Intellectica, 2019/1, **70**, pp. 7-39

Éthique et sciences cognitives

Mehdi KHAMASSI[*], Raja CHATILA[#] & Alain MILLE[◊]

RÉSUMÉ. Ce volume présente un ensemble de contributions visant à discuter de questions éthiques liées au champ des sciences cognitives. Il s'agit à la fois d'aborder des questions éthiques qui sembleraient spécifiques des sciences cognitives, et des questions éthiques générales auxquelles les sciences cognitives pourraient apporter un éclairage particulier. Sont soulevés notamment les problèmes liés à un certain nombre d'applications des sciences cognitives, comme celles des neurosciences au cadre juridique, de l'intelligence artificielle aux systèmes artificiels autonomes ou au traitement des données personnelles, ou encore des sciences comportementales aux politiques publiques. Parmi les questions générales, il s'agit de voir dans quelle mesure les sciences cognitives, pour qui la cognition humaine est un des objets d'étude centraux, incluant la manière dont nous prenons des décisions et nous nous posons des dilemmes moraux, peut apporter une vision nouvelle ou éventuellement une caractérisation plus précise des processus cognitifs sous-jacents. Notre vision de ce qu'est l'éthique au XXI[e] siècle doit-elle être révisée au regard des nouvelles connaissances en sciences cognitives ? L'interdisciplinarité des points de vue esquissés dans ce numéro contribue à souligner que loin de chercher à remplacer les autres disciplines pour tenter de répondre seules aux questions éthiques générales, les sciences cognitives se doivent de souligner avec précaution et modestie des questionnements éthiques sur lesquelles elles peuvent apporter de nouvelles connaissances qui contribueraient avec d'autres disciplines à actualiser la caractérisation des comportements et raisonnements éthiques chez l'humain.

Mots-clés : Éthique, sciences cognitives, décision, raisonnement, dilemmes moraux, influence comportementale, systèmes artificiels cognitifs, algorithmes, traitement des données personnelles.

ABSTRACT. Ethics and Cognitive Sciences. This special issue presents a set of contributions aimed at discussing ethical questions related to the cognitive science field. The goal is dual: raising ethical questions that may appear specific to cognitive science, as well as more general ethical issues on which cognitive science knowledge could shed a new light. Among the specific questions, this issue raises a number of questions related to potential applications of cognitive science research, such as the application of neuroscience to the juridical domain, of artificial intelligence to the development of autonomous artificial systems or to the processing of personal data, and of behavioral sciences to public policies. Among the general questions is raised the question whether cognitive science can bring a new perspective on humans' ethical decisions and moral dilemmas, given the fact that decision-making and cognition are core objects of research of the field. Shall our understanding of what ethics is at the XXI[st] century be revised given the new knowledge generated by

[*] Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique, Paris, France. mehdi.khamassi<at>upmc.fr
[#] Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique, Paris, France.
[◊] Liris UMR 5205, Université Lyon1

cognitive science research? The interdisciplinary points of views expressed in this special issue contribute to highlight that, rather than trying to give a sole response to ethical debates in replacement of other disciplines, cognitive science has the responsibility to underline with precaution and modesty some ethical questions to which the field can bring new knowledge which could contribute with other disciplines to update the characterization of ethical behaviors and reasoning in humans.

Keywords: Ethics, cognitive science, decision-making, reasoning, moral dilemmas, behavioral influence, artificial cognitive systems, algorithms, personal data processing.

INTRODUCTION

À moins d'être philosophe ou anthropologue avec l'éthique comme objet central de recherche, contribuer à des débats et questionnements sur l'éthique semble toujours une tâche ardue pour laquelle on n'a jamais l'impression d'avoir assez de temps à dédier au milieu de l'écriture de nos publications et demandes de financements. Coordonner un numéro spécial sur l'éthique s'en retrouve une aventure particulière, avec son lot d'espoir et d'enrichissement intellectuel d'un côté, de déception et de frustration de l'autre[1]. Rien ne semble inciter le chercheur à s'engager dans un travail actif sur l'éthique au sein de sa discipline, malgré une motivation et un intérêt qui semblent largement partagés.

Pourtant, la demande sociétale pour que les problèmes éthiques soient abordés et considérés par les chercheurs, et que ceux-ci développent une recherche socialement responsable (André, 2013), est telle que nous ressentons le besoin sans cesse renouvelé de nous emparer de ces questions, de nous les approprier, de nous questionner sur nos propres pratiques comme sur celles des contributions sociétales (positives ou négatives) sur lesquelles pourraient potentiellement déboucher les résultats de notre domaine de recherche dans un avenir plus ou moins proche. La pression à la publication ou la course au prestige sont telles que des cas de fraude scientifique sont mis au jour régulièrement (de Pracontal, 2001), contribuant à une exacerbation des risques d'une perte de confiance envers le milieu académique (Stengers *et al*., 2013). Par ailleurs, l'histoire a montré certains exemples d'engouement mêlé de naïveté pour certaines recherches ayant par la suite débouché sur des applications nocives pour la société, suggérant qu'une réflexion éthique en amont aurait pu aider à mieux anticiper voire prévenir les dérives (*e.g*., la phrénologie ou encore l'eugénisme, parmi d'autres exemples soulevés par Gouyon, 2010).

Si l'existence de comités d'éthique[2] permet d'évaluer le caractère éthique des protocoles expérimentaux, ou de faire des recommandations périodiques

[1] Ce numéro aura pour ses coordinateurs été l'occasion d'un nombre particulièrement inhabituel de refus de contribution pour diverses raisons et d'un nombre tout aussi inhabituel d'acceptations suivies d'un abandon en cours d'écriture.

[2] Voir par exemple le COMETS, la CERNA, le CEI de l'INSERM (https://www.inserm.fr/recherche-inserm/ethique/comite-ethique-inserm/missions-comite-ethique), le CCNE (https://www.ccne-ethique.fr/), les CPP (https://www.ars.sante.fr/comite-de-protection-des-personnes-1), le CERNI, POLÉTHIS.

sur les pratiques éthiques du chercheur (*e.g.*, COMETS 2014), tout suggère qu'au niveau individuel ce dernier ne doit jamais cesser de penser (Arendt, 1991), de penser sa pratique, son objet d'étude, son domaine au niveau collectif et son impact sociétal, ses applications possibles, et ce qu'est l'éthique elle-même pour lui. Ce numéro était donc tout d'abord, pour ses coordinateurs ainsi que pour le comité de rédaction de la revue Intellectica, une occasion de renouveler cette réflexion éthique, en bénéficiant de l'éclairage interdisciplinaire et de la richesse des angles de vue qui caractérisent toute question soulevée au sein des sciences cognitives.

Au-delà de la question des pratiques éthiques du chercheur, les sciences cognitives ont-elles quelque chose de particulier à apporter aux débats sur l'éthique, à notre compréhension même de ce qu'est un comportement humain éthique et de ce qui n'est est pas un, ou de comment un humain peut juger moralement du caractère éthique de telle ou telle pratique ou activité ? Les sciences cognitives ayant pour objet central, notamment, la cognition humaine, la façon dont l'humain décide de ses actions, les pense, envisage leurs conséquences possibles, voire y apporte des jugements moraux, il est tentant d'envisager une remise à plat même du concept d'éthique au regard des connaissances récentes en sciences cognitives. Mais cette ambition est-elle raisonnable ? Les quelques contributions à ce dossier qui abordent cette question (Andler, 2019 ; Penavayre *et al.*, 2019) soulignent à la fois que les sciences cognitives peuvent apporter un éclairage nouveau sur les raisons pour lesquelles les humains peuvent juger tel ou tel comportement éthique ou pas, et à la fois qu'il y a un danger à ce que les sciences cognitives veuillent contribuer seules à une redéfinition de l'éthique. Les conséquences sociétales peuvent être dévastatrices, ne serait-ce qu'en donnant au citoyen l'illusion qu'une réponse simple aux problèmes éthiques peut être trouvée au regard des connaissances en sciences cognitives, ou en réduisant la question de l'éthique à une question de compréhension des ressorts de la cognition humaine, ou encore en incitant le politique à prendre des décisions sociétales sur la seule base d'un éclairage issu des sciences cognitives.

Au fond, le débat est ici proche de celui qui a eu lieu sur l'application des sciences cognitives à l'éducation (Gaussel & Reverdy, 2013) : les sciences cognitives contribuent à une meilleure compréhension des mécanismes d'apprentissage chez les humains qu'on aurait tort de négliger dans l'élaboration de programmes éducatifs ; tout en risquant de réduire la compréhension de ce qui fait qu'un individu apprend ou n'apprend pas aux seules données encore trop jeunes issues du cadre restreint du laboratoire, et en négligence de toutes les connaissances qu'apportent d'autres disciplines des sciences humaines et sociales comme la sociologie sur ce qui caractérise la diversité des contextes sociaux dans lesquels l'apprenant est situé et contraint par de multiples forces.

L'interdisciplinarité intrinsèque, et qui semblerait presque incompressible dans le temps, des sciences cognitives semblerait pourtant mettre ce champ dans une position privilégiée pour ne pas tomber dans le piège de la réponse

---

Voir aussi https://www.who.int/ethics/partnerships/globalsummit/en/, ainsi qu'une liste d'autres organismes internationaux : https://www.cnrs.fr/comets/spip.php?article105

mono-disciplinaire à un problème complexe comme celui de l'éthique. Et pourtant… Au sein même des sciences cognitives, la tentation est parfois grande d'essayer d'apporter des réponses tranchées ou définitives à certaines questions au seul regard de sa propre discipline, sans adopter une perspective pluraliste[3]. Est-ce parce que le domaine des sciences cognitives est encore jeune (Andler, 2019) ?

Dans cet article d'introduction au dossier, nous allons tenter de dégager certains des enjeux éthiques importants liés aux sciences cognitives, notamment au titre de leurs applications en cours ou potentielles. Nous pointerons vers certains des éléments de réponse qu'apportent chacune des contributions au dossier, mais également d'autres points de vue qui auraient pu faire partie de ce dossier mais ne s'y retrouvent pas finalement, ou des points de vue d'autres disciplines. Nous conclurons en essayant de souligner certains des enseignements que le chercheur en sciences cognitives nous semble pouvoir tirer de ce dossier, et sur quelques perspectives pour continuer de faire vivre cette réflexion sur l'éthique en sciences cognitives.

1 – DÉFINITIONS, FONDEMENTS PHILOSOPHIQUES, ET TAXONOMIES ANTHROPOLOGIQUES

L'éthique possède une diversité de sens importante. Quand on cherche à en parler, il est donc essentiel de définir de quoi on parle, en particulier dans le domaine des connaissances scientifiques et de leur utilisation dans la société. Même si l'on rechigne à rapprocher les termes, morale et éthique sont liés d'un point de vue étymologique (le mot grec originel, *êthikós*, signifie « morale ») mais aussi historique. Comme l'écrit l'anthropologue Raymond Massé (2009) : « Valeurs morales, principes éthiques, vertus et règles de conduites sont autant de types de normes définissant le bien et le mal, l'acceptable et l'intolérable dans la quotidienneté autant que dans les moments critiques. ». La distinction semble alors reposer principalement sur le fait de considérer l'éthique comme le processus de questionnement et de critique au niveau individuel de diverses règles morales établies au niveau collectif – ces règles permettant à la société ou au groupe en tant qu'entité d'endosser une certaine autorité dans le guidage des individus vers les bonnes actions (Durkheim, 1895). Ce processus de questionnement et de critique est particulièrement important lorsque diverses règles morales entrent en conflit ou en contradiction. Pour Massé (2009), l'éthique est donc « l'espace (toujours limité) de liberté et d'autonomie qui permet aux êtres moraux de déconstruire les règles de conduite apprises [en vue d'être] des sujets moraux […] dans [leur] lutte constante contre [leurs] désirs, [leurs] intérêts, souvent en conflits avec les normes. » Étant donné que le questionnement éthique est ainsi en partie centré sur l'individu – mais pas seulement, nous y reviendrons –, il semble alors pouvoir être rapproché de la notion d'intégrité en tant que quête de l'individu d'une cohérence dans la manière dont ses actes respectent un certain corpus sans cesse retrié

[3] Voir notamment à ce sujet le dossier 69 d'Intellectica, coordonné par Cyril Monier et Alessandro Sarti, sur « Les neurosciences au sein des sciences de la cognition, entre neuroenthousiasme et neuroscepticisme ». Parmi les exemples discutés dans ce dossier, on peut citer le Humain Brain Project, qui semble avoir survendu à la société la possibilité qu'il aurait de réussir au bout de 10 ans à modéliser le cerveau humain.

(idéalement de façon consciente et active) de règles morales. L'éthique articule ainsi « la myriade de pratiques à travers lesquelles les gens engagent leurs propres actes, désirs et sentiments en tant qu'objets de délibération et de critique, d'enrichissement et de transformation » (Pandian, 2008).

Pour les membres de notre communauté scientifique, il est intéressant de noter qu'en dehors du domaine des sciences cognitives, l'anthropologie a constitué deux champs de recherche distincts (mais poreux), l'un abordant la question des moralités, et l'autre celle de l'éthique. Le premier s'intéresse à décrire les règles morales de chaque société, et à les lier aux normes des pratiques sociales, religieuses, politiques et économiques au sein de ces sociétés. Le deuxième vise « une comparaison transculturelle des morales et [une] recherche de potentiels universaux » (Massé, 2009). C'est ici que les sciences cognitives peuvent être tentées d'apporter leur touche. En particulier les connaissances qu'elles ont accumulées sur certains universaux des processus psychologiques humains en la matière, ou des processus neurobiologiques sous-jacents. Nous pensons que cet apport pourrait être particulièrement fertile, non pas en remplacement des travaux actuels en anthropologie, mais dans une démarche de modestie constructive de la part de chaque partie consistant à soumettre au regard critique de l'autre ce que nos données expérimentales et modèles pourraient apporter à la vision développée dans l'autre domaine. Nous espérions que des contributions à ce dossier aient pu participer de cette démarche, ou de toute autre démarche liant sciences cognitives et anthropologie. Ce n'est que partie remise. Signalons néanmoins l'existence de certains ponts déjà amorcés entre ces disciplines sur la question de la morale (Baumard & Sperber, 2007).

