

## L'éthique des sciences cognitives a-t-elle quelque chose de particulier ?

Daniel ANDLER\*

**RÉSUMÉ.** Comme toute science, les sciences cognitives soulèvent des problèmes éthiques, notamment à cause de leurs applications potentielles dans une variété considérable de domaines. L'éthique des sciences cognitives a-t-elle pour autant quelque chose de particulier ? Nous montrerons qu'elle se distingue sur trois points. En premier lieu, les problèmes sont nombreux et difficiles, en raison du triple caractère de l'interdisciplinarité du domaine, de sa jeunesse et de sa teneur humaine. En deuxième lieu, les sciences cognitives examinent sous l'angle des sciences de la nature certains aspects centraux de la condition humaine, et créent de ce fait un conflit potentiel avec des conceptions traditionnelles. Enfin, parce que parmi ces aspects figurent l'éthique et ses principaux corrélats, tels que le libre arbitre, la responsabilité, le mérite..., une question sans équivalent dans d'autres domaines est posée : celle de savoir si les recherches sur ces questions doivent, sur le plan éthique, être poursuivies, et si c'est le cas, comment parer aux éventuelles conséquences qu'elles auraient dans la société.

*Mots-clés :* Éthique de la science, neuroéthique, éthique des sciences cognitives.

**ABSTRACT. Is there something special about the ethics of cognitive science?** Like any science, cognitive science raises ethical problems, in part due to its potential applications to a considerable variety of domains. Is there however something special about the ethics of cognitive science? I show that there is. First, the problems are particularly numerous and difficult, due to three features of the field: it is interdisciplinary, it is young, and the topics it attacks are strongly connected to human affairs. Second, among the latter, some touch on certain central aspects of the human condition, and thus potentially conflict with some traditional views. Finally, because ethics itself is targeted, alongside correlative concepts such as free will, responsibility, desert..., a question is raised which has no analog in other domains: should research on these matters be pursued, in the ethical sense, and if so, how should one face the possible consequences that it may have on society?

*Keywords:* Ethics of science, neuroethics, ethics of cognitive science.

L'éthique d'une science quelconque a pour objet le bien et le mal que cette science peut causer, soit par ses conséquences, soit dans son exercice. Quand on choisit au hasard un manuel de sciences cognitives<sup>1</sup>, et qu'on y cherche une

---

\* Département d'études cognitives, École normale supérieure, Université PSL & UMR Sciences, normes, démocratie, Sorbonne Université. daniel.andler@ens.fr

<sup>1</sup> Pour prévenir tout malentendu dès la première phrase, les sciences cognitives sont prises ici en leur sens historique, celui d'une fédération de programmes de recherche relevant de six disciplines principales (psychologie, linguistique, anthropologie, neurosciences, intelligence artificielle, philosophie), auxquelles s'est adjointe plus récemment la biologie évolutionnaire, l'IA étant quant à elle désormais plongée dans un ensemble plus vaste, la modélisation computationnelle. Elles ne se

référence à l'éthique, on ne trouve généralement rien, sauf si d'aventure il contient un chapitre *sur* l'éthique. Si, en revanche, on tape « neuro-éthique » sur un moteur de recherche, généraliste ou académique, on obtient d'innombrables références, dont certaines renvoient à l'éthique en tant qu'objet d'étude. De même, l'IA et la robotique donnent lieu actuellement à des débats largement diffusés. La Commission européenne a récemment émis des recommandations visant à rendre l'IA de demain digne de confiance, et notamment éthique<sup>2</sup>. Tout le monde (ou presque) a son avis sur le « problème du trolley »<sup>3</sup>, sur la question de l'empathie chez les robots domestiques, sur le déploiement (ou l'interdiction) de robots militaires ou sur la question des *fake news* sur les réseaux sociaux. Ce triple constat suggère trois conjectures :

(C1) Les problèmes éthiques soulevés par les sciences cognitives se ramènent à ceux qui concernent les neurosciences.

(C2) Les problèmes éthiques soulevés par les sciences cognitives sont ceux que suscite la mise à l'étude, par ces sciences, de l'éthique elle-même.

(C3) Les problèmes éthiques soulevés par les sciences cognitives se limitent à ceux qui concernent leurs applications potentielles dans les technologies numériques.

Sans souscrire à aucune de ces trois conjectures, on peut chercher à tirer de chacune son grain de vérité et combiner le résultat. L'éthique des sciences cognitives se diviserait alors en trois secteurs : la neuro-éthique, la constitution de l'éthique en objet d'enquête par les sciences cognitives, et les applications technologiques (IA, robotique, transformation digitale). Un premier objectif du présent article est de montrer qu'il n'en est rien : une partie importante de l'éthique des sciences cognitives échappe à cette tripartition, laquelle repose sur une perception erronée, car trop restrictive, de ce qu'est l'éthique d'une science quelconque.

Je proposerai donc, à titre préparatoire, une cartographie du domaine, pour m'attaquer ensuite à mon objectif principal, à savoir déterminer si les questions éthiques relatives aux sciences cognitives ont un caractère particulier, qui les distinguent des questions éthiques relatives à d'autres disciplines, et si, parmi ces questions il en est qui soient générales, se rapportant au champ entier des sciences cognitives considéré comme un tout.

---

confondent donc avec aucune de ces disciplines. Pour plus de précisions, voir Andler, 2018, ainsi que les remarques à la fin de la section 1 *infra*.

<sup>2</sup> Disponible sur [https://ec.europa.eu/commission/news/artificial-intelligence-2019-apr-08\\_fr](https://ec.europa.eu/commission/news/artificial-intelligence-2019-apr-08_fr).

<sup>3</sup> Ce problème (on dit aussi le dilemme du trolley) a pour origine une expérience de pensée proposée par une philosophe anglaise, Philippa Foot, pour éclairer la notion de double effet (Foot, 1967/1978). Il s'agit, en l'occurrence, d'une situation dans laquelle un agent est placé devant une alternative : s'il agit, il sauve cinq personnes de la mort mais en condamne une ; s'il n'agit pas, les cinq premières mourront, la cinquième sera indemne. La question a été posée dans le cadre du développement de véhicules autonomes, et fait l'objet d'une abondante littérature dans l'éthique de l'IA et de la robotique (voir par exemple Bonnefon *et al.*, 2019). Voir aussi le site <http://moralmachine.mit.edu/> qui présente des tests de comportement en situation de dilemme éthique.

## 1 – QU'ENTENDRE PAR ÉTHIQUE DES SCIENCES, ET PAR SCIENCES COGNITIVES ?

Quand on parle d'éthique dans les études sur la science, on pense généralement soit à la question des applications, soit à ce qu'on appelle souvent désormais l'intégrité scientifique, soit encore à la gouvernance de la science. Ce sont en effet trois pôles de questionnement, qui méritent d'être examinés en tant que tels, et qui sont à bon droit l'objet d'une masse considérable de travaux. Cependant, la littérature récente porte l'empreinte de nouveaux courants, qui quoique distincts, trouvent leur origine dans une double prise de conscience : celle de la gravité et de l'étendue des enjeux portés directement et indirectement par la science, celle de la nécessité de donner la parole à des catégories d'intervenants qui en étaient exclus jusque-là. Gravité des effets réels et possibles des applications ; gravité des atteintes à l'intégrité scientifique ; gravité des décisions concernant le financement des recherches et l'organisation des institutions scientifiques. Étendue des enjeux : l'acquisition de connaissances scientifiques est en elle-même porteuse de valeurs, indépendamment des applications possibles, et ne peut se prévaloir d'une quelconque neutralité axiologique.

Quant aux nouveaux intervenants, ce sont d'une part les citoyens, d'autre part les philosophes des sciences. Les premiers refusent de se voir imposer des choix qui engagent leur avenir et façonnent leur existence, et qui de surcroît impliquent des dépenses qu'ils assument via l'impôt. Quant aux philosophes des sciences, ils se sont avisés qu'ils ne pouvaient plus se cantonner à l'étude des sciences sous un angle purement épistémologique. Les raisons en sont à la fois théoriques et pratiques. Sur le premier plan, la séparation étanche entre épistémologie (qui concerne les modalités d'accès à la connaissance) et éthique (qui concerne les actions guidées par les valeurs) ne semble plus tenable : la quête de la vérité est menée dans un contexte saturé de contraintes, que les scientifiques naviguent, qu'ils en soient conscients ou non, en s'appuyant sur des valeurs. Sur le second plan, sous la pression d'enjeux tels que le changement climatique, la transition énergétique, la manipulation de l'opinion et la montée des irrationalismes..., certains philosophes des sciences se sont sentis obligés de s'engager sur le plan directement éthique et politique, revenant à une tradition progressiste perdue par la génération précédente (Howard, 2009).

Tout cela concourt à constituer un champ de réflexion très vaste, dans lequel interviennent des chercheurs de cultures très variées, impliqués dans des projets très différents. Il n'existe aujourd'hui aucune manière canonique de le cartographier, pas plus du reste qu'il n'existe un accord sur la terminologie. On parle aujourd'hui d'intégrité scientifique, de déontologie, d'éthique de la recherche, de « RCR » (*responsible conduct of research*), de « RRI » (*responsible research and innovation*), et d'éthique tout court, dans le contexte des nombreux « comités d'éthique » relevant des États ou des organismes de recherche et d'enseignement supérieur. Je m'appuierai sur une taxinomie proposée par Kenneth Pimple (Pimple, 2002), qui a le mérite d'être simple, de permettre à chacun de repérer sans peine ses sujets de prédilection, et de mettre en évidence, en creux, ce qui manque.

