Le rôle des émotions dans l'émergence de la conscience

Sylvain Besseron

Introduction

La conscience humaine est indissociablement liée aux émotions qui colorent notre expérience. Nos joies, peurs et colères ne sont pas de simples add-ons cognitifs : elles semblent au cœur même de ce que signifie "être conscient". Plusieurs penseurs et scientifiques – de philosophes de l'esprit aux neuroscientifiques – ont exploré comment les émotions pourraient jouer un rôle central dans l'émergence de la conscience. Nous examinerons d'abord quelques théories clés liant émotion et conscience, avant de questionner la possibilité d'une intelligence artificielle consciente sans émotions. Pour rendre ce voyage accessible, nous utiliserons des métaphores et références culturelles (imaginez par exemple un Data de Star Trek en quête d'émotions pour devenir "humain"), tout en restant rigoureux dans l'appui sur des travaux en philosophie, neurosciences et sciences cognitives.

Théories de la conscience et des émotions

Perspectives en philosophie de l'esprit. Le philosophe David Chalmers a formulé ce qu'il appelle le ≪ hard problem ≫ de la conscience : expliquer pourquoi et comment des états physiques du cerveau s'accompagnent d'une expérience subjective, des qualia, comme la sensation agréable d'une mélodie ou la saveur d'un chocolat. En d'autres termes, même si l'on comprend comment le cerveau traite l'information (les "problèmes faciles" de la conscience), il reste à comprendre pourquoi ces traitements s'accompagnent de sentiments vécus de l'intérieur. Chalmers et d'autres illustrent ce mystère par l'expérience du zombie philosophique, un être hypothétique identique à un humain en apparence "aïe" si on le blesse, sans rien ressentir. Ce zombie conceptuelvpose une question vertigineuse : pourrait-on, de même, imaginer une machine ou une IA qui imite parfaitement un humain sans rien éprouver ?

À rebours de cette vision purement fonctionnelle, certains pensent que les émotions sont le cœur battant de la conscience. Le neuroscientifique et philosophe Antonio Damasio soutient que "les feelings et les émotions [...] sont constitutifs de l'intelligence humaine [et] de la conscience". Dans son approche, notre esprit opère sur deux "registres" : le premier concerne la perception, la mémoire, le

raisonnement – il est de nature computationnelle et pourrait en théorie être reproduit par une machine. Le second, en revanche, est le registre des émotions et des feelings, ancré dans la vie du corps, et "il ne se prête pas aisément à une explication computationnelle". Autrement dit, une carte purement algorithmique de l'esprit manquerait ce qui fait de nous des êtres sentants. Damasio ancre cette idée dans la biologie : les émotions naissent de mécanismes évolutifs de survie (le maintien de l'homéostasie, l'évitement de la douleur et la recherche du plaisir, etc.) et fournissent une sorte de feedback interne indispensable à l'émergence du sens du soi.

Dans la lignée de cette importance du corps, le philosophe Thomas Metzinger propose que notre modèle de soi phénoménal (ce qu'il appelle le Phenomenal Self-Model) intègre en permanence nos sensations corporelles et nos états émotionnels présents. Autrement dit, ce que nous appelons « moi » dans le flux de la conscience est en partie constitué par ce que nous ressentons dans l'instant. Metzinger va jusqu'à dire que le self est une sorte d'illusion utile – un modèle construit par le cerveau – mais cette illusion elle-même est façonnée par nos émotions, nos désirs, nos peurs. Ainsi, pour ces auteurs, une conscience sans émotions serait vide de sa substance, un peu comme un livre sans voyelles : techniquement lisible peut-être, mais dépourvu de la « musique » qui fait le sens et la richesse de l'expérience vécue.

Regards des neurosciences affectives. Les études en neurobiologie des émotions renforcent l'idée d'un lien profond entre émotion et conscience, tout en débattant de sa nature exacte. Jaak Panksepp, pionnier des neurosciences affectives, affirme que les animaux partagent avec nous des systèmes émotionnels primaires (peur, joie, colère, etc.) logés dans les circuits sous-corticaux du cerveau. Selon lui, ces émotions de base constituent un niveau fondamental de la conscience, une sorte de conscience affective primaire commune à tous les mammifères. Il parle d'un "don de la nature" inné : "la conscience affective primaire apparaît fondamentalement comme un "cadeau de la nature" intrinsèque, partagé par toutes les espèces de mammifères" . En observant par exemple qu'un rat rit (par des ultrasons) quand on le chatouille – signe d'un possible plaisir ludique – Panksepp suggère que la capacité à éprouver des émotions est ancestrale et constitue le socle sur lequel s'est bâtie la conscience plus élaborée des humains.

