

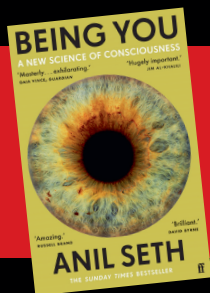
ÊTRE SOI

Une nouvelle science de la conscience



ANIL SETH

Traduit par Michel Le Bellac



Traduction du best-seller
« BEING YOU »

INTRODUCTION UNE INTRODUCTION À TATION

Anil Seth
ÊTRE SOI
UNE NOUVELLE SCIENCE DE LA CONSCIENCE

Qu'est-ce que la « conscience de soi » ? Dans cet ouvrage, Anil Seth, un des scientifiques les plus en vue dans les recherches sur la conscience, se propose de répondre à cette question et à d'autres en s'appuyant sur les données les plus récentes de la biologie, des neurosciences et de l'imagerie cérébrale, tout en intégrant les contributions de philosophes et de psychologues.

Être soi est un voyage à travers les différentes facettes de l'expérience consciente. Nous ne percevons pas un monde extérieur objectif, mais nos cerveaux sont des machines à prédire qui confrontent en permanence leurs prédictions aux données sensorielles externes et internes, un processus que l'on peut interpréter comme une suite d'hallucinations contrôlées.

Être soi nous projette dans un monde d'idées nouvelles sur le fonctionnement du cerveau et sur le mystère de la conscience humaine, idées qui suggèrent par exemple de nouvelles façons d'évaluer le niveau de conscience de patients dans le coma. Finalement l'auteur s'interroge sur les différences entre conscience et intelligence et sur la possibilité d'une conscience chez les animaux non-humains et éventuellement dans les machines « intelligentes » comme les robots conversationnels.

« La réalité est réelle, mais la façon dont nos cerveaux construisent une image de la réalité — percevoir, intégrer, prédire — est loin d'être directe. C'est un fouillis terriblement compliqué et fascinant que les neuroscientifiques commencent seulement à décoder. Le livre d'Anil Seth, *Être soi*, est un compte rendu merveilleusement accessible et compréhensible de la manière dont nos esprits captent le monde, et de comment cela fait de nous ce que nous sommes. »

Sean CARROLL, auteur de *Le Grand Tout*.

Anil Seth est professeur de neurosciences cognitives et computationnelles à l'Université du Sussex où il dirige le Centre du Sussex pour la science de la conscience.

Michel Le Bellac est professeur émérite de physique à l'Université Côte d'Azur.

ISBN : 978-2-7598-3235-4



9 782759 832354

edp sciences
www.edpsciences.org

La collection « UNE INTRODUCTION À... » se propose de faire connaître à un large public les avancées les plus récentes de la science. Les ouvrages sont rédigés sous une forme simple et pédagogique par les meilleurs experts français.

**Collection *Une introduction à* dirigée par
Michèle Leduc et Michel Le Bellac**

Être soi

Une nouvelle science de la conscience

ANIL SETH

Traduit par Michel Le Bellac

“Being You – A New Science of Consciousness” was originally published in English in 2021. This translation is published by arrangement with Janklow & Nesbit (UK) Ltd. Copyright © 2021, Anil Seth. All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from the Publishers.

Imprimé en France
ISBN (papier) : 978-2-7598-3235-4
ISBN (ebook) : 978-2-7598-3239-2

Tous droits de traduction, d'adaptation et de reproduction par tous procédés, réservés pour tous pays. La loi du 11 mars 1957 n'autorisant, aux termes des alinéas 2 et 3 de l'article 41, d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective », et d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (alinéa 1^{er} de l'article 40). Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles 425 et suivants du code pénal.

© EDP Sciences, 2023

Pour ma mère, Ann Seth
et à la mémoire de mon père Bhola Nath Seth

*Le Cerveau est plus grand que le Ciel
Mettez-les côte à côte
L'un contient l'autre
Sans problème et Vous en plus*

EMILY DICKINSON, traduction PATRICK REMAUX

SOMMAIRE

Partie I Niveau	7
<i>Prologue</i>	9
1 Le problème réel	19
2 Mesurer la conscience	43
3 Phi	73
Partie II Contenu	89
4 Percevoir le monde de l'intérieur	91
5 Le sorcier du hasard	115
6 La part du spectateur	139
Partie III Soi	165
7 Délire	167
8 Attendez-vous à vous-même	173
9 Être un animal-machine	199
10 Comme un poisson dans l'eau	225
11 Degrés de liberté	239
Partie IV Autres	257
12 Au-delà de l'humain	259
13 Machines pensantes	281
Épilogue	305
Remerciements	315
Notes	319
Références	375
Index	409

Partie I

NIVEAU

Prologue

Il y a cinq ans, pour la troisième fois de ma vie, j'ai cessé d'exister. Je subissais une opération chirurgicale mineure et mon cerveau se remplissait de produits anesthésiants. Je me souviens de sensations d'obscurité, de détachement et de désintégration. . .

Une anesthésie générale est très différente du sommeil. Heureusement que c'est le cas : si vous étiez simplement endormi, le bistouri du chirurgien vous réveillerait en un clin d'œil. Les états d'anesthésie profonde ont davantage en commun avec des états catastrophiques tels que le coma et l'état végétatif, où la conscience est totalement effacée. Dans une anesthésie profonde, l'activité électrique du cerveau est presque entièrement au repos – quelque chose qui ne se produit jamais dans la vie courante, que l'on soit éveillé ou endormi. C'est un des miracles de la médecine moderne que les anesthésistes peuvent de façon routinière altérer les cerveaux humains de sorte que ceux-ci sont capables d'entrer dans de tels états inconscients. . . et d'en sortir. C'est un acte de transformation, une sorte d'acte de magie : l'anesthésie est l'art de transformer des personnes en objets.

Bien sûr les objets redeviennent des personnes. Ainsi je refis surface, ensommeillé et désorienté, mais bel et bien *présent*. Il me semblait que le temps ne s'était pas écoulé. Au contraire, lorsque je me réveille

d'un sommeil profond, j'ai parfois du mal avec l'heure mais j'ai toujours l'impression qu'au moins un certain temps s'est écoulé, qu'il y a une continuité entre ma conscience d'avant l'endormissement et ma conscience au réveil. Sous anesthésie générale, les choses sont différentes. Je pourrais avoir été absent cinq minutes, cinq heures, cinq ans ou même cinquante. Et « absent » n'est pas exactement le bon terme. Je n'étais simplement pas là, un avant-goût de l'oubli total de la mort doublé d'un sentiment étrangement réconfortant procuré par cette absence de quoi que ce soit.

L'anesthésie générale n'agit pas seulement sur votre cerveau, sur votre esprit. Elle agit sur votre conscience. En altérant le délicat équilibre électrochimique entre les circuits neuronaux à l'intérieur de votre crâne, l'état fondamental de ce que veut dire « être » est – temporairement – aboli. C'est dans ce processus que réside un des grands mystères que la science doit encore élucider, et également un des grands mystères de la philosophie.

D'une certaine façon, à l'intérieur de chacun de nos cerveaux, l'activité combinée de milliards de neurones, où chacun de ces neurones est une minuscule machine biologique, donne naissance à une expérience consciente. Et ce n'est pas n'importe quelle expérience consciente, c'est *votre* expérience consciente, juste maintenant et juste ici même. *Comment cela peut-il se produire ? Et pourquoi notre expérience consciente, notre expérience de la vie, est-elle à la première personne ?*

J'ai un souvenir d'enfance où je me regarde dans un miroir de salle de bains, et je me rends compte pour la première fois que ce que je ressens à ce moment précis – l'expérience d'être moi – allait se terminer un jour et que « Je » mourrais. Je devais avoir huit ou neuf ans, et comme tout souvenir d'enfance, ce souvenir n'est pas vraiment fiable. Mais c'est peut-être également à cet instant que je me suis aussi rendu compte que, si ma conscience pouvait avoir une fin, alors cette conscience devait d'une certaine manière dépendre du matériau dont j'étais fait – de la substance matérielle de mon corps et

de mon cerveau. Il me semble que je me suis depuis lors confronté en permanence à ce mystère, d'une façon ou d'une autre.