Il est tout aussi intéressant d'examiner d'un point de vue historique comment, au sein de la philosophie, la notion d'éthique s'est établie à l'antiquité et a évolué au cours du temps. Si l'on reprend les considérations d'Axel Kahn dans la préface de l'ouvrage de synthèse coordonné par Ludivine Thiaw-Po-Une intitulé « Questions d'éthique contemporaine » (Thiaw-Po-Une, 2006), les philosophes se sont emparés très tôt de la question et ne l'ont pas lâchée au travers du temps. Le sens habituel de l'éthique – savoir déterminer la conduite correcte dans les activités humaines – trouve ses racines dans la philosophie morale grecque (comme les ouvrages Éthique à Nicomaque, et Éthique à Eudème, d'Aristote) et a continué d'être étudié, élaboré, raffiné au fil des siècles (ex : Éthique de Spinoza). L'Éthique à Nicomaque d'Aristote (-330) propose ainsi une analyse finaliste des conduites et activités humaines, en cela que chacune d'elle est supposée viser un but ultime, qui est un bien en soit (« la fin ultime, c'est le bonheur », p. 67). En préface de l'ouvrage, son traducteur Richard Bodéüs souligne qu'« Aristote tenait ferme à l'idée que le bien ultimement visé par l'être humain pour lui-même, parce qu'il donne sens à sa propre vie, est nécessairement ce qu'il souhaite aussi pour ses semblables. » (*op. cit*., p. 36). En cela, « l'intérêt commun » est intimement lié à la quête d'éthique de l'individu pour Aristote, qui appelle « juste les prescriptions susceptibles de produire et de garder le bonheur et ses parties constituantes au profit de la communauté des citoyens. » (*op. cit*., p. 229).

Pour Aristote, le processus central pour l'homme préoccupé d'éthique est celui de la décision. Décider, « délibérer, c'est chercher les moyens d'atteindre

une fin » (*op. cit.*, p. 147), écrit Aristote en présentant une sorte de schéma décisionnel qui ressemble à ce qu'on pourrait appeler aujourd'hui un arbre de décision. De préciser plus loin : « l'objet de la décision est ce que la délibération a retenu comme désirable parmi les actes à notre portée [, …] la décision doit être le désir délibératif de ce qui est à notre portée. » (*op. cit.*, p. 150). « La décision ne peut se passer ni d'intelligence et de pensée, ni d'un état moral […, et] est le principe constitutif de l'homme » (voir la partie 1.3.2 sur « les principes de la décision », *op. cit.*, p. 294). C'est en cela qu'Aristote lie étroitement la question de l'éthique à celle de la décision, puisque « c'est à partir de sa décision que nous pouvons juger de la qualité d'un individu, c'est-à-dire en observant ce pourquoi il agit, non l'action qu'il exécute. » (*op. cit.*, II, 11).

De façon intéressante, la méthode d'Aristote consiste à organiser une classification des vertus et des actions (*e.g.*, le généreux, l'avare, le prodigue) ainsi que de leur mesure (évitant excès et insuffisance), et à bien les définir pour les distinguer, tout en adoptant souvent l'argument de l'usage de louer telle ou telle action pour considérer qu'elle est bonne et relève de la vertu. De conclure : « nous pourrons également mieux nous convaincre que les vertus sont des moyennes [entre insuffisance et excès] après avoir constaté qu'il en est ainsi dans tous les cas. » (*op. cit.*, p. 214).

Dans la quête d'éthique, Aristote met l'accent sur la distinction entre trois types d'activité : la vie méditative, simple étude de la vérité ; la vie politique qui est consacrée aux belles actions, celles qui viennent de la « vertu » ; la vie de jouissance vouée aux plaisirs corporels. Dans ce cadre, Aristote critique à la fois les tendances des hédonistes qui ont selon lui « l'illusion due au plaisir, car celui-ci n'est pas un bien mais paraît l'être » *(op. cit.*, p. 152), et à l'opposé une tendance qui deviendra celle des stoïciens consistant à « extirper » ou « négliger » le plaisir dans la conduite humaine (*op. cit.*, p. 107). La tempérance semble être un principe clé de la vertu (*op. cit.*, p. 172-175), puisque pour Aristote, les plaisirs corporels sont « bons jusqu'à un certain point » (*op. cit.*, p. 402). On pourrait se demander, non sans autodérision, si seul l'excès de plaisirs corporels est vil (car « bestial »), ou si au contraire l'excès de plaisirs intellectuels n'est pas lui aussi signe d'intempérance et source de risque d'addiction et d'aliénation.

De façon particulièrement intéressante par rapport à la distinction entre les modes délibératifs de la décision et les habitudes comportementales considérée dans le champ des sciences cognitives (Dickinson, 1985 ; Boraud, 2015 ; Khamassi & Pacherie, 2018 ; nous y reviendrons à la prochaine section), Aristote considère qu'il y a nécessité d'acquérir de bonnes habitudes pour permettre l'action éthique (« Travailler, par les habitudes, l'âme de celui qui écoute », *op. cit.*, p. 540). En effet, pour Aristote, les arguments et l'enseignement ne sont pas suffisants pour rendre les gens honnêtes, « l'honnête homme » étant pour Aristote « celui qui est porté à décider et à exécuter […] ce qui est juste et vaut mieux qu'une certaine forme du juste » (*op. cit.*, p. 282)[4]. Il y a pour Aristote nécessité de lois pour créer de bonnes habitudes. Ceci touche directement à la question éthique de comment

[4] À noter que la notion d' « honnête homme » existe déjà dans l'Enquête d'Hérodote (-425).

aujourd'hui les politiques publiques peuvent utiliser des connaissances de sciences cognitives et de sciences comportementales pour influer sur les comportements habituels des citoyens, ce qui est discuté dans l'article de Chammat et Giraud (2019) dans ce dossier.

Une question importante et liée à l'éthique que soulève Aristote est celle de la responsabilité vis-à-vis de nos propres actes. « Si nous pouvons faire remonter nos actes à d'autres points de départ que ceux qu'on trouve en nous, alors les forfaits qui ont en nous leur point de départ sont, eux aussi, des choses qui dépendent de nous et ils sont consentis. » (*op. cit.*, p. 153). Ceci pose directement la question du déterminisme et de la causalité des actions réalisées par les humains, qui constitue un débat central des sciences cognitives à l'heure actuelle. Sur ce plan, Aristote se situe dans une approche dualiste distinguant corps et âme, et considère que cette dernière a deux parties, une rationnelle et une irrationnelle (*op. cit.*, p. 291), donnant lieu à des facultés distinctes qui « permettent à l'âme d'énoncer la vérité sous forme d'affirmation ou de négation », et donc de décider. Ceci suppose l'existence d'un libre arbitre qui, à première vue, est nécessaire pour envisager une quelconque responsabilité des humains par rapport à leurs actes, et donc pour considérer la question de l'éthique comme pertinente[5]. Pourtant, en restant dans l'antiquité, ceci contredit l'hypothèse de Platon, lui aussi dualiste, selon laquelle nul ne fait le mal de son plein gré.

Ceci est également en contraste avec d'autres éthiques philosophiques plus récentes, comme celle de Spinoza, qui balaie la notion de libre arbitre en concevant la nature de l'esprit comme « idée du corps ». Les deux, corps et esprit, « chose pensante et étendue » (*i.e.*, pensée et matière, Spinoza (1677), E II, définitions et propositions 1 et 2, p. 115-118), intelligible et sensible, étant deux manières différentes d'exprimer (ou de nous permettre de concevoir) la même chose. Ils sont en effet considérés comme deux attributs de la substance unique, infinie et éternelle (*op. cit.*, E I, définition 6, p. 65), en d'autres termes, deux attributs de la Nature (*op. cit.*, E IV, proposition 4, p. 273). « L'esprit et le corps, c'est une seule et même chose, qui se conçoit sous l'attribut, tantôt de la pensée, tantôt de l'étendue. D'où vient que l'ordre et l'enchaînement des choses est un, qu'on conçoive la nature sous l'un ou l'autre de ces attributs, par conséquent que l'ordre des actions et passions de notre corps va par nature de pair avec l'ordre des actions et passions de notre esprit. » (*op. cit.*, E III, proposition 2, scolie, p. 183). Ceci conduit à supprimer toute notion de causalité entre corps et esprit : « le corps ne peut déterminer l'esprit à penser, ni l'esprit déterminer le corps au mouvement, ni au repos, ni à quelque chose d'autre (si ça existe) » (*op. cit.*, E III, proposition 2, p. 183). Les deux sont la même chose, deux façons de représenter le même événement ou le même état dans deux espaces de description différents. Nous sommes tentés ici de faire une analogie avec la modélisation robotique, tout en gardant beaucoup de précaution car il ne s'agit pas de dire que l'esprit humain est comme un robot, mais bien au contraire de souligner que de réfléchir à la manière dont un robot peut se représenter ses propres actions (et donc de faire tendre le robot vers l'humain) peut parfois donner un éclairage nouveau sur la façon dont l'humain

[5] Pour une contre-argumentation, le lecteur peut se référer à Atlan, 2002.

pourrait lui-même former de telles représentations. En l'occurrence, la possibilité est donnée en modélisation robotique de décrire (et de faire raisonner le robot sur) sa propre action dans l'espace cartésien du corps ou dans l'espace de la tâche, chacun de ces espaces ayant des propriétés différentes et permettant au robot d'appréhender des contraintes différentes sur l'action[6]. La philosophie de Spinoza le conduit en tout cas à éviter (et à critiquer) la notion de libre arbitre et d'autonomie de la volonté en ce qu'ils induisent souvent une vision finaliste de la nature (*op. cit.*, E I, p. 65, préface de E V, p. 351, E II, proposition 35, scolie, p. 153)[7]. Spinoza interpelle ainsi les « hommes qui se trompent en ce qu'ils se pensent libres, opinion qui consiste seulement en ceci, qu'ils sont conscients, de leurs actions, et ignorants des causes qui les déterminent » (E II, proposition 35, scolie, p. 153).

Tout ceci amène Spinoza à construire une éthique du déterminisme absolu, partant « de la puissance de l'intellect, autrement dit de la liberté humaine » qui amène la joie de comprendre comme « libre nécessité ». Ceci permet à Spinoza de fonder sa théorie de la connaissance : ce déterminisme pour Spinoza n'empêche pas chez l'humain une puissance d'agir, ni une volonté de comprendre, qui l'amènent à utiliser sa raison et à pouvoir ainsi orienter son action dans le sens de la vertu. « Par vertu et puissance, j'entends la même chose, c'est-à-dire, la vertu, en tant qu'elle se rapporte à l'homme, est l'essence même ou nature de l'homme, en tant qu'il a le pouvoir de faire certaines choses qui peuvent se comprendre par les seules lois de sa nature. » (*op. cit.*, E IV, définition 8, p. 270). C'est notamment dans l'effort de comprendre, dans ce désir en tant qu'affect et puissance, que l'humain peut ressentir de la joie, affect qui a la propriété d'augmenter la puissance de penser de l'esprit (*op. cit.*, E III, proposition 11, scolie, p. 192), et que ceci produit selon Atlan (2018, p. 597) un « cheminement de l'homme, à partir de son asservissement aux passions, vers [ce que Spinoza décrit comme] 'l'homme libre, sous la conduite de la raison'. ». On touche ici au concept de « libre nécessité » qui fait que dans la connaissance de soi-même, de ce qui l'entoure, et dans une meilleure compréhension des causes qui déterminent ses propres désirs et actions, l'humain peut devenir plus sage, réfléchir et connaître ce qui est désirable et non désirable pour lui, et donc sciemment prendre ses décisions. Il s'agit donc d'une liberté différente que celle considérée habituellement sous le chapeau du « libre arbitre », mais d'une liberté en tout cas suffisante pour permettre à l'humain d'être responsable de ses actes, car apte dans la connaissance à réfléchir à leurs possibles conséquences avant de décider d'agir (Atlan, 2002). Comme le souligne Atlan (2018, p. 50), cette suppression du libre arbitre « n'impliquait pas la fin de toute philosophie morale ; bien au contraire, la connaissance adéquate 'par les causes' des déterminismes de la nature, y compris la nature humaine, impliquait la possibilité de progresser 'sous

[6] Le lecteur intéressé peut se référer au dossier récent d'Intellectica sur les « Nouvelles approches en robotique cognitive », coordonné par Khamassi & Doncieux (2016).

