

Qui suis-je ?

Le libre arbitre à la lumière des neurosciences cognitives

Que penser aujourd'hui des rapports entre corps et esprit ? Est-il possible de décoder le contenu de mes pensées à partir de l'activité de mon cerveau ? Et comment concevoir le libre arbitre à partir du moment où ce dernier semble laisser la place au déterminisme biologique ? Les neurosciences cognitives apportent aujourd'hui à ces questions un éclairage nuancé qui doit nous amener à reconsidérer notre conception de la liberté. Si notre vie mentale ne peut en effet pas être autre chose qu'un aspect de l'activité biologique de nos neurones, l'activité de ces derniers est à son tour entièrement déterminée par l'histoire de nos interactions avec le monde et les autres. Si la liberté de pensée est bel et bien une liberté biologique, la pensée n'en est pas pour autant moins libre...

AXEL CLEEREMANS

La perspective que l'homme entretient sur sa propre vie mentale et, d'une manière plus générale, sur les rapports entre corps et esprit, s'est aujourd'hui profondément modifiée, d'abord en vertu des progrès remarquables des méthodes dont nous disposons pour apprécier l'activité cérébrale, ensuite en vertu des moyens dont nous disposons pour en modifier le fonctionnement. Ces développements, essentiellement technologiques — de la résonance magnétique nucléaire à la stimulation intracrânienne, nous suggèrent ainsi que l'ensemble de notre expérience du monde et des autres est inscrite dans notre biologie. Pas un jour ne passe en effet qui ne voit paraître l'un ou l'autre article apportant d'irréfutables arguments qui soutiennent l'idée qu'aucun évènement mental n'échappe au déterminisme biologique : ici, on localise précisément l'empathie ; là, on démonte le sentiment religieux, là encore on vous démontre que votre cerveau est parfaitement capable de prendre les décisions importantes de votre vie sans qu'il soit nécessaire que vous soyez conscient de ses délibérations.

De telles conclusions suscitent souvent l'inquiétude, voire l'agressivité, tant il est vrai qu'il nous semble absurde de considérer que notre vie mentale se trouverait ainsi entièrement déterminée par l'activité de notre cerveau. On

retrouve ainsi les débats suscités il y a déjà plus d'un quart de siècle par les conclusions de *L'Homme neuronal*, ouvrage pionnier du neurologue français Jean-Pierre Changeux, publié en 1983.

Cette perspective est néanmoins la seule qui est compatible avec une vision matérialiste de notre vie mentale. Et cette perspective matérialiste est à son tour la seule qui fait sens à partir du moment où l'on se propose d'abandonner le dualisme. Car en effet, en rejetant l'idée que quelque chose de notre vie mentale échappe nécessairement à la biologie, on se voit forcé d'accepter que ce « quelque chose » dépende alors d'autre chose que de l'organisation de la matière.

DU CERVEAU À L'ESPRIT : COMMENT SE DÉBARRASSER DU DUALISME ?

Deux facteurs essentiels expliquent sans doute pourquoi l'idée que notre vie mentale est intégralement réductible à l'activité de notre cerveau est tellement difficile à accepter.

Le premier facteur est que notre expérience du monde ne nécessite en rien une référence à l'activité de notre cerveau. La manière dont ce dernier traite l'information nous est totalement opaque : nous n'avons jamais conscience des processus neurocognitifs qui sont à l'œuvre quand nous réalisons des tâches aussi diverses que chercher en mémoire le nom d'un ami, planifier la journée de demain ou déterminer que le feu est vert. Nos facultés cognitives — la mémoire, l'action et la perception — obéissent à des lois dont nous ignorons encore quasiment tout et sont déterminées par des processus dont nous n'avons aucun moyen de prendre conscience, sauf peut-être dans certaines circonstances particulières lors desquelles on peut avoir l'impression, par exemple, d'un « effort mental ».

