

des clés pourrait remplacer les services de délivrance des passeports. Avant d'attribuer le premier identificateur, le service des clés élabore ses propres clés de signature : un nombre secret et la version publique, obtenue à l'aide d'une exponentielle modulaire. Qui-conque désire vérifier les identificateurs attribués par le service des clés a juste besoin d'une copie de la clé publique.

Le service des clés fournit un certificat (une puce électronique résistant aux altérations et placée dans une carte de crédit ou dans une montre spéciale) à chaque personne, terminal ou ordinateur de son réseau. Ce certificat peut être utilisé dans toutes les transactions. Lorsque Alice demande un numéro d'identification au service des clés, elle obtient un certificat dont la mémoire contient le message suivant : «Le propriétaire de cette carte est Alice Wunderkind, professeur d'informatique à l'Université de Paris», suivi d'une signature du type ElGamal fondée sur ce message.

Ultérieurement, afin de convaincre Bob que son numéro d'identification est valide, Alice utilise son certificat pour lui envoyer un message ainsi que le témoin du jeton utilisé dans la signature. En utilisant ces valeurs et la clé publique du service des clés, Bob calcule la partie de la signature que le certificat n'a pas révélé. Pour s'assurer que l'identificateur est valide, Bob dialogue avec le certificat en lui posant des questions dont les réponses sont justes seulement s'il connaît la partie secrète de la signature.

Mes collègues et moi-même avons conçu un protocole qui intègre l'authentification et la vérification avec l'échange des clés. Ce protocole est exécuté en une fraction de seconde par un microprocesseur placé dans une carte à puce. Si leurs numéros d'identification sont bons, les deux partenaires terminent leur premier échange par une clé commune qui servira dans les transactions futures. Si l'authentification échoue, chacun reçoit une séquence différente, sans utilité aucune.

Chaîne de confidentialité

Cette méthode d'identification repose sur la confiance que Bob accorde au service de délivrance des clés d'Alice. Pour vérifier le numéro d'identification, on devra connaître la signature de tous les services des clés. Dans le cas des transactions internationales,

cette méthode n'est utilisable que si les services des clés sont hiérarchisés : chacun d'entre eux devrait se légitimer en se procurant un identificateur d'un service des clés supérieur.

Une station de distribution centrale devrait garantir la confiance que l'on accorde à chacun des services des clés universitaires, qui, à leur tour, garantiraient les services des clés internes. Ce schéma est extensible à un nombre quelconque de niveaux. Une station centrale française pourrait se légitimer en utilisant l'identificateur numérique d'un centre européen qui, à son tour, serait validé par une organisation mondiale. La carte à puce détenue par une personne pourrait contenir une cascade de messages : une identification individuelle signée par le service des clés local, une identification de l'institution signée par l'autorité régionale, et ainsi de suite. La vérification d'un identificateur ne nécessite de connaître que la signature numérique d'un petit nombre d'autorités et de leur faire confiance. Ainsi pourrait-on échanger des informations à distance avec autant de confiance que si le correspondant était présent.

Thomas BETH est professeur d'informatique à l'Université de Karlsruhe et directeur de l'Institut européen des systèmes de sécurité.

Taher ELGAMAL, *A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms*, in *IEEE Transactions on Information Theory*, vol. 31, n° 4, pp. 469-472, juillet 1985.

T. BETH, *Efficient Zero-Knowledge Identification Scheme for Smart Cards*, in *Advances in Cryptology : EUROCRYPT '88*, sous la direction de Christoph Günther, Springer-Verlag, 1988.

Public-Key Cryptography : State of the Art and Future Directions, sous la direction de T. Beth, M. Frisch et G.J. Simmons, Springer-Verlag, 1992.

Gustavus J. SIMMONS, *Contemporary Cryptology : The Science of Information Integrity*, IEEE Press, 1992.

T. BETH, *Keeping Secrets a Personal Matter or the Exponential Security System*, in *Cryptography and Coding III*, sous la direction de M.J. Ganley, Oxford University Press, 1993.

H. DANISCH, *RFC 1824 : The Exponential Security System TESS* in European Institute for System Security World Wide Web site at [http : // avalon.ira. uka.de / eiss / indexe.html](http://avalon.ira.uka.de/eiss/indexe.html).

Qu'est-ce que la conscience ?

DAVID CHALMERS

Les neurobiologistes et les philosophes explorent l'un des plus grands mystères de l'existence. La connaissance des mécanismes physiques du cerveau semble insuffisante pour comprendre la nature de la conscience.

La conscience est la chose à la fois la plus familière et la plus mystérieuse. Nous la connaissons plus directement que quoi que ce soit d'autre, mais cette connaissance est difficile à concilier avec nos autres savoirs. Pourquoi existe-t-elle ? Que fait-elle ? Comment résulte-t-elle de mécanismes neuronaux ?

D'un point de vue objectif, nous comprenons assez bien le fonctionnement du cerveau : lorsque nous regardons cette page, des photons frappent nos rétines, des signaux électriques sont transmis par les nerfs optiques à différentes régions de notre cerveau et, finalement, nous réagissons par un sourire, un froncement de sourcil ou une remarque. L'aspect subjectif est plus complexe : quand nous regardons la page, nous en sommes conscients, et nous ressentons directement les images et les mots comme une partie de notre vie mentale intime. Les couleurs des fleurs ou du ciel nous émeuvent, tandis que nous ressentons peut-être une émotion et que nous formulons des pensées. Ensemble, ces expériences constituent la conscience, la vie interne, subjective, de l'esprit.