[7] Pour Spinoza, « la Nature n'a aucune fin qui lui soit d'avance fixée, et […] toutes les causes finales ne sont que des fictions humaines. » (*op. cit.*, E I, appendice, p. 107), et les humains ne sont pas libres car il n'existent pas d'après la seule nécessité de leur nature et ne sont ainsi pas déterminés par eux seuls à agir. *(op. cit.*, E I, définition 7, p. 66).

l'emprise de la raison' vers plus de 'perfection' jusqu'à atteindre une forme nouvelle et plus authentique de liberté. »

De façon particulièrement intéressante pour ce dossier, cette vision du lien corps-esprit et des causes déterministes des décisions humaines fait écho aux recherches actuelles en sciences cognitives, et à la façon dont elles ont évolué vers une articulation des sciences humaines et des sciences de la nature, une articulation des sciences de l'esprit, de la psychologie, avec les sciences physiques et biologiques. « Nul ne pourra comprendre l'esprit humain lui-même de manière adéquate, autrement dit distincte, s'il ne connaît d'abord de manière adéquate la nature de notre corps. » (*op. cit.*, E II, proposition 13, scolie, p. 129). Ceci peut être vu, comme le suggère Atlan (2018, p. 234) comme « l'anticipation d'un programme de recherches […] qui se poursuit actuellement dans les sciences physiques étendues à la biologie [… et] repris par les sciences cognitives comme celui d'une naturalisation de l'esprit ». Gardons-nous toutefois de ne voir que le versant matérialiste et réductionniste de ce programme, et d'en oublier le « caractère programmatique […] et méthodologique » (Atlan, 2018). Ceci serait en effet oublier que Spinoza considère esprit et corps comme deux manières d'exprimer la même chose, et que de façon symétrique, on ne peut donc comprendre parfaitement tous les mouvements du corps humain si on ne connaît de manière adéquate la nature de l'esprit humain. On voit ici l'importance de discuter de la place qu'ont pris les neurosciences dans les sciences cognitives, et dans cette entreprise de naturalisation de l'esprit humain, sujet amplement discuté dans le numéro précédent d'Intellectica (Monier & Sarti, 2018). On voit également ici comment Spinoza joue le rôle de précurseur dans sa façon de suggérer qu'aucune des deux manières différentes de décrire la nature (et donc aussi l'humain, sa cognition, ses décisions, et son éthique) ne doit être négligée, que ce soit sous la forme de l'enchaînement des pensées, ou sous celle de l'enchaînement des mouvements du corps. Il prend pour cela l'exemple de la représentation mentale d'un cercle : « l'être formel de l'idée de cercle ne peut se concevoir que par une autre manière de penser, comme cause prochaine, et celle-ci à son tour par une autre, et ainsi à l'infini ; en sorte que, aussi longtemps qu'on considère les choses comme des manières de penser, nous devons expliquer l'ordre de la nature tout entière, autrement dit l'enchaînement des causes, par le seul attribut de la pensée, et en tant qu'on les considère comme des manières de l'étendue, l'ordre de la nature tout entière doit également s'expliquer par le seul attribut de l'étendue. » (*op. cit.*, E II, proposition 7, scolie, p. 121).

Ce sont tous ces éléments qui permettent à Spinoza de rejeter l'idée de libre arbitre en cela qu'il considère qu'il ne peut y avoir de causalité croisée de l'esprit sur le corps, critiquant la conception de Descartes d'une volonté humaine libre qui constitue comme une source d'énergie intérieure permettant d'activer la machine, et ainsi de causer les volitions particulières qui seraient des « décrets libres ». Symétriquement, il considère qu'il ne peut y avoir de causalité croisée du corps sur l'esprit, qui correspondrait à la vision opposée formulée notamment par Cabanis : « le cerveau sécrète la pensée comme le foie sécrète la bile ». Atlan (2018) discute de façon abondante de cette théorie philosophique au regard de certaines données expérimentales en neurosciences,

et notamment des travaux de Benjamin Libet (1992) suggérant que la décision consciente d'agir des sujets de ses expériences ne pouvait être la cause du mouvement puisqu'elle se produisait quelques centaines de millisecondes plus tard que l'activité cérébrale inconsciente d'initiation du mouvement. Ceci suggérait donc une inversion temporelle entre intention consciente rapportée par le sujet et activité cérébrale d'initiation de l'action, la première se produisant après la seconde. Ces résultats ont suscité de forts débats dans le domaine des sciences cognitives, et revêtent à première vue une importance particulière pour la discussion sur l'éthique possible des décisions humaines, en particulier dans les domaines juridiques et philosophiques. En effet, Libet et collaborateurs prétendent que leurs résultats démontrent que les actions volontaires sont en fait initiées de manière inconscientes dans le cerveau. Certains chercheurs parlent de déterminants inconscients dans le cerveau humain de décisions dites « libres » (Soon *et al*., 2008). Néanmoins, la surinterprétation de ces résultats expérimentaux a été très critiquée (*e.g*., Frith & Haggard, 2018). Une des premières critiques vient du fait que les actions impliquées dans les expériences de Libet et collaborateurs sont relativement triviales (lever un doigt), et ne demandent donc pas beaucoup de temps de préparation ni d'anticipation des conséquences possibles de l'action. Une deuxième critique concerne le fait que la plupart des décisions prises par les humains dans le monde réel impliquent une comparaison entre plusieurs actions possibles en compétition, ce qui n'est pas le cas dans l'expérience de Libet et collaborateurs. Or, devoir choisir entre plusieurs actions implique de comparer les conséquences de ces différentes actions de façon à évaluer lesquelles sont désirables pour le sujet. Une dernière critique vient du fait que c'est après avoir choisi/décidé quelle action effectuer que se pose la question du moment de l'initiation de l'action, et que dans l'action spontanée sans fort impératif d'agir ressenti par le sujet, ce moment peut fluctuer en fonction d'une activité spontanée stochastique dans le cerveau (Schurger *et al*., 2012) que Libet et collaborateurs ont parfois interprétée à tort comme le précurseur neural d'initiation de l'action volontaire (Frith & Haggard, 2018).

La notion de délai entre planification de l'action (et donc anticipation de ses conséquences) et initiation de l'action semble également jouer un rôle fondamental dans la possibilité qu'il offre de préparer l'exécution ou la non-exécution d'une action, et donc dans la question de l'éthique de la décision d'action. Comme le souligne Atlan (*op. cit*., p. 241), certains commentateurs des travaux de Libet et collaborateurs ont cherché à voir dans l'intervalle de temps entre l'initiation neurale du mouvement et son exécution une fenêtre d'opportunité d'intervention d'un libre arbitre ultimement secouru par la possibilité d'exercer un veto en inhibant les actions non désirées (*i.e*., interruption volontaire de l'action). De plus, lorsque l'intention d'action vise une exécution différée, il est alors permis au sujet d'élaborer une représentation consciente du but à atteindre et d'un plan d'actions possibles pour l'atteindre. En d'autres termes, si un certain nombre de données expérimentales ayant répliqué et étendu les résultats de Libet[8] ont confirmé qu'il y a bien une inversion temporelle entre l'initiation de l'action par des processus neuraux non conscients et sa prise de conscience, il semble néanmoins que plus le sujet

[8] Les ouvrages de Jeannerod (2009) et Atlan (2018) en donnent des discussions approfondies.

dispose de temps pour délibérer, planifier son action, et anticiper ses conséquences possibles, plus ce sujet devient apte à sortir des automatismes de l'action pour entrer dans des modes délibératifs de la décision (Viejo *et al.*, 2015). Le sujet peut ainsi participer consciemment et activement à une grande part du processus de décision d'agir plutôt qu'à son simple constat *a posteriori*. En particulier du fait que le temps de délibération et la connaissance qu'il apporte sur le processus décisionnel lui-même permettent de modifier la décision initiale (Resulaj *et al*., 2009), et font ainsi que la décision finale est le résultat d'un processus dynamique, rétroactif, qui fait en sorte que cette décision ne dépend plus seulement des causes initiales mais également des informations apportées (injectées) tout au long du processus dynamique de délibération. On touche à nouveau à la distinction entre délibération et automatismes (ou habitudes), qui fait directement écho à la distinction proposée par Daniel Kahneman (2011) entre un système 1 prenant des décisions rapides, automatiques, liées aux affects, et un système 2 capables de décisions lentes, délibératives et réflexives[9]. On peut y voir également ce lien entre réflexion sur ses propres actions, connaissance de leurs causes et de leurs effets, et aptitude à tendre vers une forme de liberté dans la raison et la vertu que prônait Spinoza. Nous pensons que ce processus de réflexion sur ses propres décisions, sur le temps long, peut permettre d'entrer dans le cadre d'une éthique de la « pensée complexe », de la « pensée systémique [qui] met alors en évidence l'importance décisive […] des modélisations pragmatiques, des conceptions induites par des buts subjectifs [ou 'projectifs' selon Fleurance 2018], qu'on place dans le futur mais qui façonnent les actions présentes. Ces buts, liés à des croyances et à des anticipations, rétroagissent sur l'action au fur et mesure que celle-ci en rapproche ou en éloigne, cependant que l'action, en se développant, modifie les buts. Il en résulte une dynamique complexe dépendante de sa propre histoire et du contexte. » (Mugur-Schächter, 1997, p. 170).

Enfin, pour terminer cette présentation historico-philosophique de l'éthique, on peut signaler que les branches de l'éthique se sont diversifiées au XVIIème siècle, où l'on voit notamment le siècle des lumières construire une vision de la société sur le profil de « l'honnête homme » développé par Aristote. Les branches de l'éthique se sont ensuite ramifiées de manière très importante au siècle dernier, avec notamment Henri Bergson qui met en exergue la notion de durée dans la planification de l'action et dans la convocation des mémoires du passé pour la préparation de la décision comme facteur clef d'exacerbation de la liberté du sujet (Bergson, 1896). Les travaux les plus récents insistent sur l'importance du contexte historique et culturel des propositions faites en matière d'éthique. À la différence de la morale, ou peut-être plutôt : encore davantage que la morale, l'éthique est nécessairement située dans un contexte culturel et politique. Ce sont sans doute les approches utilitaristes/pragmatiques qui aujourd'hui sont mises en avant, comme dans le dorénavant célèbre *dilemme du trolley* : dévier le trolley pour minimiser le nombre de décès est supposé classiquement représenter un calcul utilitaire de maximisation du bien-être collectif. Mais comment mesurer l'utilité au moment d'une décision à prendre ? Et cela a-t-il du sens de vouloir dégager une conception des décisions

[9] Pour une vision plus nuancée et moins binaire entre deux systèmes, voir notamment Osman (2018).

éthiques purement utilitariste et détachée des émotions, alors même qu'une décision dite utilitaire repose sur une comparaison de représentations mentales de valeurs, qui se sont construites en interaction avec nos émotions, comme le discute l'article de Marie Penavayre et collègues (Penavayre *et al*., 2019) dans ce numéro ?

2 – ÉTHIQUE ET INTÉGRITÉ SCIENTIFIQUE, POUR LE DÉVELOPPEMENT D'UNE RECHERCHE SOCIALEMENT RESPONSABLE

On pourra remarquer que la question de l'éthique au sein d'une communauté scientifique se rapporte le plus souvent à la question de l'intégrité scientifique et à la déontologie des pratiques des chercheurs. Ces questions sont importantes à aborder, à rappeler, et nous aurions aimé pouvoir compter dans ce dossier des contributions de membres de divers comités d'éthique institutionnels. En effet, si la question de l'intégrité scientifique pourrait être vue comme n'ayant ni plus ni moins d'enjeux que celle de l'intégrité au sein de n'importe quel autre corps de métiers, l'intégrité dans les processus liés aux connaissances scientifiques (production, diffusion, exploitation) ont une part de singularité en ce qu'elles ont un impact sociétal direct avec notamment le rôle des connaissances scientifiques dans la société, et la convocation des experts scientifiques pour prendre des décisions sociales, politiques, écologiques, etc. À ce titre, il aurait été intéressant d'examiner comment le lien du chercheur avec l'éthique évolue avec le temps, ce qui pose des enjeux sociétaux, scientifiques et politiques. Le domaine des sciences cognitives constitue-t-il une spécificité dans la manière d'aborder les questions éthiques ? Si cette question devra ici rester en suspens, nous ne pouvons que donner au lecteur quelques pointeurs vers les recommandations existantes en termes de pratiques de la recherche et d'intégrité scientifique issues de comités d'éthique liés à différentes communautés scientifiques, et laisser le lecteur en juger (*e.g*., Bouchard *et al*., 2002 ; Déclaration de Singapour sur l'Intégrité de la Recherche, 2010 ; COMETS, 2014, 2017 ; Charte nationale de la déontologie des métiers de la recherche, 2015 ; Science Europe 2015 ; Corvol, 2016 ; European Science Foundation, 2017).

Les connaissances issues du domaine des sciences cognitives peuvent-elles alors apporter une vision particulière de la question de l'intégrité des pratiques du chercheur ? Nous pensons ici à tous les comportements humains que touche le concept de banalité du mal décrit par la philosophe Hannah Arendt (1961). Ce dernier éclaire tous les petits arrangements quotidiens que le chercheur peut être tenté de faire dans son travail en réponse aux pressions à la publication, ou ses habitudes d'action qui lui permettent de faire l'économie de penser. Ici toutes les recherches en sciences cognitives sur l'importance des automatismes dans nos habitudes d'action (Dickinson, 1985 ; Boraud, 2015 ; Khamassi & Pacherie, 2019), y compris dans nos mauvaises habitudes, soulèvent un certain nombre de questions sur notre aptitude à (nous) questionner consciemment et régulièrement sur l'éthique de nos comportements. Les modèles computationnels développés dans le domaine (*e.g*., Renaudo *et al*., 2014 ; Viejo *et al*., 2015) permettent d'apporter une compréhension plus précise de certains mécanismes par lesquels les humains décident en fonction des circonstances (et de façon consciente ou non) de reposer sur leurs habitudes, et quels pourraient être les facteurs qui favorisent l'économie momentanée ou

prolongée de la délibération (*e.g*., sous-estimation de l'incertitude liée aux conséquences à long-terme des actions, sur-confiance en ses choix, surestimation de la stabilité temporelle des valeurs, pression temporelle à décider). Les sciences cognitives apportent également des connaissances sur la façon dont les humains prennent des décisions, évaluent l'utilité (anticipée aussi bien qu'*a posteriori*) de leurs actions dans l'environnement, et peuvent montrer non seulement des automatismes, mais aussi des troubles ou même parfois des pathologies de l'action (que ce soit au niveau de la décision préalable d'action, de sa mise en œuvre, de la surveillance de son exécution, ou de l'évaluation de ses conséquences ; Khamassi & Pacherie, 2018). En particulier, les sciences cognitives peuvent apporter un éclairage sur la notion de biais, que ce soit au niveau de nos perceptions, de nos interprétations, de nos décisions (Kahneman, 2011 ; Boraud, 2015 ; Bavard *et al.,* 2018), ou de nos explications *a posteriori* pour justifier nos comportements, en particulier en contexte social (Mercier & Sperber, 2017). Ceci fait directement écho au constat de Spinoza (1677) à propos du jugement moral : « quand nous nous efforçons à une chose, quand nous la voulons, ou aspirons à elle, ou la désirons, ce n'est jamais parce que nous jugeons qu'elle est bonne ; mais au contraire, si nous jugeons qu'une chose est bonne, c'est précisément parce que nous nous y efforçons, nous la voulons, ou aspirons à elle, ou la désirons. » (*op. cit*., E III, proposition 9, scolie, p. 191). Ces éléments, et notamment les connaissances les plus récentes en sciences cognitives sur la manière dont nous trouvons des justifications *a posteriori* de nos actions, peuvent servir d'aide à davantage d'autocritique et de prévention de nos biais décisionnels en vue de mieux éviter les comportements que nous jugerions non-éthique avec du recul, et que nous serions amenés à regretter plus tard.