Pimple divise l'éthique de la recherche en six secteurs, organisés comme réponse à trois grandes questions. La première : « Est-ce vrai ? » commande le premier secteur, qu'il appelle intégrité de la recherche en lui donnant un sens particulier<sup>4</sup>. Il s'agit des pratiques qui favorisent les fins épistémiques de la science (se rapprocher de la vérité, fournir des explications et prédictions aussi exactes que possible, etc.), et de ce qui leur fait obstacle, et qui va de la méthodologie générale, par exemple en matière de recueil et de traitement statistique des données, jusqu'à la fraude et les biais involontaires.

La deuxième question, « Est-ce juste (*fair*) ? » donne lieu à quatre secteurs : la collégialité (tout ce qui touche aux interactions entre scientifiques, depuis les questions de signature et de plagiat au devoir de sincérité et aux rapports entre directeur et dirigé) ; la protection des personnes (tout ce qui touche au respect des sujets et patients) ; le respect des animaux ; et enfin, ce que Pimple appelle l'*intégrité institutionnelle*, qui concerne les rapports entre chercheurs, leurs établissements de rattachement, les agences de moyens, les organismes d'État, et qui va des conflits d'intérêt au respect des codes de bonne conduite.

La troisième question, « Est-ce sage ? » commande le dernier secteur, que l'auteur appelle « Responsabilité sociale » et qui rassemble certaines des principales questions relevant du domaine « Science et société » : priorités de la recherche, impact environnemental, responsabilité sociale, communication, éducation, savoirs interdits (en raison, sans doute, de leur dangerosité potentielle, qu'elle soit matérielle ou culturelle), etc.

On le voit, les deux premières questions permettent de regrouper l'essentiel de ce qui relève, dans le vocabulaire actuel, de l'intégrité scientifique, et la troisième une partie importante de ce qu'on appelle éthique de la recherche. Cette mise en ordre a le mérite de faire ressortir les valeurs impliquées, et de montrer (peut-être involontairement) qu'il est difficile de séparer les valeurs épistémiques et les valeurs éthiques, contrairement à ce qui a longtemps été admis (Kincaid *et al.*, 2007).

En revanche, elle laisse dans l'ombre plusieurs aspects essentiels, comme le font, à sa décharge, la quasi-totalité des grilles d'analyse proposées actuellement<sup>5</sup> :

#### 1. Deux distinctions sont occultées.

*Primo*, Pimple ne distingue pas avec netteté la responsabilité individuelle et la responsabilité collective des chercheurs. Or, nombre de décisions parmi les plus importantes sur le plan éthique résultent de processus collectifs. La notion de responsabilité collective soulève de délicats problèmes, qu'il s'agisse de responsabilité épistémique, politique, historique..., mais il n'est pas nécessaire de les résoudre ici. Ce qu'il s'agit de saisir, c'est la différence qui existe, et qui est généralement très claire, entre ce qui dépend du chercheur (respecter la déontologie et ses collègues, etc.) et ce qui n'en dépend pas. Il est des gestes

<sup>4</sup> Comme l'auteur le signale, l'expression désigne le plus souvent le respect des règles, normes et bonnes pratiques en vigueur dans la discipline concernée. Les deux acceptions ne sont évidemment pas sans rapport – essentiellement un rapport de moyens à fins, sachant qu'en la matière les moyens ne suffisent pas à garantir les fins.

<sup>5</sup> Une exception est celle de Heather Douglas, (Douglas, 2014) dont je m'inspire dans le premier point *infra*.

que seule une communauté peut accomplir ou s'abstenir de le faire, pour des raisons non pas matérielles (comme lorsqu'il s'agit de pousser à plusieurs une voiture en panne trop lourde pour être déplacée par une seule personne), ni pratiques (l'indispensable division du travail scientifique) mais constitutives : créer une société savante ou une revue, déclarer un moratoire sur des recherches potentiellement néfastes, édicter un code de bonne conduite, fixer une terminologie, déterminer le périmètre d'une spécialité ou un programme d'étude, etc. : tout cela ne peut être accompli que par un groupe de chercheurs en tant que groupe.

*Secundo*, pour ce qui est de l'individu, Pimple laisse dans l'ombre sa responsabilité générale en tant qu'être humain, plus large que ses responsabilités en tant que chercheur<sup>6</sup>. Elles sont plus particulièrement engagées aux deux extrémités de ce qu'on peut appeler la chaîne décisionnelle du chercheur, celle de l'immédiat et celle du lointain.

Premièrement, toute situation particulière appelle une décision qui doit prendre en compte le contexte dans sa totalité, pouvant impliquer davantage que la responsabilité strictement statutaire du chercheur. Il en va ainsi de situations atypiques, qu'elles soient particulièrement complexes, qu'elles fassent intervenir des aspects singuliers, non prévus ou totalement inédits, ou qu'elles donnent lieu à des conflits d'obligations. Comment se comporter relève alors en dernière analyse de notre responsabilité générale.

Deuxièmement, la responsabilité du chercheur en tant que personne s'étend aux effets à long terme de son travail, qu'il s'agisse de conséquences directes de ses propres contributions, ou de sa participation, si minime qu'elle puisse paraître au regard des milliers voire dizaines de milliers de chercheurs de sa discipline, à une entreprise collective. Ces questions sont trop vastes pour se poser dans le cadre professionnel quotidien – elles relèvent de l'éthique des sciences en un sens plus large qu'il n'est couramment entendu, et dont il est temps de parler

2. Que ce second type d'interrogation reste peu visible illustre un fait général : l'éthique scientifique, dont la taxinomie de Pimple est un assez fidèle reflet, tend à se focaliser sur les questions localisées dans le temps et dans l'espace. Ce sont celles qui typiquement se posent au chercheur inséré dans un contexte institutionnel, situé à un moment précis d'une trajectoire professionnelle, et placé devant des choix – depuis les choix courants relevant des principes de bonne conduite, jusqu'aux choix les plus graves : participer à un projet financé par l'armée ou par une entreprise privée, contribuer à un programme d'armes bactériologiques, solliciter un poste dans une université confessionnelle ou située dans un pays peu respectueux des droits humains, dénoncer un plagiat, fermer les yeux sur une violation de la déontologie dans son laboratoire, etc.

Or si l'on s'affranchit de la localité, on voit surgir de nouvelles questions, qui portent sur les fins générales de la personne, d'une part, de la société d'autre part. Pourquoi choisir une carrière scientifique ? À quelles conditions s'y engager ? Pourquoi choisir telle discipline, ou telle orientation ou école au

---

<sup>6</sup> Douglas renvoie, pour ce concept, à Hardimon, 1994.

sein d'une discipline ? Que faire lorsque la discipline à laquelle on appartient prend un tournant que l'on désapprouve ? S'agissant de l'obligation générale d'honorer les valeurs et de servir les objectifs supérieurs de « la » société, à quelle société, sous quelle description, choisir de se référer ? Quelle est la communauté dont les normes doivent primer ? Qui en décide ? Quel horizon temporel assigner à ces objectifs ? Quel prix accorder aux résultats espérés, et quels sont les moyens qu'il est raisonnable de leur accorder ?

Rien n'oblige à considérer l'éthique scientifique de manière aussi extensive : tout dépend de ce qu'on cherche à comprendre. Les problèmes classiques étudiés aujourd'hui par les comités d'éthique et les orientations RCR ou RRI nourrissent un ordre du jour substantiel et cohérent. Mais certaines questions plus générales méritent également d'être examinées, d'autant qu'elles constituent l'arrière-fond de la problématique classique, sans qu'une frontière nette ne les en sépare.

C'est le cas s'agissant des sciences cognitives, en lesquelles il faut voir – c'est le moment d'y insister – non la simple addition de ses différentes composantes, considérées comme essentiellement indépendantes, mais une entreprise de recherche organique, dans laquelle chaque programme, même lorsqu'il est essentiellement mono-disciplinaire – cas le plus fréquent –, se conçoit comme une facette de l'enquête, un *work package* du projet global. Cette précision, qui semblera sans doute superflue à beaucoup de lecteurs de la présente revue, est indispensable pour deux raisons. *Primo*, elle met en garde contre un malentendu factuel : quelles que soient leurs difficultés et leurs chances de succès, les sciences cognitives ne sont pas un simple rassemblement de recherches mono-disciplinaires. *Secundo*, elle met l'accent sur l'objectif du présent article : mettre au jour les spécificités, s'il en existe, des questions éthiques soulevées par les sciences cognitives dans leur ensemble. Ce qui peut s'entendre de deux façons : de manière distributive – caractères partagés par certaines des questions soulevées par les différents programmes de recherche, ou bien collective – spécificités de l'entreprise en tant que totalité. Nous verrons que les deux cas se présentent.