Cette opacité des processus impliqués, ou, pour le dire autrement, le fait que l'activité du cerveau ne nous apparaît pas comme transparente, explique que nous considérons l'activité de notre cerveau comme distincte de l'activité de notre « moi » : je ne suis pas mon cerveau. Or ce « je » ne peut être rien d'autre que l'activité de mon cerveau, à moins que l'on soit prêt à accepter une vision dualiste des rapports entre corps et esprit. Il est significatif de remarquer qu'une forme de dualisme linguistique continue d'empreindre le discours que nous avons à propos de nous-même, et ce y compris dans le domaine de ces sciences neurocognitives qui cherchent précisément à nous en débarrasser. C'est ainsi que l'on peut dire sans sourciller des choses telles que « Votre cerveau a décidé à votre place », ou encore « Votre cerveau a déjà lui-même déterminé la position que la balle occupera dans quelques instants avant même que vous n'en preniez conscience ». Mais qui peut bien être ce « vous » qui semble être autre chose que l'activité de votre cerveau ?

Il est évidemment tout à fait naturel de se vivre et de se penser comme différent de l'activité de notre cerveau. Mais cela n'est en rien différent de ce que nous disons quand nous caractérisons les états de notre corps. De la même

manière que nous disons « J'ai mal à l'estomac » plutôt que « La muqueuse de mon fundus produit trop d'acide chlorhydrique », nous dirons « Je me sens un peu déprimé » plutôt que « Mon cerveau est en manque de sérotonine ». Nous adoptons ce langage en raison du fait que les mécanismes biologiques qui déterminent nos états mentaux n'ont aucune incidence sur notre souhait de communiquer ces états mentaux à d'autres personnes, et que ces actes de parole ont précisément pour fonction de communiquer des états mentaux et émotionnels plutôt que d'en expliquer les mécanismes. Mais rien de cela n'a d'incidence sur le fait qu'en ultime ressort, « je » ne peut être rien d'autre que l'activité de mon cerveau. Notons en outre qu'il n'y a en principe aucune raison d'arrêter la réduction à la biologie : de la même manière que l'on peut réduire l'activité mentale à l'activité biologique du cerveau, on pourrait réduire cette dernière aux processus physico-chimiques dont elle dépend. Cette deuxième réduction n'est pas plus pertinente que la première en ce qui concerne l'appréciation que nous pouvons avoir de nos états mentaux.

Le deuxième facteur qui explique pourquoi nous éprouvons tellement de difficultés à accepter cette idée que notre vie mentale peut se réduire à l'activité de notre cerveau est qu'il y a une véritable forme de désenchantement dans le constat que nous proposent aujourd'hui les neurosciences. À quoi bon se « casser la tête » à progresser s'il est vrai que je suis ainsi réductible ? Et combien sinistre est la conclusion que la beauté de mes sentiments peut se voir être réduite à l'activité de quelques centaines de millions de neurones ! Il y a néanmoins dans ces assertions une erreur fondamentale : celle de considérer le cerveau comme un simple organe exécutif, qui prendrait ses ordres d'un ailleurs mal défini, mais situé en tous les cas en dehors de la biologie. On retrouve ici le mirage du dualisme persistant évoqué plus haut.

Or cette perspective, certes profondément intuitive, que le cerveau est un simple organe exécutif est encore une fois totalement incompatible avec ce que les neurosciences modernes nous donnent à voir. Il apparaît ainsi que même de simples décisions, comme celle de pousser sur un bouton, sont prises « par notre cerveau » avant même que « nous » ne prenions conscience de l'intention d'effectuer l'action. En d'autres termes, la conscience que j'ai de mes actions semble suivre, et non pas précéder comme le voudraient nos intuitions, nos actions elles-mêmes. C'est également la conclusion qu'offre l'ouvrage récent de Daniel Wegner, ouvrage intitulé *The illusion of Conscious Will*, dans lequel l'auteur développe une théorie qu'il a appelée la « théorie de la causation mentale apparente ». Selon cette théorie, tant nos intentions que nos actions elles-mêmes trouvent toutes deux leur origine dans l'activité inconsciente du cerveau. Ce dernier établit alors un lien causal (mais illusoire) entre la survenue de l'intention et la survenue de l'action. Notre conscience du monde, dans cette perspective, ne serait donc pas grand-chose de plus qu'une reconstruction narrative des chaînes causales inconscientes qui déterminent notre comportement.

