

**La neuroéthique,
une nouvelle frontière pour les sciences humaines**

Stanislas DEHAENE

Recensés :

- Bernard Baertschi, *La neuroéthique : ce que les neurosciences font à nos conceptions morales*. Éditions La découverte, 2009 ;
- Kathinka Evers, *Neuroéthique : quand la matière s'éveille*. Editions Odile Jacob, 2009.

Que nous enseigne la science du cerveau sur nos jugements et nos comportements moraux ? Deux livres traitent la question en présentant une discipline nouvelle : la neuro-éthique.

Chimistes et biologistes le savent bien : les réactions les plus actives et les évolutions les plus radicales se situent toujours aux frontières, là où deux milieux jusque-là isolés entrent soudainement en contact. Les idées jaillissent selon la même loi – c'est souvent en marge de deux disciplines que naissent des propositions foncièrement nouvelles. Dans la science contemporaine, l'une des frontières qui se déplace le plus vite est celle des neurosciences cognitives – la branche des neurosciences qui tente de comprendre les mécanismes neuronaux

des compétences humaines, en combinant les méthodes de la psychologie et de l'imagerie cérébrale. La conscience de soi, le jugement moral, l'impact de l'éducation et tant d'autres questions qui appartenaient autrefois à la philosophie morale, sont devenus de nouveaux objets de recherche pour les neurosciences. Une crainte légitime naît de cette hyperactivité qui peut paraître expansionniste : l'acide matérialiste des neurosciences ne ronge-t-il pas les fondements même de notre société, en fragilisant voire en anéantissant ses piliers que sont le libre arbitre, la responsabilité, l'identité individuelle et le jugement moral ?

À la frontière de la philosophie et des neurosciences, la discipline émergente qu'il est convenu d'appeler « neuroéthique » s'est emparée de ces questions. Deux excellents ouvrages récents en français en dressent un panorama détaillé. Bernard Baertschi, maître d'enseignement et de recherches à l'Institut d'éthique biomédicale et au département de philosophie de Genève, œuvre essentiellement en philosophe classique. Le sous-titre de son livre, « ce que les neurosciences font à nos conceptions morales », donne le ton. Il s'agit avant tout de mettre en perspective les principales théories philosophiques du jugement moral, de la responsabilité et de l'identité personnelle à la lumière des découvertes récentes des neurosciences. Comme il l'assène à ses collègues philosophes, « si déjà au XVIII^e siècle, il était judicieux d'aborder les questions philosophiques en tenant compte de ce que les sciences nous apprenaient, c'est devenu une exigence absolue de nos jours ».

Katinka Evers, professeur au centre de recherche en éthique et bio-éthique de l'Université d'Uppsala (Suède), bien qu'elle s'appuie sur une base empirique et philosophique très semblable, se donne un objectif plus ambitieux encore. Issu d'une série de conférences au Collège de France, son livre esquisse une vision philosophique nouvelle de la condition humaine, un « matérialisme éclairé » qui tente de réconcilier notre nature indéniablement

neurobiologique avec une vision humaniste respectueuse de la personne et de sa vie harmonieuse en société. Amplifiant les propos de Jean-Pierre Changeux dans *L'homme neuronal* (1983) puis *L'homme de vérité* (2004), deux ouvrages que l'on peut légitimement considérer comme des pionniers de la neuroéthique, elle note avec justesse : « nous sommes des hommes et des femmes neuronaux, au sens où tout ce que nous faisons, pensons et sentons est une fonction de l'architecture de nos cerveaux ; et pourtant ce fait n'est pas encore totalement intégré à nos conceptions générales du monde, ni à nos conceptions de nous-mêmes » (p. 32). Puissante litote ! En réalité, un très ancien dualisme, qui situe les fonctions mentales sur un plan distinct de celles du corps et donc du cerveau, prévaut encore à tous les échelons de notre société, qu'il s'agisse bien sûr de nos églises et religions, mais aussi de notre système judiciaire, de notre organisation sociale ou de nos écoles (les concepts de mérite ou de talent résistent-ils aux avancées des neurosciences ?). C'est donc à un grand bousculement des idées reçues que nous invitent ces deux livres exigeants, parfois difficiles mais toujours vivement stimulants.