## **2 – LA SITUATION ÉTHIQUE PARTAGÉE PAR LES DIFFÉRENTES PARTIES DES SCIENCES COGNITIVES**

Qui s'attellerait à la tâche de dresser un inventaire des problèmes éthiques soulevés par les différents secteurs des sciences cognitives, pris un à un, risquerait de ne voir d'abord qu'un fouillis hétérogène. Quel rapport y a-t-il entre la question de savoir s'il faut munir les enfants nés sourds d'implants cochléaires, et celle du bien-fondé de l'étude d'éventuelles disparités cognitives entre différentes populations, ou celle du statut de théories évolutionnaires de la moralité ou de la religion ? Et quelle différence peut-il y avoir entre ces questions prises dans leur ensemble, et n'importe quel dilemme rencontré par d'autres sciences ? La même question se poserait à propos des obligations relevant de l'intégrité scientifique.

Il est exact qu'à certains égards, les sciences cognitives ont, prises une à une, des problèmes éthiques qui leur sont propres, tout comme n'importe quelle branche de la recherche, et inversement des obligations – par exemple

en matière d'intégrité scientifique – qui ne diffèrent pas de celles de toutes les branches.

Il y a cependant quelque chose qui distingue, de manière générale, les problèmes éthiques que rencontrent les sciences cognitives, et qui découle d'une triple spécificité : elles sont intrinsèquement interdisciplinaires, elles sont jeunes, et elles portent sur l'humain. Les problèmes qu'elles soulèvent sont plus pressants et plus nombreux que ceux de la plupart des disciplines. C'est d'autant plus remarquable qu'en leur sein, ces problèmes ne sont généralement pas mis en avant. Dans l'ensemble, les chercheurs ont bonne conscience, estimant qu'ils œuvrent pour une connaissance objective de l'humain, connaissance qui, outre sa valeur intrinsèque, ne peut qu'améliorer les choses pour les individus et pour la société. Elle promet en effet d'expliquer certaines imperfections de l'appareil cognitif de l'humain générique ou de certains humains, et d'imaginer des moyens de les pallier. Elle met au jour en particulier certaines erreurs fondamentales dans notre compréhension de notre vie mentale, erreurs dont nous pouvons limiter les effets délétères en en prenant conscience, etc. On y reviendra dans la dernière section.

a. Partons de *l'interdisciplinarité*, et limitons-nous pour le moment au cas des chercheurs qui l'acceptent sans arrière-pensée : ils y voient un trait constitutif du champ tel qu'ils le conçoivent, et une règle du jeu – la violer serait, à leur yeux, renoncer à contribuer, en tout cas directement, aux sciences cognitives. D'où une obligation collective : celle de délimiter les composantes concernées au sein de disciplines constituées (Andler, 2018), et d'asseoir leur légitimité sur un vocabulaire conceptuel commun. Sans doute, ce genre de décision est à l'œuvre dans la constitution et le positionnement de toute discipline et sous-discipline académique. Mais les sciences cognitives, du fait de leur jeunesse et du caractère particulièrement problématique de leur domaine, ne peuvent en déléguer la responsabilité à l'histoire ou à la tradition : elles doivent fréquemment, notamment à la faveur de décisions institutionnelles et de controverses sur le cadre à adopter, repenser collectivement leur architecture générale.

On objectera que ce sont là des questions scientifiques, ou plus largement épistémologiques, et l'on demandera quelle est leur portée proprement éthique. La réponse est double. D'une part, la réflexion sur la structuration d'un champ en émergence engage les valeurs d'ouverture d'esprit, de tolérance, de lucidité ou de modestie. D'autre part, elle renvoie aux questions générales évoquées plus haut, portant sur les fins de la recherche, les priorités de la société, les moyens mobilisés, le sens pour l'individu concerné de s'engager de cette manière, à ce moment, sur cette voie.

Il est vrai que les chercheurs ne se rapportent que rarement à ce genre de problème. Ils sont engagés dans une dynamique dont les moteurs sont l'originalité, la concurrence, l'attrait de la découverte, et dans laquelle les contingences jouent un rôle souvent décisif ; et on ne saurait exiger d'eux une parfaite vigilance éthique<sup>7</sup>. Les questions demeurent, et il n'est pas exclu

---

<sup>7</sup> L'analyse de Douglas (2014) fait droit aux écarts entre conformité réelle et conformité parfaite à l'ensemble des impératifs auxquels tout chercheur est soumis. Écarts inévitables, non seulement en

qu'elles fassent l'objet d'une attention plus grande dans la période qui s'ouvre. Cette connexion entre épistémologie et éthique de la recherche s'observe souvent, nous en verrons d'autres exemples.

Quelle que soit l'architecture du champ fixée à une phase de son développement, accepter sans arrière-pensée le principe même de l'interdisciplinarité impose au chercheur de garder à l'esprit le caractère partiel de ses recherches, non pas seulement au sens banal où tout chercheur ne fait qu'apporter une brique à l'édifice, mais au sens où la perspective qu'il adopte est nécessairement partielle. Cette pensée doit l'accompagner dans ses recherches, elle doit conditionner la manière dont il rapporte ses résultats à la communauté scientifique, et dont il accueille des perspectives différentes voire divergentes, et plus encore, s'il est possible, dont il présente ses résultats au public et aux décideurs.

Une objection à l'idée que la pluralité des approches déployées dans les sciences cognitives crée une obligation spéciale dans l'interprétation et la communication des résultats serait que d'autres disciplines connaissent un feuilletage comparable : ainsi, les sciences du vivant sont-elles présentées comme un empilement de « niveaux », depuis la molécule jusqu'à l'organisme en passant par la cellule, le tissu, l'organe... ; la physique reconnaît de même une pluralité d'échelles, tout comme les sciences sociales. En sorte que tout chercheur sait parfaitement qu'il ne décrit que l'un de ces « niveaux » et que cette description prend sa place, plus ou moins confortablement, parmi des descriptions à d'autres « niveaux » du même ordre de phénomènes. Cette objection passe à côté de la situation très particulière des sciences cognitives : les ordres de réalité appréhendés par les différentes composantes et approches en leur sein ne sont pas uniquement ordonnés selon le grain des entités fondamentales, et leurs relations ne font l'objet d'aucun consensus. Les rapports entre cerveau et organisme, entre organisme et individu, entre facultés et pensée, entre information et mécanisme... n'ont pas la belle simplicité (d'ailleurs elle-même relative) des rapports entre l'eau liquide et la molécule H<sub>2</sub>O, ou entre la cellule et le tissu, ou entre thermodynamique et mécanique statistique.

*b. La jeunesse* des sciences cognitives, combinée à la nature particulière de leur domaine, contribue à une certaine fragilité de leurs résultats : force est de constater que les théories développées au sein de l'une ou l'autre des branches établies du domaine ont une durée de vie assez courte. Scientifiques et philosophes des sciences ne s'en émeuvent pas outre mesure : toute science, même mûre, est sujette à des révisions plus ou moins radicales (Popper, 1963) – seul change le rythme. Ils savent aussi que les théories, même touchées, ne coulent généralement pas, et connaissent, plus ou moins transformées, des vies ultérieures (Kitcher 1993 ; Solomon 2001). En revanche, les non-spécialistes attendent des sciences davantage de stabilité. En sorte que tant l'honnêteté que l'intérêt bien compris exigent que les résultats actuels soient présentés non comme des acquis gravés dans le marbre, mais comme

---

raison de l'imperfection humaine, mais parce qu'on ne peut généralement pas satisfaire simultanément au plus haut degré à l'ensemble de ses obligations, éthiques et autres.



l'état présent des connaissances laissant place à un certain degré, et même un degré certain, d'incertitude.

Les sciences cognitives se retrouvent ici en bonne compagnie : la climatologie est dans une situation comparable, et même pire encore car il s'y mêle une volonté mauvaise, celle des *denialists* mus par des intérêts particuliers, par le mépris à l'égard de la science et par une mise en cause purement frauduleuse de leurs résultats<sup>8</sup>. La résistance générale aux sciences cognitives, on y reviendra dans la dernière section, existe bien mais elle est d'une autre nature et n'est pas animée par des motifs du même ordre. Ce que montre l'exemple de la climatologie est, d'abord, que le problème ne se limite pas aux sciences de l'homme<sup>9</sup> ; ensuite, que la question de la communication des résultats au public et aux responsables politiques et économiques n'est ni simplement pédagogique ni simplement épistémologique : elle est également de nature éthique, l'impératif de rigueur scientifique entrant en conflit avec l'objectif d'efficacité pragmatique, ce qui nécessite par exemple d'accepter ou de refuser de taire certaines divergences au sein de la discipline<sup>10</sup>.

c. Venons-en à ce qu'on peut appeler, pour faire bref, la *teneur humaine* des sciences cognitives. Elle alourdit considérablement leurs responsabilités, à la fois en extension et en intensité.

En extension, d'abord. La cognition est au cœur de l'action et de la pensée, donc, pour paraphraser saint Augustin, rien d'humain n'est étranger aux sciences cognitives. Toute activité humaine est potentiellement éclairée par elles, ce qui implique autant de responsabilités locales qu'on compte de secteurs d'activité, individuelle ou collective.

En intensité, ensuite. En proposant telle ou telle conception scientifique d'une activité, d'une institution, d'une pratique, d'une norme, le chercheur en sciences cognitives entre inévitablement au contact de conceptions ancrées dans le sens commun ou dans telle ou telle tradition, dont les tenants voient d'un mauvais œil ce qu'ils ressentent comme une intrusion. Il lui échoit de traiter son interlocuteur avec respect, sans céder sur ses convictions de scientifique. Un psychologue-anthropologue évolutionnaire doit s'efforcer de présenter ses conceptions de la religion sans manifester de mépris à l'égard des croyants : il touche à des convictions profondes sur lesquelles ses propres certitudes scientifiques n'ont peut-être aucune prise.