DÉCODER PENSÉES ET INTENTIONS

De nombreuses expériences récentes sont ainsi venues remettre en cause, et d'une manière difficilement réfutable, ce constat particulièrement désenchanté. On peut citer ici par exemple le travail de John-Dylan Haynes, dont les recherches se situent à l'avant-plan d'un nouveau domaine d'étude appelé *Mindreading*. La première expérience de Haynes et ses collaborateurs (Haynes & Rees, 2005) proposait aux sujets de détecter un stimulus visuel — un grillage texturé orienté soit vers la gauche, soit vers la droite — présenté très rapidement et masqué. À chaque essai, le sujet, qui passait l'expérience en résonance magnétique nucléaire, devait indiquer l'orientation du stimulus en enfonçant l'une des deux touches. Les données comportementales indiquèrent que les sujets étaient incapables de décider de l'orientation du stimulus, la précision moyenne de leurs réponses ne dépassant pas le niveau du hasard. Haynes et ses collègues analysèrent ensuite l'activité cérébrale des sujets en vue de déterminer s'il était possible d'identifier, dans l'aire cérébrale V1 — la première région corticale du système visuel — des configurations d'activité neuronale permettant de reconstruire le stimulus présenté aux sujets. Pour ce faire, ils avaient préalablement entraîné des réseaux de neurones artificiels à décider quel stimulus était présent sur la base de l'activité cérébrale de sujets à qui on avait montré les mêmes stimuli, mais dans des conditions telles que ceux-ci étaient invisibles. Et les résultats montrèrent que ces réseaux de neurones entraînés étaient en effet capables de reconstruire, avec un taux de réussite d'environ 80 %, les stimuli invisibles montrés aux sujets !

En d'autres mots : on trouve dans V1 des configurations d'activité neuronale qui permettent d'identifier un stimulus présenté à des sujets qui sont eux-mêmes incapables de décider de l'orientation de ces stimuli et qui disent par ailleurs n'avoir rien vu. L'expérimentateur en sait donc paradoxalement plus à propos du stimulus que le sujet lui-même... une conclusion dont la toute nouvelle discipline qu'est la neuroéthique commence à peine à explorer les implications légales et morales car elle soulève la possibilité qu'il sera bientôt possible de décoder l'activité neurale d'une manière suffisamment fine que pour reconstruire, en quelque sorte, les pensées d'un individu. Cette possibilité, parce qu'elle démolirait le dernier rempart de notre vie privée, à savoir le caractère fondamentalement privé de nos états mentaux, soulève une inquiétude justifiée qui exigera indubitablement une prise de position sociétale.

Ultérieurement, Haynes et ses collaborateurs (Soon, Brass, Heinze, Haynes, 2008) ont répliqué, en utilisant les mêmes méthodes d'imagerie cérébrale, les fameuses expériences concernant le libre arbitre initialement réalisées par le neurophysiologiste américain Benjamin Libet. Dans la situation expérimentale de Libet (Libet, Gleason, Wright, Pearl, 1983), on demande aux sujets d'entamer un mouvement volontaire simple (lever un doigt) au moment de leur choix. On enregistre simultanément leur activité cérébrale par électroencéphalographie, en recherchant dans le signal électrique un potentiel particulier signalant la préparation motrice (le « *readiness potential* »). On de-

mande en outre aux sujets de repérer la position d'un stimulus se déplaçant rapidement de manière circulaire sur un dispositif ressemblant au cadran d'une montre au moment où ils prennent conscience pour la première fois de l'intention d'agir. En réalignant toutes les données (le moment où le sujet effectue le mouvement; le moment où il dit prendre conscience de l'intention d'agir, et le décours temporel du potentiel de préparation), Libet montra ainsi que l'activité cérébrale indiquant l'imminence d'un mouvement précède de quelques centaines de millisecondes le moment où le sujet dit prendre conscience pour la première fois de l'intention d'agir.

