and Its Loss in Unilateral Neglect and Extinction », *Cognition*, 79 (1-2) : 39-88.

DUBOIS, J., G. DEHAENE-LAMBERTZ, M. PERRIN, J. F. MANGIN, Y. COINTEPAS, E. DUCHESNAY, D. LE BIHAN et L. HERTZ-PANNIER, 2007, « Asynchrony of the Early Maturation of White Matter Bundles in Healthy Infants : Quantitative Landmarks Revealed Noninvasively by Diffusion Tensor Imaging », *Human Brain Mapping*, 29 (1) : 14-27.

DUNBAR, R., 1996, *Grooming, Gossip and the Evolution of Language*, Londres, Faber and Faber.

DUPOUX, E., V. DE GARDELLE et S. KOUIDER, 2008, « Subliminal Speech Perception and Auditory Streaming », *Cognition*, 109 (2) : 267-273.

EAGLEMAN, D. M. et T. J. SEJNOWSKI, 2000, « Motion Integration and Postdiction in Visual Awareness », *Science*, 287 (5460) : 2036-2038.

–, 2007, « Motion Signals Bias Localization Judgments : A Unified Explanation for the Flash-lag, Flash-drag, Flash-jump, and Frohlich Illusions », *Journal of Vision* 7 (4) : 3.

ECCLES, J. C., 1997, *Comment la conscience contrôle le cerveau*, Paris, Fayard.

EDELMAN, G., 1987, *Neural Darwinism*, New York, Basic Books.

–, 1994, *Biologie de la* conscience, Paris, Odile Jacob.

EHRSSON, H. H., 2007, « The Experimental Induction of Out-of-body Experiences », *Science*, 317 (5841) : 1048.

EHRSSON, H. H., C. SPENCE et R. E. PASSINGHAM, 2004, « That's My Hand ! Activity in Premotor Cortex Reflects Feeling of Ownership of a Limb », *Science*, 305 (5685) : 875-877.

ELIASMITH, C., T. C. STEWART, X. CHOO, T. BEKOLAY, T. DEWOLF, Y. TANG et D. RASMUSSEN, 2012, « A Large-scale Model of the Functioning Brain », *Science*, 338 (6111) : 1202-1205.

ELLENBERGER, H. F., 2001, *Histoire de la découverte de l'inconscient*, Paris, Fayard.

ELSTON, G. N., 2000, « Pyramidal Cells of the Frontal Lobe : All the More Spinous to Think With », *Journal of Neuroscience*, 20 (18) : RC95.

ELSTON, G. N., 2003, « Cortex, Cognition and the Cell : New Insights into the Pyramidal Neuron and Prefrontal Function », *Cerebral Cortex*, 13 (11) : 1124-1138.

ELSTON, G. N., R. BENAVIDES-PICCIONE et J. DEFELIPE, 2001, « The Pyramidal Cell in Cognition : A Comparative Study in Human and Monkey », *Journal of Neuroscience*, 21 (17) : RC163.

ENARD, W., S. GEHRE, K. HAMMERSCHMIDT, S. M. HOLTER, T. BLASS, M. SOMEL, M. K. BRUCKNER *et al.*, 2009, « A Humanized Version of Foxp2 Affects Cortico-basal Ganglia Circuits in Mice », *Cell*, 137 (5) : 961-971.

ENARD, W., M. PRZEWORSKI, S. E. FISHER, C. S. LAI, V. WIEBE, T. KITANO, A. P. MONACO et S. PAABO, 2002, « Molecular Evolution of FOXP2, a Gene Involved in Speech and Language », *Nature*, 418 (6900) : 869-872.

ENGEL, A. K. et W. SINGER, 2001, « Temporal Binding and the Neural Correlates of Sensory Awareness », *Trends in Cognitive Sciences*, 5 (1) : 16-25.

ENNS, J. T. et V. DI LOLLO, 2000, « What's New in Visual Masking », *Trends in Cognitive Sciences*, 4 (9) : 345-352.

EPSTEIN, R., R. P. LANZA et B. F. SKINNER, 1981. « "Self-awareness" in the Pigeon », *Science*, 212 (4495) : 695-696.

FAHRENFORT, J. J., H. S. SCHOLTE et V. A. LAMME, 2007, « Masking Disrupts Reentrant Processing in Human Visual Cortex », *Journal of Cognitive Neuroscience*, 19 (9) : 1488-1497.

FAUGERAS, F., B. ROHAUT, N. WEISS, T. A. BEKINSCHTEIN, D. GALANAUD, L. PUYBASSET, F. BOLGERT *et al.*, 2011, « Probing Consciousness with Event-related Potentials in the Vegetative State », *Neurology*, 77 (3) : 264-268.

–, 2012, « Event Related Potentials Elicited by Violations of Auditory Regularities in Patients with Impaired Consciousness », *Neuropsychologia*, 50 (3) : 403-418.

FEDORENKO, E., J. DUNCAN et N. KANWISHER, 2012, « Language-selective and Domain-general Regions Lie Side by Side Within Broca's Area », *Current Biology*, 22 (21) : 2059-2062.

FELLEMAN, D. J. et D. C. VAN ESSEN, 1991, « Distributed Hierarchical Processing in the Primate Cerebral Cortex », *Cerebral Cortex*, 1 (1) : 1-47.

FERRARELLI, F., M. MASSIMINI, S. SARASSO, A. CASALI, B. A. RIEDNER, G. ANGELINI, G. TONONI et R. A. PEARCE, 2010, « Breakdown in Cortical Effective Connectivity During Midazolam-induced Loss of Consciousness », *Proceedings of the National Academy of Sciences*, 107 (6) : 2681-2686.

FFYTCHE, D. H., R. J. HOWARD, M. J. BRAMMER, A. DAVID, P. WOODRUFF et S. WILLIAMS, 1998, « The Anatomy of Conscious Vision : An fMRI Study of Visual Hallucinations », *Nature Neuroscience*, 1 (8) : 738-742.

FINGER, S., 2001, *Origins of Neuroscience : A History of Explorations into Brain Function*, Oxford, Oxford University Press.

FINKEL, L. H. et G. M. EDELMAN, 1989, « Integration of Distributed Cortical Systems by Reentry : A Computer Simulation of Interactive Functionally Segregated Visual Areas », *Journal of Neuroscience* 9 (9), 3188-3208.

FISCH, L., E. PRIVMAN, M. RAMOT, M. HAREL, Y. NIR, S. KIPERVASSER, F. ANDELMAN *et al.*, 2009, « Neural "Ignition" : Enhanced Activation Linked to Perceptual Awareness in Human Ventral Stream Visual Cortex », *Neuron*, 64 (4) : 562-574.

FISCHER, C., J. LUAUTE, P. ADELEINE et D. MORLET, 2004, « Predictive Value of Sensory and Cognitive Evoked Potentials for Awakening from Coma », *Neurology*, 63 (4) : 669-673.

FLEMING, S. M., R. S. WEIL, Z. NAGY, R. J. DOLAN et G. REES, 2010, « Relating Introspective Accuracy to Individual Differences in Brain Structure », *Science*, 329 (5998) : 1541-1543.

FLETCHER, P. C. et C. D. FRITH, 2009, « Perceiving Is Believing : A Bayesian Approach to Explaining the Positive Symptoms of Schizophrenia », *Nature Reviews Neuroscience*, 10 (1) : 48-58.

FORSTER, K. I., 1998, « The Pros and Cons of Masked Priming », *Journal of Psycholinguistic Research*, 27 (2) : 203-233.

FORSTER, K. I. et C. DAVIS, 1984, « Repetition Priming and Frequency Attenuation in Lexical Access », *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 10 (4) : 680-698.

FRANSSON, P., B. SKIOLD, S. HORSCH, A. NORDELL, M. BLENNOW, H. LAGERCRANTZ et U. ADEN, 2007, « Resting-state Networks in the Infant Brain », *Proceedings of the National Academy of Sciences*, 104 (39) : 15531-1536.

FRIED, I., K. A. MACDONALD et C. L. WILSON, 1997, « Single Neuron Activity in Human Hippocampus and Amygdala During Recognition of Faces and Objects », *Neuron*, 18 (5) : 753-765.

FRIEDERICI, A. D., M. FRIEDRICH et C. WEBER, 2002, « Neural Manifestation of Cognitive and Precognitive Mismatch Detection in Early Infancy », *NeuroReport*, 13 (10) : 1251-1254.

FRIES, P., 2005, « A Mechanism for Cognitive Dynamics : Neuronal Communication Through Neuronal Coherence », *Trends in Cognitive Sciences*, 9 (10) : 474-480.

FRIES, P., D. NIKOLIC et W. SINGER, 2007, « The Gamma Cycle », *Trends in Neurosciences*, 30 (7) : 309-316.

FRIES, P., J. H. SCHRODER, P. R. ROELFSEMA, W. SINGER et A. K. ENGEL, 2002, « Oscillatory Neuronal Synchronization in Primary Visual Cortex as a Correlate of Stimulus Selection », *Journal of Neuroscience*, 22 (9) : 3739-3754.

FRISTON, K., 2005, « A Theory of Cortical Responses », *Philosophical Transactions of the Royal Society B : Biological Sciences*, 360 (1456) : 815-836.

FRITH, C., 1996, « The Role of the Prefrontal Cortex in Self-consciousness : The Case of Auditory Hallucinations », *Philosophical Transactions of the Royal Society B : Biological Sciences*, 351 (1346) : 1505-1512.

–, 1979, « Consciousness, Information Processing and Schizophrenia », *British Journal of Psychiatry* 134 : 225-235.

–, 2010, *Comment le cerveau crée notre univers mental*, Paris, Odile Jacob.

FUJII, N. et A. M. GRAYBIEL, 2003, « Representation of Action Sequence Boundaries by Macaque Prefrontal Cortical Neurons », *Science*, 301 (5637) : 1246-1249.

FUNAHASHI, S., C. J. BRUCE et P. S. GOLDMAN-RAKIC, 1989, « Mnemonic Coding of Visual Space in the Monkey's Dorsolateral Prefrontal Cortex », *Journal of Neurophysiology*, 61 (2) : 331-349.

FUSTER, J. M., 1973, « Unit Activity in Prefrontal Cortex During Delayed-response Performance : Neuronal Correlates of Transient Memory », *Journal of Neurophysiology*, 36 (1) : 61-78.

–, 2008, *The Prefrontal* Cortex, Londres, Academic Press, 4e éd.

GAILLARD, R., S. DEHAENE, C. ADAM, S. CLEMENCEAU, D. HASBOUN, M. BAULAC, L. COHEN et L. NACCACHE, 2009, « Converging Intracranial Markers of Conscious Access », *PLOS Biology*, 7 (3) : e61.

GAILLARD, R., A. DEL CUL, L. NACCACHE, F. VINCKIER, L. COHEN et S. DEHAENE, 2006, « Nonconscious Semantic Processing of Emotional Words Modulates Conscious Access », *Proceedings of the National Academy of Sciences*, 103 (19) : 7524-7529.

GAILLARD, R., L. NACCACHE, P. PINEL, S. CLEMENCEAU, E. VOLLE, D. HASBOUN, S. DUPONT *et al.*, 2006, « Direct Intracranial, fMRI, and Lesion Evidence for the Causal Role of Left Inferotemporal Cortex in Reading », *Neuron*, 50 (2) : 191-204.

GALANAUD, D., L. NACCACHE et L. PUYBASSET, 2007, « Exploring Impaired Consciousness : The MRI Approach », *Current Opinion in Neurology*, 20 (6) : 627-631.

GALANAUD, D., V. PERLBARG, R. GUPTA, R. D. STEVENS, P. SANCHEZ, E. TOLLARD, N. M. DE CHAMPFLEUR *et al.*, 2012, « Assessment of White Matter Injury and Outcome in Severe Brain Trauma : A Prospective Multicenter Cohort », *Anesthesiology*, 117 (6) : 1300-1310.

GALLUP, G. G., 1970, « Chimpanzees : Self-recognition », *Science*, 167 : 86-87.

GASER, C. et G. SCHLAUG, 2003, « Brain Structures Differ Between Musicians and Non-musicians », *Journal of Neuroscience*, 23 (27) : 9240-9245.

GAUCHET, M., 1992, *L'Inconscient cérébral*, Paris, Seuil.

GEHRING, W. J., B. GOSS, M. G. H. COLES, D. E. MEYER et E. DONCHIN, 1993, « A Neural System for Error Detection and Compensation », *Psychological Science*, 4 (6) : 385-390.

GELSKOV, S. V. et S. KOUIDER, 2010, « Psychophysical Thresholds of Face Visibility During Infancy », *Cognition*, 114 (2) : 285-292.

GIACINO, J., J. J. FINS, A. MACHADO et N. D. SCHIFF, 2012, « Central Thalamic Deep Brain Stimulation to Promote Recovery from Chronic

Posttraumatic Minimally Conscious State : Challenges and Opportunities », *Neuromodulation*, 15 (4) : 339-349.

GIACINO, J. T., 2005, « The Minimally Conscious State : Defining the Borders of Consciousness », *Progress in Brain Research*, 150 : 381-395.

GIACINO, J. T., K. KALMAR et J. WHYTE, 2004, « The JFK Coma Recovery Scale-Revised : Measurement Characteristics and Diagnostic Utility », *Archives of Physical Medicine and Rehabilitation*, 85 (12) : 2020-2029.

GIACINO, J. T., M. A. KEZMARSKY, J. DELUCA et K. D. CICERONE, 1991, « Monitoring Rate of Recovery to Predict Outcome in Minimally Responsive Patients », *Archives of Physical Medicine and Rehabilitation*, 72 (11) : 897-901.

GIACINO, J. T., J. WHYTE, E. BAGIELLA, K. KALMAR, N. CHILDS, A. KHADEMI, B. EIFERT *et al.*, 2012, « Placebo-controlled Trial of Amantadine for Severe Traumatic Brain Injury », *New England Journal of Medicine*, 366 (9) : 819-826.

GIESBRECHT, B. et V. DI LOLLO, 1998, « Beyond the Attentional Blink : Visual Masking by Object Substitution », *Journal of Experimental Psychology : Human Perception and Performance*, 24 (5) : 1454-1466.

GILBERT, C. D., M. SIGMAN et R. E. CRIST, 2001, « The Neural Basis of Perceptual Learning », *Neuron*, 31 (5) : 681-697.

GOBET, F. et H. A. SIMON, 1998, « Expert Chess Memory : Revisiting the Chunking Hypothesis », *Memory*, 6 (3) : 225-255.

GOEBEL, R., L. MUCKLI, F. E. ZANELLA, W. SINGER et P. STOERIG, 2001, « Sustained Extrastriate Cortical Activation Without Visual Awareness Revealed by fMRI Studies of Hemianopic Patients », *Vision Research*, 41 (10-11) : 1459-1474.

GOLDFINE, A. M., J. D. VICTOR, M. M. CONTE, J. C. BARDIN et N. D. SCHIFF, 2011, « Determination of Awareness in Patients with Severe Brain Injury Using EEG Power Spectral Analysis », *Clinical Neurophysiology*, 122 (11) : 2157-2168.

–, 2012, « Bedside Detection of Awareness in the Vegetative State », *The Lancet*, 379 (9827) : 1701-1702.

GOLDMAN-RAKIC, P. S., 1988, « Topography of Cognition : Parallel Distributed Networks in Primate Association Cortex », *Annual Review of Neuroscience*, 11 : 137-156.

–, 1995, « Cellular Basis of Working Memory », *Neuron*, 14 (3) : 477-485.

GOODALE, M. A., A. D. MILNER, L. S. JAKOBSON et D. P. CAREY, 1991, « A Neurological Dissociation Between Perceiving Objects and Grasping Them », *Nature*, 349 (6305) : 154-156.

GOULD, S. J., 1974, « The Origin and Function of "Bizarre" Structures : Antler Size and Skull Size in the "Irish Elk", Megaloceros giganteus », *Evolution*, 28 (2) : 191-220.

GOULD, S. J. et R. C. LEWONTIN, 1979, « The Spandrels of San Marco and the Panglossian Paradigm : A Critique of the Adaptationist Programme », *Proceedings of the Royal Society B : Biological Sciences*, 205 (1161) : 581-598.

GREENBERG, D. L., 2007, « Comment on "Detecting Awareness in the Vegetative State" », *Science*, 315 (5816) : 1221 ; et réponse de l'auteur : 1221.

GREENWALD, A. G., R. L. ABRAMS, L. NACCACHE et S. DEHAENE, 2003, « Long-term Semantic Memory Versus Contextual Memory in Unconscious Number Processing », *Journal of Experimental Psychology : Learning, Memory, Cognition*, 29 (2) : 235-247.

GREENWALD, A. G., S. C. DRAINE et R. L. ABRAMS, 1996, « Three Cognitive Markers of Unconscious Semantic Activation », *Science*, 273 (5282) : 1699-1702.