Qu'est-ce que la neuro-éthique ?

La neuro-éthique recouvre en fait au moins deux domaines, que Baertschi appelle respectivement « l'éthique des neurosciences » et « la neuroscience de l'éthique », tandis qu'Evers adopte une distinction similaire entre neuro-éthique fondamentale et appliquée. « La neuroéthique fondamentale s'interroge sur la manière dont la connaissance de l'architecture fonctionnelle du cerveau et de son évolution peut approfondir la compréhension que nous avons de notre identité personnelle, de la conscience et de l'intentionnalité, ce qui inclut le développement de la pensée morale et du jugement moral ; la neuroéthique appliquée étudie l'éthique des neurosciences, par exemple les problèmes éthiques que soulèvent les techniques

d'imagerie neuronale, l'amélioration cognitive ou la neuropharmacologie. » (K. Evers, p. 204-205)

Les deux ouvrages s'accordent également sur les grands axes selon lesquels la réflexion neuroéthique doit s'organiser : (1) Comment envisager les concepts de responsabilité et de libre arbitre, notamment du point de vue juridique, si l'homme n'est qu'une machine neuronale ? (2) Quels fondements donner aux principes de l'éthique si notre sens moral n'est issu que d'un cerveau bricolé par l'évolution, où coexistent des mécanismes de jugement émotionnel et rationnel parfois contradictoires ? (3) Comment réviser les normes sociales en vue d'une « vie bonne » pour tous, et quelles limites devons-nous imposer aux développements neurotechnologiques, alors que les neurosciences offriront bientôt la possibilité d'altérer le fonctionnement cérébral, en bien comme en mal ?

Il importe de dire d'emblée que, face à ces perspectives parfois effrayantes, nos deux auteurs font généralement preuve d'une grande retenue et d'un optimisme que je partage, et qui contraste agréablement avec la paralysante prévalence du principe de précaution. « L'éthique n'a rien à craindre des neurosciences : pas de panique morale ! » dit Braetschi. « La signification, le sens, la dignité, la *raison d'être*, et autres notions si chères aux humains ne sont pas perdues pour nous du fait d'être des constructions qui dépendent de notre architecture cérébrale », lui répond Evers.

Résumer pourquoi il en va ainsi serait une gageure, tant les deux livres sont denses en arguments. Je tenterai plutôt d'en dégager quelques points essentiels, en y ajoutant mon propre grain de sel. Tout d'abord, les neurosciences ont perdu un peu de l'arrogance qui leur faisait envisager d'éliminer du champ de la science tout le vocabulaire de la psychologie.

Inversement, la psychologie cognitive ne craint plus de s'intéresser à l'architecture du cerveau, ni d'aborder des questions autrefois jugées taboues et impossibles à analyser scientifiquement telles que la conscience, le sens du soi, ou les émotions. L'ultra-réductionnisme des behavioristes a vécu, tout comme le cognitivisme naïf fondé sur la seule métaphore de l'ordinateur. Leur succède une neuroscience cognitive à la fois plus homogène et plus franchement pluridisciplinaire, qu'Evers qualifie de « science psychophile », qui n'évacue plus les états mentaux ou émotionnels comme des fictions immatérielles aux antipodes de la réalité physique, mais qui tente au contraire d'élucider quelles architectures neuronales les sous-tendent.