Cette responsabilité est au moins autant celle d'une collectivité que celle de chercheurs individuels. C'est l'anthropologie évolutionnaire dans son ensemble qui doit s'interroger sur la question du respect à l'égard du point de vue des croyants – peut-être estimera-t-elle qu'elle ne peut satisfaire pleinement à cette exigence sans se renier, mais au moins doit-elle tenter de le faire.

On aurait tort de penser que ce genre de problème ne se pose que dans quelques cas particulièrement délicats, impliquant des croyances, des

---

<sup>8</sup> Voir Oreskes & Conway, 2010. Il y a bien d'autres exemples de disciplines fragiles, mais la climatologie mêle de manière singulière incomplétude et robustesse.

<sup>9</sup> Dont Wittgenstein disait qu'elles restent éternellement au stade de la petite enfance. Si la climatologie doit tenir compte des facteurs sociaux, elle porte essentiellement sur des phénomènes physiques.

<sup>10</sup> Comment le GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat) fait face à ce problème est examiné dans Andler *et al.* (soumis).

convictions essentielles. Il se pose en réalité pour toutes les fonctions et facultés humaines, du langage au raisonnement, des émotions aux préférences alimentaires, esthétiques ou sexuelles. Il s'agit d'un trait distinctif que partagent sinon toutes, du moins une grande proportion des questions éthiques propres aux sciences cognitives.

Enfin, dans toutes ces questions sont engagées non seulement la responsabilité de rôle des chercheurs, mais leur responsabilité générale. Ainsi, le psychologue-anthropologue évolutionnaire assume son rôle en défendant avec énergie sa conception éliminativiste du surnaturel, mais c'est sa responsabilité générale qui lui impose de traiter le croyant avec un minimum d'égards. C'est d'autant plus important que les indications fournies par les sciences cognitives ne sont pas seulement théoriques : elles ont aussi des incidences pratiques. Le psychologue du développement qui défend une certaine théorie en rapport avec l'apprentissage a la responsabilité de déterminer si une école, un maître, un système éducatif, un haut responsable de l'éducation sont en mesure d'appliquer cette théorie sans faire encourir aux élèves ou au système un risque disproportionné, en raison soit d'incertitude sur la justesse de la théorie, soit d'effets secondaires dommageables qu'entraînerait son adoption dans le contexte considéré (Pasquinelli, 2013).

Toute application dans l'un des innombrables domaines d'activité des sociétés contemporaines implique une responsabilité de ce genre. Et une fois de plus, si les sciences cognitives sont logées à la même enseigne que d'autres disciplines, elles sont particulièrement exposées, toujours pour les mêmes raisons : leur jeunesse, qui accentue l'incertitude de leurs théories ; leur interdisciplinarité, qui implique que le point de vue d'une des composantes disciplinaires peut être sérieusement incomplet, et bien évidemment ce que j'ai appelé leur teneur humaine, qui fait que pratiquement aucun des domaines de l'activité humaine n'est à l'abri d'une intervention possible de leur part.

Remarquons enfin que ces mêmes caractéristiques, qui élèvent le seuil d'exigence éthique pour les sciences cognitives, rendent aussi plus difficile d'y satisfaire. Une discipline jeune doit donner de la voix pour espérer se faire entendre de l'extérieur, chaque spécialité doit faire de même au sein de domaine interdisciplinaire, et enfin la complexité et l'incertitude des effets possibles, à moyen et long terme, sur les activités et pratiques humaines rendent particulièrement difficile, pour le chercheur, d'évaluer ce qui relève de sa responsabilité.

### **3 – L'INTERDISCIPLINARITÉ, UN ENGAGEMENT ÉTHIQUE COLLECTIF DES SCIENCES COGNITIVES**

Après avoir examiné certains traits distinctifs partagés par les programmes de recherche abrités par les sciences cognitives, venons-en à la situation éthique des sciences cognitives en tant que système ou projet constitué. Ce sont l'interdisciplinarité et la teneur humaine qui jouent, séparément et conjointement, le rôle décisif.

Commençons par l'interdisciplinarité, en remarquant d'abord qu'elle fait en tant que telle l'objet d'une décision de la part des chercheurs. Si certains l'acceptent sans arrière-pensée comme trait constitutif et permanent du cadre théorique dans lequel leur propre travail se place et prend sens, ce choix ne va

pas de soi, et au moins trois autres attitudes sont concevables. La première, peut-être la plus fréquente, consiste à considérer l'interdisciplinarité comme provisoire, soit qu'elle conduise à une nouvelle discipline, la science cognitive, ce Graal d'une théorie pleinement intégrée de l'esprit et du cerveau, soit qu'elle se réduise à une discipline existante, à savoir les neurosciences. La deuxième y voit une simple obligation pratique sans aucune portée ni conséquence théorique. La troisième, une hypocrisie collective, obligeant chacun à faire semblant d'y souscrire pour pouvoir bénéficier d'avantages institutionnels auxquels les disciplines traditionnelles ne donnent pas accès.

Cette dernière attitude participe d'un cynisme assez répandu à l'égard de l'interdisciplinarité : elle ne serait au fond qu'un slogan propagé par les bureaucrates de la recherche et par quelques scientifiques ou philosophes fatigués en quête de « vision ». S'agissant d'une lubie intellectuelle, rien n'interdirait de la détourner à ses propres fins, ou de la traiter comme un simple bruit. Or l'histoire montre que l'interdisciplinarité dans un domaine déterminé (et non comme impératif universel) résulte d'efforts conceptuels considérables et non d'une décision arbitraire. En témoignent les débats qui ponctuent l'émergence de champs interdisciplinaires. Pour ce qui nous concerne, on peut invoquer la décision de l'Association pour la recherche cognitive de se constituer sur la base du rapport Lentin, lequel excluait les neurosciences de son périmètre<sup>11</sup> ; ou la décision (parfois implicite mais toujours discutée<sup>12</sup>) des neuroscientifiques de subordonner à leur discipline, lexicalement et conceptuellement, l'ensemble du champ. Ces épisodes de nature restrictive prennent place à côté de moments fondateurs du champ, depuis les conférences Macy (Heims, 1991, Dupuy, 2005), les rencontres de 1956 à Dartmouth et à MIT (Gardner, 1985), jusqu'au colloque de Royaumont dit « Chomsky-Piaget » (Piatelli Palmarini, 1979) ou la création de sociétés savantes telle la Cognitive Science Society ou les Sociétés pour la psychologie et la philosophie nord-américaine et européenne, à quoi il faut ajouter la création des neurosciences elles-mêmes, à partir de spécialités largement indépendantes jusque-là<sup>13</sup>. On est loin de l'inspiration subite d'un responsable de la recherche ou d'un penseur en quête de notoriété.

La conception dépréciative de l'interdisciplinarité étant ainsi écartée, deux questions restent posées. La première porte sur le caractère provisoire ou permanent de l'interdisciplinarité, la seconde sur sa teneur éthique intrinsèque, *a priori* inexistante. Elles sont en partie liées, comme nous allons le voir.

L'interdisciplinarité est généralement présentée comme une contrainte épistémologique : il existe des familles de phénomènes dont aucune discipline existante ne permet à elle seule de rendre compte. Qui cherche à comprendre une famille de ce genre – appelons-la  $\Phi$  – n'a pas d'autre choix que l'aborder

<sup>11</sup> Voir Lentin 1982, Andler 2019 ; bien entendu l'ARC, et son excellente revue, ont mis fin il y a longtemps à cet interdit, dont il faut remarquer qu'il n'était pas, dans le contexte de l'époque, une pure et simple aberration.

<sup>12</sup> Voir le récent numéro spécial d'*Intellectica*, Monier & Sarti, 2019.

<sup>13</sup> On dispose d'une étude détaillée de la manière dont les neurosciences ont émergé en tant que telles au Royaume-Uni (Abi-Rached, 2012) et en France (Debru *et al.* 2008) ; le processus a été très semblable aux États-Unis.

sous plusieurs angles, sans perspective immédiate d'une unification des connaissances ainsi obtenues. Deux hypothèses se présentent alors :

- (U)  $\Phi$  constitue (au prix d'éventuelles modifications) un domaine naturel susceptible à terme de donner lieu à une représentation scientifique unifiée.
- (H)  $\Phi$  ne constitue pas un tel domaine, car il se situe à l'intersection de plusieurs domaines naturels indépendants ; il ne donnera jamais lieu, au mieux, qu'à une représentation hybride.

Ce qu'il faut entendre par un domaine naturel et par une représentation unifiée est une question centrale pour la philosophie des sciences, qu'il n'est pas possible de présenter ici<sup>14</sup>. Ce n'est pas non plus absolument nécessaire : le lecteur peut se référer à des exemples qui lui sont familiers. Il se rendra compte à cette occasion que ces notions ne sont pas strictement dichotomiques : les domaines sont plus ou moins naturels, et les représentations plus ou moins unifiées. Les feux de forêts, les accidents de la route, les traditions alimentaires sont hybrides ; la dynamique des fluides est unifiée. La physique dans son ensemble est proche de l'unification, sans l'atteindre ; la biologie est moins unifiée qu'elle, mais plus que l'économie, etc. La démographie, la géographie sont nettement hybrides, et on voit mal comment elles pourraient cesser de l'être.

