L'expérience consciente est-elle indubitable ?

Sylvain Montalvo

L'EXPÉRIENCE CONSCIENTE EST-ELLE INDUBITABLE ?

Mémoire de master 1

Mention philosophie

Université de Lille

Année universitaire 2018-2019

Sylvain Montalvo

Sous la direction de M. Alexandre Billon

RÉSUMÉ

Qu'y a-t-il de plus certain que l'existence et la nature de nos expériences conscientes ? Elles semblent nous accompagner au quotidien : enrichir notre perception du monde et fonder notre identité en nous distinguant des objets qui en sont dépourvus. Elles nous paraissent si immédiates que nous regardons avec suspicion toute remise en cause de leur réalité. Cependant, ces expériences posent un problème : elles trouvent difficilement leur place dans une théorie matérialiste du monde qui, quant à elle, enrichit notre compréhension de celui-ci tout en relativisant notre anthropocentrisme. Pour réconcilier ces perspectives dissonantes, une tradition philosophique, le matérialisme réductionniste, propose de rendre compte de nos expériences conscientes dans des termes conformes à la théorie matérialiste du monde. Nous allons dans cet essai soutenir que cette réconciliation est impossible, nous laissant en définitive face à un dilemme : inclure des propriétés non matérielles dans notre théorie du monde ou rejeter, en dépit de son caractère indubitable, l'existence des expériences conscientes. Afin de proposer une résolution à ce dilemme, nous allons explorer minutieusement ce que sont le doute et la certitude. Une caractérisation satisfaisante de ces notions nous permettra d'évaluer leur valeur épistémique concernant nos expériences conscientes. Si l'expérience consciente est indubitable, cela requiert-il d'ériger son existence en prémisse indispensable à toute description sérieuse *du monde ?*

TABLE DES MATIÈRES

INTRODUCTION : IRRÉDUCTIBLES QUALIA

L'approche réductionniste, quand elle est possible, consiste à reformuler une question appelant explication dans un vocabulaire plus fondamental, dans lequel les lois explicatives (causale, téléologique, historique, etc.) sont connues. Néanmoins, tout problème ne peut bénéficier de cette technique. On le qualifie alors d'irréductible. Dans ce cas, la tâche s'annonce plus ardue : il devient nécessaire de concevoir de nouveaux outils théoriques, de nouvelles entités ou propriétés, de nouvelles lois susceptibles de rendre compte des phénomènes considérés. Il faut ensuite les tester. Parfois, des lois existent également pour un phénomène pourtant réductible. La chimie dispose d'un corpus propre de lois, mais elles sont en définitive réductibles à celles de la physique. Pourtant, les lois et les entités postulées par la chimie gardent une légitimité pratique. C'est d'ailleurs le cas de la plupart de nos sciences, puisqu'alors qu'elles ont toutes leurs spécificités, il semble qu'elles puissent toutes être, plus ou moins directement, réduites à la physique. Ce point est discuté pour certaines sciences, telles que la sociologie, dans le cadre desquelles on entend dire que *le tout est plus que la somme des parties* : le tout ne serait pas réductible à la somme de parties reformulables dans un lexique plus fondamental. Cependant, nous pouvons penser que la somme des parties dont il est fait mention ici n'inclut pas certaines données restées implicites et dont la non-prise en compte donne l'illusion d'une irréductibilité. Un démon de Laplace, connaissant tout ce qu'il y a à connaître à propos du monde physique, ne serait-il pas en d'anticiper les prédictions de la sociologie à partir de l'évolution prévisible des particules qui composent les personnes qui composent les sociétés ? Ce qui lui manquerait c'est une explication sociologique et non une explication tout court.[1] Selon David Chalmers (1996), seul un phénomène semble poser un problème tout particulier à l'approche réductionniste : celui de la conscience. Dans cette introduction, nous allons explorer la méthode réductionniste et son applicabilité à la conscience. Nous allons tenter de dresser un contour des phénomènes de la conscience qui appellent à une explication et présenter sous forme d'un dilemme les deux types de solutions envisageables si la méthode réductionniste ne s'applique pas.

[1] Cette apparente supériorité du tout sur les parties pourrait également trouver son origine dans des définitions ambiguës du tout ou de la partie. Pour reprendre notre exemple, la *société* serait le tout dont les parties seraient des *personnes*. La pertinence de la notion de société peut être discutée : peut-être n'est-elle rattachée à aucune réalité et n'est que le fruit de l'imagination de l'homme, prompte à stipuler des concepts là où il voit des motifs récurrents. Le concept de personne est lui aussi sujet à interprétations, comme le montre Derek Parfit (1987). Il est envisageable que l'une ou l'autre de ces notions puisse être reformulée d'une manière telle que le tout concorde avec la somme des parties.

Reducto

« — Reducto ! dit-il.

Le sortilège de Réduction jaillit de la baguette et traversa le nuage de brume en le laissant intact. Harry se rendit compte de son erreur : le sortilège de Réduction n'avait d'effet que sur les objets solides. » (Rowling, 2000, *Harry Potter et la coupe de feu*, p. 553)

Réduire une notion signifie la reformuler en termes plus fondamentaux et l'étudier avec les outils de la science qui emploie ces termes. Cela n'est possible que si la notion est complètement explicable par ces outils élémentaires. Dans le cas contraire, tout ou partie de la notion reste obscur : nous n'avons pas fait d'avancée explicative satisfaisante. Le risque peut même être d'avoir l'impression erronée d'avoir apporté une explication là où la notion garde en réalité tout son mystère. Pour savoir quand une réduction est possible, Chalmers propose de s'appuyer le concept de survenance logique (« *logical supervenience* ») :

> « Les propriétés B surviennent logiquement sur les propriétés A, si aucune situation logique possible n'est identique sur le plan de ses propriétés A tout en étant distincte sur le plan de ses propriétés B. »[2] (Chalmers, 1996, p. 35, notre traduction)

Selon Chalmers, une propriété est réductible lorsqu'elle survient logiquement sur les propriétés du niveau auquel on souhaite la réduire (*ibid.*, p. 47)[3]. Par exemple, les propriétés de la chimie surviennent sur celles de la physique dans la mesure où il n'y a pas de monde logiquement possible identique sur le plan de ses propriétés physiques, mais qui serait différent sur le plan de ses propriétés chimiques.

Un exemple classique de réduction est celle du vivant à la physique. Il a été envisagé qu'une propriété particulière, l'*élan vital* ou la *force vitale*, distinguait les êtres vivants des objets inertes. Définir la vie reste un exercice périlleux, mais certaines propriétés semblent moins contingentes (p. ex. requérir une source d'énergie) que d'autres (p. ex. être basé sur la chimie du carbone). L'une de ces propriétés semble être une candidate intéressante pour le statut de propriété tant nécessaire que suffisante : la reproduction (David et Samadi, 2011, p. 12). Il est probable que, rencontrant d'étranges entités sur une nouvelle planète, ce soit le critère de la

[2] « B-properties supervene *logically* on A-properties if no two *logically possible* situations are identical with respect to their A-properties but distinct with respect to their B-properties. » (Chalmers, 1996, p. 35)

[3] Chalmers distingue survenance logique et survenance naturelle. Cette dernière peut être contingente à un monde particulier, dans lequel l'expérience montrerait qu'il n'y a pas de situation identique sur le plan des propriétés A, mais distincte sur le plan des propriétés B, alors qu'il serait conceptuellement envisageable que ce soit le cas. (*ibid.*, p. 35)

reproduction qui nous permette de déterminer si nous avons affaire à des êtres vivants. Certains trouveront ce critère trop inclusif, autrement dit insuffisant. Il est en effet applicable aux virus informatiques, aux prions, etc. Néanmoins, il est envisageable de simplement accepter ces implications (*to bite the bullet*). Toutefois, nous nous focaliserons sur le cas des êtres vivants conventionnels tels qu'on les connait sur Terre[4]. Le progrès des sciences biologiques a permis de donner une explication exhaustive de leur reproduction en termes biochimiques, donc en termes physiques. Là encore, il ne s'agit pas de remettre en cause la légitimé de la biologie, dans la mesure où expliquer un phénomène tel qu'une division cellulaire avec le seul usage des entités et des lois de la physique reste techniquement inenvisageable et intellectuellement moins satisfaisant, car laissant de côté d'autres dimensions explicatives (notamment téléologiques lorsque cette division cellulaire s'inscrit dans l'embryogenèse). Cependant, tous les mécanismes de la reproduction sont *in fine* de nature physique si bien qu'un monde en tout point similaire au nôtre sur le plan physique le serait également sur le plan de la reproduction des êtres vivants. Un vitaliste qui soutiendrait malgré tout la nécessité d'une propriété telle que l'élan vital pour caractériser la vie rétorquerait que la réduction de cette dernière à la physique est impossible dans la mesure où les propriétés de l'entité théorique postulée, l'élan vital, ne sont par définition pas réductibles à des propriétés physiques. Si on accepte ce postulat, le vitaliste aura parfaitement raison d'affirmer que la réduction ne peut avoir lieu. En effet, la propriété « élan vital » ne surviendrait ainsi pas logiquement sur les propriétés physiques. On pourrait alors imaginer une situation identique sur le plan des propriétés physiques, mais distincte sur le plan des propriétés vitales : dépourvu d'élan vital. La prémisse de l'existence d'un élan vital nécessaire à la vie est néanmoins devenue difficile à accepter. En effet, elle reviendrait à admettre que des êtres en tout point similaires à la faune et à la flore terrestre, puisse exister sur une planète physiquement jumelle, qui pourtant n'accueillerait aucune créature vivante.

Un problème tout à fait analogue se pose concernant la conscience. Si l'ensemble des propriétés de cette dernière survient logiquement sur les propriétés physiques, une explication matérialiste réductionniste est suffisante. Dans le cas contraire, d'autres explications sont requises.

[4] La classification actuelle reconnait sept règnes : les archées, les bactéries, les protistes, les chromistes, les mycètes, les végétaux et les animaux.

Vers une réduction de la conscience à la physique ?

Avant de se demander si une réduction des propriétés de la conscience aux propriétés physiques est possible, il convient de définir clairement la conscience et ses propriétés. Pour ce faire, nous allons nous appuyer sur la proposition de Chalmers, qui distingue les propriétés psychologiques et les propriétés phénoménales de la conscience.

Définir la conscience : la dichotomie de Chalmers

Qu'est-ce que la *conscience* ? Pour un réanimateur, il s'agit d'un niveau de vigilance, mesuré par l'échelle de Glasgow, qui s'oppose au coma. Un zoologiste peut parler de conscience de soi pour désigner un niveau d'élaboration propre à certaines espèces. La conscience s'entend aussi dans un sens éthique : avoir conscience ou non de ses actes. Parfois, il s'agit de la capacité à avoir une pensée réflexive (ou métacognitive). Dans tous ces exemples, ce sont des propriétés fonctionnelles qui définissent la conscience : la disposition, actualisée ou non, à réaliser telle ou telle action : réagir à son environnement, à ses propres pensées ou s'envisager en tant qu'individu. David Chalmers propose de parler de *conscience psychologique* pour désigner toutes les propriétés fonctionnelles de la conscience : percevoir, penser, croire, désirer, imaginer, se souvenir, prévoir, s'interroger, choisir, agir et bien d'autres, selon toutes les combinaisons possibles. Il ne semble pas y avoir d'obstacle conceptuel à l'implémentation de ces propriétés dans un système non biologique (p. ex. un ordinateur sophistiqué), dans la mesure où ce qui les définit se limite à la fonction qu'elles produisent. Ces propriétés surviennent sur les propriétés physiques si bien que si la conscience n'était que la somme, si longue soit-elle, de ces propriétés, une explication matérialiste réductionniste de la conscience serait satisfaisante. Cependant, ces propriétés semblent laisser quelque chose de côté. Notre perception, par exemple, ne parait pas se limiter au fait qu'un objet se situe dans notre champ perceptif, que nous agissons de façon adaptée à sa présence, que nous affirmons le percevoir. D'aucuns diront qu'il y a quelque chose de plus : *ce que cela fait de percevoir cet objet*. Or, ce que cela fait ne peut pas être retranscrit en propriété fonctionnelle : il n'y a pas de différence dispositionnelle ou comportementale spécifique à vivre *ce que cela fait*. Chalmers propose donc une deuxième famille de propriétés de la conscience : celles qui relèvent de la *conscience phénoménale*. Une propriété est phénoménale s'il y a *quelque chose que cela fait d'avoir cette propriété* (Nagel, 1974). La conscience phénoménale désigne le caractère

intrinsèque de l'expérience. Nous allons parler d'expérience consciente ou de qualia pour désigner les propriétés de cette famille. Uriah Kriegel défend l'idée qu'outre les expériences sensorielles, il existe des expériences conscientes non sensorielles, comme les cognitions (il peut y avoir quelque chose que cela fait de croire que *p*) et les conations (de désirer que *p*). De plus, les phénoménologies peuvent se combiner pour former l'ensemble du répertoire des expériences phénoménales que nous semblons être amenés à vivre. Ainsi, la phénoménalité des émotions serait par exemple une « combinaison de phénoménologies proprioceptives, algohédoniques, cognitives et conatives » (« *a combination of proprioceptive, algedonic, cognitive, and conative phenomenology* », Kriegel, 2015, p. 158).

Le progrès des sciences de l'esprit

Afin d'expliquer exhaustivement la conscience, il convient donc d'expliquer ses propriétés psychologiques et ses propriétés phénoménales. L'explication des premières est, selon Chalmers, à notre portée. Ce sont les *problèmes faciles de la conscience*. Il s'agit par exemple de l'explication de la mémoire, de l'apprentissage, du cycle veille-sommeil, de théorie de l'esprit (l'attribution à autrui d'états mentaux), du langage, etc. Si nous n'avons pas encore le détail du fonctionnement de ces mécanismes, le progrès des sciences de l'esprit nous donne une idée de ce à quoi nous attendre. Broca et Wernicke ont découvert les régions du cerveau impliquées dans le langage. Les sciences cognitives contiennent de solides théories de l'apprentissage. Le rôle de la mélatonine dans le rythme circadien est maintenant bien connu. La potentialisation synaptique à long terme pourrait être le corrélat neural de la mémoire. Nous semblons ainsi nous rapprocher toujours davantage d'une vision transversale des fonctions de l'esprit, de l'échelle biomoléculaire à l'échelle comportementale en passant par la neuroanatomie, au moyen de sciences réductibles à la physique.

À l'inverse, toujours selon Chalmers, aucun progrès majeur n'a été réalisé concernant le *problème difficile de la conscience*, c'est-à-dire l'explication des expériences conscientes. Il est possible de contester ce point. Nous pouvons supposer que la compréhension de l'aspect phénoménal de l'esprit finira par émerger de la compréhension de ses propriétés fonctionnelles. Tout serait question de temps et nous serions en bonne voie. Toutefois, avec Chalmers, nous réfutons cette hypothèse. L'exploration des fonctions de l'esprit laissera toujours de côté la dimension expérientielle. Chalmers fait remarquer que les arguments pour une explication de

cette dernière par les premières sont issus d'une substitution implicite de la conscience phénoménale par une variété de conscience psychologique (Chalmers, *op.cit.*, p. 111). Il est en effet difficile ne serait-ce que d'imaginer une description cognitive, neurophysiologique voire évolutionniste de l'esprit en mesure d'expliquer pourquoi nous avons des expériences plutôt que pas et pourquoi nous avons ce type d'expériences et pas d'autres. L'explication doit se trouver ailleurs et pour l'instant les progrès laissent à désirer.

Pourquoi n'avance-t-on pas vers une résolution du problème difficile ? Selon Chalmers, cela est dû à l'absence d'outils conceptuels appropriés. Comme il semble s'agir de propriétés ne survenant pas sur les propriétés physiques, il parait nécessaire de mettre au point un nouveau corpus théorique, *ad hoc*, ce que Chalmers essaye d'initier dans *The Conscious mind*. Deux autres explications de cette absence d'avancée peuvent être envisagées. Premièrement, il est possible que le problème difficile soit trop difficile. Il serait cognitivement clos pour l'être humain, soit du fait d'un défaut quantitatif de nos capacités cognitives, soit à cause d'un défaut qualitatif de ces dernières : notre cerveau ne serait pas « câblé » pour résoudre ce genre de problème, l'évolution ne nous ayant dotés que d'outils cognitifs appropriés à nos besoins d'antan et non à la résolution de problèmes métaphysiques. Cette explication est à double tranchant, puisqu'elle est aussi invoquée à l'appui de la thèse matérialiste réductionniste : nous ne serions simplement pas en mesure de comprendre comment les propriétés phénoménales surviennent sur les propriétés physiques. Deuxièmement, il est possible que nous ne progressions pas, car il n'y a en fait rien vers quoi progresser. C'est la position matérialiste éliminativiste selon laquelle les expériences conscientes n'existent pas et qu'il n'y a donc tout simplement rien à expliquer. Nous soutenons que montrer l'absence de la nécessité d'une explication au sens restreint (proximale ou distale) peut constituer une forme d'explication *satisfaisante* et *sérieuse* de l'expérience consciente ; au même titre que l'éliminativisme concernant l'élan vital a pu expliquer de façon acceptable et non détournée les propriétés de la vie.

Pour déterminer si le matérialisme éliminativiste peut jouer un rôle analogue dans l'analyse de l'expérience conscience, il est temps de nous intéresser de plus près aux supposées propriétés essentielles de ces dernières.

La nature supposée des expériences conscientes

Expériences et qualia

Qu'est-ce qui est supposé différencier un état mental accompagné d'une expérience d'un état mental qui en est dépourvu ? Les qualia désignent ce quelque chose de plus supposé accompagner le premier et absent du second. Elles seraient des propriétés des sensations ou perceptions dont on ne peut rendre compte avec des informations purement physiques (Jackson, 1982). Les qualia auraient pour propriétés essentielles d'être (Dennett, 1988) :

Ineffables. Il parait impossible d'expliquer à une personne née sourde ce que cela fait d'écouter de la musique. Elle serait en mesure de comprendre de quoi il s'agit sur un plan physique : une succession temporellement ordonnée d'ondes de pression dans l'air, accompagnées d'harmoniques, venant stimuler l'oreille interne où a lieu la transduction (codage électrochimique). L'information, transmise au cerveau via le nerf VIII, est intégrée à d'autres données pour générer des souvenirs et des dispositions comportementales telles que l'envie de danser. Toutefois, il semble que même la description la plus détaillée de l'ensemble des processus qui accompagne la musique et sa perception, laisse de côté *ce que cela fait* d'écouter de la musique.

Intrinsèques. Selon Dennett, le caractère ineffable des qualia vient notamment du fait que ce sont des propriétés supposées intrinsèques. Alors que nous définissons les entités du monde physique par leurs propriétés relationnelles (leurs interactions potentielles ou effectives avec d'autres objets du monde), les qualia seraient non-relationnelles, concerneraient la nature interne des objets qui en possèdent (p. ex. les esprits) et non leurs interactions fonctionnelles.

Privées. Du caractère ineffable des qualia découle également qu'il est impossible de comparer ce que cela nous fait d'écouter de la musique par rapport à ce que cela fait à quelqu'un d'autre. Ainsi, toute comparaison interpersonnelle des qualia serait, selon ce critère, impossible.

Directement appréhendables. Les qualia seraient immédiatement accessibles à la conscience, si bien qu'elles ne pourraient faire l'objet d'une erreur. Nous pouvons nous tromper lorsque nous affirmons entendre le son d'une trompette au loin, mais il semble qu'on ne puisse pas faire erreur sur le fait que nous faisons *l'expérience de ce* son. Ainsi, l'accessibilité immédiate des qualia est supposée nous assurer de la *certitude* de leur existence et de leur

nature. Nous insistons sur le fait que cette certitude se limite à l'expérience en elle-même et n'inclut pas la perception qui la produit ou l'interprétation qu'en fait le sujet.[5]

C'est en particulier ce caractère certain des qualia, dans le sens restreint de leur appréhension directe, qui va nous intéresser. C'est en effet sur cette certitude que certains philosophes se basent pour affirmer l'existence de telles expériences. Ainsi, pour Chalmers, c'est cette « intuition qui est la raison d'être même du problème de la conscience. »[6] (Chalmers, *op.cit*., p. 110, notre traduction)

L'échec des explications réductionnistes

Nous évoquions préalablement le fait que les progrès des neurosciences et des sciences cognitives (toutes deux en définitive réductibles à la physique) ne semblaient pas nous permettre d'avancer dans la compréhension des expériences conscientes. Nous allons ici reformuler cette hypothèse sous la forme suivante : *aucune science réductible à la physique ne peut nous apprendre quelque chose sur la nature des qualia*. Cela est implicite dans la définition des qualia proposée par Dennett, étant donné leur caractère intrinsèque. En effet, comme le fait remarquer Chalmers : « les théories physiques ne caractérisent que leurs entités basiques en termes de relations, c'est-à-dire leurs liens causaux et autres relations avec d'autres entités »[7] (Chalmers, *ibid.*, p.153, notre traduction). S'appuyant sur les travaux de Bertrand Russell, il précise que ces théories ne nous disent rien de la nature intrinsèque des objets physiques. Nous allons ici décrire trois arguments classiques qui viennent appuyer l'hypothèse selon laquelle les qualia ne sont pas réductibles aux propriétés de la physique.

La possibilité d'un monde zombie (Chalmers, *ibid.* p. 94). Selon le principe de survenance, si toutes les propriétés du monde sont réductibles à la physique, alors un monde

[5] Notre lointain air de trompette est peut-être en fait issu d'un saxophone, mais la distance, les bruits environnants et notre niveau attentionnel ont pu altérer notre perception. On peut aussi supposer qu'il s'agit d'une hallucination. La perception est donc incertaine. Par ailleurs, quelqu'un ne disposant pas dans son lexique du mot « trompette » ne pourra pas interpréter correctement l'expérience. De plus, le mot « trompette » est bien insuffisant pour décrire l'entièreté de l'expérience. Une description exhaustive de l'expérience impliquerait une quantité faramineuse d'informations exactes dont la formulation en langue naturelle n'est qu'un bref résumé (Greg Egan (2009) imagine le *LAMA* une langue qui aurait cette puissance descriptive et permettrait donc à ses locuteurs de partager leurs expériences de façon exhaustive). Ainsi, l'interprétation est également faillible. Toutefois, l'expérience, étant directement appréhendée, est supposée certaine dans son existence et sa nature.

[6] « The « intuition » at work here is the very *raison d'être* of the problem of consciousness » (Chalmers, 1996, p.110)

[7] « physical theory only characterizes its basic entities *relationally*, in terms of their causal and other relations to other entities. » (Chalmers, 1996, p.153)

qui partagerait toutes les propriétés physiques d'un monde contenant des qualia ne pourrait pas différer selon d'autres propriétés. Or, un monde zombie, tel que physiquement indiscernable de ce monde qualitatif, mais dont les états mentaux des habitants sont dépourvus de qualia est concevable. Ainsi, les qualia ne sont pas des propriétés réductibles aux propriétés physiques.

Le gouffre explicatif (Levine, 1984). Un problème classique soulevé par Saul Kripke (1982) est celui de l'identification de la douleur avec « l'activation des fibres nerveuses »[8]. L'hypothèse selon laquelle la douleur serait identique à l'activation des fibres, au même titre que la chaleur est identique à l'agitation des molécules, pose un problème dans la mesure où l'instanciation de la douleur semble possible dans un monde dépourvu de ces fibres. L'approche fonctionnaliste, qui tend à désigner la douleur en termes fonctionnels semble contourner le problème puisqu'elle avance que la douleur est réalisable de nombreuses manières dès lors que les propriétés fonctionnelles associées à la douleur sont présentes. La douleur serait alors identifiée à état fonctionnel donné propice à produire les effets attendus de la douleur. Néanmoins, comme le souligne Joseph Levine, quelque chose semble laissé de côté : pourquoi cet état fonctionnel ferait ce que cela fait et pas autre chose ou même rien ? C'est l'écart entre une définition fonctionnelle de la douleur et tel qu'elle est expérimentée qui constitue le gouffre explicatif supposé par Levine.