Enfin, certaines des contributions au dossier rappellent que l'éthique de la recherche est avant tout l'éthique, et qu'elle est donc une responsabilité générale du chercheur en tant que citoyen, qui est plus large que celle du chercheur seul (Andler, 2019). C'est donc une affaire quotidienne, et non pas une simple capacité à respecter un ensemble de normes concernant les bonnes pratiques du chercheur. Elle touche donc à nos habitudes, encore une fois, en cela que nos décisions les plus automatiques impliquent une comparaison de valeurs même lorsque nous n'avons pas conscience d'être en train de manipuler des valeurs. Elle touche également à l'inscription de nos actes dans la société, pouvant permettre, lorsque nous en développons une réflexion consciente et active, de s'inscrire dans une « recherche socialement responsable » (André, 2013). Être responsable, c'est en particulier prendre une décision quand il n'y a pas cohérence entre différentes normes morales ou entre différentes contraintes, ce qui oblige à une certaine hiérarchisation des choix.

3 – ÉTHIQUE ET APPLICATION DES CONNAISSANCES SUR LA COGNITION NATURELLE

Une des questions centrales de ce dossier aura consisté à se demander pourquoi le domaine des sciences cognitives serait-il particulièrement intéressant à interroger du point de vue des questions éthiques ? Un des éléments de réponse touche aux applications spécifiques du domaine des sciences cognitives. Nous discuterons ici des applications liées à des connaissances sur la cognition naturelle, tandis que la section suivante

discutera de celles sur la cognition artificielle (algorithmes, intelligence artificielle, robotique autonome).

Toutes les connaissances sur les biais décisionnels des humains peuvent notamment donner lieu à un premier questionnement éthique concernant leur application à des fins d'influence du comportement. L'article de Mariam Chammat et Stephan Giraud (2019) dans ce dossier aborde l'utilisation des sciences comportementales par/pour les politiques publiques. Il s'intéresse en particulier à la notion de nudge (« coup de pouce » ou « incitation douce favorisant une décision souhaitée ») qui réfère à l'incitation d'autrui à effectuer un comportement considéré comme bénéfique à l'intéressé et/ou à la collectivité. Les justifications du nudge sont souvent prises sous l'angle de l'exemple du rapport parent/enfant, ce qui soulève la question de sa légitimité lorsque le nudge est appliqué sur un adulte. Chammat & Giraud (2019) plaident pour un usage responsable d'une approche comportementale au service des politiques publiques. L'efficacité des politiques publiques dépend en effet de leur capacité à faire évoluer les comportements des citoyens de façon à promouvoir l'intérêt collectif. Néanmoins, se servir de connaissances en sciences cognitives pour tenter d'influer sur le comportement des citoyens à leur insu, même dans l'intérêt collectif, pose clairement un problème éthique non trivial. Étant donné que la puissance publique cherche à accroître non seulement son efficacité, mais aussi sa légitimité perçue par les citoyens, elle doit faire attention aux accusations possibles de paternalisme excessif voire de pratiques manipulatrices.

Les auteurs proposent dans ce dossier deux critères d'évaluation des nudges, afin de pouvoir mieux en appréhender la dimension éthique : la transparence vis-à-vis du sujet du nudge (*i.e.*, la ou les personnes ciblée(s) par le nudge) ; la préservation de l'autonomie du sujet du nudge. Pour Chammat & Giraud (2019), « un nudge transparent est un nudge visible mais surtout explicite : il laisse apparaître l'intention dans laquelle il a été conçu. » De manière orthogonale, « le second critère d'évaluation proposé consiste à distinguer les nudges selon leur mode d'action sur le comportement et donc les processus cognitifs sous-jacents qui sont visés. Hansen et Jespersen distinguent ainsi des nudges mettant en jeu essentiellement les modes de pensée automatique (dits de type 1) et d'autres qui, tout en intervenant sur notre système cognitif automatique, mobilisent plus d'attention et de pensée réflexive (type 2). Le changement de comportement est dans ce dernier cas issu d'un processus de l'ordre de la délibération ou du choix. » On retrouve à nouveau ici la distinction entre le sytème 1 et le sytème 2 proposée par Daniel Kahneman (2011), et l'idée que si on influence le premier (automatique et habituel), on ne permet pas facilement au sujet du nudge de réfléchir consciemment à son action (ni à son caractère éthique ou non-éthique).

De façon intéressante, selon certaines définitions du nudge (Thaler & Sunstein, 2008), « l'intervention doit être simple et facile à esquiver.» Mais ceci rend les choses ambigües et contradictoires avec la volonté d'influencer efficacement. Du coup se pose la question : doit-on faire le bonheur des gens à leur insu, et modifier des comportements sans se prévaloir d'un consentement éclairé ? Pour Chammat & Giraud (2019) : « cela suppose *a minima* une rationalité du nudgeur bien supérieure à la moyenne. Or, rappelons-nous notre

constat initial : l'*homo œconomicus* est une chimère. Décideurs et praticiens du nudge, même avec toute la lucidité et l'humilité souhaitables, sont aussi des êtres de contexte, sans doute perclus de biais. » De par la structure de nos démocraties actuelles, les décideurs sont peu nombreux et sont peu représentatifs de la diversité sociale. On peut donc se demander si cela ne les rend pas plus facilement influençables, ne serait-ce que par des modes de pensée qui seraient spécifiques à leur classe sociale. Si par exemple les décideurs sont majoritairement éduqués dans des institutions où l'on répète à l'envie que le public est intrinsèquement moins efficace que le privé, alors ces décideurs pourraient être tentés de vouloir « nudger » la population en ce sens tout en pensant œuvrer pour une plus grande efficacité de la société et peut-être naïvement dans le sens d'un bien-être accru au niveau collectif. Mais où situer le périmètre légitime d'intervention de la puissance publique dans la sphère individuelle ? Faut-il nudger le nudgeur, surtout quand il s'agit d'une puissance publique en position d'influencer l'ensemble des citoyens dans le sens de l'intérêt collectif ? Comment également garantir une préservation de l'autonomie des personnes ? Voici certaines des questions soulevées par la contribution de Chammat & Giraud à ce numéro, qui concluent que « toute démarche nudge dans le domaine de l'action publique se doit d'être elle-même conçue en toute transparence. Ses objectifs, ses mécaniques, ses résultats doivent faire l'effort d'une publicité totale. La conséquence peut en être une atténuation des effets recherchés dans certains cas, mais c'est un impératif éthique qui peut même, par cet intermédiaire, contribuer à une pédagogie sur l'existence de certaines limitations cognitives et les moyens d'y remédier. Une chose est en tout cas sûre : ils doivent *a minima* être en situation de pouvoir demander des comptes et des réponses sur les tenants et aboutissants d'interventions touchant à leur comportements individuels. »

Enfin, Chammat & Giraud (2019) pointent vers une similarité entre les techniques du nudge et les techniques de marketing : « on ne saurait enfin oublier, s'agissant de la nécessité de contextualiser, que les méthodes, outils, connaissances, mobilisables pour « nudger » ont très largement à voir avec ceux du marketing. De cette proximité technique découle un évident principe prudentiel (la puissance publique ne saurait fonctionner comme un publicitaire prescripteur), mais on doit tout autant souligner l'importance des finalités recherchées. Les techniques restent des techniques, appréciables en tant que telles, mais plus encore en fonction de leur objectif d'usage. Ceci conforte l'idée, développée par ailleurs, qu'il est indispensable *a minima* de procéder à des évaluations *ad hoc* des interventions comportementales envisagées. » Ici nous est donnée l'occasion d'élargir le débat éthique sur l'utilisation de connaissances sur les biais cognitifs des humains dans le champ du marketing et de la publicité. En effet, les chercheurs étudiant les mécanismes de décision, en particulier dans les situations économiques, peuvent être tentés de conseiller les publicitaires sur la manière d'amener plus facilement le consommateur à l'acte d'achat. C'est ce qui a donné lieu à la naissance récente du champ du « neuromarketing ». Mais cette pratique a rapidement donné lieu à polémique, notamment au sein des sciences cognitives, puisque même si elle semble pour certain une contribution des chercheurs à la mission de valorisation des connaissances scientifiques, elle porte en elle le risque de contribuer à aggraver la réduction de l'autonomie décisionnelle des individus par le marketing

publicitaire à grande échelle. En effet, le calibrage de publicités sur la base de leur aptitude à activer certains réseaux cérébraux plutôt que d'autres peut, avec l'amélioration progressive de la rigueur des techniques d'imagerie cérébrale, amener à un renforcement de l'efficacité de l'influence publicitaire sur le comportement d'autrui (Dumas *et al*., 2012). Pour tenter d'échapper à la polémique, des chercheurs du domaine ont proposé de restreindre le terme de neuromarketing aux recherches en entreprise, tout en qualifiant leur recherche académique de « consumer marketing ». Mais la frontière entre les deux est-elle si nette et les conséquences éthiques sont-elles si distinctes ? Ces questions semblent appeler des réponses non triviales, et la réflexion éthique être particulièrement souhaitable ici.

Un autre débat éthique important concerne l'application de connaissances de neurosciences dans le domaine juridique (Oullier & Sauneron, 2012), et en particulier sur la caractérisation et la compréhension que peuvent avoir les citoyens de ce que sont la morale et l'éthique elles-mêmes. En effet, les sciences cognitives contribuant à une meilleure compréhension du pourquoi et du comment de nos jugements moraux, il pourrait être tentant d'y voir une opportunité de redéfinir la morale et l'éthique. Dans leur contribution au dossier, Marie Penavayre, Cédric Brun & Thomas Boraud (2019) questionnent ainsi ce que les neurosciences peuvent apporter à la caractérisation de nos jugements moraux. Contre la tentation du chercheur en sciences cognitives de redéfinir la morale ou l'éthique grâce à des données expérimentales sur les jugements moraux ou éthiques de sujets humains, ils rappellent, suivant David Hume, que « l'on ne pourrait dériver une norme d'un fait ou réduire une norme à un fait ». Mais ce positionnement a conduit pendant longtemps à un anti-naturalisme de principe, parfois caricatural, dans l'abord des questions éthiques, ce que les auteurs cherchent à éviter. Ils rappellent néanmoins qu' « une éthique naturalisée ne peut se réduire à l'élaboration d'une morale normative sur des données empiriques. […] L'entreprise de fondation d'une morale normative sur la base des seules neurosciences cognitives et comportementales conduit à ignorer la dimension historique et sociale de notre moralité. » Les neurosciences combinées à l'anthropologie cognitive ou encore à la psychologie évolutionniste peuvent aider à mieux comprendre, et donc à naturaliser, ces facteurs historiques et sociaux sur l'évolution de nos intuitions morales. Ceci reste quoiqu'il en soit dissocié des décisions normatives sur la morale.

Les auteurs discutent en particulier de manière approfondie de l'exemple du dilemme du Trolley étudié par Joshua Greene et ses collègues grâce à une série d'expériences où les sujets devaient décider de dévier ou non un trolley pour le faire tuer un plus petit nombre de personnes que celles mises en danger par la trajectoire initiale du trolley. Les auteurs critiquent en particulier l'utilisation par Joshua Greene du terme d'« utilitarisme » pour « désigner de manière générale tout jugement d'une action fondée sur l'évaluation de ses conséquences » et son opposition au « déontologisme [qui] est la position en éthique normative selon laquelle une action est morale si elle est conforme à certains devoirs (dans sa version kantienne, à des impératifs catégoriques) […] indépendamment des conséquences en termes d'accroissement de bien-être global. […] L'opposition entre utilitarisme et déontologisme vise généralement

une version kantienne du déontologisme moniste selon laquelle les devoirs moraux dérivent d'un principe universel unique permettant à la raison d'identifier de manière autonome les normes morales qu'elle doit suivre pour juger moralement. » (Penavayre *et al*., 2019). Greene semble dénigrer les jugements de type déontologiques, en les considérant comme trop liés aux affects et trop peu à la rationalité. Néanmoins, ceci revient à négliger que même un jugement dit utilitaire repose sur une comparaison de valeurs qui ont pu être construites mentalement en interaction forte avec les émotions (Damasio, 1995). Les auteurs soulignent ainsi qu'il y a une forme de confusion conceptuelle dans ce type de travaux, mêlée à une confusion inférentielle qui vise à généraliser des conclusions sur les jugements moraux à partir d'une situation expérimentale (du trolley) simpliste.