Lorsque j'étais étudiant de licence à l'Université de Cambridge dans les années 1990, une romance d'adolescent avec la physique et la philosophie s'est développée en une fascination pour la psychologie et les neurosciences, bien que ces domaines, à cette époque, aient évité et même proscrit toute mention de la notion de conscience. Mon travail de thèse m'a fait faire un détour assez long et productif dans les domaines de l'intelligence artificielle et de la robotique, avant qu'un passage de six années à l'Institut des Neurosciences de San Diego (Californie), sur les rives du Pacifique, ne me donne finalement la chance de me plonger directement dans les mécanismes cérébraux à la base de la conscience. Dans cet institut, j'ai pu travailler directement avec Gerald Edelman, lauréat Nobel – un des scientifiques les plus en vue parmi ceux qui ont remis sur les rails la conscience comme objet de recherche scientifique légitime.

Aujourd'hui, depuis plus de dix ans, je suis le co-directeur d'un centre de recherches, l'Institut pour la science de la conscience de l'Université du Sussex, niché au milieu des collines en pente douce et verdoyantes des South Downs près de la cité balnéaire de Brighton. Notre centre rassemble des spécialistes des neurosciences, des psychologues, des psychiatres, des experts de l'imagerie cérébrale, des sorciers de la réalité virtuelle, des mathématiciens et des philosophes, avec l'objectif commun d'ouvrir de nouvelles perspectives sur les mécanismes cérébraux à la base de la conscience.

Que vous soyez ou non un scientifique, la conscience est un mystère qui doit vous interpeller. Pour chacun d'entre nous, l'expérience consciente est *la totalité de ce qui existe*. Sans cette conscience, rien n'existe : ni le monde, ni le soi, intérieur ou extérieur.

Imaginez qu'une version future de moi-même, peut-être même dans un futur pas si lointain, vous offre l'affaire de votre vie. Je peux

remplacer votre cerveau par une machine qui lui soit équivalente en tout point de sorte que, de l'extérieur, personne ne peut voir de différence. Cette nouvelle machine possède plusieurs avantages, en particulier elle est immunisée contre toute détérioration et peut-être va-t-elle vous assurer une vie éternelle.

Mais il y a un piège. Puisque même le moi-futur n'a aucune certitude sur la manière dont un cerveau réel donne naissance à la conscience, je ne peux pas garantir que vous aurez une quelconque expérience consciente si vous acceptez mon offre. Il est possible que vous ayez cette expérience, s'il s'avère que la conscience dépend uniquement des capacités fonctionnelles du cerveau, de la puissance et de la complexité de ses circuits, mais peut-être ce ne sera pas le cas, si la conscience dépend en fait d'un matériau biologique spécifique, des neurones par exemple. Bien sûr, comme votre cerveau-machine conduit à un comportement en tout point identique, si je vous demande « êtes-vous conscient ? », le nouveau-vous va répondre « oui ». Mais qu'en sera-t-il en dépit de cette réponse, si la vie – pour vous – n'est plus ressentie à la première personne ?

Je soupçonne que vous n'allez pas conclure le marché. En l'absence de conscience, peu importe que vous viviez cinq ans de plus ou même cinq siècles. Pendant toutes ces années, *il n'y aurait rien qui ressemble à être vous*.

En dehors de tout jeu philosophique, il est facile d'apprécier l'importance pratique de comprendre la façon dont la conscience est engendrée par des mécanismes internes du cerveau, en d'autres termes de comprendre quelle est la base neuronale de la conscience. L'anesthésie générale est à ranger parmi les plus grandes inventions de tous les temps. Si nous passons à un côté plus sombre, des dysfonctionnements douloureux de la conscience peuvent accompagner des lésions du cerveau et des maladies mentales pour un nombre croissant d'entre nous, et j'en fais partie, qui font face à ces problèmes. Et pour chacun d'entre nous les expériences conscientes se modifient tout au long de la vie, depuis la confusion bourgeonnante et bourdonnante

des débuts de la vie, en passant par la clarté de l'âge adulte, probablement illusoire et certainement non universelle et en finissant par notre dérive finale dans une dissolution graduelle – mais pour certains d'une rapidité confondante – du soi lorsque s'installe la désintégration neurodégénérative. Vous existez indubitablement à chaque étape de ce processus, mais la notion qu'il existe un seul moi conscient (une âme ?) qui persiste au cours du temps est probablement grossièrement erronée. De fait, un des aspects les plus incontournables du mystère de la conscience est la nature du *moi*. Est-ce que la conscience peut exister sans la conscience de soi et, si c'est le cas, est-ce que cela aurait vraiment une importance ?

Les réponses aux questions difficiles que nous venons d'évoquer soulèvent plusieurs problèmes dans la façon dont nous pensons le monde et la vie qu'il sous-tend. Quand la conscience apparaît-elle dans le développement ? Est-ce qu'elle émerge à la naissance, ou bien est-elle déjà présente dans l'utérus ? Que dire de la conscience des animaux non humains – et pas uniquement des primates et autres mammifères, mais même de celle de créatures venues d'un autre monde comme les pieuvres, voire d'organismes aussi simples que les nématodes ou les bactéries ? Est-ce que cela ressemble à quelque chose d'être une *Escherichia coli* ou une daurade ? Et que dire des machines futures ? Dans ce cas, nous devrions nous poser la question non seulement de la puissance des nouvelles formes de l'intelligence artificielle qui sont en train de nous dépasser, mais aussi la question de quand et comment il *nous* sera nécessaire d'adopter une position éthique à leur égard. En ce qui me concerne, ces questions évoquent la sympathie troublante que j'éprouve pour Dan Bowman lorsque je le vois détruire la personnalité de l'ordinateur HAL dans le film *2001 : Odyssée de l'espace*, par le simple fait de le dépouiller de ses banques de données, une par une. Le fait qu'une plus grande empathie soit suscitée par les épreuves endurées par les répliquants de Ridley Scott dans *Blade Runner*

donne une clé sur l'importance de notre nature en tant que *machines vivantes* en ce qui concerne notre expérience d'être un soi conscient.

Ce livre traite des neurosciences de la conscience : la tentative de comprendre comment notre univers intérieur de l'expérience subjective est relié à, et peut être expliqué en termes de processus physiques et biologiques qui se déploient dans les cerveaux et les corps. C'est le projet qui m'a captivé tout au long de ma carrière et j'ai la conviction que ce projet a atteint une maturité telle que des bribes de réponses commencent à émerger.

Ces bribes de réponses modifient déjà, et même de façon dramatique, la manière dont nous pensons l'expérience consciente qui nous entoure, et dont nous-mêmes faisons partie. La façon dont nous envisageons la conscience a un impact sur chaque aspect de notre vie. La science de la conscience n'est pas autre chose qu'un compte rendu de ce que nous sommes, de ce que c'est d'être moi, ou vous, et pourquoi il y a quelque chose qui ressemble à simplement « être ».

Dans l'histoire que je vais raconter, j'adopte un point de vue personnel, qui a pris forme au cours de nombreuses années de recherches, de contemplation et de conversations. Autant que je le puisse le comprendre, le problème de la conscience ne sera pas « résolu » de la façon dont a été décodé le génome humain ou de la façon dont a été établie la réalité du changement climatique. Je ne pense pas non plus que le mystère sera résolu par un coup de baguette magique, une seule intuition lumineuse de type « eureka » – un mythe confortable mais en général erroné sur la manière dont la science progresse.

Selon moi, une science de la conscience devrait pouvoir expliquer comment les différentes propriétés de la conscience dépendent de, et sont reliées à, la façon dont opère le matériau neuronal humide ¹

1. N.d.T. L'expression utilisée par l'auteur est *neuronal wetware*, qui fait manifestement référence à *software*, le logiciel et *hardware*, le matériel. Mais ce jeu de mots est intraduisible en français.