Les qualia inversées (Shoemaker, 1982). Le caractère privé des qualia laisse ouvert la possibilité que la manière dont deux individus, appelons les Paul et Chani, font l'expérience des couleurs ou des sons puisse être différente. Par exemple, si les spectres sonores de Paul et Chani sont qualitativement inversés, l'expérience de Paul lorsqu'il entend un son aigu peut être similaire à celle de Chani lorsqu'elle entend un son grave, et vice versa. Néanmoins, Paul et Chani s'accorderaient sur toutes les propriétés relationnelles de sons, par exemple qu'un *fa* est plus grave qu'un *sol*. La possibilité qu'ils aient toutefois une expérience différente des sons ne peut être expliquée en termes physiques et suppose donc l'existence des qualia.

De nombreuses variantes de ces arguments foisonnent dans la littérature et bien d'autres ont été avancés. Nous allons ici nous contenter d'accorder un certain crédit aux trois arguments

[8] On spécifie généralement : l'activation de fibres C, mais l'association fibres C-douleur est aujourd'hui désuète. D'une part, tout le contingent des fibres C n'est pas dévolu à la transmission de la douleur alors que d'autres contingents de fibres y participent (les fibres Aδ). D'autre part, la douleur est considérée comme un processus bien plus complexe que la seule activation de fibres nerveuses périphériques, comme en témoignent les douleurs psychogènes, les douleurs des membres fantômes, etc. La prise en compte du contexte, des expériences antérieures et des représentations du sujet est incontournable dans la compréhension du processus douloureux. Kripke reconnait cela, précisant que nous sommes libres de remplacer « fibre nerveuse » par un terme qui nous semble plus approprié. Ceci n'a donc pas de répercussion sur le fond de l'argument de Kripke, dans la mesure où ces éléments sont eux aussi de nature fonctionnelle.

susmentionnés et, avec Chalmers, conclure qu'ils remettent sérieusement en question la possibilité d'une réduction des qualia aux propriétés du monde physique.

L'étonnante corrélation

Cependant, si les qualia sont irréductibles aux propriétés physiques, une étrange coïncidence semble apparaître : les mêmes qualia semblent toujours accompagner les mêmes états physiques. Ainsi, lorsqu'il y a quelque chose que cela nous fait d'entendre un air de trompette au loin, il y a également des cognitions du type « croire qu'il y a une trompette qui joue au loin », ou a minima « croire entendre une trompette jouer au loin ». Lorsqu'il y a quelque chose que cela nous fait d'avoir mal, il y a des comportements de retrait, la formation d'un souvenir douloureux, etc. Alors qu'on peut imaginer un état mental dépourvu de qualia (le désir inconscient freudien en serait un exemple paradigmatique), il semble difficile d'imaginer un état mental constitué exclusivement de qualia. Ned Block (1995) suggère néanmoins que ces états existent, par exemple lorsqu'il semble que nous prenions conscience d'entendre un air de trompette que nous entendions déjà sans y prêter attention. Il semble également difficile d'imaginer la substitution d'une quale par une autre. Lorsque toutes les conditions physiques d'apparition de la douleur sont réunies, il semble que ce soit toujours une quale de douleur qui accompagne notre expérience, jamais une qualia de plaisir, et encore moins de couleur ou de son.[9]

Le dilemme de Chalmers

Selon le point de vue avec lequel on observe le problème, il semble y avoir deux conclusions totalement contradictoires. Du point de vue subjectif, l'expérience consciente est au « centre de notre univers épistémique » (« the center of our epistemic universe », Chalmers, *op.cit.,* p. 196). Elle parait indubitable et toutes les croyances et connaissances que nous établissons sur le monde semblent se faire à travers notre expérience. Hume soutient une

[9] Il est envisageable de discuter ce point en s'appuyant notamment sur le cas de synesthésies, mais nous pouvons faire abstraction de ces cas marginaux.

conception de prime abord similaire : toutes nos idées (états mentaux) seraient soit des perceptions, soit des souvenirs de perceptions, soit en serait la combinaison (Hume, 1995). Inversement, d'un point de vue objectif, une description exhaustive du monde semble possible en laissant les qualia à la marge. Aucune théorie physique ne fait appel à ces dernières, et il semblerait qu'un démon-zombie[10] ait l'impression d'avoir dit tout ce qu'il y avait à dire sur l'univers une fois acquises toutes les connaissances sur les entités physiques, les lois physiques et les conditions de base qui s'y trouvent. Notre démon-zombie serait probablement très sceptique sur l'existence d'autres phénomènes tant les prédictions qu'il ferait sur l'évolution du monde seraient justes en l'absence de recours à une théorie incluant ces phénomènes. Il apparait donc que nous nous trouvons face à un dilemme.[11]

Nous tenons l'existence des qualia pour vraie. C'est la position de Chalmers. Elle est notamment motivée par le caractère indubitable d'un point de vue subjectif[12]. Dans ce cas, il nous revient la charge de construire une théorie non réductionniste à même de rendre compte de l'étonnante corrélation (le pur hasard, si on ne peut formellement l'exclure, paraissant trop improbable pour en constituer une explication satisfaisante). Le coût de cette théorie serait d'admettre l'existence de propriétés non physiques nécessaires à une description exhaustive du monde. Ce qu'on y gagne, c'est la possibilité d'expliquer les qualia, mais dans une forme restreinte, épiphénoménale. En effet, l'ensemble des phénomènes observables[13] du monde physique est déjà déterminé par les propriétés et lois physiques du monde : c'est la fermeture causale du monde physique. Dès lors, il n'y a pas *besoin* de faire appel au concept de qualia pour rendre compte de l'enchaînement des causes et effets de nos comportements (qui après

[10] Nous proposons de définir le démon-zombie comme un démon de Laplace (omniscient sur le plan physique), mais aveugle à tout ce qui ne survient pas logiquement sur la physique.

[11] Frankish (2016) présente ce problème sous la forme d'un trilemme entre :

- *Réalisme radical.* « qui traite les phénomènes conscients comme réels et inexplicables sans une innovation théorique radicale » (« which treats phenomenal consciousness as real and inexplicable without radical theorical innovation », *ibid.*)
- *Réalisme conservateur*. « qui accepte la réalité des phénomènes conscients, mais cherche à les expliquer en termes physiques » (« which accepts the reality of phenomenal consciousness but seeks to explain it in physical terms », *ibid.*)
- *Illusionnisme*. « qui nie que le phénomène est réel et se focalise sur l'explication de son apparence » (« deny that the phenomenon is real and focus on explaining the appearance of it », *ibid.*)

Avec David Chalmers, nous rejetons la seconde option du trilemme de Frankish, car nous rejetons l'idée que les qualia (dont nous acceptons la possibilité conceptuelle) sont réductibles aux propriétés physiques. Les deux branches du dilemme de Chalmers peuvent tout à fait être reformulées en réalisme radical et illusionnisme.

[12] Ce caractère indubitable concerne les qualia de façon synchronique (l'éventuel souvenir de nos qualia peut bien sûr être l'objet d'une altération mnésique ; ce souvenir peut également être envisagé comme ayant sa propre qualité différente de l'expérience qui l'a produit) et tel qu'elles sont supposées apparaître à l'esprit (nous pouvons nous tromper lorsque nous tentons de les décrire, par erreur ou insuffisance conceptuelle, cf. note 5).

[13] La chute des corps, la désintégration des noyaux atomiques, la division des cellules, le son d'une trompette et le comportement des hommes qui sont parvenus à décrire ces phénomènes.

tout sont similaires dans un monde qualitatif et dans un monde zombie). Or, la psychologie naïve intègre couramment ces dernières dans le processus causal : nous sommes tentés de dire que c'est l'expérience du son qui nous fait juger qu'une trompette joue au loin, que c'est l'expérience de la douleur qui nous fait retirer la main de la plaque chauffante, que c'est l'expérience de la peur qui conduit au développement d'un trouble de stress post-traumatique. Si le monde physique est effectivement causalement clos[14] et que nous attribuons aux qualia une efficacité, il s'agit donc au mieux d'un cas de surdétermination : les événements physiques qu'on met en lien avec les qualia auraient simultanément deux causes suffisantes : les qualia et leurs corrélats fonctionnels. Cependant, ces deux causes n'auraient pas de valeur explicative équivalente : alors que déposséder les processus mentaux fonctionnels de leur efficacité causale rentrerait en contradiction avec la complétude du domaine physique, soustraire cette efficacité aux qualia ne pose pas de problème théorique particulier (en dehors d'une contradiction avec la psychologie naïve). Le principe de parcimonie nous prescrit de limiter autant que possible nos hypothèses si bien que nous sommes invités à faire l'économie de celle d'une efficacité causale des qualia.[15] Ainsi, l'hypothèse des qualia n'est peut-être pas si attractive, dans la mesure où leur retirer leur efficacité parait les priver d'une part significative de leur attrait théorique.[16]

Nous nions l'existence des qualia. C'est la position matérialiste éliminativiste. Elle semble difficile à envisager tant l'existence d'expérience consciente nous semble certaine et primitive. Cependant, cette posture a pour elle l'avantage de la parcimonie : elle limite l'appareil théorique aux propriétés et lois décrites par la physique. Elle répond avec élégance aux principaux arguments anti-réductivistes : les zombies sont concevables, à vrai dire nous en sommes nous-mêmes. Il n'y a pas de gouffre explicatif, car de l'autre côté du supposé gouffre, il n'y a rien. Il n'y a pas de problème des qualia inversées, puisqu'il n'y a pas de qualia. Il n'y a pas non plus d'étonnante corrélation à expliquer. Qu'est-ce que cela fait d'être une chauve-souris ? Absolument rien. En revanche, le coût de cette posture peut paraitre exorbitant : elle

[14] Aucun cas d'efficacité causale sur le monde physique d'un phénomène non physique n'ayant jamais été observé, y compris à l'endroit de l'univers le plus réputé pour sa densité de qualia : la Terre.

[15] Le refus du principe de parcimonie nous met face au risque de l'inflation théorique : nous pouvons supposer que la réaction au stimulus douloureux est (sur)déterminée par le processus fonctionnel implémenté dans le système nerveux, la quale de la douleur *et* l'action d'un ange gardien. Il n'est pas plus légitime de faire l'économie de la troisième cause que de la deuxième.

[16] Des propositions ont été faites pour réconcilier fermeture causale du monde physique et efficacité des qualia, par exemple la théorie panpsychiste ou le monisme neutre, mais ces considérations hypothétiques débordent du cadre de cette introduction. Nous y reviendrons néanmoins succinctement en début de partie 2 (page 39).

implique d'éteindre la lumière, de considérer que la richesse de nos expériences phénoménales n'est en fait qu'illusoire.

Si nous voulons prendre le problème difficile de la conscience au sérieux, comme le souhaite David Chalmers, il apparait qu'aucune explication basée sur une réduction des propriétés phénoménales de la conscience à des propriétés physiques ne pourra être satisfaisante. Il s'agit donc de trouver une autre explication. Chalmers poursuit cet objectif en proposant des pistes vouées à établir une théorie de la conscience compatible avec les dernières théories physiques et rendant compte de ce que nous avons appelé l'étonnante corrélation : le fait que les qualia, ces propriétés des expériences conscientes, semblent systématiquement associées à des propriétés précises du monde physique. En effet, si cette corrélation est contingente (il est possible d'imaginer un monde, par exemple le monde zombie où elle n'a pas lieu), elle est pour Chalmers *une propriété naturelle de notre monde*. Il n'y a, suppose Chalmers, jamais de douleur en l'absence d'un système fonctionnel analogue à un système nerveux ressentant de la douleur (c'est-à-dire disposé, dans des circonstances similaires à produire des sorties comportementales similaires à des stimuli douloureux). Ainsi, il propose les « principes de cohérence » (« principles of coherence », Chalmers, *op.cit.*, p. 218) : une loi de notre univers qui associe qualia et propriétés physiques fonctionnelles. Afin d'expliquer pourquoi ce sont toujours les mêmes qualia qui s'associent aux mêmes propriétés physiques, il ajoute un second principe : celui « d'invariance organisationnelle » (« the principle of organizational invariance », *ibid.*, p. 247). L'alternative éliminativiste, selon laquelle les qualia n'existent pas, est attrayante. Elle est parcimonieuse, ne nécessitant pas autant d'entités théoriques que celle de Chalmers. Elle semble plus objective, ne dépendant pas d'un observateur, mais compréhensible par tous, humains comme démons-zombies. Cependant, Chalmers rejette formellement cette alternative pour deux raisons. D'une part, elle lui semble entrer « en conflit avec les faits manifestes » (« in conflict with the manifest facts », *ibid.*, p. 164). Étant directement appréhendables, elles ne seraient pas sujettes à l'examen sceptique. D'autre part, il affirme n'avoir « jamais rencontré un argument vaguement en mesure d'établir que ce fait [personne n'est conscient au sens phénoménal] »[17] (*ibid.*, p. 164, notre traduction). On peut émettre à l'encontre de cette seconde affirmation une objection basée sur la charge de la preuve, qui nous semble incomber à celui qui affirme l'existence d'un phénomène.[18]

[17] « never seen an argument that comes remotely close to making this case » (Chalmers, 1996, p. 164)

[18] Aux yeux d'aucun, l'existence des qualia est à tel point évidente qu'il reviendrait à l'éliminativiste de prouver leur inexistence. Nous considérons que ce serait lui demander l'impossible. En effet, le dualiste peut toujours pratiquer la politique de la terre brûlée : concéder à l'éliminativiste autant de contre-exemples qu'il en propose

Nous allons au cours de cet essai nous focaliser sur la première affirmation, portant sur la phénoménalité comme fait manifeste. Plus précisément, nous allons nous intéresser à la question de l'indubitabilité de l'expérience consciente. Que sont le doute et la certitude ? Nous appuyant sur la considération de cas pathologiques, prototypes du dérèglement d'adaptations évolutives, nous allons en proposer une approche naturaliste, suggérant qu'il s'agit ni plus ni moins que *d'émotions épistémiques*, c'est-à-dire de phénomènes affectifs complexes porteurs d'une valeur épistémique et portant sur l'état épistémique du sujet. Nous allons pour ce faire tenter une caractérisation des émotions épistémiques, dont nous allons défendre l'intérêt conceptuel malgré le flou qui entoure la notion d'émotion, fondée sur une analogie avec d'autres affects épistémiques, tels que les sentiments épistémiques. Nous irons jusqu'à proposer que ces conceptions du doute et de la certitude comme émotions épistémiques soient les seules à rendre compte de ces états mentaux de façon satisfaisante en les comparants à d'autres conceptions traditionnelles telles que le doute hyperbolique cartésien.

Cette analyse des émotions épistémiques va nous permettre d'étudier la valeur de celles-ci dans l'argumentation en faveur de l'existence des expériences conscientes. Anticipant ainsi une réponse affirmative à la question titre de cet essai, « l'expérience consciente est-elle indubitable ? », nous allons envisager les implications d'une telle certitude sur une théorie du monde se voulant exhaustive. En particulier, est-il plus parcimonieux d'élaborer une telle théorie incluant ces expériences, dont nous sommes certains de l'existence, ou d'expliquer cette certitude au sein d'une théorie purement matérialiste ? Si nous nous orientons vers la deuxième solution, il nous reviendra de proposer l'ébauche d'une telle explication fonctionnelle de la certitude. La caractérisation naturaliste des émotions épistémiques pourra permettre une telle esquisse.

Bien évidemment, d'aucuns pourront affirmer douter de l'expérience consciente, répondant ainsi par la négative à la question titre. Le cas échéant, la tentation éliminativiste n'en est que plus grande. Néanmoins, nous nous proposons dans cet essai, comme nous en enjoint David Chalmers, de prendre au sérieux une expérience consciente certaine et

tout en affirmant qu'il reste d'autres situations où des propriétés non physiques interviennent. Il peut aller jusqu'à trouver refuge dans un dualisme minimaliste p. ex. un dualisme solipsiste, affirmant être l'unique être phénoménalement conscient dans un monde de zombies (cf. page 63). Nous estimons pour cette raison que faire reposer la charge de la preuve sur l'éliminativiste revient à nier qu'une discussion *sérieuse* sur le problème difficile est possible. Par la suite, nous nous intéresserons à l'approche contextualiste de l'épistémologie proposée par David Lewis (1996) et qui nous semble permettre d'éviter le spectre d'un scepticisme absolu auquel pourrait nous exposer le refus d'admettre une existence fondamentale telle que celle supposée de l'expérience consciente (cf. page 70).

irréductible aux phénomènes physiques, nous laissant face à ce que nous avons appelé le dilemme de Chalmers : une théorie non réductionniste ou un éliminativisme en mesure de rendre compte de cette certitude. Nous allons défendre que, contrairement au physicalisme réductionniste, l'éliminativisme puisse fournir une explication sérieuse de l'expérience consciente.

PARTIE 1 : LE BÉNÉFICE DU DOUTE

VERS UNE CARACTÉRISATION DES ÉMOTIONS ÉPISTÉMIQUES

Les pathologies du doute : vers une vision naturaliste du doute et de la certitude

Les obsessions de Ernst

Lorsqu'il consulte pour la première fois, Ernst Lanzer explique souffrir depuis plusieurs années d'intenses angoisses, dont les objets lui semblent insensés. Il peut s'inquiéter d'être un criminel et, n'obtenant pas réassurance dans l'exploration de ses souvenirs, rechercher le soulagement auprès d'un confident apte à lui confirmer son innocence. Il vit avec la peur de commettre des actes impulsifs au contenu mortifère. Il redoute que certaines de ses pensées puissent causer de terribles souffrances aux personnes qu'il aime, notamment à son père, pourtant décédé depuis plusieurs années. Pour conjurer ses angoisses obsédantes, Ernst se sent contraint d'accomplir des rituels répétitifs, en apparence dénués de lien avec l'objet de ses tourments. Ses troubles sont apparus dans l'enfance, initialement sous la forme d'une idée glaçante : on pouvait lire dans ses pensées, deviner ses plus sombres secrets. Bientôt, cette idée s'accompagna d'une lutte intérieure pour empêcher ses pensées les plus honteuses de se former ; lutte perdue d'avance tant il lui était impossible de contrôler les travers de sa psyché. Impuissant face aux contenus de son esprit, étant persuadé de leurs répercussions sur le monde, Ernst se mit à élaborer de fastidieuses liturgies conjuratoires. Elles pouvaient, au moins un temps, soulager ses effrois.

Lorsqu'il publie que le cas de Ernst, dit « l'homme aux rats », Sigmund Freud évoque un « cas de névrose obsessionnelle » (Freud, 2017, p. 383). Ce trouble serait la résultante d'un conflit psychique inconscient, s'exprimant par des compulsions, mentales ou motrices (pensées redondantes, rituels) et par un mode de pensée caractérisé par un doute envahissant et des ruminations anxieuses. Il aboutirait à l'inhibition de la pensée et de l'action (Laplanche et Pontalis, 2007, p. 284). On retrouve ces deux dimensions dans l'histoire clinique présentée par Freud. Il décrit chez son patient des comportements imposés et idiosyncrasiques, à l'origine d'une intense souffrance psychique. Il remarque que ces attitudes sont souvent précédées d'une vaine lutte pour empêcher leur apparition. Il dépeint des compulsions de comptage, de prières, mais aussi des séquences comportementales plus élaborées. Ainsi, un jour, Ernst « se sentit

absolument obligé » (Freud, *op.cit*., p. 439) d'ôter une pierre du bas-côté de la route : l'idée lui était venue qu'elle causerait un accident à son aimée. Une fois le rocher retiré, il lui vint l'idée contraire et la nécessité de remettre le caillou exactement à sa place. Les compulsions de Ernst ont un sens aussi obscur qu'ambivalent. Ses comportements sont souvent contradictoires et traduisent une intense tension interne. Par ailleurs, Freud décrit son patient comme scrupuleux et enclin au doute excessif. On retrouve là une analogie avec le concept de psychasthénie décrit par Pierre Janet (cité par Laplanche et Pontalis, *op. cit.*, p. 285). Freud en a repris l'idée d'une « paralysie de la décision » (Freud, *op. cit.*, p. 493), qu'il met en lien avec l'envahissement psychique du patient par le doute. Freud définit le doute comme une « perception interne de l'indécision » (*ibid.*). Il témoignerait d'un « état d'inhibition insupportable » (*ibid.*, p. 496). Dans la pensée de Freud, la compulsion est donc secondaire, elle est une « tentative de compenser le doute » (*ibid.*). Le doute, quant à lui, s'apparente à un sentiment de valence affective particulièrement négative.

D'après la description de Freud, le tableau clinique présenté par Ernst s'apparente à ce que nous qualifierions aujourd'hui de trouble obsessionnel compulsif (TOC). Les critères du trouble obsessionnel compulsif dans la cinquième édition du *Manuel diagnostique et statistique des troubles mentaux* (*DSM-5*, American Psychiatric Association, 2015) sont la présence d'obsessions ou de compulsions à l'origine d'un retentissement important sur la vie du patient (en termes de temps perdu, de difficultés sociales, de détresse psychologique…)[19]. Il est classiquement considéré que le patient doit avoir conscience du caractère déraisonnable de ses pensées ou actions.[20] Plusieurs formes cliniques de TOC sont décrites, dont le type « vérificateur », enclin au doute excessif et prompt à vérifier maintes et maintes fois la même chose (Cottraux, 2016). Or, nous vivons tous des périodes de doute, nous ressentons tous l'incertitude face à un choix ambigu, nous réfléchissons à deux fois avant de nous engager et changeons parfois d'opinion après avoir pesé le pour et le contre. Nous ressentons de l'insatisfaction face à une décision dont nous ne sommes pas sûrs du bien fondé. Qu'est-ce qui différencie ce doute de celui du vérificateur obsessionnel ? Nous émettons l'hypothèse que le doute est un affect, porteur d'une valeur heuristique[21] : celle d'informer l'individu sur un état épistémique insatisfaisant, l'invitant à rechercher davantage d'informations. On pourrait ainsi

[19] Sans que les symptômes ne soient mieux expliqués par les effets d'une substance psychoactive ou par une autre pathologie (notamment un autre trouble mental).
[20] Ce critère est nuancé à partir du DSM-5. Il y est proposé de spécifier le niveau d'insight, inversement corrélé au degré d'adhésion du patient à ses idées obsessionnelles ou au bienfondé de ses comportements compulsifs.
[21] Une heuristique est « une procédure simple qui permet de trouver des réponses adéquates, bien que souvent imparfaites, à des questions difficiles. » (Kahneman, 2016, p. 153)

caractériser le doute par un seuil, en dessous duquel nous détectons que les informations à notre disposition sont trop peu significatives pour une prise de décision adaptée, déclenchant ce sentiment d'inconfort. Dans le TOC, ce seuil serait trop élevé, si bien que l'individu ressentirait le doute alors qu'il disposerait d'un corpus de données normalement compatible avec une prise de décision adaptée.

Freud souligne le rôle de la mémoire dans l'émergence de la pensée obsessionnelle. En effet, Ernst semble accorder une bien maigre confiance à ses souvenirs, si bien que ces derniers ne constituent pas un témoignage suffisant à l'établissement d'une croyance sur certains aspects du monde : ceux qui sont perçus par Ernst comme source de danger, à risque de catastrophe et dont il se tient pour responsable. Ainsi, les souvenirs de ses propres actions, p. ex. les rituels mentaux, n'attestent pas de façon fiable de leur bonne réalisation. Ernst doit donc déployer et redéployer ses routines jusqu'à réassurance. On peut dès lors se demander si c'est sa mémoire qui est altérée ou la confiance qu'il lui accorde. Or, comme nous l'avons vu, chez les patients présentant des TOC, l'insight est fréquemment préservé. Ils sont en mesure de juger leurs comportements de vérifications absurdes, ils savent que le gaz a été fermé, que la liturgie a été prononcée, que leurs mains ont été lavées. Mais le doute persiste, obsédant, si bien que devenant intolérable, il force le sujet à recontrôler, répéter, reproduire. Ce vécu imposé des rituels témoigne de la préservation de l'encodage mnésique des comportements déjà effectués. Les informations à disposition devraient être suffisantes à éloigner le doute. Ce n'est donc pas la mémoire qui fait défaut, mais sa probité.

