

# Mémoire et conscience, un essai à la croisée de l'intelligence artificielle, des neurosciences cognitives et de la philosophie

Yann Cherdo

## I. PRÉAMBULE

Cet essai constitue une partie de mon mémoire de doctorat dans le domaine de l'Intelligence Artificielle (IA) bio-inspirée<sup>1</sup> où je suis entré notamment animé de curiosité au sujet de la nature humaine.

## II. DÉFINITIONS

Voici les définitions utilisées dans cet essai portant sur certains termes clés et récurrents:

Le “soi” ou le “moi” constituent tout ce qui permet à un individu de se définir comme étant lui-même et le distinguant indépendamment du reste du monde et d'autrui.

Le libre arbitre constitue la capacité d'un individu à faire un choix et à maîtriser sa destinée indépendamment du reste du monde et d'autrui.

La nature humaine se réfère à tout ce qui peut définir et caractériser l'être humain, de sa structure physique à ses comportements sociaux, cognitifs, intellectuels et ses objets mentaux aussi abstraits soient-ils.

La réalité, ou la vérité, constituent dans cet essai les objets de la science ou plus largement de la pensée, qu'ils soient considérés comme objectifs, universels ou subjectifs.

La raison est une faculté qu'a l'esprit humain d'organiser ses relations avec le réel<sup>2</sup>. Elle s'opère notamment au travers de modes de pensée comme le déterminisme, la causalité, la logique ou le système axiomatique.

Je proposerai une définition de la conscience dans la section VII à partir d'un ensemble de considérations la concernant et qui sont développés dans la section VI.

## III. MOTIVATION ET SUJET

L'IA de par son nom renvoie naturellement chacun d'entre nous à un questionnement d'ordre personnel. Dans la culture occidentale l'intelligence est associée à la nature humaine, et semble faire partie de la définition du “moi”, depuis le début de l'histoire en passant par la religion et la période des lumières<sup>3</sup>. Mais alors qu'est-ce que ce domaine de l'IA a à dire ou à apporter sur le sujet de la nature humaine?

La nature humaine se révèle par exemple au travers de l'histoire de l'évolution des techniques et des machines.

Aujourd'hui, l'IA s'inscrit dans la continuation de ce processus et reflète également les activités humaines qu'une machine peut reproduire comme l'écriture ou la conduite automobile grâce au véhicule autonome. Ce processus dépasse aussi les seules activités dont l'humain est capable par lui-même et permet de l'augmenter avec notamment l'aviation ou le traitement de données massives.

Aussitôt que l'humain invente un système permettant de satisfaire une tâche à sa place, cette tâche peut apparaître comme d'autant moins humaine ou faisant d'autant moins partie constitutive de la nature humaine. Plus les machines reproduisent des activités humaines, moins ces activités semblent nous éclairer directement sur la nature de l'humain. Cette idée vaut aussi pour les tâches réalisées par l'IA. De façon similaire, mes lectures en neuroscience cognitive relatives aux mécanismes corticaux permettant l'apprentissage, la mémoire ou l'agentivité<sup>4</sup> m'ont amené à questionner la place de ces propriétés cognitives dans la nature humaine.

La relation intime entre l'IA et les neurosciences a été un fil conducteur tout au long de ma thèse de doctorat. Par exemple j'ai exploité et modélisé des neurones impulsifs et les caractères parcimonieux de l'activité et de la connectivité du cortex humain pour les modèles d'IA que j'ai développés. Ceci notamment dans le but de réduire leur consommation énergétique<sup>5</sup>.

Plutôt que de parler du concept d'intelligence, j'aimerais me concentrer sur ceux de la conscience et du “moi” puisqu'ils sont aussi des sujets prégnants et couramment employés dans les discussions portant sur l'IA. Il pourrait être proposé que le questionnement d'ordre personnel mentionné précédemment amène à une définition de la conscience indépendante de la mémoire. C'est cette thèse que je propose de développer dans cet essai.

Comme il en sera question plus bas dans cet essai, s'attaquer frontalement à la définition de la conscience peut se montrer difficile. En conséquence une méthode utilisée en physique m'a semblé appropriée. Elle consiste à définir une particule non pas par sa structure ou ses constituants mais par ses symétries ou transformations invariantes. En effet, une compréhension complémentaire de la conscience pourrait émerger de la définition de ses frontières conceptuelles. Dans

<sup>1</sup><https://www.college-de-france.fr/fr/agenda/seminaire/modeles-multi-echelles-et-reseaux-de-neurones-convolutifs/intelligence-artificielle-et-intelligence-naturelle-vers-ia-bio-inspiree>

<sup>2</sup><https://www.cnrtl.fr/definition/raison>

<sup>3</sup>Histoire de la philosophie occidentale, Russel, les belles lettres, 2011

<sup>4</sup>Agency theory, Shapiro, 2005

<sup>5</sup>Time series prediction and anomaly detection with recurrent spiking neural networks, Y Cherdo, B Miramond, A Pegatoquet, 2023 International Joint Conference on Neural Networks (IJCNN)

cet essai je propose que les neurosciences cognitives puissent permettre d'esquisser une telle définition et en donne un exemple avec le concept de la mémoire dans la section VIII.

#### IV. LA COMPRÉHENSION PAR LA RAISON, LE SYSTÈME AXIOMATIQUE ET LE SYSTÈME DE CAUSALITÉ

Une éducation classique, moderne et occidentale en physique et mathématiques peut forger une puissante capacité de raisonnement et d'induction permettant la résolution de problèmes complexes. La méthode de résolution se base le plus souvent sur une compréhension logique ou causale d'un système, édiflée parfois par corrélations. Le raisonnement axiomatique est une clé de cette méthode: s'appuyant sur un ensemble d'hypothèses l'on déduit une théorie ou connaissance d'un système et la mettons à l'épreuve de tests empiriques et de l'analyse critique de la communauté scientifique.

