Université de Liège – Séminaire de phénoménologie 2011-2012 :
L'expérience subjective : approches phénoménologiques de la conscience phénoménale

Introduction au problème de la conscience phénoménale[1]

Approches métaphysiques et épistémologiques

Arnaud Dewalque, ULg

Mon intention est de proposer une introduction au problème de la conscience phénoménale, que l'on associe traditionnellement au problème des qualia. Je me pencherai essentiellement sur le contexte dans lequel ce problème a vu le jour. Ce qui est très caractéristique, à cet égard, c'est que le problème de la conscience phénoménale a immédiatement donné lieu à de multiples prises de positions métaphysiques (relatives à ce qui existe, à l'« ameublement du monde ») et épistémologiques (relatives à ce que nous pouvons connaître). Cette prédominance des approches métaphysiques et épistémologiques est indissociable de la manière même de poser le problème de la conscience phénoménale, si bien que ce problème est lui-même apparu prioritairement comme un problème métaphysique ou épistémologique. Si l'étude de la conscience phénoménale se présente effectivement comme un « défi » à relever, c'est, d'abord et avant tout, parce qu'elle semble difficilement conciliable avec une attitude métaphysique dominante en philosophie de l'esprit : le physicalisme, soit la thèse selon laquelle le monde entier est composé de choses physiques dont le comportement est intégralement explicable, en principe, par les sciences physiques (physique, chimie, biologie, etc.). Comme on va le voir, prendre au sérieux le problème de la conscience phénoménale, cela semble précisément impliquer l'adoption d'une position critique à l'égard du physicalisme.

J'aborderai donc ici le débat sur son terrain métaphysique-épistémologique initial, et je focaliserai plus particulièrement mon attention sur l'un des arguments contre le physicalisme les plus discutés en philosophie de l'esprit. Cet argument est traditionnellement baptisé « argument de la connaissance »

[1] Ceci est le texte remanié de l'intervention faite lors de la séance du 6 octobre 2011 (« L'argument de la connaissance. Introduction au problème de la conscience phénoménale »), Université de Liège, Unité de recherche Phénoménologie*s*. Pour toutes questions et remarques : a.dewalque@ulg.ac.be.

(*Knowledge Argument*). Il repose sur l'idée suivante : la connaissance physique n'est pas toute la connaissance que l'on peut avoir à propos du monde ; il y a un genre de connaissance qui ne peut pas être obtenu par investigation des structures physiques, mais qui est intrinsèquement lié au point de vue « en première personne ». L'argument, sous sa forme canonique, a été formulé par Frank Jackson dans un article intitulé « Epiphenomenal Qualia » (*The Philosophical Quarterly*, 1982). Mais l'idée générale apparaît déjà, moyennant certains divergences, chez des auteurs plus anciens, notamment chez Thomas Nagel (« What is it like to be a Bat? », 1974).

Je retracerai le contexte général dans lequel est apparu le problème de la conscience phénoménale, j'indiquerai succinctement les principales attitudes métaphysiques et épistémologiques adoptées face à ce problème, puis je reconstruirai l'argument de la connaissance et je terminerai en mentionnant quelques objections qui lui ont été adressées. On verra que, si l'on peut douter que l'argument de la connaissance soit réellement efficace pour réfuter le physicalisme, il y a en revanche de bonnes raisons de penser qu'il met en évidence, de façon pressante, notre besoin d'une théorie plus objective et plus rigoureuse de l'expérience subjective, une théorie qui, surtout, ne soit pas empêtrée d'emblée dans des alternatives métaphysiques et épistémologiques. Je suggérerai, dans une séance ultérieure[2], que la phénoménologie a précisément l'avantage de présenter une certaine *neutralité* face aux alternatives métaphysiques et épistémologiques.

1. Arrière-plan historique : les paradigmes dominants en philosophie de l'esprit

L'étude de la conscience constitue un axe de recherche *bien défini* depuis un peu plus d'une vingtaine d'années. Cet axe de recherche rassemble une série de disciplines comme neuroscience, psychologie, philosophie, intelligence artificielle et linguistique. Toutes ces disciplines constituent le champ très large de ce que l'on appelle aujourd'hui les *Consciousness Studies*. Si cette appellation est relativement jeune, le problème de la conscience lui-même possède un riche passé historique. On fait traditionnellement remonter son origine à Descartes**,** qui a posé un dualisme métaphysique entre corps et esprit. Le dualisme cartésien est à l'origine de ce qui a formé le problème central de la philosophie de l'esprit entre – approximativement – 1950 et 1980 : le *Mind-Body Problem*, le problème de l'articulation de l'esprit et du corps. Ce problème est motivé avant tout par la volonté de produire une théorie *scientifique* de l'esprit, c'est-à-dire une théorie qui soit (a) fondée sur des données observables

[2] Cf. la séance à venir le 17/11/11 : « La thèse de l'intentionnalité phénoménale ».

et (b) en accord avec la conception du monde comme ensemble d'objets naturels étudiés par les sciences physiques, chimiques et biologiques.