GREICIUS, M. D., V. KIVINIEMI, O. TERVONEN, V. VAINIONPAA, S. ALAHUHTA, A. L. REISS et V. MENON, 2008, « Persistent Default-mode Network Connectivity During Light Sedation, » *Human Brain Mapping*, 29 (7) : 839-847.

GREICIUS, M. D., B. KRASNOW, A. L. REISS et V. MENON, 2003, « Functional Connectivity in the Resting Brain : A Network Analysis of the Default Mode Hypothesis », *Proceedings of the National Academy of Sciences*, 100 (1) : 253-258.

GRIFFITHS, J. D., W. D. MARSLEN-WILSON, E. A. STAMATAKIS et L. K.TYLER, 2013, « Functional Organization of the Neural Language System : Dorsal and Ventral Pathways Are Critical for Syntax », *Cerebral Cortex*, 23 (1) : 139-147.

GRILL-SPECTOR, K., T. KUSHNIR, T. HENDLER et R. MALACH, 2000, « The Dynamics of Object-Selective Activation Correlate with Recognition Performance in Humans », *Nature Neuroscience*, 3 (8) : 837-843.

GRINDAL, A. B., C. SUTER et A. J. MARTINEZ, 1977, « Alpha-pattern Coma : 24 Cases with 9 Survivors », *Annals of Neurology*, 1 (4) : 371-377.

GROSS, J., F. SCHMITZ, I. SCHNITZLER, K. KESSLER, K. SHAPIRO, B. HOMMEL et A. SCHNITZLER, 2004, « Modulation of Long-range Neural Synchrony Reflects Temporal Limitations of Visual Attention in Humans », *Proceedings of the National Academy of Sciences*, 101 (35) : 13050-13055.

HADAMARD, J., 1945, *An Essay on the Psychology of Invention in the Mathematical Field*, Princeton, NJ, Princeton University Press.

HAGMANN, P., L. CAMMOUN, X. GIGANDET, R. MEULI, C. J. HONEY, V. J. WEDEEN et O. SPORNS, 2008, « Mapping the Structural Core of Human Cerebral Cortex », *PLOS Biology*, 6 (7) : e159.

HALELAMIEN, N., D.-A. WU et S. SHIMOJO, 2007, « TMS Induces Detail-Rich “Instant Replays” of Natural Images », *Journal of Vision*, 7 (9).

HALLETT, M., 2000, « Transcranial Magnetic Stimulation and the Human Brain », *Nature*, 406 (6792) : 147-150.

HAMPTON, R. R., 2001, « Rhesus Monkeys Know When They Remember », *Proceedings of the National Academy of Sciences*, 98 (9) : 5359-5362.

HAN, C. J., C. M. O'TUATHAIGH, L. VAN TRIGT, J. J. QUINN, M. S. FANSELOW, R. MONGEAU, C. KOCH et D. J. ANDERSON, 2003, « Trace But Not Delay Fear

Conditioning Requires Attention and the Anterior Cingulate Cortex », *Proceedings of the National Academy of Sciences*, 100 (22) : 13087-13092.

HANSLMAYR, S., J. GROSS, W. KLIMESCH et K. L. SHAPIRO, 2011, « The Role of Alpha Oscillations in Temporal Attention », *Brain Research Reviews*, 67 (1-2) : 331-343.

HASSON, U., Y. NIR, I. LEVY, G. FUHRMANN et R. MALACH, 2004, « Intersubject Synchronization of Cortical Activity During Natural Vision », *Science*, 303 (5664) : 1634-1640.

HASSON, U., J. I. SKIPPER, H. C. NUSBAUM et S. L. SMALL, 2007, « Abstract Coding of Audiovisual Speech : Beyond Sensory Representation », *Neuron*, 56 (6) : 1116-1126.

HAYDEN, B. Y., D. V. SMITH et M. L. PLATT, 2009, « Electrophysiological Correlates of Default-mode Processing in Macaque Posterior Cingulate Cortex », *Proceedings of the National Academy of Sciences*, 106 (14) : 5948-5953.

HAYNES, J. D., 2009, « Decoding Visual Consciousness from Human Brain Signals », *Trends in Cognitive Sciences*, 13 : 194-202.

HAYNES, J. D., R. DEICHMANN et G. REES, 2005, « Eye-specific Effects of Binocular Rivalry in the Human Lateral Geniculate Nucleus », *Nature*, 438 (7067) : 496-499.

HAYNES, J. D., J. DRIVER et G. REES, 2005, « Visibility Reflects Dynamic Changes of Effective Connectivity Between V1 and Fusiform Cortex », *Neuron*, 46 (5) : 811-821.

HAYNES, J. D. et G. REES, 2005a, « Predicting the Orientation of Invisible Stimuli from Activity in Human Primary Visual Cortex », *Nature Neuroscience*, 8 (5) : 686-691.

–, 2005b, « Predicting the Stream of Consciousness from Activity in Human Visual Cortex », *Current Biology*, 15 (14) : 1301-1307.

HAYNES, J. D., K. SAKAI, G. REES, S. GILBERT, C. FRITH et R. E. PASSINGHAM, 2007, « Reading Hidden Intentions in the Human Brain », *Current*

Biology, 17 (4) : 323-328.

He, B. J. et M. E. Raichle, 2009, « The fMRI Signal, Slow Cortical Potential and Consciousness », *Trends in Cognitive Sciences*, 13 (7) : 302-309.

He, B. J., A. Z. Snyder, J. M. Zempel, M. D. Smyth et M. E. Raichle, 2008, « Electrophysiological Correlates of the Brain's Intrinsic Large-scale Functional Architecture », *Proceedings of the National Academy of Sciences*, 105 (41) : 16039-16044.

He, B. J., J. M. Zempel, A. Z. Snyder et M. E. Raichle, 2010, « The Temporal Structures and Functional Significance of Scale-free Brain Activity », *Neuron*, 66 (3) : 353-369.

He, S. et D. I. MacLeod, 2001, « Orientation-selective Adaptation and Tilt After-effect from Invisible Patterns », *Nature*, 411 (6836) : 473-476.

Hebb, D. O., 1949, *The Organization of Behavior*, New York, Wiley. (Traduction française : *Psychophysiologie du comportement*, Paris, Presses universitaires de France, 1958.)

Heit, G., M. E. Smith et E. Halgren, 1988, « Neural Encoding of Individual Words and Faces by the Human Hippocampus and Amygdala », *Nature*, 333 (6175) : 773-775.

Henson, R. N., E. Mouchlianitis, W. J. Matthews et S. Kouider, 2008, « Electrophysiological Correlates of Masked Face Priming », *NeuroImage*, 40 (2) : 884-895.

Herrmann, E., J. Call, M. V. Hernandez-Lloreda, B. Hare et M. Tomasello, 2007, « Humans Have Evolved Specialized Skills of Social Cognition : The Cultural Intelligence Hypothesis », *Science*, 317 (5843) : 1360-1366.

Hochberg, L. R., D. Bacher, B. Jarosiewicz, N. Y. Masse, J. D. Simeral, J. Vogel, S. Haddadin *et al.*, 2012, « Reach and Grasp by People with Tetraplegia Using a Neurally Controlled Robotic Arm », *Nature*, 485 (7398) : 372-375.

Hofstadter, D., 2013, *Je suis une boucle étrange. Que veut dire « je » ?*, Paris, Dunod.

HOLENDER, D., 1986, « Semantic Activation Without Conscious Identification in Dichotic Listening Parafoveal Vision and Visual Masking : A Survey and Appraisal », *Behavioral and Brain Sciences*, 9 (1) : 1-23.

HOLENDER, D. et K. DUSCHERER, 2004, « Unconscious Perception : The Need for a Paradigm Shift », *Perception and Psychophysics*, 66 (5) : 872-881 ; discussion : 888-895.

HOPFIELD, J. J., 1982, « Neural Networks and Physical Systems with Emergent Collective Computational Abilities », *Proceedings of the National Academy of Sciences*, 79 (8) : 2554-5258.

HORIKAWA, T., M. TAMAKI, Y. MIYAWAKI et Y. KAMITANI, 2013, « Neural Decoding of Visual Imagery During Sleep », *Science*, 340 (6132) : 639-642.

HOWARD, I. P., 1996, « Alhazen's Neglected Discoveries of Visual Phenomena », *Perception*, 25 (10) : 1203-1217.

HOWE, M. J. A. et J. SMITH, 1988, « Calendar Calculating in Idiots Savants : How Do They Do It », *British Journal of Psychology*, 79 (pt. 3) : 371-386.

HURON, C., J. M. DANION, F. GIACOMONI, D. GRANGE, P. ROBERT et L. RIZZO, 1995, « Impairment of Recognition Memory With, but Not Without, Conscious Recollection in Schizophrenia », *American Journal of Psychiatry*, 152 (12) : 1737-1742.

IZARD, V., C. SANN, E. S. SPELKE et A. STRERI, 2009, « Newborn Infants Perceive Abstract Numbers », *Proceedings of the National Academy of Sciences*, 106 (25) : 10382-10385.

IZHIKEVICH, E. M. et G. M. EDELMAN, 2008, « Large-scale Model of Mammalian Thalamocortical Systems », *Proceedings of the National Academy of Sciences*, 105 (9) : 3593-3598.

JAMES, W., 1890, *The Principles of Psychology*, New York, Holt.

JAYNES, J., 1994, *La Naissance de la conscience dans l'effondrement de l'esprit*, Paris, Presses universitaires de France.

JENKINS, A. C., C. N. MACRAE et J. P. MITCHELL, 2008, « Repetition Suppression of Ventromedial Prefrontal Activity During Judgments of Self and Others », *Proceedings of the National Academy of Sciences* 105 (11) : 4507-4512.

JENNETT, B., 2002, *The Vegetative State : Medical Facts, Ethical and Legal Dilemmas*, New York, Cambridge University Press.

JENNETT, B. et F. PLUM, 1972, « Persistent Vegetative State After Brain Damage. A Syndrome in Search of a Name », *The Lancet* 1 (7753) : 734-737.

JEZEK, K., E. J. HENRIKSEN, A. TREVES, E. I. MOSER et M. B. MOSER, 2011, « Theta-paced Flickering Between Place-cell Maps in the Hippocampus », *Nature*, 478 (7368) : 246-249.

JI, D. et M. A. WILSON, 2007, « Coordinated Memory Replay in the Visual Cortex and Hippocampus During Sleep », *Nature Neuroscience*, 10 (1) : 100-107.

JOHANSSON, P., L. HALL, S. SIKSTROM et A. OLSSON, 2005, « Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task », *Science*, 310 (5745) : 116-119.

JOHNSON, M. H., S. DZIURAWIEC, H. ELLIS et J. MORTON, 1991, « Newborns' Preferential Tracking of Face-like Stimuli and Its Subsequent Decline », *Cognition*, 40 (1-2) : 1-19.

JOLICŒUR, P., 1999, « Concurrent Response-selection Demands Modulate the Attentional Blink », *Journal of Experimental Psychology : Human Perception and Performance*, 25 (4) : 1097-1113.

JORDAN, D., G. STOCKMANNS, E. F. KOCHS, S. PILGE et G. SCHNEIDER, 2008, « Electroencephalographic Order Pattern Analysis for the Separation of Consciousness and Unconsciousness : An Analysis of Approximate Entropy, Permutation Entropy, Recurrence Rate, and Phase Coupling of Order Recurrence Plots », *Anesthesiology*, 109 (6) : 1014-1022.

JOUVET, M., 1993, *Le Sommeil et le Rêve*, Paris, Odile Jacob.

KAHNEMAN, D. et A. TREISMAN, 1984, « Changing Views of Attention and Automaticity », *in* R. Parasuraman, R. Davies et J. Beatty (éd.), *Varieties of Attention*, New York, Academic Press : 29-61.

KANAI, R., T. A. CARLSON, F. A. VERSTRATEN et V. WALSH, 2009, « Perceived Timing of New Objects and Feature Changes », *Journal of Vision*, 9 (7) : 5.

KANAI, R., N. G. MUGGLETON et V. WALSH. 2008, « TMS over the Intraparietal Sulcus Induces Perceptual Fading », *Journal of Neurophysiology*, 100 (6) : 3343-3350.

KANE, N. M., S. H. CURRY, S. R. BUTLER et B. H. CUMMINS, 1993, « Electrophysiological Indicator of Awakening from Coma », *The Lancet*, 341 (8846) : 688.

KANWISHER, N., 2001, « Neural Events and Perceptual Awareness », *Cognition*, 79 (1-2) : 89-113.

KARLSGODT, K. H., D. SUN, A. M. JIMENEZ, E. S. LUTKENHOFF, R. WILLHITE, T. G. VAN ERP et T. D. CANNON, 2008, « Developmental Disruptions in Neural Connectivity in the Pathophysiology of Schizophrenia », *Development and Psychopathology*, 20 (4) : 1297-1327.

KENET, T., D. BIBITCHKOV, M. TSODYKS, A. GRINVALD et A. ARIELI, 2003, « Spontaneously Emerging Cortical Representations of Visual Attributes », *Nature*, 425 (6961) : 954-956.

KENTRIDGE, R. W., T. C. NIJBOER et C. A. HEYWOOD, 2008, « Attended but Unseen : Visual Attention Is Not Sufficient for Visual Awareness », *Neuropsychologia*, 46 (3) : 864-869.

KERSTEN, D., P. MAMASSIAN et A. YUILLE, 2004, « Object Perception as Bayesian Inference », *Annual Review of Psychology*, 55 : 271-304.

KIANI, R. et M. N. SHADLEN, 2009, « Representation of Confidence Associated with a Decision by Neurons in the Parietal Cortex », *Science*, 324 (5928) : 759-764.

KIEFER, M., 2002, « The N400 Is Modulated by Unconsciously Perceived Masked Words : Further Evidence for an Automatic Spreading

Activation Account of N400 Priming Effects », *Brain Research : Cognitive Brain Research*, 13 (1) : 27-39.

KIEFER, M. et D. BRENDEL, 2006, « Attentional Modulation of Unconscious "Automatic" Processes : Evidence from Event-Related Potentials in a Masked Priming Paradigm », *Journal of Cognitive Neuroscience*, 18 (2) : 184-198.

KIEFER, M. et M. SPITZER, 2000, « Time Course of Conscious and Unconscious Semantic Brain Activation », *NeuroReport*, 11 (11) : 2401-2407.

KIESEL, A., W. KUNDE, C. POHL, M. P. BERNER et J. HOFFMANN, 2009, « Playing Chess Unconsciously », *Journal of Experimental Psychology : Learning, Memory, Cognition*, 35 (1) : 292-298.

KIHARA, K., T. IKEDA, D. MATSUYOSHI, N. HIROSE, T. MIMA, H. FUKUYAMA et N. OSAKA, 2010, « Differential Contributions of the Intraparietal Sulcus and the Inferior Parietal Lobe to Attentional Blink : Evidence from Transcranial Magnetic Stimulation », *Journal of Cognitive Neuroscience*, 23 (1) : 247-256.

KIKYO, H., K. OHKI et Y. MIYASHITA, 2002, « Neural Correlates for Feeling-of-knowing : An fMRI Parametric Analysis », *Neuron*, 36 (1) : 177-186.

KIM, C. Y. et R. BLAKE, 2005, « Psychophysical Magic : Rendering the Visible "Invisible" », *Trends in Cognitive Sciences*, 9 (8) : 381-388.

KING, J. R., F. FAUGERAS, A. GRAMFORT, A. SCHURGER, I. EL KAROUI, J. D. SITT, C. WACONGNE *et al.*, 2013, « Single-trial Decoding of Auditory Novelty Responses Facilitates the Detection of Residual Consciousness », *NeuroImage,* 83 : 726-738.

KING, J. R., J. D. SITT, F. FAUGERAS, B. ROHAUT, I. EL KAROUI, L. COHEN, L. NACCACHE et S. DEHAENE, 2013, « Long-distance Information Sharing Indexes the State of Consciousness of Unresponsive Patients », soumis.

KNOCHEL, C., V. OERTEL-KNOCHEL, R. SCHONMEYER, A. ROTARSKA-JAGIELA, V. VAN DE VEN, D. PRVULOVIC, C. HAENSCHEL *et al.*, 2012. « Interhemispheric Hypoconnectivity in Schizophrenia : Fiber Integrity and Volume

Differences of the Corpus Callosum in Patients and Unaffected Relatives », *NeuroImage*, 59 (2) : 926-934.

KOCH, C. et F. CRICK, 2001, « The Zombie Within », *Nature*, 411 (6840) : 893.

KOCH, C. et N. TSUCHIYA, 2007, « Attention and Consciousness : Two Distinct Brain Processes », *Trends in Cognitive Sciences*, 11 (1) : 16-22.