En d'autres termes, la conscience d'une intention semble suivre, et non précéder comme le voudrait notre intuition, l'initiation par le cerveau de l'activité neuronale préparatrice de l'action. Ce renversement de causalité suscita une discussion qui continue encore aujourd'hui, notamment en raison de certaines difficultés méthodologiques qui rendaient problématique l'interprétation proposée par Libet.

C'est ici que l'expérience de Soon et de ses collaborateurs apporte de précieuses innovations. Ces auteurs ont adapté la méthodologie proposée par Libet à une étude en résonance magnétique nucléaire, qui permet, par contraste avec l'électroencéphalographie, d'apprécier l'activité cérébrale dans un ensemble de régions beaucoup plus étendu et de mieux localiser cette activité. À chaque essai, les participants devaient décider librement d'appuyer sur l'une des deux touches plutôt que d'effectuer un simple mouvement — une autre différence importante avec le paradigme de Libet. Pendant cette période, différentes lettres étaient affichées à l'écran. De la même manière que dans le protocole de Libet, on demandait aux participants de se rappeler la lettre qui était affichée à l'écran au moment où ils prenaient conscience de leur décision. Ils devaient ensuite rapporter cette lettre à la fin de l'essai. En suivant la même procédure que dans l'expérience précédente, c'est-à-dire en utilisant des méthodes de reconnaissance de patrons, les auteurs ont ensuite cherché à identifier les régions cérébrales dont l'activité était prédictive de la décision prise par les sujets.

Deux régions cérébrales s'avèrent ainsi particulièrement importantes : une région du cortex préfrontal (BA10) et une région pariétale impliquant le précuneus et la partie postérieure du cortex cingulaire. De manière cruciale, l'activité de ces deux régions permettait de prédire la décision libre prise le sujet (pousser à gauche ou à droite) sept secondes avant que le sujet n'ait déclaré avoir pris conscience de sa décision ! Autrement dit, il est possible de savoir quelle action le sujet va réaliser avant même que le sujet n'en prenne conscience... Ce résultat récent renforce donc encore les conclusions de Libet, et ce d'autant plus que le délai entre la présence d'une activité cérébrale prédictive et la prise de conscience correspondante se compte maintenant non plus en millisecondes, mais bien en secondes.

VOTRE CERVEAU EST PLASTIQUE

Ces deux expériences, ainsi que d'autres, de plus en plus nombreuses, indiquent clairement ce que nous savions quelque part déjà, à savoir que notre conscience est une conséquence plutôt qu'une cause de l'activité cérébrale menant à nos actions. Encore une fois, ce résultat n'est pas surprenant (comment pourrait-il en être autrement ?), mais vient malgré tout bouleverser nos intuitions, et surtout, le sentiment que nous disposons d'un libre arbitre totalement affranchi du déterminisme de la biologie (un sentiment de libre arbitre qui est par ailleurs illusoire pour d'autres raisons également, voir à ce propos l'excellent ouvrage de Daniel Dennett, 1985).

Et pourtant, quel est le problème exactement ? Quel est le problème, à partir du moment où nous acceptons que « je » ne suis pas différent de mon cerveau ? La menace que nous percevons dans les progrès des neurosciences cognitives disparaissent dès que nous reconnaissons que le cerveau est un organe profondément plastique plutôt que le simple « *hardware* » biologique que nous imaginons. Nous sommes en cela trompés par les métaphores que nous ont données les sciences de l'informatique et de l'intelligence artificielle. Les ordinateurs avec lesquels la plupart d'entre nous interagissent quotidiennement fonctionnent selon des principes radicalement différents de ceux qui régissent le traitement de l'information par la machine biologique qu'est le cerveau. En particulier, il existe chez les premiers une distinction forte entre matériel et logiciel, alors que cette distinction est totalement absente chez ce dernier. Ce que fait un ordinateur est à peu près complètement défini par le programme qu'il est en train d'exécuter, et l'existence de ce programme est indépendante de la machine qui l'exécute. Ce que nous savons, par contre, n'existe qu'en vertu de l'inscription biologique de ce savoir : la machine-cerveau ne fait qu'un avec notre « logiciel » mental.