Les chercheurs qui étudient le cerveau et la pensée ont longtemps laissé la conscience de côté : la science, qui repose sur l'objectivité, ne pouvait traiter de quelque chose d'aussi subjectif que la conscience. Le courant béhavioriste, qui a dominé la psychologie au début du siècle, se concentrait sur le comportement externe et interdisait de parler des processus mentaux internes. Plus tard, l'émergence des sciences cognitives a focalisé l'attention sur les mécanismes cérébraux, mais l'étude de la conscience restait exclue ; à peine évoquait-on le problème après les dîners bien arrosés.

Depuis quelques années, un nombre croissant de neurobiologistes, de psychologues et de philosophes a décidé de violer ces interdits et d'explorer les secrets de la conscience. Ils ont proposé de nombreuses théories, variées et parfois opposées, souvent fondées sur des acceptions différentes des concepts fondamentaux. Seul le raisonnement philosophique permettra de sortir de cette confusion.

Aux deux extrêmes, on trouve la position réductionniste, selon laquelle la conscience est explicable par les méthodes des neurosciences et de la psychologie, et la position « mystérieuse », selon laquelle on ne comprendra jamais la conscience. L'analyse montre que ces deux positions sont fausses et que la vérité est probablement à mi-chemin.

Contrairement aux réductionnistes, je pense que les outils des neurosciences, aussi performants soient-ils, ne peuvent pas rendre pleinement compte de la conscience. Contre la position « mystérieuse », je soutiens que la conscience est explicable par un nouveau type de théorie ; nous ne pouvons pas encore élaborer une telle théorie en détail, mais nous pouvons l'esquisser. Cette théorie contiendra de nouvelles lois fondamentales, et le concept d'information y jouera un rôle central. Elle aura des conséquences surprenantes sur notre façon de voir l'Univers, et de nous voir nous-mêmes.

Le mot « conscience » a de nombreuses acceptions. Quels problèmes

1. UNE NEUROBIOLOGISTE qui connaîtrait tout sur la manière dont le cerveau traite les couleurs mais vivrait dans une pièce en noir et blanc, ne saurait pas ce que c'est que de voir les couleurs. La connaissance du cerveau n'apporte pas la connaissance complète de la conscience.



sont regroupés derrière ce mot? Distinguons d'abord des «problèmes faciles» et un «problème difficile». Les problèmes faciles sont aussi ardues que la plupart de ceux que traitent la psychologie et la biologie, mais le cœur du mystère est la solution du problème difficile.

Problèmes faciles et problème difficile

Les problèmes faciles sont, par exemple, les suivants : comment un sujet humain perçoit-il des stimuli sen-

soriels et comment y réagit-il de manière appropriée? Comment le cerveau rassemble-t-il des informations provenant de différentes sources et comment utilise-t-il ces informations pour commander le comportement? Pourquoi pouvons-nous verbaliser nos états internes? Toutes ces questions sont faciles, parce qu'elles concernent le fonctionnement objectif du système cognitif. La poursuite des études de psychologie cognitive et de neurosciences devrait les résoudre.

Le problème difficile consiste à comprendre comment les processus phy-

siques cérébraux engendrent la conscience subjective. Ce problème concerne les aspects profonds de la pensée et de la perception ; il se rapporte aux sentiments que les choses suscitent en nous. La sensation d'un bleu vif, le son ineffable d'un hautbois lointain, une douleur intense, une étincelle de bonheur ou la qualité méditative d'un instant de réflexion, tout cela fait partie de ce que je nomme la conscience. Ces phénomènes constituent le véritable mystère de l'esprit.

Une expérience de pensée, due au philosophe australien Franck Jackson,



illustre la distinction entre la connaissance objective et la conscience : Marie, qui vit au XXIII^e siècle, est la meilleure spécialiste mondiale des mécanismes cérébraux responsables de la vision des couleurs, mais elle a passé sa vie dans une pièce noire et blanche, sans jamais voir de couleur. Elle sait tout ce qu'on peut savoir sur la biologie, la structure et le fonctionnement du cerveau. Cette connaissance lui permet de résoudre les problèmes faciles : comment le cerveau discrimine les stimuli, rassemble les informations et produit des comptes rendus verbaux. Marie connaît la correspondance entre les noms des couleurs et les longueurs d'onde de la

lumière, mais il reste quelque chose de crucial que Marie ne connaît pas : le sentiment que l'on a quand on regarde une couleur comme le rouge. Des faits relatifs à la conscience ne peuvent pas être déduits du fonctionnement physique du cerveau.