Le monde de William

Nous avons présenté le TOC comme exemple paradigmatique du doute pathologique. Le sentiment d'incertitude y est décrit comme excessif. Le seuil épistémique de son apparition semble déréglé, revu démesurément à la hausse. Nous pouvons nous interroger sur l'existence de phénomènes inverses, c'est-à-dire de pathologies qui seraient caractérisées par un défaut quantitatif du doute. Que se passerait-il si un individu était incapable de ressentir le doute ? Qu'adviendrait-il si son seuil d'exigence était si bas que la moindre idée, même peu étayée, était acceptée sans remise en cause ?

Oliver Sacks nous présente un cas qui semble illustrer cette situation : celui de William Thomson, un patient placé en institut neurologique pour un syndrome de Korsakov (Sacks,

1988). Parmi ses symptômes, on retrouve la tétrade classique du syndrome de Korsakov : une amnésie antérograde, de fausses reconnaissances, des confabulations et une anosognosie (Martini, Dehmas et Paille, 2017). Sacks décrit les confabulations de William comme une suite d'idées s'enchaînant sans l'ombre d'un doute apparent. Le monde de William se construit dans une pseudo-cohérence émergeant d'explications *ad hoc* aux événements perçus. Lorsqu'il croit reconnaître un ouvrier en la personne du docteur Sacks avec qui il discute, il lui adresse : « vous autres les mécaniciens, vous commencez toujours par jouer aux médecins, avec vos vestes blanches et vos stéthoscopes » (Sacks, *op.cit.*, p. 145). Il élabore des scénarios si farfelus que Sacks décrit son état comme un « délire affabulatoire frénétique (que l'on appelle parfois la « psychose de Korsakov », bien qu'il ne s'agisse pas du tout d'une psychose) » (Sacks, *ibid.*, p. 147) tant la production du discours est pseudo-délirante.

William Hirstein définit la confabulation comme le « rapport confiant de souvenir d'événements qui soit n'ont jamais eu lieu, soit ont eu lieu bien plus tôt dans la vie »[22] (Hirstein, 2006, p. 1, notre traduction). Il souligne deux différences importantes entre le mensonge et la confabulation. Dans le premier cas, il y a intention de tromper. De manière formalisée, où A est le trompeur, B le trompé et *p* un énoncé quelconque :[23]

1. A dit à B que *p*
2. *p* est faux
3. A sait que *p* est faux
4. A à l'intention de générer la croyance que *p* chez B en disant que *p*

Dans le cas de la confabulation, le critère 3 du mensonge n'est pas rempli, le confabulateur ayant, selon la définition de Hirstein, confiance en la véracité de son assertion. Par ailleurs, dans le cadre du mensonge, il y a l'intention de tromper. Le critère 4 peut être respecté dans le cadre de la confabulation. Il y a en effet un acte de langage dont l'intention semble être de générer la croyance que *p* chez l'interlocuteur. Mais dans le cas de la confabulation, il n'y a pas l'intention de générer une croyance erronée. L'analyse de Sacks est que la neurotoxicité chronique de l'alcool a détérioré de façon irréversible les capacités mnésiques de William, si bien qu'afin de préserver un semblant de cohérence dans la narration

[22] « [confident] report as memories events that either did not happen […] or that happened to him, but much earlier in life » (Hirstein, 2006, p. 1)
[23] Reformulation d'une citation de Hirstein : « I lie to you when (and only when) 1. I claim *p* to you 2. *p* is false 3. I believe that *p* is false 4. I intend to cause you to believe *p* is true by claiming that *p* is true. » (*ibid.*, p. 16)

de sa biographie, il comble les manques avec « une prolifération de faux récits, dans une fausse continuité, de faux mondes peuplés de fausses personnes, habités de fantômes » (Sacks, *op.cit.*, p. 148). Daniel Dennett postule que le soi est un « centre de gravité narratif » (« center of narrative gravity », Dennett, 1992). Dans cette optique, la décomposition des souvenirs de William l'empêcherait de maintenir un sentiment d'unité psychique s'il ne comblait pas ses lacunes par des idées extérieurement perçues comme extravagantes. Alors que Freud soulignait les doutes exagérés de Ernst, notamment concernant l'authenticité de ses souvenirs, le patient de Sacks nous est décrit comme dépourvu de la capacité à douter. William ne doute de rien, ni de ses souvenirs, ni de son état de santé. Ses confabulations sont énoncées avec l'indifférente certitude de celui qui est inaccessible à toute remise en cause.

Les syndromes miroirs

Hirstein propose de considérer la confabulation et le trouble obsessionnel comme des syndromes miroirs (Hirstein, *op. cit.*, p. 97). Un exemple paradigmatique de syndromes miroirs est la symétrie existante entre le délire d'illusion des sosies de Capgras et la prosopagnosie. Il s'agit respectivement d'un syndrome dans lequel le patient à la conviction que ses proches ont été remplacés par des sosies leur ressemblant parfaitement et d'un trouble lié l'incapacité à reconnaître explicitement des visages connus. La reconnaissance des visages implique en effet deux circuits cérébraux distincts : l'un conscient, celui de la reconnaissance explicite, l'autre inconscient et affectif, associé au sentiment de familiarité.[24] Dans le cas du délire des sosies, la reconnaissance explicite semble préservée, mais associée à une perte du sentiment de familiarité : reconnaissant les visages de ses proches, mais ne ressentant pas les affects normalement associés à leur présence, le patient en tire la conclusion qu'il s'agit d'imposteur. De façon symétrique, si les patients prosopagnosiques dénient reconnaître le visage de leur proche, ils sont en mesure d'éprouver un sentiment de familiarité face à un visage connu. À la présentation de photographies de proches et d'inconnus, ils font mieux que le hasard pour deviner qui est connu, sans parvenir à expliquer de quelle manière. Ces résultats sont corroborés par la mesure de paramètres physiologiques corrélés à la réponse émotionnelle, tels que la conductance cutanée et la fréquence cardiaque. L'hypothèse de Hirstein est que nous sommes face à une situation analogue dans le cadre de la confabulation et de l'obsession : dans le

[24] La distinction conscient-inconscient est ici strictement fonctionnelle. L'événement conscient, contrairement à l'événement inconscient, peut faire l'objet d'un rapport verbal.

premier cas, la capacité à douter est diminuée, voire abolie, dans le second elle est exacerbée. Il propose un continuum entre confabulations et obsessions, entre lesquelles on retrouve différents seuils d'apparition du doute. Cette idée d'une continuité entre une norme médiane et deux pôles pathologiques définis par l'excès ou le défaut quantitatif d'un état mental nous amène à émettre l'hypothèse, par analogie avec l'anxiété ou la thymie, que ce dernier est de nature affective. Nous allons donc explorer plus en avant cette nature du doute (et de la certitude).

Doute, certitude et émotions épistémiques

Qu'est-ce qu'une émotion ?

La riche littérature concernant la nature et les propriétés des émotions s'avère souvent discordante, si bien qu'aucune conception ferme et définitive de ces dernières n'a encore émergé. Nous allons ici tenter d'en extraire certaines caractéristiques utiles dans notre étude du doute et de la certitude.

Les sensations corporelles. Nos émotions les plus familières sont largement accompagnées de sensations physiques relativement spécifiques. Nous tremblons de peur, transpirons et sentons notre pouls s'accélérer. Notre mâchoire se crispe de colère tandis que notre respiration semble s'emballer. Le dégoût provoque des haut-le-cœur, mais la joie réchauffe ce dernier. William James postule que ces sensations corporelles, loin d'être les conséquences de nos émotions en constituent en fait la nature. « Ma théorie […] est que les changements corporels suivent directement la perception du fait excitateur et que notre ressenti de ces changements es*t* les émotions »[25] (James, 1950, p. 449, notre traduction). Pour défendre cette hypothèse, James argue que si nous soustrayions de notre champ de conscience toutes les sensations corporelles il ne resterait rien de l'émotion sinon « un froid et neutre état de perception intellectuelle » (« a cold and neutral state of intellectual perception », *ibid.*, p. 451). Cette proposition est partiellement infirmée par la constatation qu'un individu dont les afférences nerveuses provenant des viscères auraient été sectionnées ne devient pas subitement

[25] « My theory, on the contrary, is that *the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur* IS *the emotion.* » (James, 1950, p. 449)

dépourvu d'émotions, ni même particulièrement alexithymique. Des défenseurs contemporains de la conception jamesienne des émotions suggèrent que l'intéroception est suffisante sans être nécessaire à l'existence des émotions. Ainsi, Antonio Damasio propose un mécanisme alternatif où « le corps est court-circuité et le cortex préfrontal et l'amygdale ne font que pousser le cortex somatosensoriel à reproduire les types d'activités neurales qu'il aurait eus, si le corps avait été placé dans un état déterminé et s'il avait envoyé les signaux correspondants. » (Damasio, 2010, p. 253). Une dernière objection à la nature strictement corporelle des émotions réside dans le fait que des variations physiologiques similaires semblent pouvoir engendrer des émotions différentes. En effet, les conséquences physiologiques de l'activation de l'axe du stress, décrites par Walter Cannon (1915), consistant en l'activation l'accélération du rythme cardiaque et respiratoire, la redistribution du flux sanguin vers les muscules et le cerveau, la mobilisation des réserves énergétiques, etc. afin d'engendrer la fameuse réponse « combat-fuite » soit sensiblement similaire en cas de peur et en cas de colère. Néanmoins, si la *théorie des marqueurs somatiques* de Damasio semble insuffisante à définir complètement nos émotions, il apparait que la présence de sensations corporelles accompagnant un état mental est un bon indice de sa nature émotionnelle. Selon Peter Goldie (2002), ces sensations corporelles sont en fait en lien avec une autre propriété des émotions : l'intentionnalité. Cette dernière serait double, d'une part dirigée vers les sensations corporelles, d'autre part dirigées vers un objet du monde.

L'intentionnalité. Outre l'intentionnalité corporelle évoquée au paragraphe précédent, Goldie (*ibid.*) souligne le fait que les émotions sont dirigées vers des objets du monde. Cette caractéristique permet notamment de différencier les émotions et les humeurs. Ces dernières sont réputées moins intenses et de durée plus longue que les émotions. L'intentionnalité constituerait une autre différence majeure puisque les humeurs ne sont pas dirigées vers un objet particulier, mais teintent globalement nos états mentaux et nos perceptions. Les psychiatres décrivent par exemple des distorsions cognitives entre une personne euthymique et un patient dépressif, ce dernier adoptant une attitude cognitive défaitiste (Cottraux, 2017, p. 90)[26]. Les émotions, quant à elles, porteraient systématiquement sur quelque chose. Nous avons peur *de*, nous sommes en colère *contre*, nous regrettons *que*.[27] Toutefois, parmi le vaste corpus de nos états mentaux, les émotions sont loin d'avoir le monopole de l'intentionnalité. Pensons

[26] L'hypothèse du réalisme dépressif stipule qu'au contraire ce sont les personnes euthymiques dont les cognitions sont biaisées (cf. note 65). Quelle que soit la perspective, une différence de traitement de l'information entre ces deux populations semble faire consensus.

[27] Cette propriété reste discutable dans la mesure où certaines émotions, telles que la joie, semblent pouvoir exister sans porter sur un objet précis. Il est néanmoins envisageable qu'une joie sans objet soit à ranger aux côtés des humeurs tandis qu'une joie émotionnelle porte sur une situation.

tout simplement aux croyances ou aux désirs qui sont aussi nécessairement de nature intentionnelle. Néanmoins, cette fois encore, la présence de l'intentionnalité est un argument (non suffisant) en faveur de l'appartenance à la famille des émotions lorsque nous enquêtons sur la nature d'un état mental.

La direction d'ajustement esprit-monde. En se limitant aux deux critères précédents, il est difficile de différencier émotion et désir. Comme nous l'évoquions, ils portent tous deux sur un objet. Il est en outre envisageable d'attribuer une sensation corporelle propre au désir. Cependant, les désirs ont toujours une direction d'ajustement monde-esprit : ils vont avoir pour fin, via l'action, de conformer le monde à l'esprit (Deonna et Teroni, 2016, p. 38). Les émotions auraient, quant à elles, une direction d'ajustement esprit-monde : elles auraient vocation à représenter le monde, à modifier l'esprit en fonction de l'état du monde afin de lui permettre (ou non) d'adopter une réaction appropriée. On pourrait comparer la dyade désir/émotion à un thermostat : tandis que l'émotion aurait le rôle du thermomètre, censé représenter de manière appropriée un aspect de l'état du monde (la température de la pièce pour le thermomètre, la présence d'un danger pour la peur), le désir serait comparable au climatiseur, censé modifier la situation (rafraichir la pièce, se mettre à l'abri). Cette opposition, qui rapprocherait davantage l'émotion d'un état cognitif, reste discutable dans la mesure où certaines théories des émotions (telles que la théorie mixte, *ibid.*) incluent l'aspect conatif au sein même de l'émotion. L'émotion serait le thermostat dans son ensemble et non uniquement le thermomètre. Cette proposition pose cependant également problème puisqu'elle suppose une nature combinatoire des émotions qui comporteraient alors à la fois une direction d'ajustement monde-esprit et esprit-monde. Quoi qu'il en soit, la présence d'une direction d'ajustement esprit-monde est un critère qui, sans être suffisant, semble au moins nécessaire à la caractérisation d'une émotion. Par ailleurs, ce critère accorde une valeur épistémique aux émotions, dans la mesure où elles servent justement à « prendre la température » de différents états du monde afin de générer des réponses appropriées.

La valence. Dans la théorie du jugement axiologique (*ibid.*, p. 46), il est possible de rendre compte de l'aspect motivationnel de l'émotion sans y inclure un désir : l'émotion serait de nature essentiellement cognitive, mais associée à une valeur positive ou négative. Par exemple, la peur serait la croyance en la présence d'un danger (valeur négative). La présence de cette dimension axiologique pourrait rendre compte de l'impact des émotions sur

l'apprentissage, notamment selon le modèle skinnérien du conditionnement opérant.[28] Limiter la caractérisation d'une émotion à un strict effet de renforcement/aversion est insuffisant, puisque cela empêche d'en expliquer la variété et de distinguer les émotions des perceptions algohédoniques. Toutefois, la valence représente une propriété supplémentaire à ajouter à notre inventaire des attributs des émotions.

L'histoire évolutive. Si les quatre critères précédents seront globalement suffisants caractériser les émotions, nous allons tout de même nous attarder sur une question primordiale dans une étude à leur sujet : pourquoi existent-elles ? Dans ses travaux sur la reconnaissance des émotions d'autrui par l'observation de leurs expressions faciales, le psychologue Paul Ekman a élaboré une liste d'émotions de bases retrouvées universellement dans l'espèce humaine.[29] Le fait que ces dernières soient accompagnées de signes apparents et facilement reconnaissables par les pairs est un argument en faveur du rôle social des émotions (Ekman, 1999). La communication non verbale de notre tristesse permettrait ainsi d'obtenir un soutien salvateur de notre proche entourage. Le caractère transculturel de ces signes émotionnels, pouvant même s'étendre à d'autres espèces de mammifères, amène à envisager, par argument phylogénétique, que les émotions ont une histoire évolutive relativement longue et présente un avantage significatif en termes de valeur sélective. Tooby et Cosmides proposent une considération des émotions basée sur une approche modulaire des fonctions cognitivo-comportementales : « […] un large corpus de découvertes empiriques en psychologie, biologie et neurosciences soutient que l'architecture mentale humaine est composée de programmes évolués fonctionnellement spécialisés »[30] (Tooby et Cosmides, 2008, notre traduction). Ils ajoutent que « [l'] existence de ces divers programmes créé un problème adaptatif : des programmes individuellement conçus pour résoudre des problèmes adaptatifs spécifiques pourraient, si activés simultanément, produire des outputs entrant en conflit les uns avec les autres »[31] (*ibid.*, notre traduction). Dans ce cadre, ils suggèrent que les émotions ont un rôle de

[28] Un comportement ayant pour conséquence une émotion positive verra sa fréquence d'apparition augmenter (renforcement positif), un agissement ayant pour résultat une émotion négative verra sa fréquence diminuer (aversion positive), un comportement ayant pour effet d'éviter l'apparition d'une émotion négative sera favorisé (renforcement négatif), et dans de plus rares cas, une action susceptible d'empêcher une émotion positive sera entravée (aversion négative).

[29] Cette liste contenait initialement la tristesse, la joie, la colère, la peur, le dégoût et la surprise, qui sont (hormis la surprise) personnifiés dans le film *Inside out* (Docter, 2015). Elle a été modifiée à plusieurs reprises et sa pertinence reste débattue.

[30] « the human mental architecture is crowded with evolved, functionally specialized programs. Each is tailored to solve a different adaptive problem » (Tooby et Cosmides, 2008).

[31] « the existence of all these diverse programs itself creates an adaptive problem: Programs that are individually designed to solve specific adaptive problems could, if simultaneously activated, deliver outputs that conflict with one another » (*ibid.*)

coordination entre ces différents modules, permettant d'en activer certains tout en en inhibant d'autres de manière cohérente. Par exemple, l'émotion *peur* pourrait schématiquement activer le module *recherche visuelle d'une menace* tout en désactivant le module *recherche de nourriture*. L'apport de la théorie darwienne de l'évolution est justement de répondre à l'argument du réglage fin des fonctions du vivant d'une façon alternative à l'explication intentionnelle auparavant prépondérante, symbolisée par l'argument de l'horloge de William Paley, supposée démontrer l'origine divine de la vie. Aussi, inclure la notion d'histoire évolutive dans la caractérisation des émotions nous invite à nous intéresser à son paramétrage.

Ainsi, nous avons retenu quatre critères en faveur la nature émotionnelle d'un état mental : la cooccurrence de sensations corporelles, l'intentionnalité, une direction d'ajustement monde-esprit et la présence d'une axiologie. Si aucun de ces éléments ne semble être suffisant à la définition d'une émotion, leur présence constitue un faisceau d'arguments robuste en vue d'attribuer à un état mental ce statut. L'imprécision de ces critères reflète probablement le flou qui entoure l'essence des émotions. Cependant, l'approche évolutionniste avancée par Tooby et Cosmides nous invite à prendre au sérieux le concept d'émotion et à aller, pour les états mentaux qui nous intéressent, chercher quels paramètres de déclenchement la nature a définis et quelle marge de plasticité elle leur a accordée.

Sentiments et émotions épistémiques

Santiago Arango-Muñoz (2014) propose l'existence d'états mentaux évaluatifs dirigés vers notre état épistémique. Le postulat de l'existence de tels sentiments épistémiques (« epistemic feelings ») est en effet utile pour rendre compte de phénomènes d'allure paradoxale, tels que le sentiment de connaissance (« feeling of knowing »). Nous parvenons généralement à savoir si nous sommes en mesure de répondre à une question (de culture générale par exemple) avant même d'avoir la réponse à l'esprit. Des jeux de rapidité sont directement basés sur cette aptitude puisque c'est la promptitude à manifester détenir la réponse qui départage les joueurs. Une fois qu'un joueur s'est signalé, il dispose de quelques secondes pour formuler une réponse qu'il n'avait pas nécessairement d'emblée à l'esprit (en mémoire de travail). Le fait qu'une telle prédiction sur notre état épistémique soit significativement plus efficace qu'une prévision hasardeuse témoigne de l'existence de systèmes qu'on pourrait qualifier de métamnésiques qui se manifestent sous la forme de ces sentiments épistémiques.

Arango-Muñoz propose une taxonomie de ces affects selon deux dimensions : l'une axiologique, l'autre chronologique. Il identifie ainsi trois sentiments épistémiques positifs : le sentiment de connaissance déjà évoqué, le sentiment de fluence, qui évalue l'efficacité du processus de récupération et le sentiment de justesse qui estime l'exactitude de la réponse fournie. Leurs pendants négatifs sont respectivement les sentiments d'incertitude, de difficulté et d'erreur (*ibid.*). L'appartenance de ces sentiments à la catégorie des émotions prête à discussion. Ronald de Sousa propose une nuance minime entre ces deux notions, stipulant que les émotions sont des sentiments ne pouvant être attribués qu'à une personne (tandis que les sentiments pourraient exister à un niveau infrapersonnel) et soient par nature plus complexe (de Sousa, 2016), ce qui s'apparente à la perspective de Tooby et Cosmides. Nous proposons ainsi de parler d'*émotions épistémiques* lorsque nous nous référons spécifiquement à un état mental de nature émotionnelle dont l'intentionnalité porte sur un état épistémique.[32]

La valeur épistémique des émotions

Si nous nous intéressons de plus près à la peur, il semble que sa fonction, pour reprendre la perspective de Tooby et Cosmides, soit de permettre à l'individu apeuré de se placer dans un état physiologique et psychologique approprié à une réponse adaptée à son environnement (ce qui est concordant avec le modèle du stress de Cannon et la cascade physiologique qu'il avait mise en évidence). Il parait donc nécessaire que la peur soit associée à système de détection du danger. Le lien entre ce système de détection et la cascade d'événements déclenchée par l'émotion étant de nature systématique, nous prendrons le parti de considérer cette dimension de détection comme faisant partie intégrante de l'émotion, ce qui n'entre pas en contradiction avec les quatre critères que nous avons établis. Ainsi, la peur serait une sorte de système d'alarme s'activant en présence d'un danger et déclenchant le système combat-fuite.