Selon les auteurs, ces confusions semblent résulter d'une motivation de Joshua Greene et collègues à aller plus loin qu'une simple caractérisation des bases neurales du jugement moral, jusqu'à viser une utilisation des résultats expérimentaux en vue de renforcer au niveau sociétale la légitimité et la crédibilité des jugements utilitaristes par rapport aux jugements déontologiques. Comme ils l'écrivent : « Joshua Greene résume cette ambition de la manière suivante : 'La science peut faire progresser l'éthique en révélant les fonctionnements internes cachés de nos jugements moraux, particulièrement ceux que nous faisons intuitivement. Une fois ces fonctionnements internes révélés, nous pourrions être moins confiants dans certains de nos jugements et vis-à-vis des théories éthiques qui sont (explicitement ou implicitement) fondées sur ceux-ci.' […] Il ne s'agit donc pas seulement, selon Greene, d'identifier des biais de raisonnements (par exemple, liés à nos émotions) systématiquement mobilisés lors de certains jugements moraux et d'informer nos procédures de jugement, à partir de la reconnaissance de ces biais, pour les éviter ('voie directe' de l'usage des neurosciences des jugements moraux (*ibidem*, pp. 711-716)). Il s'agit également (surtout ?) de déterminer vers quelle théorie éthique normative se tourner pour juger de la moralité d'un acte lorsque nous cherchons à 'résoudre un désaccord moral pratique' (*ibidem*, p. 717). Cette double ambition est au cœur d'une discussion très vive dont le principal enjeu est de déterminer quelle contribution les neurosciences des jugements moraux peuvent réellement espérer apporter à l'éthique normative. » On voit donc qu'il y a un débat éthique important à poser à propos de la tentation des sciences empiriques de contribuer à redéfinir l'éthique. Certains auteurs considèrent plutôt que les deux questions sont orthogonales, et qu'il n'y a même pas besoin de discuter de la validité scientifique des travaux sur les bases neurobiologiques de la morale pour décider si les neurosciences peuvent contribuer ou pas à modifier la manière dont nous jugeons les autres (Bigenwald & Chambon, 2019). Selon eux, n'importe quelle modification apportée par la science dans un domaine normatif (le droit, l'éthique) a besoin d'être intégrée par les mentalités pour être effective (quand bien même la science aurait raison !). C'est la raison pour laquelle même si la science prouve que le libre-arbitre est une illusion, tant que l'illusion perdure (dans les mentalités), nous pouvons être tenus pour responsables.

D'une façon plus générale, il est important de se poser régulièrement des questions d'éthique à chaque fois qu'il est question de généraliser des données empiriques obtenues dans un cadre expérimental (par nature restreint) à des problématiques sociétales. Ceci ne veut pas dire qu'il ne soit pas souhaitable de transférer voire d'appliquer des connaissances empiriques de ce type à des débats sociétaux. Ceci peut au contraire apporter des connaissances utiles au débat. Comme le souligne Marie Penavayre et collègues (2019) : « Toutefois, il nous paraît utile que les neurosciences des jugements moraux poursuivent leurs recherches afin de nourrir notre 'vie éthique', au sens où identifier les processus neurobiologiques sous-tendant nos jugements moraux permet d'attirer notre attention sur des biais systématiques, des phénomènes de sensibilité au contexte ou de saisir comment l'évolution a contraint certaines structures neurobiologiques impliquées dans nos jugements moraux. »[10] Mais il s'agit plutôt de se poser systématiquement la question de la limite des réponses qui peuvent être apportées aux débats sociétaux de cette façon. Au fond, il semble que de parvenir à transmettre des connaissances scientifiques accompagnées de toutes les précautions qui semblent nécessaires quant aux limites de leur interprétation soit, d'un côté, une tâche plus ardue pour le chercheur en quête d'interaction avec la société, mais de l'autre, une façon d'apporter un bénéfice plus grand sur le long terme. Ceci pourrait permettre aux non-scientifiques de montrer davantage d'esprit critique pour d'une part traiter les paroles des scientifiques avec scepticisme constructif (Khamassi & Decremps, 2019), notamment à cause du « neuroréalisme » (McCabe & Castel, 2008 ; Weisberg *et al.*, 2008), et d'autre part de tomber moins facilement dans la crise de confiance envers les scientifiques dès qu'un résultat scientifique est nuancé ou contredit par un autre (Stengers *et al.*, 2013).

De façon similaire, la contribution de Daniel Andler au dossier examine les conséquences sociétales qu'il peut y avoir à utiliser les connaissances issues des sciences cognitives pour assassiner publiquement le libre arbitre. Naturaliser l'humain pour finir peut-être par démontrer qu'il n'y a pas de libre arbitre ne risque-t-il pas de déresponsabiliser les humains dans leurs décisions et leurs impacts sur la société ? Daniel Andler se demande alors si ces questionnements éthiques doivent nous conduire à abandonner les recherches sur le libre arbitre, ou si ce n'est pas le cas, à les accompagner d'une pédagogie accrue pour que les résultats de ces recherches ne soient pas interprétés comme rendant obsolète la notion de libre arbitre dans la société humaine, notamment en termes juridiques. En effet, comme le soulignent Bigenwald & Chambon (2019), la question du libre-arbitre n'a rien à voir avec l'institution de la responsabilité pénale : on juge quelqu'un parce qu'il rapporte subjectivement avoir été conscient de son action et de ses conséquences ; on ne juge pas la présence ou l'absence d'antécédents neurologiques non-conscients à l'action elle-même.

Parallèlement à cela, si les recherches en sciences cognitives sur les principes psychologiques et neurobiologiques sous-jacents aux jugements

[10] On revient ici d'une certaine manière sur la proposition de Spinoza que la réflexion et la connaissance sur notre propre processus de délibération conduit à une augmentation de la puissance d'agir et donc à une forme de liberté « authentique » dans la décision éthique.

éthiques montrent que ceux-ci reposent principalement sur l'impact de nos émotions sur notre raison, ceci peut amener certains à rejeter ou minimiser la portée des jugements éthiques. Dans leur contribution, Marie Penavayre et collègues examinent également cette question et ses conséquences possibles dans le domaine juridique.

3 – ÉTHIQUE ET APPLICATION DES CONNAISSANCES SUR LA COGNITION ARTTIFICIELLE

Ce dossier thématique d'Intellectica est également l'occasion d'examiner des questions éthiques soulevées par une autre série d'applications issues de recherches sur la cognition, en particulier celles sur la cognition artificielle.

a - Algorithmes et intelligence artificielle

Le premier numéro d'Intellectica, en 1985, se titrait « Les interactions homme/ordinateur ». Trente-quatre ans plus tard, « Les interactions homme/ algorithmes », « Les interactions homme/robot », « Les interactions homme/ intelligence artificielle » soulèvent des questions éthiques inédites et mobilisent les différentes disciplines des sciences cognitives. Ce numéro aborde donc naturellement ce qui relève des algorithmes et de l'intelligence artificielle « à l'ère du web et de l'internet ». Ce qui relève de la robotique est présenté dans la section suivante.

La notion d'intelligence artificielle a été imaginée très tôt dans l'histoire de l'humanité et l'on sait qu'Alan Turing a imaginé son « test de Turing » en 1950 : « Si une machine peut mener une conversation qu'on ne puisse différencier d'une conversation avec un être humain, alors la machine pouvait être qualifiée d'intelligente ». On voit que dans ce test, c'est l'utilisateur *abusé* par une interaction bien menée qui projette sur la machine le don d'intelligence. Le monde de la recherche s'est étonné que les média qualifient d'intelligence artificielle un nombre croissant d'algorithmes et puis insensiblement associent la notion d'algorithme à la notion d'intelligence artificielle. Les *applications* souvent qualifiées *d'intelligences artificielles* auraient donc passé avec succès le test de Turing ? En quelque sorte oui, puisque les observateurs (utilisateurs, journalistes, blogueurs, …) considèrent ces applications comme intelligentes ! Les chercheurs étaient incrédules, car la recherche en intelligence artificielle s'est formellement mise en place à Darmouth en 1956, en considérant les différentes fonctions cognitives et étudier comment elles pouvaient être mises en œuvre avec des algorithmes les plus généraux possibles, tandis que les sciences cognitives adoptaient parfois le modèle computationnel pour étudier les mêmes fonctions cognitives. Les chercheurs n'imaginaient pas utiliser le test de Turing (subjectif) pour valider leurs travaux. Le terme d'intelligence artificielle se justifie également par un nombre croissant d'applications numériques actuelles qui s'intéressent à réguler des fonctions cognitives qui, jusqu'à il y a peu, étaient réputées hors du champ de l'automatisation. Un marché de la cognition est apparu. Il se décline en recommandation comportementale, orientation de la décision, décision de ce qui est *bien ou non* pour l'utilisateur, jugeant des capacités cognitives lors des apprentissages, classant inlassablement les profils cognitifs au profit de

stratégies marchandes ou sécuritaires, et s'engageant à tout faire *pour le bien-être des utilisateurs*.

Les terminaux mobiles, les objets connectés, le développement du web, la généralisation des protocoles internet pour les communications, l'automatisation galopante de tâches réputées cognitives contribuent à donner à ce marché de la cognition un développement tellement rapide qu'il est considéré comme prioritaire par la plupart des économistes et des responsables politiques. Ce développement galopant ne laisse pas beaucoup de place à la prise de recul, à la réflexion et l'inquiétude grandit dans la société tandis que les questions éthiques sont considérées comme essentielles et urgentes à étudier. Les termes de *transparence, loyauté, traçabilité, équité et explicabilité* reviennent souvent dans les recommandations faites par les institutions comme le CNNUM, le CIGREF, la CNIL avec le rapport « Comment permettre à l'Homme de garder la main ? », mais aussi par des associations comme la Quadrature du Net. La question est posée clairement aujourd'hui aussi bien au niveau national[11], européen (Groupe Européen D'Éthique des Sciences et des Nouvelles Technologies, 2018) qu'international (IEEE Public Discussion, 2019).

Dans ce numéro thématique, l'article de Camille Roth s'intéresse à la qualification précise des effets cognitifs des algorithmes à l'œuvre pour donner aux utilisateurs une vision du paysage informationnel correspondant à ses besoins. C'est sur le Web que le paysage informationnel est façonné pour chacun.e selon ses *préférences* implicites ou/et explicites et aussi selon des *indicateurs* évaluant la *qualité* du paysage proposé. Deux formes idéales d'algorithmes sont distinguées : les algorithmes qui tendent à « lire dans nos pensées » (Read our Mind : ROM) et les algorithmes qui tendent à « nous faire changer d'avis » (Change our Mind : COM). Une autre façon de distinguer les effets produits par les algorithmes est également retenue, selon le niveau auquel ils agissent : un effet de « réarrangement » de l'information, en amont, lié à la réduction ou projection de l'information retenue ou filtrée avec un certain objectif ; un effet d' « arrangement » de l'information, en aval, plus psychologique, lié à la façon de présenter les résultats de la recherche, dans leur mise en forme mais aussi par la révélation d'indicateurs (de succès par exemple) pour d'autres utilisateurs. Les algorithmes de recommandation, en particulier, s'appuient sur des « préférences » utilisateur implicites en exploitant les interactions et les données échangées pour chercher des similarités de comportement et en déduire des informations utiles. Ils exploitent aussi plus classiquement des préférences explicites que l'utilisateur précise volontairement. De ce point de vue, ils tendent vers le pôle « ROM ». Néanmoins, ils sont aussi susceptibles d'influer sur la manière dont le paysage informationnel est sélectionné pour l'utilisateur, débouchant possiblement sur une « bulle de filtrage » très personnalisée : c'est ainsi que les études montrent que même en mode « protégé » (sans cookies en particulier) les réponses à la même requête par plusieurs utilisateurs donne des résultats chaque fois différents. La facette COM se réalise ainsi dans le contexte de la facette ROM

[11] Voir les nombreux avis sur le sujet du Comité Consultatif National d'Éthique : https://www.ccne-ethique.fr/fr/type_publication/avis (dernière consultation le 4 juillet 2019).

et, d'une certaine manière, en recommandant d'autres choses, peut offrir une part de sérendipité autorisant à sortir de la « bulle de filtrage ». La question est alors de savoir si cette recommandation va dans le sens de ce que cherchait l'utilisateur ou dans un sens d'orientation vers d'autres choix, de consommation par exemple, ou parfois même de vote politique.

La façon de présenter et d'environner un paysage informationnel joue un rôle déterminant dans la manière de fournir à l'utilisateur un indicateur de la « valeur » de son choix, avec par exemple des scores de pertinence. Ce processus a tendance à orienter vers les articles les plus sélectionnés mais certains algorithmes cherchent aussi à fournir des « coups de coude » (ou *nudge*) pour proposer des choix moins fréquents mais pouvant convenir en termes de besoin, ouvrant l'espace informationnel sur une zone non éclairée dans la bulle de l'utilisateur.

L'article se conclut sur la nécessité de développer les études cherchant à établir à quel point l'utilisateur comprend ce qu'il se passe, à quel point il « devine » l'algorithme ou les algorithmes à l'œuvre. Ce n'est pas encore très étudié, mais les premiers résultats portant sur l'analyse de la manière dont 500 utilisateurs de Facebook considéraient le rôle éventuel de l'algorithme de curation dans la présentation (ou pas) d'une information, donnent : *« pas d'idée : 20% ; c'est l'utilisateur qui décide : 25% ; j'ai vu des symptômes de la curation (des posts publiés mais non présentés) (80%) ; je crois qu'il fait comme ceci ou comme cela (45%) »*. Ces chiffres montrent à quel point l'utilisateur tente de comprendre pour agir correctement ! Il n'en a évidemment pas les moyens et est réduit à émettre des hypothèses impossibles à vérifier. Il s'agit alors de ce que l'auteur appelle « folk algorithmics » (algorithmie populaire), par analogie avec la folk physics ou folk biology.