Des modèles bien établis tels que les lois de Kepler ou de Newton sont valides seulement dans leur système axiomatique, tout comme les mathématiques. De ce fait, une vérité universelle semble difficile à atteindre et il est douteux de prétendre pouvoir jamais parfaitement comprendre ce qu'est la conscience. Il pourrait donc être difficile de l'implémenter dans une IA. A cet égard le philosophe Platon est une référence pour montrer la fragilité de l'humain lorsqu'il prétend détenir une connaissance universelle à partir d'un raisonnement<sup>678</sup>.

En découle que le raisonnement ne peut aboutir à une certitude et que le besoin de cette dernière répond à d'autres motivations. Une des fins du raisonnement pourrait être non pas la certitude mais bien le doute. Le doute et la curiosité sont vecteurs de progrès et semblent participer à la joie et l'équilibre d'une activité intellectuelle comme la science.

Néanmoins, une compréhension plus fine de la conscience peut être obtenue grâce notamment à l'expérience que l'on en fait, aux sciences humaines et aux neurosciences. Ainsi, des définitions limitées de la conscience peuvent être proposées et implémentées dans une IA.

#### V. DÉTERMINISME ET INDÉTERMINISME, DEUX HYPOTHÈSES POUR ABORDER LE "MOI" ET LA CONSCIENCE

Les concepts de conscience et de "moi" peuvent être perçus comme singuliers. S'ils pouvaient être expliqués par un raisonnement axiomatique, utilisant des relations de causalité tout comme les neurosciences expliquent certains comportements cognitifs, que deviendraient ces concepts et en particulier le "moi"? Si un modèle déterministe pouvait prédire les actions, les choix ou la pensée d'un individu que resterait-il de son "moi" et de son libre arbitre? Et de quoi aurions-nous conscience? Au fur et à mesure que je lisais des ouvrages et travaux de recherche notamment en

neuroscience cognitive à des fins d'exploitations de certains mécanismes corticaux pour l'IA, il m'a semblé de moins en moins trivial que la mémoire et l'apprentissage soient des composants constitutifs du "moi". Ces concepts, quoique présentant un aspect souvent statistique, semblaient soudain relever plus d'un système mécanique déterministe que d'un être doué de conscience et du sens du "moi".

Si par exemple la communauté scientifique défendait qu'un tel modèle déterministe puisse exister, cela impliquerait une perte d'agentivité pour tout individu avec des conséquences probablement singulières sur l'organisation sociale et ce concept de "moi" tel qu'il a été forgé depuis l'antiquité en occident. Justice, démocratie et libre arbitre devraient alors surement être redéfinis. Il peut aussi être envisagé que dans l'hypothèse déterministe alors le libre arbitre, le "moi" et la conscience pourraient être des erreurs conceptuelles causées par une compréhension limitée du monde.

Bien que les neurosciences expliquent de façon déterministe certains comportements cognitifs, il est possible de choisir l'hypothèse où le libre arbitre, le "moi" et la conscience existent bien. De nombreux arguments peuvent être trouvés pour la supporter notamment dans le fonctionnement des sociétés actuelles, les arts, la philosophie et les sciences. La simple sensation permanente du "moi" et d'une expérience objective, que le lecteur peut éprouver maintenant, fait partie de ces arguments. Dans ce qui suit je propose d'utiliser cette hypothèse.

#### VI. DÉFINITIONS DE LA CONSCIENCE

Ayant proposé de chercher une meilleure compréhension de la conscience, il semble nécessaire de rappeler des définitions ou des approches existantes. Elles sont complexes et varient d'un domaine ou d'un auteur à l'autre.

##### A. Psychologie

En psychologie<sup>9</sup> la conscience est, par exemple, associée à la dualité qui apparaît pendant les premières années d'existence d'un humain. En se percevant elle-même une personne en quelque sorte se dédouble, d'où découle notamment le concept de problème. En effet, l'idée de problème semble inexistante si un individu ne pouvait pas se projeter dans plusieurs possibilités. Par exemple la notion de problème peut apparaître lorsqu'un individu regrette un choix passé. Cela lui est permis parce qu'il peut se projeter dans un autre choix qui d'après lui aurait pu le mener à une meilleure fin. Il compare bien plusieurs versions de lui-même et se faisant semble se dédoubler. A noter que ce mécanisme de conceptualisation de soi, de souvenir d'évènements passés et de leur projection dans des évènements hypothétiques sont permis par la mémoire. La notion de temps en est aussi le résultat puisque sans mémoire l'individu ne pourrait se souvenir de qui il est et donc de se projeter dans le passé ou le futur. Il vivrait alors dans la seule instantanéité. A noter que cela ne l'empêcherait peut être pas de faire l'expérience

<sup>6</sup>Gorgias, Platon oeuvres complètes tome I, traduction par Léon Robin, Editions Gallimard, bibliothèque de la pléiade, 1950

<sup>7</sup>Phèdre, Platon oeuvres complètes tome II, traduction par Léon Robin, Editions Gallimard, bibliothèque de la pléiade, 1950

<sup>8</sup>Théétète, Platon oeuvres complètes tome II, traduction par Léon Robin, Editions Gallimard, bibliothèque de la pléiade, 1950

<sup>9</sup>The spiritual problem of modern man, Jung, 1928

de la conscience ou même du “moi” bien qu’il ne soit plus possible de construire ce dernier autour du souvenir de l’individu.