(wetware) à l'intérieur de nos crânes. L'objectif d'une science de la conscience ne devrait pas être – du moins principalement – d'expliquer pourquoi la conscience se trouve être en premier lieu une partie de l'Univers. Il ne devrait pas non plus consister à comprendre le fonctionnement du cerveau dans toute sa complexité, tout en balayant sous le tapis le mystère de la conscience. Ce que j'espère vous montrer est qu'en rendant compte des propriétés de la conscience en termes de mécanismes qui se déroulent dans le cerveau et dans le corps, le mystère métaphysique profond du pourquoi et du comment de la conscience devient peu à peu moins mystérieux.

J'ai utilisé l'expression « matériau humide » pour souligner le fait que les cerveaux ne sont pas des ordinateurs que l'on aurait fabriqués avec des tissus biologiques. Ce sont à la fois des machines chimiques et des réseaux électriques. Tout cerveau qui a existé un jour a fait partie d'un corps vivant, incorporé dans et interagissant avec son environnement – un environnement qui inclut souvent plusieurs autres cerveaux incarnés. Expliquer les propriétés de la conscience en termes de mécanismes biophysiques exige que l'on comprenne les cerveaux – et les esprits conscients – comme des systèmes *immergés* et *incarnés*.

Finalement, je voudrais vous suggérer une nouvelle conception du soi – cet aspect de la conscience qui pour chacun de nous a probablement le plus de signification. Une tradition influente, remontant au moins au dix-septième siècle et à Descartes, a soutenu que les animaux non humains ne possédaient pas une personnalité consciente parce qu'ils ne disposaient pas d'un esprit rationnel pour guider leur comportement. C'étaient des « animaux-machines » : des automates faits de chair et d'os incapables de réfléchir à leur propre existence.

Je ne partage pas ce point de vue. Selon moi, la conscience est plus liée à la vie qu'à l'intelligence. Nous sommes des moi-conscients précisément *parce que* nous sommes des animaux-machines. Je vais soutenir que les expériences *d'être soi*, ou *d'être moi*, émergent de la façon dont le cerveau prédit et contrôle l'état interne du corps. L'essence du soi, de son identité, n'est ni un esprit rationnel, ni une

âme immatérielle. C'est l'incarnation ultime d'un processus biologique, un processus qui sous-tend le simple sentiment d'être en vie et qui est la base exclusive de toutes nos expériences du soi. Être soi est littéralement une propriété de votre corps.

Ce livre se divise en quatre parties. Dans la première partie, je développe mon approche personnelle à l'étude scientifique de la conscience. Cette partie traite également de la question du « niveau » de conscience – quel est le degré de conscience de quelqu'un ou de quelque chose – et des progrès dans les tentatives de « mesurer » ce niveau de conscience. La seconde partie est consacrée au « contenu » conscient – c'est-à-dire de ce dont vous êtes conscient, lorsque vous êtes. . . conscient. La troisième partie inverse la focalisation en s'orientant vers le soi et toutes les expériences conscientes que l'identité implique. La quatrième et dernière partie explore ce qu'implique cette nouvelle approche dans la compréhension de la conscience pour les animaux non humains et la possibilité de machines sensibles. En refermant ce livre, vous aurez compris que nos expériences conscientes du monde et du soi sont des formes de prédictions faites par le cerveau – des hallucinations contrôlées – qui émergent par l'intermédiaire et en *raison* de nos corps vivants.

En dépit de sa mauvaise réputation chez les neuroscientifiques, Sigmund Freud avait raison sur de nombreux points. En revenant sur l'histoire des sciences, il identifiait trois « coups décisifs » assénés à la suprématie autoproclamée de l'espèce humaine, chacun de ces coups correspondant à une avancée scientifique majeure dont chacune fut vigoureusement combattue à son époque. Le premier coup fut porté par Copernic, dont la théorie héliocentrée montra que c'est la Terre qui tourne autour du Soleil, et non l'inverse. Tout d'un coup nous avons pris conscience que nous n'étions pas au centre de l'Univers. Nous ne sommes qu'une tache minuscule quelque part dans l'immensité, une minuscule tache bleue suspendue au-dessus de l'abîme. Ensuite vint Darwin, qui révéla que nous avions un ancêtre commun avec tous les autres organismes vivants, un énoncé qui, de façon

surprenante, est encore aujourd'hui combattu dans certaines parties du monde. Le troisième coup porté à l'exceptionnalisme humain identifié par Freud, faisant fi de toute modestie, était sa propre théorie de l'inconscient, qui défiait l'idée que nos processus mentaux étaient sous notre contrôle conscient et rationnel. Bien qu'il ait raté sa cible dans les détails, Freud avait absolument raison de mettre en évidence le fait qu'une explication biologique de l'esprit et de la conscience serait une manière supplémentaire, et peut-être finale, de déloger l'espèce humaine de son trône.

Ces modifications dans la façon dont nous nous voyons nous-mêmes doivent être bien accueillies. Chaque avancée dans notre compréhension correspond à un nouvel émerveillement et à une nouvelle capacité de nous voir comme moins *différents de* et plus une partie du reste de la Nature.

Nos expériences conscientes font partie de la Nature, tout comme nos corps et notre monde. Et quand la vie se termine, notre conscience s'évapore elle aussi. Lorsque j'y pense, je suis à nouveau transporté dans mon expérience – plus exactement ma non-expérience – de l'anesthésie, à l'effacement qu'elle implique, peut-être réconfortant – mais néanmoins effacement. Le romancier Julian Barnes, dans sa méditation sur notre caractère mortel, le dit à la perfection. Quand vient la fin de la conscience, il n'y a rien – réellement *rien* – dont nous devrions avoir peur.

1

Le problème réel

Qu'est-ce que la conscience ?

Pour une créature consciente, il existe quelque chose qui ressemble à *être* cette créature. Il existe quelque chose qui ressemble à être moi, quelque chose qui ressemble à être vous et probablement quelque chose qui ressemble à être un mouton ou un dauphin. Toutes ces créatures sont douées d'expériences subjectives. Cela ressemble à quelque chose d'être moi. Mais il n'y a à peu près certainement rien qui ressemble à être une bactérie, une touffe d'herbe ou un robot jouet. Dans ces différents cas, il n'y a (probablement) aucune expérience subjective en cours : ni univers intérieur, ni conscience ¹.

Cette façon de voir les choses est étroitement liée au point de vue du philosophe Thomas Nagel, qui a publié en 1974 un article aujourd'hui légendaire intitulé *What is it like to be a bat ?*, ou « Qu'est-ce que cela fait d'être une chauve-souris ? » Dans cet article, Thomas Nagel

1. N.d.T. La phrase anglaise contient *no awareness, no consciousness*. Il existe en anglais deux mots pour conscience : *awareness* et *consciousness*. Pour certains auteurs, ces deux mots sont strictement synonymes.

soutient que si nous, les humains, ne pourrions jamais faire l'expérience de ce qu'est d'être une chauve-souris, il devrait néanmoins exister quelque chose à quoi cela ressemble pour la chauve-souris, d'être une chauve-souris². J'ai toujours eu un faible pour l'approche de Nagel parce qu'elle souligne la *phénoménologie* : les propriétés subjectives de l'expérience consciente, par exemple pourquoi une expérience visuelle possède la forme, la structure et les qualités que l'on éprouve, si on la compare aux propriétés subjectives d'une expérience émotionnelle, ou à celles d'une expérience olfactive. En philosophie, ces propriétés sont souvent appelées des *qualia* : la rougeur du rouge, l'accès de rage de la jalousie, la douleur lancinante ou l'élancement monotone d'une rage de dents.

Pour qu'un organisme soit conscient, il doit posséder une sorte de phénoménologie qui lui est propre. *N'importe quelle* sorte d'expérience – *n'importe quelle* propriété phénoménologique – bénéficie d'un poids identique. Chaque fois qu'il existe une expérience, la phénoménologie est présente ; et chaque fois qu'il y a phénoménologie, il y a conscience. Une créature qui accède à l'existence juste pour un moment sera consciente aussi longtemps qu'il existe quelque chose qui ressemble à être cette créature, même si tout ce qui se passe est un sentiment fugace de douleur ou de plaisir.