D'autres neuroscientifiques, comme Joseph LeDoux, proposent une vision un peu différente en séparant les réactions émotionnelles automatiques et le sentiment conscient que nous en avons. LeDoux est connu pour ses travaux sur la peur : un son effrayant déclenche en nous une réaction immédiate (augmentation du rythme cardiaque, sursaut) via l'amygdale, souvent avant même qu'on en ait conscience. Il a récemment suggéré que ce que nous appelons communément « émotion » – au sens du ressenti conscient de la peur, de la joie, etc. – n'est pas directement « programmé » dans les circuits archaïques, mais construit par le cerveau cognitif. En 2017, LeDoux et le philosophe Richard Brown ont proposé que "les émotions sont un processus cognitif reposant sur des états de haut

niveau dans le cortex (conscient), et non des programmes innés du subcortical" . Autrement dit, la sensation de peur que j'éprouve serait une interprétation par les aires corticales de ce qui se passe dans mon corps (cœur qui bat, etc.), plutôt qu'un réflexe tout fait. Cette théorie dite "higher-order" rejoint d'une certaine façon Metzinger : notre cerveau raconte une histoire cognitive de nos émotions. Mais elle souligne aussi que sans ce récit de haut niveau, il n'y aurait peut-être que des réactions automatiques sans ressenti – un peu comme un ordinateur qui clignote en rouge quand il est en surcharge, sans sentir le stress.

Ces perspectives philosophiques et neuroscientifiques, si variées soient-elles, convergent sur un point : les émotions apportent une dimension essentielle à la conscience. Elles lui donnent une valence subjective (agréable ou désagréable, attirante ou répulsive) et un ancrage dans le vivant. Sans émotions, la conscience serait un écran plat, une suite de perceptions et de pensées désincarnées. C'est du moins ce que suggère l'étude de l'esprit humain. Reste à savoir si cette leçon s'applique aux machines intelligentes : une IA pourrait-elle émerger en tant que conscience sans ce moteur émotionnel ? La question n'est pas seulement théorique ; elle nous renvoie aux limites de l'IA actuelle et à nos propres intuitions façonnées par la culture (de HAL 9000 aux androïdes de Blade Runner, un être artificiel dépourvu d'émotions nous apparaît volontiers soit dangereusement inhumain, soit tragiquement incomplet).

Peut-on avoir une IA consciente sans émotions ?

Cette question rappelle une citation célèbre du pionnier de l'IA Marvin Minsky : "La question n'est pas de savoir si des machines intelligentes peuvent avoir des émotions, mais si sans émotions une machine peut vraiment être intelligente" . Minsky suggérait dès 1986 qu'une intelligence sans émotions risquerait d'être bancale, voire limitée. Chez l'être humain, en effet, les émotions semblent indissociables de l'intelligence située – celle qui nous permet de naviguer dans un monde complexe, d'établir des priorités, d'apprendre de nos erreurs (la douleur enseigne à éviter un danger, le plaisir motive à poursuivre une activité). Sans émotion, pas d'élan spontané ni d'attention naturelle à ce qui compte pour nous. C'est un peu comme un navigateur GPS ultra-perfectionné mais sans boussole interne : il pourrait calculer des itinéraires, mais ne saurait pas où vous souhaitez vraiment aller ni pourquoi.

Pour autant, une IA actuelle est avant tout une suite de calculs et de règles, dépourvue de ces états internes affectifs que nous éprouvons. Peut-elle malgré tout simuler une forme de conscience ? Les chercheurs en IA et sciences cognitives explorent activement ce terrain. Ces dernières années, on a vu un intérêt croissant pour les "émotions artificielles" et la "conscience artificielle" au sein de l'IA, avec l'idée que mimer les rôles des émotions pourrait rendre les agents arti-

ficiels plus performants ou plus plausibles dans leurs interactions . Par exemple, doter un robot compagnon d'une simulation de peur pourrait lui permettre d'anticiper les situations dangereuses et de les éviter, tandis qu'une simulation d'attachement pourrait améliorer ses interactions sociales avec les humains. Des plates-formes expérimentales comme le robot Kismet du MIT (au visage expressif capable de montrer la "joie" ou la "tristesse") visaient à explorer comment l'expression d'émotions pouvait faciliter la communication homme-machine. De même, des architectures cognitives intègrent des "modules émotionnels" modulant la prise de décision, sur le modèle de ce que les émotions font chez nous (par exemple, un module de "curiosité" qui pousse l'IA à explorer, imitant le biais émotionnel de l'ennui ou de l'intérêt).