L’idée que cette dualité est à l’origine du concept de problème peut se retrouver dans le mot diable dont le préfixe “di” signifie deux ou double dans son origine grecque ancienne<sup>10</sup>. Chez Homère<sup>11</sup> et Platon<sup>12</sup>, Ulysse est craint parce qu’il est capable de feindre et de manipuler, il est double. A contrario, Achille est adoré notamment parce qu’il est cohérent et fidèle à lui-même, il est prédictible et incapable d’utiliser la ruse comme Ulysse.

### B. *Étymologie*

L’étymologie du mot conscience provient du mot latin “conscientia”<sup>13</sup><sup>14</sup>, contraction des termes “cum”<sup>15</sup><sup>16</sup> et “scientia”<sup>17</sup>, pouvant être traduit par “connaissance en commun” ou littéralement par “avec connaissance”. La différence avec le mot “scientia” seul pourrait provenir de l’interaction entre le “moi” d’un humain, ou sa conscience, et la connaissance. Il pourrait sembler trop simple qu’un robot ou une IA puisse en fait avoir de la connaissance en utilisant une unité de mémoire chargée d’information car il ne semble pas y avoir de différence fondamentale entre cette unité de mémoire, un bras robotisé, le processeur ou encore le silicium le constituant, tout cela est de la matière organisée. En conséquence l’étymologie du mot conscience semble invoquer une approche plus singulière de la connaissance liée à sa création même et à sa manifestation dans l’esprit humain.

Il est intéressant de constater ici une certaine perversion du langage, comme dans l’usage des mots mémoire, connaissance ou conscience de façon indifférenciée entre un humain et un robot. En utilisant les mots “mémoire” ou “connaissance” pour décrire un robot, par exemple dans la phrase “Ce robot est doté d’une mémoire dans laquelle de la connaissance a été téléchargée.”, ce robot se voit implicitement attribué à travers ces définitions des qualités humaines qu’il ne possède pas. En effet, les mots mémoire et connaissance pour un humain vont très au-delà de ce qu’ils désignent dans un robot, qui n’est que matière organisée. Des concepts, comprenant notamment des idées complexes et liées à l’humain, reflets et produits de sociétés et langages en formation durant des milliers d’années, sont ici projetés sur un système contemporain qu’est le robot.

Enfin, ces considérations appellent à essayer de comprendre la relation entre le “moi” et la conscience d’une part et le concept de matière organisée d’autre part.

### C. *Philosophie*

Dans les oeuvres de Platon, et sauf erreur de ma part, le mot conscience n’apparaît pas. Peut être était-ce un terme non existant dans l’ancien grec ou une conséquence de sa traduction. Un autre terme y est par contre fréquemment utilisé et constitue un concept clé de ses oeuvres: l’âme<sup>18</sup>. La description de l’âme chez Platon semble en fait très similaire à celle de la conscience. La définition qu’il donne de l’âme est cohérente avec les définitions précédentes de la conscience de par la dualité qu’elle propose entre le “moi” ou l’âme et le monde physique, corps inclus. L’âme y est singulière puisqu’elle est présentée comme immatérielle et ne répond pas vraiment aux lois de la physique, ce qui semble-t-il permet au libre arbitre d’exister. Cette approche de la dualité corps et âme, ou parfois dite corps et esprit, n’est pas présente seulement dans les oeuvres de Platon mais fait partie d’une tradition dont on peut trouver certaines origines chez divers auteurs antiques comme Homère ou Hérodote ainsi que dans le christianisme.

Dans cette dualité, une frontière nette est tracée entre l’âme ou le “moi” et le monde physique où le corps peut être perçu comme une frontière floue entre un individu et le reste du monde.

Cette approche singulière de la conscience qui défie le raisonnement a notamment fait l’objet de travaux du philosophe Michel Bitbol<sup>19</sup>. La conscience y est proposée comme une expérience individuelle: “l’expérience consciente”. Comprendre la conscience apparaît comme un paradoxe puisque l’acte de comprendre lui-même est une activité de la conscience...

Cette difficulté circulaire du raisonnement ne semble par ailleurs pas propre au concept de la conscience puisque, par exemple, le néant ne peut être défini explicitement sous peine que l’objet de cette définition ne soit pas le néant. Ou bien l’affirmation “Tous les humains sont des menteurs.” aboutit à une circularité du raisonnement.

### D. *Neuroscience*

En neuroscience<sup>20</sup><sup>21</sup> la conscience est associée à un espace de travail global qui échange des stimulus avec des neurones ou des assemblées de neurones dits inconscients. Cette proposition est notamment supportée expérimentalement par la mise en évidence d’activités neuronales situées dans cet espace de travail, en particulier dans la région préfrontale du cortex, lorsque l’individu étudié effectue une tâche consciemment. Cet espace de travail permettrait d’accéder à la conscience. Une dualité est observée entre d’une part un certain ensemble de neurones présentant une activité globale et corrélée aux actes conscients et d’autre part le reste des neurones constituant le cerveau.