Dans cette perspective, la peur peut être modélisée comme un test prédictif, dont les qualités intrinsèques sont la sensibilité (Se) et la spécificité (Sp). La sensibilité correspond à la proportion de vrais positifs (VP) parmi le nombre total de cas (présence d'un réel danger), c'est-à-dire la somme des vrais positifs et des faux négatifs (FN). $Se = VP/(VP + FN)$. Plus la peur

[32] Comme nous allons le voir par la suite, on peut accorder à l'ensemble des émotions une valeur épistémique. La spécificité des émotions épistémiques est qu'elles portent elles-mêmes sur un état épistémique du sujet (la mémoire comme l'envisageait Freud, cf. page 23) contrairement à des émotions non épistémiques telles que le dégoût, dont la valeur épistémique porte sur l'état d'un objet extérieur (dans le cas du dégoût, il peut par exemple s'agir d'un aliment dont il est nécessaire d'évaluer la potentielle toxicité).

est sensible, plus il probable qu'elle soit activée à juste titre et déclenche donc de façon appropriée le système combat-fuite. La spécificité correspond quant à elle à la proportion de vrais négatifs (VN) parmi le nombre total de non-cas (absence d'un réel danger), soit la somme des vrais négatifs et des faux positifs (FP). $Sp = VN/(VN + FP)$. Plus la peur est spécifique, moins il est probable que le système combat-fuite soit activé en l'absence de menace. L'importance de la spécificité est de freiner la suractivation d'un système coûteux en énergie dans des situations où les ressources sont limitées et seront mieux employées à d'autres fins (la prospection alimentaire, la digestion, la reproduction…). On imagine aisément que ces paramètres varient d'un individu à l'autre. Chez *le couard*, la spécificité est basse : il est peu probable qu'il reste serein face à un vrai danger (peu de FN), mais au prix d'une inquiétude souvent déclenchée pour rien (davantage de FP). Chez *l'intrépide*, c'est la sensibilité qui est basse : alors qu'il limite ses dépenses énergétiques en évitant l'activation intempestive de son système combat-fuite (peu de FP), il court aussi le risque d'ignorer un danger bel et bien présent (davantage de FN). *Le couard* et *l'intrépide* sont ici deux archétypes représentant deux stratégies d'adaptation différentes selon la balance bénéfice/coût associée au système de réponse au danger. Il est à noter que dans notre espèce, ils ne représentent pas des cas très éloignés l'un de l'autre. Nous avons globalement tous la même tendance à déclencher notre système de peur en présence d'un prédateur, marchant au bord d'un précipice ou menacés d'un licenciement. Les différences de sensibilité et de spécificité se font à la marge. On remarquera aussi que ces paramètres ne sont pas rigides, mais évoluent au cours de la vie, au gré des expériences. Il est possible d'apprendre à déclencher des réactions de peur dans certaines situations, notamment par conditionnement classique[33], mais également à les faire disparaître[34]. On retrouve ici les deux échelles temporelles de l'adaptation, celle du groupe, où génération après génération les sensibilités et spécificités de tests émotionnels ont été réglés par les mécanismes de l'évolution et celle de l'individu, où l'évolution a également sélectionné une certaine marge d'ajustement par apprentissage permettant de faire face à un environnement changeant. Comment la sensibilité et la spécificité de la peur ont-elles été paramétrées au gré de l'évolution ? Quand on évalue la sensibilité et la spécificité d'un test diagnostic, il est

[33] Dans l'expérience dite « du petit Albert », le psychologue John Watson a conditionné un enfant de neuf mois à avoir une phobie des souris blanches en associant la présentation de la souris (stimulus neutre) à un bruit sourd (stimulus inconditionnel générant une réponse inconditionnelle de peur). Cette association a transformé la présentation de la souris en stimulus conditionnel, générant l'angoisse indépendamment du bruit (réponse conditionnelle). (Cottraux, 2017, p. 23)

[34] La disparition peut se faire de façon automatique, via le phénomène d'extinction : après des présentations itératives du stimulus conditionnel en l'absence du stimulus inconditionnel, la réponse conditionnelle s'atténue et finit par disparaître. Elle peut également être favorisée par des techniques de thérapie comportementales telles que la désensibilisation systématique. (*ibid.*, p. 121)

possible de chercher un optimum à l'aide d'une courbe représentant la sensibilité en fonction du nombre de faux positifs (courbe ROC). Néanmoins, selon l'enjeu du test diagnostique, il est possible de favoriser la sensibilité au prix d'une moins bonne spécificité : par exemple s'il s'agit d'un dépistage, il est préférable de ne pas passer à côté d'un diagnostic (bonne sensibilité) quitte à surdiagnostiquer (mauvaise spécificité) une pathologie dont des explorations ultérieures permettront de confirmer ou non la présence. De même, une optimisation du « test danger » qu'est la peur a pu se construire au fil des générations par le fait qu'un meilleur réglage aura une meilleure valeur sélective.[35]

Est-il possible de caractériser l'ensemble du répertoire émotionnel en termes de sensibilité et de spécificité ? S'il parait plus difficile d'estimer la valeur épistémique d'émotions telles que la joie et la tristesse, nous n'excluons pas que ce soit possible. Une piste pourrait être de faire appel aux notions de synchronie et diachronie. Comme nous l'avons vu, la peur a une fonction principalement synchronique : adapter son comportement à une situation donnée. Elle a néanmoins également un rôle diachronique dans l'apprentissage de conduites d'évitement opportunes (par le biais de l'aversion positive, cf. note 28). Peut-être qu'à l'inverse, les émotions telles que la joie et la tristesse, outre leur rôle communicatif, jouent ce rôle épistémique diachronique.

Les émotions auxquelles nous nous sommes intéressées jusqu'à présent ont, abstraction faite de l'intentionnalité somatique décrite par Goldie, une intentionnalité portant sur le monde extérieur. Nous allons à présent nous intéresser à une autre gamme d'émotions, cette fois-ci dirigées vers nos états internes. À noter que ce changement d'orientation n'entre pas en contradiction avec le principe d'ajustement esprit-monde : il s'agit simplement de garder à l'esprit que nos états et processus internes font partie intégrante du monde. Nous allons plus spécifiquement nous intéresser à des émotions dédiées à l'évaluation de notre propre état épistémique : la certitude et le doute.

[35] À noter que nous ne parlons pas d'un optimum absolu, mais d'un optimum local, l'évolution ne permettant pas de franchir des vallées de valeurs sélectives en vue d'atteindre un « meilleur » optimum que l'actuel. Toutefois dans le cas d'une modélisation simplifiée d'une émotion telle que la peur (avec pour seule variable prise en compte le seuil de déclenchement), on peut considérer qu'il n'y a qu'un seul optimum.

Doutes et certitudes comme émotions épistémiques

Disposant à présent d'un modèle convenable des émotions, nous allons nous intéresser à une caractérisation du doute et de la certitude selon cette perspective. Nous avons défendu l'idée que le doute et la certitude répondaient au critère d'intentionnalité. Nous avons proposé que cette dernière porte en réalité sur nos états internes, bien qu'on puisse également considérer qu'elle porte indirectement sur un objet du monde. Dans cette conception, nous ne doutons pas que la lumière soit éteinte, nous doutons de notre souvenir concernant l'état de la lumière. Ainsi, la direction d'ajustement esprit-monde ne s'envisage qu'en considérant nos souvenirs et autres états internes comme objets du monde. Concernant les sensations corporelles, le doute semble être accompagné d'une tension physique apparentée à un état anxieux. Enfin, en suivant les typologies de sentiments épistémiques d'Arango-Muñoz, nous avons souligné la valence négative du doute.

Avec le cas des troubles obsessionnels, nous sommes tout d'abord intéressés au doute dans ses excès. Une dimension centrale que nous avons mise en évidence dans les TOC est l'intolérance à l'incertitude, concernant notamment les souvenirs, amenant à des comportements de vérification à type de persévérations et à des ritualisations mentales ayant vocation à diminuer une charge émotionnelle corrélée à ce sentiment d'indétermination. Des travaux de neurosciences viennent appuyer l'hypothèse émotionnelle du doute et l'importance des zones cérébrales responsables de la mémoire dans l'activation de ces réponses émotionnelles (Kepecs, 2013). À l'opposé, nous avons étudié l'abolition du doute chez les confabulateurs, au premier rang desquels les malades de Korsakov, où une abrasion émotionnelle semble avoir eu lieu sous l'effet de lésions neurologiques d'origines toxiques et carentielles. À l'inverse des premiers, ces patients se présentent comme impulsifs, irréfléchis ou spontanés. Mais ce qui saute aux oreilles est leur discours fantaisiste et la conviction avec laquelle ils le tiennent. Ils paraissent avoir perdu l'aptitude à en caractériser certaines idées comme souvenirs et d'autres comme actes d'imagination. D'un côté à l'autre de ce spectre, c'est le seuil d'activation de l'émotion-doute qui semble varier. Si dans les troubles anxieux, c'est le seuil de déclenchement de la peur qui se dérègle, nous pouvons supposer que dans les pathologies du doute c'est celui du doute. Une première possibilité, serait de rendre compte de ces troubles en nous appuyant sur le principe de Broussais, selon lequel « toutes les maladies [consistent] dans l'excès ou le défaut de l'excitation des divers tissus au-dessus et au-dessous du degré qui constitue l'état normal. » (Canguilhem, 2013, p. 24). Cependant, Canguilhem juge

lui-même ce principe non satisfaisant, puisqu'il omet une dimension essentielle de la pathologie : le trouble qu'elle engendre dans les capacités d'adaptation à l'environnement. Cette seconde considération reste appropriée à nos troubles du doute, puisque ce sont également les possibilités d'adaptation dont nous a dotés la nature qu'ils viennent entraver. En résumé, le doute peut être caractérisé comme une émotion épistémique, dont les seuils de déclenchement sont censés être adaptés à notre état épistémique : notre ressenti du doute venant nous inciter à adopter des comportements de recherche d'informations. Les troubles émotionnels concernant le doute illustrent une désadaptation synchronique (le doute présente plus de faux positifs ou de faux négatifs que prévu) et diachronique (le doute perd sa susceptibilité à s'ajuster). Concernant la certitude, il est possible d'en proposer une définition négative par rapport au doute (de Sousa, *op.cit.*) ou de lui attribuer une positivité à l'instar des sentiments épistémiques de valence positive (Arango-Muñoz, *op.cit.*). Dans les deux cas, il apparait une certaine symétrie axiologique entre les deux notions.

Nous proposons de qualifier le doute et la certitude d'*émotions épistémiques*, non pas du fait de leur valeur épistémique (valeur que nous estimons retrouver dans l'ensemble du répertoire émotionnel), mais du fait que leur intentionnalité est directement dirigée vers notre état épistémique, c'est-à-dire la valeur épistémique de nos états mentaux. Néanmoins, d'autres caractérisations du doute et de la certitude, en apparence dénuées de dimension affective, peuvent être envisagées. Nous allons à présent les considérer.

Le doute et la certitude ne sont-ils qu'affectifs ?

L'enjeu de ce travail étant de discuter de la valeur épistémique du caractère indubitable de nos expériences conscientes (tant en termes d'existence que de nature), il nous apparait nécessaire d'explorer une conception du doute en apparence alternative à celle que nous avons jusqu'à présent envisagée. Nous avons défendu une interprétation affective du doute, ressenti comme situation d'inconfort épistémique, ayant pour rôle fonctionnel de favoriser les comportements de recherche d'information et de vérification. Cette fonction, issue de notre histoire biologique et individuelle, nous permet de nous adapter de façon appropriée à notre environnement. Dans les cas pathologiques, cette heuristique affective se dérègle et l'individu s'en trouve perturbé à différents degrés (dont les extrêmes peuvent être illustrés par les cas de

Ernst et William). Néanmoins, même dans les cas non pathologiques, on observe une tendance de ces heuristiques à fonctionner de façon biaisée (Kahneman, 2016). L'horloger aveugle[36] qui a étalonné nos affects ne s'est pas basé sur leur fidélité, mais sur leur utilité. Or, il semble exister une conception alternative du doute : ce dernier, systématique et rationnel, serait fondamental dans l'établissement de nos connaissances.

Certitude et doute cartésiens

De prime abord, la méthode cartésienne du doute systématique semble effectivement s'opposer à nos jugements habituels, que nous avons décrits comme basés sur des heuristiques affectives : les émotions épistémiques. Alors que ces dernières seraient avant tout pratiques, sélectionnés par les mécanismes de l'évolution pour leur valeur sélective, le doute cartésien serait essentiellement rationnel, remettant systématiquement en cause l'ensemble de nos certitudes, à la recherche de celle ou celles qui seraient en mesure de résister à cet examen méthodique. Pour fonder une véritable épistémologie, dont le doute le plus hyperbolique ne pourrait ébranler l'édifice, Descartes a besoin de deux ingrédients. D'une part, une méthode déductive : « ces longues chaînes de raisons, toutes simples et faciles, dont les géomètres ont coutume de se servir » (Descartes, 2016, p. 50), permettant de construire une connaissance exhaustive « pourvu seulement qu'on s'abstienne d'en recevoir aucune pour vraie qui ne le soit, et qu'on garde toujours l'ordre qu'il faille pour les déduire les unes des autres » (*ibid.*). Outre cette méthodologie édificatrice, Descartes a besoin d'un socle immunisé contre toute remise en cause. Nos sens n'étant pas infaillibles, comme en témoignent les rêves, les illusions et les hallucinations, ils ne peuvent nous fournir cette certitude première. Il en va de même pour nos intuitions logiques, puisque l'enfant, prompt aux paralogismes, ou l'influence des substances psychotropes, susceptibles d'obscurcir nos raisonnements, nous fournissent des exemples de leur faillibilité. Dans sa première méditation, Descartes propose l'expérience de pensée du malin génie « non moins rusé et trompeur que puissant, qui a employé toute son industrie à [le] tromper » (Descartes, 2009, p. 89). Cette réflexion l'amène à identifier ce qui lui parait être une fondation inébranlable : sa propre existence en tant qu'elle est nécessaire à ce qu'il puisse être trompé sur tout le reste. Disposant ainsi d'une base et d'une méthode, Descartes reconstruit ensuite son épistémologie au fil de ses six méditations.

[36] Expression empruntée à Richard Dawkins (2003), désignant les mécanismes de l'évolution

La nature du doute présentée par Descartes semble bien différente de celle que nous explorions jusqu'ici. Le doute émotionnel ne parait avoir sa place que dans des situations spécifiques : « les actions de la vie ne souffrant souvent aucun délai, c'est une vérité très certaine que lorsqu'il n'est pas en notre pouvoir de discerner les plus vraies opinions, nous devons suivre les plus probables » (Descartes, 2016, p. 57). Nous invitant à ouvrir davantage le champ de notre défiance, la méthode proposée par Descartes semble imposer d'aller à contre-courant de nos affects. Ainsi, après avoir exposé sa méthode dubitative, il ajoute : « il ne suffit pas d'avoir fait ces remarques, il faut encore que je prenne soin de m'en souvenir ; car ces anciennes et ordinaires opinions me reviennent souvent en la pensée, le long et familier usage qu'elles ont eu avoir moi leur donnant droit d'occuper mon esprit contre mon gré, et de se rendre presque maîtresses de ma créance » (Descartes, 2009, p. 87). Nos inclinations naturelles nous pousseraient à dévier de la rigueur exigée par la méthode cartésienne. À quel point la certitude et le doute cartésiens diffèrent-ils de nos émotions épistémiques ? Pour fonder la certitude, Descartes s'appuie en définitive sur une conception de ces notions bel et bien apparentée à la nôtre. Lorsqu'il énonce les règles de sa méthode, il délimite les contours de la certitude de la façon suivante : « ne comprendre rien de plus en mes jugements, que ce qui se présenterait si clairement et si distinctement à mon esprit, que je n'eusse aucune occasion de la mettre en doute » (Descartes, 2016, p. 49). Clarté et distinction sont des propriétés qui s'attribuent aux informations dont nous disposons. Cette conception de la certitude comme évaluation de notre état épistémique ne diffère donc pas radicalement de nos estimations émotionnelles.

Doute 1/doute 2

William James propose de distinguer chez l'homme deux systèmes cognitifs distincts. Le premier, partagé avec les êtres vivants aux capacités plus limitées serait de nature associationniste : « Une grande partie de nos pensées consistent en trains d'images d'enchainant les unes les autres » (« Much of our thinking consists of trains of images suggested one by another », James, *op.cit.*, p. 325). Cet enchaînement d'idée, si on rapproche cette description associationniste de la pensée de celle de David Hume, remonte ultimement à des perceptions sensorielles ou à la trace mnésique laissée par ces dernières. Ainsi, pour Hume, « Croire, c'est en ce cas, éprouver une impression immédiate des sens ou la répétition de cette impression dans la mémoire. Ce sont purement et simplement la force et la vivacité de la perception qui constituent l'acte initial du jugement et qui posent le fondement du raisonnement » (Hume,

1995, p. 148). Notons qu'il n'est pas nécessaire d'inclure une dimension phénoménale dans les perceptions évoquées par Hume, dans la mesure où seul leur rôle fonctionnel entre ici en ligne de compte. La dimension affective de cette modalité d'établissement des connaissances est évoquée par Hume : « la croyance doit plaire à l'imagination, grâce à la force et à la vivacité qui l'accompagnent, puisque l'on constate que toute idée dotée de force et de vivacité est agréable à cette faculté. » (*ibid.*, p. 191). Un rapprochement peut ici être effectué entre ces notions de force et vivacité chez Hume et les idées de clarté et distinction chez Descartes.

James oppose à ce premier système, un second système, typiquement humain, de « raisonnement [qui] nous vient en aide dans les situations inédites – situations pour lesquelles notre sagesse associative commune […] nous laisse sans ressource. »[37] (James, *op.cit.*, p. 330, notre traduction). Ce second système peut être caractérisé par de nombreuses propriétés de la pensée méthodologique proposée par Descartes : il est contrôlé, volontaire, il requiert un effort, il est déductif. La théorie psychologique du double processus (respectivement système 1 et système 2) a notamment été affinée par les travaux de Kahneman et Tversky, dont nous évoquions déjà les découvertes concernant les biais cognitifs[38]. Ces derniers seraient principalement le fait de raccourcis du système 1 auquel on attribue les propriétés d'être rapide, involontaire, peu coûteux en énergie et intuitif.

Ainsi, nous pourrions parler de doute 1 pour désigner le doute-émotion et de doute 2 pour désigner le doute cartésien (et symétriquement d'une certitude 1 et d'une certitude 2). Cependant, comme nous l'avons vu, la frontière entre ces deux formes de doute ne semble pas si nette, puisqu'on retrouve en définitive une dimension affective dans la caractérisation de la certitude chez Descartes. De plus, il semblerait étonnant qu'un système cognitif aussi complexe que le système 2 soit apparu *de novo* et brutalement (à l'échelle évolutive, et par argument phylogénétique, puisqu'il semble être exclusif aux grands singes, voire à l'homme). Peter Carruthers (2012) propose à l'inverse que le système 2 émerge à partir de boucles d'opérations du système 1 (ce qui expliquerait sa lenteur relative et son coût cognitif). Schématiquement, le doute 2 pourrait être l'aboutissement de doutes 1 portant sur d'autres doutes 1 portant eux-mêmes sur d'autres doutes 1 et étant associés à chaque étape à des combinaisons d'impressions et d'idées. La complexité des raisonnements élaborés de l'homme et sa capacité à douter 2 pourraient en définitive provenir des mécanismes simples du système 1, tout comme les

[37] « Reasoning helps us out of unprecedented situations – situations for which all our common associative wisdom, all the education which we share in common with the beasts, leaves us without resource. » (James, 1950, p. 330)
[38] Travaux synthétisés par Daniel Kahneman après le décès d'Amos Tversky (Kahneman, 2016)

capacités de calcul les plus sophistiquées d'un ordinateur peuvent être réduites des successions et associations de calculs simples.[39]

Le doute et la certitude sont par essence émotionnels

Nous avons tout d'abord démontré qu'une caractérisation du doute et de la certitude comme émotions épistémiques était possible. En effet, ils partagent avec d'autres émotions telles que la peur, les propriétés d'intentionnalité, de direction d'ajustement esprit-monde et une dimension axiologique. L'approche évolutionniste des émotions a permis d'en caractériser des qualités épistémiques telles que la sensibilité, la spécificité et des seuils de déclenchement susceptibles de s'adapter aux variations de l'environnement à différentes échelles de temps. Cependant, pour pouvoir analyser la valeur justificative du doute et de la certitude en philosophie de l'esprit, centrale dans l'argumentation en faveur de la nécessité de postuler l'existence des expériences conscientes, il nous est apparu important de montrer que non seulement cette caractérisation émotionnelle était appropriée au doute et à la certitude, mais qu'elle était aussi la seule réellement pertinente. Les autres conceptualisations du doute et de la certitude n'étant, en définitive, que des distorsions de ces propriétés affectives, liées à la complexification de l'appareil cognitif humain et de l'émergence d'un système déductif à partir de systèmes associatifs ancestraux. C'est ainsi que, dans la suite de ce développement, lorsque nous parlerons de doute et de certitude, il s'agira toujours de processus affectifs.

Dès lors, si l'indubitabilité de l'expérience consciente constitue l'argument princeps en faveur de la nécessité d'élaborer une théorie du monde incluant des propriétés (les qualia) et des lois (cohérence, invariance organisationnelle) spécifiques et non réductibles à celles de la physique, nous pouvons raisonnablement comparer les bénéfices philosophiques d'une telle théorie comparée à une théorie plus minimaliste, reposant exclusivement sur les lois et propriétés de la physique, mais susceptible d'expliquer de manière satisfaisante cette certitude. La caractérisation du doute et de la certitude comme émotions épistémiques est donc essentielle

[39] Si nous sommes attachés au critère de de Sousa (2016) selon lequel une émotion ne peut être attribuée qu'à une personne tandis qu'un sentiment peut exister à un niveau infrapersonnel, le doute 1 serait à considérer comme un sentiment. Néanmoins ce critère ne nous est pas apparu essentiel dans notre développement.

à l’élaboration d’une telle explication, indispensable pour considérer le dilemme de Chalmers sans recourir à une esquive réductionniste.

PARTIE 2 : CERTITUDE ET TREMBLEMENT

LES QUALIA À L'ÉPREUVE DES ÉMOTIONS ÉPISTÉMIQUES

Selon David Chalmers, les tentatives d'explication physicaliste réductionniste des expériences conscientes sont vouées à l'échec, car les propriétés de la conscience phénoménale ne surviennent pas logiquement sur les propriétés physiques. Ce n'est, à l'inverse, pas le cas des propriétés de la conscience psychologique. Celles-ci sont de nature fonctionnelle donc explorables via les outils des sciences cognitives, des neurosciences, de la chimie, sciences elles-mêmes réductibles à la physique. « À peu près tout survient logiquement sur le physique » (« Almost everything is logically supervenient on the physical », Chalmers, *op.cit.* p. 71). Cet *à peu près* n'exclut en définitive que les qualia qui représenteraient dès lors une exception particulièrement notable face à tous les autres événements, propriétés et entités présents dans l'univers. Si on se borne à n'accorder la possession de qualia qu'aux objets disposant de propriétés fonctionnelles bien spécifiques (selon le principe de cohérence), cela revient à dire que l'homme et ses analogues fonctionnels sont un empire dans un empire, pour reprendre cette image de Spinoza. Pour résoudre ce problème, Chalmers envisage une solution proche du monisme spinoziste, soutenant que l'expérience, loin d'être confinée dans la boîte crânienne de bipèdes sans plumes, est en fait présente en toute chose. « Partout où il y a interaction causale, il y a information, et partout où il y a information, il y a expérience. » (« wherever there is a causal interaction, there is information, and wherever there is information, there is experience. », *ibid.*, p. 297).[40] En définitive, quelle qu'en soit la forme (monisme spinoziste, panpsychisme, dualisme) l'existence d'une phénoménalité de la conscience non réductible à la physique représente l'une des branches de ce que nous avons présenté comme le dilemme de Chalmers. Elle s'opposait à la position matérialiste éliminativiste, selon laquelle l'expérience consciente est une notion dont on peut en fait se passer dans une description du monde ayant pourtant vocation à être exhaustive. L'élimination de la phénoménalité présente les atouts indéniables de la parcimonie et de l'élégance. À l'inverse, la ramification du dilemme choisie par Chalmers est coûteuse sur un plan théorique : elle implique de postuler un large corpus de propriétés nouvelles (les qualia) accompagnées de lois fondamentales pour rendre compte des

[40] Il précise néanmoins : « le panpsychisme n'est pas le fondement métaphysique de mon idée : le fondement est plutôt un dualisme naturaliste avec des lois psychophysiques. » (« panpsychism is not at the metaphysical foundation of my view: what is rather at the foundation is naturalistic dualism with psychophysical laws », Chalmers, 1996, p. 299)

événements du monde. Cependant, si Chalmers est prêt à payer ce prix, c'est parce que l'existence de l'expérience consciente lui parait indubitable. Or, nous avons défendu l'idée que la certitude, tout comme le doute, étaient par essence des émotions épistémiques. Cela nous amène à discuter de la valeur justificative des émotions épistémiques. En effet, si Chalmers s'appuie sur sa certitude de l'expérience consciente pour établir la connaissance de l'existence de cette dernière et qu'on définit la connaissance comme une croyance vraie et justifiée, il nous incombe d'évaluer la valeur justificative de ce caractère certain.