Ces deux derniers exemples nous montrent au passage qu'il faut distinguer le plan institutionnel du plan épistémologique : sur le premier démographie et géographie sont mono-disciplinaires ; sur le second elles sont hybrides, c'est-à-dire intrinsèquement interdisciplinaires. Qu'en est-il alors des sciences cognitives ? On dit souvent qu'il est vain de spéculer sur la forme qu'elles prendront dans dix, vingt ou cent ans, puisque cela dépend d'évolutions sociologiques et de découvertes scientifiques qui sont par essence imprévisibles. Dans la perspective institutionnelle, cette attitude est justifiée : si aujourd'hui on ne voit guère se confirmer les prédictions d'une inéluctable convergence, rien ne dit qu'à plus ou moins longue échéance la tendance ne se renversera pas, et qu'une dynamique unificatrice irrésistible ne se mettra pas en place.

En revanche, le plan épistémologique laisse la place pour une argumentation, dont je ne peux que rappeler quelques éléments. En faveur de l'unification sont avancées trois idées : (1) La cognition est pour l'essentiel produite par le cerveau (ou plus largement par le système nerveux), dont la compréhension, qui ne peut que progresser, nous fournira un jour une description explicative et prédictive à peu près complète des phénomènes cognitifs. (2) Plus généralement, la cognition est un phénomène naturel relevant in fine, via la biologie, des sciences physiques. Elle est donc par principe connaissable dans les termes des sciences physiques<sup>15</sup>. (3) De manière

<sup>14</sup> Voir Andler, 2011b ; Dupré, 1993 ; Galison & Stump 1996.

<sup>15</sup> Cette position reçoit dans la littérature philosophique le nom de *physicalisme*, selon lequel toutes les sciences de la nature se réduisent à la physique fondamentale. Se posent alors des questions délicates : en quoi consiste cette réduction, et quelle physique – celle de Newton, celle de la relativité et des quanta, celle de demain, est-elle investie du rôle de base possible de réduction ? Toutes ces questions

plus précise, la cognition est un ensemble de fonctions biologiques complexes de caractère adaptatif, et la biologie évolutive en fournira à terme une description s'articulant avec la description neuroscientifique, comme elle le fait pour n'importe quelle autre fonction biologique.

En faveur d'une interdisciplinarité pérenne, on invoque (1) l'absence d'indices empiriques ou d'arguments conceptuels en faveur d'une unification, qui semble devoir passer par l'élimination ou la réduction, hors de portée pour l'essentiel, des phénomènes intentionnels et normatifs ; (2) le recours constant aux concepts et méthodes de toutes les branches des sciences cognitives, dont aucune ne semble superflue ou proche de l'obsolescence ; (3) le maintien, à travers l'histoire des sciences de la nature, de l'indépendance des sciences particulières telles que la géologie ou la biochimie, en dépit du fait qu'elles soient en un sens subordonnées à la physique fondamentale (Fodor, 1974) ; (4) l'idée que la cognition n'est pas un pur produit du système nerveux, mais qu'elle n'émerge que dans le contexte du corps et/ou du groupe social, qui y contribuent à part peut-être égale, en tout cas importante, que ce soit sur le plan conceptuel ou causal.

Ce rappel beaucoup trop sommaire d'une question familière de fondements constitue le contexte dans lequel je souhaite examiner une hypothèse : c'est que le choix entre l'option U et l'option H engage des enjeux éthiques, malgré ou par-delà leur apparence purement factuelle. Ces enjeux se manifestent à plusieurs niveaux. D'abord, sur le plan des préférences méthodologiques, l'option H préserve l'autonomie relative de phénomènes centraux pour les humains, donc affirme l'importance à leur accorder dans nos tentatives pour construire une représentation scientifique du monde incluant l'humain. Ensuite, sur le plan pragmatique, elle a de meilleures chances que l'option U de fournir des explications nouvelles de nos comportements, décisions, délibérations, émotions, volitions... que nous sommes en mesure d'assimiler ; il est donc naturel de préférer voir les sciences cognitives se développer selon l'option H. Enfin, sur le plan de la responsabilité générale des individus et des communautés, on peut préférer contribuer à une discipline conforme à l'option H qu'à une discipline conforme à l'option U ; ou encore, dans une perspective objective, on peut préférer voir se développer une culture disciplinaire plus résistante aux tentations récurrentes de l'élimination et de la simplification outrancière des phénomènes humains<sup>16</sup>.

Il ne manque pas de philosophes pour se gausser de ce genre de considérations, qu'ils estiment tributaires d'une vision théologique faisant de la nature, du monde tel qu'il est, la demeure des humains<sup>17</sup>. Cette objection est à

---

sont controversées. Voir par exemple Papineau, 1993 ; Stoljar, 2017, et un bref résumé de la discussion dans Andler, 2016, pp. 51-52.

<sup>16</sup> Je suis redevable de cette idée à Mehdi Khamassi.

<sup>17</sup> Joseph Rouse offre ainsi une description éloquente d'un certain type de philosophe naturaliste intransigeant : « [pour lui], les êtres humains vivent dans un monde qui est indifférent voire hostile à nos intérêts, à nos désirs, à nos valeurs, aux perspectives que nous privilégions, et les sciences nous procurent notre principal accès à ce monde à la périphérie duquel nous nous situons et auquel c'est à nous de nous accommoder. [...] [Ce n'est qu'en nous tenant résolument à] une auto-compréhension purement scientifique que nous nous affranchirons des résidus de nos illusions autoglorificatrices. » (Rouse, 2008, cité dans Andler, 2016, p. 395).

son tour de nature éthique, et doublement. *Primo*, elle mise sur une stratégie d'élimination de processus que la plupart des gens estiment essentiels à leur existence d'êtres humains, et elle le fait – c'est l'aspect éthique – en l'absence d'argument scientifique décisif. *Secundo*, elle adresse aux partisans de H le reproche de s'accrocher à un mythe, dans le seul but de sauver des préférences subjectives et irrationnelles. Ils sont coupables de *wishful thinking*, et compromettent ainsi l'épanouissement humain, qui (s'il a encore un sens) ne peut se fonder que sur une appréhension correcte de la réalité : c'est un manquement à l'éthique qui leur est imputé.

Il ne m'appartient pas de trancher le débat : mon objectif est d'en mettre en évidence le caractère proprement éthique. L'interdisciplinarité ne serait pas seulement un fait, ni seulement un pari méthodologique, mais une prise de position éthique, et relèverait ainsi de l'éthique des sciences cognitives au sens collectif et non distributif.

#### 4 – LA PORTÉE ANTHROPOLOGIQUE DES SCIENCES COGNITIVES

De toute science humaine (dans le sens large qui inclut toutes les sciences de l'homme et de la société) on doit se demander si, en tentant d'éclairer tel ou tel aspect de l'humanité, elle ne court pas le risque d'y porter atteinte. Qu'elle mette au jour tel mécanisme, telle régularité, cela ne peut qu'ébranler certaines de nos conceptions spontanées, quelles qu'en soient l'origine, concernant par exemple le libre arbitre en général, ou l'autonomie dont nous jouissons dans certains de nos choix, ou l'individualité, ou la rationalité ; et qu'en modifiant de la sorte notre conception de l'humanité en nous, nous mettons en péril cette humanité.

Dans ce débat vieux comme les sciences humaines elles-mêmes, plusieurs ripostes ont été avancées, qui sont autant de parades à l'imputation de déshumanisation, de mécanisation de l'humain, de déterminisme, de réification, etc. La première est d'assigner à ces sciences une fonction descriptive et éventuellement herméneutique : elles doivent ne viser à rien d'autre que décrire la dynamique des significations que les individus et les sociétés attribuent à leurs actions et à leurs discours ; elles peuvent aller jusqu'à contribuer, dans un processus en spirale, à la mise au jour voire à la création des significations – le repérage par les moyens scientifiques de *patterns* ou d'invariants dans les transactions de significations, pouvant conduire à une prise de conscience et à une intériorisation par les acteurs. Ce genre d'enquête laisse ainsi intouchées les notions cardinales sur lesquelles se fondent les conceptions traditionnelles de l'individu et de la condition humaine, et les lois que certaines de leurs approches cherchent à mettre en évidence ne contraignent l'individu que par l'intermédiaire de ses croyances, valeurs, préférences, etc., prises en tant que telles. La deuxième parade consiste à protéger l'individu par la considération que toute loi ou régularité portant sur les humains est statistique : ainsi, appartenir à un ensemble qui manifeste telle préférence, même massivement, ne contraint aucun de ses membres en particulier à s'y conformer ; être de disposition colérique n'oblige pas à se mettre en colère en toute occasion, etc. La troisième, qui est une version radicale et philosophique ou théologique de la précédente, recourt à l'idée du libre arbitre érigé en évidence ou en principe intangible, en vertu duquel

aucune contrainte, d'aucun ordre que ce soit, ne détermine les pensées, intentions, actions de l'individu à un moment donné.