La contribution d'Alain Mille « Vers des dispositifs techniques numériques orientés éthiques ? » pose la question de manière plus générale en s'intéressant aux conditions à réunir pour que les dispositifs techniques numériques diffusés à l'ère du web et des intelligences artificielles, constituant le problème puissent également faire partie de la solution pour que la société puisse gérer les questions éthiques soulevées par leur utilisation. Cette position s'oppose aux postures de défiance absolue vis-à-vis des dispositifs techniques numériques adoptées par certains théoriciens de la technoscience comme Ellul (1977) par exemple. Cette posture de défiance est considérée par l'auteur comme improductive pour réfléchir aux pistes à explorer dans la situation actuelle. Pour définir les bonnes propriétés que doivent posséder les dispositifs techniques numériques, un tour d'horizon des prises de position philosophiques repérées par l'auteur (sans prétention de production de connaissance en la matière) est réalisé sous une forme d'interviews rétro-futuristes qui sont l'occasion de rappeler la complexité et la vitalité de la question éthique quand il s'agit de dispositifs techniques. Deux propriétés principales sont dégagées : Propriété I de construction et mémorisation de l'activité située du technicien (en référence à Simondon) en tant que rétentions tertiaires (en référence à Stiegler) ; Propriété II de soutien au processus de discussion sur l'activité et sa régulation. Ces propriétés sont ensuite étudiées dans le détail au travers d'un état de l'art de travaux, réalisations, dispositifs qui les exhiberaient et les illustreraient concrètement. Les qualités d'un dispositif technique numérique

pouvant démontrer la propriété I sont passées en revue : capacité à documenter les fonctions assurées, accessibilité et lisibilité de la documentation (Open Source) ; reposant sur des sciences et techniques largement enseignées et apprises (informatique à l'école, tout au long de la vie, publications ouvertes, …) ; capacités d'inscription des connaissances encapsulées sous une forme accessible à toutes et tous, typiquement en texte ; capacités à fournir les traces d'interaction avec les outils de leur analyse au moment de l'interaction même, plusieurs travaux de l'équipe de l'auteur illustrant cette partie. Une catégorisation en types de documentarisation différents est proposée pour le monde de l'intelligence artificielle : le type *historique* cognitiviste qui proposait déjà de fournir les bases de règles, de faits et les traces de raisonnement sous une forme textuelle ; le type *ontologique*, généralisation de phase précédente en proposant un *niveau connaissance* indépendant du niveau *opérationnel ;* le type *non symbolique* reposant sur l'analogie neuronale, la modélisation en agents réactifs ou encore d'agents en apprentissage interactionnel toutes formes aujourd'hui très difficiles à équiper de modules d'explication des fonctions assurées ; le cas du *deep learning* est étudié plus spécifiquement puisque aujourd'hui largement considéré comme l'intelligence artificielle par excellence avec ses difficultés intrinsèques à offrir des capacités d'explication et de documentarisation de ces explications. La propriété II de soutien au processus de discussion sur l'activité et sa régulation est considérée en examinant l'importance accordée à la capacité de réflexivité et d'affordance dans les interactions homme-machine. Deux illustrations tirées des travaux de l'équipe de l'auteur sont utilisées pour montrer comment les mécanismes d'explication peuvent être interactifs avec l'utilisateur et constituent un potentiel pour alimenter des processus de discussion. Les dispositifs de discussion collaborative qui se développent actuellement n'ont pas été traités dans l'article faute de place et de recul également sur leurs capacités à soutenir des délibérations éthiques. Ces travaux sont considérés comme essentiels dans la discussion du travail.

b - Robotique autonome

De son côté, la communauté robotique a également pris à bras le corps les enjeux éthiques notamment liés au développement de systèmes artificiels autonomes, et en particulier de robots autonomes (*e.g*., Pham *et al*., 2018). Le développement de robots autonomes, dits « cognitifs » en cela qu'ils sont programmés pour réaliser des tâches impliquant des processus similaires à certaines fonctions cognitives chez l'humain comme la planification de l'action, la navigation, l'interaction sociale (Khamassi & Doncieux, 2016), a en effet un impact direct sur la société en termes d'emploi et de conditions de travail. Plus précisément, par leur contribution à l'automatisation de la production industrielle, les robots conduisent à une disparition des métiers les moins qualifiés, tout en ouvrant la possibilité de créer de nouveaux métiers pour la conception et le maintien de ces robots. Bien que le chercheur en robotique aime souvent se rassurer en se disant que les robots peuvent aider à libérer les humains des tâches ingrates, répétitives, aliénantes ou dangereuses pour l'humain, le chercheur ne doit pas éluder les questions des conséquences sociétales du déploiement des robots. En effet, Pham et collaborateurs (2018)

soulignent que malgré ce potentiel, les robots sont perçus par une part de l'opinion publique comme une menace pour l'emploi. Ceci est dû notamment au fait que les profits réalisés grâce à l'automatisation de la production par le biais du robot ne semblent pas s'accompagner systématiquement d'une augmentation de programmes éducatifs (ou de leurs moyens financiers) permettant aux ouvriers non-qualifiés d'atteindre un niveau de qualification plus élevé. D'une façon générale, les auteurs soulignent que le chercheur ne doit pas négliger la question politique (et donc éthique) de qui possède les robots déployés dans la société, et de comment est organisé et structuré le système économique actuel, pour pouvoir envisager sereinement une contribution sociétale positive sur le long-terme des recherches actuelles en robotique.

Un autre débat éthique important concernant le développement de robots autonomes concerne leur utilisation dans le domaine militaire. En effet, l'autonomie croissante des systèmes d'armements est devenue une problématique centrale débattue au niveau international dès 2013 avec notamment un rapport de l'ONU sur le développement d'armes autonomes « pouvant, une fois activées, poursuivre leur cible et passer à exécution sans intervention humaine » (Heyns, 2013). À cette occasion, comme le décrivent Righetti et collaborateurs (2018), un ensemble de chercheurs en robotique ont lancé une campagne intitulée « stop killer robots ». Cette initiative peut être considérée comme la première discussion internationale multilatérale sur les systèmes autonomes, et visait à faire prendre conscience à un large public des enjeux éthiques importants liés aux développements technologiques actuels issus de la recherche en robotique. Ceci a permis aux États de convoquer en 2016 un groupe d'experts ayant pour mission d'évaluer les conséquences de ces développements juridiques en termes de droit. Néanmoins, comme le soulignent Righetti et collaborateurs (2018), il n'y a pas encore de consensus entre États à l'heure actuelle pour savoir s'il doit y avoir régulation et ce qui doit être régulé. Un point important du débat concerne la distinction, rarement faite en robotique, entre « systèmes autonomes » et « systèmes automatisés ». Les premiers impliquent un algorithme de prise de décision difficile à prédire, tandis que les seconds réfèrent à un système programmé pour répondre de façon déterministe à des événements bien identifiés à l'avance[12]. Cette distinction souligne qu'un enjeu juridique important des systèmes artificiels autonomes concerne leur capacité à choisir leur cible et à décider de façon autonome de tirer. Les auteurs défendent à ce sujet la thèse que même si des robots étaient bientôt en capacité technique de prendre une décision de tuer qui soit éventuellement en accord avec la loi, ceci ne serait moralement pas acceptable car « cela saperait la notion d'humanité et serait un affront pour la dignité humaine ». Les auteurs soulignent que le chercheur en robotique se doit de ne pas rester silencieux sur ce sujet, de contribuer au débat, voire même de participer activement à l'établissement de régulations et de normes éthiques en la matière qui n'existent pas encore. On pourrait ajouter que le rôle du chercheur est également d'apporter un éclairage au large public sur les limites et potentialités techniques en cours et à venir des robots. En effet, il y a à

[12] Les auteurs mentionnent le cas des mines antipersonnel, qui ont déjà été bannies par le traité d'Ottawa.

nouveau ici une question politique du fait que même si les robots ne sont pas encore techniquement avancés, adaptatifs ou « intelligents », ils peuvent déjà être utilisés pour tuer. Il y a en effet déjà plusieurs exemples d'humains qui ont été blessés ou tués par des armes automatiques programmées par de simples algorithmes réactifs aux stimuli extérieurs. La question est donc non seulement celle des potentialités techniques des robots, qui doivent être connues du grand public pour lui permettre d'anticiper et de réfléchir sur l'avenir de la société, mais également à nouveau sur qui possède les robots et qui décide de comment les utiliser.

Ces exemples de contributions aux débats éthiques sur le développement et le déploiement sociétal de robots autonomes soulignent en tout cas l'idée que même si les recherches d'un chercheur en robotique ne contribuent pas directement au développement d'armes, ni à d'autres applications nocives pour la société, le chercheur est néanmoins un des mieux placés pour savoir quels types de connaissances de son domaine ont vocation à être exploitées en ce sens, ou dans quels sens les développements en la matière ont le plus de chance d'évoluer. Il y a donc un devoir éthique à communiquer envers le plus grand nombre, pour faire savoir quel est l'état des connaissances et quels sont les risques, et ainsi permettre à chaque membre de la société, en tant que citoyen, chercheurs et non-chercheurs, de se faire suffisamment tôt une opinion informée sur les applications technologiques souhaitables ou non-souhaitables.

5 – DISCUSSION

Ce dossier aura été l'occasion d'aborder un ensemble de questions éthiques qui se posent au chercheur en sciences cognitives, de l'intégrité des pratiques de recherche aux réflexions sur les conséquences sociétales des applications issues des sciences cognitives. Mais également de la question plus générale de l'éthique du chercheur en tant qu'humain et citoyen responsable, et de la manière dont les connaissances en sciences cognitives peuvent éclairer notre compréhension des processus par lesquels les humains prennent des décisions éthiques ou non, et portent des jugements notamment moraux.

En guise de conclusion partielle et provisoire, nous sommes tentés de citer la préface de Pierre Macherey au « Cours de philosophie biologique et cognitiviste. Spinoza et la biologie actuelle » (2018) d'Henri Atlan, qui est lui-même médecin, biologiste et philosophe : en analysant l'Éthique de Spinoza, « Henri Atlan s'intéresse à la pensée de Spinoza avant tout parce qu'il considère qu'elle fournit une arme imparable pour combattre la philosophie spontanée des savants [notamment le dualisme et les monismes que sont l'idéalisme et le néo-matérialisme de l'esprit produit par la matière] et la conviction qu'il est possible de parvenir à un savoir achevé et définitif des réalités du monde qu'elle entretient. C'est dans ce sens que, dans son grand ouvrage sur Les Étincelles de hasard, il caractérisait déjà sa propre tâche : « La question n'est pas de croire en un contenu de connaissance scientifique, mais de délibérer sur le domaine de ses applications pertinentes, pour savoir comment s'y référer et comment l'intégrer à l'orientation de notre pensée et de notre existence » (t. I, p. 193) » (Atlan, 2018, p. 14). Il s'agit donc ici directement de philosopher en tant que se questionner sur l'éthique en soi, « sans se substitue[r] au travail de la connaissance scientifique » (*op. cit.*,

p. 15). Ceci souligne que malgré la pression à la publication qui pousse sans cesse le chercheur à se focaliser uniquement sur sa production scientifique, les questionnements sur les applications à long-terme de ses recherches et sur l'éthique ont tout autant d'importance pour la société. Ceci rappelle également que la philosophie et le questionnement sur l'éthique sont historiquement liés, et peut aider à mieux comprendre pourquoi certains ouvrages majeurs de philosophie portent tout simplement le titre d'« Éthique » (comme celle de Spinoza ou celles d'Aristote, sus mentionnées). Ceci ne doit toutefois pas conduire à négliger que d'autres courants de la philosophie s'intéressent à des problématiques sans lien avec l'éthique, comme la métaphysique, ou encore la philosophie des sciences.

Avant de conclure définitivement cette introduction, il nous semble important de discuter une distinction importante qui a été faite, notamment dans l'article de Daniel Andler (2019) dans ce numéro, entre les responsabilités éthiques individuelles et collectives des chercheurs, car elle permet au chercheur de prendre sa part de responsabilité dans les choix éthiques qui dépassent le simple cadre de ses décisions individuelles. Daniel Andler parle certes de la responsabilité plus large du chercheur en tant que citoyen qui va au-delà de la responsabilité en tant que chercheur seul. Il distingue également les responsabilités individuelles du chercheur (dans ses pratiques, dans sa communication sur ses résultats, dans la priorisation des contributions à la société qu'il ou elle peut faire) de ses responsabilités collectives. Par exemple, la décision de fonder un nouveau journal, un nouveau domaine, ou une nouvelle section disciplinaire au niveau institutionnel ne peut que se prendre en collectivité, et non pas par un chercheur seul. Or l'impact sociétal que peuvent avoir des décisions au niveau collectif contient son risque de dilution des impressions de responsabilité de chaque individu qui constitue ce collectif (El Zein *et al*., 2019). Cela peut conduire à un détachement du chercheur, ou à une absence de réflexion éthique sur les conséquences de l'action collective auquel il contribue.