KOECHLIN, E., L. NACCACHE, E. BLOCK et S. DEHAENE, 1999, « Primed Numbers : Exploring the Modularity of Numerical Representations with Masked and Unmasked Semantic Priming », *Journal of Experimental Psychology : Human Perception and Performance*, 25 (6) : 1882-1905.

KOIVISTO, M., M. LAHTEENMAKI, T. A. SORENSEN, S. VANGKILDE, M. OVERGAARD et A. REVONSUO, 2008, « The Earliest Electrophysiological Correlate of Visual Awareness ? », *Brain and Cognition*, 66 (1) : 91-103.

KOIVISTO, M., T. MANTYLA et J. SILVANTO, 2010, « The Role of Early Visual Cortex (V1/V2) in Conscious and Unconscious Visual Perception », *NeuroImage*, 51 (2) : 828-834.

KOIVISTO, M., H. RAILO et N. SALMINEN-VAPARANTA, 2010, « Transcranial Magnetic Stimulation of Early Visual Cortex Interferes with Subjective Visual Awareness and Objective Forced-choice Performance », *Consciousness and Cognition*, 20 (2) : 288-298.

KOMURA, Y., A. NIKKUNI, N. HIRASHIMA, T. UETAKE et A. MIYAMOTO, 2013, « Responses of Pulvinar Neurons Reflect a Subject's Confidence in Visual Categorization », *Nature Neuroscience*, 16 : 749-755.

KONOPKA, G., E. WEXLER, E. ROSEN, Z. MUKAMEL, G. E. OSBORN, L. CHEN, D. LU *et al.*, 2012, « Modeling the Functional Genomics of Autism Using Human Neurons », *Molecular Psychiatry*, 17 (2) : 202-214.

KORNELL, N., L. K. SON et H. S. TERRACE, 2007, « Transfer of Metacognitive Skills and Hint Seeking in Monkeys », *Psychological Science*, 18 (1) : 64-71.

KOUIDER, S., V. DE GARDELLE, J. SACKUR et E. DUPOUX, 2010, « How Rich Is Consciousness ? The Partial Awareness Hypothesis », *Trends in Cognitive Sciences*, 14 (7) : 301-307.

KOUIDER, S. et S. DEHAENE, 2007, « Levels of Processing During Nonconscious Perception : A Critical Review of Visual Masking », *Philosophical Transactions of the Royal Society B : Biological Sciences* 362 (1481) : 857-875.

–, 2009, « Subliminal Number Priming Within and Across the Visual and Auditory Modalities », *Experimental Psychology,* 56 (6) : 418-433.

KOUIDER, S., S. DEHAENE, A. JOBERT et D. LE BIHAN, 2007, « Cerebral Bases of Subliminal and Supraliminal Priming During Reading », *Cerebral Cortex*, 17 (9) : 2019-2029.

KOUIDER, S. et E. DUPOUX, 2004, « Partial Awareness Creates the "Illusion" of Subliminal Semantic Priming », *Psychological Science*, 15 (2) : 75-81.

KOUIDER, S., E. EGER, R. DOLAN et R. N. HENSON, 2009, « Activity in Face-responsive Brain Regions Is Modulated by Invisible, Attended Faces : Evidence from Masked Priming », *Cerebral Cortex*, 19 (1) : 13-23.

KOUIDER, S., C. STAHLHUT, S. V. GELSKOV, L. BARBOSA, M. DUTAT, V. DE GARDELLE, A. CHRISTOPHE *et al.*, 2013, « A Neural Marker of Perceptual Consciousness in Infants », *Science*, 340 (6130) : 376-380.

KOVACS, A. M., E. TEGLAS et A. D. ENDRESS, 2010, « The Social Sense : Susceptibility to Others' Beliefs in Human Infants and Adults », *Science*, 330 (6012) : 1830-1834.

KOVACS, G., R. VOGELS et G. A. ORBAN, 1995, « Cortical Correlate of Pattern Backward Masking », *Proceedings of the National Academy of Sciences*, 92 (12) : 5587-5591.

KREIMAN, G., I. FRIED et C. KOCH, 2002, « Single-neuron Correlates of Subjective Vision in the Human Medial Temporal Lobe », *Proceedings of the National Academy of Sciences*, 99 (12) : 8378-8383.

KREIMAN, G., C. KOCH et I. FRIED, 2000a, « Category-specific Visual Responses of Single Neurons in the Human Medial Temporal Lobe », *Nature Neuroscience*, 3 (9) : 946-953.

–, 2000b, « Imagery Neurons in the Human Brain », *Nature*, 408 (6810) : 357-361.

KREKELBERG, B. et M. LAPPE, 2001, « Neuronal Latencies and the Position of Moving Objects », *Trends in Neurosciences*, 24 (6) : 335-339.

KROLAK-SALMON, P., M. A. HENAFF, C. TALLON-BAUDRY, B. YVERT, M. GUENOT, A. VIGHETTO, F. MAUGUIERE et O. BERTRAND, 2003, « Human Lateral Geniculate Nucleus and Visual Cortex Respond to Screen Flicker », *Annals of Neurology*, 53 (1) : 73-80.

KRUGER, J. et D. DUNNING, 1999, « Unskilled and Unaware of It : How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-assessments », *Journal of Personality and Social Psychology*, 77 (6) : 1121-1134.

KUBICKI, M., H. PARK, C. F. WESTIN, P. G. NESTOR, R. V. MULKERN, S. E. MAIER, M. NIZNIKIEWICZ *et al.*, 2005, « DTI and MTR Abnormalities in Schizophrenia : Analysis of White Matter Integrity », *NeuroImage*, 26 (4) : 1109-1118.

LACHTER, J., K. I. FORSTER et E. RUTHRUFF, 2004, « Forty-five Years After Broadbent (1958) : Still No Identification Without Attention », *Psychology Review*, 111 (4) : 880-913.

LAGERCRANTZ, H. et J. P. CHANGEUX, 2009, « The Emergence of Human Consciousness : From Fetal to Neonatal Life », *Pediatric Research*, 65 (3) : 255-260.

–, 2010, « Basic Consciousness of the Newborn, » *Seminars in Perinatology*, 34 (3) : 201-206.

LAI, C. S., S. E. FISHER, J. A. HURST, F. VARGHA-KHADEM et A. P. MONACO, 2001, « A Forkhead-domain Gene Is Mutated in a Severe Speech and Language Disorder », *Nature*, 413 (6855) : 519-523.

LAMME, V. A., 2006, « Towards a True Neural Stance on Consciousness », *Trends in Cognitive Sciences*, 10 (11) : 494-501.

LAMME, V. A. et P. R. ROELFSEMA, 2000, « The Distinct Modes of Vision Offered by Feedforward and Recurrent Processing », *Trends in Neurosciences*, 23 (11) : 571-579.

LAMME, V. A., K. ZIPSER et H. SPEKREIJSE, 1998, « Figure-ground Activity in Primary Visual Cortex Is Suppressed by Anesthesia », *Proceedings of the National Academy of Sciences*, 95 (6) : 3263-368.

LAMY, D., M. SALTI et Y. BAR-HAIM, 2009, « Neural Correlates of Subjective Awareness and Unconscious Processing : An ERP Study », *Journal of Cognitive Neuroscience*, 21 (7) : 1435-1446.

LANDMAN, R., H. SPEKREIJSE et V. A. LAMME, 2003, « Large Capacity Storage of Integrated Objects Before Change Blindness », *Vision Research*, 43 (2) : 149-164.

LAU, H. et D. ROSENTHAL, 2011, « Empirical Support for Higher-order Theories of Conscious Awareness », *Trends in Cognitive Sciences*, 15 (8) : 365-373.

LAU, H. C. et R. E. PASSINGHAM, 2006, « Relative Blindsight in Normal Observers and the Neural Correlate of Visual Consciousness », *Proceedings of the National Academy of Sciences*, 103 (49) : 18763-18768.

–, 2007, « Unconscious Activation of the Cognitive Control System in the Human Prefrontal Cortex », *Journal of Neuroscience*, 27 (21) : 5805-5811.

LAUREYS, S., 2005, « The Neural Correlate of (Un)Awareness : Lessons from the Vegetative State », *Trends in Cognitive Sciences*, 9 (12) : 556-559.

LAUREYS, S., M. E. FAYMONVILLE, A. LUXEN, M. LAMY, G. FRANCK et P. MAQUET, 2000, « Restoration of Thalamocortical Connectivity After Recovery from Persistent Vegetative State », *The Lancet*, 355 (9217) : 1790-1791.

LAUREYS, S., C. LEMAIRE, P. MAQUET, C. PHILLIPS et G. FRANCK, 1999, « Cerebral Metabolism During Vegetative State and After Recovery to Consciousness », *Journal of Neurology, Neurosurgery and Psychiatry*, 67 (1) : 121.

LAUREYS, S., A. M. OWEN et N. D. SCHIFF, 2004, « Brain Function in Coma, Vegetative State, and Related Disorders », *The Lancet Neurology*, 3 (9) : 537-546.

LAUREYS, S., F. PELLAS, P. VAN EECKHOUT, S. GHORBEL, C. SCHNAKERS, F. PERRIN, J. BERRE *et al.*, 2005, « The Locked-in Syndrome : What Is It Like to Be Conscious but Paralyzed and Voiceless ? », *Progress in Brain Research*, 150 : 495-511.

LAWRENCE, N. S., F. JOLLANT, O. O'DALY, F. ZELAYA et M. L. PHILLIPS, 2009, « Distinct Roles of Prefrontal Cortical Subregions in the Iowa Gambling Task », *Cerebral Cortex*, 19 (5) : 1134-1143.

LEDOUX, J., 2005, *Le Cerveau des émotions. Les mystérieux fondements de notre vie émotionnelle*, Paris, Odile Jacob.

LENGGENHAGER, B., M. MOUTHON et O. BLANKE, 2009, « Spatial Aspects of Bodily Self-consciousness », *Consciousness and Cognition*, 18 (1) : 110-117.

LENGGENHAGER, B., T. TADI, T. METZINGER et O. BLANKE, 2007, « Video Ergo Sum : Manipulating Bodily Self-consciousness », *Science*, 317 (5841) : 1096-1099.

LEON-CARRION, J., P. VAN EECKHOUT, R. DOMINGUEZ-MORALES MDEL et F. J. PEREZ-SANTAMARIA, 2002, « The Locked-in Syndrome : A Syndrome Looking for a Therapy », *Brain Injury*, 16 (7) : 571-582.

LEOPOLD, D. A. et N. K. LOGOTHETIS, 1996, « Activity Changes in Early Visual Cortex Reflect Monkeys' Percepts During Binocular Rivalry », *Nature*, 379 (6565) : 549-553.

–, 1999, « Multistable Phenomena : Changing Views in Perception », *Trends in Cognitive Sciences*, 3 (7) : 254-264.

LEROY, F., H. GLASEL, J. DUBOIS, L. HERTZ-PANNIER, B. THIRION, J. F. MANGIN et G. DEHAENE-LAMBERTZ, 2011, « Early Maturation of the Linguistic Dorsal Pathway in Human Infants », *Journal of Neuroscience*, 31 (4) : 1500-1506.

LEVELT, W. J. M., 1989, *Speaking : From Intention to Articulation*, Cambridge, Mass., MIT Press.

LEVY, J., H. PASHLER et E. BOER, 2006, « Central Interference in Driving : Is There Any Stopping the Psychological Refractory Period ? », *Psychological Science*, 17 (3) : 228-235.

LEWIS, J. L., 1970, « Semantic Processing of Unattended Messages Using Dichotic Listening », *Journal of Experimental Psychology*, 85 (2) : 225-228.

LIBET, B., 1965, « Cortical Activation in Conscious and Unconscious Experience », *Perspectives in Biology and Medicine*, 9 (1) : 77-86.

–, 1991, « Conscious vs Neural Time », *Nature*, 352 (6330) : 27-28.

–, 2004, *Mind Time : The Temporal Factor in Consciousness*, Cambridge, Mass., Harvard University Press.

LIBET, B., W. W. ALBERTS, E. W. WRIGHT, Jr., L. D. DELATTRE, G. LEVIN et B. FEINSTEIN, 1964, « Production of Threshold Levels of Conscious Sensation by Electrical Stimulation of Human Somatosensory Cortex », *Journal of Neurophysiology*, 27 : 546-578.

LIBET, B., W. W. ALBERTS, E. W. WRIGHT, Jr. et B. FEINSTEIN, 1967, « Responses of Human Somatosensory Cortex to Stimuli Below Threshold for Conscious Sensation », *Science*, 158 (808) : 1597-1600.

LIBET, B., C. A. GLEASON, E. W. WRIGHT et D. K. PEARL, 1983, « Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-potential). The Unconscious Initiation of a Freely Voluntary Act », *Brain*, 106 (pt. 3) : 623-642.

LIBET, B., E. W. WRIGHT, Jr., B. FEINSTEIN et D. K. PEARL, 1979, « Subjective Referral of the Timing for a Conscious Sensory Experience : A

Functional Role for the Somatosensory Specific Projection System in Man », *Brain*, 102 (1) : 193-224.

Liu, Y., M. Liang, Y. Zhou, Y. He, Y. Hao, M. Song, C. Yu *et al.*, 2008, « Disrupted Small-world Networks in Schizophrenia », *Brain*, 131 (pt. 4) : 945-961.

Logan, G. D. et M. J. Crump, 2010, « Cognitive Illusions of Authorship Reveal Hierarchical Error Detection in Skilled Typists », *Science*, 330 (6004) : 683-686.

Logan, G. D. et M. D. Schulkind, 2000, « Parallel Memory Retrieval in Dual-task Situations : I. Semantic Memory », *Journal of Experimental Psychology : Human Perception and Performance*, 26 (3) : 1072-1090.

Logothetis, N. K., 1998, « Single Units and Conscious Vision », *Philosophical Transactions of the Royal Society B : Biological Sciences*, 353 (1377) : 1801-1818.

Logothetis, N. K., D. A. Leopold et D. L. Sheinberg, 1996, « What Is Rivalling During Binocular Rivalry ? », *Nature*, 380 (6575) : 621-624.

Louie, K. et M. A. Wilson, 2001, « Temporally Structured Replay of Awake Hippocampal Ensemble Activity During Rapid Eye Movement Sleep », *Neuron*, 29 (1) : 145-156.

Luck, S. J., R. L. Fuller, E. L. Braun, B. Robinson, A. Summerfelt et J. M. Gold, 2006, « The Speed of Visual Attention in Schizophrenia : Electrophysiological and Behavioral Evidence », *Schizophrenia Research*, 85 (1-3) : 174-195.

Luck, S. J., E. S. Kappenman, R. L. Fuller, B. Robinson, A. Summerfelt et J. M. Gold, 2009, « Impaired Response Selection in Schizophrenia : Evidence from the P3 Wave and the Lateralized Readiness Potential », *Psychophysiology*, 46 (4) : 776-786.

Luck, S. J., E. K. Vogel et K. L. Shapiro, 1996, « Word Meanings Can Be Accessed but Not Reported During the Attentional Blink », *Nature*, 383 (6601) : 616-618.

LUMER, E. D., G. M. EDELMAN et G. TONONI, 1997a, « Neural Dynamics in a Model of the Thalamocortical System. I : Layers, Loops and the Emergence of Fast Synchronous Rhythms », *Cerebral Cortex*, 7 (3) : 207-227.

–, 1997b, « Neural Dynamics in a Model of the Thalamocortical System. II : The Role of Neural Synchrony Tested Through Perturbations of Spike Timing », *Cerebral Cortex*, 7 (3) : 228-236.

LUMER, E. D., K. J. FRISTON et G. REES, 1998, « Neural Correlates of Perceptual Rivalry in the Human Brain », *Science*, 280 (5371) : 1930-1934.

LYNALL, M. E., D. S. BASSETT, R. KERWIN, P. J. MCKENNA, M. KITZBICHLER, U. MULLER et E. BULLMORE, 2010, « Functional Connectivity and Brain Networks in Schizophrenia », *Journal of Neuroscience*, 30 (28) : 9477-9487.

MACK, A. et I. ROCK, 1998, *Inattentional Blindness*, Cambridge, Mass., MIT Press.

MACKNIK, S. L. et M. M. HAGLUND, 1999, « Optical Images of Visible and Invisible Percepts in the Primary Visual Cortex of Primates », *Proceedings of the National Academy of Sciences*, 96 (26) : 15208-15210.

MACLEOD, D. I. et S. HE, 1993, « Visible Flicker from Invisible Patterns », *Nature*, 361 (6409) : 256-258.

MAGNUSSON, C. E. et H. C. STEVENS, 1911, « Visual Sensations Created by a Magnetic Field », *American Journal of Physiology*, 29 : 124-136.