Il est utile de bien faire la distinction entre les propriétés phénoménologiques de la conscience et ses propriétés *fonctionnelles* et *comportementales*. Ces dernières font référence au rôle que peut jouer la conscience dans le fonctionnement de nos esprits et de nos cerveaux et dans le comportement dont un organisme est capable, du fait qu'il a des expériences conscientes. Bien que les fonctions et les comportements associés à la conscience soient des sujets importants, ce n'est pas le meilleur endroit pour chercher des définitions. La conscience

2. Cet article fait partie de ceux qui ont eu la plus grande influence sur toute la philosophie de l'esprit. Selon Nagel « un organisme possède des états mentaux conscients s'il existe quelque chose à quoi cela ressemble *d'être* cet organisme – quelque chose à quoi cela ressemble pour cet organisme ». Nagel (1974), p. 2, les italiques sont conformes à l'original.

est d'abord et avant tout le sujet de l'expérience subjective – elle traite de phénoménologie.

Cela peut sembler évident, mais cela n'a pas toujours été le cas. Plusieurs fois dans le passé, être conscient a été confondu avec maîtriser le langage, ou présenter un comportement d'un type particulier. Mais la conscience ne dépend pas d'un comportement visible, comme cela est confirmé durant les rêves, ou pour les personnes souffrant de paralysie totale. Affirmer que le langage est un prérequis pour la conscience reviendrait à dire que les bébés, ou les personnes qui ont perdu la faculté du langage et la plupart, sinon la totalité des animaux non humains, ne sont pas doués de conscience. La pensée complexe et abstraite est juste une petite fraction – mais il est possible que ce soit un trait distinctif de l'espèce humaine – de ce que cela représente d'être conscient.

Certaines théories très en vogue de la science de la conscience continuent à mettre l'accent sur la fonction et le comportement plutôt que sur la phénoménologie. Parmi ces théories se détache tout particulièrement celle de « l'espace de travail neuronal global », qui a été développée depuis plusieurs années par le psychologue Bernard Baars et le neuroscientifique Stanislas Dehaene, entre autres. Selon cette théorie, le contenu mental (les perceptions, les pensées, les émotions, *etc.*) devient conscient lorsqu'il accède à un « espace de travail neuronal global » qui, d'un point de vue anatomique, est réparti sur les régions frontales et pariétales du cortex (le cortex cérébral est la surface externe du cerveau, massivement plissée et faite de neurones très serrés les uns sur les autres³). C'est lorsqu'un contenu mental est diffusé à l'intérieur de cet espace de travail cortical que nous en prenons conscience, et ce contenu peut alors être utilisé pour orienter le comportement de façon bien plus flexible que dans le cas de la

3. Chacun des hémisphères du cortex cérébral possède quatre lobes. Les lobes frontaux sont devant, les lobes occipitaux derrière et les lobes temporaux sur les côtés, près des oreilles. Certains spécialistes identifient un cinquième lobe – le lobe limbique – dans les profondeurs du cerveau.

perception inconsciente. Par exemple, supposons que je perçoive de manière consciente un verre d'eau posé sur la table en face de moi. Je pourrais alors saisir ce verre et le boire, ou le jeter sur mon ordinateur (une tentation presque irrésistible) écrire un poème sur ce verre ou le rapporter à la cuisine après m'être rendu compte qu'il était posé sur la table depuis une quinzaine de jours. La perception inconsciente ne permet pas ce degré de flexibilité dans le comportement.

Une autre théorie très en vogue, appelée « théorie de la pensée d'ordre supérieur », propose qu'un contenu mental devient conscient lorsqu'il existe un processus cognitif d'ordre supérieur qui est d'une certaine façon orienté vers ce contenu, le rendant conscient. Dans cette théorie, la conscience est étroitement liée à des processus comme la *métacognition* – ce qui veut dire connaissance sur la connaissance – qui à nouveau mettent l'accent sur les propriétés fonctionnelles plutôt que sur la phénoménologie, quoique dans une moindre mesure que la théorie de l'espace de travail neuronal global. Comme dans le cas de l'espace de travail neuronal global, les théories de la pensée d'ordre supérieur mettent aussi l'accent sur les régions frontales du cerveau comme régions clés de la conscience.

Bien que ces théories soient intéressantes et influentes, je n'en dirai pas beaucoup plus sur l'une ou l'autre dans ce livre. En effet, elles mettent sur le devant de la scène les aspects fonctionnels et comportementaux de la conscience, alors que l'approche que je vais suivre part de la phénoménologie – de l'expérience elle-même – et c'est seulement à partir de là que l'on peut dire des choses sur la fonction et le comportement.

On peut considérer que la définition de la conscience comme étant « n'importe quelle expérience subjective quelle qu'elle soit » est simple, voire simpliste, mais après tout c'est tant mieux. Lorsqu'un phénomène complexe n'est que partiellement compris, des définitions prématurément précises peuvent se révéler contraignantes et même trompeuses. L'histoire des sciences a maintes fois démontré que les définitions pertinentes évoluent conjointement avec la compréhension

scientifique et servent d'échafaudage au progrès scientifique, plutôt que de point de départ ou de points d'appui autonomes. En génétique, par exemple, la définition d'un « gène » s'est considérablement modifiée à mesure que la biologie progressait. De même, à mesure que notre compréhension de la conscience progresse, sa définition – ou ses définitions – va aussi évoluer. Si nous acceptons pour l'instant que la conscience concerne en premier lieu et avant tout la phénoménologie, alors nous pouvons passer à l'étape suivante.

Comment la conscience émerge-t-elle ? Comment les expériences conscientes sont-elles reliées à la machinerie biophysique à l'intérieur de nos cerveaux et de nos corps ? Comment sont-elles reliées au tourbillon des atomes, des quarks ou des supercordes, ou de quoi que ce soit dont l'Univers est fabriqué en dernière analyse ?

La formulation classique de cette question est connue comme « le problème difficile de la conscience ». Cette expression fut inventée par le philosophe australien David Chalmers au début des années 1990, et elle a depuis défini l'agenda pour une large part de la science de la conscience. Voici la manière dont il la décrit :

« On ne peut pas nier que certains organismes ressentent des expériences. Mais la question de comprendre comment ces systèmes ressentent leurs expériences est source de perplexité. Comment se fait-il que, lorsque nos systèmes cognitifs s'engagent dans le traitement de l'information des processus visuels ou auditifs, nous ayons des expériences visuelles ou auditives : la qualité d'un bleu profond ou la sensation d'un la mineur ? Comment pouvons-nous expliquer qu'il existe quelque chose qui ressemble à accueillir une image mentale ou à faire l'expérience d'une émotion ? Il existe un consensus quasi-unanime sur le fait que les expériences émergent d'une base physique, mais nous n'avons pas d'explication convaincante de pourquoi et comment cela se

produit ainsi. Pourquoi un processus physique serait-il susceptible de donner lieu à une riche vie intérieure ? Objectivement, cela semble tout à fait déraisonnable, et pourtant c'est ce qui se produit. »

Chalmers oppose ce problème difficile de la conscience à ce qu'il appelle le problème facile de la conscience – ou les problèmes faciles de la conscience – qui se préoccupe d'expliquer la façon dont des systèmes physiques comme nos cerveaux peuvent donner naissance à nombre de propriétés fonctionnelles ou comportementales. Ces propriétés fonctionnelles abordent par exemple des sujets comme le traitement des signaux sensoriels, la sélection des actions et le contrôle du comportement, le fait de prêter attention, la génération du langage et ainsi de suite. Les problèmes faciles recouvrent tout ce que les êtres que nous sommes faisons et tout ce que l'on peut caractériser en termes d'une fonction – comment une donnée est traduite en un résultat – ou en termes d'un comportement.