Nous allons dans un premier temps revenir sur la balance coûts/bénéfices théoriques de la branche chalmersiennes du dilemme. En effet, montrer que ce rapport n'est pas si attractif renforcera l'attrait pour la branche éliminativiste. Par ailleurs, pour donner une légitimité à cette dernière, il faut montrer que l'argument fondamental de Chalmers est contestable. Or, attribuer une valeur justificative à la certitude de l'expérience consciente, revient à affirmer que l'existence de celle-ci est rendue très probable par son indubitabilité. Si nous parvenons à montrer que d'autres mécanismes peuvent être à l'origine de cette certitude, c'est-à-dire proposer des explications crédibles à l'existence de cette certitude en l'absence d'expérience consciente, nous aurons considérablement réduit la portée de l'argument. Bien évidemment, même en supposant l'existence de ces explications alternatives, cela ne suffit pas à prouver que l'explication de Chalmers est incorrecte. Cette dernière reste une explication valable, voire elle-même plutôt parcimonieuse abstraction faite du coût théorique des qualia. C'est donc de l'évaluation du coût théorique des qualia que dépend l'inclinaison pour l'un ou l'autre bras du dilemme de Chalmers. En d'autres termes, être en mesure de rendre compte de la certitude de l'expérience consciente tout en postulant l'absence d'expérience consciente (A) *et* montrer que l'hypothèse de l'existence de cette dernière représente un coût théorique prohibitif (B) pourra nous amener à considérer le matérialisme éliminativiste comme la réponse la plus légitime au problème difficile de la conscience (C), envisagé avec tout le sérieux que David Chalmers attend de nous, c'est-à-dire sans chercher à déguiser le problème difficile de la conscience en problèmes faciles par une pirouette réductionniste.

Les qualia comblent-elles un vide théorique ?

À l'aube du XX[e] siècle, l'achèvement des sciences fondamentales semblait approcher. Toutefois, « la beauté et la clarté de la théorie dynamique [restait] obscurcie par deux nuages » (Kelvin, 1901). Ces deux ombres dans le firmament des sciences[41] allaient donner naissance à deux branches majeures de la physique contemporaine : la relativité générale et la physique quantique. Le rêve d'une théorie complète de la physique synthétisant les lois et propriétés énoncées par ces deux champs de recherche ne s'est pas encore concrétisé, mais d'aucuns suspectent que le prochain grand pas en avant dans le domaine de la physique théorique sera la fondation d'une telle « théorie du tout » (Hawking, 2007). Nous pourrions toutefois souligner que d'autres nuages semblent toujours obscurcir notre compréhension du monde, le plus opaque étant probablement celui qui entoure son origine. Deux questions semblent en effet rester à ce jour hors de portée de toute explication scientifique : « pourquoi y a-t-il quelque chose ? pourquoi cela ? » (« Why anything? Why this? », Parfit, 2004). Selon David Chalmers, un autre mystère vient s'ajouter aux limites actuelles de nos connaissances sur le monde : celui de la conscience. Ainsi, selon lui, « une véritable théorie finale requiert un composant additionnel » (« a truly final theory needs an additional component », Chalmers, *op.cit.*, p. 126). Ce composant devrait être un corpus de propriétés non physiques, pouvant soit être fondamentalement de nature phénoménale, soit protophénoménales et sur lesquelles les propriétés phénoménales surviendraient logiquement. Afin de ne pas subir immédiatement le couperet du rasoir d'Occam, cet ensemble de propriétés doit prouver sa nécessité dans l'élaboration de ladite théorie finale. C'est de cette nécessité que nous allons dans un premier temps discuter : ces propriétés viennent-elles effectivement combler un vide théorique ? Chalmers défend cette nécessité avec l'argument que l'existence des qualia est une incontestable certitude. C'est plus précisément la validité de l'argument selon lequel la certitude de l'existence des qualia rend très probable leur existence que nous allons donc tenter d'évaluer.

[41] L'expérience de Michelson-Morley, incompatible avec l'éther, et le problème du rayonnement du corps noir.

Une certitude venue d'ailleurs

L'implication suivante est-elle acceptable dans le contexte d'un être humain amené à réfléchir à des questions philosophiques avec les outils épistémiques dont la nature l'a doté ? Le cas échéant, pour quelles valeurs de *p* ?[42]

(I) [*p* est certain] → *p*

Si l'on souhaite démontrer la fausseté de l'implication (I), il faut montrer que le conséquent peut être faux alors que l'antécédent est vrai. Formulé ainsi, le problème semble relever davantage d'une question de psychologique expérimentale que d'une question de philosophie. En effet, pour invalider l'implication, il suffit de montrer la possibilité d'une proposition *p* telle que la conjonction suivante, équivalente à (I), soit vraie pour *p* vrai :

(II) ([*p* est certain] ∧ ~*p*)

Le sens commun nous assure de la véracité de (II) pour certaines valeurs de *p*. La vie est émaillée de certitudes contrariées. Cependant, il est possible que pour certaines valeurs de *p* et dans certains contextes la conjonction (II) puisse être systématiquement fausse. Cela peut être le cas si on prend pour *p* une nécessité (ce qui n'implique pas que la réciproque soit vraie et qu'une nécessité suppose sa certitude). Bien entendu, ce qui nous intéresse dans le cadre de cet essai, est d'explorer (I) ou (II) pour la proposition *p : les qualia existent*. Ainsi, nous allons discuter de la proposition suivante : « *l'existence des qualia* est certaine implique *l'existence des qualia* » (I). Pour se faire, nous allons envisager plusieurs possibilités qui rendraient vraie *l'existence des qualia est certaine* en conjonction avec *les qualia n'existent pas* (II). Si au moins l'une d'entre elles est convaincante, alors l'argument de la certitude des qualia en faveur de leur existence est amoindri. Cela ne signifie pas qu'il perde toute valeur : l'existence des qualia reste *a priori* une excellente raison d'être certains de leur existence. Cela montre simplement que des explications du monde ne postulant pas leur existence ne sont pas d'emblée disqualifiées. Les propositions qui vont suivre ne sont probablement pas aussi convaincantes les unes que les autres (et il n'est pas exclu que d'autres propositions non envisagées ici le soient en fait

[42] L'emploi de la logique modale épistémique ne nous est pas apparu pertinent ici, car nous maintenons une distinction entre certitude et connaissance. Nous ne remettons évidemment pas en question le caractère tautologique de □*p* → *p*. Notre conception émotionnelle de la certitude lui accorde une marge d'erreur qu'on ne retrouve pas dans la notion de connaissance.

davantage). Cependant, l'argument requiert seulement qu'un certain crédit soit accordé à leur disjonction.

La certitude en l'existence des qualia est le produit d'un délire. Selon Karl Jaspers, le délire est une idée tenue avec « une extraordinaire conviction, avec une certitude subjective incomparable. » (« They are held with an extraordinary conviction, with an incomparable subjective certainty », Jaspers, cité par Walker, 1991). Le DSM-5 les définit comme des croyances figées, mais souligne qu'il est parfois difficile de le différencier d'une idée fermement tenue pour vraie. Ainsi, l'idée que la certitude de l'existence des qualia est délirante n'implique pas qu'elle soit l'apanage de personnes souffrant de pathologies mentales. La définition de Jaspers, mettant en avant cette certitude irrévocable, pourrait en définitive s'appliquer à de nombreuses idées jugées non délirantes dans un contexte culturel donné. Or, le DSM-5 précise également que l'écart entre l'idée et la norme culturelle est central dans la notion de délire. Baignant certes dans un bain culturel dans lequel la certitude en l'existence des qualia est la norme, elle peut néanmoins s'apparenter à un délire si l'on s'en tient au sens de Jaspers. Si on se refuse à qualifier cette certitude de délirante, on peut reformuler la proposition ainsi : *l'expérience des qualia est une croyance irrationnelle*. Lisa Bortolotti, qui défend une thèse selon laquelle la frontière entre croyance irrationnelle et idée délirante est arbitraire, explore trois critères de la rationalité :

- La rationalité procédurale : « [elle] concerne la façon dont les croyances interagissent et se rapportent les unes aux autres »[43] (Bortolotti, 2010, p. 16, notre traduction). Elle précise que « pour un sujet, être rationnel consiste à avoir des croyances conformes aux meilleurs standards de raisonnement disponibles »[44] (*ibid.*). La croyance en l'existence des qualia se conforme-t-elle à ce critère ? Si on prend pour standard de raisonnement l'approche bayésienne de la connaissance, aucune croyance ne peut atteindre une probabilité de cent pour cent, si bien que la certitude est à exclure. Bien entendu, le choix du raisonnement bayésien comme meilleur standard est en lui-même discutable.

[43] « Procedural rationality concerns the way beliefs interact with and relate to one another » (Bortolotti, 2010, p. 16)
[44] « According to the standard picture, for a subject to be rational is to have beliefs that conform to the best available standards of correct reasoning » (*ibid.*)

- La rationalité épistémique : « [elle] concerne la relation entre les croyances et les preuves disponibles »[45] (*ibid.*, p. 17). La question concerne ici essentiellement ce qui est acceptable comme preuve. Nous retombons ainsi sur la dichotomie de point de vue où l'existence de l'expérience consciente répondra très bien ou très mal aux putatives preuves, respectivement subjectives et objectives.
- La rationalité agentielle : Bortolotti en considère deux aspects : « si un sujet est en mesure de donner de bonnes raisons intersubjectives au contenu de sa croyance, et s'il manifeste son endossement de la croyance en agissant selon elle »[46] (*ibid.*, p. 18). Sur ce critère, la certitude de l'existence des qualia semble être prise en défaut. Aucune raison intersubjective n'est par définition recevable. En ce qui concerne l'action conforme, c'est un problème que nous étudierons plus en avant dans le cadre du paradoxe des jugements phénoménaux.

Malgré cela, affirmer que la certitude concernant l'existence des qualia est irrationnelle peut sembler fallacieux à deux égards. D'une part, l'irrationalité de cette certitude ne semble pas réfutable (*a fortiori* dans le cadre d'une conception normative de la rationalité), ce qui entrave gravement la portée de l'argument. D'autre part, il apparait que la certitude concernant l'inexistence des qualia peut faire l'objet d'une critique symétrique. Cependant, soulignons que notre objectif n'est pas, à ce stade, d'affirmer avec certitude que les qualia n'existent pas, mais de montrer que la certitude vécue de leur existence ne s'élève pas au rang de connaissance indiscutable.

La certitude en l'existence des qualia est une tromperie de soi. Les mécanismes de la tromperie de soi restent abondamment discutés (Deweese-Boyd, 2017). La tromperie de soi s'illustre par exemple dans le fait qu'une majorité d'automobilistes estiment être meilleurs conducteurs que la moyenne (Svenson, 1981), dans le refus actif d'entendre toute preuve attestant de la culpabilité d'un proche accusé d'un crime ou de la réalité d'un diagnostic grave, dans l'échec d'une institution à prendre en compte un risque pourtant reconnu individuellement par une partie, parfois majoritaire, de ses membres[47], etc. Dans chaque cas, une dimension

[45] « Epistemic rationality concerns the relation among beliefs and the available evidence, and depends upon a subject's capacity to form new beliefs that are firmly grounded on the available evidence and to update existing beliefs when relevant evidence becomes available » (*ibid.*, p. 17)

[46] « There are two aspects of agential rationality that I shall consider in this project: whether a subject is in a position to give intersubjectively good reasons for the content of a reported belief, and whether she manifests her endorsement of the belief content by acting on it in the relevant circumstances » (*ibid.*, p. 18)

[47] Cas dramatiquement bien illustré dans la série *Chernobyl* (Mazin, 2019) qui décrit la catastrophe nucléaire de Tchernobyl survenue en avril 1986 et s'intéresse notamment à l'aveuglement motivé de nombreux responsables à tous les niveaux hiérarchiques.

motivationnelle différencie la tromperie de soi de l'erreur de jugement. Une conception traditionnelle de cette dernière est calquée sur la tromperie d'autrui que nous évoquions en introduction et durant laquelle un individu croyant *~p* cherche à induire chez autrui la croyance *p* pour une raison donnée. Dans le cas de la tromperie de soi, il semble donc qu'un individu croyant *~p* façonne lui-même sa croyance que *p* pour des raisons précises (par exemple le caractère intolérable de la croyance *~p*). Alfred Mele souligne deux paradoxes concernant la tromperie de soi (Mele, 2001, p. 59) :

- Le paradoxe statique : le fait d'entretenir simultanément deux croyances contradictoires : *p* et *~p*
- Le paradoxe dynamique : le fait d'employer intentionnellement une stratégie de tromperie tout en étant dupé par cette dernière

Nous allons plus particulièrement nous intéresser au premier. Plusieurs solutions ont été avancées. Mele lui-même propose que dans certains cas, la croyance *~p* ne soit jamais entretenue par le trompeur de soi, mais qu'un traitement biaisé et orienté des données disponibles génère directement la croyance *p (ibid.*). Donald Davidson suggère qu'il n'est pas paradoxal d'entretenir la croyance *p,* la croyance *~p,* à condition de ne pas entretenir la croyance *[p ∧ ~p]*, ce qui suppose un certain cloisonnement de l'esprit (Davidson, 1986). Rendre compte ainsi de la certitude de l'existence des qualia requiert deux choses. D'une part, il convient de proposer une caractérisation de la tromperie de soi non pas en termes de croyance, mais en termes de certitude. Cette transition ne nous semble pas incompatible avec les différentes descriptions du phénomène déjà évoquées. D'autre part, il est nécessaire de rendre compte de l'aspect motivationnel de cette certitude trompeuse. Nous reviendrons sur les motifs qui peuvent la susciter, mais considérons pour le moment qu'ils existent. Si nous acceptons ces deux propositions, la tromperie de soi pourrait élégamment rendre compte de la véracité de la conjonction (II) pour *p : les qualia existent*.

La certitude en l'existence des qualia est une confabulation. Dans la première partie, nous avions déjà utilisé le cas de la confabulation pour appuyer l'idée que l'émotion-certitude puisse s'activer à tort, en nous reposant sur l'exemple du syndrome de Korsakov. Un autre cas de confabulation étudié par Hirstein nous parait plus directement lié à la question des qualia : le syndrome d'Anton. Ce dernier est lié à une cécité corticale[48] associée à des lésions frontales

[48] La cécité corticale est liée à une lésion bilatérale du cortex occipital responsable du traitement de l'information destinée à la perception visuelle consciente. Le reste du système visuel est préservé, de l'œil aux régions cérébrales chargées de fonctions annexes, telles que l'hypothalamus chargé du rythme nycthéméral, le colliculus supérieur

pouvant engendrer une anosognosie (Hirstein, *op.cit.*, p. 146). Le patient, aveugle, mais non conscient de ses troubles, confabule sur ses expériences visuelles. « Par exemple, si on leur demande de décrire quels habits leur médecin porte, ils vont fournir la description détaillée d'un médecin générique. »[49] (Hirstein, *op.cit.*, p. 12, notre traduction). La difficulté posée par le syndrome d'Anton est que les lésions responsables de l'anosognosie sont généralement associées à d'autres troubles cognitifs en lien avec l'atteinte frontale (désinhibition, trouble du jugement, labilité émotionnelle, etc.). De ce fait, une évaluation de leur certitude quant à l'existence de qualia associées à la perception visuelle alléguée est compromise. Il semble toutefois conceptuellement possible d'associer une cécité corticale à une anosognosie isolée de tout autre déficit cognitif. Par ailleurs, les fonctions visuelles annexes étant préservées dans la cécité corticale (cf. note 48), un tel patient pourrait présenter un phénomène de vision aveugle (*blindsight*), voire devenir un « superblindsighter », en mesure de se servir des indices fournis par ces voies accessoires pour se comporter d'une manière indifférenciable d'un individu voyant (Block, 1995). Ned Block suggère que le superblindsighter pourrait représenter un cas de zombie visuel (quasi-zombie) philosophique. Qu'en serait-il de l'Anton-superblindsighter (ASBS) ?[50] Peut-être que l'ASBS représenterait un cas encore plus exact de zombie visuel, puisque contrairement au superblindsighter, il serait indiscernable y compris sur le plan du report verbal (voire sur un plan fonctionnel). Quelle conclusion en tirer ? L'ASBS pourrait représenter un cas paradigmatique de qualia confabulées. Il aurait la certitude d'avoir une expérience visuelle et cette certitude se verrait en quelque sorte confirmée par un observateur extérieur, incapable de différencier l'ASBS d'un autre individu. Pourrait-on rendre compte de toutes les expériences conscientes de la même manière ? Certes, les exemples de confabulation que nous avons invoqués jusque-là étaient tous de nature pathologique, mais Hirstein souligne que la confabulation n'est pas l'apanage des malades de l'encéphale. De nombreux cas de confabulation normale sont décrits : ils seraient fréquents chez les enfants et l'expérience de Nisbett et Wilson a mis en évidence un mécanisme de reconstruction confabulatoire des motivations liées à une prise de décision retrouvé chez une proportion importante de volontaires dénués de troubles neurologiques.[51]

---

en charge de la coordination des mouvements et le prétectum responsable du contrôle réflexe de la pupille et du cristallin.

[49] « For instance, if asked to describe what their doctor is wearing, they will provide a full description of a generic doctor. » (Hirstein, 2006, p. 12)

[50] L'ASBS partagerait la cécité corticale et l'anosognosie (supposément isolée) du syndrome d'Anton et les capacités du superblindsighter.

[51] Hirstein nous résume l'expérience de la façon suivante : « Nisbett et Wilson (1977) ont disposé sur une table dans magasin des paires de bas de nylon et ont demandé aux clients de choisir la paire qu'ils préféraient. À l'insu

Si l'une de ces propositions, ou toute autre en mesure de rendre compte d'une façon naturaliste de la certitude en l'existence de l'expérience consciente, est ne serait-ce qu'acceptable, cela amoindrit de fait la force de l'argument selon lequel cette certitude impose de construire un arsenal théorique à même de rendre compte de cette existence.

Un doute qui ne nous atteint pas

Nous avons suggéré plusieurs propositions qui pourraient rendre compte de la certitude en l'existence des qualia même en leur absence. Nous allons à présent explorer la question sous un angle légèrement différent, en proposant des explications de l'impossibilité de douter de cette dernière. Sur un plan formel, le raisonnement est tout à fait analogue, puisqu'il s'agit d'étudier l'implication :

(I') [*p* est indubitable] → *p*

Si on considère la symétrie entre le doute et la certitude comme parfaite (I) et (I') sont interchangeables et les propositions qui vont suivre peuvent directement s'ajouter à la liste précédente. Dans la mesure où cette symétrie peut être discutée, il nous est apparu pertinent de les présenter séparément.

L'absence de doute concernant les qualia est liée à un défaut cognitivo-émotionnel. Au même titre qu'en introduction nous évoquions la possibilité que la résolution du problème difficile de la conscience soit entravée par un défaut qualitatif ou quantitatif de nos capacités cognitives, l'absence de doute pourrait s'expliquer par des défauts similaires concernant nos capacités à douter. Notre cerveau ne serait pas capable de générer l'émotion du doute à propos de l'existence des qualia. Cette explication peut paraître *ad hoc* dans la mesure où il serait étonnant que la question de l'expérience consciente soit une exception notable au milieu d'un océan de propositions philosophiques sujettes au doute (de l'existence d'un monde extérieur à l'existence d'un soi, en passant par l'écoulement du temps). Pourquoi aurait l'existence d'une

des sujets, les paires étaient identiques. Les gens avaient tendance à choisir la paire la plus à droite pour une raison floue, mais lorsqu'on leur demandait la raison de leur choix, les clients mettaient en avant la couleur et les textures des nylons. Quand on leur dît que les nylons étaient identiques et leur expliqua l'effet de position, les sujets tendaient tout de même à résister à ces explications et à maintenir leurs raisons initiales. » (« Nisbett et Wilson (1977) set up a table in a department store with pairs of nylon stockings and asked shoppers to select the pair they preferred. Unbeknown to the shoppers, the pairs where identical. People tended to choose the rightmost pair for reasons that are not clear, but when asked the reason for their choice, the shoppers commented on the color and texture of the nylons. When they were told that the nylons were identical, and about the position effects, the shoppers nevertheless tended to resist this explanation and stand by their initial reasons. », Hirstein, 2006, p. 3)

expérience consciente cette particularité ? Une explication pourrait être son caractère immédiat : tandis que l'existence du monde, de soi ou de l'écoulement du temps semblent être médiée par nos sens ou notre raison, l'existence de l'expérience en elle-même loin d'être médiée serait justement le médiateur. C'est le critère d'appréhension directe de Dennett (cf. page 13). Or, même Chalmers accepte que nos perceptions et nos idées puissent être décrites en termes strictement fonctionnels, si bien que le caractère médiateur de l'expérience consciente peut être absent du processus tout en étant subjectivement indubitable.

L'absence de doute est liée à une anorexie épistémique. La partie de la rétine où s'insère le nerf optique est dépourvue de photorécepteurs, si bien qu'un objet situé dans la région du champ visuel correspondante sera invisible : c'est le phénomène de la tache aveugle. Cependant, même en vision monoculaire, nous ne percevons pas de « vide » dans notre champ visuel. Il est généralement considéré que le cerveau remplit les trous, expliquant ainsi l'apparente continuité du champ visuel. Or, « une absence d'information n'est pas l'information d'une absence » (« An absence of information is not the same as information about an absence. », Dennett, 1993, p. 324). Une explication alternative serait donc que notre esprit loin de combler les manques aurait tendance à les ignorer, à moins d'avoir de bonnes raisons (c'est-à-dire que cela soit porteur d'une valeur sélective) et la possibilité d'en faire autrement. Dennett décrit cette tendance à la négligence, dont les cas d'anosognosies, comme une « perte d'appétit épistémique » (« loss of epistemic appetite », *ibid.*, p. 356). Suivant cette hypothèse, il se peut que l'absence de doute concernant l'existence des qualia soit elle-même une forme d'anorexie épistémique, une incapacité à générer l'émotion du doute produisant une apparente certitude analogue à l'apparente continuité de notre champ visuel.

Nous pouvons traiter ces deux propositions de deux manières. Nous pouvons tout d'abord les considérer comme des possibilités à part entière devant être ajoutées à la disjonction des propositions concernant la certitude. Remarquons toutefois qu'elles sont dans le fond assez similaires. L'anorexie épistémique pouvant être liée au défaut cognitivo-émotionnel préalablement évoqué. De plus, il est possible de s'en servir pour renforcer la proposition selon laquelle la certitude en l'existence des qualia est une confabulation. En effet, elles supposent toutes deux des aspects de la confabulation qu'on retrouvait dans le cas de William, respectivement l'abrasion émotionnelle et le désintérêt pour l'objet de la confabulation.

Quel vide théorique les qualia viennent-elles combler ?

Chalmers défend l'idée que l'expérience consciente est à tel point indubitable, dans son existence comme dans sa nature, que l'existence des qualia *doit* être intégrée à une théorie du monde ayant prétention à l'exhaustivité. Il affirme donc que toute théorie strictement physicaliste laisserait de côté certains aspects du monde. Cependant, la possibilité d'explications alternatives à cette indubitabilité diminue la portée de l'argument. Il est clair que l'existence d'une entité ou d'une propriété est une bonne raison d'être certain de son existence. Comme nous l'avons vu, si l'on parvient à proposer une explication alternative à cette certitude (A) *et* que nous montrons que ces explications sont plus parcimonieuses que de postuler l'existence (B), alors nier l'existence peut finalement être la meilleure des explications (C).

Les explications (A) possibles que nous avons proposées se sont focalisées sur la faillibilité de la certitude elle-même, en tant qu'émotion épistémique. Le programme illusionniste suggère que l'erreur puisse se situer à d'autres niveaux fonctionnels (Frankish, 2016)[52]. Il nous incombe à présent d'évaluer le coût théorique de l'existence des qualia afin de pouvoir affirmer (B). Le cas échéant, (C) en découlera.

Reste-t-il une place pour d'irréductibles qualia ?