MAIA, T. V. et J. L. MCCLELLAND, 2004, « A Reexamination of the Evidence for the Somatic Marker Hypothesis : What Participants Really Know in the Iowa Gambling Task », *Proceedings of the National Academy of Sciences*, 101 (45) : 16075-16080.

MAIER, A., M. WILKE, C. AURA, C. ZHU, F. Q. YE et D. A. LEOPOLD, 2008, « Divergence of fMRI and Neural Signals in V1 During Perceptual

Suppression in the Awake Monkey », *Nature Neuroscience*, 11 (10) : 1193-1200.

MARCEL, A. J., 1980, « Conscious and Preconscious Recognition of Polysemous Words : Locating the Selective Effect of Prior Verbal Context », *in* R. S. Nickerson (éd.), *Attention and Performance*, Hillsdale, NJ, Lawrence Erlbaum, vol. 8.

–, 1983, « Conscious and Unconscious Perception : Experiments on Visual Masking and Word Recognition », *Cognitive Psychology*, 15 : 197-237.

MAROIS, R., D. J. YI et M. M. CHUN, 2004, « The Neural Fate of Consciously Perceived and Missed Events in the Attentional Blink », *Neuron*, 41 (3) : 465-472.

MARSHALL, J. C. et P. W. HALLIGAN, 1988, « Blindsight and Insight in Visuospatial Neglect », *Nature*, 336 (6201) : 766-767.

MARTI, S., J. SACKUR, M. SIGMAN et S. DEHAENE, 2010, « Mapping Introspection's Blind Spot : Reconstruction of Dual-task Phenomenology Using Quantified Introspection », *Cognition*, 115 (2) : 303-313.

MARTI, S., M. SIGMAN et S. DEHAENE, 2012, « A Shared Cortical Bottleneck Underlying Attentional Blink and Psychological Refractory Period », *NeuroImage*, 59 (3) : 2883-2898.

MARTICORENA, D. C., A. M. RUIZ, C. MUKERJI, A. GODDU et L. R. SANTOS, 2011, « Monkeys Represent Others' Knowledge but Not Their Beliefs », *Developmental Science*, 14 (6) : 1406-1416.

MASON, M. F., M. I. NORTON, J. D. VAN HORN, D. M. WEGNER, S. T. GRAFTON et C. N. MACRAE, 2007, « Wandering Minds : The Default Network and Stimulus-independent Thought », *Science*, 315 (5810) : 393-395.

MASSIMINI, M., M. BOLY, A. CASALI, M. ROSANOVA et G. TONONI, 2009, « A Perturbational Approach for Evaluating the Brain's Capacity for Consciousness », *Progress in Brain Research*, 177 : 201-214.

MASSIMINI, M., F. FERRARELLI, R. HUBER, S. K. ESSER, H. SINGH et G. TONONI, 2005, « Breakdown of Cortical Effective Connectivity During Sleep »,

Science, 309 (5744) : 2228-2232.

MATSUDA, W., A. MATSUMURA, Y. KOMATSU, K. YANAKA et T. NOSE, 2003, « Awakenings from Persistent Vegetative State : Report of Three Cases with Parkinsonism and Brain Stem Lesions on MRI », *Journal of Neurology, Neurosurgery, and Psychiatry*, 74 (11) : 1571-1573.

MATTLER, U., 2005, « Inhibition and Decay of Motor and Nonmotor Priming », *Attention, Perception and Psychophysics*, 67 (2) : 285-300.

MAUDSLEY, H., 1868, *The Physiology and Pathology of the Mind*, Londres, Macmillan.

MAY, A., G. HAJAK, S. GANSSBAUER, T. STEFFENS, B. LANGGUTH, T. KLEINJUNG et P. EICHHAMMER, 2007, « Structural Brain Alterations Following 5 Days of Intervention : Dynamic Aspects of Neuroplasticity », *Cerebral Cortex*, 17 (1) : 205-210.

MCCARTHY, M. M., E. N. BROWN et N. KOPELL, 2008, « Potential Network Mechanisms Mediating Electroencephalographic Beta Rhythm Changes During Propofol-induced Paradoxical Excitation », *Journal of Neuroscience*, 28 (50) : 13488-13504.

MCCLURE, R. K., 2001, « The Visual Backward Masking Deficit in Schizophrenia », *Progress in Neuro-psychopharmacology and Biological Psychiatry*, 25 (2) : 301-311.

MCCORMICK, P. A., 1997, « Orienting Attention Without Awareness », *Journal of Experimental Psychology : Human Perception and Performance*, 23 (1) : 168-180.

MCGLINCHEY-BERROTH, R., W. P. MILBERG, M. VERFAELLIE, M. ALEXANDER et P. KILDUFF, 1993, « Semantic Priming in the Neglected Field : Evidence from a Lexical Decision Task », *Cognitive Neuropsychology*, 10 : 79-108.

MCGURK, H. et J. MACDONALD, 1976, « Hearing Lips and Seeing Voices », *Nature*, 264 (5588) : 746-748.

MCINTOSH, A. R., M. N. RAJAH et N. J. LOBAUGH, 1999, « Interactions of Prefrontal Cortex in Relation to Awareness in Sensory Learning », *Science*, 284 (5419) : 1531-1533.

Mehler, J., P. Jusczyk, G. Lambertz, N. Halsted, J. Bertoncini et C. Amiel-Tison, 1988, « A Precursor of Language Acquisition in Young Infants », *Cognition*, 29 (2) : 143-178.

Melloni, L., C. Molina, M. Pena, D. Torres, W. Singer et E. Rodriguez, 2007, « Synchronization of Neural Activity Across Cortical Areas Correlates with Conscious Perception », *Journal of Neuroscience*, 27 (11) : 2858-2865.

Meltzoff, A. N. et R. Brooks, 2008, « Self-experience as a Mechanism for Learning About Others : A Training Study in Social Cognition », *Developmental Psychology*, 44 (5) : 1257-1265.

Merikle, P. M., 1992, « Perception Without Awareness : Critical Issues », *American Psychologist*, 47 : 792-796.

Merikle, P. M. et S. Joordens, 1997, « Parallels Between Perception Without Attention and Perception Without Awareness », *Consciousness and Cognition*, 6 (2-3) : 219-236.

Meyer, K. et A. Damasio, 2009, « Convergence and Divergence in a Neural Architecture for Recognition and Memory », *Trends in Neurosciences*, 32 (7) : 376-382.

Miller, A., J. W. Sleigh, J. Barnard et D. A. Steyn-Ross, 2004, « Does Bispectral Analysis of the Electroencephalogram Add Anything but Complexity ? », *British Journal of Anaesthesia*, 92 (1) : 8-13.

Milner, A. D. et M. A. Goodale, 1995, *The Visual Brain in Action*, New York, Oxford University Press.

Monti, M. M., A. Vanhaudenhuyse, M. R. Coleman, M. Boly, J. D. Pickard, L. Tshibanda, A. M. Owen et S. Laureys, 2010, « Willful Modulation of Brain Activity in Disorders of Consciousness », *New England Journal of Medicine*, 362 (7) : 579-589.

Moray, N., 1959, « Attention in Dichotic Listening : Affective Cues and the Influence of Instructions », *Quarterly Journal of Experimental Psychology*, 9 : 56-60.

MORENO-BOTE, R., D. C. KNILL et A. POUGET, 2011, « Bayesian Sampling in Visual Perception », *Proceedings of the National Academy of Sciences*, 108 (30) : 12491-12496.

MORLAND, A. B., S. LE, E. CARROLL, M. B. HOFFMANN et A. PAMBAKIAN, 2004, « The Role of Spared Calcarine Cortex and Lateral Occipital Cortex in the Responses of Human Hemianopes to Visual Motion », *Journal of Cognitive Neuroscience*, 16 (2) : 204-218.

MORO, S. I., M. TOLBOOM, P. S. KHAYAT et P. R. ROELFSEMA, 2010, « Neuronal Activity in the Visual Cortex Reveals the Temporal Order of Cognitive Operations », *Journal of Neuroscience*, 30 (48) : 16293-162303.

MORRIS, J. S., B. DEGELDER, L. WEISKRANTZ et R. J. DOLAN, 2001, « Differential Extrageniculostriate and Amygdala Responses to Presentation of Emotional Faces in a Cortically Blind Field », *Brain*, 124 (pt. 6) : 1241-1252.

MORRIS, J. S., A. OHMAN et R. J. DOLAN, 1998, « Conscious and Unconscious Emotional Learning in the Human Amygdala », *Nature*, 393 (6684) : 467-470.

–, 1999, « A Subcortical Pathway to the Right Amygdala Mediating “Unseen” Fear », *Proceedings of the National Academy of Sciences*, 96 (4) : 1680-1685.

MORUZZI, G. et H. W. MAGOUN, 1949, « Brain Stem Reticular Formation and Activation of the EEG », *Electroencephalography and Clinical Neurophysiology*, 1 (4) : 455-473.

NAATANEN, R., P. PAAVILAINEN, T. RINNE et K. ALHO, 2007, « The Mismatch Negativity (MMN) in Basic Research of Central Auditory Processing : A Review », *Clinical Neurophysiology*, 118 (12) : 2544-2590.

NACCACHE, L., 2006a, « Is She Conscious ? », *Science*, 313 (5792) : 1395-1396.

–, 2006b, *Le Nouvel Inconscient*, Paris, Odile Jacob.

NACCACHE, L., E. BLANDIN et S. DEHAENE, 2002, « Unconscious Masked Priming Depends on Temporal Attention », *Psychological Science*, 13 :

416-424.

NACCACHE, L. et S. DEHAENE, 2001a, « The Priming Method : Imaging Unconscious Repetition Priming Reveals an Abstract Representation of Number in the Parietal Lobes », *Cerebral Cortex*, 11 (10) : 966-974.

–, 2001b, « Unconscious Semantic Priming Extends to Novel Unseen Stimuli », *Cognition* 80 (3) : 215-229.

NACCACHE, L., R. GAILLARD, C. ADAM, D. HASBOUN, S. CLÉMENCEAU, M. BAULAC, S. DEHAENE et L. COHEN, 2005, « A Direct Intracranial Record of Emotions Evoked by Subliminal Words », *Proceedings of the National Academy of Sciences*, 102 : 7713-7717.

NACCACHE, L., L. PUYBASSET, R. GAILLARD, E. SERVE et J. C. WILLER, 2005, « Auditory Mismatch Negativity Is a Good Predictor of Awakening in Comatose Patients : A Fast and Reliable Procedure », *Clinical Neurophysiology*, 116 (4) : 988-989.

NACHEV, P. et M. HUSAIN, 2007, « Comment on "Detecting Awareness in the Vegetative State" », *Science*, 315 (5816) : 1221 ; et réponse de l'auteur : 1221.

NELSON, C. A., K. M. THOMAS, M. DE HAAN et S. S. WEWERKA, 1998, « Delayed Recognition Memory in Infants and Adults as Revealed by Event-related Potentials », *International Journal of Psychophysiology*, 29 (2) : 145-165.

NEW, J. J. et B. J. SCHOLL, 2008, « "Perceptual Scotomas" : A Functional Account of Motion-Induced Blindness », *Psychological Science*, 19 (7) : 653-659.

NIEDER, A. et S. DEHAENE, 2009, « Representation of Number in the Brain », *Annual Review of Neuroscience*, 32 : 185-208.

NIEDER, A. et E. K. MILLER, 2004, « A Parieto-frontal Network for Visual Numerical Information in the Monkey », *Proceedings of the National Academy of Sciences*, 101 (19) : 7457-7462.

NIEUWENHUIS, S., M. S. GILZENRAT, B. D. HOLMES et J. D. COHEN, 2005, « The Role of the Locus Coeruleus in Mediating the Attentional Blink : A

Neurocomputational Theory », *Journal of Experimental Psychology : General*, 134 (3) : 291-307.

NIEUWENHUIS, S., K. R. RIDDERINKHOF, J. BLOM, G. P. BAND et A. KOK, 2001, « Error-related Brain Potentials Are Differentially Related to Awareness of Response Errors : Evidence from an Antisaccade Task », *Psychophysiology*, 38 (5) : 752-760.

NIMCHINSKY, E. A., E. GILISSEN, J. M. ALLMAN, D. P. PERL, J. M. ERWIN et P. R. HOF, 1999, « A Neuronal Morphologic Type Unique to Humans and Great Apes », *Proceedings of the National Academy of Sciences*, 96 (9) : 5268-5273.

NISBETT, R. E. et T. D. WILSON, 1977, « Telling More Than We Can Know : Verbal Reports on Mental Processes », *Psychological Review*, 84 (3) : 231-259.

NØRRETRANDERS, T., 1999, *The User Illusion : Cutting Consciousness Down to Size*, Londres, Penguin.

NORRIS, D., 2006, « The Bayesian Reader : Explaining Word Recognition as an Optimal Bayesian Decision Process », *Psychological Review*, 113 (2) : 327-357.

NORRIS, D., 2009, « Putting It All Together : A Unified Account of Word Recognition and Reaction-time Distributions », *Psychological Review*, 116 (1) : 207-219.

OCHSNER, K. N., K. KNIERIM, D. H. LUDLOW, J. HANELIN, T. RAMACHANDRAN, G. GLOVER et S. C. MACKEY, 2004, « Reflecting upon Feelings : An fMRI Study of Neural Systems Supporting the Attribution of Emotion to Self and Other », *Journal of Cognitive Neuroscience*, 16 (10) : 1746-1772.

OGAWA, S., T. M. LEE, A. R. KAY et D. W. TANK, 1990, « Brain Magnetic Resonance Imaging with Contrast Dependent on Blood Oxygenation », *Proceedings of the National Academy of Sciences*, 87 (24) : 9868-9872.

OVERGAARD, M., J. ROTE, K. MOURIDSEN et T. Z. RAMSOY, 2006, « Is Conscious Perception Gradual or Dichotomous ? A Comparison of Report

Methodologies During a Visual Task », *Consciousness and Cognition*, 15 (4) : 700-708.

Owen, A. M., M. R. Coleman, M. Boly, M. H. Davis, S. Laureys et J. D. Pickard, 2006, « Detecting Awareness in the Vegetative State », *Science*, 313 (5792) : 1402.

Owen, A., M. R. Coleman, M. Boly, M. H. Davis, S. Laureys, D. Jolles et J. D. Pickard, 2007, « Response to Comments on "Detecting Awareness in the Vegetative State" », *Science*, 315 : 1221.

Pack, C. C., V. K. Berezovskii et R. T. Born, 2001, « Dynamic Properties of Neurons in Cortical Area MT in Alert and Anaesthetized Macaque Monkeys », *Nature*, 414 (6866) : 905-908.

Pack, C. C. et R. T. Born, 2001, « Temporal Dynamics of a Neural Solution to the Aperture Problem in Visual Area MT of Macaque Brain », *Nature*, 409 (6823) : 1040-1042.

Pallier, C., A. D. Devauchelle et S. Dehaene, 2011, « Cortical Representation of the Constituent Structure of Sentences », *Proceedings of the National Academy of Sciences*, 108 (6) : 2522-2527.

Palva, S., K. Linkenkaer-Hansen, R. Naatanen et J. M. Palva, 2005, « Early Neural Correlates of Conscious Somatosensory Perception », *Journal of Neuroscience*, 25 (21) : 5248-5258.

Parvizi, J. et A. R. Damasio, 2003, « Neuroanatomical Correlates of Brainstem Coma », *Brain*, 126 (pt. 7) : 1524-1536.

Parvizi, J., C. Jacques, B. L. Foster, N. Withoft, V. Rangarajan, K. S. Weiner et K. Grill-Spector, 2012, « Electrical Stimulation of Human Fusiform Face-selective Regions Distorts Face Perception », *Journal of Neuroscience*, 32 (43) : 14915-14920.

Parvizi, J., G. W. Van Hoesen, J. Buckwalter et A. Damasio, 2006, « Neural Connections of the Posteromedial Cortex in the Macaque », *Proceedings of the National Academy of Sciences*, 103 (5) : 1563-1568.

Pascual-Leone, A., V. Walsh et J. Rothwell, 2000, « Transcranial Magnetic Stimulation in Cognitive Neuroscience – Virtual Lesion, Chronometry

et Functional Connectivity », *Current Opinion in Neurobiology*, 10 (2) : 232-237.

PASHLER, H., 1984, « Processing Stages in Overlapping Tasks : Evidence for a Central Bottleneck », *Journal of Experimental Psychology : Human Perception and Performance*, 10 (3) : 358-377.

–, 1994, « Dual-task Interference in Simple Tasks : Data and Theory », *Psychological Bulletin*, 116 (2) : 220-244.

PEIRCE, C. S., 1901, « The Proper Treatment of Hypotheses : A Preliminary Chapitre, toward an Examination of Hume's Argument Against Miracles, in Its Logic and in Its History », *Historical Perspectives*, 2 : 890-904.