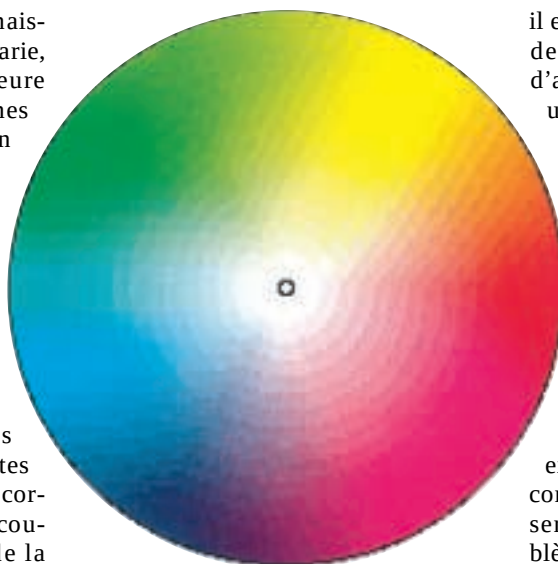
Personne ne sait pourquoi ces processus physiques sont accompagnés de la conscience. Pourquoi, lorsque notre cerveau traite la lumière d'une certaine longueur d'onde, éprouvons-nous une sensation de violet profond ? Pourquoi ressentons-nous des choses ? Un automate inconscient ferait-il, aussi bien que nous, tout ce que nous faisons ? Une théorie de la conscience devrait répondre à ces questions.

Je ne nie pas que la conscience soit produite par le cerveau. Nous savons que la sensation subjective de vision est étroitement associée à des opérations qui se déroulent dans le cortex visuel. C'est précisément cette association qui nous laisse perplexes : l'expérience subjective semble émerger d'un processus physique, mais nous ne savons pas comment ni pourquoi.

Vers la résolution des problèmes faciles

Le développement de la neurologie et de la psychologie est si rapide que l'on pourrait espérer que l'énigme est en passe d'être résolue, mais la plupart des travaux ne concernent que les problèmes faciles. La confiance des réductionnistes vient des progrès réalisés dans ces domaines, mais le problème difficile reste sans solution.

Selon Francis Crick, de l'Institut Salk, et Christof Koch, de l'Institut



2. DANS CE DISQUE DE COULEURS, les nuances sont disposées de telle sorte que celles que l'on ressent comme similaires sont voisines. Les couleurs proches correspondent aussi à des représentations perceptives similaires dans le cerveau.

de technologie de Californie, la conscience émergerait d'une synchronisation des décharges neuronales, dans le cerveau, à une fréquence de 40 hertz. Ce phénomène expliquerait comment la couleur et la forme d'un objet, qui sont traitées par des régions cérébrales différentes, sont unifiées en un tout cohérent. Deux éléments d'information seraient associés lorsqu'ils sont représentés par des séquences de décharges neuronales synchronisées.

Cette théorie résoudra peut-être l'un des problèmes faciles, mais elle n'aborde pas le problème difficile : pourquoi des oscillations synchronisées engendreraient-elles une sensation visuelle consciente, quelle que soit la manière dont les perceptions sont traitées et intégrées ? F. Crick et C. Koch ne prennent pas position sur la possibilité de la science de résoudre le problème difficile (voir l'encadré des pages 62 et 63).

Le même type de critique s'applique à presque tous les travaux récents sur la conscience. Dans son livre *La conscience expliquée*, le philosophe Daniel Dennett propose une théorie perfectionnée pour expliquer comment, dans le cerveau, de nombreux processus indépendants produisent ensemble une réaction cohérente à un événement perçu. Cette théorie explique peut-être comment nous explicitons verbalement nos états internes, mais elle ne nous dit pas grand-chose de la raison pour laquelle

il existerait une expérience subjective derrière cette verbalisation. Comme d'autres théories réductionnistes, c'est une théorie des problèmes faciles.

Les problèmes faciles ont un point commun : ils concernent l'accomplissement des fonctions cognitives ou comportementales. Il s'agit toujours de savoir comment le cerveau exécute une tâche donnée : comment il capte et analyse les stimuli, comment il assimile les informations, comment il produit des rapports. Lorsque la neurobiologie aura expliqué les mécanismes neuronaux correspondants, les problèmes faciles seront résolus. En revanche, le problème difficile ne le sera pas. Même si

toutes les fonctions comportementales et cognitives en rapport avec la conscience étaient expliquées, il resterait un mystère : pourquoi ces fonctions s'accompagnent-elles de conscience ? C'est cette énigme supplémentaire qui rend difficile le problème difficile.

De nouveaux outils d'explication physique, telles la dynamique non linéaire ou la mécanique quantique, ne permettent pas non plus de résoudre le problème difficile. Selon Stuart Hameroff, de l'Université d'Arizona, et Roger Penrose, de l'Université d'Oxford, la conscience résulterait de processus physiques quantiques qui se produiraient dans des organites des neurones nommés microtubules. Une telle hypothèse pourrait expliquer comment le cerveau prend des décisions ou comment il prouve des théorèmes mathématiques, mais elle ne dit rien sur la manière dont ces processus engendreraient la conscience. Toutes les théories qui ne reposent que sur une approche physique ont le même inconvénient.

Le handicap de ces théories physiques est constitutif : elles expliquent pourquoi les systèmes ont une certaine structure physique et comment ils accomplissent diverses fonctions. La plupart des problèmes examinés par la science ont cette forme : pour expliquer la vie, par exemple, on cherche à décrire comment un système physique peut se reproduire, s'adapter, métaboliser de l'énergie. La conscience pose un problème complètement différent, parce qu'elle n'est ni structure ni fonction.