Zombie-Mary et le paradoxe du jugement phénoménal

En introduction, nous avons succinctement présenté trois arguments classiques pour l'existence des qualia : la possibilité d'un monde zombie, le gouffre explicatif et les qualia inversées. Nous allons à présent nous intéresser à un quatrième argument, avancé par Franck

[52] Notamment au niveau des perceptions ou de l'introspection. Nous ne rejetons pas ces éventualités sur lesquelles nous ne nous sommes cependant pas focalisés. Elles ne sont d'ailleurs pas mutuellement exclusives et peuvent même s'envisager comme d'autres explications de la certitude (A). Inversement, nous pouvons considérer que notre analyse vient alimenter le programme illusionniste. Cela suppose qu'expliquer la certitude reste malgré tout insuffisant. *Non seulement nous sommes certains que l'expérience consciente est là, mais en plus elle* est *là !* L'illusionnisme pourrait apporter des clefs de compréhension de cet *en plus*. Toutefois, nous remettons en question la nécessité de ces explications additionnelles. La proposition n'est-elle pas trompeuse, laissant implicite un élément qui la rendrait pléonastique ? *Non seulement nous sommes certains que l'expérience consciente est là, mais en plus elle* est *là,* nous en sommes certains !
L'approche illusionniste entretient une étroite proximité avec la nôtre, étayant le matérialisme éliminativiste tout en prenant au sérieux le dilemme de Chalmers, cf. note 11

Jackson : l'expérience de Mary, la spécialiste des couleurs. Jackson la présente de la façon suivante :

> « Mary est une brillante scientifique qui est, pour quelque raison, forcée d'étudier le monde dans une pièce en noir et blanc via un moniteur en noir et blanc. Elle se spécialise dans la neurophysiologie de la vision et acquiert, supposons-le, toutes les informations physiques à propos de ce qu'il se passe quand nous voyons des tomates mûres, ou le ciel, et que nous utilisons des termes tels que « rouge », « bleu », etc. Elle découvre, par exemple, quelle combinaison de longueurs d'onde du ciel stimule la rétine, et exactement comment cela produit, via le système nerveux central, la contraction des cordes vocales et l'expulsion d'air depuis les poumons qui résultent en l'énonciation de la phrase « le ciel est bleu ». […]. Qu'arrive-t-il lorsque Mary est délivrée de sa pièce noire et blanche ou lorsqu'il lui est donné un moniteur coloré ? Apprend-elle quelque chose, ou non ? Il semble évident qu'elle apprend quelque chose à propos du monde et de notre expérience de ce dernier. Mais il est alors inexplicable que sa connaissance préalable soit incomplète. Elle avait toutes les informations physiques. Ainsi, il y a plus que cela, ce qui implique que le physicalisme est faux. »[53] (Jackson, 1982, notre traduction).

Supposons à présent que vive dans le monde zombie l'alter ego de Mary, Zombie-Mary. Par définition, Zombie-Mary est placée dans une situation physiquement similaire à celle de Mary. Dès lors, nous pouvons comparer les destins de Mary et Zombie-Mary lors de leur sortie de la pièce. Selon Jackson, Mary apprend quelque chose de nouveau : ce que cela fait de voir des couleurs. Cela pourra par exemple se manifester par une exclamation « c'est donc cela voir du bleu !? ». Dans le monde de Zombie-Mary, cela ne fait rien si bien que, pour sa part, elle n'apprend rien de nouveau. Comment Zombie-Mary va-t-elle se comporter à sa sortie de la pièce ? Si Zombie-Mary est indifférente à la vue de ciel lors de sa sortie, son comportement différera de celui de Mary, ce qui est en contradiction avec le postulat selon lequel le monde de Mary et de monde zombie sont physiquement indifférenciables. Cela contreviendrait par ailleurs au principe de fermeture causale du monde physique, puisqu'il faudrait faire appel à une cause non physique pour expliquer la différence de comportement physique entre nos deux protagonistes. Il est donc plus cohérent de considérer que le comportement de Zombie-Mary se

[53] « Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor. She specializes in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like 'red', 'blue', and so on. She discovers, for example, just which wave-length combinations from the sky stimulate the retina, and exactly how this produces via the central nervous system the contraction of the vocal chords and expulsion of air from the lungs that results in the uttering of the sentence 'The sky is blue'. […] What will happen when Mary is released from her black and white room or is given a color television monitor? Will she *learn* anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had *all* the physical information. *Ergo* there is more to have than that, and Physicalism is false. » (Jackson, 1982)

superpose à celui de Mary ; qu'elle s'exclame également « c'est donc cela voir du bleu !? ». Mais dans ce cas, quelle est en définitive la portée de l'argument de Jackson ? En quoi est significatif le fait que Mary semble apprendre quelque chose si Zombie-Mary parait en faire de même en l'absence de qualia ?[54] Ce qu'il se passe dans le cas de Zombie-Mary, c'est que nous ajoutons au développement de Jackson une prémisse supplémentaire : *le physicalisme est vrai*. Avec cette nouvelle prémisse, la conclusion reste identique, *le physicalisme est faux*.[55] Mais elle entre de fait en contradiction avec cette prémisse additionnelle.

Zombie-Mary nous place donc face à un choix : soit nous acceptons l'argument du monde zombie de Chalmers, soit nous acceptons le « knowledge argument » de Jackson. Mais accepter les deux implique des contradictions telles que celle que nous venons d'évoquer. Chalmers parle de « paradoxe du jugement phénoménal » pour désigner des problèmes analogues (Chalmers, *op.cit.*, p. 172).

Le paradoxe du jugement phénoménal et la place des qualia

Chalmers formalise le paradoxe du jugement phénoménal de la manière suivante :

> « (1) Le domaine physique est causalement fermé ; (2) les jugements concernant la conscience surviennent logiquement sur la physique ; (3) la conscience ne survient pas logiquement sur la physique ; (4) nous savons que nous sommes conscients. Des prémisses (1) et (2) il suit que les jugements concernant la conscience sont explicables réductivement. En combinaison avec la prémisse (3), cela implique que la conscience est sans importance pour nos jugements, ce qui entre en conflit avec la prémisse (4). »[56] (Chalmers, *op.cit.*, p. 183, notre traduction).

Par souci de clarté, nous allons reformuler les deuxième et troisième prémisses : (2') *les jugements concernant l'expérience consciente surviennent logiquement sur la physique* ; (3') *l'expérience consciente ne survient pas logiquement sur la physique*. Chalmers consacre la suite

[54] Il pourrait nous être objecté que Mary apprend réellement quelque chose tandis que Zombie-Mary n'a que l'illusion d'apprendre quelque chose ; que bien qu'elles aient les mêmes jugements, elles ne sont pas dans le même état épistémique. Cependant, comment pourrait un observateur extérieur faire cette distinction et s'en servir dans une démonstration ?

[55] Physicalisme est ici entendu dans le sens de matérialisme éliminativisme. Nous supposons en effet que c'est dans ce sens-là que Jackson emploie ici le terme physicalisme.

[56] « The paradox is a consequence of the facts that (1) the physical domain is causally closed; (2) judgments about consciousness are logically supervenient on the physical; (3) consciousness is not logically supervenient on the physical; and (4) we know we are conscious. From premises (1) and (2) it follows that judgments about consciousness can be reductively explained. In combination with premise (3), this implies that consciousness is explanatorily irrelevant to our judgments, which lies in tension with premise (4) » (Chalmers, 1996, p. 183)

de son chapitre à soutenir qu'en dépit d'une tension apparente, ces quatre prémisses ne sont pas incompatibles. Or, il nous semble qu'une prémisse implicite soit en fait à l'origine de cette tension : (5) *savoir comme nous sommes conscients implique que nous soyons conscients* (au sens de conscience phénoménale)[57]. En effet, la conjonction de (1), (2'), (3') et (4) ne semble pas paradoxale si on minimise la portée de notre supposé savoir concernant la conscience. Seule la prémisse (5) génère la tension qu'évoque Chalmers et l'amène à un développement visant à réconcilier cet ensemble de prémisses d'allure incompatibles. À noter que Chalmers parle ici de « savoir », mais que cette transition implicite de la certitude au savoir implique que la certitude ait une valeur justificative dans l'établissement de connaissances. Nous pouvons ainsi ajouter une autre prémisse implicite : (6) *nous savons que nous sommes conscients, car l'expérience consciente est indubitable*.

La prémisse (6) soulève donc le problème de la valeur justificative de l'émotion épistémique dans la formation d'une connaissance. Nulle question ici d'en contester toute portée ; nous avons au contraire défendu l'idée que les émotions épistémiques avaient un rôle central dans l'édification de nos savoirs. Néanmoins, puisque ces émotions sont apparues grâce aux processus de l'évolution, ce n'est pas leur fidélité au réel, mais leur utilité qui les ont façonnées. Lorsqu'une représentation exacte de la réalité présente un avantage sélectif, elle sera favorisée.[58] Cependant, lorsqu'il s'agit de questions plus métaphysiques, dont l'utilité biologique est moins évidente, il se peut que d'autres qualités que la fidélité aient été privilégiées ; ou encore qu'aucune ne l'ait particulièrement été, en l'absence de pression de sélection dans ce contexte. Ainsi, nous ne contestons pas de la valeur justificative de la certitude en général, mais spécifiquement dans le cas présent. Selon le sens que Chalmers donne à « savoir », nous pouvons ainsi, soit contester (6), ce qui implique de rejeter (4), soit conserver (4), en reformulant (4) et (5) de la façon suivante : (4') *nous sommes certains que nous avons une expérience consciente* ; (5') *être certains que nous avons une expérience consciente implique que nous ayons une expérience consciente*. Étant donné que nous acceptons (4'), nous allons donc conserver (1), (2), (3'), (4') et (5').

[57] Implication que nous ne remettons pas en cause sur le plan formel (cf. note 42).

[58] McKay et Dennett nous rappellent néanmoins que « bien que la survie soit la seule monnaie forte de la sélection naturelle, le taux de change avec la vérité est probablement honnête dans la plupart des circonstances. » (« Although survival is the only hard currency of natural selection, the exchange rate with truth is likely to be fair in most circumstances. », McKay et Dennett, 2009). La survie mentionnée dans ce contexte signifie bien évidemment la persistance de copies du réplicateur (cf. page 59), ce qui inclut sa reproduction. Il semble que la fidélité au réel de nos représentations constitue généralement un net avantage. Toutefois, comme nous le verrons, cette règle peut comporter des exceptions.

Par ailleurs, force est de constater que (5') requiert la mobilisation d'un appareil théorique qui dépasse de très loin celui fourni par la seule physique (et les sciences qui y sont réductibles). Outre les entités ou propriétés que sont les qualia, des lois naturelles sont requises pour expliquer leur relation aux entités et propriétés du monde physique (par exemple l'étonnante corrélation de laquelle découlent les lois de cohérence et d'invariance organisationnelle). Nous pouvons admettre que la légitimité d'une théorie scientifique se mesure à sa capacité à faire des prédictions plus correctes que les théories concurrentes. Or, d'un point de vue objectif, la seule prédiction que nous apporte cette théorie est justement celle de la *formulation* de jugements phénoménaux. L'existence des qualia permet, nous ne le nions pas, d'expliquer de façon parcimonieuse pourquoi nous sommes si certains de leur existence.[59] Expliquer cette certitude par d'autres biais nécessite également l'élaboration d'un appareil théorique non négligeable. Néanmoins, cette dernière ne requiert pas une révision complète de notre compréhension du monde avec l'ajout d'un ensemble théorique distinct de la physique. Les seuls présupposés supplémentaires qui sont requis sont psychologiques. Or, d'après (2), ces jugements surviennent sur la physique. Il y a donc un profond déséquilibre entre les coûts théoriques présentés par une explication psychologique (des jugements concernant la conscience) et une explication ontologique telle que (5').[60] Le principe de parcimonie nous invite donc à rejeter (5'). En conservant (1), (2') et (3'), nous aboutissons à une conception matérialisme éliminativiste de la conscience. Étant donné que nous n'avons pas souhaité disqualifier (4'), il nous incombe encore de proposer une explication de (4') qui soit compatible avec l'éliminativisme.

[59] Ce point reste toutefois discutable : compte tenu de l'inertie causale des qualia, même leur existence ne constituerait pas une explication suffisante des jugements phénoménaux.

[60] Notons que l'existence des qualia n'exclut par ailleurs pas les explications alternatives à la certitude de leur existence. Cela reviendrait cependant à payer à la fois le prix de l'explication ontologique *et* celui de l'explication psychologique. Il se peut tout à fait qu'elles existent sans jouer de rôle dans notre certitude de leur existence. C'est d'ailleurs possiblement cela qui différencierait un monde qualitatif et un monde zombie en tout point identiques sur le plan physique. Cependant, de façon analogue à la surdétermination que les qualia présenteraient sur l'enchaînement causal du monde physique si on leur attribuait une efficacité causale, ne serions-nous pas face à un cas de *surexplication* vis-à-vis l'appareil explicatif du monde physique si on leur attribuait une valeur explicative ?

Simplifier le problème difficile[61]

Notre travail nous a conduits à une reformulation du paradoxe du jugement phénoménal tel que :

- (1) le domaine physique est causalement fermé
- (2') les jugements concernant l'expérience consciente surviennent logiquement sur la physique
- (3') l'expérience consciente ne survient pas logiquement sur la physique
- (4') nous sommes certains que nous avons une expérience consciente

Précisons que par « jugements » nous entendons tout simplement l'enchaînement causal qui génère notamment le comportement verbal « formuler des affirmations »[62]. C'est ce processus causal qui survient logiquement sur la physique. Il suit de (2') et (3') que même si l'expérience consciente existe, elle ne nous permet pas de porter de jugement sur son existence ou sa nature (ce que Chalmers souligne, cf. citation page 54). La transition entre l'expérience et le jugement impliquerait en effet soit l'efficacité causale de cette expérience sur le jugement, ce qui enfreint (1), soit un cas de surdétermination. C'est pourquoi le principe de parcimonie nous invite donc à nous passer purement et simplement de ce concept, permettant d'ajouter à nos prémisses : (0) *nous n'avons pas d'expérience consciente*. On peut dès lors formuler la conjonction : *nous sommes certains que nous avons une expérience consciente* (4') et *nous n'avons pas d'expérience consciente* (0). Nous retrouvons ici la conjonction (II). Cette conjonction d'apparence paradoxale requiert elle-même une explication. Une partie du travail a déjà été effectué. En effet, nous avons suggéré plusieurs propositions alternatives qui pourraient expliquer (4') tout en respectant (0). Cependant, ces dernières laissent un goût d'inachevé dans la mesure où elles paraissent *ad hoc*. Pourquoi serions-nous tous sujets à une tromperie de nous ? Pourquoi partagerions-nous un délire collectif ? Pourquoi tiendrions-nous le même discours confabulatoire ? La question reste analogue quelles que soient l'explication ou les explications auxquelles on adhère : pourquoi partageons-nous universellement la

[61] D'une manière analogue, Frankish affirme que « l'illusionnisme remplace le problème difficile par le problème de l'illusion » (« Illusionism replaces the hard problem with the illusion problem », Frankish, 2016). Il propose pour méthode de « former des hypothèses sur les mécanismes cognitifs sous-jacents et leurs bases neurophysiologiques et neuroanatomique » (« to form hypotheses about the underlying cognitive mechanisms and their bases in neurophysiology and neuroanatomy », *ibid.*). Nous nous focalisons ici sur le versant psychologique.
[62] Comportement auquel un fonctionnaliste pourra, sans contrevenir à l'argument, ajouter la disposition à générer ce comportement, la croyance en la validité de ces jugements et tous les autres ajouts qu'il estime nécessaire à une caractérisation plus précise de ces jugements, sous réserve que ces ajouts surviennent logiquement sur la physique.

certitude de l'existence d'un inexistant ? Rappelons-nous que nos croyances et certitudes ne sont qu'une *représentation* du monde et que la nature ne nous a pas fourni les ressources requises pour que ces représentations soient parfaites[63]. Néanmoins, une explication basée sur un « fonctionnement normal générant des erreurs qui, sans être en elles-mêmes adaptatives, semblent tolérables »[64] (McKay et Dennett, 2009, notre traduction) ou *a fortiori* sur un dysfonctionnement pathologique ne nous parait pas satisfaisante face à l'ubiquité de la certitude. McKay et Dennett suggèrent que, bien que dans l'écrasante majorité des situations il est avantageux d'entretenir des croyances avérées (cf. note 58), il est possible que parfois une tendance systématique à se tromper soit bénéfique. D'une part, ce biais peut être le prix à payer en échange d'un avantage adaptatif supérieur, ce que McKay et Dennett appellent des « sous-produits » (« by-products », *ibid.*). D'autre part, il semble que dans certains cas, ces certitudes erronées puissent être en elles-mêmes avantageuses (ils retiennent en particulier le cas des biais d'optimisme[65]). Quelle qu'en soit l'explication évolutive, une explication psychologique complémentaire s'impose. Nous proposons que cette dernière repose sur le fait que l'expérience phénoménale représente pour nous une richesse, une part de nous-mêmes et un socle moral. L'aversion à la perte (Kahneman, 2016, p. 435) génère un effroi à l'idée d'y renoncer.

Nous nous étions servis de l'exemple de la peur pour illustrer la valeur épistémique de nos émotions. Cela impliquait de leur attribuer des qualités intrinsèques de sensibilité et de spécificité, dont le réglage serait le fruit d'une adaptation à deux vitesses : l'évolution de l'espèce et l'expérience de l'individu. Les cas pathologiques suggèrent que le seuil de déclenchement de la peur déterminé par ces qualités intrinsèques peut s'avérer désadapté : un trouble de l'adaptation à l'origine de pathologies anxieuses telles que l'anxiété généralisée ou le trouble panique[66]. Néanmoins, même dans les cas non pathologiques, un seuil de

[63] Nous préférons parler de certitude erronée tandis que McKay et Dennett étudient les « erreurs de croyance » (« misbelief », McKay et Dennett, 2009). Croyances et certitudes partagent la direction d'ajustement esprit-monde, et nous n'avons pas eu besoin de faire un appel direct à la notion de croyance dans cet essai, si bien que nous proposons une lecture de l'article de McKay et Dennett fondée sur la notion de certitude erronée plutôt que d'erreur de croyance. Cela ne devrait pas affecter notre développement.

[64] « by functioning normally [create] families of errors that are, if not themselves adaptative, apparently tolerable » (McKay et Dennett, 2009)

[65] McKay et Dennett les appellent « illusions positives » (« positive illusions », McKay et Dennett, 2009). Il s'agit notamment des « auto-évaluations positives irréalistes, les perceptions exagérées de contrôle et de maîtrise et d'un optimiste irréaliste concernant le futur » (« unrealistically positive self-evaluations, exaggerated perceptions of personal control or mastery, and unrealistic optimism about the future », *ibid.*). Cette considération va dans le sens de la note 26.

[66] Le DSM-5 définit l'anxiété généralisée comme un trouble lié à des soucis excessifs quasi permanents et incontrôlables associés à des symptômes tels que des difficultés d'endormissement, une tension musculaire, une irritabilité, etc., et le trouble panique comme la survenue récurrente et inattendue d'attaques de panique suivies du développement d'une appréhension de ces récurrences pouvant se manifester par des comportements d'évitement. (American Psychiatric Association, 2015)

déclenchement de la peur privilégiant la sensibilité au détriment de la spécificité a pu s'avérer d'un précieux secours du fait de conséquences potentielles asymétriques entre un faux positif (dépense énergétique inutile) et un faux négatif (ignorer la présence d'un prédateur). Ainsi, au même titre que les autres émotions, la valeur épistémique de la peur dépend moins de sa probité que de son utilité. Nous proposons ici que la perte des trois qualités de l'expérience consciente (richesse, socle, identité) génère une crainte inadaptée à l'origine de notre tendance psychologique à affirmer la certitude de l'expérience consciente. Dans le cas où l'on attribue celle-ci à un mécanisme de tromperie de soi, ces peurs peuvent être assimilées aux motivations de la duperie.

La peur du noir

> « Il y a environ un siècle, la littérature du courant de conscience commença à être promulguée par des auteurs tels que James Joyces, Marcel Proust et Arthur Schnitzler. En lieu et place d'une traditionnelle présentation propre, travaillée et organisée des pensées et conduites du protagoniste, ces auteurs tentèrent de décrire la vie intérieure d'une manière réaliste et par conséquent chaotique et confuse. On pourrait dire qu'ils cherchaient un rendu plus approprié de la vie interne sur un plan phénoménologique. La technique gagna rapidement popularité et renom. Cela soulève la question de ce qui la rend si attractive, mais intuitivement on pourrait penser qu'elle nous offre un aperçu de « ce que cela fait d'être une telle personne dans une telle situation ». Hors de la fiction, nous n'avons d'aperçu direct que de notre propre conscience. Cette solitude épistémique est brisée par les narrations de courants de conscience qui nous offre l'aperçu d'une autre conscience. »[67] (Kriegel, *op.cit*., p. 1, notre traduction).

Cet extrait illustre l'importance que nous accordons à l'expérience consciente. C'est d'elle que découlerait la richesse de notre vie intérieure : la splendeur d'un ciel étoilé loin des villes, la délectation de la première gorgée de bière, la sensualité d'une caresse ou le frisson des premiers accords de *Breathe* (Pink Floyd, 1973). C'est ainsi que pas même la plus fine des

[67] « About a century ago, stream-of-consciousness literature started being promulgated by such writers as James Joyce, Marcel Proust, and Arthur Schnitzler. Instead of the traditional well-organized, cleaned-up, highly processed presentation of characters' thought and conduct, these writers attempted to describe inner life in a realistic, hence somewhat chaotic and confused, fashion. We might say they were seeking a more phenomenologically adequate rendering of inner life. The technique has quickly gained popularity and renown. It is a good question what is so compelling about it, but one immediate thought is that it offers insight into questions of the form "What is it like to be this kind of person, in this kind of situation?" Outside fiction, we have direct insight only into our own consciousness; this epistemic loneliness is broken by stream-of-consciousness narratives that offer a believable peep into another consciousness. » (Kriegel, 2015, p. 1)

descriptions de ces sensations ne semble rendre compte de *ce que cela fait*. Nous acceptons la possibilité d'avoir un double zombie, mais serions terrorisés par la perspective d'en devenir un.

Cependant, notre hypothétique double zombie ne se poserait-il pas exactement les mêmes questions ? Toutes ces sensations ne peuvent-elles pas être décrites en termes strictement fonctionnels ? Les auteurs que cite Uriah Kriegel semblent pourtant y parvenir ! Quand nous avons peur « d'éteindre la lumière », de quoi avons-nous réellement peur ? De perdre ce qui fait le sel de l'existence ou de perdre la chimère à laquelle on attribuait ce sel ? Un zombie philosophique ne se plaindrait pas de manquer de quelque chose, mais il aurait très certainement la même peur du noir.

La peur de l'absence

> « Le simple fait que les organismes ont des expériences conscientes signifie, au fond, qu'il y a quelque chose que cela fait d'*être* cet organisme. […] Nous pouvons appeler cela le caractère subjectif de l'expérience. Il n'est capturé par aucune des récentes analyses réductives du mental, ces dernières étant toutes compatibles avec son absence. Il n'est pas analysable en termes d'états fonctionnels ou d'états intentionnels puisque ceux-ci pourraient être attribués à des robots ou des automates se comportant comme des personnes, mais ne faisant l'expérience de rien. Pour des raisons similaires, il n'est pas analysable en termes de rôle causal des expériences sur les comportements humains. Je ne nie pas que les états et événements mentaux causent des comportements, ni qu'ils puissent être caractérisés fonctionnement. Je nie simplement que ces choses suffisent à leur analyse. »[68] (Nagel, *op.cit.*, notre traduction).