Bien évidemment ces problèmes sont tout sauf faciles. Leur solution va occuper des neuroscientifiques pour les décennies ou les siècles à venir. Le point de vue de Chalmers est que les problèmes faciles sont simples à résoudre en principe, alors que l'on ne peut pas en dire autant du problème difficile. Plus précisément, Chalmers considère que l'on ne rencontre pas d'obstacle conceptuel si l'on veut expliquer comment ces problèmes faciles finiront par trouver une solution en termes de mécanismes physiques. Au contraire, il semble pour Chalmers qu'aucune explication du problème difficile en termes de processus physiques ne soit à la hauteur de la tâche. Pour être précis, un « mécanisme » peut être défini comme une chaîne de parties en interactions causales capables de produire l'effet recherché. « Même une fois que tous les problèmes faciles auront été résolus l'un après l'autre, le problème difficile n'en aura pas pour autant progressé. » Même lorsque nous aurons expliqué comment s'accomplissent toutes les fonctions qui se rapportent à l'expérience – la discrimination

des perceptions, la catégorisation, l'accès interne, le compte rendu verbal – il restera encore très vraisemblablement une dernière question sans réponse : « *Pourquoi le fait que ces fonctions puissent s'accomplir est-il accompagné d'une expérience consciente ?* »

Les racines du problème difficile remontent à la Grèce antique, et peut-être même avant, mais elles sont particulièrement évidentes au dix-septième siècle chez René Descartes, qui a scindé l'Univers en matière de l'esprit, *res cogitans*, et matière concrète, *res extensa*. Cette distinction a été le point de départ de la philosophie du dualisme, qui a depuis lors rendu toutes les discussions sur la conscience compliquées et confuses. Cette confusion est particulièrement évidente dans la prolifération et la diversité des cadres philosophiques qui ont servi à réfléchir sur la conscience.

Attachez vos ceintures, c'est maintenant qu'arrivent les « ismes » !

La position philosophique qui me guide, et qui est aussi l'hypothèse par défaut de nombreux neuroscientifiques, est le *physicalisme*. La base du physicalisme est que l'Univers est fait de matière physique et que les états de la conscience sont soit identiques à des, ou émergent de, configurations spécifiques de cette matière physique. Certains philosophes utilisent le terme *matérialisme* au lieu de physicalisme, mais en ce qui nous concerne ces deux termes sont équivalents.

À l'opposé du physicalisme se situe l'*idéalisme*. C'est l'idée – fréquemment associée à l'évêque du dix-huitième siècle George Berkeley – selon laquelle la conscience ou l'esprit sont les sources ultimes de la réalité, et non une quelconque matière ou essence physique. Le problème n'est plus de comprendre comment l'esprit émerge de la matière mais comment la matière émerge de l'esprit.

Se positionnant quelque peu inconfortablement entre ces deux extrêmes, les *dualistes* comme Descartes croient que la conscience (l'esprit) et la matière physique sont des modes séparés de l'existence, ce qui pose le problème épineux de la façon dont ils interagissent. Aujourd'hui peu de philosophes ou de scientifiques endosseraient cette façon de voir. Mais pour bien des personnes, au moins en

Occident, le dualisme reste attrayant. L'intuition séduisante que les expériences conscientes *semblent* non physiques encourage un « dualisme naïf », où ce « semble » nous attire vers des croyances sur ce que sont réellement les choses. Ainsi que nous le verrons à maintes reprises dans ce livre, l'apparence des choses est souvent un guide très peu fiable si l'on veut comprendre ce qu'elles sont réellement.

Une des variantes les plus influentes du physicalisme est le *fonctionnalisme*. Tout comme le physicalisme, le fonctionnalisme est une hypothèse courante et en général implicite faite par les neuroscientifiques. Nombre de ceux qui tiennent le physicalisme pour évident font de même pour le fonctionnalisme. Mon point de vue personnel, cependant, est d'être agnostique et un peu méfiant.

Pour le fonctionnalisme, la conscience ne dépend pas de ce dont un système est fabriqué, de son contenu physique, mais uniquement de ce que le système fait, des fonctions qu'il accomplit, de la façon dont il transforme des données en résultats. L'intuition qui sous-tend le fonctionnalisme est que l'esprit et la conscience représentent des formes de traitement de l'information qui peuvent certes être mises en œuvre par un cerveau, mais pour lesquelles un cerveau biologique n'est pas strictement indispensable.

Remarquez la façon dont l'expression « traitement de l'information » vient de se faufiler sans crier gare, comme elle l'a aussi fait dans la citation de Chalmers quelques pages plus haut. Ce terme est si courant dans les discussions sur l'esprit, le cerveau et la conscience qu'il est aisé de l'introduire sans précaution. Ce serait une erreur, car la suggestion que le cerveau « traite l'information » dissimule quelques hypothèses fortes. Selon la personne qui fait les hypothèses, cela va de l'idée que le cerveau est une sorte d'ordinateur (le matériel, *hardware*) tandis que l'esprit et la conscience sont le logiciel⁴ (*mindware*), en passant par des hypothèses sur ce que l'information elle-même est réellement. Toutes ces hypothèses nous entraînent sur un terrain

4. N.d.T. Le texte anglais contient un jeu de mots intraduisible sur *software* et *mindware*.

glissant. Un cerveau est complètement différent d'un ordinateur, du moins si l'on se réfère aux ordinateurs qui nous sont familiers. Et la question de ce qu'est vraiment l'information est au moins aussi épineuse que ce qu'est vraiment la conscience, ainsi que nous le verrons ultérieurement. C'est pour ces raisons que je me méfie du fonctionnalisme.

Prendre le fonctionnalisme au pied de la lettre, comme beaucoup le font, est porteur d'une implication incontournable : la conscience est quelque chose qui peut être *simulée* sur un ordinateur. Rappelez-vous que, pour le fonctionnalisme, la conscience dépend uniquement de ce qu'un système fait, pas de ce dont il est fabriqué. Cela veut dire que si vous reproduisez correctement les relations fonctionnelles – si vous vous assurez que le système possède l'application correcte qui mène de l'entrée vers le résultat – cela sera suffisant pour donner naissance à la conscience. Autrement dit, pour les fonctionnalistes, *simulation* est équivalent à *instanciation* – ce qui veut dire accéder à la réalité, prendre la forme d'une réalité.

Est-ce bien raisonnable ? Dans certains cas, la simulation est certainement équivalente à l'instanciation. Un ordinateur qui joue au Go, comme l'ordinateur Alpha Zero de la compagnie britannique Deep Mind, celui qui a vaincu le champion du monde humain de Go, joue *réellement* au Go. Mais dans de nombreuses situations ce n'est pas le cas. Prenez le cas de la prévision météorologique. Les simulations sur ordinateur, quelle que soit leur sophistication actuelle, ne sont ni mouillées ni venteuses. Est-ce que la conscience ressemble davantage au Go ou davantage à la météo ? Ne vous attendez pas à une réponse, il n'y en a pas, du moins pas encore au jour d'aujourd'hui. Il suffit de se rendre compte qu'il s'agit d'une question valable. C'est pourquoi je reste agnostique sur le fonctionnalisme.

Encore deux autres « ismes » et nous en aurons terminé.

Le premier sur la liste est le *panpsychisme*. Le panpsychisme repose sur l'idée que la conscience est une propriété fondamentale de l'Univers, tout comme la masse/énergie ou la charge électrique :

elle est présente à un certain point partout et dans toute chose. On se moque parfois du panpsychisme en affirmant que, dans cette vision philosophique, des objets comme les pierres ou les cuillers sont conscients, de la même façon que vous et moi sommes conscients, mais ce sont là des interprétations délibérément fausses, conçues de manière à discréditer le panpsychisme, de le faire apparaître comme stupide. Il existe des versions plus sophistiquées de cette conception, dont nous verrons quelques exemples dans les chapitres suivants, mais le problème principal avec le panpsychisme n'est pas qu'il repose sur une idée folle – après tout il arrive que des idées folles se révèlent exactes, ou au moins utiles. Le problème principal est qu'il n'explique pas vraiment quoi que ce soit et il ne peut pas conduire à des hypothèses que l'on pourrait tester. C'est une échappatoire trop facile au mystère apparent de la conscience posé par le problème difficile de Chalmers, et l'adopter comme une ouverture possible vers la science de la conscience mène de fait à une impasse.