Bien entendu, les neurosciences seront utiles, parce qu'elles pourraient

expliquer les mécanismes neuronaux associés à la conscience et qu'elles pourraient même détailler les correspondances entre des événements cérébraux et des éléments de conscience. Toutefois, nous ne franchirons pas ce que le philosophe Joseph Levine a nommé le fossé explicatif entre les processus physiques et la conscience tant que nous ne saurons pas pourquoi ces processus secrètent la conscience. Pour faire ce saut, nous devons utiliser un nouveau type de théorie.

Une vraie théorie de tout

De quelle théorie avons-nous besoin ? Observons tout d'abord qu'en physique tous les systèmes ne sont pas décomposables en éléments plus fondamentaux. L'espace-temps, la masse et la charge sont considérés comme des caractéristiques fondamentales du monde : ils ne sont pas réductibles à quelque chose de plus simple. Malgré cette irréductibilité, des théories détaillées et utiles relient ces éléments par des lois fondamentales, expliquant de nombreux phénomènes complexes.

L'idée que la physique contient tous les éléments fondamentaux et toutes

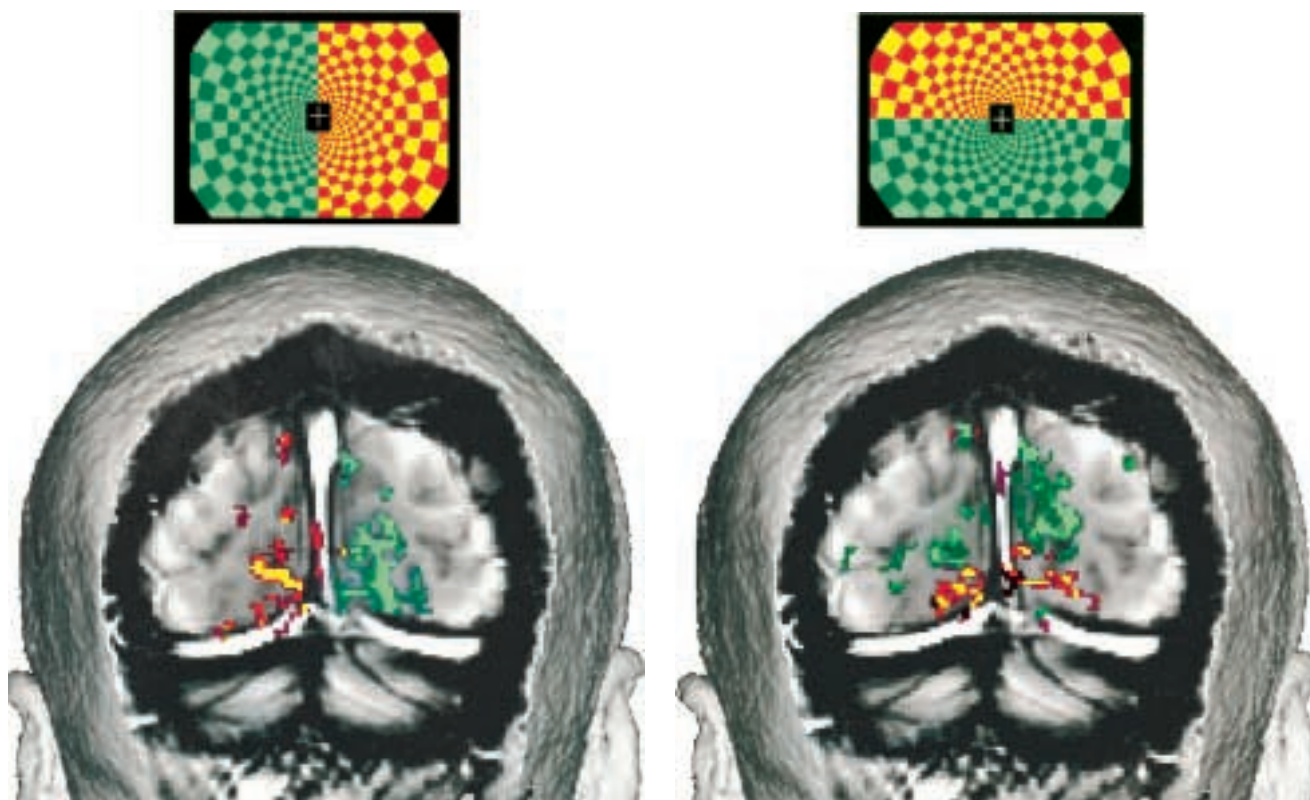
les lois fondamentales de l'Univers est largement répandue. Comme le physicien Steven Weinberg l'a écrit, le rêve du physicien est une «théorie de tout», à partir de laquelle on déduirait tout ce qu'il faut savoir sur l'Univers. S. Weinberg concède toutefois que la conscience pose un problème : malgré la puissance explicative de la théorie physique, on ne peut apparemment pas en déduire l'existence de la conscience. Même si la physique permettait d'expliquer les mécanismes cérébraux objectifs associés à la conscience, elle n'expliquerait pas la conscience elle-même, de sorte qu'aucune théorie physique ne sera une vraie théorie de tout : pour élaborer cette dernière, il faut introduire une composante supplémentaire.