Prenons l'exemple de la conscience, aujourd'hui objet d'intenses recherches empiriques. Aucun chercheur ne nie plus que la prise de conscience d'une information corresponde à une transition majeure dans l'état d'activité de nos réseaux de neurones. En identifier les mécanismes nécessite de recueillir simultanément des informations d'ordre subjectif et psychologique (le rapport verbal du sujet qui décrit sa perception) et des données d'ordre objectif et neuronal (l'activité cérébrale mesurée par l'électro-encéphalographie ou l'imagerie par résonance magnétique). Selon cette vision matérialiste mais non réductionniste de la conscience, le cerveau humain est un organe biologique que l'évolution darwinienne a doté d'une architecture structurée qui projette des représentations mentales, des évaluations, une perspective personnelle (un « soi ») et des plans d'action sur le monde qui l'entoure. Cette organisation neuronale est variable et plastique, non soumise au seul diktat des gènes, mais capable d'incorporer les conventions culturelles et les règles morales. La conscience fait partie de ses propriétés matérielles les plus fondamentales.

Persistance du libre arbitre

Il n'y a aucune contradiction, dans le cadre de ce « matérialisme éclairé », à parler des fondements cérébraux de la responsabilité morale. Le concept de choix libre et volontaire ne fond pas à la chaleur des avancées neuroscientifiques. C'est une erreur profonde de le nier au prétexte d'un éventuel déterminisme de l'activité nerveuse, ou d'en rechercher les fondements dans l'incertitude quantique fondamentale. Ceux qui pensent que le libre arbitre nécessite un jet de dés quantiques commettent une erreur de niveau comparable à celle d'un physicien qui recherchait l'explication de la solidité d'une charpente au niveau atomique. En réalité, ce que nous appelons – légitimement ! – le libre arbitre n'est que la description fonctionnelle d'un cerveau qui, par son organisation même, dispose d'une capacité d'envisager plusieurs voies d'action, d'en évaluer les conséquences, et d'en choisir l'une sans que ce choix soit intégralement prévisible de l'extérieur ou même de l'intérieur. Ainsi, affirme Baertschi, « pas besoin de libre arbitre au sens fort [i.e. une brèche de la causalité physique] pour évoquer la responsabilité ; il suffit d'une action intentionnelle lucidement voulue ». L'homme neuronal reste donc une personne responsable, du moins si son cerveau n'a pas subi d'altérations ou de lésions.

Ce mode de fonctionnement normal du cerveau, intentionnel, capable de réfléchir lucidement et en toute conscience, les neurosciences contribuent elles-mêmes à le définir. L'imagerie cérébrale a en effet révélé des distinctions massives entre celles de nos actions qui relèvent d'un traitement automatique, lequel peut parfois se dérouler en l'absence de toute conscience, et celles qui impliquent un choix réfléchi, sous contrôle. Le cortex préfrontal occupe une place centrale dans le dispositif de contrôle exécutif, également appelé « administrateur central » qui régule, choisit ou inhibe en fonction des buts que nous nous fixons.¹ Il s'agit de la fraction des hémisphères cérébraux qui se situe juste en dessous du

¹ Passingham, R. (1993). *The frontal lobes and voluntary action* (Vol. 21). New York: Oxford University Press.

front et constitue jusqu'à 30% de la surface du cortex chez l'homme, une fraction inégalée parmi tous les primates. Son bon fonctionnement est essentiel à toute décision réfléchie. Même si nos connaissances sur son organisation restent fragmentaires, on peut légitimement parler, à son sujet, du fondement cérébral de la raison, de la planification et de la synthèse mentale.

Le facteur émotionnel

Dire que nos choix sont rationnels et conscients n'exclut aucunement qu'ils soient, au moins en partie, guidés par nos émotions, parfois d'une manière non-consciente. Dans le nouveau cadre conceptuel qui émerge des recherches neuroscientifiques les plus récentes, tant les circuits émotionnels responsables de nos réactions corporelles rapides (peur, joie, colère...) que les grandes voies qui libèrent des agents chimiques neuromodulateurs (dopamine, noradrénaline, acétylcholine...) constituent les briques d'algorithmes sophistiqués d'évaluation de notre situation passée et future, à différentes échelles temporelles. Leurs signaux mitigés constituent les piliers d'un système de valeur sur la base duquel nous choisissons de maintenir ou de changer notre comportement.²