PENROSE, R. et S. HAMEROFF, 1998, « The Penrose-Hameroff "Orch OR" Model of Consciousness », *Philosophical Transactions of the Royal Society London (A)*, 356 : 1869-1896.

PERIN, R., T. K. BERGER et H. MARKRAM, 2011, « A Synaptic Organizing Principle for Cortical Neuronal Groups », *Proceedings of the National Academy of Sciences*, 108 (13) : 5419-5424.

PERNER, J. et M. AICHHORN, 2008, « Theory of Mind, Language and the Temporoparietal Junction Mystery », *Trends in Cognitive Sciences*, 12 (4) : 123-126.

PERSAUD, N., M. DAVIDSON, B. MANISCALCO, D. MOBBS, R. E. PASSINGHAM, A. COWEY et H. LAU, 2011, « Awareness-related Activity in Prefrontal and Parietal Cortices in Blindsight Reflects More than Superior Visual Performance », *NeuroImage*, 58 (2) : 605-611.

PESSIGLIONE, M., P. PETROVIC, J. DAUNIZEAU, S. PALMINTERI, R. J. DOLAN et C. D. FRITH, 2008, « Subliminal Instrumental Conditioning Demonstrated in the Human Brain », *Neuron*, 59 (4) : 561-567.

PESSIGLIONE, M., L. SCHMIDT, B. DRAGANSKI, R. KALISCH, H. LAU, R. J. DOLAN et C. D. FRITH, 2007, « How the Brain Translates Money Into Force : A Neuroimaging Study of Subliminal Motivation », *Science*, 316 (5826) : 904-906.

Petersen, S. E., H. Van Mier, J. A. Fiez et M. E. Raichle, 1998, « The Effects of Practice on the Functional Anatomy of Task Performance », *Proceedings of the National Academy of Sciences*, 95 (3) : 853-860.

Peyrache, A., M. Khamassi, K. Benchenane, S. I. Wiener et F. P. Battaglia, 2009, « Replay of Rule-Learning Related Neural Patterns in the Prefrontal Cortex During Sleep », *Nature Neuroscience*, 12 (7) : 919-926.

Piazza, M., V. Izard, P. Pinel, D. Le Bihan et S. Dehaene, 2004, « Tuning Curves for Approximate Numerosity in the Human Intraparietal Sulcus », *Neuron*, 44 (3) : 547-555.

Piazza, M., P. Pinel, D. Le Bihan et S. Dehaene, 2007, « A Magnitude Code Common to Numerosities and Number Symbols in Human Intraparietal Cortex », *Neuron*, 53 : 293-305.

Picton, T. W., 1992, « The P300 Wave of the Human Event-Related Potential », *Journal of Clinical Neurophysiology*, 9 (4) : 456-479.

Pinel, P., F. Fauchereau, A. Moreno, A. Barbot, M. Lathrop, D. Zelenika, D. Le Bihan *et al.*, 2012, « Genetic variants of FOXP2 and KIAA0319/TTRAP/THEM2 Locus Are Associated with Altered Brain Activation in Distinct Language-Related Regions », *Journal of Neuroscience*, 32 (3) : 817-825.

Pins, D. et D. Ffytche, 2003, « The Neural Correlates of Conscious Vision », *Cerebral Cortex*, 13 (5) : 461-474.

Pisella, L., H. Grea, C. Tilikete, A. Vighetto, M. Desmurget, G. Rode, D. Boisson et Y. Rossetti, 2000, « An “Automatic Pilot” for the Hand in Human Posterior Parietal Cortex : Toward Reinterpreting Optic Ataxia », *Nature Neuroscience*, 3 (7) : 729-736.

Plotnik, J. M., F. B. de Waal et D. Reiss, 2006, « Self-recognition in an Asian Elephant, » *Proceedings of the National Academy of Sciences*, 103 (45) : 17053-17057.

Portas, C. M., K. Krakow, P. Allen, O. Josephs, J. L. Armony et C. D. Frith, 2000, « Auditory Processing Across the Sleep-wake Cycle :

Simultaneous EEG and fMRI Monitoring in Humans », *Neuron*, 28 (3) : 991-999.

POSNER, M. I., 1994, « Attention : The Mechanisms of Consciousness », *Proceedings of the National Academy of Sciences*, 91 : 7398-7403.

POSNER, M. I. et M. K. ROTHBART, 1998, « Attention, Self-regulation and Consciousness », *Philosophical Transactions of the Royal Society B : Biological Sciences*, 353 (1377) : 1915-1927.

POSNER, M. I. et C. R. R. SNYDER, 1975/2004, « Attention and Cognitive Control », *in* D. A. Balota et E. J. Marsh (éd.), *Cognitive Psychology : Key Readings,* New York, Psychology Press : 205-223.

POSNER, N. I. et C. R. R. SNYDER, 1975, « Attention and Cognitive Control », *in* R. L. Solso (éd.), *Information Processing and Cognition : The Loyola Symposium*, Hillsdale, Lawrence Erlbaum : 55-85.

PRIOR, H., A. SCHWARZ et O. GUNTURKUN, 2008, « Mirror-induced Behavior in the Magpie (Pica Pica) : Evidence of Self-recognition », *PLOS Biology*, 6 (8) : e202.

QUIROGA, R. Q., G. KREIMAN, C. KOCH et I. FRIED, 2008, « Sparse But Not "Grandmother-cell" Coding in the Medial Temporal Lobe », *Trends in Cognitive Sciences*, 12 (3) : 87-91.

QUIROGA, R. Q., R. MUKAMEL, E. A. ISHAM, R. MALACH et I. FRIED, 2008, « Human Single-neuron Responses at the Threshold of Conscious Recognition », *Proceedings of the National Academy of Sciences*, 105 (9) : 3599-3604.

QUIROGA, R. Q., L. REDDY, C. KOCH et I. FRIED, 2007, « Decoding Visual Inputs from Multiple Neurons in the Human Temporal Lobe », *Journal of Neurophysiology*, 98 (4) : 1997-2007.

QUIROGA, R. Q., L. REDDY, G. KREIMAN, C. KOCH et I. FRIED, 2005, « Invariant Visual Representation by Single Neurons in the Human Brain », *Nature* 435 (7045) : 1102-1107.

RAICHLE, M. E., 2010, « Two Views of Brain Function », *Trends in Cognitive Sciences*, 14 (4) : 180-190.

RAICHLE, M. E., J. A. FIESZ, T. O. VIDEEN et A. K. MACLEOD, 1994, « Practice-related Changes in Human Brain Functional Anatomy During Nonmotor Learning », *Cerebral Cortex*, 4 : 8-26.

RAICHLE, M. E., A. M. MACLEOD, A. Z. SNYDER, W. J. POWERS, D. A. GUSNARD et G. L. SHULMAN, 2001, « A Default Mode of Brain Function », *Proceedings of the National Academy of Sciences*, 98 (2) : 676-682.

RAILO, H. et M. KOIVISTO, 2009, « The Electrophysiological Correlates of Stimulus Visibility and Metacontrast Masking », *Consciousness and Cognition*, 18 (3) : 794-803.

RAMACHANDRAN, V. S. et R. L. GREGORY, 1991, « Perceptual Filling in of Artificially Induced Scotomas in Human Vision », *Nature*, 350 (6320) : 699-702.

RAYMOND, J. E., K. L. SHAPIRO et K. M. ARNELL, 1992, « Temporary Suppression of Visual Processing in an RSVP Task : An Attentional Blink ? », *Journal of Experimental Psychology : Human Perception and Performance*, 18 (3) : 849-860.

REDDY, L., R. Q. QUIROGA, P. WILKEN, C. KOCH et I. FRIED, 2006, « A Single-neuron Correlate of Change Detection and Change Blindness in the Human Medial Temporal Lobe », *Current Biology*, 16 (20) : 2066-2072.

REED, C. M. et N. I. DURLACH, 1998, « Note on Information Transfer Rates in Human Communication », *Presence : Teleoperators and Virtual Environments*, 7 (5) : 509-518.

REISS, D. et L. MARINO, 2001, « Mirror Self-recognition in the Bottlenose Dolphin : A Case of Cognitive Convergence », *Proceedings of the National Academy of Sciences*, 98 (10) : 5937-5942.

RENSINK, R. A., J. K. O'REGAN et J. CLARK, 1997, « To See or Not to See : The Need for Attention to Perceive Changes in Scenes », *Psychological Science*, 8 : 368-373.

REUSS, H., A. KIESEL, W. KUNDE et B. HOMMEL, 2011, « Unconscious Activation of Task Sets », *Consciousness and Cognition*, 20 (3) : 556-

567.

REUTER, F., A. DEL CUL, B. AUDOIN, I. MALIKOVA, L. NACCACHE, J. P. RANJEVA, O. LYON-CAEN *et al.*, 2007, « Intact Subliminal Processing and Delayed Conscious Access in Multiple Sclerosis », *Neuropsychologia*, 45 (12) : 2683-2691.

REUTER, F., A. DEL CUL, I. MALOKOVA, L. NACCACHE, S. CONFORT-GOUNY, L. COHEN, A. A. CHERIF *et al.*, 2009, « White Matter Damage Impairs Access to Consciousness in Multiple Sclerosis », *NeuroImage*, 44 (2) : 590-599.

REYNVOET, B. et M. BRYSBAERT, 1999, « Single-digit and Two-digit Arabic Numerals Address the Same Semantic Number Line », *Cognition*, 72 (2) : 191-201.

–, 2004, « Cross-notation Number Priming Investigated at Different Stimulus Onset Asynchronies in Parity and Naming Tasks », *Journal of Experimental Psychology*, 51 (2) : 81-90.

REYNVOET, B., M. BRYSBAERT et W. FIAS, 2002, « Semantic Priming in Number Naming », *Quarterly Journal of Experimental Psychology A*, 55 (4) : 1127-1139.

REYNVOET, B., W. GEVERS et B. CAESSENS, 2005, « Unconscious Primes Activate Motor Codes Through Semantics », *Journal of Experimental Psychology : Learning, Memory, Cognition*, 31 (5) : 991-1000.

RICŒUR, P., 1990, *Soi-même comme un autre*, Paris, Seuil.

RIGAS, P. et M. A. CASTRO-ALAMANCOS, 2007, « Thalamocortical Up States : Differential Effects of Intrinsic and Extrinsic Cortical Inputs on Persistent Activity », *Journal of Neuroscience*, 27 (16) : 4261-4272.

ROCKSTROH, B., M. MÜLLER, R. COHEN et T. ELBERT, 1992, « Probing the Functional Brain State During P300 Evocation », *Journal of Psychophysiology*, 6 : 175-184.

RODRIGUEZ, E., N. GEORGE, J.-P. LACHAUX, J. MARTINERIE, B. RENAULT et F. J. VARELA, 1999, « Perception's Shadow : Long-distance Synchronization of Human Brain Activity », *Nature*, 397 (6718) : 430-433.

ROELFSEMA, P. R., 2005, « Elemental Operations in Vision », *Trends in Cognitive Sciences*, 9 (5) : 226-333.

ROELFSEMA, P. R., P. S. KHAYAT et H. SPEKREIJSE, 2003, « Subtask Sequencing in the Primary Visual Cortex », *Proceedings of the National Academy of Sciences*, 100 (9) : 5467-5472.

ROELFSEMA, P. R., V. A. LAMME et H. SPEKREIJSE, 1998, « Object-based Attention in the Primary Visual Cortex of the Macaque Monkey », *Nature*, 395 (6700) : 376-381.

ROPPER, A. H., 2010, « Cogito Ergo Sum by MRI », *New England Journal of Medicine*, 362 (7) : 648-649.

ROSANOVA, M., O. GOSSERIES, S. CASAROTTO, M. BOLY, A. G. CASALI, M. A. BRUNO, M. MARIOTTI *et al.*, 2012, « Recovery of Cortical Effective Connectivity and Recovery of Consciousness in Vegetative Patients », *Brain*, 135 (pt. 4) : 1308-1320.

ROSENTHAL, D. M., 2008, « Consciousness and Its Function », *Neuropsychologia*, 46 (3) : 829-840.

ROSS, C. A., R. L. MARGOLIS, S. A. READING, M. PLETNIKOV et J. T. COYLE, 2006, « Neurobiology of Schizophrenia », *Neuron*, 52 (1) : 139-153.

ROUGIER, N. P., D. C. NOELLE, T. S. BRAVER, J. D. COHEN et R. C. O'REILLY, 2005, « Prefrontal Cortex and Flexible Cognitive Control : Rules Without Symbols », *Proceedings of the National Academy of Sciences*, 10 (*2*20) : 7338-7343.

ROUNIS, E., B. MANISCALCO, J. C. ROTHWELL, R. PASSINGHAM et H. LAU, 2010, « Theta-burst Transcranial Magnetic Stimulation to the Prefrontal Cortex Impairs Metacognitive Visual Awareness », *Cognitive Neuroscience*, 1 (3) : 165-175.

SACKUR, J. et S. DEHAENE, 2009, « The Cognitive Architecture for Chaining of Two Mental Operations », *Cognition*, 111 (2) : 187-211.

SACKUR, J., L. NACCACHE, P. PRADAT-DIEHL, P. AZOUVI, D. MAZEVET, R. KATZ, L. COHEN et S. DEHAENE, 2008, « Semantic Processing of Neglected Numbers », *Cortex*, 44 (6) : 673-682.

SADAGHIANI, S., G. HESSELMANN, K. J. FRISTON et A. KLEINSCHMIDT, 2010, « The Relation of Ongoing Brain Activity, Evoked Neural Responses, and Cognition », *Frontiers in Systems Neuroscience*, 4 : 20.

SADAGHIANI, S., G. HESSELMANN et A. KLEINSCHMIDT, 2009, « Distributed and Antagonistic Contributions of Ongoing Activity Fluctuations to Auditory Stimulus Detection », *Journal of Neuroscience*, 29 (42) : 13410-13417.

SAGA, Y., M. IBA, J. TANJI et E. HOSHI, 2011, « Development of Multidimensional Representations of Task Phases in the Lateral Prefrontal Cortex », *Journal of Neuroscience*, 31 (29) : 10648-10665.

SAHRAIE, A., L. WEISKRANTZ, J. L. BARBUR, A. SIMMONS, S. C. R. WILLIAMS et M. J. BRAMMER, 1997, « Pattern of Neuronal Activity Associated with Conscious and Unconscious Processing of Visual Signals », *Proceedings of the National Academy of Sciences*, 94 : 9406-9411.

SALIN, P. A. et J. BULLIER, 1995, « Corticocortical Connections in the Visual System : Structure and Function », *Physiological Reviews*, 75 (1) : 107-154.

SAUR, D., B. SCHELTER, S. SCHNELL, D. KRATOCHVIL, H. KUPPER, P. KELLMEYER, D. KUMMERER *et al.*, 2010, « Combining Functional and Anatomical Connectivity Reveals Brain Networks for Auditory Language Comprehension », *NeuroImage*, 49 (4) : 3187-3197.

SAXE, R., 2006, « Uniquely Human Social Cognition », *Current Opinion in Neurobiology*, 16 (2) : 235-239.

SAXE, R. et L. J. POWELL, 2006, « It's the Thought That Counts : Specific Brain Regions for One Component of Theory of Mind », *Psychological Science*, 17 (8) : 692-699.

SCHENKER, N. M., D. P. BUXHOEVEDEN, W. L. BLACKMON, K. AMUNTS, K. ZILLES et K. SEMENDEFERI, 2008, « A Comparative Quantitative Analysis of Cytoarchitecture and Minicolumnar Organization in Broca's Area in Humans and Great Apes », *Journal of Comparative Neurology*, 510 (1) : 117-128.

SCHENKER, N. M., W. D. HOPKINS, M. A. SPOCTER, A. R. GARRISON, C. D. STIMPSON, J. M. ERWIN, P. R. HOF et C. C. SHERWOOD, 2009, « Broca's Area Homologue in Chimpanzees (*Pan troglodytes*) : Probabilistic Mapping, Asymmetry, and Comparison to Humans », *Cerebral Cortex*, 20 (3) : 730-742.

SCHIFF, N., U. RIBARY, F. PLUM et R. LLINAS, 1999, « Words Without Mind », *Journal of Cognitive Neuroscience*, 11 (6) : 650-656.

SCHIFF, N. D. 2010. « Recovery of Consciousness After Brain Injury : A Mesocircuit Hypothesis », *Trends in Neurosciences*, 33 (1) : 1-9.

SCHIFF, N. D., J. T. GIACINO, K. KALMAR, J. D. VICTOR, K. BAKER, M. GERBER, B. FRITZ *et al.*, 2007, « Behavioural Improvements with Thalamic Stimulation After Severe Traumatic Brain Injury », *Nature*, 448 (7153) : 600-303.