Je propose donc que l'on considère la conscience comme une caractéristique fondamentale, irréductible à quoi que ce soit de plus élémentaire. L'idée peut paraître étrange, mais la cohérence l'exige et l'histoire nous donne quelques précédents : au XIX^e siècle, par exemple, quand les phénomènes électromagnétiques étaient inexplicables à partir des principes connus, les physiciens ont introduit une nouvelle entité fondamentale,

la charge électromagnétique, et ils en ont étudié les lois fondamentales. Ne peut-on reprendre cette méthodologie dans le cas de la conscience ? Si les théories actuelles ne l'expliquent pas, alors il faut introduire de nouveaux éléments.

Un nouvel élément fondamental requiert de nouvelles lois fondamentales : celles-ci relieront la conscience aux éléments de la théorie physique, sans interférer avec les lois physiques, car le monde physique forme un tout fermé. Les nouvelles lois serviront plutôt à franchir le fossé explicatif : elles indiqueront comment la conscience dépend des processus physiques sous-jacents. Une théorie complète aura deux composantes : des lois physiques, qui décriront le comportement des systèmes physiques, de l'infiniment petit à l'infiniment grand, et des lois «psychophysiques», qui diront comment certains systèmes physiques sont associés à une conscience.

Comment découvrir les lois psychophysiques ? L'obstacle majeur, dans cette recherche, est le manque de données. Telle que je l'ai décrite ici, la conscience est subjective, de sorte que nous ne pouvons pas analyser directement celle d'autrui. Cet obstacle



3. LES VARIATIONS DU DÉBIT SANGUIN dans le cortex visuel montrent comment le cerveau d'un sujet répond à une configuration qu'il voit. Les couleurs de cette illustration représentent les zones

cérébrales activées par la vision de chaque moitié du modèle. Cette expérience montre les réactions neuronales associées à la sensation visuelle, qui seraient à la base de la conscience.

n'est toutefois pas insurmontable : chacun d'entre nous a accès à ses propres expériences, une source abondante, utilisable pour formuler des théories. Nous pouvons utiliser des informations indirectes, telles les descriptions par d'autres de leur conscience. Les arguments philosophiques et les expériences de pensée apportent aussi des éléments. Ces méthodes ont leurs limites, mais elles nous fournissent largement de quoi commencer la recherche.

Comme les théories psychophysiques ne pourront pas être testées de manière indiscutable, elles seront plus spéculatives que celles des disciplines scientifiques classiques. Toutefois, nos expériences intimes, ainsi que les données extraites des témoignages d'autres, sont des données assez contraignantes : si nous trouvons une théorie qui en rende mieux compte que d'autres, nous l'accepterons. Comme aucune théorie psychophysique n'existe aujourd'hui, il serait de toute façon prématuré de se préoccuper de la réfutabilité de ces théories.

Conscience et présence à soi

Commençons à chercher des lois générales qui relient les processus physiques à l'expérience consciente. Lorsque nous sommes conscients de quelque chose, nous sommes généralement capables d'agir dessus et d'en parler (grâce à des fonctions objectives, physiques). Réciproquement, une information dont on peut parler et que l'on peut utiliser pour agir est généralement consciente. Par conséquent, la conscience est reliée avec ce que nous pourrions nommer la «présence à soi», le processus par lequel l'information, dans le cerveau, est rendue globalement disponible pour des opérations motrices telles que la parole ou l'action corporelle.

La notion de «présence à soi» semble banale, mais, telle que nous l'avons définie ici, elle est objective et physique, alors que la conscience ne l'est pas. On devra préciser la relation entre la présence à soi et la conscience pour analyser le cas des animaux et des bébés, qui ne parlent pas, mais, dans

les cas simples, nous voyons les grandes lignes d'une loi psychophysique : il y a conscience quand il y a présence à soi, et réciproquement.

Poursuivons cette idée en analysant le phénomène de perception visuelle. Nous ressentons notre champ de vision comme une mosaïque constamment changeante de couleurs, de formes et de configurations qui, en tant que telle, possède une structure géométrique détaillée. Puisque nous pouvons décrire cette structure, nous diriger vers ses composantes et réaliser d'autres actions qui dépendent d'elle, nous sommes incités à admettre que la structure de la conscience correspond directement à la structure de l'information donnée au cerveau par les processus neuronaux de la perception visuelle.

Nos sensations de la couleur ont aussi une dimension tridimensionnelle intrinsèque, traduite par la structure du traitement de l'information dans le cortex visuel. Cette structure s'observe sur les disques de couleurs utilisés par les artistes : les couleurs y sont disposées dans un ordre précis, du rouge

Pourquoi les neurosciences peuvent expliquer la conscience

Nous pensons qu'aujourd'hui la meilleure démarche pour expliquer la conscience est l'étude des mécanismes cérébraux qui en sont le plus directement responsables. En localisant les neurones qui correspondent le plus à la conscience et en comprenant comment ils sont associés à des neurones d'autres parties du cerveau, nous découvrirons de nouveaux éléments de réponse à ce que David Chalmers nomme le «problème difficile» : une explication complète de la manière dont la conscience subjective émerge de ces mécanismes cérébraux.