Les avancées des neurosciences contribuent ainsi à mieux discerner les limites à l'intérieur desquelles s'exercent nos choix libres et responsables. Le fonctionnement harmonieux des systèmes cérébraux de l'émotion apparaît essentiel aux conduites rationnelles. Leur perturbation réduit l'éventail des choix disponibles à l'organisme, parfois jusqu'à dénier à la personne toute capacité de contrôler son comportement. L'addiction, par exemple, intervient lorsqu'un agent pharmaco-chimique pervertit de leur but premier les circuits dopaminergiques et cholinergiques de prise de décision qui, au fil de l'apprentissage, adaptent

² Yu, A. J., & Dayan, P. (2005). "Uncertainty, neuromodulation, and attention". *Neuron*, 46(4), 681-692.

nos choix en fonction des signaux de récompense que nous recevons de l'environnement. Des modèles théoriques de ces circuits neuronaux³ expliquent comment les drogues, même lorsqu'elles cessent d'engendrer du plaisir, continuent de biaiser la prise de décision jusqu'à faire de la personne droguée un être, non pas sans volonté, mais dont toutes les décisions sont réorientées vers un objectif pathologique. Evers et Baertschi auraient pu développer cet exemple, ils choisissent plutôt d'explorer, à la suite de Damasio, le cas bien connu des patients atteints, tel le célèbre Phineas Gage, de lésions du cortex orbitofrontal mésial, et devenus incapables d'éprouver dans leur corps et dans leur cerveau les signaux émotionnels associés à l'anticipation des conséquences, bonnes ou mauvaises, des actions envisagées.⁴ Ces personnes font de mauvais choix de vie, leur conduite devenue amoral les conduit souvent à perdre leur travail, leurs amis ou leur famille, parfois à terminer en prison, alors même qu'ils restent capables d'énoncer verbalement les règles morales que leur cerveau malade ne leur permet plus d'appliquer. Cette dernière compétence, que l'on pourrait qualifier de théorie de la morale, peut même s'évanouir si la lésion survient très tôt dans l'enfance, engendrant alors des adolescents à la merci de poussées utilitaristes et égocentriques les plus antisociales.

On le voit, la réflexion neuroéthique conduit à préciser ce que l'on pourrait appeler des « circonstances neuronales atténuantes » pour certaines personnes dont le cerveau ne fonctionne pas normalement, sans toutefois, et c'est là un point crucial, appeler à éliminer la notion même de personne responsable. Il s'agit simplement d'accepter, et d'intégrer à nos sociétés, le fait que notre nature humaine s'écarte radicalement de l'image d'Épinal dualiste qu'en donne notre système judiciaire actuel. Nous sommes le résultat d'un bien étrange

³ Gutkin, B. S., Dehaene, S., & Changeux, J. P. (2006). « A neurocomputational hypothesis for nicotine addiction ». *Proc Natl Acad Sci U S A*, 103(4), 1106-1111; Redish, A. D. (2004). "Addiction as a computational process gone awry". *Science*, 306(5703), 1944-1947.

⁴ Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. G.P. Putnam. New York: NY. Traduction française : *L'erreur de Descartes*, Odile Jacob, 1997.

bricolage évolutif, un empilement de circuits autrefois utiles à la survie, dont rien ne garantit la cohérence, et qui ne cessent de tirer à hue à dia.

