

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.

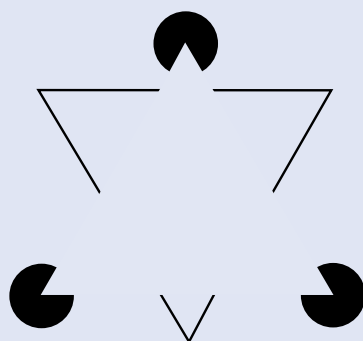
Nous admirons l'audace de D. Chalmers, qui identifie d'emblée le problème difficile et tente de le résoudre, mais nous ne sommes pas convaincus par certaines de ses expériences de pensée. Selon nous, le problème difficile peut être décomposé en plusieurs questions. Pourquoi avons-nous une conscience ? 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