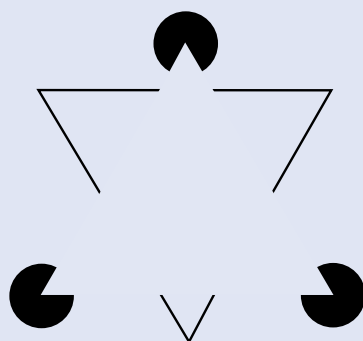
Nous admirons l'audace de D. Chalmers, qui identifie d'emblée le problème difficile et tente de le résoudre, mais nous ne sommes pas convaincus par certaines de ses expériences de pensée. Selon nous, le problème difficile peut être décomposé en plusieurs questions. Pourquoi avons-nous une conscience ? Comment avons-nous des sensations conscientes particulières (telle la sensation du bleu de la couleur bleue) ? Pourquoi certains éléments de l'expérience subjective sont-ils impossibles à communiquer à d'autres personnes (pourquoi sont-ils personnels) ? Nous pensons avoir trouvé une réponse à la dernière question, et nous avons des propositions pour les deux premières, sur la base d'un phénomène nommé la représentation neuronale explicite.

Illustrons le terme «explicite» par un exemple. En réaction à l'image d'un visage, des cellules ganglionnaires sont activées dans toute la rétine, à peu près comme les pixels d'un écran de télévision, et engendrent une représentation implicite du visage. Simultanément, elles réagissent à d'autres caractéristiques de l'image, telles des ombres, des lignes

ou les variations d'éclairage. Des neurones, à des niveaux supérieurs du cortex visuel, réagissent surtout au visage, ou au visage vu sous un certain angle. Cette catégorie de neurones permet au cerveau de représenter le visage de manière explicite. Leur mort, à la suite d'un accident vasculaire cérébral ou d'une lésion, entraîne la prosopagnosie, l'incapacité pour un sujet de reconnaître des visages familiers, y compris le sien, de manière consciente, bien que la personne reste capable d'identifier un visage comme étant un visage. De la même manière, des lésions touchant d'autres parties du cortex visuel peuvent provoquer la perte de la sensation des couleurs, excepté le noir et blanc, et sans qu'il y ait aucun déficit des récepteurs rétiniens de la couleur.

À chaque étape de sa progression dans le cortex, l'information visuelle est codée, généralement de manière semi-hiérarchique : les cellules ganglionnaires de la rétine réagissent à un point lumineux ; les neurones du cortex visuel primaire réagissent préférentiellement à des lignes ou à des arêtes ; des neurones supérieurs détectent les contours en mouvement, etc. Ceux qui réagissent aux visages ou aux objets familiers sont encore plus haut dans la hiérarchie. Au sommet, les neurones sont reliés aux structures prémotrices et motrices du cerveau, où ils activent les neurones qui déclenchent des actions telles que parler ou éviter un véhicule qui passe.

D. Chalmers pense, comme nous, que les aspects subjectifs des perceptions sont étroitement liés à l'activité des neurones qui leur correspondent. Il décrit une expérience de pensée très connue, où une neurobiologiste hypothétique, Marie, spécialisée dans la perception des couleurs, n'a jamais vu une couleur. Nous pensons que Marie ne sait



Le triangle de Kanizsa stimule des neurones qui codent explicitement un contour qui n'existe pas.

au vert selon un axe, du bleu au jaune le long d'un autre, et du noir au blanc selon un troisième (voir la figure 2). Les couleurs géométriquement proches sur la palette sont ressenties comme similaires. Elles correspondent aussi probablement à des représentations perceptives similaires dans le cerveau, dans le cadre d'un codage neuronal tridimensionnel que nous ne comprenons pas encore. Le concept fondamental sous-jacent est un principe de cohérence structurale : la structure de la conscience est l'image de la structure de l'information que nous percevons, et réciproquement.

Une autre loi psychophysique possible est un principe d'invariance organisationnelle : les systèmes physiques qui ont la même organisation abstraite seraient à l'origine du même type de sensation consciente, quelle que soit leur constitution. De ce fait, si des puces en silicium reproduisaient toutes les interactions de nos neurones, elles produiraient la même conscience. L'idée est controversée, mais elle est solidement étayée par des expériences de

pensée où l'on remplace progressivement les neurones par des circuits micro-électroniques (voir l'encadré de la page 64) : on peut donc penser qu'un jour des machines seront conscientes.

L'objectif ultime d'une théorie de la conscience est de trouver un ensemble simple et performant de lois fondamentales, analogues aux lois fondamentales de la physique. Toutefois, les principes précédents ne sont probablement pas fondamentaux ; ils semblent être des analogues psychophysiques des lois macroscopiques de la thermodynamique ou de la cinématique. Quelles seraient alors les lois fondamentales sous-jacentes ?

L'information, élément de base

Je propose que les lois psychophysiques élémentaires soient fondées sur le concept d'information. Claude Shannon, de l'Institut de technologie du Massachusetts, a défini l'information, dans les années 1940, comme un ensemble d'états séparés dont la structure est fon-

dée sur des analogies et des différences. Un nombre binaire à dix chiffres, par exemple, est un état d'information. De tels états se concrétisent dans le monde physique à chaque fois qu'ils correspondent à des états physiques (tels des potentiels électriques) ; leurs différences peuvent être transmises, par exemple, par des lignes téléphoniques.