<sup>18</sup>Phédon, Platon oeuvres complètes tome I, traduction par Léon Robin, Editions Gallimard, bibliothèque de la pléiade, 1950

<sup>19</sup>Aux sources de la conscience, <https://lejournal.cnrs.fr/articles/aux-sources-de-la-conscience>, 2014

<sup>20</sup>What is Consciousness?, Wallace, 2004

<sup>21</sup>Global workspace theory of consciousness: toward a cognitive neuroscience of human experience, Baars, 2005

<sup>10</sup><https://www.cnrtl.fr/definition/di->

<sup>11</sup>Iliade - Odyssée, Homère, traduction par Victor Bérard et Robert Flacelière, Editions Gallimard, bibliothèque de la pléiade, 1955

<sup>12</sup>Le petit Hippias, Platon oeuvres complètes tome I, traduction par Léon Robin, Editions Gallimard, bibliothèque de la pléiade, 1950

<sup>13</sup><https://www.littre.org/definition/conscience>

<sup>14</sup><https://gaffiot.fr/conscientia>

<sup>15</sup><https://gaffiot.fr/con>

<sup>16</sup><https://gaffiot.fr/cum>

<sup>17</sup><https://gaffiot.fr/scientia>

A noter que les progrès en neurosciences dans la compréhension de la conscience sont importants mais laissent encore une certaine incertitude quant au lien entre le mécanisme neuronal dit conscient, ou d'accès à la conscience, et le phénomène de cette expérience consciente vécue par chaque individu.

## VII. UNE PROPOSITION DE DÉFINITION DE LA CONSCIENCE

Je n'ai pas l'expertise suffisante pour apporter de plus amples références quant à la définition de la conscience. Il est possible d'en trouver davantage dans les domaines des neurosciences, de la psychologie et de la phénoménologie de l'esprit.

A partir des définitions précédemment exposées je propose en synthèse d'en retenir la suivante: par conscience, j'entends le système qui permet à un humain de faire l'expérience du "moi", du libre arbitre et de sa dualité.

Dans cet essai, les concepts de conscience et de "moi" sont parfois présentés côte à côte. Cela reflète l'intime relation qui semble exister entre ces deux concepts puisque "l'expérience consciente" pourrait bien être avant tout une expérience du "moi". C'est bien un sujet qui prend conscience d'un objet.

## VIII. RÉSULTAT PORTANT SUR LA CONSCIENCE ET LA MÉMOIRE

En neurosciences cognitives<sup>222324</sup>, il existe plusieurs modèles permettant de décrire les mécanismes de la mémoire. La plasticité neuronale explique en partie la capacité particulière des mammifères, et en particulier de l'humain, d'apprendre tout au long de la vie d'un individu et ainsi de transmettre une culture. Tandis que les espèces reptiliennes évoluent et s'adaptent plutôt grâce aux mutations génétiques. En particulier, la "plasticité fonction du temps d'occurrence des impulsions" (STDP), qui dérive de la loi de Hebb, décrit cette plasticité à l'échelle de la synapse qui permet à un neurone de transmettre un potentiel électrique à un autre. La dynamique de cette plasticité ainsi que le fonctionnement des neurones expliquent par exemple pourquoi l'humain est capable d'apprendre des séquences très longues comme des textes ou des morceaux de musique, parfois même inconsciemment. L'activité parcimonieuse des neurones permet d'expliquer la grande robustesse de la mémoire ainsi que sa grande capacité de rétention d'information. La hiérarchie et l'organisation du cerveau expliquent les capacités d'abstraction et de corrélation entre plusieurs sens comme la vue et le toucher et ainsi de mémoriser et représenter le monde à la façon des humains. Ces mécanismes rendent un individu capable de considérer ses contraintes compte tenu de son expérience vécue et de ce qu'il en a retenu, et ainsi de peser dans ses décisions. D'autres mécanismes comme la compétition (le

neuro-darwinisme), l'attention, les interactions avec les parties plus anciennes du cerveau pourraient aussi être développées.

Enfin, il est intéressant de noter que le cortex humain, lieu privilégié de la plasticité neuronale qui concentre environ 80% des neurones du cerveau, est une mémoire mais aussi la matrice qui opère ou contraint majoritairement les sens, le contrôle moteur, le langage, l'imagination, etc.

Ces modèles de neurones et d'apprentissage sont en partie utilisés aujourd'hui dans les modèles d'IA.

Dans les sections V et VI-A, les relations entre le "moi" et la mémoire ou le déterminisme ont été questionnées. Il semble intuitif que la mémoire participe à la définition du "moi" grâce aux souvenirs mais aussi aux acquis culturels comme le langage et les relations sociales. Pourtant, les mécanismes de la mémoire précédemment décrits sont déterminés et ne rendent pas compte de la conscience. De plus, la définition de la conscience repose sur des concepts en partie contraires au déterminisme comme le libre arbitre ou le "moi". Paradoxalement, définir le "moi" grâce à la mémoire et ses mécanismes enferme le "moi" dans un déterminisme qui en fait la négation comme déjà mentionné dans la section V. Cela révèle bien deux formes possibles du moi, celui qui repose sur la mémoire et l'environnement d'un individu, et le "moi" qui semble associé à l'expérience consciente.

En conséquence, il apparaît que la mémoire n'est pas un élément constitutif du "moi" ou de la conscience.

Il peut être troublant de penser que la conscience ne dépendrait pas de la mémoire. En effet, comment nous définirions nous ou comment agirions-nous en tant qu'individu sans la mémoire? Que seraient les sensations, la volonté et l'activité intellectuelle sans la mémoire? Ces questions semblent toutes donner quelques indices quant à la relation de ces concepts avec la conscience.

A partir de ces arguments il est possible de questionner ces relations ainsi que certaines conceptions, ou parfois certitudes, portant sur la conscience.

Il convient de dire que le concept de mémoire n'est peut être pas réductible aux seuls modèles proposés en neuroscience et mentionnés précédemment et que d'autres relations pourraient exister entre mémoire et conscience. Cependant ces différents modèles rendent étonnamment bien compte du fonctionnement de la mémoire chez l'humain.

## IX. DIFFICULTÉS DE L'INDÉTERMINISME

Ayant repoussé à plusieurs reprises le déterminisme de la définition de la conscience, que dire de l'indéterminisme? Bien qu'il soit utilisé dans le langage et dans les sciences, son existence n'est pas certaine.