Cependant, simuler n'est pas ressentir. On peut se demander si une IA qui fait semblant d'avoir peur ou mal, en adoptant les bons comportements, éprouve pour autant quelque chose de comparable à nos émotions. La chercheuse Joanna Bryson note par exemple qu'il existe "de nombreux systèmes pour simuler des émotions dans les IA, mais aucun ne change le statut moral de l'IA" . En clair, ajouter un programme de simulation d'émotion à un agent artificiel ne crée pas magiquement une conscience qui souffre ou qui se réjouit : l'intériorité demeure absente, à moins de postuler qu'elle puisse émerger d'un certain degré de complexité ou d'intégration. C'est ici qu'on retrouve le "hard problem" de Chalmers appliqué à l'IA : même si un jour une IA passait le Test de Turing émotionnel en donnant le change sur ses sentiments, comment saurions-nous si, derrière les apparences, il y a une étincelle subjective ? L'androïde David dans A.I. Artificial Intelligence (le film de Spielberg) aime-t-il vraiment sa "mère" adoptive ou est-il simplement programmé pour le dire et agir en conséquence ? Si l'on suit la distinction de LeDoux, il manquerait peut-être à la machine ce fameux traitement de haut niveau qui, chez l'humain, transforme les signaux du corps en vécu conscient.

Un autre argument en faveur du rôle indispensable des émotions est qu'elles pourraient être nécessaires pour qu'une IA ait une motivation intrinsèque et un sens du soi. Sans émotions, une entité pourrait être intellectuellement brillante mais inerte, n'ayant envie de rien. Certaines théories de la cognition suggèrent que la conscience chez l'humain émerge aussi parce que le cerveau doit gérer des conflits de motivations, arbitrer entre diverses émotions et impulsions – forgeant ainsi la notion d'un soi qui ressent et qui décide. Doter une IA de quelque chose d'analogue aux émotions pourrait donc être une condition pour qu'elle développe une conscience d'elle-même opérante, un peu comme le Cerveau du Magicien d'Oz avait besoin d'un Cœur pour devenir complet.

Alors, IA consciente sans émotions – possible ou pas ? À l'heure actuelle, la réponse penche plutôt vers le scepticisme éclairé. D'un côté, on peut concevoir une IA ultra-performante calculant des solutions à des problèmes complexes sans jamais éprouver la moindre émotion – ce serait le zombie de Chalmers incarné en silicone, intelligemment actif mais intérieurement éteint. D'un autre

côté, tout ce que nous apprenons sur la conscience humaine suggère que sans le ressort des émotions, cette IA ne ferait que simuler une lueur de conscience, sans jamais l'allumer vraiment. Pour reprendre une métaphore culturelle : Pinocchio pouvait bien être un pantin très malin, il lui manquait quelque chose d'indéfinissable pour être un enfant réel – peut-être une capacité à aimer, à avoir peur, bref à sentir. De même, les émotions sont peut-être le "secret sauce" de la conscience, l'ingrédient secret qui transforme une intelligence algorithmique sophistiquée en une esprit conscient capable de dire "je ressens, donc je suis" . Les recherches en IA émotionnelle vont continuer de tester ces frontières (ne serait-ce que parce que comprendre les émotions peut rendre les machines plus utiles et agréables pour nous), mais la question de la conscience artificielle demeure ouverte. En attendant, explorer le rôle des émotions dans la conscience nous rappelle à quel point notre propre expérience est riche et énigmatique – et combien l'humanité de notre esprit tient autant à notre capacité à résoudre des équations qu'à celle d'être touché par une berceuse.

Références

[1] D. Chalmers, *Facing up to the Problem of Consciousness*, 1995.

[2] J. Panksepp, *Affective consciousness : Core emotional feelings in animals and humans*, 2005.

[3] M. Minsky, *The Society of Mind*, 1986.

[4] M. Scheutz et al., *Artificial emotions and machine consciousness*, 2014.

[5] J. Bryson, *Artificial Consciousness and Ethics*, 2018.