Ces parades, même si on les accepte, n'exonèrent pas les sciences humaines de toute question éthique, tant s'en faut. Mais elles sont inopérantes dans le cas des sciences cognitives, qui visent à être, si c'est possible et autant que possible, une science *naturelle* de l'humain pour lesquelles les significations, l'action, le libre arbitre, le sujet ou la personne-même ne sont pas des concepts primitifs, mais des phénomènes (réels ou illusoire) dont il faut rendre compte en s'en tenant, si possible et autant que possible, aux entités et mécanismes naturels. Surgit ainsi le second problème éthique spécifique aux sciences cognitives dans leur ensemble. Problème qui se dédouble : le premier aspect concerne les effets que du seul fait de leur existence et de leurs progrès ces sciences peuvent ou risquent d'avoir sur la conception que les humains se font de leur propre nature ; le second porte sur les actions, autorisées ou encouragées par les nouvelles théories scientifiques, visant à modifier cette nature.

Il s'agit là d'un cas particulier, quoique particulièrement aigu, d'un problème général : celui de savoir si toute connaissance est bonne à avoir ; si, lorsqu'elle l'est ou à supposer qu'elle le soit, il faut l'acquérir quel qu'en soit le prix ; et si, enfin, le risque que l'on prend de ne pas voir le programme de recherche dans lequel on s'engage déboucher sur une connaissance authentique est acceptable. Le débat entre celui qui répond oui à ces questions et celui qui répond non comporte une dimension éthique irréductible.

Certains débats de ce genre sont en pratique dénués d'intérêt, car ils sont tranchés avec netteté et de la même manière par tout le monde ou presque : qui songerait à contester sérieusement la légitimité de la plupart des recherches en biologie humaine, en sciences de la Terre ou en météorologie ? Ou parce qu'ils sont sans espoir d'aboutir, l'expérience montrant qu'on n'est toujours pas parvenu, malgré de méritoires efforts, à dégager un consensus et que l'enjeu est relativement limité (qu'on songe à maintes spécialités académiques), ou encore du fait d'une incertitude fondamentale (dans le cas, par exemple, de l'étude d'un problème dont rien n'indique qu'on puisse le résoudre, ou d'une recherche dont il est impossible de prévoir, si approximativement que ce soit, les éventuelles conséquences). Mais posées à propos des sciences cognitives, ces questions sont au contraire de la plus haute importance. Entreprendre de naturaliser l'humain aussi complètement que possible, est-ce un programme qui a des chances raisonnables d'aboutir ? s'il aboutit, le savoir acquis est-il une bonne chose ? s'il n'aboutit pas, ou pas complètement, aura-t-il néanmoins, ou par ce fait même, contribué à nous éclairer ? et enfin, dans ces différentes hypothèses, quelles conséquences peut-on attendre d'un tel savoir quant à l'amélioration ou la détérioration de notre condition, ou du sort de certains d'entre nous ?

Face à ces questions, deux attitudes sont possibles. La première est d'y voir autant de raisons de rejeter les sciences cognitives. Elles sont de fait une source importante, sinon la plus importante, de l'opposition aux sciences cognitives. En prendre pleinement conscience est indispensable pour qui veut désarmer

cette opposition ou du moins engager avec elle un dialogue constructif<sup>18</sup>. La seconde attitude est d'y voir non un motif de se distancier des sciences cognitives, mais d'en examiner la portée anthropologique et les effets possibles sans en sous-estimer par myopie ou désinvolture la gravité.

La neuroéthique se donne pour objet de réfléchir à ces questions, principalement sous l'angle pratique. Elles concernent toutes ce que nous conviendrons d'appeler, pour les besoins de l'exposé, les « fonctions sensibles », à savoir celles qui se rattachent à la notion traditionnelle de sujet, telles que le soi (ou la personne), le caractère ou la personnalité, le libre arbitre, la conscience en ses différentes acceptions, y compris la conscience morale, l'éthique (comprise comme régime de normativité et ensemble de normes dont se saisit la conscience morale), la responsabilité, l'estime de soi, le ressentiment et autres émotions morales, le blâme et la louange... Cette énumération ne prétend pas être fondée sur une analyse rigoureuse, et le terme de « fonction » ne doit pas être interprété trop littéralement. La notion de fonction sensible permet néanmoins de caractériser de manière concise les deux grandes branches de la neuroéthique.

La première porte sur les représentations et conceptions relatives aux fonctions sensibles, la seconde sur les interventions matérielles effectuées soit directement sur le cerveau, soit indirectement par le biais du comportement, en vue de modifier d'une manière ou d'une autre telle ou telle fonction sensible. Des exemples de la première catégorie sont la question de savoir s'il est légitime de prouver que le libre arbitre est une illusion, celle de savoir s'il est légitime de mettre au point des techniques « cérébroscopiques » permettant de détecter des préférences ou des pensées d'un individu qu'il dissimule ou peut même ignorer, celle des conséquences possibles de telles découvertes. De la seconde catégorie relève la question de savoir s'il est légitime d'administrer, par exemple à un suspect ou à un ennemi, un « sérum de vérité »<sup>19</sup>, à supposer qu'on ait au préalable mis au point un tel sérum ; ou de pousser un individu, en stimulant telle zone de son cerveau, à faire un choix qu'il ne ferait pas spontanément<sup>20</sup>, ou à sortir d'un état de deuil qu'il ne semble pas vouloir abandonner de lui-même ; ou encore, de lui conférer des aptitudes cognitives ou physiques exceptionnelles en lui administrant un dopant ou en stimulant son cerveau...

À ces deux catégories correspondent les deux branches de la neuroéthique (Farah, 2005 ; Levy, 2007 ; Farah 2010). Disons d'abord pourquoi ce terme

<sup>18</sup> Voir Andler (2010, 2011a, 2016), et pour un large aperçu des résistances de certains scientifiques aux neurosciences et, directement ou à travers elles, aux sciences cognitives, voir Johnson & Erneling, 1997 ; Erneling & Johnson, 2005 ; Choudhury & Slaby, 2012.

<sup>19</sup> Ce que l'on pensait être la scopolamine au cours de la Deuxième Guerre mondiale : voir, parmi bien des exemples réels ou fictionnels, le roman et le film *Les Canons de Navarone*.

<sup>20</sup> On sait que par simulation cérébrale profonde, on peut induire des comportements ou des préférences (par exemple musicales ou sexuelles) étrangères au sujet non stimulé, et cela de manière réversible. Signalons que la technique, attribuée généralement au neurochirurgien Alim-Louis Benabid en collaboration avec le neurologue Pierre Pollak et datée des années 1980, a été pratiquée dès les années 1950 par un psychiatre américain largement oublié, Robert Heath (1915-1999). Le flou scientifique et éthique dans lequel opérait Heath a été avec raison dénoncé, et peut servir aujourd'hui en quelque sorte d'exercice préparatoire ou d'avertissement en matière de neuroéthique. Voir. Frank, 2018.



convient et ne convient pas pour l'ensemble des questions qui nous intéressent. Il convient dans la mesure où beaucoup de ces questions se posent dans le cadre théorique et pratique des neurosciences *stricto sensu*, mais aussi du fait que, indépendamment des méthodes employées pour acquérir une connaissance ou intervenir sur un individu, la connaissance ou l'intervention impliquent en dernière instance un processus neuronal, qu'il soit ou non bien identifié. Il ne convient cependant pas pour au moins deux raisons. La première est que les problèmes portant sur les concepts relatifs aux fonctions sensibles ne relèvent ni exclusivement ni même principalement des compétences neuroscientifiques : ce sont des questions à fort contenu philosophique pour lesquelles, outre les philosophes eux-mêmes, les psychologues, anthropologues, sociologues, éthologues voire économistes sont au moins aussi bien placés que les neuroscientifiques. Pour prendre le cas le plus emblématique, celui de l'éthique elle-même, en tant que fonction sensible, les anthropologues et psychologues évolutionnaires, les éthologues et les économistes conduisent des recherches naturalistes aussi éclairantes que celles des neuroscientifiques, et ne devant rien à ces dernières. Aucun résultat des neurosciences ne peut à lui seul apporter de lumière sur ces questions, et cela pour des raisons logiques sans rapport avec l'état d'avancement de la science du cerveau ou la précision de la neuroimagerie ; nous y reviendrons. La seconde raison pour laquelle le terme neuroéthique est trompeur est que pour une bonne part, les fonctions sensibles émergent à l'interface de l'individu et de son environnement : elles ne sont pas « implémentées » dans le cerveau<sup>21</sup>.

La distinction entre concevoir ou comprendre et intervenir est analytiquement utile, mais les deux questionnements sont étroitement mêlés. Selon la conception que l'on a du libre arbitre, on sera plus ou moins réticent à infléchir les choix du sujet en intervenant directement sur ses processus neurocognitifs, sans passer par une prise de conscience de sa part ; ou à intervenir d'une manière qui puisse limiter son autonomie de manière générale. Selon l'idée qu'on se fait du lien entre personne et caractère, on sera plus ou moins enclin à modifier le caractère d'un patient ou d'un criminel, ou à accepter une modification comme effet secondaire d'un acte thérapeutique critique, ou encore à modifier notre jugement sur son comportement, jugement qui peut entraîner l'action concrète d'une punition ou d'une condamnation. Si l'on ne croit pas à la conscience, on peut adopter une attitude à l'égard des patients dans diverses formes de coma<sup>22</sup> qui diffère de celui qui y croit, et si l'on y croit, la conception qu'on s'en fait influera sur les décisions de fin de vie (Kahane & Savulescu, 2009 ; Levy, 2014 ; Fins & Schiff, 2017). Inversement, des interventions, à des fins médicales ou autres, sur le système nerveux central ne peuvent pas ne pas avoir, et ont effectivement un retentissement parfois considérable sur les conceptions tant scientifiques que profanes des fonctions sensibles.