La physique quantique a été utilisée notamment pour élaborer des générateurs de nombres aléatoires<sup>25</sup>. Sauf erreur de ma part, il existe bien des générateurs de nombres pseudo-aléatoires et non purement aléatoires. En effet, dans le livre référencé ci-dessus, il est expliqué que la réalisation

<sup>22</sup>Neuronal dynamics: From single neurons to networks and models of cognition, Gerstner, 2014

<sup>23</sup>On Intelligence, Hawkins, 2004

<sup>24</sup>The cerebral code: Thinking a thought in the mosaics of the mind, Calvin, 1998

<sup>25</sup>L'Impensable Hasard: Non-localité, téléportation et autres merveilles quantiques, Gisin, 2012

matérielle de l'expérience du jeu de Bell sensée prouver la nature aléatoire du générateur est elle-même difficile et pourrait être imparfaite. Dans un tel cas, un résultat positif perdrait sa signification. L'auteur défend que, bien qu'aillant obtenu des résultats positifs, son équipe n'était pas en mesure de démontrer que l'expérience était parfaite. Il conclut donc qu'il n'est pas certain que les nombres produits par un générateur quantique soient bien indéterminés.

Par ailleurs, le principe d'indétermination d'Heisenberg énonce qu'il est impossible de connaître simultanément la fonction d'onde et la position d'une particule. Mais le fait que l'on ne puisse pas les connaître simultanément ne veut pas dire que l'indéterminisme existe.

L'existence du choix comme conséquence du libre arbitre fait partie de ces difficultés. Une proposition de preuve expérimentale pourrait être de confronter un même humain plusieurs fois au même choix à partir de conditions parfaitement identiques entre chaque essai. Cette expérience semble irréaliste puisqu'elle nécessiterait la mise en oeuvre de la simultanéité ou de la réversibilité du temps.

#### X. DIFFICULTÉS DU DÉTERMINISME ET DE LA RAISON

Il ne serait pas surprenant que le déterminisme échoue à permettre d'expliquer la conscience. Dans l'histoire des sciences, des convictions et certitudes ont dominé puis ont été remplacées par d'autres tout en participant au progrès de la connaissance. Par ailleurs, la science semble découvrir plus de mystères que de certitudes à mesure qu'elle avance.

Le déterminisme ne pourrait permettre qu'incomplètement la conceptualisation d'une connaissance. Il pourrait aussi être une conséquence et l'image du système physique neuronal. Dans ce cas le concept de déterminisme et ce qui en découle pourrait être réduit à un phénomène physique, celui du cerveau humain, en plaçant la conscience comme spectatrice ou peut être utilisatrice.

Erwin Schrödinger<sup>26</sup> a questionné les modes de raisonnements scientifiques en les supposant en partie causes des difficultés conceptuelles singulières rencontrées en physique quantique. Il propose que ces modes de raisonnements ne sont pas universels et sont le produit de leur histoire depuis l'antiquité en Europe occidentale. Ainsi, la façon de raisonner est elle-même une hypothèse qui conditionne et fait partie de la connaissance produite.

En physique, le concept selon lequel la matière possède une masse s'est par exemple ancré dans l'inconscient collectif et la culture occidentale grâce au consensus scientifique dont il a bénéficié pendant des siècles. Cependant, un autre concept est apparu selon lequel la matière ne posséderait pas de masse. Par exemple, dans ce nouveau concept, les particules sans masse acquièrent de la masse grâce à leur interaction avec leur champ associé<sup>27</sup>.

Ettore Majorana<sup>28</sup> a questionné le caractère déterminé du monde physique. Le concept d'aléatoire peut en effet être

perçu comme un moyen d'approximer un système insuffisamment compris pour le déterminer. Cet argument suppose que le système est effectivement déterminé. Par exemple, le mouvement brownien peut être perçu soit comme un mouvement indéterminé et indéterminable soit comme un mouvement déterminé mais difficilement déterminable. Il propose que l'aléatoire pourrait être un constituant du monde physique et non plus une approximation conceptuelle résultant d'un manque de connaissance. Ce faisant, il offre un moyen de questionner le déterminisme et la façon dont est utilisée la raison dans les sciences.

La compréhension de la raison elle-même n'est pas triviale. Son fonctionnement est complexe et peut mener à des paradoxes comme le montrent les travaux du chercheur Hugo Mercier<sup>29</sup>: "La raison ne sert pas tant à guider nos choix qu'à inférer des raisons pour les justifier."

Cela dit l'utilité du déterminisme et les progrès qu'il a permis sont peu questionnables.

#### XI. RÉALISME ET CONSTRUCTIVISME

Ayant confronté l'idée que la mémoire puisse être constitutive du "moi" ou de la conscience avec la manière dont les neurosciences expliquent le fonctionnement de la mémoire, il m'a semblé important d'introduire ici un parallèle avec le livre de Jean Pierre Changeux et Alain Connes<sup>30</sup> portant notamment sur le réalisme et le constructivisme.

Dans ce livre, le mathématicien Alain Connes décrit son intuition selon laquelle les mathématiques sont un langage permettant de découvrir une réalité indépendante de l'humain, de sa culture ou de sa pensée. Cette intuition a pris naissance dans l'importante robustesse du contenu des mathématiques et dans la façon dont ce contenu évolue au travers des cultures, de l'histoire et de consensus scientifiques internationaux. A cet égard, cette idée peut être associée au réalisme selon Platon qui propose l'existence d'une réalité, ou vérité, qui dépasse l'humain et en est indépendante.