Un exemple que l'on pourrait discuter ici, et qui est directement en lien avec le champ des sciences cognitives, est celui de l'engouement récent pour l'apprentissage machine (parfois dit « profond » (en anglais, *deep learning*) lorsqu'il est réalisé via des méthodes initialement neuro-inspirées, appelées des réseaux de neurones formels à multiples couches dits « profonds », LeCun *et al*., 2015) appliqué au traitement de grandes quantités de données (« big data »)[13]. Suite à cet engouement, de multiples entreprises privées ont proposé ces cinq dernières années à des chercheurs de venir travailler chez elles, soit pour appliquer ces méthodes en tant que « data scientist », soit pour participer à des équipes de Recherche & Développement en leur sein. Si on peut se réjouir que cela ait ouvert de nouveaux débouchés pour les non-permanents du secteur académique (jeunes docteurs, post-doctorants, ingénieurs en CDD), on peut néanmoins déplorer que cela ait conduit un grand nombre de chercheurs permanents à quitter (au moins partiellement) leur poste académique pour rejoindre ces entreprises.

[13] Cet engouement pour l'apprentissage machine dit « profond » est brièvement abordé et resitué par rapport à d'autres courants de l'intelligence artificielle dans de Loor *et al*. (2015).

Or, ici se posent un certain nombre de questions éthiques aux niveaux individuel et collectif. Au niveau collectif tout d'abord, puisque ces embauches se sont faites à grande échelle, nous avons assisté à un manque soudain pour les universités : départ de nombreux professeurs qui n'assurent plus leur enseignement habituel ; déstructuration d'un certain nombre d'équipes de recherches universitaires qui n'atteignent plus la masse critique suffisante pour perdurer en tant qu'équipe ; réduction de la participation à des projets collaboratifs de recherches (notamment européens) dont le financement avait été obtenu avant le départ des chercheurs concernés. Si certaines grandes entreprises elles-mêmes se sont rendu compte d'un possible problème sur le long-terme du fait que le milieu académique allait avoir du mal à continuer de bien former la génération d'après (potentiellement recrutable par ces mêmes entreprises) (Sample, 2017), d'autres problèmes en découlent, comme l'affaiblissement momentané d'un domaine entier de recherche (celui de l'apprentissage machine) de façon trans-universitaire et transnationale, ou encore la concentration des experts de ce domaine hors du milieu de la recherche publique. Ceci a donc eu un coût au niveau collectif pour le milieu universitaire (public).

Certes, l'existence d'équipes de recherche dans le milieu privé peut être vue comme une diversification des contributions potentielles à la connaissance, et comme une utilisation de moyens financiers privés (en plus des moyens publics) pour contribuer à l'acquisition des connaissances. Néanmoins, pour rester bénéfiques pour la société, nous recommandons que ces contributions restent minoritaires par rapport à la recherche publique. En effet, un problème éthique important se pose lorsque pour un domaine de recherche donné, les équipes de recherche privées se retrouvent en position de suprématie par rapport aux équipes de recherche publique (notamment avec plus de moyens, plus d'effectifs et un accès à davantage de données), alors même que le milieu universitaire est en crise de moyens dans de nombreux pays depuis plusieurs années (Tugend, 2016), et que certaines des grandes entreprises qui embauchent ces chercheurs (comme les GAFAM) sont régulièrement accusées par les instances politiques des pays (Ducourtieux, 2018) ou par les associations de contribuer à cette crise des finances publiques en ne contribuant pas de façon juste à l'impôt au regard des profits qu'elles réalisent (Sample, 2017). De plus, cela conduit au déplacement du public vers le privé d'une grande partie de ceux qui sont parmi les mieux placés pour penser aux questions éthiques liées aux potentielles conséquences de ces recherches. Ceci pose un problème de conflit d'intérêt à grande échelle puisque les recherches sur les questions éthiques liées aux applications de l'apprentissage machine (notamment la question de la non-transparence du traitement des données personnelles et de l'influence sur les individus qui peut en résulter, voir à ce sujet les articles de Roth et Human et collaborateurs dans ce numéro) pourraient déboucher sur des conclusions (comme la recommandation de ne pas appliquer des algorithmes non-transparents sur des données personnelles, ou encore de ne pas développer des capacités d'autonomie décisionnelle létale en matière militaire pour les robots (Righetti *et al.*, 2018)) qui soient en contradiction avec les intérêts financiers de ces entreprises. Or, comme l'écrivait Dante (1314) : « là où les ressources de l'esprit s'unissent au mal vouloir et à la force, on ne peut trouver aucun recours. ». Pour l'instant, avec

l'engouement décrit plus haut, les forces de l'esprit, dont de nombreux collègues brillants qui essaient de faire au mieux leurs recherches dans les conditions favorables et confortables que leur offrent ces entreprises, se sont seulement unies à un cadre structurel qui n'est pas nécessairement une force néfaste, mais en tout cas une force dirigée principalement vers le profit plutôt que vers l'intérêt commun. Les entreprises privées étant le plus souvent organisées de manière pyramidale, il suffit ensuite d'une décision d'un petit nombre au sommet de la pyramide pour que les frontières de l'éthique soient franchies malgré toute la bonne volonté des individus ayant rejoint la pyramide. En effet, « la personnalité d'un homme ne pèse jamais très lourd face à la tyrannie des structures » (Halimi, 2010).

On se retrouve donc ici dans un cas où des décisions individuelles de chercheurs de partir du public vers le privé (pour avoir de meilleurs conditions de direction d'une équipe de recherche, moins d'obligations administratives absurdes, combinées à un meilleur salaire et à un accès à des ressources de calcul imbattables) ont des conséquences éthiques au niveau collectif : cela contribue à affaiblir encore davantage les universités alors mêmes que celles-ci auraient pu offrir de meilleurs conditions pour les équipes de recherches si les grandes entreprises ayant capté ces chercheurs n'avaient pas contribué de façon aussi importante à l'optimisation fiscale (pour ne pas parler d'évasion fiscale dans certains cas) au détriment des états (Sample, 2017).

Enfin, au-delà du clivage public/privé, on se retrouve de plus ici dans une situation de concentration accrue des moyens de recherche dans un petit nombre d'équipes de recherche (ici privée, mais la question s'applique aussi aux équipes des laboratoires publics) au détriment des autres équipes. Si l'on fait partie des « meilleurs » d'un point de vue bibliométrique, comment ne pas considérer qu'on « mérite » d'être à la tête d'une équipe encore plus grande avec encore plus de moyens, pour tenter de viser encore plus loin dans ses objectifs en termes de recherche, et aussi en termes de carrière ? Que l'on soit dans le public ou le privé, comment renoncer à demander la somme la plus élevée possible pour des demandes de financement, ou à renoncer à accepter plusieurs financements sollicités en parallèle, lorsque les réponses sont positives pour plusieurs, alors que les demandes multiples visaient initialement à maximiser les chances d'obtenir au moins un financement ? Et quelles conséquences cela a-t-il sur les autres équipes ? Que penser lorsque certains de nos collègues, « très bons » mais parfois en deçà d'un certain seuil purement bibliométrique, nous annoncent qu'ils se retrouvent parfois sans financement pour leur équipe de recherche pendant une certaine période ? Il est difficile au niveau individuel, parce que c'est au détriment de notre propre intérêt immédiat, de se souvenir que cette concentration des moyens est peut-être même néfaste au niveau collectif. En effet, il suffit de considérer par exemple que la rentabilité d'une trop forte somme investie sur les deux bras d'un seul chef d'équipe (*e.g.*, 5 M€) ne pourra jamais être la même qu'avec cette même somme mieux répartie parmi tous les dossiers jugés « très bons » (*e.g.*, 500 K€ pour dix équipes de recherche, moyennant éventuellement 10-20% de bonus pour le « meilleur » dossier). Toutes ces questions n'ont pas de réponses simples, mais elles méritent de ne jamais être écartées de nos esprits à toutes et

tous, pour éviter de se retrouver surpris(e) *a posteriori* par les conséquences de nos choix au niveau collectif.

On touche donc bien ici à un ensemble de questions éthiques à un niveau collectif, et qui touchent à ce que Daniel Andler (2019, même numéro), dans la suite de Heather Douglas (2014), appelle la responsabilité du chercheur en tant qu'être humain concernant tout ce qui touche aux applications ou aux conséquences à long-terme de ses choix de recherches.

La course à la publication semble maximiser une certaine productivité scientifique. Mais seulement à court-terme. Il semble évident pour de nombreux critiques de cette démarche, dont le physicien Serge Haroche (2019), que ceci va au détriment d'intérêts collectifs à long-terme, comme la possibilité de déboucher sur des découvertes non prévues, ou appuyées sur des hypothèses qui semblent aujourd'hui à contre-courant et donc difficilement publiables. Mais est-il possible de convaincre maintenant ceux qui bénéficient le plus du système actuel des effets néfastes d'une démarche utilitariste à court-terme ? Peut-on se permettre d'attendre les décennies nécessaires pour pouvoir mesurer si l'augmentation du nombre moyen de publications par chercheur aura éventuellement été accompagné de moins de changements de paradigmes, de plus de recherche incrémentale et de moins de grandes avancées, d'un plus grand taux de rétractations ? Peut-être même que cet examen pourrait déboucher sur la conclusion qu'au niveau purement comptable et restreint de la production scientifique, les effets collatéraux (rétractations, cas de fraude, concentration des financements, pressions psychologiques et abandons de la part de collègues que l'on jugerait très bons et méritants (Richardson, 2019 ; Susi *et al*., 2019), etc.) ne sont qu'un coût à payer pour une production scientifique accrue ou, au pire, maintenue au même niveau dans un contexte de forte compétition internationale (néanmoins, voir Morin, 2019). Mais cela serait-il acceptable pour autant ?

Ce serait oublier que cette pression à la publication aura conduit chacun d'entre nous à réduire le temps de la réflexion éthique, ou encore le temps consacré à la lecture d'autre chose que les articles et ouvrages scientifiques de notre domaine, qui pourraient pourtant nous aider à mieux penser, à mieux comprendre, à mieux maîtriser l'inscription de nos actes dans les perspectives sociétales à long-terme. Il en résulte que nous nous sentons démunis : il nous semble que nous n'avons aucun impact sur les transformations actuelles de la société, et en son sein, de l'organisation même des structures où se fait la recherche scientifique. Nous ressentons une baisse du sentiment d'agentivité (Haggard & Chambon, 2012). Ce qui veut dire une perte d'autonomie : ce n'est plus à chacun de penser à l'éthique, de ce qui dans notre travail peut être bon ou mauvais pour la société, on le pense pour nous. Et une déresponsabilisation : puisque les décisions se prennent ailleurs, nous avons l'illusion de ne pas en partager une part de responsabilité, et du coup l'illusion que nous pouvons simplement nous concentrer sur ce sur quoi nous excellons.

On voit ici à nouveau, mais cette fois-ci au niveau collectif, le besoin d'entretenir ces débats et réflexions sur l'éthique pour être penseurs actifs des multiples conséquences de nos actes de chercheurs et de citoyens sur la société.

Cette exploration de certains fondements philosophiques de l'éthique et de leurs liens possibles avec les données expérimentales récentes en sciences

cognitives suggère en tout cas que le temps de la décision est un facteur clef pour permettre à l'humain de prendre des décisions les plus conformes possibles avec son éthique. Le temps permet notamment la simulation mentale (et donc l'anticipation) des conséquences potentielles de l'action (Redish, 2016), que ce soit avant l'exécution de ses propres actions, ou dans la capacité à se simuler mentalement en train de réaliser des actions observées chez autrui (notamment grâce aux neurones miroirs, Gallese *et al*., 2004), ou encore de se plonger grâce à la fiction dans la simulation mentale d'une expérience subjective de situations non vécues (Rosat, 2012). Être confronté en amont à une connaissance de situations historiques ou fictionnelles où des personnes ont pu faire des actions que l'on jugerait non conforme à notre éthique (voir une discussion de quelques exemples historiques mêlant sciences et guerre dans Khamassi & Decremps, 2019) ne semble que pouvoir favoriser la simulation mentale en avance de l'action pour se rendre compte de son caractère non éthique. Ceci est particulièrement crucial pour pouvoir mieux anticiper des situations complexes où l'évaluation du caractère éthique n'est pas triviale et nécessite le temps de la réflexion. Ceci peut ainsi amener à une forme de connaissance préalable qui pourrait aider à prévenir la réalisation d'une action non éthique lorsque l'on est confronté soi-même à la décision d'agir ou de ne pas agir dans un certain sens.

REMERCIEMENTS

Les auteurs tiennent à remercier Virginie Beaucousin et Valerian Chambon pour leurs commentaires constructifs sur ce texte, Marie-Jo Lécuyer pour son édition, et tous les contributeurs à ce numéro pour les multiples réflexions qu'ils ont suscité et qui ont contribué à nourrir ce texte.

RÉFÉRENCES

Andler, D. (2019). L'éthique des sciences cognitives a-t-elle quelque chose de particulier ? *Intellectica*, *70*, 2019/1, pp. 41-62

André, J.-C. (2013). Towards a Socially Responsible Research (SRR) charter in Engineering Sciences at CNRS level. *International Journal of Techno-ethics*, *4*, 39-51.

Arendt, H. (1991). *Eichmann à Jérusalem : rapport sur la banalité du mal*. Paris, Éditions Gallimard.

Aristote [-330]. *Éthique à Nicomaque*. Traduction de Richard Bodéüs, Paris, Éditions Flammarion, 2004, 560 p.
(ISBN 978-2080709479, notice BnF no FRBNF39129950).

Atlan, H. (1999). *Les Étincelles du hasard*. Paris, Éditions Seuil, t. I.

Atlan, H. (2002). *La science est-elle inhumaine ? Essai sur la libre nécessité*. Paris, Éditions Bayard.

Atlan, H. (2018). *Cours de philosophie biologique et cognitiviste. Spinoza et la biologie actuelle*. Paris, Éditions Odile Jacob.

Baumard, N. & Sperber, D. (2007). La morale. *Terrain. Anthropologie & sciences humaines*, (*48*), 5-12.