–, 2008, « Behavioural Improvements with Thalamic Stimulation After Severe Traumatic Brain Injury », *Nature*, 452 (7183) : 120.

SCHIFF, N. D., U. RIBARY, D. R. MORENO, B. BEATTIE, E. KRONBERG, R. BLASBERG, J. GIACINO *et al.*, 2002, « Residual Cerebral Activity and Behavioural Fragments Can Remain in the Persistently Vegetative Brain », *Brain*, 125 (pt. 6) : 1210-1234.

SCHILLER, P. H. et S. L. CHOROVER, 1966, « Metacontrast : Its Relation to Evoked Potentials », *Science*, 153 (742) : 1398-1400.

SCHMID, M. C., S. W. MROWKA, J. TURCHI, R. C. SAUNDERS, M. WILKE, A. J. PETERS, F. Q. YE et D. A. LEOPOLD, 2010, « Blindsight Depends on the Lateral Geniculate Nucleus », *Nature*, 466 (7304) : 373-377.

SCHMID, M. C., T. PANAGIOTAROPOULOS, M. A. AUGATH, N. K. LOGOTHETIS et S. M. SMIRNAKIS, 2009, « Visually Driven Activation in Macaque Areas V2 and V3 Without Input from the Primary Visual Cortex », *PLOS One*, 4 (5) : e5527.

SCHNAKERS, C., D. LEDOUX, S. MAJERUS, P. DAMAS, F. DAMAS, B. LAMBERMONT, M. LAMY *et al.*, 2008, « Diagnostic and Prognostic Use of Bispectral

Index in Coma, Vegetative State and Related Disorders », *Brain Injury*, 22 (12) : 926-931.

SCHNAKERS, C., A. VANHAUDENHUYSE, J. GIACINO, M. VENTURA, M. BOLY, S. MAJERUS, G. MOONEN et S. LAUREYS, 2009, « Diagnostic Accuracy of the Vegetative and Minimally Conscious State : Clinical Consensus Versus Standardized Neurobehavioral Assessment », *BMC Neurology*, 9 : 35.

SCHNEIDER, W. et R. M. SHIFFRIN, 1977, « Controlled and Automatic Human Information Processing. I : Detection, Search, and Attention », *Psychological Review*, 84 (1) : 1-66.

SCHOENEMANN, P. T., M. J. SHEEHAN et L. D. GLOTZER, 2005, « Prefrontal White Matter Volume Is Disproportionately Larger in Humans Than in Other Primates », *Nature Neuroscience*, 8 (2) : 242-252.

SCHURGER, A., F. PEREIRA, A. TREISMAN et J. D. COHEN, 2009, « Reproducibility Distinguishes Conscious from Nonconscious Neural Representations », *Science*, 327 (5961) : 97-99.

SCHURGER, A., J. D. SITT et S. DEHAENE, 2012, « An Accumulator Model for Spontaneous Neural Activity Prior to Self-initiated Movement », *Proceedings of the National Academy of Sciences*, 109 (42) : E2904-13.

SCHVANEVELDT, R. W. et D. E. MEYER, 1976, « Lexical Ambiguity, Semantic Context, and Visual Word Recognition », *Journal of Experimental Psychology : Human Perception and Performance*, 2 (2) : 243-256.

SELF, M. W., R. N. KOOIJMANS, H. SUPER, V. A. LAMME et P. R. ROELFSEMA, 2012, « Different Glutamate Receptors Convey Feedforward and Recurrent Processing in Macaque V1 », *Proceedings of the National Academy of Sciences*, 109 (27) : 11031-11036.

SELFRIDGE, O. G., 1959, « Pandemonium : A Paradigm for Learning », *in* D. V. Blake et A. M. Uttley (éd.), *Proceedings of the Symposium on Mechanisation of Thought Processes*, Londres, H. M. Stationery Office : 511-529.

SELIMBEYOGLU, A. et J. PARVIZI, 2010, « Electrical Stimulation of the Human Brain : Perceptual and Behavioral Phenomena Reported in the Old and

New Literature », *Frontiers of Human Neuroscience*, 4 : 46.

SERGENT, C., S. BAILLET et S. DEHAENE, 2005, « Timing of the Brain Events Underlying Access to Consciousness During the Attentional Blink », *Nature Neuroscience*, 8 (10) : 1391-1400.

SERGENT, C. et S. DEHAENE, 2004, « Is Consciousness a Gradual Phenomenon ? Evidence for an All-or-none Bifurcation During the Attentional Blink », *Psychological Science*, 15 (11) : 720-728.

SERGENT, C., V. WYART, M. BABO-REBELO, L. COHEN, L. NACCACHE et C. TALLON-BAUDRY, 2013, « Cueing Attention After the Stimulus Is Gone Can Retrospectively Trigger Conscious Perception », *Current Biology*, 23 (2) : 150-155.

SHADY, S., D. I. MACLEOD et H. S. FISHER, 2004, « Adaptation from Invisible Flicker », *Proceedings of the National Academy of Sciences*, 101 (14) : 5170-5173.

SHALLICE, T., 1972, « Dual Functions of Consciousness », *Psychological Review*, 79 (5) : 383-393.

–, 1979, « A Theory of Consciousness », *Science*, 204 (4395) : 827.

–, 1995, *Symptômes et modèles en neuropsychologie. Des schémas aux réseaux*, Paris, Presses universitaires de France.

SHANAHAN, M. et B. BAARS, 2005, « Applying Global Workspace Theory to the Frame Problem », *Cognition*, 98 (2) : 157-176.

SHAO, L., Y. SHUAI, J. WANG, S. FENG, B. LU, Z. LI, Y. ZHAO *et al.*, 2011, « Schizophrenia Susceptibility Gene Dysbindin Regulates Glutamatergic and Dopaminergic Functions via Distinctive Mechanisms in Drosophila », *Proceedings of the National Academy of Sciences*, 108 (46) : 18831-18836.

SHERMAN, S. M., 2012, « Thalamocortical Interactions », *Current Opinion in Neurobiology*, 22 (4) : 575-579.

SHIFFRIN, R. M. et W. SCHNEIDER, 1977, « Controlled and Automatic Human Information Processing. II : Perceputal Learning, Automatic Attending, and a General Theory », *Psychological Review*, 84 (2) : 127-190.

SHIMA, K., M. ISODA, H. MUSHIAKE et J. TANJI, 2007, « Categorization of Behavioural Sequences in the Prefrontal Cortex », *Nature*, 445 (7125) : 315-318.

SHIRVALKAR, P., M. SETH, N. D. SCHIFF et D. G. HERRERA, 2006, « Cognitive Enhancement with Central Thalamic Electrical Stimulation », *Proceedings of the National Academy of Sciences*, 103 (45) : 17007-17012.

SIDAROS, A., A. W. ENGBERG, K. SIDAROS, M. G. LIPTROT, M. HERNING, P. PETERSEN, O. B. PAULSON *et al.*, 2008, « Diffusion Tensor Imaging During Recovery from Severe Traumatic Brain Injury and Relation to Clinical Outcome : A Longitudinal Study », *Brain*, 131 (pt. 2) : 559-572.

SIDIS, B., 1898, *The Psychology of Suggestion*, New York, D. Appleton.

SIEGLER, R. S., 1987, « Strategy Choices in Subtraction », *in* J. Sloboda et D. Rogers (éd.), *Cognitive Processes in Mathematics*, Oxford, Clarendon Press : 81-106.

–, 1988, « Strategy Choice Procedures and the Development of Multiplication Skill », *Journal of Experimental Psychology : General*, 117 (3) : 258-275.

–, 1989, « Mechanisms of Cognitive Development », *Annual Review of Psychology*, 40 : 353-379.

SIEGLER, R. S. et E. A. JENKINS, 1989, *How Children Discover New Strategies*, Hillsdale, NJ, Lawrence Erlbaum.

SIGALA, N., M. KUSUNOKI, I. NIMMO-SMITH, D. GAFFAN et J. DUNCAN, 2008, « Hierarchical Coding for Sequential Task Events in the Monkey Prefrontal Cortex », *Proceedings of the National Academy of Sciences*, 105 (33) : 11969-11974.

SIGMAN, M. et S. DEHAENE, 2005, « Parsing a Cognitive Task : A Characterization of the Mind's Bottleneck », *PLOS Biology*, 3 (2) : e37.

–, 2008, « Brain Mechanisms of Serial and Parallel Processing During Dual-task Performance », *Journal of Neuroscience*, 28 (30) : 7585-7598.

Silvanto, J. et Z. Cattaneo, 2010, « Transcranial Magnetic Stimulation Reveals the Content of Visual Short-term Memory in the Visual Cortex », *NeuroImage*, 50 (4) : 1683-1689.

Silvanto, J., A. Cowey, N. Lavie et V. Walsh, 2005, « Striate Cortex (V1) Activity Gates Awareness of Motion », *Nature Neuroscience*, 8 (2) : 143-144.

Silvanto, J., N. Lavie et V. Walsh, 2005, « Double Dissociation of V1 and V5/MT Activity in Visual Awareness », *Cerebral Cortex*, 15 (11) : 1736-1741.

Simons, D. J. et M. S. Ambinder, 2005, « Change Blindness : Theory and Consequences », *Current Directions in Psychological Science*, 14 (1) : 44-48.

Simons, D. J. et C. F. Chabris, 1999, « Gorillas in Our Midst : Sustained Inattentional Blindness for Dynamic Events », *Perception*, 28 (9) : 1059-1074.

Singer, P., 1993, *Practical Ethics*, Cambridge, Cambridge University Press, 2e éd.

Singer, W., 1998, « Consciousness and the Structure of Neuronal Representations », *Philosophical Transactions of the Royal Society B : Biological Sciences*, 353 (1377) : 1829-1840.

Sitt, J. D., J. R. King, I. El Karoui, B. Rohaut, F. Faugeras, A. Gramfort, L. Cohen *et al.*, 2014, « Large Scale Screening of the Neural Signatures of Consciousness in Vegetative and Minimally Conscious State Patients », *Brain*, 137 (Pt. 8) : 2258-2270.

Sklar, A. Y., N. Levy, A. Goldstein, R. Mandel, A. Maril et R. R. Hassin, 2012, « Reading and Doing Arithmetic Nonconsciously », *Proceedings of the National Academy of Sciences*, 109 (48) : 19614-19619.

Smallwood, J., E. Beach, J. W. Schooler et T. C. Handy, 2008, « Going AWOL in the Brain : Mind Wandering Reduces Cortical Analysis of External Events », *Journal of Cognitive Neuroscience*, 20 (3) : 458-469.

SMEDIRA, N. G., B. H. EVANS, L. S. GRAIS, N. H. COHEN, B. LO, M. COOKE, W. P. SCHECTER *et al.*, 1990, « Withholding and Withdrawal of Life Support from the Critically Ill. », *New England Journal of Medicine*, 322 (5) : 309-315.

SMITH, J. D., J. SCHULL, J. STROTE, K. MCGEE, R. EGNOR et L. ERB, 1995, « The Uncertain Response in the Bottlenosed Dolphin (*Tursiops truncatus*) », *Journal of Experimental Psychology : General*, 124 (4) : 391-408.

SOTO, D., T. MANTYLA et J. SILVANTO, 2011, « Working Memory Without Consciousness », *Current Biology*, 21 (22) : R912-R913.

SPORNS, O., G. TONONI et G. M. EDELMAN, 1991, « Modeling Perceptual Grouping and Figure-ground Segregation by Means of Active Reentrant Connections », *Proceedings of the National Academy of Sciences*, 88 (1) : 129-133.

SQUIRES, K. C., C. WICKENS, N. K. SQUIRES et E. DONCHIN, 1976, « The Effect of Stimulus Sequence on the Waveform of the Cortical Event-related Potential », *Science*, 193 (4258) : 1142-1146.

SQUIRES, N. K., K. C. SQUIRES et S. A. HILLYARD, 1975, « Two Varieties of Long-latency Positive Waves Evoked by Unpredictable Auditory Stimuli in Man », *Electroencephalography and Clinical Neurophysiology*, 38 (4) : 387-401.

SRINIVASAN, R., D. P. RUSSELL, G. M. EDELMAN et G. TONONI, 1999, « Increased Synchronization of Neuromagnetic Responses During Conscious Perception », *Journal of Neuroscience*, 19 (13) : 5435-5448.

STANIEK, M. et K. LEHNERTZ, 2008, « Symbolic Transfer Entropy », *Physical Review Letters*, 100 (15) : 158101.

STAUNTON, H., 2008, « Arousal by Stimulation of Deep-Brain Nuclei », *Nature*, 452 (7183) : E1 ; et discussion : E1-2.

STEPHAN, K. E., K. J. FRISTON et C. D. FRITH, 2009, « Dysconnection in Schizophrenia : From Abnormal Synaptic Plasticity to Failures of Self-monitoring », *Schizophrenia Bulletin*, 35 (3) : 509-527.

STEPHAN, K. M., M. H. THAUT, G. WUNDERLICH, W. SCHICKS, B. TIAN, L. TELLMANN, T. SCHMITZ *et al.*, 2002, « Conscious and Subconscious Sensorimotor Synchronization – Prefrontal Cortex and the Influence of Awareness », *NeuroImage*, 15 (2) : 345-352.

STETTLER, D. D., A. DAS, J. BENNETT et C. D. GILBERT, 2002, « Lateral Connectivity and Contextual Interactions in Macaque Primary Visual Cortex », *Neuron*, 36 (4) : 739-750.

STEYN-ROSS, M. L., D. A. STEYN-ROSS et J. W. SLEIGH, 2004, « Modelling General Anaesthesia as a First-order Phase Transition in the Cortex », *Progress in Biophysics and Molecular Biology*, 85 (2-3) : 369-385.

STRAYER, D. L., F. A. DREWS et W. A. JOHNSTON, 2003, « Cell Phone-induced Failures of Visual Attention During Simulated Driving », *Journal of Experimental Psychology : Applied*, 9 (1) : 23-32.

STRIEM-AMIT, E., L. COHEN, S. DEHAENE et A. AMEDI, 2012, « Reading with Sounds : Sensory Substitution Selectively Activates the Visual Word Form Area in the Blind », *Neuron*, 76 (3) : 640-652.

SUDDENDORF, T. et D. L. BUTLER, 2013, « The Nature of Visual Self-recognition », *Trends in Cognitive Sciences*, 17 (3) : 121-127.

SUPÈR, H., H. SPEKREIJSE et V. A. LAMME, 2001a, « Two Distinct Modes of Sensory Processing Observed in Monkey Primary Visual Cortex (V1) », *Nature Neuroscience*, 4 (3) : 304-310.

–, 2001b, « A Neural Correlate of Working Memory in the Monkey Primary Visual Cortex », *Science*, 293 (5527) : 120-124.

SUPÈR, H., C. VAN DER TOGT, H. SPEKREIJSE et V. A. LAMME, 2003, « Internal State of Monkey Primary Visual Cortex (V1) Predicts Figure-ground Perception », *Journal of Neuroscience*, 23 (8) : 3407-3414.

SUPP, G. G., M. SIEGEL, J. F. HIPP et A. K. ENGEL, 2011, « Cortical Hypersynchrony Predicts Breakdown of Sensory Processing During Loss of Consciousness », *Current Biology*, 21 (23) : 1988-1993.

TAINE, H., 1870, *De l'intelligence*, Paris, Hachette.

TANG, T. T., F. YANG, B. S. CHEN, Y. LU, Y. JI, K. W. ROCHE et B. LU, 2009, « Dysbindin Regulates Hippocampal LTP by Controlling NMDA Receptor Surface Expression », *Proceedings of the National Academy of Sciences*, 106 (50) : 21395-21400.

TAYLOR, P. C., V. WALSH et M. EIMER, 2010, « The Neural Signature of Phosphene Perception », *Human Brain Mapping*, 31 (9) : 1408-1417.

TELFORD, C. W., 1931, « The Refractory Phase of Voluntary and Associative Responses », *Journal of Experimental Psychology*, 14 (1) : 1-36.

TERRACE, H. S. et L. K. SON, 2009, « Comparative Metacognition », *Current Opinion in Neurobiology*, 19 (1) : 67-74.

THOMPSON, S. P., 1910, « A Physiological Effect of an Alternating Magnetic Field », *Proceedings of the Royal Society B : Biological Sciences*, B82 : 396-399.

TOMBU, M. et P. JOLICŒUR, 2003, « A Central Capacity Sharing Model of Dual-task Performance », *Journal of Experimental Psychology : Human Perception and Performance*, 29 (1) : 3-18.

TONONI, G., 2008, « Consciousness as Integrated Information : A Provisional Manifesto », *Biological Bulletin*, 215 (3) : 216-242.

TONONI, G. et G. M. EDELMAN, 1998, « Consciousness and Complexity », *Science*, 282 (5395) : 1846-1851.