Finalement on trouve le *mystérianisme*, associé à la philosophie de Colin McGinn. Le mystérianisme repose sur l'idée qu'il peut exister une explication physique complète de la conscience – une solution complète au problème difficile de Chalmers – mais nous, pauvres humains, ne sommes juste pas assez intelligents pour découvrir cette solution, ou même pour reconnaître la pertinence de cette solution si elle nous était offerte sur un plateau par des extraterrestres super-intelligents. Il existe une compréhension physique de la conscience, mais elle ne nous est pas accessible, exactement comme les grenouilles ne peuvent pas accéder au concept de cryptomonnaie. Cette solution est hors de portée des capacités cognitives de l'espèce humaine, nous avons une intelligence limitée.

Que peut-on dire du mystérianisme ? Il est bien possible qu'il existe des choses que nous ne pourrions jamais comprendre en raison des limitations de nos cerveaux et de nos esprits. De fait, personne n'est capable de comprendre la globalité du fonctionnement d'un Airbus A380, et pourtant j'ai beaucoup apprécié d'avoir pris place sur

un siège d'A380 un jour où je revenais de Dubaï. Il existe certainement des sujets hors de portée cognitive pour la plupart d'entre nous, même s'ils sont compréhensibles *en principe* par des humains, par exemple certains points de détail de la théorie des supercordes en physique. Comme les cerveaux sont des systèmes finis disposant de ressources finies, et comme certains cerveaux semblent incapables de comprendre certaines choses, il semble inévitable que ce soit aussi le cas pour l'ensemble des humains en ce qui concerne certains sujets. Cependant, c'est faire preuve d'un pessimisme injustifié que d'inclure préventivement la conscience dans ces domaines non cartographiés de l'ignorance spécifique à notre espèce⁵.

Un des points les plus remarquables de la recherche scientifique est qu'elle est cumulative et qu'elle progresse pas à pas. Aujourd'hui, nombre d'entre nous peuvent comprendre des choses qui auraient semblé complètement incompréhensibles, *même en principe*, à nos ancêtres, et même à des scientifiques et à des philosophes en activité il y a quelques décennies. Au fil du temps, un mystère après l'autre a cédé à l'application systématique du raisonnement et de l'expérience. Si nous prenions le mystérianisme comme une option sérieuse, nous ferions mieux de tout abandonner et de rentrer chez nous. Mais nous n'allons pas le faire.

Les « ismes » fournissent différentes options pour réfléchir sur le lien entre la conscience et le reste de l'Univers pris dans son ensemble. Lorsque l'on essaie d'attribuer un poids aux mérites et aux insuffisances de ces « ismes », il est important de reconnaître que ce qui importe le plus n'est pas de déterminer lequel de ces cadres est « correct », dans le sens où l'on peut démontrer qu'il approche la vérité, mais de déterminer le cadre qui sera le plus utile pour progresser dans notre compréhension de la conscience. C'est pourquoi

5. N.d.T. On se souvient de la déclaration de Donald Rumsfeld, secrétaire à la défense de George Bush, en 2002 : « ... il y a le *connu connu*, c'est-à-dire les choses que nous savons ; nous savons aussi qu'il y a le *connu inconnu*, c'est-à-dire les choses que nous savons que nous ne savons pas ; mais il y a aussi l'*inconnu inconnu*, les choses que nous ne savons pas que nous ne savons pas ».

je penche en faveur d'une sorte de variante fonctionnellement agnostique du physicalisme. Pour moi, c'est l'état d'esprit le plus pragmatique et le plus productif que l'on puisse adopter pour progresser dans l'étude de la science de la conscience. Et, en ce qui me concerne, c'est aussi ce que je trouve de plus honnête intellectuellement.

En dépit de son pouvoir d'attraction, le physicalisme est loin d'être universellement accepté dans la communauté des scientifiques travaillant sur la conscience. Un des défis les plus courants au physicalisme est ce que l'on appelle l'expérience de pensée du zombie. Les zombies dont il s'agit ne sont pas les cadavres ambulants dévoreurs de cerveaux des films d'horreur, nos zombies sont des « zombies philosophiques ». Mais il nous faut tout aussi bien nous en débarrasser parce que, dans le cas contraire, une explication physique de la conscience est mort-née avant même que nous puissions la démarrer.

Un zombie philosophique est une créature que l'on ne peut pas distinguer d'une créature consciente, mis à part le fait qu'il est dépourvu de conscience. Un zombie d'Anil Seth me ressemblerait en tout point, agirait comme moi, marcherait comme moi et parlerait comme moi, mais il n'aurait rien qui ressemble à être moi, il ne posséderait ni univers intérieur ni expérience vécue. Demandez au zombie d'Anil Seth s'il est conscient, il vous répondra « Oui, je suis conscient ». Le zombie d'Anil Seth aurait même écrit divers essais sur les neurosciences de la conscience, en incluant quelques réflexions douteuses sur la pertinence des zombies philosophiques pour ce sujet. Mais rien de tout cela n'impliquerait une expérience consciente.

Voici maintenant pourquoi l'idée du zombie est supposée fournir un argument contredisant une explication physique de la conscience. Si vous pouvez imaginer un zombie, cela veut dire que vous pouvez concevoir un monde qui ne peut pas être distingué de notre monde, mais dans lequel aucune conscience n'est présente. Et si vous pouvez concevoir un tel monde, alors la conscience ne peut pas être un phénomène physique.

Voici pourquoi cette argumentation est erronée. L'argument du zombie, comme nombre des arguments qui s'attaquent au physicalisme, est un argument de concevabilité, et les arguments de concevabilité sont intrinsèquement faibles. Comme beaucoup de ces arguments, ils ont une plausibilité inversement proportionnelle au degré de connaissance que nous avons atteint.

Est-ce que vous pouvez imaginer un Airbus A380 volant à l'envers ? Bien sûr vous le pouvez. Il suffit d'imaginer un gros avion flottant dans l'air et se déplaçant à reculons. Un tel scénario est-il *réellement* concevable ? Eh bien, plus vous en savez sur l'aérodynamique et l'ingénierie aéronautique, et moins ce scénario vous paraît concevable. Dans ce cas précis, même des connaissances sommaires sur le sujet suffisent à rendre clair le fait que les avions ne peuvent pas voler à reculons. Cela ne peut simplement pas exister ⁶.

C'est la même chose avec les zombies. En un sens, c'est trivial d'imaginer un zombie philosophique. Il faut juste que je me représente une image de moi déambulant ici et là et qui n'ait aucune perception consciente. Mais est-ce que je peux *réellement* le concevoir ? Ce que l'on me demande de faire, *en réalité*, est de prendre en considération les capacités et les limitations d'un vaste réseau de plusieurs milliards de neurones et de zillions de synapses (les connexions entre neurones), pour ne pas mentionner les cellules gliales, les gradients de neurotransmetteurs et autres friandises neurobiologiques, toutes enveloppées dans un corps qui interagit avec un monde incluant lui-même d'autres cerveaux dans d'autres corps. Est-ce que je peux faire cela ? Est-ce que quelqu'un peut le faire ? Juste comme dans le cas de l'A380, plus vous en savez sur le cerveau et ses liens avec les expériences conscientes, et moins un zombie devient concevable. Que quelque chose soit ou non concevable est souvent une observation

6. Un hélicoptère, qui peut voler à reculons, n'est pas un avion. J'ai été étrangement heureux de découvrir que l'origine du mot « hélicoptère » n'était pas une combinaison de « héli » et « copter », ce que j'avais toujours supposé, mais plutôt de « helico » (spirale) et « pter » (aile). Pour moi, le mot fait vraiment sens maintenant.

psychologique sur la personne qui met en avant la concevabilité, et non un éclairage sur la nature de la réalité. Voilà la faiblesse du zombie. On nous demande d'imaginer ce qui n'est pas imaginable, et à travers cet acte de compréhension illusoire, on tire des conclusions sur les limites de l'explication par le physicalisme.⁷

Nous sommes maintenant prêts à attaquer ce que j'appelle le *problème réel* de la conscience. C'est une façon de penser la conscience qui a pris forme pour moi sur de nombreuses années, en assimilant et en m'appuyant sur les intuitions d'autres scientifiques et philosophes. Aborder le problème réel est, je le pense, l'approche qui offre avec la plus grande probabilité les meilleures perspectives de succès à une science de la conscience.