Bavard, S., Lebreton, M., Khamassi, M., Coricelli, G. & Palminteri, S. (2018). Reference point and range-adaptation produce both rational and irrational choices in human reinforcement learning. *Nature Communications*, *9(1)*, 4503.

Bergson, H. [1896]. *Matière et Mémoire*. Paris, Presses Universitaires de France, Quadrige, 2009.

Bigenwald, A. & Chambon, V. (2019). Criminal Responsibility and Neuroscience: No Revolution Yet. *Frontiers in Psychology*, *10*:1406.

Bontems, V. (2018). La machine respectueuse. L'éthique des techniques de Simondon à l'ère des robots. *Rev. Fr. d'éthique Appliquée*, *1(5)*, 22-33.

Boraud, T. (2015). *Matière à décision*. Paris, CNRS Éditions, février 2015, 260 p.

Bouchard, C. (2002). Structures mises en place en matière d'éthique et de déontologie de la recherche scientifique au niveau mondial (UNESCO) au sein de l'Union européenne, et en France au niveau national et régional. Volume 1 : recensement, mission, composition, fonctionnement, publications. Rapport pour le Comité d'éthique pour les sciences du CNRS (COMETS) du 12 juillet 2002.

Chammat, M. & Giraud, S. (2019). L'éthique du nudge : pour un usage responsable d'une approche comportementale au service des politiques publiques. I*ntellectica*, *70,* 2019/1, pp. 83-96.

Charte nationale de la déontologie des métiers de la recherche (2015). Publiée le 26/01/2015 et ratifiée le 22/01/2019. https://www.hceres.fr/sites/default/files/media/downloads/2015_Charte_fran%C3%A7aise_IS.pdf (consulté le 24 juin 2019).

COMETS (2014). Promouvoir une recherche intègre et responsable : un guide. Rapport du Comité d'éthique pour les sciences du CNRS (COMETS), juillet 2014.

COMETS (2017). Pratiquer une recherche intègre et responsable : un guide. Rapport du Comité d'éthique pour les sciences du CNRS (COMETS), mars 2017. http://www.cnrs.fr/comets/IMG/pdf/guide2017-fr.pdf (consulté le 24 juin 2019).

Corvol, P. (2016). Bilan et propositions de mise en œuvre de la charte nationale d'intégrité scientifique. INSERM, https://www.inserm.fr/recherche-inserm/integrite-scientifique (consulté le 24 juin 2019).

Damasio, A.R. (1995). *L'erreur de Descartes : La raison des émotions*. Paris, Odile Jacob.

Dante, A. [1314]. *L'Enfer*. Extrait de La Divine Comédie. Traduction de Jacqueline Risset, Paris, Éditions Flammarion, 2010.

Déclaration de Singapour sur l'Intégrité de la Recherche (2010). https://wcrif.org/documents/313-ss-french/file (consulté le 24 juin 2019).

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B*, *Biological Sciences*, *308*(1135), 67-78.

Douglas, H. (2014). The moral terrain of science. *Erkenntnis*, *79(5)*, 961-979.

Ducourtieux, C. (2018). Taxation des GAFA : le Parlement européen recommande de moderniser l'impôt sur les sociétés. *Le Monde*, 19 février 2018. https://www.lemonde.fr/economie/article/2018/02/19/taxation-des-gafa-le-parlement-europeen-recommande-de-moderniser-l-impot-sur-les-societes_5259090_3234.html (consulté le 6 juillet 2019).

Dumas, G., Khamassi, M., N'Diaye, K., Foubert, L., Jouffe, Y. & Roth, C. (2012). Procès des Déboulonneurs de pub : et la liberté de (non) réception ? *Le Monde.fr*.

Durkheim, E. [1895]. *Les Règles de la Méthode sociologique*. Nouvelle Édition, Paris, Éditions Flammarion, 2010.

Ellul, J. (1977). *Le système technicien*. Paris, Calmann-Lévy.

El Zein, M., Bahrami, B. & Hertwig, R. (2019). Shared responsibility in collective decisions. *Nature human behaviour*, *1*.

European Science Foundation (2017). The European Code of Conduct for Research Integrity -2017. http://archives.esf.org/coordinating-research/mo-fora/research-integrity.html (consulté le 24 juin 2019).

Fleurance, P. (2018). Introduction à la journée « Agir et penser en complexité appelle à la raison ouverte et ouvrante : pourquoi aujourd'hui cette interpellation

collective ? », du 30 mars 2018, CNAM, Paris. Éditorial de l'*Interlettre Chemin Faisant du Réseau Intelligence de la Complexité*, n° *86*, Sept.-Oct.2018 (http://www.intelligence-complexite.org/fileadmin/docs/il86.pdf, consulté le 21 juin 2019).

Frith, C.D. & Haggard, P. (2018). Volition and the brain–revisiting a classic experimental study. *Trends in neurosciences*, *41(7)*, 405-407.

Gallese, V., Keysers, C. & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8(9)*, 396-403.

Gaussel, M. & Reverdy, C. (2013). Neurosciences et éducation : la bataille des cerveaux. Dossier d'actualité Veille et Analyses IFÉ, n° 86, septembre. Lyon : ENS de Lyon. En ligne : http://ife.ens-lyon.fr/vst/DA/detailsDossier.php?parent=accueil&dossier=86&lang=fr

Gouyon, P.-H. (2010). Présentation sur l'éthique lors de la journée des nouveaux entrants au CNRS.

Haggard, P. & Chambon, V. (2012). Sense of agency. *Current Biology*, *22(10)*, R390-R392.

Halimi, S. (2010). Peut-on réformer les États-Unis ? *Le Monde Diplomatique*, n°*670*, janvier 2010.

Haroche, S. (2019). Pourquoi la recherche française décroche ? France Culture, émission La grande table des idées, 26 mars 2019. https://www.franceculture.fr/emissions/la-grande-table-2eme-partie/pourquoi-la-recherche-francaise-decroche (consulté le 6 juillet 2019).

Hérodote, L. [-425]. *Enquête, Livres I à IV*. Traduction d'Andrée Barguet. Paris, Éditions Gallimard, Collection Folio/Classique, 1964.

C. Heyns. (2013). Annual report of the Special Rapporteur on extrajudicial, summary or arbitrary executions. United Nations Human Rights Council. http://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/23/47 (consulté le 5 juillet 2019).

Human, S., Neumann, G. & Peschl, M. (2019). [How] can pluralist approaches to computational cognitive modeling of human needs and values save our Democracies? *Intellectica*, *70*, 2019/1, pp 165-180.

Kahneman, D. (2011). *Thinking, fast and slow*. Londres, Macmillan.

Khamassi, M. & Decremps, F. (2019). Apprentissage de la démarche scientifique et de l'esprit critique : un enseignement de Sorbonne Université pour les étudiants d'aujourd'hui, citoyens de demain. In S. Bertezene & D. Vallat (eds.), *Guider la raison qui nous guide : Agir et penser en complexité*, Caen, France: Management Prospective Editions.

Khamassi, M. & Doncieux, S. (2016). Nouvelles approches en robotique cognitive. *Intellectica*, *65*, vol. 2016/1.

Khamassi, M. & Pacherie, E. (2019). Action. In D. Andler, T. Collins & C. Tallon-Baudry (éds.) *La cognition : du neurone à la société*. Paris, Gallimard, 2018.

Jeannerod, M. (2009). *Le cerveau volontaire*. Paris, Éditions Odile Jacob.

de Loor, P., Mille, A. & Khamassi, M. (2015). Intelligence artificielle : l'apport des paradigmes incarnés. *Intellectica*, *64*, 27-52.

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436.

Libet, B. (1992). Models of conscious timing and the experimental evidence. *Behavioral and Brain Sciences*, *15(2)*, 213-215.

Massé, R. (2009). Anthropologie des moralités et de l'éthique - Essai de définitions. *Anthropologie et Sociétés*, *33(3)*, 21-42. doi:10.7202/039679ar.

McCabe, D.P. & Castel, A.D. (2008). Seeing is believing: the effect of brain images on judgments of scientific reasoning. *Cognition*, *107(1)*, 343-352.

Mercier, H. & Sperber, D. (2017). *The Enigma of Reason. A New Theory of Human Understanding*, Bristol, Allen Lane, mars 2017, 416 pages.

Mille, A. (2019). Construction "orientée éthique" de l'environnement numérique. *Intellectica*, *70*, 2019/1, pp. 119-163.

Monier, C. & Sarti, A. (2018). Les neurosciences au sein des sciences de la cognition, entre neuroenthousiasme et neuroscepticisme. Introduction au dossier. *Intellectica*, *69*, 2018/1-2.

Morin, H. (2019). La France recule dans le classement mondial de la science. *Le Monde*, 5 juillet 2019. https://www.lemonde.fr/sciences/article/2019/07/04/la-france-recule-dans-le-classement-mondial-de-la-science_5485433_1650684.html (consulté le 5 juillet 2019).

Mugur-Schächter, M. (1997). Les Leçons de la mécanique quantique. Vers une épistémologie formelle. *Le Débat*, n° *94*, mars-avril 1997.

Osman, M. (2018). Persistent Maladies: The Case of Two-Mind Syndrome. *Trends in cognitive sciences*, *(22/4)*, 276-277.

Oullier, O., & Sauneron, S. (2012). Le cerveau et la loi : éthique et pratique du neurodroit. Centre d'analyse stratégique.

Pandian A. (2008). Traditions in Fragments: Inherited Forms and Fractures in the Ethics of South India. *American Ethnologist*, *35(3)*, 466-480.

Penavayre, M., Brun, C. & Boraud, T. (2019). Neurobiologie des jugements moraux, avancée épistémique ou voie sans issue ? *Intellectica*, *70*, 2019/1, pp. 63-82.

Pham, Q. C., Madhavan, R., Righetti, L., Smart, W. & Chatila, R. (2018). The Impact of Robotics and Automation on Working Conditions and Employment [Ethical, Legal, and Societal Issues]. *IEEE Robotics & Automation Magazine*, *25(2)*, 126-128.

de Pracontal, M. (2001). *L'imposture scientifique en dix leçons*. Paris, Éditions La Découverte.

Redish, A.D. (2016). Vicarious trial and error. *Nature Reviews Neuroscience*, *17(3)*, 147.

Renaudo, E., Girard, B., Chatila, R. & Khamassi, M. (2014). *Design of a control architecture for habit learning in robots. Biomimetic & Biohybrid Systems*, Third International Conference, Living Machines 2014, 8608, 249-260.

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*(7261), 263.

Richardson, H. (2019). University counselling services 'inundated by stressed academics'. BBC, 23 Mai 2019. https://www.bbc.com/news/education-48353331?SThisFB&fbclid...R0LqaRn2KkNtgXmfIY4cMPZtMkp2v9I65LWORJ3BghqotLpad9aTmcIMmM (consulté le 6 juillet 2019).

Righetti, L., Pham, Q.C., Madhavan, R. & Chatila, R. (2018). Lethal Autonomous Weapon Systems [Ethical, Legal, and Societal Issues]. *IEEE Robotics & Automation Magazine*, *25(1)*, 123-126.

Rosat, J.J. (2012). Éducation politique et art du roman. Réflexions sur 1984. La philosophie de la connaissance au Collège de France.

Roth, C. (2019). Algorithmic distortion of informational landscapes. *Intellectica*, *70*, 2019/1, pp 97-118.

Sample, I. (2017). 'We can't compete': why universities are losing their best AI scientists. *The Guardian*, 1 Nov 2017. https://www.theguardian.com/science/2017/nov/01/cant-compete-universities-losing-best-ai-scientists (consulté le 6 juillet 2019).

Science Europe (2015). Seven reasons to care about integrity in research. https://www.scienceeurope.org/wp-content/uploads/2015/06/20150617_Seven-Reasons_web2_Final.pdf (consulté le 24 juin 2019).

Soon, C.S., Brass, M., Heinze, H.J. & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature neuroscience*, *11(5)*, 543.

Spinoza, B. [1677] *Éthique.* Traduction de Bernard Pautrat, Paris, Éditions Seuil, 1988. Les numéros de pages indiqués au fil du texte sont tirés de la traduction de Roland Caillois, Paris, Éditions Gallimard, 1954.

Stengers, I., James, W. & Drumm, T. (2013). *Une autre science est possible ! Manifeste pour un ralentissement des sciences* suivi de *Le poulpe du doctorat*. Paris, Éditions La découverte.

Susi, T., Shalvi, S. & Srinivas, M. (2019). 'I'll work on it over the weekend': high workload and other pressures faced by early-career researchers. *Nature*, 17 Juin 2019. https://www.nature.com/articles/d41586-019-01914-z?fbclid=IwAR1A7tDdnFBsWWgjRAE7tfRfSUmFMNbR8mt4NA7xF43xm5wZbJjcxky6Kw8 (consulté le 6 juillet 2019).

Thaler, R.H. & Sunstein, C.R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, Yale University Press.

Thiaw-Po-Une, L. (2006). *Questions d'éthique contemporaine*. Paris, Éditions Stock.

Tugent, A. (2016). How Public Universities Are Addressing Declines in State Funding. *New-York Times*, 22 June 2016. https://www.nytimes.com/2016/06/23/education/how-public-universities-are-addressing-declines-in-state-funding.html (consulté le 6 juillet 2019).

Viejo, G., Khamassi, M., Brovelli, A. & Girard, B. (2015). Modelling choice and reaction time during arbitrary visuomotor learning through the coordination of adaptive working memory and reinforcement learning. *Frontiers in Behavioral Neuroscience*, *9*, 225.

Weisberg, D.S., Keil, F.C., Goodstein, J., Rawson, E. & Gray, J.R. (2008). The seductive allure of neuroscience explanations. *J. Cogn. Neurosci.*, *20(3)*, 470-477.