Cette réalité a (outre le fait un peu regrettable que l'on ne puisse pas se télécharger un logiciel de maîtrise du tennis, par exemple) une implication cruciale qui nous échappe souvent : chaque expérience que nous vivons trouve une inscription biologique. Le simple fait de lire ce texte modifie vos connexions synaptiques... et l'implication centrale de ce simple fait est que l'activité de votre cerveau reflète tant la richesse que la diversité de votre propre expérience du monde et des autres. Nous apprenons tout le temps, que nous le voulions ou non. Si les contenus de notre vie mentale sont entièrement déterminés par l'activité de notre cerveau, l'activité de ce dernier est à son tour quasiment entièrement déterminée par l'histoire de nos interactions avec le monde et les autres. Cette boucle magnifique qui connecte notre monde intérieur à la réalité inscrit notre activité mentale dans un substrat biologique changeant constamment. Au déterminisme inflexible que conjure la métaphore maintenant obsolète de l'esprit en tant qu'ordinateur, il faut donc opposer un déterminisme ouvert, plastique, en dialogue constant avec le monde. En se modifiant constamment, le cerveau, loin d'être ce simple « *hardware* » que nous laisse imaginer notre familiarité avec les ordinateurs, est en réalité simultanément la mémoire de nos

Bibliographie

- Dennett D.C. (1993), *La conscience expliquée*, Odile Jacob.
- Changeux J.-P. (1983), *L'Homme neuronal*, Fayard.
- Dennett D. C. (1985), *Elbow Room. The varieties of free will worth wanting*, Oxford University Press.
- Frith C. (2007), *Making up the mind: How the brain creates our mental world*, Blackwell.
- Haynes J.-D., Rees G. (2005), « Predicting the orientation of invisible stimuli from activity in human primary visual cortex », *Nature Neuroscience*, 8(5), 686-691.
- Libet B., Gleason C. A., Wright E. W., Pearl D. K. (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(623-642).
- Soon C. S., Brass M., Heinze H.-J., Haynes J.-D. (2008). « Unconscious determinants of free decisions in the human brain », *Nature Neuroscience*, 11, 543-545.
- Wegner D. M. (2002), *The Illusion of Conscious Will*, Bradford Books, MIT Press.

expériences passées ainsi que le siège de nos actions futures. La machine et son logiciel ne font qu'un.

Mais ce cerveau est avant tout votre cerveau, un organe dans lequel se sont inscrites progressivement toutes les expériences qui ont compté dans votre vie. Nous pouvons donc faire l'économie du renoncement: rien n'est perdu (ni gagné, par ailleurs), en acceptant l'idée que notre vie mentale est intégralement causée par le cerveau, en interaction avec notre environnement et avec les autres. Ce qui est peut-être plus problématique est d'accepter que la conscience que nous avons de nous-même est une sorte de post-diction, comme l'avait déjà suggéré le philosophe américain Daniel Dennett dans son ouvrage de 1993 intitulé *La conscience expliquée*. On peut sans doute trouver une forme de résolution ici en enracinant la fonction de la conscience dans la communication: si je suis conscient des états de mon cerveau, c'est avant tout pour pouvoir en parler aux autres et pour, en retour, m'expliquer mon propre fonctionnement. « La conscience, c'est pour les autres », comme me le disait récemment le psychologue anglais Chris Frith (voir aussi son ouvrage récent *Making up the mind*). La conscience aurait donc, selon cette perspective en plein essor actuellement, une origine sociale: « Je suis un autre », comme disait le poète.

Dès lors, l'impression de désenchantement que l'on peut éprouver à la lumière des travaux récents en neurosciences cognitives s'efface pour laisser la place à un sentiment d'émerveillement devant la sophistication presque magique de notre machine biologique, non seulement plastique, mais également tournée vers les autres. Et les neurosciences cognitives ont bien compris, aujourd'hui, le formidable défi que représente le projet de faire le pont, en quelque sorte, entre ces deux points extrêmes de notre savoir — l'expérience subjective que chacun d'entre nous a du monde et des autres, et notre compréhension de plus en plus fine des mécanismes biologiques qui en sont responsables. ■