Les multiples bases cérébrales de nos jugements moraux

Dans la mesure où une bonne part de notre évolution s'est faite en société, une partie de notre attirail cérébral concerne les conduites sociales. Adultes, enfants, et même certains autres primates non-humains, nous disposons tous d'un sens aigu de la justice qui nous fait refuser une répartition inégalitaire des biens même s'il nous en coûte personnellement – un test devenu classique, le « jeu de l'ultimatum », le montre sans détours. Le choix altruiste de partager, pour le plus grand bien commun, et d'étendre la sympathie au delà de notre famille biologique jusqu'au groupe social tout entier, pourrait résulter d'une combinaison de ce sens de la justice et de notre compétence particulière pour la représentation des états mentaux d'autrui – ce que les psychologues appellent notre « théorie de l'esprit » des autres. Cependant, d'autres structures neuronales militent contre ces tendances sociables. L'amygdale cérébrale, un noyau bilatéral de neurones spécialisés dans l'évaluation rapide des situations émotionnelles ou dangereuses, réagit avec une hyper-sensibilité aux visages inconnus ou de mauvaise réputation. Un véritable racisme subliminal s'inscrit dans ce circuit dont les décharges rapides engendrent un état de peur avant même que nous ayons pris conscience de la présence d'un visage. Les nouvelles recherches d'Elizabeth Spelke, à l'Université Harvard, suggèrent que, dans la première année de vie de l'enfant, c'est la langue parlée, plus encore que le visage, qui induit une réaction spontanée de rejet de l'autre.

Comme le notent Katinka Evers et Bernard Baertschi, cet état contradictoire de notre cerveau, mélange de circuits altruistes et xénophobes, ne laisse guère d'espoir de légitimer telle ou telle de nos conceptions éthiques sur une base neuronale. Comme d'habitude, la

science dit ce qui est, non pas ce qui devrait être. Il revient aux sociétés humaines de se constituer des systèmes de jugement moral. Cependant, ceux-ci gagneront en pertinence et en humanité à mesure qu'ils prendront mieux en compte le portrait nuancé de notre condition humaine que dépeignent les neurosciences.

Evers ajoute un point essentiel: la personne humaine n'est entièrement déterminée ni par ses gènes ni par son histoire. Une variabilité permanente caractérise l'activité neuronale qui peut donc, d'une façon qui restera toujours imprévisible, basculer vers l'honneur ou vers l'horreur. Cette variabilité, particulièrement évidente dans les neurones corticaux, ne disparaît pas au cours du développement, mais elle est modulée par l'immense potentiel de plasticité du cerveau humain. L'homme est le seul primate qui naît avec une immaturité telle que, jusqu'à la fin de l'adolescence et au delà, des millions de synapses continuent d'être créées et éliminées chaque jour. Variabilité et plasticité lui permettent d'apprendre à se réguler lui-même sur la base de règles épigénétiques issues du groupe social. L'éducation peut être vue comme un processus de canalisation de la variabilité neuronale dans des directions jugées désirables. C'est ici, dans le domaine de la neuro-éducation, que la réflexion neuroéthique acquiert, à mon sens, son plus grand intérêt. Par l'éducation, notre espèce est la seule à pouvoir se choisir l'éthique qu'elle se donne. En élucidant les mécanismes neuronaux de nos jugements moraux, et en comprenant mieux comment ils sont modulés par l'éducation que nous avons reçue, nous gagnerons un nouveau pouvoir d'action sur nous-mêmes. Selon Baertschi, non seulement nous aurions tort d'être frileux dans nos explorations de ce nouveau pouvoir, mais d'un point de vue éthique nous aurons bientôt le *devoir* d'agir lorsque notre science permettra de modifier le cerveau d'une façon qui assure une vie heureuse ou décente à ceux que leur architecture cérébrale entraînerait fatalement sur la mauvaise pente morale.⁵

⁵ Voir également Dennett, D. (2003). *Freedom evolves*. London: Allen Lane.

Attention aux « neuro-mythes » !

