

Certes, les IA battent les humains à beaucoup de jeux. Mais elles utilisent des stratégies bien particulières, comme le déroulement d'une quantité énorme de calculs pour prévoir plusieurs coups à l'avance. Elles ne se fondent donc pas sur un ensemble de règles et de connaissances présentes de façon plus ou moins consciente dans l'esprit des meilleurs joueurs. Selon Chater, ce revers ne tient pas au fait que l'on n'aurait pas trouvé ces précieux savoirs, qui seraient trop profondément enfouis, mais plus essentiellement au fait qu'on ait cru à tort à leur existence.

Comment les experts parviennent-ils à être si performants, dans ce cas ? La réponse est peut-être à chercher du côté des succès récents des réseaux de neurones artificiels. Dans ces modèles, des cellules virtuelles mimant des neurones sont mises en lien, et s'activent ou se désactivent les unes les autres de manière purement automatique. En faisant varier la force des liens entre les neurones, on finit par créer des réseaux capables de prouesses inédites, comme la reconnaissance et la classification d'images. De même, le cerveau d'un champion d'échecs deviendrait de plus en plus efficace pour mener une partie, sans que cela ne soit formalisable par des règles logiques ; cela se traduirait plutôt par des souvenirs diffus, par des habitudes et des réflexes bien plus performants que la moyenne, et par une capacité extraordinaire à imaginer des enchaînements de coups à partir de ces éléments.

Plus généralement, il n'y a dans les réseaux de neurones artificiels aucun principe universel inscrit au départ, aucune définition formelle des catégories, seulement un mécanisme dépourvu de sens, un réglage *ad hoc* des paramètres et beaucoup d'apprentissage. Le comportement apparemment intelligent qu'ils affichent est une « propriété émergente », certes observable de l'extérieur, mais qui n'est gravée nulle part. En réalité, lorsque le réseau fonctionne bien, on ne sait pas vraiment pourquoi : les paramètres sont juste correctement ajustés. Il n'empêche qu'une fois stabilisé, ce minicerveau artificiel agira de manière similaire d'une fois sur l'autre – de la même manière que nous avons une certaine constance dans nos réactions émotionnelles ou intellectuelles. Peut-on pour autant dire qu'il a des idées, des concepts stables, voire une personnalité ? Probablement pas.

L'INCONSCIENT DES PSYS

Dans la quête des profondeurs de l'âme humaine, la psychanalyse a sans doute été l'erreur la plus têtue. Les pionniers de cette discipline ont imaginé un inconscient particulièrement riche. Loin de celui des cognitivistes, qui désigne simplement l'ensemble des processus qui se déroulent hors du champ de la conscience, l'inconscient des psychanalystes est complexe. Il

s'agit pour ainsi dire d'un autre moi à l'intérieur de moi, avec sa personnalité, sa volonté propre, ses raisons. Un être capable de s'opposer à moi, de décider de me cacher des choses.

Par divers moyens, comme le rêve ou les associations libres, les psychanalystes pensent trouver un accès détourné à cette réalité des gouffres. Ces méthodes n'ont pourtant rien de scientifique, notamment parce que les mêmes observations peuvent être interprétées de façons différentes et souvent contradictoires. Si vous rêvez que vous avez peur d'une araignée, par exemple, tel psychanalyste y verra une angoisse d'ordre sexuel, là où tel autre le prendra comme un signe que vous avez du mal à assumer un rôle

QUE NOUS N'AYONS PAS À SUBIR LA TYRANNIE IMPLACABLE D'UN AUTRE MOI CACHÉ AU FOND DE NOUS EST PLUTÔT RÉJOUISSANT, CAR CELA SIGNIFIE QUE NOUS AVONS LE POUVOIR DE CHANGER

de dirigeant dans votre entreprise (car vous avez peur d'être appelé « à régner »...).

On pourrait croire que l'échec du décryptage de l'inconscient tient au fait qu'il est si rusé qu'il ne nous laisse parvenir que des bribes insuffisantes. Mais il est plus probable que la raison soit plus profonde : si l'on échoue à découvrir les raisons cachées, les valeurs et les désirs de l'inconscient, c'est plus simplement parce qu'ils n'existent pas. Il y a bien sûr des processus inconscients complexes, des analogies automatiques, des mouvements émotionnels incontrôlés, des interprétations machinales – bref un traitement de l'information sophistiqué, mais opaque, comme dans un réseau de neurones artificiels. Pourtant, tout ceci n'est pas le fruit d'une personnalité riche gisant au fond de notre esprit.