Si l'on se réfère au problème réel, l'objectif premier d'une science de la conscience est *d'expliquer, de prédire* et de *contrôler* les propriétés phénoménologiques de la perception consciente. Cela veut dire expliquer *pourquoi* une expérience consciente spécifique est ce qu'elle est – pourquoi elle présente ces propriétés phénoménologiques spécifiques – en termes de mécanismes physiques et de processus dans le cerveau et dans le corps. Ces explications devraient nous rendre capables de prédire quand des expériences subjectives spécifiques vont se produire et permettre de les contrôler par une intervention sur les mécanismes sous-jacents. En bref, prendre à bras-le-corps le problème réel exige d'expliquer *pourquoi* un motif particulier de l'activité cérébrale – ou d'autres processus physiques – correspond à une catégorie particulière d'expérience consciente, en ne se contentant pas simplement d'établir qu'il le fait.

7. Un cerveau humain adulte contient environ 86 milliards de neurones et à peu près mille fois plus de connexions. Si vous comptiez une connexion par seconde, cela vous prendrait environ 3 millions d'années pour terminer le décompte. Et par-dessus le marché, il est de plus en plus clair que des neurones uniques peuvent effectuer des opérations complexes par eux-mêmes.

Le problème réel est différent du problème difficile parce que son objectif n'est pas – au moins au départ – d'expliquer en premier lieu pourquoi et comment la conscience fait partie de l'Univers. Il n'est pas à la poursuite de l'ingrédient spécial qui pourrait faire émerger par magie la conscience de tel ou tel mécanisme (ou inversement). Il est aussi différent du (des) problème(s) facile(s), parce qu'il se focalise sur la phénoménologie plutôt que sur la fonction ou le comportement. Il ne balaie pas sous le tapis les aspects subjectifs de la conscience. Et, en raison de l'accent mis sur les mécanismes et les processus, le problème réel est naturellement en phase avec une vision du monde physicaliste de la relation entre l'esprit et la matière.

Afin de clarifier ces distinctions, demandons-nous comment les différentes approches pourraient essayer d'expliquer l'expérience subjective du « rouge ».

Si l'on se place du point de vue du problème facile, le défi est d'expliquer toutes les propriétés mécanistes, fonctionnelles et comportementales associées à l'expérience du rouge : comment des longueurs d'onde spécifiques de la lumière activent le système visuel, les conditions dans lesquelles nous affirmons des choses comme « cet objet est rouge », le comportement typique en face de feux de signalisation et comment il peut arriver que le rouge suscite des réponses émotionnelles d'un type particulier, et ainsi de suite.

Ce que l'approche du problème facile laisse de côté, par construction, est l'explication de pourquoi et comment ces propriétés mécanistes, fonctionnelles et comportementales sont accompagnées par une quelconque phénoménologie – dans ce cas le phénomène du « rouge ». L'existence d'une expérience subjective, par opposition à l'absence d'une telle expérience, est le domaine de prédilection du problème difficile. Quelle que soit l'information mécaniste dont vous disposez, il ne sera jamais déraisonnable de poser la question « d'accord, mais pourquoi ce mécanisme est-il associé à une expérience consciente ? » Si vous prenez le problème difficile à cœur, vous

aurez toujours le soupçon d'un fossé entre les explications mécanistes et l'expérience subjective de « voir du rouge ».

Le problème réel accepte que les expériences conscientes existent et se focalise en premier lieu sur leurs propriétés phénoménologiques. Par exemple : les expériences du rouge sont visuelles, elles sont généralement, mais pas obligatoirement, liées à des objets, elles semblent être des propriétés des surfaces, elles ont différents niveaux de saturation, elles définissent une catégorie parmi les différentes expériences de la couleur bien qu'elles puissent varier graduellement à l'intérieur d'une catégorie, et ainsi de suite. Ce qui est important, c'est que ce sont toutes des propriétés de *l'expérience elle-même* et non – du moins pas au départ – des propriétés fonctionnelles associées à cette expérience. Le défi, pour le problème réel, est d'expliquer, de prédire et de contrôler ces propriétés phénoménologiques en termes de ce qui se passe dans le cerveau et le corps. Nous aimerions savoir ce qu'il en est en termes de motifs spécifiques d'activité dans le cerveau – tels que l'activité complexe en boucle dans le cortex visuel⁸ – qui explique, prédit et contrôle la raison pour laquelle une expérience, comme l'expérience du rouge, a cet aspect spécifique et pas un autre. Pourquoi ce n'est pas comme l'expérience d'un bleu, ou d'une rage de dents, ou de la jalousie.

Explication, prédiction et contrôle. Voilà les critères suivant lesquels les projets scientifiques sont évalués, aussi mystifiants qu'aient pu apparaître au départ les phénomènes qu'ils ciblaient. Les physiciens ont parcouru un chemin considérable dans le décryptage des secrets de l'Univers – en expliquant, prédisant et contrôlant ses propriétés – mais ils restent muets si on leur demande de quoi est fait l'Univers ou pourquoi il existe. Exactement de la même manière, la science de la conscience peut faire de grands progrès en faisant la lumière sur les propriétés et la nature de l'expérience consciente, sans qu'il soit nécessaire d'expliquer comment ou pourquoi elles se trouvent faire partie de l'Univers dans lequel nous vivons.

8. Le cortex visuel est situé dans le lobe occipital, à l'arrière du cerveau.

Nous ne devrions pas non plus nous attendre nécessairement à ce que les explications scientifiques soient toujours intuitivement satisfaisantes. En physique, la mécanique quantique a la réputation justifiée d'être non intuitive, mais elle est néanmoins acceptée comme la théorie qui appréhende de la façon la plus satisfaisante la nature de la réalité physique. Il est également fort possible qu'une science de la conscience arrivée à maturité soit capable d'expliquer, prédire et contrôler les propriétés phénoménologiques sans jamais donner ce sentiment d'évidence intuitive, « oui, c'est vrai, *bien sûr* il fallait qu'il en soit ainsi ! »

Il est important de comprendre que mettre en avant le problème réel de la conscience ne revient pas à tracer une croix sur le problème difficile. Le problème réel s'attaque au problème difficile indirectement, mais il ne l'évite pas. Afin de comprendre pourquoi c'est le cas, je vais introduire la notion de « corrélations neuronales de la conscience ».

Je suis toujours stupéfait lorsque je me rappelle à quel point la science de la conscience était discréditée il y a seulement trente ans. En 1989, une année avant que je commence mes études de licence à l'Université de Cambridge, une des sommités en psychologie, Stuart Sutherland, écrivait :

« La conscience est un phénomène fascinant, mais insaisissable. Il est impossible de définir ce que c'est, ce qu'elle fait ou pourquoi elle est le produit de l'évolution. Rien de ce qui a été écrit sur le sujet ne vaut la peine d'être lu ».

Cet énoncé accablant est paru dans pas moins que le *Dictionnaire international de psychologie* et il saisis l'attitude envers la conscience à laquelle j'ai été confronté au début de ma carrière académique.

Loin de Cambridge, et bien que je ne l'aie pas su à l'époque, la situation était plus prometteuse. Francis Crick, co-découvreur de la structure moléculaire de l'ADN avec Rosalind Franklin et James Watson, et son collègue Christof Koch, tous deux basés à San Diego en

Californie, développaient ce qui allait devenir la méthode dominante dans l'émergence de la science de la conscience : la recherche des corrélations neuronales de la conscience.