Il y a de l'information dans la conscience. La disposition des taches de couleur dans le champ visuel est comparable à celle de pixels sur un écran de télévision. Curieusement, nous trouvons les mêmes états d'information dans la conscience et dans les processus physiques sous-jacents du cerveau. Le codage tridimensionnel des espaces de couleur suggère que l'état d'information, dans la conscience que nous avons des couleurs, correspond directement à un état d'information dans le cerveau. Nous pourrions même considérer les deux états comme des aspects distincts d'un même état d'information, qui s'exprime simultanément dans le traitement physique et dans la conscience.

pas ce que c'est que de voir une couleur parce qu'elle n'a jamais eu de représentation neuronale explicite d'une couleur dans son cerveau : elle n'a que celles des mots et des idées qui sont associés aux couleurs.

Pour que nous puissions décrire une sensation visuelle subjective, l'information doit être transmise jusqu'aux neurones moteurs qui commandent la verbalisation ou d'autres actions. Cette transmission implique toujours le codage de l'information : l'information explicite exprimée par les neurones moteurs est liée, mais n'est pas identique, à l'information explicite exprimée par la décharge des neurones associés à la sensation visuelle, à un niveau donné de la hiérarchie visuelle.

Par conséquent, nous ne pouvons pas rendre compte, avec des mots et des idées, de la nature exacte d'une sensation subjective. Nous pouvons transmettre une différence entre deux sensations subjectives, telle la différence entre le rouge et l'orange : une différence dans une aire visuelle de niveau élevé est toujours associée à une différence au niveau des étapes motrices. Nous ne pouvons donc jamais expliquer à d'autres personnes la nature d'une sensation consciente, mais seulement ses relations à d'autres sensations.

Les réponses aux deux autres questions (Pourquoi avons-nous une conscience ? Qu'est-ce qui crée une sensation consciente spécifique ?) semblent plus complexes. Pour D. Chalmers, elles requièrent l'introduction de la conscience comme nouvel élément fondamental du monde, en rapport avec la capacité de traitement de l'information d'un organisme. Mais quel type d'informations neuronales la conscience produit-elle ? Et qu'est-ce qui fait qu'un certain type d'information correspond au bleu de la couleur bleue plutôt qu'au vert de la couleur verte ? Ces problèmes sont parmi les plus difficiles de l'étude de la conscience.

Nous préférons une autre démarche, fondée sur le concept de «signification». Dans quel sens peut-on dire que les neurones qui codent explicitement un visage transmettent la signification d'un visage au reste du cerveau ? Une telle propriété doit être en relation avec le champ de projection de la cellule, la configuration de ses connexions synaptiques avec

les neurones qui codent explicitement des concepts apparentés. Finalement, ces connexions atteignent la sortie motrice. Les neurones qui réagissent à un certain visage doivent être reliés à ceux qui expriment le nom de la personne dont c'est le visage, d'autres pour sa voix, pour les mémoires dans lesquelles elle est impliquée. Ces associations entre neurones doivent être utiles du point de vue comportemental, cohérentes avec les rétroactions provenant du corps et du monde extérieur.

La signification provient de liens entre ces représentations et d'autres, qui sont dispersées dans le système cortical en un vaste réseau associatif, semblable à un dictionnaire ou à une base de données relationnelles. Plus ces connexions sont variées, plus riche est la signification. Si, comme dans l'exemple précédent de prosopagnosie, les sorties synaptiques de tels neurones du visage étaient bloquées, les cellules continueraient à réagir au visage de la personne, mais il n'y aurait pas de signification associée et, par conséquent, bien moins de sensation. Un visage continuerait d'être vu, mais pas reconnu comme tel.

Bien entendu, des groupes de neurones peuvent acquérir de nouvelles fonctions, permettant au cerveau d'apprendre de nouvelles catégories (de nouveaux visages) et d'associer ces nouvelles catégories à d'autres déjà existantes. Certaines associations élémentaires, telle la douleur, sont, à un certain degré, innées, mais elles sont affinées au cours de la vie.

L'information pourrait être un concept clef, comme D. Chalmers le pense. Pour plus de certitude, nous devons considérer des flux d'information hautement parallèles, associés, comme les neurones, en des réseaux complexes. Il serait utile de rechercher les caractéristiques nécessaires à un réseau de neurones (ou de toute autre construction informatique analogue) pour créer la signification : nous trouverions peut-être les bases neuronales de la signification. Le problème difficile de la conscience apparaîtrait alors sous un jour totalement nouveau. Il pourrait même disparaître.

Francis CRICK et Christof KOCH

La danse de la conscience dans un cerveau synthétique

Un système artificiel complexe pourrait-il avoir une conscience ? La question est fascinante. Un tel système ne sera pas construit dans les prochaines décennies, ni dans les prochains siècles, mais une simple expérience de pensée montre qu'un cerveau artificiel correctement organisé aurait le même type de conscience qu'un être humain.

Considérons un système constitué de puces en silicium qui ont la même organisation et le même fonctionnement que les neurones de notre cerveau. Chaque puce fait exactement ce que fait son analogue naturel, tout en étant reliée de la même manière aux éléments environnants. Le comportement de ce système artificiel sera exactement le même que le nôtre. Sera-t-il conscient de la même manière que nous ?

Supposons qu'il ne le soit pas : soit il a une conscience différente (il ressent du bleu lorsque nous ressentons du rouge), soit il n'a pas de conscience du tout. Nous allons examiner le premier cas (le raisonnement est le même dans les deux cas).