Ce constat ne doit pas nous abattre. Que nous n'ayons pas à subir la tyrannie implacable d'un autre moi caché au fond de nous est plutôt réjouissant. Que nos habitudes façonnent plus nos comportements que des valeurs qui nous seraient inaccessibles est rassurant, car cela signifie que nous avons le pouvoir de changer, sans nous lancer dans la spéléologie psychologique : les habitudes peuvent être modifiées de manière très concrète par l'effort et la répétition. La volonté a son mot à dire, bien plus qu'un alter ego capricieux et insaisissable. ■

BIBLIOGRAPHIE

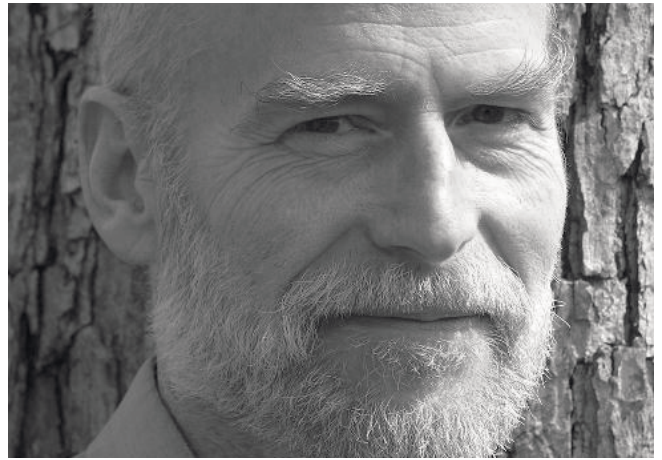
N. CHATER,
*Et si le cerveau
était bête ?*, Plon, 2018.

E. TROUCHE ET AL.,
*The selective laziness of
reasoning*, *Cognitive Science*,
vol. 40, pp. 2122-2136, 2016.

L. HALL ET AL., *How the polls
can be both spot on and dead
wrong : Using choice
blindness to shift
political attitudes and voter
intentions*, *Plosone*, vol. 8(4),
e60554, 2013.

L. ROZENBLIT ET F. KEIL,
*The misunderstood limits of
folk science : An illusion of
explanatory depth*, *Cognitive
Science*, vol. 26, pp. 521-562, 2002.

PETER CARRUTHERS



« Le vrai travail de l'esprit reste inconscient, seul ce qu'il produit apparaît à notre regard intérieur »

Un de vos récents articles s'intitule : « L'illusion de la pensée consciente ». Qu'est-ce que cela signifie ?

Peter Carruthers : Tout simplement, que l'idée d'après laquelle nous pensons consciemment est selon moi une erreur. Cette impression subjective repose sur un phénomène que j'appelle « l'illusion d'immédiateté ». Le point de départ de mes réflexions a été une tentative pour cerner de façon plus précise deux approches de la théorie de la conscience.

D'une part, la théorie de l'espace de travail global (la *Global workspace theory*), défendue par les neuroscientifiques Stanislas Dehaene et Bernard Baars. La thèse centrale est qu'un état mental conscient doit être accessible à d'autres fonctions psychiques comme la mémoire de travail, la prise de décision ou le langage. Les états conscients, selon cette conception, rayonnent de manière globale sur l'ensemble du cerveau.

BIO EXPRESS

16 JUIN 1952
Naissance

1980
Thèse de philosophie au Balliol College d'Oxford, en Grande-Bretagne.

2001
Professeur de philosophie à l'université du Maryland, à College Park, aux États-Unis.

Selon une autre théorie, défendue notamment par le psychologue Michael Graziano, de l'université de Princeton, et le philosophe David Rosenthal, de l'université de New York, les états mentaux conscients sont simplement ceux que nous connaissons, auxquels nous avons accès de façon immédiate, sans avoir à lire nos propres pensées ou à les interpréter. Cette conception est aussi qualifiée de « théorie d'ordre supérieur » (*Higher order theory*). Mon argument est essentiellement le suivant : peu importe que l'on privilégie telle ou telle interprétation, nous restons incapables de saisir consciemment nos pensées, souhaits ou jugements. Ils ne sont ni à notre disposition dans notre mémoire de travail ni directement accessibles à notre conscience.

Les sources de notre pensée seraient alors inconscientes ?

Peter Carruthers : Chaque jour, nous disons naturellement des choses comme