La référence ultime en matière de corrélation neuronale de la conscience, ou CNC, tient dans la définition la suivante : c'est « l'ensemble des mécanismes neuronaux minimaux qui, pris conjointement, sont suffisants pour rendre possible n'importe quelle perception consciente spécifique. » L'approche CNC affirme qu'il existe un motif spécifique d'activité neuronale qui est responsable de n'importe quelle expérience sensible, comme celle de la perception du « rouge ». Chaque fois que ce motif est présent, une expérience du « rouge » surviendra, et si ce motif est absent, il n'y aura aucune perception de cette sorte.

Le grand mérite de l'approche CNC est qu'elle fournit un cadre pratique pour développer des recherches. Pour identifier une CNC, tout ce que vous avez à faire est de concocter une situation dans laquelle des personnes ont à certains moments une expérience consciente particulière, et à d'autres moments ne l'ont pas, tout en s'assurant que toutes les autres conditions soient aussi identiques que possible. Étant donné une telle situation, vous pouvez comparer l'activité cérébrale correspondant à ces deux expériences sensorielles, en utilisant des méthodes d'imagerie du cerveau comme l'imagerie par résonance magnétique fonctionnelle (IRMf) ou l'électro-encéphalographie (EEG)⁹. L'activité neuronale spécifique à un état de conscience identifie la CNC pour cette expérience particulière.

Le phénomène de la « rivalité binoculaire » fournit un exemple éclairant. Dans la rivalité binoculaire, une image différente est proposée à chacun des deux yeux, par exemple un visage à l'œil droit et une

9. L'imagerie par résonance magnétique fonctionnelle (IRMf) mesure un signal métabolique (l'oxygénation du sang) lié à l'activité cérébrale, qui fournit une bonne résolution spatiale mais a un lien seulement indirect avec ce que font les neurones. L'EEG mesure les minuscules signaux électriques générés par l'activité d'une population importante de neurones près de la surface du cortex. Cette méthode piste l'activité cérébrale plus directement que l'IRMf, mais avec une moins bonne résolution spatiale.

maison à l'œil gauche. Dans cette situation, la perception consciente ne se fixe pas sur une chimère étrange visage-maison. Elle bascule alternativement entre le visage et la maison en se fixant sur l'une des deux images pendant quelques secondes. D'abord vous voyez une maison, puis un visage, et à nouveau une maison. . . et ainsi de suite. Le résultat crucial de cette expérience est que la perception consciente change alors que la donnée sensorielle, la donnée fournie au système visuel, est invariable. En examinant ce qui se passe dans le cerveau, il est donc possible de distinguer l'activité qui piste la perception consciente de celle qui piste une donnée sensorielle quelle qu'elle soit. L'activité cérébrale qui correspond à une perception consciente identifie la CNC pour cette perception.

La stratégie des CNC a produit des résultats impressionnants au cours de ces dernières années, en fournissant des moissons de découvertes fascinantes, mais ses limitations deviennent apparentes. Un des problèmes est qu'il est difficile, sinon impossible, de démêler une « véritable » CNC d'un ensemble de facteurs avec lesquels on peut potentiellement la confondre, les plus importants étant les phénomènes neuronaux qui sont soit des prérequis, soit des conséquences d'une CNC elle-même. Dans le cas de la rivalité binoculaire, l'activité cérébrale qui accompagne la perception consciente peut en fait pister en amont (le prérequis) des processus comme « concentrer son attention » et pister, en aval, le comportement verbal qui consiste à « faire un rapport verbal », à verbaliser que vous avez vu, un visage ou une maison. Bien qu'ils soient liés au flux de la perception consciente, les mécanismes neuronaux responsables de l'attention et du rapport verbal – ou bien d'autres prérequis en amont ou d'autres conséquences en aval – ne doivent pas être confondus avec les mécanismes responsables de la perception consciente elle-même.

Mais un problème encore plus fondamental est que les *corrélations* ne sont pas des *explications*. Nous savons tous que de simples corrélations n'établissent pas un lien de causalité, mais il est également vrai que les corrélations ne sont pas des explications. Même avec des

dispositifs expérimentaux ingénieux et une imagerie du cerveau toujours plus performante, les corrélations ne peuvent jamais par elles-mêmes servir d'explication. Dans cette perspective, la stratégie CNC et le problème difficile sont des compagnons de route naturels. Si nous nous restreignons à collecter les corrélations entre ce qui se produit dans le cerveau et ce qui se produit dans notre expérience, nous découvrirons sans surprise que nous allons toujours soupçonner l'existence d'un fossé dans l'explication du phénomène par la physique. Mais si au contraire nous allons au-delà de l'établissement de corrélations pour mettre au jour des explications qui relient les propriétés des mécanismes neuronaux à celles de l'expérience subjective, ainsi que le préconise l'approche du problème réel, alors ce fossé va se combler et pourrait même disparaître totalement. Lorsque nous serons capables de prédire (et d'expliquer et de contrôler) pourquoi l'expérience du « rouge » est ce qu'elle est, et non une expérience de couleur bleue ou de jalousie, le mystère de comment le « rouge » se produit deviendra moins mystérieux et perdra peut-être même tout son mystère.

L'ambition de l'approche fondée sur le problème réel est la suivante : à mesure que nous jetons des ponts explicatifs entre le physique et le phénoménologique, l'intuition suggérée par le problème difficile selon laquelle la conscience ne peut jamais être comprise en termes de phénomènes physiques va finalement se volatiliser en s'évaporant en une bouffée de fumée métaphysique. Quand ce sera le cas, nous aurons entre nos mains une science de la conscience satisfaisante et totalement convaincante.

Qu'est-ce qui peut justifier cette ambition ? Prenons l'exemple de la façon dont la compréhension physique de la vie a mûri au cours de ces deux derniers siècles.

Il n'y a pas si longtemps, la vie semblait aussi mystérieuse que la conscience l'est aujourd'hui. Les scientifiques et les philosophes doutaient que des mécanismes physiques ou chimiques puissent un jour expliquer la propriété d'être en vie. La différence entre le vivant

et le non-vivant, entre l'animé et l'inanimé, semblait si fondamentale que l'on considérait comme tout à fait non plausible l'idée que l'on puisse jeter un pont explicatif entre les deux grâce à un mécanisme physique de quelque nature que ce soit.

Cette philosophie du *vitalisme* a atteint son apogée au dix-neuvième siècle. Elle était soutenue par des biologistes de pointe comme Johannes Müller ou Louis Pasteur, et elle a perduré jusqu'au début du vingtième siècle. Les vitalistes considéraient que la propriété d'être vivant pouvait être expliquée uniquement si l'on faisait appel à un ingrédient particulier : une étincelle de vie, un élan vital. Mais nous savons que nous n'avons besoin d'aucun ingrédient spécial. Le vitalisme est aujourd'hui rejeté par la communauté scientifique. Bien que nombre de domaines se rapportant à la vie restent encore inconnus – par exemple comment fonctionne une cellule – l'idée qu'être vivant exige une sorte d'ingrédient surnaturel a perdu toute crédibilité. Le vice de forme fatal du vitalisme a été d'interpréter un manque d'imagination comme l'intuition d'une nécessité. C'est le même vice de forme qui rend inopérant l'argument du zombie.

Les sciences de la vie ont été capables d'évoluer au-delà de la myopie vitaliste grâce à une focalisation sur les progrès pratiques, en mettant l'accent sur le « problème réel » de ce que veut dire être vivant. Sans se laisser décourager par le pessimisme vitaliste, les biologistes se sont attelés à la description des propriétés des systèmes vivants et ensuite sur l'explication (et aussi sur la prédiction et le contrôle) de ces propriétés en termes de processus physiques et chimiques. La reproduction, le métabolisme, la croissance, le développement, l'autoréparation ou l'autorégulation homéostatique sont tous devenus individuellement et collectivement susceptibles d'explications mécanistes. À mesure que les détails ont été compris, et on continue à les comprendre, non seulement le mystère de base de ce qu'« est la vie » s'est estompé mais le concept même de vie s'est tellement ramifié que « être vivant » n'est plus considéré comme une propriété de tout ou rien. Des zones grises se font jour, avec bien sûr les virus en tête