Deux critères classiques sont censés caractériser l'identité personnelle : la continuité physique et la continuité mnésique. La matière qui compose notre corps se renouvelle sans cesse, si bien que nous n'avons plus grand-chose en commun avec celui que nous étions il y a plusieurs années. De plus, si on remplaçait un par un chacun des neurones de notre cerveau par leur équivalent fonctionnel en silicium, à partir de quand ne serions-nous plus nous-mêmes

[68] « the fact that an organism has conscious experience *at all* means, basically, that there is something it is like to *be* that organism. […] We may call this the subjective character of the experience. It is not captured by any of the familiar, recently devised reductive analyses of the mental, for all of them are logically compatible with its absence. It is not analyzable in terms of any explanatory system of functional states, or intentional states, since these could be ascribed to robots or automata that behaved like people though they experienced nothing. It is not analyzable in terms of the causal role of experiences in relation to typical human behavior – for similar reasons. I do not deny that conscious mental states and events cause behavior, nor that they may be given functional characterizations. I deny only that this kind of thing exhausts their analysis. » (Nagel, 1974)

(Parfit, 1987, p. 474) ? Lorsque nous empruntons comme chaque matin un téléporteur pour nous rendre au travail sur Mars, la copie qui en sort à soixante-seize millions de kilomètres est-elle *nous* bien qu'aucune des particules qui composent son corps n'a foulé la planète bleue et alors qu'un corps qui nous semblait être *nous* a été détruit sur cette dernière (*ibid.* p. 199) ? Dans chacun de ces cas, nous le resterions si nous nous référons au deuxième critère : celui de la continuité mnésique. Selon celui-ci, ce sont nos souvenirs qui constituent notre identité. Mais alors, qui était cette personne dans notre corps le soir dont un excès d'alcool ne nous a laissé aucun souvenir (*ibid.* p. 166) ? Et le jour où le téléporteur vers Mars a dysfonctionné si bien qu'une copie martienne est apparue sans que l'original terrestre soit détruit : laquelle de ces deux mêmes continuités mnésiques étions-*nous* ? Avec de multiples expériences de pensées telles que celles-ci, Derek Parfit parvient à nous convaincre que les deux principaux critères de l'identité personnelle sont compromis. Toutefois, comme l'évoque Nagel, il semble rester *quelque chose que cela fait d'être* nous.

L'expérience consciente de soi peut être invoquée pour expliquer ce supposé lien qui transcende le temps et qui justifie que nous soyons la même personne tout au long de notre existence (ou du moins toute la durée d'un flux de conscience ininterrompu, ce qui est nettement plus bref). Chalmers la décrit comme « une sorte de bruit de fond […] qui est d'une certaine manière fondamental, présent même quand les autres composants de la conscience ne le sont pas. »[69] (Chalmers, *op.cit.*, notre traduction). Cependant, sommes-nous bien certains de la réalité d'un tel « bruit de fond » ? Pour Hume, « à aucun moment je ne puis me saisir moi sans saisir une perception, ni ne puis observer autre chose que ladite perception » (Hume, 1995, p. 343). Ainsi, de manière analogue à l'expérience consciente, la meilleure explication de l'identité personnelle est peut-être simplement d'en nier la réalité, aussi certaine qu'elle peut paraitre.

Cette perspective est suffisamment angoissante à bien des égards : pourquoi préservons-nous notre santé si ce n'est pas nous qui en bénéficierons demain ? Pourquoi ne devenons pas d'invétérés court-termistes : les « théoriciens du but présent » (« *present-aim theorists* ») que présente Parfit (*op.cit.*, p. 177) ? Quelles en seraient les implications morales, si l'on ne pouvait juger coupable une personne pour les crimes commis par celles avec qui elle ne partage que

[69] « a kind of background hum, for instances, that is somehow fundamental to consciousness and that is there even when the other components are not. » (Chalmers, 1996, p. 10)

continuités physique et mnésique, bien insuffisantes à établir l'identité ? Ce ne sont pas les seules implications moralement dérangeantes du matérialisme éliminativiste.

La peur du vide

> « La nature a placé l'humanité sous la souveraineté de deux maîtres, la douleur et le plaisir. C'est à eux et eux seuls de désigner ce que nous devons faire et ce que nous devrions faire. Les notions de bien et de mal tout comme l'enchaînement des causes et des effets sont soumis à leur empire. Ils nous gouvernent dans tout ce que nous faisons, disons, pensons : chaque effort que nous faisons pour nous soustraire à leur sujétion ne fait que démontrer et confirmer cette dernière. Dans son discours, un homme peut prétendre abjurer leur autorité : en réalité il va rester aliéné tout du long. Le principe d'utilité reconnait cette soumission et l'assume pour fonder un système dont l'objet est d'élever l'édifice du bonheur avec les outils de la raison et de la loi. »[70] (Bentham, 2017, p. 11, notre traduction)

Selon Jeremy Bentham, la douleur et le plaisir doivent être à l'origine de principes moraux et législatifs. Est moralement bon ce qui est propice à engendrer du plaisir ou à éviter de la douleur, et réciproquement. Un choix moral doit être pris en fonction de ses conséquences en termes de peines et de joies et un système législatif convenable devrait s'appuyer sur ces principes. Ainsi, douleur et plaisir sont le socle de la morale et de la loi. On peut retrouver des racines bien plus anciennes à ces principes. On attribue ainsi à Hippocrate le principe connu sous sa forme latine *primum non nocere* : avant tout ne pas faire mal, resté central dans la déontologie médicale contemporaine. La douleur est devenue un sujet de préoccupation à part entière, pris en charge indépendamment de son étiologie (en témoignent les progrès de l'anesthésiologie et de la médecine palliative en tant que disciplines à part entière). C'est généralement à la quale de la douleur qu'il est fait référence. Torturer un « superspartiate »[71]

[70] « Nature has places mankind under the governance of two sovereign masters, pain et pleasure. It is for them alone to point out what we ought to do, as well as to determine what we shall do. On the one hand the standard of right and wrong, on the other the chain of causes and effects, are fastened to their throne. They govern us in all we do, in all we say, in all we think: every effort we make to throw off our subjection, will serve but to demonstrate and confirm it. In words a man may pretend to abjure their empire: but in reality he will remain subject, and assumes it for the foundation of that system, the object of which is to rear the fabric of felicity by the hands of reason and law. » (Bentham, 2017, p.11)

[71] Hilary Putnam propose l'expérience de pensée du superspartiate pour illustrer l'insuffisance du comportementalisme. Ce dernier a acquis l'aptitude à rester parfaitement stoïque face à la douleur : aucune sortie comportementale ne vient trahir sa souffrance. Cependant, l'existence de cette dernière montre que le stimulus douloureux vient changer l'état dispositionnel de ce dernier, bien que ces sorties comportementales restent indiscernables. (Putnam, 1968) Pour la suite du développement, supposons que nous avons affaire à un superspartiate fonctionnel, qui n'aurait donc même pas cette distinction dispositionnelle. S'il y a une différence chez celui-ci entre l'expérience de la douleur et son absence, elle ne peut pas survenir logiquement sur la physique.

(Putnam, 1968) nous semble moralement répréhensible quand bien même il ne présenterait aucune des caractéristiques fonctionnelles de la douleur. Ce point est à relativiser puisque pour Hilary Putnam, le superspartiate algique est dans un état mental différent que son homologue ataraxique, mais nous pourrions pousser l'argument jusqu'à lui soustraire cette disposition. Nous avons par ailleurs tendance à disconvenir de la gravité morale de provoquer des réactions de douleur à une entité à laquelle nous n'attribuons pas de qualia, quand bien même elle aurait des réactions comportementales et des changements d'état interne associés à la souffrance : une amibe, un robot, un personnage de jeu vidéo, etc.[72] C'est ainsi que la sentience, définie comme l'aptitude à vivre des expériences conscientes, semble fondamentale pour reconnaitre à certaines entités et non à d'autres une importance morale. Dès lors, les conséquences de l'éliminativisme paraissent vertigineuses. Il semble anéantir les racines de nos considérations morales les plus profondes.

Toutefois, rappelons qu'il est également possible de définir douleur et plaisir en termes fonctionnels, comme dispositions à produire des comportements de fuite ou de recherche, à apprendre des conséquences de nos comportements antérieurs selon le modèle du conditionnement opérant ou à modifier nos états internes fonctionnels. Richard Dawkins propose, en accord avec une définition de la vie fondée sur la reproduction (cf. page 8), une histoire de la biologie débutant avec de simples « réplicateurs » : un assemblage de molécules complexes apparues dans le hasard de la soupe primordiale et ayant pour propriété particulière de transformer d'autres molécules en copies d'elles-mêmes (Dawkins, 2003). La complexification des organismes (associée à l'augmentation de leur valeur sélective par d'accidentelles erreurs de réplications), devenus capables de comportements simples, a généré un point de vue sur le monde « pouvant être grossièrement partitionné en [situations] favorables, défavorables et neutres. » (« from which the world's events can be roughly partitioned into the favorable, the unfavorable, and the neutral », Dennett, 1993, p. 174). Ainsi, Dennett suggère que l'origine non arbitraire des notions de bien et de mal se situe dans l'intérêt (définitionnel) de ces réplicateurs à éviter les situations défavorables et rechercher les situations favorables à leur réplication.

[72] L'attribution de qualia à chacun de ces exemples reste possible. Chalmers l'accepterait certainement en regard de la loi de cohérence. Néanmoins, peu de gens défendraient qu'il soit immoral de « faire mal » à un personnage de jeu vidéo.

Ce que nous avons présenté comme la peur du noir, la peur de l'absence et la peur du vide, liées au sentiment de perdre respectivement la richesse de notre vie expérientielle, ce qui fait notre identité personnelle et le socle de nos principes moraux, a de fait des répercussions sur notre évaluation de la crédibilité de la théorie éliminativiste. Néanmoins, comme le rappelle Hume, « il n'y a pas de méthode de raisonnement plus blâmable que de tenter de réfuter une hypothèse par le danger de ses conséquences pour la morale » (Hume, 2006, p. 151). Nous soutenons qu'il en est de même pour ses conséquences existentielles et esthétiques. Cependant, ces angoisses peuvent être des pistes nous orientant vers une appréhension psychologique de la conjonction entre l'absence d'expérience consciente (0) et notre certitude dans leur existence (4'). Cette certitude s'apparenterait, selon cette perspective, à une tromperie de soi motivée par l'angoisse associée à ses répercussions.

CONCLUSION : LA BOUSSOLE, LE MICROSCOPE ET LE MIROIR

Au cours de ce travail, nous avons proposé de caractériser la certitude afin d'explorer ce jugement porté sur l'existence des qualia. Il nous est apparu que la certitude, au même titre que le doute, pouvait être analysée en termes d'émotion épistémique. Nous avons en effet soutenu que le doute et la certitude partageaient avec le reste du registre des émotions les propriétés d'ajuster l'esprit sur le monde, de se rapporter à quelque objet, de porter une valence et de s'accompagner de modifications physiologiques (ou ressenties comme telles). Par analogie avec les autres émotions, nous avons également proposé qu'ils dussent être le fruit d'une histoire évolutive, autrement dit le résultat des deux forces de l'évolution : dérive génétique et sélection naturelle. Dès lors, nous avons supposé que le doute et la certitude étaient porteurs d'une valeur épistémique : un seuil de déclenchement réglé pour optimiser la valeur sélective de l'individu. C'est ainsi qu'elles se sont avérées efficaces dans des situations susceptibles d'affecter la survie et la reproduction des gènes égoïstes dont les individus émotifs étaient les véhicules (Dawkins, *op. cit.*)[73]. À l'inverse, elles n'ont jamais eu vocation directe à nous guider vers une représentation fidèle du monde. Leur dérèglement, particulièrement visible dans le cas des TOC ou de la confabulation, nous affecte tous lorsqu'il s'agit de douter ou d'être certains de propositions éloignées des intérêts desdits gènes. Cela ne signifie pas qu'ils n'aient plus aucune valeur dans ces contextes : vivre avec des TOC ou confabuler n'est pas sans répercussions ; et se représenter le monde de manière appropriée représente un avantage certain.[74] Cependant, le réglage est trop imparfait en regard du niveau de précision requis par l'examen philosophique (ce que nous allons appeler le microscope épistémologique). L'indubitabilité de l'existence des qualia est un argument significatif en faveur de leur existence, mais il est loin d'être sans appel. Dans une seconde partie, nous avons en effet montré que d'autres explications à cette indubitabilité pouvaient être envisagées. Elles ne constituent bien évidemment pas des preuves *contre* l'existence des qualia, mais elles minimisent la portée de notre certitude dans la justification de notre croyance en la réalité de ces dernières. Cette

[73] Richard Dawkins défend une thèse selon laquelle ce n'est pas à l'individu qu'il faut attribuer une valeur sélective, mais aux gènes qu'il porte. Si les intérêts de l'individu et de ses gènes convergent dans de nombreux cas, ce n'est pas systématique. (Dawkins, 2003)

[74] Comme le soulignent McKay et Dennett, le dérèglement d'une fonction d'évaluation dans un cas pathologique ne préjuge pas de sa vocation à représenter fidèlement ou non l'objet de son évaluation dans les cas non pathologiques. Comme nous l'avons vu, il semble de plus que dans l'immense majorité des cas la fidélité au réel soit utile (McKay et Dennett, 2009). Ainsi les émotions épistémiques auraient une vocation *indirecte* à nous orienter vers une représentation exacte du réel.

remise en cause ne serait pas si attractive si elle ne permettait pas de résoudre avec une élégante simplicité un bon nombre de problèmes posés par le postulat initial de l'existence des qualia. D'une part, elle rend caduque la nécessité d'un appareil conceptuel complexe censé rendre compte du rapport entre qualia et jugements phénoménaux. D'autre part, elle met fin aux complications théoriques invoquées pour sauver Mary de sa prison épistémique, édifier un pont entre les deux berges d'un putatif gouffre explicatif ou encore postuler une indicible phénoménologie de l'écholocation. Néanmoins, si l'on souhaite défendre l'importance d'une théorie gardant une place pour les qualia, nous pouvons nous interroger sur les prédictions qu'elle nous permet d'énoncer (nous allons parler de boussole expérientielle susceptible de nous orienter vers les bonnes prédictions).

L'expérience consciente : une boussole finement réglée ?

Une théorie est supposée générer des prédictions, nécessairement vérifiables et si possible exactes. Peut-être le monde est-il peuplé d'ectoplasmes n'ayant absolument aucune interaction avec la matière : on ne démontrera jamais qu'ils n'existent pas. Néanmoins, le principe de parcimonie nous enjoint à en douter, sous peine de voir le monde surpeuplé de toutes les entités éthérées que nous sommes en mesure d'imaginer et davantage encore. L'un de ces fantômes pose néanmoins un problème particulier. Tandis que nous n'avons aucun mal à douter de l'existence d'un vaste bestiaire immatériel et inopérant, l'existence des qualia semble quant à elle indubitable. Pourtant, les qualia s'apparentent aux autres chimères : elles n'interagissent pas avec la matière et leur existence parait irréfutable. Mais la certitude persiste.

Des mondes qualitativement différents

Avec l'argument de la possibilité d'un monde zombie, Chalmers avance la possibilité d'une pluralité de mondes concevables qui seraient identiques sur le plan physique, mais différents sur le plan des expériences conscientes. Une façon de distinguer ces différents mondes est de les ordonner en fonction de leur densité de qualia. Ainsi, le monde zombie correspond à l'absence de qualia et à l'inverse le monde maximalement qualitatif en foisonne autant qu'il est concevable (si tant est qu'il y ait conceptuellement une limite haute, dans le cas

contraire considérons qu'il s'agit d'un monde *très* qualitatif). Envisageons ce qu'il se passe dans certains de ces mondes.

Le monde zombie. Ce monde est totalement dépourvu de qualia. Les entités et propriétés postulées par la physique y sont présentes et il est habité, au moins sur une planète, par une population de zombies dont certains sont des zombies-philosophes qui débattent de l'existence des qualia. Pour certains d'entre eux, cette existence est indubitable et ils vont jusqu'à proposer des lois qui régissent le lien entre les qualia et des propriétés fonctionnelles de leurs esprits. C'est le monde dans lequel un matérialiste éliminativiste affirmera que nous nous trouvons.

Le monde du dualiste solipsiste. Ce monde est similaire en tout point au monde zombie à une exception près : l'un de ses habitants n'est pas un zombie, mais fait l'expérience de qualia. Par hasard, ce dernier est un philosophe d'obédience dualiste solipsiste : il est certain d'avoir une expérience conscience, mais nie qu'autrui en dispose. Dans *ce* monde, il a raison. Cependant, il a des analogues dans tous les autres mondes envisagés qui ont une croyance fausse de même contenu.

Le monde légiféré. Dans ce monde, les qualia existent et elles sont strictement corrélées aux propriétés fonctionnelles des esprits, selon les lois de cohérence et d'invariance organisationnelle. Ce monde contient également une population dans laquelle se trouvent les contreparties des zombies-philosophes du monde zombie. Elles se posent les mêmes questions, avancent les mêmes arguments et écrivent les mêmes livres. Néanmoins dans ce monde, comme dans les suivants, le matérialisme éliminativiste est faux.

Le monde du sixième sens.[75] Les habitants de ce monde partagent eux aussi toutes les propriétés physiques de leurs analogues des deux mondes précédents et ont donc les mêmes doutes et les mêmes jugements. Ici, les qualia existent également, mais elles ne respectent pas strictement la loi de cohérence : un groupe de qualia supplémentaire est expérimenté par les habitants, mais ces qualia ont la particularité d'être strictement dénuées de corrélat fonctionnel. De ce fait, personne n'en parle, ne porte de jugement dessus et ne les mentionne dans des livres ou des discussions (si ce n'est dans le cadre de débats philosophiques et de façon purement

[75] Les humains ont évidemment bien plus que cinq sens. Outre la vue, l'ouïe, le toucher, l'odorat et le goût, il y a la proprioception, l'équilibroception, la nociception, le prurit, la sensation associée au besoin d'uriner, etc. Par ailleurs, les cinq sens traditionnels peuvent être décomposés en différentes modalités, le toucher réunissant par exemple le tact fin, le tact grossier, la thermoception, etc. Par ailleurs, le *sixième sens* du monde du sixième sens est une quale flottante qui ne peut être assimilée à un *sens* au sens fonctionnel du terme.

hypothétique). Elles ont la caractéristique d'être strictement épiphénoménales et dénuées de corrélats physiques.

Le monde très qualitatif. Dans cet univers, les qualia fourmillent. Il y en a bien plus que ce que prédit la loi de cohérence. Cependant, puisque ce monde respecte lui aussi les lois de la physique et notamment le principe de fermeture causale, là encore les habitants n'en parlent pas. Ils ne discutent que des qualia qui sont corrélées à leurs états mentaux. On y trouve bien, comme dans chacun des mondes suscités, quelques habitants qui suggèrent que davantage de qualia que celles qui sont corrélées aux états mentaux fonctionnels peuvent potentiellement exister, mais sans vraiment y croire.

Ces cinq mondes sont physiquement indiscernables, mais diffèrent selon la loi qui lie les qualia aux propriétés fonctionnelles (appelée *loi de cohérence* dans le monde légiféré)[76]. Puisque le démon-zombie (cf. note 10) est incapable de se repérer d'un monde à l'autre, étant aveugle à tout ce qui ne survient pas sur la physique, il semble qu'affirmer nous situer dans tel ou tel monde revient à se servir d'un outil auquel il n'a pas accès : la boussole expérientielle, réglée par les lois de cohérence et d'invariance.

Le réglage fin de la boussole expérientielle

Selon Chalmers, nous nous situons dans ce que nous avons appelé le *monde légiféré*.[77] Nous avons, nous aussi, l'intuition de nous situer dans cet univers-là. C'est celui vers lequel notre boussole expérientielle semble pointer : nos états mentaux fonctionnellement conscients seraient corrélés à nos expériences conscientes selon les principes énoncés par Chalmers. Cependant, force est de constater que nous n'avons aucune raison, hormis cette intuition, de stipuler que nous nous trouvons dans le monde légiféré plutôt que dans n'importe lequel des autres mondes possibles dont nous n'avons cité que quatre exemples. Cette intuition, partagée par tous les analogues des autres mondes, reste difficile à considérer autrement que sur un plan fonctionnel. Elle correspond à la boussole expérientielle et pour que la direction qu'elle indique

[76] Nous pourrions également proposer des variantes du *monde légiféré* (respectant la loi de cohérence) qui se différencierait selon la loi d'invariance organisationnelle : le *monde des qualia inversées*, le *monde des qualia s'estompant* ou encore le *monde des qualia dansantes* (Chalmers, 1996, p. 253, 263 et 266). Le démon-zombie reste incapable de les différencier.

[77] Rappelons toutefois que dans chacun des autres mondes, des alter ego de Chalmers tiennent le même discours. De même, des copies identiques de cet essai existent dans chaque monde et les contreparties de nos lecteurs en sont au même point de leur lecture.

soit la bonne, il semble nécessaire qu'elle soit finement réglée avec la loi de cohérence. Or, comment pourrions-nous expliquer ce réglage fin ? Il y a généralement trois principes pour expliquer de tels paramétrages :

L'intentionnalité. Une pendule est finement réglée, car l'artisan qui l'a conçue avait l'intention qu'elle le soit. La moindre altération de sa structure aura de grandes conséquences sur son fonctionnement, mais l'horloger a fait de son mieux pour qu'elle remplisse sa fonction aussi bien que possible. C'est une explication téléologique, puisque c'est l'usage final de l'objet qui détermine son réglage. Certains philosophes défendent une thèse selon laquelle seule l'explication intentionnelle pourrait expliquer le réglage fin des constantes de la physique, nécessaires à ce que notre univers puisse accueillir la vie (Swinburne, 2009).

Le hasard et la sélection. Un corps humain est finement réglé, si bien qu'à l'instar de la montre, la moindre modification est susceptible d'en entrainer la mort ou la non-viabilité. L'immense majorité des mutations non neutres intéressant les cellules germinales entrainent la mort du gamète, de l'œuf ou de l'embryon. Une minorité de ces mutations n'empêche pas la fécondation et la naissance, mais en son sein, la plus grande partie a des conséquences désastreuses sur la santé du nouveau-né. Enfin, une part infinitésimale de ces mutations peut apporter un avantage sélectif. La théorie de l'évolution permet d'expliquer la formation d'organismes complexes par les seuls faits du hasard (mutations sporadiques et dérive génétique) et de la sélection naturelle. Ainsi, parmi tous les mutants qui échappent aux accidents à l'origine de la dérive, la minuscule partie qui bénéficie d'un avantage sélectif sera favorisée. Or, parfois cette mutation peut être à l'origine d'un nouveau degré de complexité pour l'organisme expliquant l'apparition de structures sophistiquées et finement réglées.

Le principe anthropique. La planète Terre semble finement réglée pour accueillir la vie telle qu'on la connait. Il aurait suffi de quelques variations dans sa composition chimique, dans la distance à son étoile ou dans sa masse pour la rendre invivable (à l'image de ses sœurs Vénus et Mars, jusqu'à preuve du contraire). Il est possible d'expliquer cette improbable conjonction de propriétés sans faire appel au principe d'intentionnalité ou aux mécanismes de l'évolution : si notre planète n'avait pas possédé toutes les caractéristiques propices à notre existence, nous ne serions pas là pour en parler. Cela suppose néanmoins qu'un très grand nombre de planètes existent pour que, de temps en temps, par hasard, l'une d'elles soit compatible avec l'émergence de la vie. Dès lors une explication basée sur un point de vue *a posteriori* du réglage fin de cette planète est suffisante pour expliquer cette propriété.