De son côté, le neuro-biologiste Jean Pierre Changeux propose que cette intuition pourrait être fautive et que cette robustesse si convaincante du contenu des mathématiques, objet d'un consensus presque unique dans l'histoire de l'humanité, serait le résultat et l'image de la structure du cerveau humain et de ses fonctions cognitives. Cet argument se nourrit notamment des neurosciences qui offrent aujourd'hui des explications biologiques et physiques des fonctions cognitives humaines telles que la logique, l'imagination ou l'inférence qui semblent nécessaires pour faire des mathématiques. Les concepts d'ordre, de causalité et de stabilité cognitive, ou d'homéostasie, utilisés pour décrire le cerveau humain semblent bien être le résultat de sa structure biologique. Cet argument peut être associé au constructivisme qui propose que la réalité est une construction humaine.

<sup>26</sup>'Nature and the Greeks' and 'Science and Humanism', Schrödinger, 1996

<sup>27</sup>The Particle at the End of the Universe: How the Hunt for the Higgs Boson Leads us to the Edge of a New World, Carroll, 2013

<sup>28</sup>The value of statistical laws in physics and social sciences, Majorana and Mantegna, Springer, 2006

<sup>29</sup>Notre raison est-elle rationnelle?, <https://lejournal.cnrs.fr/articles/notre-raison-est-elle-rationnelle>, 2017

<sup>30</sup>Matière à pensée, Changeux and Connes, Odile Jacob, 1989

La confrontation de ces deux modes de pensée permet notamment de mieux comprendre et de définir les frontières de l'intuition d'Alain Connes. Suivant la méthode de la dialectique, arguments après exemples, le dialogue mené par ces deux auteurs offre une compréhension plus claire de ce que sont ou non les mathématiques, et par la même de cette intuition.

A noter que constructivisme et réalisme ne semblent pas incompatibles. Tout en supposant qu'il existe une réalité indépendante de l'humain il est permis de supposer que la connaissance soit un produit de la pensée et du raisonnement humains dont l'objet est cette réalité. La connaissance serait alors bien dépendante de l'humain tandis que la réalité qu'elle tend à approcher serait indépendante de l'humain. Platon défendait d'ailleurs que la connaissance, ou la vérité, se trouve en "soi" et n'est accessible que par le moyen de "soi" grâce au courage, à la sagesse et à la dialectique notamment<sup>31,32,33</sup>. Dans cette vision la connaissance est donc bien le produit de l'être humain qui vise la vérité, qui existe selon Platon, mais qui lui est inatteignable.

Dans cet essai, je tente avec mes modestes moyens de structurer une argumentation de façon similaire à ce livre de Jean Pierre Changeux et Alain Connes, repoussant les frontières de concepts complexes et intuitifs comme la conscience ou le "moi" en utilisant les neurosciences et en confrontant des idées ancrées dans la culture comme l'âme ou le libre arbitre avec le matérialisme.

## XII. CONSCIENCE ET OBJECTIVITÉ DANS LES SCIENCES

Le travail que j'ai pu produire tout au long de mes études scientifiques, de ma thèse et au sein des postes d'ingénieur que j'ai occupés semble notamment permis par ma volonté d'objectivité et le mécanisme de ma conscience.

En science l'acte de résoudre un problème s'appuie souvent implicitement sur l'indépendance entre le sujet, le scientifique, et l'objet du problème. Grâce à cette indépendance une solution objective ou une compréhension de la réalité en jeu peut émerger. La solution peut être considérée comme objective et la connaissance acquise est parfois perçue comme indépendante de l'humain, ce qui semble hypothétique. Il est en effet difficile de défendre que le cheminement logique de la pensée du scientifique soit indépendant de sa culture, son activité neuronale ou de sa conscience, et ce malgré l'effort important et singulier des sciences pour rendre son contenu universel. Cette hypothèse de séparation entre sujet et objet a entre-autres été étudiée par le philosophe Edgar Morin<sup>34</sup>. Tout comme Erwin Schrödinger<sup>35</sup> questionna le substrat culturel de la science moderne, cette hypothèse pourrait handicaper la science si elle n'était pas questionnée et peut être même intégrée en tant qu'hypothèse claire de la méthode scientifique.

<sup>31</sup>Ménon, Platon oeuvres complètes tome I, traduction par Léon Robin, Editions Gallimard, bibliothèque de la pléiade, 1950

<sup>32</sup>Phèdre, Platon oeuvres complètes tome II, traduction par Léon Robin, Editions Gallimard, bibliothèque de la pléiade, 1950

<sup>33</sup>République, Platon oeuvres complètes tome I, traduction par Léon Robin, Editions Gallimard, bibliothèque de la pléiade, 1950

<sup>34</sup>Introduction à la pensée complexe, Edgar Morin, 2015

<sup>35</sup>'Nature and the Greeks' and 'Science and Humanism', Schrödinger, 1996

A noter que la séparation entre sujet et objet peut se retrouver dans les étymologies des mots conscience "cum scientia" et science "scientia". Le mot science, ou "scientia", ne comprend pas le "cum" qui semble faire référence au sujet.