Comme les puces et les neurones ont les mêmes fonctions, ils sont interchangeables, à condition d'avoir une interface adaptée. Les puces peuvent donc remplacer les neurones, et l'on produit un continuum de situations où une proportion croissante de neurones est remplacée par des puces. Dans ce continuum, la conscience du système change aussi. Nous pourrions remplacer

tous les neurones de notre cortex visuel par des puces. Le cerveau résultant, dont le cortex visuel est artificiel, a une conscience différente de l'original : là où précédemment nous voyions du rouge, nous verrions du violet.

Imaginons alors que nous relions les deux cortex visuels au cerveau, par l'intermédiaire d'un commutateur à deux positions. Lorsque le commutateur est dans une position, nous utilisons le cortex visuel naturel ; dans l'autre position, c'est le cortex artificiel qui est activé. Notre sensation changerait alors du rouge au violet, ou réciproquement. Lorsqu'on actionne le commutateur de manière répétitive, nos sensations "dansent" entre les deux états conscients différents (rouge et violet).

Comme l'organisation de notre cerveau ne change pas, il ne peut y avoir de changement comportemental quand on bascule le commutateur. Par conséquent, lorsqu'on nous demande ce que nous voyons, nous disons que rien n'a changé : nous ne voyons que du rouge, bien que les deux couleurs dansent sous nos yeux. Cette conclusion est absurde : l'hypothèse qu'un système artificiel, dont l'organisation et le fonctionnement sont les mêmes que ceux d'un cerveau neuronal, a des sensations conscientes différentes est donc absurde. Les systèmes qui ont la même organisation ont donc la même conscience.



Si un cerveau synthétique en tout point semblable au nôtre avait une conscience différente, une pomme pourrait nous apparaître alternativement rouge ou bleue.

leurs dansent sous nos yeux. Cette conclusion est absurde : l'hypothèse qu'un système artificiel, dont l'organisation et le fonctionnement sont les mêmes que ceux d'un cerveau neuronal, a des sensations conscientes différentes est donc absurde. Les systèmes qui ont la même organisation ont donc la même conscience.

L'information, ou du moins certaines informations, posséderait donc deux aspects de base : l'un physique et l'autre intime. Cette hypothèse est un principe fondamental, sur lequel reposeraient les relations qui unissent les opérations physiques et la conscience. Partout où nous trouvons de la conscience, elle est l'un des aspects d'un état d'information, dont l'autre aspect est codé dans un processus physique à l'intérieur du cerveau. Il faudrait étoffer cette proposition pour en faire une théorie satisfaisante, mais elle remplit convenablement les principes précédemment mentionnés (des systèmes qui ont la même organisation matérialisent la même information, par exemple) et elle expliquerait de nombreux aspects de la conscience.

L'idée est au moins compatible avec plusieurs autres, telle la proposition du physicien John Wheeler, selon laquelle l'information est un fondement de la physique de l'Univers. Les lois de la physique pourraient être reformulées en termes d'information : les lois physiques et psychophysiques se rejoindraient alors. Les théories de la physique et de la conscience pourraient même être réunies en une super théorie de l'information.

La généralité excessive du concept d'information pose toutefois un problème : un thermostat, qui contient de l'information, a-t-il une conscience ? Il y a au moins deux réponses possibles. Nous pourrions inclure dans les lois fondamentales le fait qu'une partie seulement de l'information a un aspect de conscience, peut-être en fonction du traitement physique qu'elle subit. Nous pourrions, au contraire, admettre que toute information a un aspect de conscience : lorsqu'il y a un traitement complexe de l'information, il y a une conscience complexe, et lorsqu'il y a un traitement simple de l'information, il y a une conscience simple. Dans ce cas, même un thermostat pourrait avoir une conscience, mais elle serait beaucoup plus simple que même la conscience fondamentale que nous avons d'une couleur, et il n'aurait aucune émotion ni pensée associée. Une telle idée semble bizarre, mais si la conscience est vraiment fondamentale, elle doit être largement répandue. De ces deux possibilités, nous devrions choisir celle qui s'intégrerait dans la théorie la plus puissante.

Ces idées produiront-elles une théorie plus vaste, qui prédira la structure précise de notre conscience à partir des

mécanismes physiques de notre cerveau ? Si le projet échoue et que mes idées se révèlent fausses, d'autres idées seront explorées, et d'autres théories fondamentales seront sans doute élaborées. Nous connaissons peut-être un jour le plus grand mystère de l'esprit.

David CHALMERS est membre du département de philosophie de l'Université de Santa Cruz.

Daniel DENNETT, *La conscience expliquée*, Éditions Odile Jacob, 1993.

Roger PENROSE, *Les ombres de l'esprit*, InterÉditions, 1995.

David CHALMERS, *Absent Qualia, Fading Qualia, Dancing Qualia*, in *Conscious Experience*, sous la direction de Thomas Metzinger, Ferdinand Schöningh, 1995.

Explaining Consciousness : The «Hard Problem», numéro spécial du *Journal of Consciousness Studies*, vol. 2, n° 3, automne 1995.

Daniel PINKAS, *La matérialité de l'esprit*, Éditions La Découverte, 1995.

The Nature of Consciousness : Philosophical and Scientific Debates, sous la direction de Ned Block, Owen Flanagan et Güven Güzeldere, MIT Press (à paraître).