---

<sup>21</sup> Je laisse de côté la question de savoir dans quelle mesure on doit suivre les doctrines externalistes qui appliquent cette idée à la cognition dans son ensemble.

<sup>22</sup> On distingue les patients en état végétatif persistant (PVS) et en état de conscience minimale (MCS), cette seconde condition donnant lieu à des distinctions supplémentaires fort débattues (Laureys *et al.*, 2004 ; Bernat & Rottenberg, 2007 ; Bayne *et al.*, 2016).

Remarquons au passage que l'expérience clinique ou la psychologie expérimentale peuvent également susciter des remises en cause de ce genre. Que l'on songe à la vision aveugle par exemple, ou au syndrome de Capgras, dont la découverte ne doit rien aux neurosciences (qui en revanche en fournissent parfois une explication). Les travaux sur la langue des signes, modifient assez profondément nos conceptions profanes du langage, une fonction qui sans être sensible au sens stipulé ici est certainement centrale dans la cognition humaine (voir Sandler & Lillo-Martin, 2006).

Les questions éthiques de la deuxième catégorie définie plus haut sont les plus brûlantes. Ce sont celles que soulèvent les interventions directes ou indirectes sur le cerveau ou les décisions fondées sur des mesures de l'activité neuronale, et elles exigent l'examen le plus attentif car elles portent sur des pratiques actuelles ou qui le deviendront à court terme. Elles ressemblent, dans leur structure générale, à maintes questions d'éthique appliquée, dans le domaine médical, juridique, social ou économique, tout en s'en distinguant par le fait qu'en faisant de l'esprit humain un objet de manipulation et d'observation directes, sans respecter la barrière du « for intérieur », c'est-à-dire sans passer par le guichet de la décision consciente du sujet, elles risquent de mettre en péril son autonomie.

Mais les questions les plus singulières sont celles de la première catégorie, celles qui concernent les représentations des fonctions sensibles. Parmi elles, trois interrogations constituent une sorte de nœud central, dont toutes les autres dépendent. La *première* porte sur l'approche naturaliste de l'éthique, qui vise à en faire une disposition mentale produite au cours de la phylogénie, étroitement liée aux émotions et réalisée dans le cerveau, mettant en jeu notamment le circuit de récompense. Est-il éthique de chercher à expliquer l'éthique ainsi ? La *deuxième* interrogation porte sur les effets, dans le public et la société en général, de cette tentative, menée par la science, pour expliquer l'éthique et ce qui lui est lié, le sentiment moral et la responsabilité. Quels que soient les résultats de cette tentative, par sa simple existence, appuyée sur l'autorité de la science, elle ébranle l'éthique en tant qu'instance normative intangible (quelles que soient ses spécifications dans la moralité et les pratiques). Il n'est pas nécessaire de penser comme un philosophe pour saisir intuitivement que si l'éthique n'est qu'une disposition psychologique, aucun principe supérieur n'oblige à s'y conformer – il existe d'ailleurs une longue tradition de nihilisme éthique ou d'amoralisme qui a préparé l'opinion publique et la culture à tirer les conséquences de l'élimination d'une instance normative transcendante, qu'elle soit théologique ou laïque. La question éthique qui se pose alors est de savoir s'il est éthiquement justifié de poursuivre ces recherches, susceptibles d'apporter un soutien scientifique au nihilisme éthique, et si c'est le cas, s'il ne faut pas, au sens éthique du terme, les accompagner d'un effort pédagogique pour montrer qu'elles ne mettent pas nécessairement en cause la responsabilité, la rétribution ni la pratique de la louange et du blâme. Cet effort est d'autant plus difficile à mener qu'il ne manque pas de philosophes compétents pour estimer que l'éthique au sens traditionnel et les pratiques qui lui sont associées n'ont effectivement plus de

sens<sup>23</sup>. D'où, *troisième* brin du nœud, la question d'une nouvelle éthique, remplaçant l'ancienne, qui non seulement concilie normativité et ancrage naturel – un défi intellectuel de taille, mais qui puisse être comprise et acceptée sans formation spécialisée – une solution philosophique inintelligible par le commun des mortels ne ferait pas l'affaire.

De cette troisième interrogation on dira peut-être que c'est en réalité un aspect d'un problème fondamental, celui d'une éthique pour notre temps. On dira que c'est là un problème pour les philosophes<sup>24</sup>, non pour les chercheurs en sciences cognitives ; ou encore, au niveau collectif, un problème pour la société, et non pour la communauté des sciences cognitives. Ce serait, d'abord, dessiner une frontière arbitraire et contre-productive entre philosophie, ou du moins une branche de la philosophie, et sciences cognitives ; ce serait ensuite oublier que la responsabilité, individuelle et collective, des chercheurs, ne se limite pas à leur rôle, et qu'elle relève aussi de ce qu'on a appelé leur responsabilité générale.

Examiner l'éthique et le libre arbitre sous la loupe des neurosciences, ou celle d'une science quelconque, ne suffit jamais à fournir une réponse unique et claire aux questions fondamentales, et cela, comme il a été dit, pour des raisons logiques au sens large. C'est que l'enquête scientifique ne prend sens que dans un cadre conceptuel qui, sans être fixé une fois pour toute, constitue néanmoins à tout moment le point archimédien sur lequel s'appuie la pensée. Cela est vrai des concepts fondamentaux par lesquels nous tentons de saisir le monde physique (voir par exemple Sklar, 2000) ; ce l'est *a fortiori* des concepts fondamentaux par lesquels nous saisissons l'existence humaine.

C'est la raison la plus profonde pour laquelle la recherche des bases naturelles des fonctions sensibles doit être à la fois encouragée et placée dans un cadre épistémologique adéquat. Elle doit être encouragée, parce qu'elle peut nous aider à y voir plus clair dans les ressorts cachés de notre dynamique mentale, de nos émotions, de nos comportements, au cours d'existences soit à peu près normales, soit s'écartant des normes. Elle doit être correctement située sur la carte des connaissances, sous peine d'engendrer des incompréhensions et des erreurs aussi dommageables que celles qu'elle prétend remplacer. À cette fin, l'interdisciplinarité est la meilleure garantie, et elle doit donc être défendue. Le « doit » en question est de nature éthique, et ce qui vient d'être dit exprime une position éthique, celle que je défends (Anderl, 2016) ; d'autres positions sont défendables et défendues, mais elles se placent également,

---

<sup>23</sup> Un représentant très en vue de la conception incompatibiliste, selon laquelle le libre arbitre est une illusion qu'il s'agit de reconnaître une bonne fois et d'expliquer est Gregg Caruso : voir Caruso, 2012 (voir aussi Smilanski, 2000 ; Wegner, 2002). Un débat récent l'a opposé à Daniel Dennett (Caruso & Dennett, 2018). Tout en ayant une conception résolument naturaliste et en particulier darwinienne de l'esprit humain, Dennett défend une position compatibiliste (voir Dennett 1984/2015), qui fait droit au libre arbitre dans un monde entièrement naturel. Ce que s'efforce de montrer le philosophe Derk Pereboom, incompatibiliste radical, c'est qu'il est possible, dans un monde déterministe excluant le libre arbitre, de donner sens aux conceptions et pratiques du blâme et de la punition, de l'éloge et de l'admiration (Pereboom, 2014).

<sup>24</sup> Comme me le rappelle opportunément un reviewer anonyme, que je remercie, ce problème est central dans la pensée et l'œuvre du philosophe allemand Hans Jonas (1903-1993) : voir Jonas, 1979.

qu'elles en soient conscientes ou non, sur le plan éthique.



Les problèmes éthiques des sciences cognitives se présentent ainsi sous la forme d'une pyramide. Le problème qui vient d'être présenté en constitue le sommet. Il leur appartient en propre au sens où l'on n'en voit pas d'équivalent dans d'autres sciences, et où il présente la particularité d'impliquer un cercle réflexif, puisqu'on s'interroge sur l'éthique des recherches sur l'éthique – réflexivité qu'on retrouve, de manière moins flagrante, à propos du libre arbitre : c'est en apparence le libre arbitre du chercheur qui l'oriente vers l'étude, et l'élimination éventuelle, du libre arbitre. Au niveau inférieur se situent un ensemble de problèmes qui ont pour spécificité commune que les sciences cognitives soumettent à examen des pratiques considérées comme constitutives de la condition humaine, et entrent de ce fait potentiellement en conflit avec des conceptions solidement ancrées dans notre auto-compréhension. À la base de la pyramide se trouvent les problèmes éthiques qu'on peut dire ordinaires : ils sont ceux de toute science, même s'ils se distinguent par leur nombre et leur intensité, en raison de certaines particularités des sciences cognitives<sup>25</sup>.