La science permet de propager une curiosité rigoureuse et fascinante à propos de la connaissance et de la réalité. Elle semble néanmoins bien faire partie des actes permis par la conscience. Il est en effet difficile d'imaginer que cette activité puisse émaner spontanément et naturellement d'un système non-conscient comme un robot ou une IA. Si la conscience est bien un moyen nécessaire et permanent à l'activité des sciences, il peut sembler étrange que les sciences aient révélé tant de connaissances dans divers domaines, parfois même sur des objets lointains comme en astrophysique, et si peu, semble-t-il, portant sur la conscience elle-même. Cet apparent déséquilibre est peut être dû aux besoins et défis sociétaux auxquels la science répond. La conscience faisant partie de la réalité intellectuelle et sensible, elle fait partie des objets de la science. Elle constitue un mystère relativement épais grâce auquel le scientifique semble paradoxalement découvrir des connaissances enveloppées parfois d'un sentiment de certitude. L'on pourrait se demander d'où viendrait une certitude lorsqu'un des moyens par lesquels elle est obtenue, la conscience, semble encore échapper à la compréhension.

## XIII. CONCLUSION

Dans cet essai, il a d'abord été supposé que la conscience, le "moi" et le libre arbitre existent bien. Cela a permis de rendre compte du conflit entre ces concepts et le déterminisme qui semble contredire leur existence. Il a donc été proposé que le déterminisme, malgré son rôle important en science, pourrait ne pas suffire pour comprendre ce qu'est la conscience. Une façon de contourner ce problème est d'affiner les frontières de la définition de la conscience avec les connaissances actuelles. C'est à dire d'affiner la compréhension de la conscience en définissant ce qu'elle n'est pas en sa proximité. Pour ce faire il a été proposé d'exclure de la définition de la conscience tout phénomène qui semble déterministe. L'exemple de la mémoire a été choisi puisqu'elle fait partie des fonctions cognitives qui semblent avoir un lien avec la conscience. La mémoire telle qu'expliquée par le biais de modèles déterministes en neuroscience a été proposée comme indépendante de la conscience.

En conclusion, ce raisonnement se termine sur la question de savoir ce que pourrait être la conscience sans la mémoire selon les neurosciences ainsi que sur une remise en question de certaines idées associées à la conscience comme le libre arbitre et le "moi".

## XIV. DISCUSSION

Cet essai aboutit à des incertitudes et des questions. Le concept de la conscience étudié au travers des neurosciences, de la science et de la philosophie soulève plusieurs difficultés, notamment l'incertitude de sa définition et les mystères singuliers et encore incompris qu'il abrite.

## A. IA

En conséquence, il est imprudent de clamer que l'IA pourrait faire l'expérience de la conscience ou que la communauté de recherche en IA s'en approcherait grâce à l'obtention seule de performances toujours plus accrues, selon les métriques usuelles en IA, sur des tâches cognitives aussi compliquées soient-elles. D'autres manières de développer l'IA, émancipées du seul objectif de performance, pourraient peut-être la mener vers l'expérience consciente à condition de disposer d'une définition implémentable et déterminée de celle-ci. Or nous avons vu dans cet essai que les différentes définitions de la conscience ne la décrivent pas de manière déterministe et donc programmable. En effet, l'essor actuel de l'IA est notamment permis grâce à la bonne définition et l'implémentation déterminée des performances recherchées.

Il est toutefois important de noter que l'IA peut produire des illusions puissantes, voire générer des addictions et ainsi avoir des effets néfastes sur la santé humaine. Il est possible de constater ces phénomènes au travers des réseaux sociaux ou d'applications comme celle visant à faire "revivre" un défunt grâce à une IA qui a appris son langage personnel<sup>36</sup>.

Dans cet essai, il a été plusieurs fois dit qu'un robot ou une IA était seulement de la matière organisée. Cela reste supposé car il pourrait aussi être défendu que les expériences de la conscience ou du "moi" puissent être exclusivement les conséquences des lois de la physique.

## B. Art et science

Il semble que la conscience se manifeste entre-autres au travers d'activités complexes et singulières. L'art serait l'une d'entre elles. La poésie, la peinture, la musique pourraient être des moyens d'exprimer une réalité, ou une vérité, sans l'entremise de la compréhension rationnelle<sup>37</sup>. Cela pourrait expliquer l'important impact de l'art sur l'humain car il permettrait de lui révéler une réalité qui peut être reconnue en "soi". La seule compréhension rationnelle peut être de ce point de vue perçue comme incomplète pour exprimer la réalité, ou la vérité. Chez Platon encore, la beauté n'est pas un concept esthétique mais peut être comprise comme la manifestation de la vérité dans l'esprit humain, ou dans son âme<sup>38</sup>. Il est intéressant de constater que dans cette perspective les mathématiques ou la physique par exemple peuvent être comprises comme des arts, révélant une part de vérité intérieure ou extérieure à l'esprit humain, sans permettre de vraiment comprendre d'où elle vient ni ce qu'elle veut dire.

A son tour l'IA peut produire des poèmes ou des peintures esthétiques avec un certain degré de perfection, exploitant l'apprentissage d'association complexes entre des expressions humaines depuis une base de données très

volumineuse. Suivant ce qui a été dit précédemment, et malgré une performance manifeste, il serait risqué de prétendre qu'une telle IA produise de l'art. En effet, il s'agit ici de distinguer quelle est l'entité qui crée l'art de celle qui reproduit une certaine compréhension des actes et des productions qui en découlent.

Par exemple, une IA peut apprendre à reproduire certains styles de peinture et même en associer plusieurs ou transformer une photo dans ces styles. Si l'ordre et la logique algorithmique lui sont donnés elle peut aussi à partir de ces styles générer d'autres styles. Ces actes sont finalement des reproductions car ils dépendent entièrement de l'humain qui a implémenté dans une IA une logique permettant à celle-ci d'apprendre des relations complexes dans une base de données illustrant des oeuvres d'arts. Dans les deux cas, la logique d'apprentissage, ou plus largement l'IA elle-même, ainsi que les oeuvres d'art qui constituent cette base de données sont les produits de l'humain. Par conséquent les actes d'une IA, c'est à dire d'un système à priori constitué exclusivement de matière organisée, qu'ils soient associés à de l'art ou à n'importe quelle autre tâche, ne sont pas indépendants de l'humain et en sont la production.