TOOLEY, M., 1972, « Abortion and Infanticide », *Philosophy and Public Affairs*, 2 (1) : 37-65.

–, 1983, *Abortion and Infanticide*, Londres, Clarendon Press.

TREISMAN, A. et G. GELADE, 1980, « A Feature-integration Theory of Attention », *Cognitive Psychology*, 12 : 97-136.

TREISMAN, A. et J. SOUTHER, 1986, « Illusory Words : The Roles of Attention and of Top-down Constraints in Conjoining Letters to Form Words », *Journal of Experimental Psychology : Human Perception and Performance*, 12 : 3-17.

TSAO, D. Y., W. A. FREIWALD, R. B. TOOTELL et M. S. LIVINGSTONE, 2006, « A Cortical Region Consisting Entirely of Face-selective Cells », *Science*,

311 (5761) : 670-674.

Tshibanda, L., A. Vanhaudenhuyse, D. Galanaud, M. Boly, S. Laureys et L. Puybasset, 2009, « Magnetic Resonance Spectroscopy and Diffusion Tensor Imaging in Coma Survivors : Promises and Pitfalls », *Progress in Brain Research*, 177 : 215-229.

Tsodyks, M., T. Kenet, A. Grinvald et A. Arieli, 1999, « Linking Spontaneous Activity of Single Cortical Neurons and the Underlying Functional Architecture », *Science*, 286 (5446) : 1943-1946.

Tsubokawa, T., T. Yamamoto, Y. Katayama, T. Hirayama, S. Maejima et T. Moriya, 1990, « Deep-brain Stimulation in a Persistent Vegetative State : Follow-up Results and Criteria for Selection of Candidates », *Brain Injury*, 4 (4) : 315-327.

Tsuchiya, N. et C. Koch, 2005, « Continuous Flash Suppression Reduces Negative Afterimages », *Nature Neuroscience*, 8 (8) : 1096-1101.

Tsunoda, K., Y. Yamane, M. Nishizaki et M. Tanifuji, 2001, « Complex Objects Are Represented in Macaque Inferotemporal Cortex by the Combination of Feature Columns », *Nature Neuroscience*, 4 (8) : 832-838.

Tsushima, Y., Y. Sasaki et T. Watanabe, 2006, « Greater Disruption Due to Failure of Inhibitory Control on an Ambiguous Distractor », *Science*, 314 (5806) : 1786-1788.

Tsushima, Y. A. R. Seitz et T. Watanabe, 2008, « Task-irrelevant Learning Occurs Only When the Irrelevant Feature Is Weak », *Current Biology*, 18 (12) : R516-R517.

Turing, A. M., 1936, « On Computable Numbers, with an Application to the Entscheidungsproblem », *Proceedings of the London Mathematical Society*, 42 : 230-265.

–, 1952, « The Chemical Basis of Morphogenesis », *Philosophical Transactions of the Royal Society B : Biological Sciences*, 237 : 37-72.

Tyler, L. K. et W. Marslen-Wilson, 2008, « Fronto-temporal Brain Systems Supporting Spoken Language Comprehension », *Philosophical*

Transactions of the Royal Society B : Biological Sciences, 363 (1493) : 1037-1054.

Tzovara, A., A. O. Rossetti, L. Spierer, J. Grivel, M. M. Murray, M. Oddo et M. De Lucia, 2012, « Progression of Auditory Discrimination Based on Neural Decoding Predicts Awakening from Coma », *Brain* 136 (pt. 1) : 81-89.

Uhlhaas, P. J., D. E. Linden, W. Singer, C. Haenschel, M. Lindner, K. Maurer et E. Rodriguez, 2006, « Dysfunctional Long-range Coordination of Neural Activity During Gestalt Perception in Schizophrenia », *Journal of Neuroscience*, 26 (31) : 8168-8175.

Uhlhaas, P. J. et W. Singer, 2010, « Abnormal Neural Oscillations and Synchrony in Schizophrenia », *Nature Reviews Neuroscience*, 11 (2) : 100-113.

Van Aalderen-Smeets, S. I., R. Oostenveld et J. Schwarzbach, 2006, « Investigating Neurophysiological Correlates of Metacontrast Masking with Magnetoencephalography », *Advances in Cognitive Psychology*, 2 (1) : 21-35.

Van den Bussche, E., K. Notebaert et B. Reynvoet, 2009, « Masked Primes Can Be Genuinely Semantically Processed », *Journal of Experimental Psychology*, 56 (5) : 295-300.

Van den Bussche, E. et B. Reynvoet, 2007, « Masked Priming Effects in Semantic Categorization Are Independent of Category Size », *Journal of Experimental Psychology*, 54 (3) : 225-235.

Van Gaal, S., L. Naccache, J. D. I. Meeuwese, A. M. Van Loon, L. Cohen et S. Dehaene, 2013, « Can the Meaning of Multiple Words be Integrated Unconsciously ? », *Phil. Trans. R. Soc. B.*, 369 (1641).

Van Gaal, S., K. R. Ridderinkhof, J. J. Fahrenfort, H. S. Scholte et V. A. Lamme, 2008, « Frontal Cortex Mediates Unconsciously Triggered Inhibitory Control », *Journal of Neuroscience*, 28 (32) : 8053-8062.

Van Gaal, S., K. R. Ridderinkhof, H. S. Scholte et V. A. Lamme, 2010, « Unconscious Activation of the Prefrontal No-go Network », *Journal of*

Neuroscience, 30 (11) : 4143-4150.

VAN OPSTAL, F., F. P. DE LANGE et S. DEHAENE, 2011, « Rapid Parallel Semantic Processing of Numbers Without Awareness », *Cognition*, 120 (1) : 136-147.

VARELA, F., J.-P. LACHAUX, E. RODRIGUEZ et J. MARTINERIE, 2001, « The Brainweb : Phase Synchronization and Large-Scale Integration », *Nature Reviews Neuroscience*, 2 (4) : 229-239.

VELMANS, M., 1991, « Is Human Information Processing Conscious ? », *Behavioral and Brain Sciences*, 14 : 651-726.

VERNES, S. C., P. L. OLIVER, E. SPITERI, H. E. LOCKSTONE, R. PULIYADI, J. M. TAYLOR et J. HO *et al.*, 2011, « Foxp2 Regulates Gene Networks Implicated in Neurite Outgrowth in the Developing Brain », *PLOS Genetics*, 7 (7) : e1002145.

VINCENT, J. L., G. H. PATEL, M. D. FOX, A. Z. SNYDER, J. T. BAKER, D. C. VAN ESSEN, J. M. ZEMPEL *et al.*, 2007, « Intrinsic Functional Architecture in the Anaesthetized Monkey Brain », *Nature*, 447 (7140) : 83-86.

VOGEL, E. K., S. J. LUCK et K. L. SHAPIRO, 1998, « Electrophysiological Evidence for a Postperceptual Locus of Suppression During the Attentional Blink », *Journal of Experimental Psychology : Human Perception and Performance*, 24 (6) : 1656-1674.

VOGEL, E. K. et M. G. MACHIZAWA, 2004, « Neural Activity Predicts Individual Differences in Visual Working Memory Capacity », *Nature*, 428 (6984) : 748-751.

VOGEL, E. K., A. W. MCCOLLOUGH et M. G. MACHIZAWA, 2005, « Neural Measures Reveal Individual Differences in Controlling Access to Working Memory », *Nature*, 438 (7067) : 500-503.

VOGELEY, K., P. BUSSFELD, A. NEWEN, S. HERRMANN, F. HAPPE, P. FALKAI, W. MAIER *et al.*, 2001, « Mind Reading : Neural Mechanisms of Theory of Mind and Self-perspective », *NeuroImage*, 14 (1 pt. 1) : 170-181.

VOSS, H. U., A. M. ULUC, J. P. DYKE, R. WATTS, E. J. KOBYLARZ, B. D. MCCANDLISS, L. A. HEIER *et al.*, 2006, « Possible Axonal Regrowth in

Late Recovery from the Minimally Conscious State », *Journal of Clinical Investigation*, 116 (7) : 2005-2011.

VUILLEUMIER, P., N. SAGIV, E. HAZELTINE, R. A. POLDRACK, D. SWICK, R. D. RAFAL et J. D. GABRIELI, 2001, « Neural Fate of Seen and Unseen Faces in Visuospatial Neglect : A Combined Event-related Functional MRI and Event-related Potential Study », *Proceedings of the National Academy of Sciences*, 98 (6) : 3495-500.

VUL, E., D. HANUS et N. KANWISHER, 2009, « Attention as Inference : Selection Is Probabilistic ; Responses Are All-or-none Samples », *Journal of Experimental Psychology : General*, 138 (4) : 546-560.

VUL, E., M. NIEUWENSTEIN et N. KANWISHER, 2008, « Temporal Selection Is Suppressed, Delayed et Diffused During the Attentional Blink », *Psychological Science*, 19 (1) : 55-61.

VUL, E. et H. PASHLER, 2008, « Measuring the Crowd Within : Probabilistic Representations Within Individuals », *Psychological Science* (Wiley-Blackwell), 19 (7) : 645-647.

WACONGNE, C., J.-P. CHANGEUX et S. DEHAENE, 2012, « A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity », *Journal of Neuroscience*, 32 (11) : 3665-3678.

WACONGNE, C., E. LABYT, V. VAN WASSENHOVE, T. BEKINSCHTEIN, L. NACCACHE et S. DEHAENE, 2011, « Evidence for a Hierarchy of Predictions and Prediction Errors in Human Cortex », *Proceedings of the National Academy of Sciences*, 108 (51) : 20754-20759.

WAGNER, U., S. GAIS, H. HAIDER, R. VERLEGER et J. BORN, 2004, « Sleep Inspires Insight », *Nature*, 427 (6972) : 352-355.

WATSON, J. B., 1913, « Psychology as the Behaviorist Views It », *Psychological Review*, 20 : 158-177.

WEGNER, D. M., 2003, *The Illusion of Conscious Will*, Cambridge, Mass., MIT Press.

WEINBERGER, J., 2000, « William James and the Unconscious : Redressing a Century-old Misunderstanding », *Psychological Science*, 11 (6) : 439-

445.

WEISKRANTZ, L., 1986, *Blindsight : A Case Study and Its Implications*, Oxford, Clarendon Press.

–, 1997, *Consciousness Lost and Found : A Neuropsychological Exploration*, New York : Oxford University Press.

WEISS, Y., E. P. SIMONCELLI et E. H. ADELSON, 2002, « Motion Illusions as Optimal Percepts », *Nature Neuroscience*, 5 (6) : 598-604.

WESTMORELAND, B. F., D. W. KLASS, F. W. SHARBROUGH et T. J. REAGAN, 1975, « Alpha-coma. Electroencephalographic, Clinical, Pathologic, and Etiologic Correlations », *Archives of Neurology*, 32 (11) : 713-718.

WHITTINGSTALL, K. et N. K. LOGOTHETIS, 2009, « Frequency-band Coupling in Surface EEG Reflects Spiking Activity in Monkey Visual Cortex », *Neuron*, 64 (2) : 281-289.

WIDAMAN, K. F., D. C. GEARY, P. CORMIER et T. D. LITTLE, 1989, « A Componential Model for Mental Addition », *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 15 : 898-919.

WILKE, M., N. K. LOGOTHETIS et D. A. LEOPOLD, 2003, « Generalized Flash Suppression of Salient Visual Targets », *Neuron*, 39 (6) : 1043-1052.

–, 2006, « Local Field Potential Reflects Perceptual Suppression in Monkey Visual Cortex », *Proceedings of the National Academy of Sciences*, 103 (46) : 17507-17512.

WILLIAMS, M. A., C. I. BAKER, H. P. OP DE BEECK, W. M. SHIM, S. DANG, C. TRIANTAFYLLOU et N. KANWISHER, 2008, « Feedback of Visual Object Information to Foveal Retinotopic Cortex », *Nature Neuroscience*, 11 (12) : 1439-1445.

WILLIAMS, M. A., T. A. VISSER, R. CUNNINGTON et J. B. MATTINGLEY, 2008, « Attenuation of Neural Responses in Primary Visual Cortex During the Attentional Blink », *Journal of Neuroscience*, 28 (39) : 9890-9894.

WOMELSDORF, T., J. M. SCHOFFELEN, R. OOSTENVELD, W. SINGER, R. DESIMONE, A. K. ENGEL et P. FRIES, 2007, « Modulation of Neuronal Interactions

Through Neuronal Synchronization », *Science*, 316 (5831) : 1609-1612.

WONG, K. F., 2002, « The Relationship Between Attentional Blink and Psychological Refractory Period », *Journal of Experimental Psychology : Human Perception and Performance*, 28 (1) : 54-71.

WONG, K. F. et X. J. WANG, 2006, « A Recurrent Network Mechanism of Time Integration in Perceptual Decisions », *Journal of Neuroscience*, 26 (4) : 1314-1328.

WOODMAN, G. F. et S. J. LUCK, 2003, « Dissociations Among Attention, Perception, and Awareness During Object-Substitution Masking », *Psychological Science*, 14 (6) : 605-611.

WYART, V., S. DEHAENE et C. TALLON-BAUDRY, 2012, « Early Dissociation Between Neural Signatures of Endogenous Spatial Attention and Perceptual Awareness during Visual Masking », *Frontiers of Human Neuroscience*, 6.

WYART, V. et C. TALLON-BAUDRY, 2008, « Neural Dissociation Between Visual Awareness and Spatial Attention », *Journal of Neuroscience*, 28 (10) : 2667-2679.

–, 2009, « How Ongoing Fluctuations in Human Visual Cortex Predict Perceptual Awareness : Baseline Shift Versus Decision Bias », *Journal of Neuroscience*, 29 (27) : 8715-8725.

WYLER, A. R., G. A. OJEMANN et A. A. WARD Jr., 1982, « Neurons in Human Epileptic Cortex : Correlation Between Unit and EEG Activity », *Annals of Neurology*, 11 (3) : 301-308.

YANG, T. et M. N. SHADLEN, 2007, « Probabilistic Reasoning by Neurons », *Nature*, 447 (7148) : 1075-1080.

YOKOYAMA, O., N. MIURA, J. WATANABE, A. TAKEMOTO, S. UCHIDA, M. SUGIURA, K. HORIE *et al.*, 2010, « Right Frontopolar Cortex Activity Correlates with Reliability of Retrospective Rating of Confidence in Short-term Recognition Memory Performance », *Neuroscience Research*, 68 (3) : 199-206.

ZEKI, S., 2003, « The Disunity of Consciousness », *Trends in Cognitive Sciences*, 7 (5) : 214-218.

ZHANG, P., K. JAMISON, S. ENGEL, B. HE et S. HE, 2011, « Binocular Rivalry Requires Visual Attention », *Neuron*, 71 (2) : 362-369.

ZYLBERBERG, A., S. DEHAENE, G. B. MINDLIN et M. SIGMAN, 2009, « Neurophysiological Bases of Exponential Sensory Decay and Top-down Memory Retrieval : A Model », *Frontiers in Computational Neuroscience*, 3 : 4.

ZYLBERBERG, A., S. DEHAENE, P. R. ROELFSEMA et M. SIGMAN, 2011, « The Human Turing Machine : A Neural Framework for Mental Program », *Trends in Cognitive Sciences*, 15 (7) : 293-300.

ZYLBERBERG, A., D. FERNANDEZ Slezak, P. R. ROELFSEMA, S. DEHAENE et M. SIGMAN, 2010, « The Brain's Router : A Cortical Network Model of Serial Processing in the Primate Brain », *PLOS Computational Biology*, 6 (4) : e1000765.

Remerciements

Mes idées sur la nature de la conscience ne sont pas nées dans la solitude. Depuis trente ans, j'ai eu la chance d'être entouré de collègues extraordinaires, souvent devenus d'excellents amis, qui m'ont abreuvé d'un foisonnement permanent d'idées. J'ai contracté une dette particulière envers deux d'entre eux : mon maître Jean-Pierre Changeux qui, dès les années 1990, m'avait soufflé que le problème de la conscience n'était peut-être pas si inaccessible et que nous pouvions l'attaquer simultanément par la théorie et par l'expérience ; et mon ami Laurent Cohen qui a su m'orienter dans la littérature neuropsychologique. C'est Laurent qui m'a présenté un jeune étudiant en médecine, Lionel Naccache, aujourd'hui devenu brillant professeur de neurologie et chercheur en neurosciences cognitives, avec qui nous avons commencé à sonder les profondeurs des opérations subliminales. Nos collaborations et nos discussions n'ont jamais cessé. Jean-Pierre, Laurent, Lionel, merci pour votre amitié, vos conseils et vos constants encouragements.