#### RÉFÉRENCES

- Abi-Rached, J.M. (2012). From Brain to Neuro: The Brain Research Association and the Making of British Neuroscience, 1965-1996. *Journal of the History of the Neurosciences*, 21/2, 189-213.
- Andler, D. (2010). Is naturalism the unsurpassable philosophy for the sciences of man in the 21<sup>st</sup> century?. In F. Stadler, S. Hartmann, D. Dieks, W.J. Gonzalez, T. Uebel & M. Weber (eds.), *The Present Situation in the Philosophy of Science*. Berlin, Springer, pp. 283-303.
- Andler, D. (2011a). Le naturalisme est-il l'horizon scientifique des sciences sociales ?. In Th. Martin (éd.), *Les sciences humaines sont-elles des sciences ?* Paris, Vuibert, 2011, pp. 15-34; repris in Y. Brailowsky & H. Inglebert (éd.), *1970-2010 : les sciences de l'Homme en débat*. Nanterre, Presses universitaires de Paris Ouest, 2013, pp. 331-356.
- Andler, D. (2011b). Unity without myths. In J. Symons, O. Pombo & J.M. Torres (eds.), *Otto Neurath and the Unity of Science*. Dordrecht, Springer, pp. 129-144.
- Andler, D. (2016). *La silhouette de l'humain. Quelle place pour le naturalisme dans le monde d'aujourd'hui ?* Paris, Gallimard, coll. Essais, 574 p.
- Andler, D. (2018). Les sciences cognitives : un tour d'horizon. In T. Collins, D. Andler & C. Tallon-Baudry (éds.), *La cognition, du neurone à la culture*. Paris, Gallimard, pp. 15-70.
- Andler, D. (2019). La pente du naturalisme. In C. Monier & A. Sarti (dir.) *Les Neurosciences au sein des sciences de la cognition entre neuroenthousiasme et neuroscepticisme*. *Intellectica*, n° 69, 149-165
- Andler, D., Barberousse, A., Drouet, I. & Jebeile, J. (soumis). IPCC Assessment Reports as an epistemological puzzle.

---

<sup>25</sup> Je remercie les deux reviewers anonymes, ainsi que Mehdi Khamassi, pour leurs très utiles remarques et questions.

- Bayne, T., Hohwy, J., & Owen, A.M. (2016). Are there levels of consciousness? *Trends in Cognitive Science*, 20/6, 405-413.
- Bernat, J.L. & Rottenberg, D.A. (2007). Conscious awareness in VS and MCS: The borderlands of neurology. *Neurology*, 68, 885-886.
- Bonnefont, J.-F., Shariff, A. & Rahwan, I. (2019). The Trolley, the Bull Bar, and Why Engineers Should Care About the Ethics of Autonomous Cars. *Proceedings of the IEEE*, vol. 107, n°3, 502-504.
- Caruso, G.D. (2012). *Free Will and Consciousness. A Determinist Account of the Illusion of Free Will*. Lanham, MD, Lexington Books.
- Caruso, G.D. & Dennett, D.C. (2018). Just deserts. An Aeon Essay, accessible à <https://aeon.co>
- Choudhury, S. & Slaby, J. (eds.) (2012). *Critical Neuroscience. A Handbook of the Social and Cultural Contexts of Neuroscience*. Maiden, MA, Wiley-Blackwell.
- Debru, C., Barbara, J.-G. & Chérici, C. (2008). *L'essor des neurosciences, France 1945-1975*. Paris, Hermann.
- Dennett, D.C. (1984/2015). *Elbow Room. The Varieties of Free Will Worth Wanting*. Cambridge, MA, MIT Press.
- Douglas, H. (2014). The moral terrain of science. *Erkenntnis*, vol. 79, n°5, 961-979.
- Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA, Harvard University Press.
- Dupuy, J.-P. (2005). *Aux origines des sciences cognitives*. Paris, La Découverte.
- Erneling, C.E. & Johnson, D.M. (eds.) (2005). *The Mind as a scientific object. Between brain and culture*. New York, Oxford University Press.
- Farah, M. (2005). Neuroethics: the practical and the philosophical. *Trends in Cognitive Sciences*, 9/1, 34-40.
- Farah, M. (ed.) (2010). *Neuroethics: an Introduction with Readings*. Cambridge, MA, MIT Press.
- Fins, J.J. & Schiff, N.D. (2017). Differences that make a difference in disorders of consciousness. *AJOB Neuroscience*, 8/3, 131-134.
- Fodor, J. (1974). Special sciences, or the disunity of science as a working hypothesis. *Synthese* n° 28; in J. Fodor, *Representations*, Cambridge, MA, MIT Press, 1981.
- Foot, P. [1967]. The problem of abortion and the doctrine of the double effect, *The Oxford Review*, 5, 5-15. Reprinted in « *Virtues and Vices* ». Oxford, Basil Blackwell, 1978.
- Frank, L. (2018). *The Pleasure Shock: The Rise of Deep Brain Stimulation and Its Forgotten Inventor*. New York, Dutton.
- Galison, P. & Stump, D. (éds.) (1996). *The Disunity of Science*. Stanford, CA, Stanford University Press.
- Gardner, H. [1985]. *The Mind's New Science*. New York, Basic Books, éd. augm., 1987).
- Hardimon, M. (1994). Role obligations. *The Journal of Philosophy*, vol. XCI, n°7, 333-363.
- Heims, S. (1991). *Constructing a Social Science for Postwar America (1946-1953). The Cybernetics Group*. Cambridge, MA, MIT Press.
- Howard, D. (2009). Better Red than Dead- Putting an End to the Social Irrelevance of Postwar Philosophy of Science. *Science and Education*, 8/2, 199-220.
- Johnson D.M. & Erneling, C.E. (eds.) (1997). *The Future of the Cognitive Revolution*. New York, Oxford University Press.
- Jonas, H. (1979). *Das Prinzip Verantwortung: Versuch einer Ethik für die technologische Zivilisation*. Frankfurt am Main, Insel-Verlag. Trad. fr. *Le principe responsabilité*. Paris, Flammarion, 1998.
- Kahane, G. & Savulescu, J. (2009). Brain damage and the moral significance of consciousness. *Journal of Medicine and Philosophy*, vol. 34, 6-26

- Kitcher, P. (1993). *The Advancement of Science*. New York & Oxford, Oxford University Press.
- Kincaid H., Dupré J., Wylie A. (éds). (2007). *Value-free science? Ideals and illusions*. New York, Oxford University Press.
- Laureys, S., Owen, A.M. & Schiff, N.D. (2004). Brain function in coma, vegetative state, and related disorders. *Lancet Neurology*, vol. 3, 537-546.
- Lentin, A. (1982), Rapport sur la recherche cognitive (Rapport introductif au Colloque « Domaines et objectifs de la recherche cognitive ». Pont-à-Mousson : Association pour la recherche cognitive). *Intellectica*, n° 3, 3-40.
- Levy, N. (2014). The Value of Consciousness. *Journal of Consciousness Studies*, 21/1-2, 127-138.
- Merton, R.K. (1942). A note on science and democracy. *Journal of Legal and Political Sociology*, 1, 115-126. Texte repris ultérieurement sous le titre plus connu « The Normative Structure of Science ».
- Monier C. & Sarti A. (dir.) (2019), Les Neurosciences au sein des sciences de la cognition entre neuroenthousiasme et neuroscepticisme. *Intellectica*, n° 69
- Oreskes, N. & Conway, E.M. (2010). *Merchants of Doubt, How a Handful of Scientists Obscured the Truth from Tobacco Smoke to Global Warming*. New York, Bloomsbury Press. Trad. fr. *Les Marchands de doute*. Paris, Le Pommier, 2012.
- Papineau, D. (1993). *Philosophical Naturalism*. Oxford, Blackwell.
- Pasquinelli, E. (2013), Slippery slopes. Some considerations for favoring a good marriage between education and the science of the mind-brain-behavior, and forestalling the risks. *Trends in Neuroscience and Education*, 2/3-4, 111-121
- Piatelli Palmarini, M. (1979). *Théories du langage, théories de l'apprentissage : le débat entre Jean Piaget et Noam Chomsky*. Paris, Seuil.
- Pereboom, D. (2014). *Free will, Agency, and Meaning of Life*. New York, Oxford University Press.
- Pimple, K.D. (2002). Six domains of research ethics. A heuristic framework for the responsible conduct of research. *Science and Engineering Ethics*, 8, 191-205.
- Popper, K.R. (1963). *Conjectures and refutations. The Growth of Scientific Knowledge*, London. Routledge & Kegan Paul. Trad. Fr.. M.I. & M.B. Delaunay, *Conjectures et réfutations: la croissance du savoir scientifique*. Paris, Payot, 1985.
- Rouse, J. (2008). Review of Georg Gasser (ed.), "How successful is naturalism?". *Notre-Dame Philosophical Reviews*. Accessible à <https://ndpr.nd.edu/news/how-successful-is-naturalism/>
- Sandler, W. & Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals*. Cambridge, Cambridge University Press.
- Sklar, L. (2000). *Theory and Truth: Philosophical Critique within Foundational Science*. New York, Oxford University Press.
- Smilanski, S. (2000). *Free Will and Illusion*. Oxford, The Clarendon Press.
- Solomon. M. (2001). *Social Empiricism*, Cambridge, MA, MIT Press.
- Stoljar, D. (2017). Physicalism. In E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/win2017/entries/physicalism/>
- Wegner, D. (2002). *The Illusion of Conscious Will*. Cambridge, MA, MIT Press.