Lorsqu'un peintre peint un tableau, il convient de dire que le tableau est la production du peintre et non du pinceau qui en est un des moyens. Lorsqu'un photographe prend une photo, il convient de dire que cette photo est la production du photographe et non de l'appareil photo qui en est un des moyens. Quand un pilote de formule 1 gagne une course, il convient de dire que le pilote a gagné la course et non son véhicule qui fait pourtant partie des moyens de cette victoire. Peindre, faire de la photographie ou des courses automobiles ne sont pas les produits des produits de l'humain comme l'IA, ils sont la création de l'humain. L'on pourrait défendre que la différence avec une IA est que celle-ci, une fois implémentée et son apprentissage fini, fonctionne indépendamment de l'humain. Cette IA a toutefois besoin de l'humain lors de sa phase de fonctionnement, dite d'inférence, puisqu'elle a besoin d'énergie, de maintenance informatique etc.

D'autre part, et par exemple, le télescope spatial James Webb a été créé par l'humain et a été envoyée dans l'espace pour notamment en prendre des clichés. Une fois le télescope lancé, l'on pourrait dire qu'il parcourt l'univers et en prend des clichés de façon relativement indépendante de l'humain bien qu'il suive son programme, tout comme l'IA. Il pourrait alors être dit que le télescope spatial James Webb a pris des clichés de l'univers. Mais il semble plus précis de dire que l'humain a pris des clichés de l'univers par le moyen du télescope spatial James Webb.

Mis à part la complexité et la difficulté d'expliquer les inférences d'une IA, il n'y a pas de différence fondamentale entre l'algorithme qui contrôle cette sonde et un algorithme d'IA. Les actions d'une IA n'en restent pas moins un produit déterminé de la pensée humaine, bien qu'utilisant souvent la génération de nombres pseudo-aléatoires. Alors l'IA pourrait dans certains cas être un moyen grâce auquel l'humain pourrait produire de l'art et non une entité qui produirait de

<sup>36</sup>Reviving The Dead With AI: Is It Really Worth It?, <https://www.forbes.com/sites/federicoguerrini/2023/04/15/reviving-the-dead-with-ai-is-it-really-worth-it/>

<sup>37</sup>Ion, Platon oeuvres complètes tome I, traduction par Léon Robin, Editions Gallimard, bibliothèque de la pléiade, 1950

<sup>38</sup>République, Platon oeuvres complètes tome I, traduction par Léon Robin, Editions Gallimard, bibliothèque de la pléiade, 1950

l'art par elle-même.

De plus la logique algorithmique qui constitue l'IA est, dans l'exemple de la peinture, un modèle de ce que l'humain a pu comprendre de l'acte artistique de peindre, incluant la façon de s'inspirer d'œuvres existantes par l'apprentissage. D'une part il a été argumenté plus haut que les modèles issus du raisonnement humain ne sont pas universels et expriment de façon incomplète l'objet qu'ils visent malgré leurs qualités. Par conséquent cette IA représenterait une idée incomplète de l'acte de produire de l'art et en serait une approximation. D'autre part, il a aussi été dit plus haut que l'art est un acte en partie irraisonné et que cette part qui échappe à la raison est une condition de son existence. Autrement dit si l'art était déterminé et entièrement modélisable, au travers d'une IA par exemple, peut-être ne serait-il plus l'art. Enfin, l'art a été décrit comme un moyen autre que celui de la raison d'accéder à la réalité ou à une vérité intérieure ou extérieure à l'humain. C'est cette relation bijective entre l'œuvre d'art et l'humain qui constitue en partie l'art et qui pourrait expliquer l'impact important qu'il a sur l'humain. L'œuvre d'art lorsqu'elle est perçue par un humain invoque des vérités qu'il reconnaît en lui sans qu'il les comprenne. Cet acte de reconnaissance est d'ordre personnel et passe nécessairement par le "moi" et la conscience. L'on pourrait se demander ce que serait la nature de la relation entre une œuvre d'art et un robot, ou une IA.

De nouveau, nous pouvons constater un des bénéfices d'une production humaine, l'IA, qui par le reflet de l'humain qu'elle projette permet une remise en question de la nature humaine et notamment des définitions de concepts complexes comme l'art ou la connaissance par exemple. Cela permet de générer le doute qui est moteur quant à l'abandon de certitudes pouvant bloquer le progrès des sciences.

Puisqu'une compréhension plus fine et large des concepts utilisés en IA peut émerger des sciences humaines ou des arts, des considérations philosophiques pourraient participer au progrès de l'élaboration d'IAs et de la communication des travaux de la communauté de l'IA. Cette approche plus large en science est défendue par le chercheur Edgar Morin au travers de la "pensée complexe"<sup>39</sup>. Finalement, la rigueur impose de questionner toute certitude et cette aptitude peut se nourrir d'une curiosité scientifique accrue.

A travers cet exercice de réflexion et de vulgarisation, j'espère avoir offert au lecteur un moment de curiosité intellectuelle et de joie. Pour tenter de le clôturer harmonieusement permettez-moi de citer ci-après ces propos de Nietzsche.

"La magnifique et sauvage déraison de la poésie vous réfute, sectateurs de l'utile. C'est justement la volonté de se délivrer de l'utile qui élève l'homme au-dessus de lui-même." Le Gai Savoir, Nietzsche.

## XV. REMERCIEMENTS

<sup>39</sup>Introduction à la pensée complexe, Morin, 2015