Qu'en est-il du putatif réglage fin de la boussole expérientielle ? L'évolution ne semble pas être une explication envisageable, à moins d'expliquer à quel moment et de quelle manière le hasard et la sélection auraient agi. L'intentionnalité supposerait l'existence d'une intelligence créatrice qui aurait volontairement créé le monde en y incorporant les principes chalmersiens. Si cette explication n'est pas disqualifiée d'emblée, elle suppose d'ajouter encore des présupposés ontologiques à une théorie qui s'est déjà beaucoup alourdie de présupposés législatifs. L'explication anthropique supposerait quant à elle que parmi un grand nombre de mondes possibles, nous nous trouvons dans celui où il y a des qualia finement réglées parce que si nous n'y étions pas nous n'en aurions pas. Cependant, la possibilité d'autres mondes dont les habitants présentent des jugements similaires annule la spécificité éventuelle du nôtre : les habitants de ces différents mondes font un appel similaire au principe anthropique pour expliquer, à tort, le réglage fin de ces deux lois dans leur monde.

Force est d'admettre qu'aucun des trois principes ne fournirait une explication convenable à un réglage fin des lois chalmersiennes constituant la boussole expérientielle. En revanche, ils sont tous trois utiles à expliquer l'apparition de la certitude en ces dernières. En effet, l'émotion épistémique requiert :

- Une planète susceptible d'accueillir la vie. Son existence signifie l'occurrence d'une conjonction improbable, mais cette dernière est rendue possible par un nombre incommensurable de planètes dans l'univers et tout autant de combinaisons existantes. *Principe anthropique.*
- L'apparition sur cette planète d'organismes suffisamment complexes pour générer de la pensée abstraite. *Hasard et sélection.*
- L'élaboration par ces organismes de théories explicatives à propos de leur perception du monde et de l'existence. *Intentionnalité.*

La possibilité de mondes physiquement identiques, mais qui diffèrent sur le plan des qualia, loin de constituer un argument en faveur de l'existence de celles-ci dans le nôtre, suggère au contraire qu'il n'y a aucune raison d'imaginer que nous nous situons dans l'un de ces mondes plutôt qu'un autre. La boussole expérientielle ne nous est d'aucune utilité. Cependant, s'il fallait parier sur le monde dans lequel nous nous trouvons, l'un d'entre eux semblerait tirer son épingle du jeu : le monde zombie. Celui-ci a l'avantage de la simplicité et semble être le moins arbitraire. Peut-être vivons-nous dans le monde du dualiste solipsiste, dans le monde légiféré,

dans le monde du sixième sens ou dans le monde très qualitatif, au même titre qu'il est possible que nous vivions dans toutes sortes de mondes peuplés de fantômes causalement inefficaces. Mais jusqu'à preuve du contraire, il est plus légitime de considérer que nous vivons dans un monde sans qualia ni âmes désincarnées ; un monde qui néanmoins remplit les conditions d'apparition *de la certitude* de l'expérience consciente.

La certitude en microscopie épistémologique

Si les émotions épistémiques ont une pertinence macroscopique indéniable, guidant nos actions et nos jugements quotidiens, l'ont-elles également dans le contexte d'une investigation philosophique minutieuse ? Les arguments proposés dans la première partie nous amènent à considérer doute et certitude comme des émotions épistémiques et seulement comme telles. Or, lorsqu'elle se focalise sur les expériences conscientes, la certitude ainsi caractérisée peut sembler ne plus suffire. L'expérience consciente constitue-t-elle une exception ? Nous allons scruter cette certitude de plus près, en disséquant l'une des perceptions auxquelles il semble de prime abord le plus difficile de nier une propriété phénoménale : la douleur.

Qu'est-ce que cela fait d'avoir mal ?

Indubitablement, il y a quelque chose que cela fait d'avoir mal : mal. Dépassons cette tautologie et tentons de faire l'inventaire de tout ce que cela fait.[78] Nous pouvons procéder par la méthode jamesienne de soustraction (cf. page 27). Dépouillons la douleur de :

- *Ses effets comportementaux directs*. La douleur ne nous motive plus à nous éloigner de sa source, nous ne souhaitons plus d'anesthésie chez le dentiste, ne voyons dans le retrait de la main posée sur la plaque chauffante rien d'autre qu'un intérêt rationnel à préserver notre intégrité physique.

[78] Nous nous intéressons ici à l'ensemble de *ce que cela fait*, ce qui dépasse donc la stricte notion nagelienne de *what it is like*. En français, l'emploi du verbe *faire*, traduction coutumière de *be like*, invite à un inventaire intégrant les propriétés fonctionnelles, les propriétés qui *font*.

- *Ses effets cognitifs et émotionnels*. Nous sommes indifférents à la douleur, nous n'avons plus tendance à nous juger souffrants et n'avons plus peur à la perspective de souffrir.
- *Ses effets aversifs.* La douleur ne nous donne plus de leçons. Peu importe le nombre de fois que nous posons la main sur la plaque de cuisson, la probabilité d'apparition de ce comportement ne diminue pas. Il n'y a plus d'apprentissage opérant.
- *Ses effets physiologiques*. La douleur ne génère plus de catécholamines susceptibles d'accélérer le rythme cardiaque, d'augmenter la tension, d'activer le système nerveux sympathique en vue d'organiser une disposition physiologique au combat ou à la fuite.

Une fois ces soustractions effectuées, que cela fait-il encore d'avoir mal ? D'aucuns affirment qu'il reste la quale de la douleur. Ainsi, le superspartiate fonctionnel, chez qui même le changement dispositionnel lié à la douleur est amendé, est supposé avoir l'expérience consciente de la douleur. Une expérience consciente bien particulière, puisqu'exclusivement phénoménale et n'ayant aucun aspect psychophysiologique.

À la question « qu'est-ce que cela fait d'avoir mal », il y a donc deux réponses envisageables. (Φ) *cela fait réagir, avoir peur, se plaindre, apprendre à éviter, changer notre état corporel* et *avoir une quale de douleur*. (Ψ) *cela fait réagir, avoir peur, se plaindre, apprendre à éviter, changer notre état corporel.* Point. En ajoutant la quale de la douleur à cette conjonction, nous sommes contraints d'admettre qu'il est concevable que cette quale puisse rester présente après soustraction de tous les autres conjoints. Si nous acceptons cela, une sinistre conclusion émerge : il est possible qu'en ce moment même nous fissions l'expérience consciente d'une atroce douleur dénuée de tout corrélat physique, cognitif et comportemental. Condamnés à ne rien dire et ne rien faire.

Nous soutenons qu'une fois ôté l'ensemble des strates fonctionnelles de la douleur, la quale qui semble y être rattachée est moins évidente à discerner. Dès lors, l'argument de Kripke (*op. cit.*, p. 133-144) selon lequel on ne peut identifier la douleur avec un ensemble de propriétés physiques perd de sa portée : certes, la conjonction [douleur] ∧ (Φ) est contingente, mais [douleur] ∧ (Ψ) parait nécessaire.[79] Kripke avance que l'identification est impossible, car

[79] Le contenu des conjonctions (Φ) et (Ψ) est bien évidemment discutable. Nous pensons que pour Kripke (Φ) ne contiendrait en fait que la « qualité phénoménologique immédiate » (Kripke, 1982, p. 141), c'est-à-dire la quale. Quant à ce qui doit être inclus dans la conjonction (Ψ), il s'agit d'une question empirique portant sur ce qui caractérise fonctionnellement la douleur (cf. note 8).

« être dans la même situation épistémique que si l'on avait mal, c'est avoir mal » (*ibid.*, p. 140). Il affirme que « la référence de « douleur » n'est pas fixée par une propriété accidentelle de la douleur, mais par sa propriété d'être une douleur, par sa qualité phénoménologique immédiate. » (*ibid.* p. 141). Son rejet de l'identification douleur = (Φ) est donc, en ce sens, parfaitement légitime. En niant ce qui est supposé distinguer (Φ) et (Ψ), c'est-à-dire l'existence d'une quale-douleur, il devient toutefois possible d'accepter l'identification douleur = (Ψ), dans la mesure où tous les termes de (Ψ) surviennent logiquement sur la physique.

Une certitude pas comme les autres ?

Un cartésien estime que s'il n'y a qu'une seule et unique certitude que nous pouvons entretenir, c'est celle de l'existence nos expériences conscientes. Un argument qui va dans ce sens est celui de la découverte diachronique du monde. Certes, nous pouvons décrire l'air lointain d'une trompette en termes fonctionnels : curiosité, disposition à énoncer « entendez-vous cette mélodie ? », tendance à le juger plaisant ou déplaisant, etc. Cependant, on pourrait avancer que toutes ces dispositions sont acquises après l'expérience initiale du son. Ce serait l'expérience de la musique qui engendrerait le panel de dispositions cognitivo-comportementales associées. La certitude de l'expérience consciente aurait cette particularité d'être primitive, par opposition à d'autres certitudes potentiellement tout aussi ancrées (« Neptune est la huitième planète du système solaire », « les célibataires ne sont pas mariés », « il n'existe pas d'entiers strictement positifs x, y et z, tels que $x^n + y^n = z^n$ pour $n > 2$ », etc.), mais acquises fonctionnellement, par transmission culturelle ou découverte empirique.[80]

Cependant, l'argument du primat de l'expérience fait-il de l'expérience consciente au sens de Chalmers une certitude toute particulière ? D'une part, quelle preuve concrète avons-nous de cette primauté chronologique ? Nos souvenirs ne remontent pas aussi loin que nos premières perceptions, si bien que nous n'avons pas la possibilité de confirmer que l'expérience précède l'usage. Bien évidemment, aucun nouveau-né n'a la bienséance de témoigner de ses premières impressions et si un jour l'un d'entre eux le faisait, ce serait nécessairement via le langage qui présuppose lui-même un usage fonctionnel de la perception. D'autre part, en admettant une primauté de la perception, cela suppose-t-il que cette perception soit qualitative ?

[80] Le cas d'éventuelles connaissances innées ne pose pas de problème particulier. Il s'agit également d'une acquisition fonctionnelle, lors de la neurogenèse et selon un plan issu de la phylogenèse. C'est pourquoi les connaissances innées ne sont pas *primitives* au sens où nous l'entendons ici.

Il est tout à fait possible de répondre au problème de Molyneux à la façon de John Locke sans supposer l'existence de qualia. Les *impressions* humiennes peuvent être décrites fonctionnellement. La priorité de la perception ne serait, par définition, pas différente entre les mondes zombie, du dualiste solipsiste, légiféré, du sixième sens ou très qualitatif, et donc n'implique pas que la perception soit accompagnée de qualia. Par ailleurs, nous avons proposé différentes modalités d'acquisition de la certitude de l'expérience consciente ne requérant pas que cette dernière soit antérieure à toute autre ou ait une quelconque spécificité par rapport à toute autre certitude.

L'éliminativisme est-il incroyable ?

Répondre à cette question requiert de préciser ce que nous entendons par incroyable. Si nous prenons incroyable comme synonyme d'inconcevable, la réponse est très aisée : l'éliminativisme n'est pas incroyable. Chalmers n'a par exemple pas de difficulté avec la concevabilité d'un monde zombie. Si nous entendons par croire tenir pour vrai, là encore aucune difficulté : ce n'est peut-être pas la croyance la plus partagée, mais d'aucuns tiennent l'éliminativisme pour vrai si bien qu'en ce sens, s'il n'est pas cru, il est au moins croyable. Si nous envisageons la croyance au sens de la certitude, il est plus difficile de nier que l'éliminativisme a quelque chose d'incroyable. Considérons une dernière forme de croyance :

> « 144. L'enfant apprend à croire une quantité de choses. C'est-à-dire qu'il apprend à agir selon ces croyances. Petit à petit se forme un système de ce qu'il croit et, dans ce système, certaines choses sont inébranlablement fixées et d'autres sont plus ou moins mobiles. Ce qui est solidement fixé ne l'est pas parce qu'intrinsèquement manifeste ou évident, mais parce que tenu immobile par tout ce qui l'entoure. » (Wittgenstein, 2006).

Ainsi, notre certitude, croyance inébranlable en l'existence de l'expérience consciente n'est-elle pas tenue immobile par ces mêmes implications esthétiques, existentielles et morales dont la remise en cause nous terrifie ? Cette importance des implications susmentionnées peut provenir elle-même de ses implications pratiques : nous attribuons à nos qualia, peut-être à tort, nos goûts musicaux, notre personnalité, notre éthique, nos valeurs et c'est conformément à ces attributions que nous agissons au quotidien. Le contextualisme épistémique suppose que nous puissions tenir une proposition pour certaine dans un contexte donné et douteuse dans un autre (Lewis, 1996). Ainsi, il serait cohérent d'avoir foi en la réalité de nos expériences conscientes

dans le contexte quotidien où ses implications le requièrent, tout en en doutant à la lumière du microscope épistémologique.

La question titre de cet essai était « l'expérience consciente est-elle indubitable ? ». En définitive, nous maintenons une réponse affirmative : en quelque sorte, l'existence d'une expérience consciente, des qualia, de la conscience phénoménale est certaine. Nos heuristiques affectives l'excluent du champ du doute. Néanmoins, c'est la portée de cette indubitabilité que nous souhaitons relativiser. Dans un contexte pragmatique et adaptatif, il a tout son sens et, effectivement, il semble plus utile de faire avec et de tenir compte de nos expériences conscientes lorsque nous assistons à un concert, prenons des antalgiques ou évaluons les implications morales de telle ou telle action. Mais comme nous le rappellent les sciences de l'évolution, utilité et fidélité ne vont pas toujours de pair et optimisation et perfection ne sont pas synonymes. Cette indubitabilité, tout opportune et bien réglée qu'elle soit, n'est pas la preuve définitive de la réalité des expériences conscientes.

L'envers d'un dilemme

Subjectivement, l'existence des qualia parait incontestable. Les expériences perceptives semblent être notre seule lucarne sur le monde, si bien que de l'observation d'un ciel étoilé à l'élaboration de la relativité générale, tous nos savoirs, toutes nos croyances, toutes nos idées semblent médiés par le mélange d'impressions que laisse le monde dans notre esprit. Il se peut, bien entendu, que certaines de nos idées soient innées, présentes depuis toujours en nous. Néanmoins, ne sont-elles pas, à leur manière, perçues au même titre que les traces mnésiques par lesquelles nos impressions s'unissent ? Un regard sur le monde extérieur, un regard sur le monde intérieur, mais toujours un regard subjectif. En définitive, cette primauté de l'expérience peut s'étendre jusqu'à l'idéalisme. De *toujours la subjectivité* à *rien d'autre que la subjectivité*. Postuler un monde extérieur parait alors être un cas de *surexplication* qui n'est pas sans rappeler celui que nous avions invoqué au sujet des qualia. Une symétrie frappante semble alors apparaître. Ce sont les lois de la nature qui ont permis aux étoiles de se former pour habiller le ciel. Ce sont le vertige et l'émerveillement face à ce spectacle céleste qui nous ont amenés à nous interroger sur la nature de la réalité et former des théories telles que celle de la relativité. Inversement, c'est donc le postulat d'un monde intérieur qui semble devenir superflu.

Ces perspectives antagonistes ne se rejoignent finalement que dans leur incapacité à répondre aux deux questions les plus difficiles de la philosophie : *pourquoi y a-t-il quelque chose ? pourquoi cela ?* Idéalistes et matérialistes n'auraient-ils pas, en définitive, deux visions en miroir d'un seul et même monde, dont l'explication pose un *problème* très *difficile* ?

RÉFÉRENCES BIBLIOGRAPHIQUES

American Psychiatric Association, 2015, *DSM-5 : manuel diagnostique et statistique des troubles mentaux*, Issy-les-Moulineaux, Elsevier Masson

Arango-Muños, S., 2014, « The Nature of epistemic feelings », *Philosophical psychology*, vol. 27, n° 2, doi : 10.1080/09515089.2012.732002

Bentham, J., 2017, *An introduction to the principles of morals and legislation*, s.l., Pantianos Classics

Block, N., 1995, « Concepts of consciousness », dans : Chalmers, D., 2002, *Philosophy of mind: classical and contemporary readings*, Oxford, Oxford University Press, p. 206-218

Bortolotti, L., 2010, *Delusions and other irrational beliefs*, Oxford, Oxford University Press

Canguilhem, G., 2013, *Le Normal et le pathologique*, Paris, Presses universitaires de France

Cannon, W., 1915, *Bodily changes in pain, hunger, fear, and rage : an account of recent researches into the function of emotional excitement*, New York, D. Appleton and company, numérisé par the Internet Archive, 2010, url : archive.org/details/bodilychangesinp00cann (consulté le 1.6.2019)

Carruthers, P., 2012, « Language in cognition », dans : Margolis, E., Samuels, R., Stich, S., *The Oxford handbook of philosophy of cognitive science*, Oxford, Oxford University Press, p. 382-401

Chalmers, D., 1996, *The Conscious mind: in search of a fundamental theory*, Oxford, Oxford University Press

Cottraux, J., 2016, « Trouble obsessionnel compulsif », *EMC – Psychiatrie*, vol. 13, n° 4, doi : 10.1016/S0246-1072(16)75242-4

Cottraux, J., 2017, *Les Psychothérapies cognitives et comportementales*, Issy-les-Moulineaux, Elsevier Masson

Damasio, A., 2010, *L'Erreur de Descartes : la raison des émotions*, Paris, Odile Jacob

David, P., Samadi, S., 2011, *La Théorie de l'évolution : une logique pour la biologie*, Paris, Flammarion

Davidson, D., 1986, « Deception and division », dans : Davidson, D., 2004, *Problems of rationality*, Oxford, Oxford University Press

Dawkins, R., 2003, *Le Gène égoïste*, Paris, Odile Jacob

de Sousa, 2016, « Epistemic feelings », dans : Brun, G., Doğuoğlu, U., Kuenzle, D., *Epistemology and emotions*, Abington (Royaume-Uni), Routledge

Dennett, D., 1988, « Quining qualia », dans : Chalmers, D., 2002, *Philosophy of mind :classical and contemporary readings*, Oxford, Oxford University Press, p. 226-246

Dennett, D., 1992, « The Self as the center of narrative gravity », dans : Kessel, F., Cole, P., Johnson, D., 2010, *Self and consciousness: multiple perspectives*, New York, Psychology Press

Dennett, D., 1993, *Consciousness explained*, Londres, Penguin

Dennett, D., 2012, *De beaux rêves : obstacles philosophiques à une science de la conscience*, Paris, Gallimard

Descartes, R., 2009, *Méditations métaphysiques*, Paris, Flammarion

Descartes, R., 2016, *Discours de la méthode*, Paris, Flammarion

Deonna, J., Teroni, F., 2008, *Qu'est-ce qu'une émotion ?*, Paris, Vrin

Deweese-Boyd, I., 2017, « Self-deception », dans : Zalta, E., *The Stanford encyclopedia of philosophy*, éd. automne 2017, url : plato.stanford.edu/archives/fall2017/entries/self-deception/ (consulté le 1.6.2019)

Docter, P., 2015, *Inside out*, Emeryville (Californie), Pixar Animation Studios/Walt Disney Pictures, format pellicule 35 mm

Egan, G., 2009, *Lama*, dans : Egan, G., *Océanique*, Saint-Mammès, le Bélial'

Ekman, P., 1999, « Basic emotions », dans : Dalgleish, T., Power, M., *Handbook of cognition and emotion*, s.l., Wiley-Blackwell

Frankish, K., 2016, « Illusionism as a theory of consciousness », *Journal of consciousness studies*, vol. 23, n° 11-12, p. 11-39

Freud, S., 2017, *Cinq psychanalyses*, Paris, Payot & Rivages

Goldie, P., 2002, « Emotions, feelings and intentionality », *Phenomenology and the cognitive science*, vol. 1, n° 3, doi : 10.1023/A:1021306500055

Hirstein, W., 2006, *Brain fiction: self-deception and the riddle of confabulation*, Cambridge (Massachusetts), The MIT Press

Hawking, S., 2007, *Une brève histoire du temps : du big bang aux trous noirs*, Paris, J'ai lu

Hume, D., 1995, *L'Entendement : traité de la nature humaine : Livre I et appendice*, Paris, Flammarion

Hume, D., 2006, *Enquête sur l'entendement humain,* Paris, Flammarion

Jackson, F., 1982, « Epiphenomenal qualia », dans : Chalmers, D., 2002, *Philosophy of mind :classical and contemporary readings*, Oxford, Oxford University Press, p. 273-280.

James, W., 1950, *The Principles of psychology*, vol. 2, New York, Dover Publication

Kahneman, D., 2016, *Système 1 – système 2 : les deux vitesses de la pensée*, Paris, Flammarion

Kelvin, 1901, « Nineteenth century clouds over the dynamical theory of heat and light », *Philosophical Magazine*, série 6, vol. 2, n° 7, doi : 10.1080/14786440109462664

Kepecs, A., 2013, « The Uncertainty of it all », *Nature neuroscience*, vol. 16, n° 6, doi : 10.1038/nn.3416

Kriegel, U., 2015, *The Varieties of consciousness*, New York, Oxford University Press

Kripke, S., 1982, *La Logique des noms propres*, Paris, Les Éditions de Minuit

Laplanche, J., Pontalis, J.-B., 2007, *Vocabulaire de la psychanalyse*, Paris, Presses universitaires de France

Levine, J., 1983, « Materialism and qualia », dans : Chalmers, D., 2002, *Philosophy of mind: classical and contemporary readings*, Oxford, Oxford University Press, p. 354-361

Lewis, D., 1996, « Elusive knowledge », *Australian journal of philosophy*, vol. 74, n° 4, p. 549-567

Martini, H., Dehmas, M., Paille, F., 2017, « Complications neurologiques de la consommation d'alcool », *EMC – Neurologie*, vol. 14, n° 3, doi : 10.1016/S0246-0378(16)39455-6

Mazin, C., 2019, *Chernobyl*, HBO/Sky UK, format numérique

McKay, R., Dennett, D., 2009, « The Evolution of misbelief », *Behavioral and brain sciences*, n° 32, doi : 10.1017/S0140525X09990975

Mele, A., 2001, *Self-deception unmasked*, Princeton (New Jersey), Princeton University Press

Nagel, T., 1974, « What is it like to be a bat? », dans : Chalmers, D., 2002, *Philosophy of mind: classical and contemporary readings*, Oxford, Oxford University Press, p. 219-226.

Parfit, D., 1987, *Reasons and persons*, Oxford, Oxford University Press

Parfit, D., 2004, « Why anything? Why this? », dans : Crane, T., Farkas, K., *Metaphysics: a guide and anthology*, Oxford, Oxford University Press

Pink Floyd, 1973, *Breathe*, auteur : Roger Waters, compositeurs : Roger Waters, Richard Wright, David Gilmour, dans : Pink Floyd, *The Dark side of the moon*, Londres, Harvest, format 33 tours.

Putnam, H., 1968, « The Nature of Mental States », dans : Chalmers, D., 2002, *Philosophy of mind: classical and contemporary readings*, Oxford, Oxford University Press, p. 73-79

Rowling, J., 2000, *Harry Potter et la coupe de feu*, Paris, Gallimard

Sacks, O., 1988, *L'Homme qui prenait sa femme pour un chapeau et autres récits cliniques*, Paris, Éd. du Seuil

Shoemaker, S., 1982, « The Inverted spectrum », *Journal of philosophy*, vol. 79, Juillet, p. 357-381, doi : 10.2307/2026213

Svenson, O., 1981, « Are we all less risky and more skillful than our fellow drivers? », *Acta Psychologica*, 47, p. 143-148

Swinburne, R., 2009, *Y a-t-il un dieu ?*, Paris, Ithaque

Tooby, J., Cosmides, L., 2008, « The Evolutionary psychology of the emotions and their relationship to internal regulatory variables », dans : Lewis, M., Haviland-Jones, M., Barrett, L., *Handbook of emotions*, New York, Guilford

Walker, C., 1991, « Delusions: what did Jaspers really say? », *British journal of psychiatry*, 159

Wittgenstein, L., 2006, *De la certitude*, Paris, Gallimard