Dans les vingt dernières années, Paris est devenu un haut lieu de la recherche sur la conscience. Mon laboratoire a grandement bénéficié de cet environnement stimulant, et particulièrement de discussions passionnantes avec Patrick Cavanagh, Sid Kouider, Jérôme Sackur, Étienne Koechlin, Kevin O'Regan et Mathias Pessiglione. De nombreux étudiants et post-doctorants, souvent financés par la Fondation Fyssen ou par l'excellent master de sciences cognitives de l'École normale supérieure,

ont insufflé à mon équipe leur énergie et leur créativité. Toute ma gratitude va à mes thésards Lucie Charles, Antoine Del Cul, Raphael Gaillard, Jean-Rémi King, Claire Sergent, Mélanie Strauss, Lynn Uhrig, Catherine Wacongne et Valentin Wyart, ainsi qu'à mes collègues post-doctorants Tristan Bekinschtein, Floris de Lange, Sébastien Marti, Kimihiro Nakamura, Moti Salti, Aaron Schurger, Jacobo Sitt, Simon Van Gaal et Filip Van Opstal. Un merci particulier à Mariano Sigman, pour dix années de collaboration fructueuse, de partage sans arrière-pensée… et d'amitié, tout simplement.

Une grande variété de disciplines, de laboratoires et de chercheurs contribuent au dynamisme de la recherche sur la conscience. J'ai ainsi bénéficié de nombreuses conversations avec Bernard Baars (à qui nous devons la première version de la théorie de l'espace de travail global), Moshe Bar, Edoardo Bisiach, Olaf Blanke, Ned Block, Antonio Damasio, Daniel Dennett, Derek Denton, Gerry Edelman, Pascal Fries, Karl Friston, Chris Frith, Uta Frith, Mel Goodale, Tony Greenwald, John-Dylan Haynes, Biyu Jade He, Nancy Kanwisher, Markus Kiefer, Christof Koch, Victor Lamme, Dominique Lamy, Hakwan Lau, Steven Laureys, Nikos Logothetis, Lucia Melloni, Earl Miller, Adrian Owen, Josef Parvizi, Dan Pollen, Michael Posner, Alex Pouget, Marcus Raichle, Geraint Rees, Pieter Roelfsema, Niko Schiff, Mike Shadlen, Tim Shallice, Kimron Shapiro, Wolf Singer, Elizabeth Spelke, Giulio Tononi, Wim Vanduffel, Larry Weiskrantz, Mark Williams, et d'autres encore.

Mes recherches ont reçu le généreux soutien financier de l'Institut national de la santé et de la recherche médicale (Inserm), du Commissariat à l'énergie atomique et aux énergies alternatives (CEA), du Collège de France, de l'université Paris-Sud, et du Conseil européen de la recherche. Le centre NeuroSpin, situé au sud de Paris et dirigé par Denis Le Bihan, m'a fourni un environnement stimulant, au sein duquel j'ai pu développer ces idées en toute liberté. Toute ma gratitude va aux collègues de ce site qui m'ont soutenu et conseillé, notamment Gilles Bloch, Jean-Robert Deverre, Lucie Hertz-Pannier, Béchir Jarraya, Andreas

Kleinschmidt, Jean-François Mangin, Bertrand Thirion, Gaël Varoquaux et Virginie Van Wassenhove.

Faute de temps, j'ai écrit ce livre par fragments, souvent à l'occasion de courts séjours à l'étranger. M'ont ainsi accueilli l'Institut d'études avancées Peter-Wall de Vancouver, l'Université Macquarie de Sydney, l'Institut d'études avancées IUSS de Pavie, la Fondation des Treilles, l'Académie pontificale des sciences au Vatican, l'Université de Tel-Aviv et l'Institut Weizmann en Israël… et La Chouannière et La Trinitaine, deux retraites familiales où bon nombre de ces lignes ont été écrites et réécrites.

Comme à chaque livre, mon éditrice Odile Jacob et sa collaboratrice Émilie Barian m'ont consacré beaucoup de temps – qu'elles soient ici remerciées pour leurs commentaires et leurs suggestions. L'ensemble du texte a également bénéficié du regard acéré de Sid Kouider et de Lionel Naccache.

Enfin, tout au long de cette enquête scientifique, mon épouse Ghislaine Dehaene-Lambertz a partagé avec moi son savoir encyclopédique sur le cerveau en développement, mais surtout l'amour et la tendresse qui rendent la vie vivable et la conscience délectable.

Crédits des illustrations

Figure 1. © Ministère de la Culture, médiathèque du Patrimoine, dist. RMN-Grand Palais/image IGN.

Figure 4 (*en haut à droite*). Par l'auteur.

Figure 4 (*en bas*). Adaptée par l'auteur de D. A. Leopold et N. K. Logothetis, 1999, « Multistable Phenomena : Changing Views in Perception », *Trends in Cognitive Sciences*, 3 : 254-264. Copyright © 1999. Reproduit avec l'autorisation d'Elsevier.

Figure 5. par l'auteur.

Figure 6 (*en haut*). D. J. Simons et C. F. Chabris, 1999, « Gorillas in Our Midst : Sustained Inattentional Blindness for Dynamic Events », *Perception*, 28 : 1059-1074.

Figure 7 (*en haut et au milieu*). Adaptée par l'auteur de S. Kouider et S. Dehaene, 2007, « Levels of Processing During Non-Conscious Perception : A Critical Review of Visual Masking », *Philosophical Transactions of the Royal Society B : Biological Sciences*, 362 (1481) : 857-875 (Figure 1, p. 859).

Figure 7 (*en bas*). Par l'auteur.

Figure 9 (*en haut*). Courtoisie Melvyn Goodale.

Figure 10. Courtoisie Edward Adelson.

Figure 11. Adaptée par l'auteur et fondée sur S. Dehaene *et al.*, 1998, « Imaging Unconscious Semantic Priming », *Nature*, 395 : 597-600.

Figure 12. Adaptée par l'auteur de M. Pessiglione *et al.*, 2007, « How the Brain Translates Money into Force : A Neuroimaging Study of Subliminal Motivation », *Science*, 316 (5826) : 904-906. Courtoisie Mathias Pessiglione.

Figure 13. Par l'auteur.

Figure 14. Par l'auteur.

Figure 15. Adaptée par l'auteur de R. Moreno-Bote, D. C. Knill et A. Pouget, 2011, « Bayesian Sampling in Visual Perception », *Proceedings of the National Academy of Sciences of the United States of America*, 108 (30) : 12491-12496. Figure 1A.

Figure 16 (*top*). Adaptée par l'auteur de S. Dehaene *et al.*, 2001, « Cerebral Mechanisms of Word Masking and Unconscious Repetition Priming », *Nature Neuroscience*, 4 (7) : 752-758 (Figure 2).

Figure 16 (*en bas*). Adaptée par l'auteur de S. Sadaghiani *et al.*, 2009, « Distributed and Antagonistic Contributions of Ongoing Activity Fluctuations to Auditory Stimulus Detection », *Journal of Neuroscience*, 29 (42) : 13410-13417. Courtoisie Sepideh Sadaghiani.

Figure 17. Adaptée par l'auteur de S. Van Gaal *et al.*, 2010, « Unconscious Activation of the Prefrontal No-Go Network », *Journal of Neuroscience*, 30 (11) : 4143-4150. (Figures 3 et 4). Courtoisie Simon Van Gaal.

Figure 18. Adaptée par l'auteur de C. Sergent *et al.*, 2005, « Timing of the Brain Events Underlying Access to Consciousness During the Attentional Blink », *Nature Neuroscience*, 8 (10) : 1391-1400.

Figure 19. Adaptée par l'auteur de A. Del Cul *et al.*, 2007, « Brain Dynamics Underlying the Nonlinear Threshold for Access to Consciousness », *PLOS Biology*, 5 (10) : e260.

Figure 20. Adaptée par l'auteur de L. Fisch, E. Privman, M. Ramot, M. Harel, Y. Nir, S. Kipervasser *et al.*, 2009, « Neural "Ignition" : Enhanced Activation Linked to Perceptual Awareness in Human Ventral Stream Visual Cortex », *Neuron*, 64 : 562-74. Reproduit avec l'autorisation d'Elsevier.

Figure 21 (*en haut*). Adaptée par l'auteur de E. Rodriguez *et al.*, 1999, « Perception's Shadow : Long-Distance Synchronization of Human Brain Activity », *Nature*, 397 (6718) : 430-433 (Figures 1 et 3).

Figure 21 (*en bas*). Adaptée par l'auteur de R. Gaillard *et al.*, 2009, « Converging Intracranial Markers of Conscious Access », *PLOS Biology*, 7 (3) : e61 (Figure 8).

Figure 22. Adaptée par l'auteur de R. Q. Quiroga, R. Mukamel, E. A. Isham, R. Malach et I. Fried, 2008, « Human Single-Neuron Responses at the Threshold of Conscious Recognition », *Proceedings of the National Academy of Sciences of the United States of America*, 105 (9) : 3599-3604 (Figure 2). © 2008 National Academy of Sciences, USA.

Figure 23 (*à droite*). © 2003 Neuroscience of Attention & Perception Laboratory, Princeton University.

Figure 24 (*en haut*) : B. J. Baars, 1989, *A Cognitive Theory of Consciousness*, Cambridge (R.-U.), Cambridge University Press. Courtoisie Bernard Baars.

Figure 24 (*en bas*). S. Dehaene, M. Kerszberg et J.-P. Changeux, 1998, « A Neuronal Model of a Global Workspace in Effortful Cognitive Tasks », *Proceedings of the National Academy of Sciences of the United States of America*, 95 (24) : 14529-14534 (Figure 1). © 1998 National Academy of Sciences, USA.

Figure 25 (*à droite*), Courtoisie Michel Thiebaut de Schotten.

Figure 26 (*en bas*). G. N. Elston, 2003, « Cortex, Cognition and the Cell : New Insights into the Pyramidal Neuron and Prefrontal Function », *Cerebral Cortex*, 13 (11) : 1124-1138. Reproduit avec l'autorisation des Oxford University Press.

Figure 27. Adaptée par l'auteur de S. Dehaene *et al.*, 2005, « Ongoing Spontaneous Activity Controls Access to Consciousness : A Neuronal Model for Inattentional Blindness », *PLOS Biology*, 3 (5) : e141.

Figure 28. Adaptée par l'auteur de S. Dehaene *et al.*, 2006, « Conscious, Preconscious, and Subliminal Processing : A Testable Taxonomy », *Trends in Cognitive Sciences*, 10 (5) : 204-211.

Figure 29. Adaptée par l'auteur de S. Laureys *et al.*, 2004, « Brain Function in Coma, Vegetative State, and Related Disorders », *Lancet Neurology*, 3 (9) : 537-546.

Figure 30. Adaptée par l'auteur de M. M. Monti, A. Vanhaudenhuyse, M. R. Coleman, M. Boly, J. D. Pickard, L. Tshibanda *et al.*, 2010, « Willful Modulation of Brain Activity in Disorders of Consciousness », *New England Journal of Medicine*, 362 : 579-589. © 2010 Massachusetts Medical Society. Reproduit avec l'autorisation de la Massachusetts Medical Society.

Figure 31. Adaptée par l'auteur de T. A. Bekinschtein, S. Dehaene, B. Rohaut, F. Tadel, L. Cohen et L. Naccache, 2009, « Neural Signature of the Conscious Processing of Auditory Regularities », *Proceedings of the National Academy of Sciences of the United States of America*, 106 (5), 1672-1677 (Figures 2 et 3).

Figure 32. Courtoisie Steven Laureys.

Figure 33. Adaptée par l'auteur de J. R. King, J. D. Sitt *et al.*, 2013, « Information Sharing in the Brain Indexes Consciousness in Noncommunicative Patients », *Current Biology*, 23 : 1914-1919, © 2013. Reproduit avec l'autorisation d'Elsevier.

Figure 34. Adaptée par l'auteur de G. Dehaene-Lambertz, S. Dehaene et L. Hertz-Pannier, 2002, « Functional Neuroimaging of Speech Perception in Infants », *Science*, 298 (5600) : 2013-2015.

Figure 35. Adaptée par l'auteur de S. Kouider *et al.*, 2013, « A Neural Marker of Perceptual Consciousness in Infants », *Science*, 340 (6130) : 376-380.

DU MÊME AUTEUR
CHEZ ODILE JACOB

Apprendre à lire. Des sciences cognitives à la salle de classe (avec G. Dehaene-Lambertz, C. Huron et L. Sprenger-Charolles), 2011.
La Bosse des maths quinze ans après, 2010.
Parole et musique. Aux origines du dialogue humain (dir. avec C. Petit), 2009.
Les Neurones de la lecture, 2007.
La Bosse des maths, 1996.

TABLE

Couverture

Titre

Copyright

Dédicace

Introduction – La matière de la pensée

Le défi de Descartes

Le dernier problème

Déchiffrer le code conscient

Voir ou ne pas voir

Faire du subjectif une science

Les signatures d'une pensée consciente

L'avenir de la conscience

CHAPITRE 1 - Quand la conscience entre au laboratoire

Les facettes de la conscience

La méthode contrastive

Images rivales

Lorsque l'attention cligne

La perception avance masquée

La primauté du subjectif

CHAPITRE 2 - Sonder la profondeur de l'inconscient

Les pionniers de l'inconscient

Le siège des opérations inconscientes

Le côté obscur du cerveau

Tisser des liens inconscients

Jouer aux échecs inconsciemment

La vision d'une voix

Sens sans conscience ?

La grande bataille de la sémantique inconsciente

Arithmétique inconsciente

Une combinaison inconsciente de concepts

Attentif, mais inconscient

La valeur d'une pièce de monnaie inconsciente

Mathématiques inconscientes

Statistiques nocturnes

La boîte à outils de l'inconscient

CHAPITRE 3 - À quoi sert la conscience ?

Statistiques inconscientes, échantillonnage conscient

Pensées durables

La machine de Turing humaine

Le premier réseau social

CHAPITRE 4 - Les signatures de la pensée consciente

L'avalanche consciente

Chronométrer l'accès à la conscience

La conscience en retard sur le monde

Un pur moment de conscience

L'embrasement de la conscience

Au plus profond du cerveau conscient

Le réseau neuronal global

Le point de non-retour et ses précurseurs

Le décodage d'une pensée consciente

Induire une hallucination

Détruire la conscience

Une chose qui pense

CHAPITRE 5 - Théoriser la conscience

La conscience est une diffusion globale

Au-delà de la modularité

Un réseau de communication évolué

Sculpter une pensée consciente

La forme d'une idée

Simuler l'ignition consciente

Pas de repos pour le cerveau

Darwin dans le cerveau

Un catalogue de l'inconscient

Les états subjectifs de la matière

CHAPITRE 6 - L'épreuve de vérité

Perdre conscience

Cortico ergo sum

Libérer le papillon de la conscience

La détection consciente de la nouveauté

Sonder le cortex

Détecter l'existence d'une pensée autonome

Vers des interventions cliniques

CHAPITRE 7 - L'avenir de la conscience

Un bébé est-il conscient ?

Une conscience chez l'animal ?

Une conscience de soi chez le singe ?

Le propre de l'homme ?

Des maladies de la conscience ?

Des machines conscientes ?

Notes

Bibliographie

Remerciements

Crédits des illustrations

Du même auteur chez Odile Jacob

Éditions Odile Jacob

Des idées qui font avancer les idées

Retrouvez tous nos livres disponibles en numérique
sur odilejacob.fr
Retrouvez-nous sur Facebook
Suivez-nous sur Twitter

Stanislas Dehaene

Le Code de la conscience

D'où viennent nos perceptions, nos sentiments, nos illusions et nos rêves ? Où s'arrête le traitement mécanique de l'information et où commence la prise de conscience ? L'esprit humain est-il suffisamment ingénieux pour comprendre sa propre existence ? La prochaine étape sera-t-elle une machine consciente de ses propres limites ?

Depuis plus de vingt ans, Stanislas Dehaene analyse les mécanismes de la pensée humaine. Dans ce livre, il invite le lecteur dans son laboratoire où d'ingénieuses expériences visualisent l'inconscient et démontent les bases biologiques de la conscience. Grâce à l'imagerie cérébrale et même à des électrodes introduites dans la profondeur du cortex, nous commençons enfin à comprendre les algorithmes qui nous font penser.

Détecter la présence de la conscience, décoder à quoi pense un individu, un bébé ou même un animal, sortir les patients du coma, doter les machines d'un début de conscience… *Le Code de la conscience* ouvre d'extraordinaires perspectives pratiques et intellectuelles, en accordant une importance égale aux implications technologiques, philosophiques, personnelles et éthiques de la résolution du dernier des mystères.

Stanislas Dehaene est professeur au Collège de France, titulaire de la chaire de psychologie cognitive expérimentale et membre de l'Académie des sciences. Il a publié *Les Neurones de la lecture* et *La Bosse des maths*, qui ont rencontré un très grand succès.

En couverture : © Science Photo Library/Corbis.