Encore faut-il, pour discuter sereinement de ces questions, s'appuyer sur une base empirique solide. C'est là, peut-être, que nos deux livres d'introduction à la neuroéthique trouvent leurs limites. En effet, les résultats empiriques qu'ils décrivent restent maigres, trop brièvement esquissés et souvent peu récents. Baertschi tombe même volontiers dans un certain sensationnalisme qui confine parfois à la science-fiction. Est-il vraiment opportun de s'interroger aujourd'hui sur l'effacement des souvenirs indésirables (p. 14), la « substance compliférante » qui annihile toute volonté (p. 15), la transplantation de cellules de cerveaux humains dans le cerveau d'animaux (p. 18), « la petite protubérance de matière grise » responsable de l'expérience religieuse ou mystique (p. 98), ou le *God Helmet* (p. 99) qui permet à tout un chacun de la partager ? Le lecteur curieux cherchera en vain des références scientifiques sérieuses sur ces questions provocantes – il n'y a là rien d'autre que des recherches balbutiantes, voire des « neuro-mythes » complaisamment répandus par certains neuroscientifiques en mal de reconnaissance médiatique, que l'auteur s'honorerait de dissiper plutôt que de propager. Le livre d'Evers, beaucoup plus rigoureux sur ce plan, n'hésite pas à ramener, en une phrase assassine, certains fantasmes aux oubliettes qu'ils n'auraient jamais dû quitter. Ainsi, selon elle, « un exemple de suggestion scientifiquement inadéquate concerne la possibilité prétendument nouvelle de "lire les esprits" au moyen d'implants électroniques dans le cerveau ou de scanners ». Le neuroscientifique que je suis ne peut qu'acquiescer – mais également regretter que Katinka Evers n'en explique pas mieux les raisons à ses lecteurs. En réalité, la technique a considérablement progressé, au point que l'imagerie par résonance magnétique, doublée de puissants algorithmes de classification des images, parvient à décoder certaines représentations mentales. D'ores et déjà, par exemple, l'enregistrement de l'activité

cérébrale permet de différencier lequel de deux chiffres est maintenu en mémoire⁶, ou lequel de deux mots a été présenté.⁷ Ces résultats sont d'un profond intérêt pour la recherche fondamentale, dans la mesure où ils révèlent l'existence d'un code neural des symboles et des mots qui est distribué à tout le cortex, détectable à l'échelle du millimètre, et en partie partagé par tous les individus – des résultats étonnants qui réfutent pratiquement tout ce que nous croyions savoir sur la représentation du sens des mots dans le cerveau humain. Toutefois, ces recherches ne permettent pas de « lire dans le cerveau » d'une façon qui doive préoccuper l'apprenti neuro-éthicien. En effet, ils ne dépassent pas 60% de succès (au lieu de 50% pour un choix binaire fait au hasard), et ce dans des conditions de coopération et d'immobilisation du volontaire qui ne pourront probablement jamais être réunies à l'insu de la personne ni contre sa volonté. Ils sont par contre d'une immense portée clinique. Grâce à ces recherches, bientôt, des patients aphasiques, hémiplegiques ou « *locked-in* », c'est-à-dire atteints du syndrome d'enfermement décrit par Jean-Dominique Bauby dans *Le scaphandre et le papillon*, réapprendront à communiquer et à contrôler leur corps sur la base d'interfaces cerveau-machine.

Dans ce domaine, comme dans bien d'autres auxquels se frotte la neuroéthique, l'urgence me semble être de sortir du fantasme pour entrer, de plain-pied, dans des interrogations concrètes et réalistes sur la place des nouvelles neurosciences cognitives dans notre société. Par delà les applications cliniques qui verront prochainement le jour, on ne peut qu'espérer que notre siècle connaîtra des réformes sociétales profondes, rendues indispensables par notre compréhension grandissante de la complexe condition de l'homme

⁶ Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., & Kleinschmidt, A. (2009). "Decoding of individual number information from spatial activation patterns in human intraparietal cortex". *Current Biology*, sous presse.

⁷ Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., et al. (2008). "Predicting human brain activity associated with the meanings of nouns". *Science*, 320(5880), 1191-1195.

neuronal. Je pense tout particulièrement à nos prisons, vestiges moyenâgeux que le neuroéthicien ne peut que considérer scandaleux, car on connaît bien la quantité de malades psychiatriques et cérébraux qu'elle recueille, tout autant qu'inefficace, car on voit mal comment le cerveau pourrait s'y réformer sereinement. C'est là un parmi quelques dizaines de grands chantiers qui attendent les neuroéthiciens de demain.

Publié dans laviedesidees.fr, le 6 octobre 2009

© laviedesidees.fr