Cette approche a donné lieux à trois grands paradigmes qui ont dominé la philosophie de l'esprit jusque dans les années 1980 : le béhaviorisme logique, le matérialisme éliminativiste et le fonctionalisme. Ces trois paradigmes ont contribué à déterminer le contexte dans lequel est apparu le problème de la conscience phénoménale. Il me faut donc en dire un mot rapidement[3].

1 / En réaction à la méthode introspectioniste défendue à la fin du XIX^e^ siècle (notamment par Brentano et son école, puis le mouvement phénoménologique), les béhavioristes John Watson et Burrhus Frederic Skinner soutiennent que la psychologie s'intéresse seulement au comportement de l'homme et des animaux, qui peut être observé de l'extérieur. Même si l'on peut douter que l'ancienne psychologie soit *intégralement* introspectionniste, comme le suggère Watson, la position défendue dans son article de 1913 est très claire : l'étude du comportement permet de traiter la psychologie comme une « branche expérimentale purement objective de la science naturelle », dans laquelle la méthode introspectioniste n'a aucun rôle essentiel à jouer (Watson 1913, p. 158). Le comportement est principalement constitué par les réactions des individus face aux *stimuli* produits par le contact de l'environnement physique avec ses récepteurs sensoriels. Cette approche méthodologique a une contrepartie philosophique : le béhaviorisme logique (associé notamment au nom de Carl Hempel, mais dont on trouve aussi des traces chez Willard van Orman Quine). Par « état mental », le béhavioriste logique entend un comportement-type qui peut être observé et théorisé. Par exemple, éprouver une douleur au contact d'une flamme, c'est manifester certains comportements-types comme retirer vivement sa main, crier, etc.

2 / À partir des années 1950, le développement des sciences cognitives va entraîner l'apparition d'un autre paradigme, le *matérialisme éliminativiste*, dont les principaux représentants sont Paul et Patricia Churchland. Selon eux, la psychologie populaire a toutes les apparences d'une théorie, mais c'est une théorie fausse, non scientifique, analogue à l'alchimie, c'est-à-dire sans support observationnel solide. Par conséquent, elle doit être *éliminée* et remplacée par les neurosciences une fois que celles-ci seront pleinement développées. Dans ette optique, un état mental comme la douleur ne serait rien de plus que l'état dans lequel se trouve un organisme à un moment *t*, état qui est intégralement déterminable par la constitution neurophysiologique de cet organisme au moment *t*. Bref, l'approche matérialiste revient à adhérer à ce que l'on appelle usuellement la *théorie de l'identité esprit-cerveau*. Cette théorie est celle que

[3] Pour un aperçu plus détaillé, je renvoie à l'excellente introduction de Denis Fisette et Pierre Poirrier dans leurs deux volumes sur la *Philosophie de l'esprit*, t. I et II, Paris, Vrin, 2002 et 2003.

défend notamment David Armstrong dans *A Materialist Theory of the Mind* : « L'esprit n'est rien d'autre que le cerveau » (*the mind is nothing but the brain* ; Armstrong 1968, p. 1).

L'éliminativisme représente une option radicale qui a été très controversée. Par la suite, certains auteurs ont avancé un argument célèbre destiné à montrer que ni le béhaviorisme ni l'éliminativisme ne pouvaient fournir une explication satisfaisante de l'esprit. Cet argument est connu sous le nom d'*argument de la réalisation multiple*. Il repose sur l'idée suivante : un même état mental peut être réalisé dans plusieurs individus qui n'ont ni le même comportement ni la même constitution neurophysiologique. Or, cela ne serait pas logiquement concevable si le concept d' « état mental » était réductible au concept de « comportement-type » ou à celui de « constitution neurophysiologique ». La formulation la plus fameuse de cet argument est due à David Lewis. Dans un article intitulé « Douleur de fou et douleur de martien » (1980), Lewis soutient que la douleur est irréductible à ses manifestations comportementales, car on peut concevoir que le fou éprouve de la souffrance sans l'exprimer par des comportements-types : au lieu de hurler de douleur, sa souffrance pourrait se traduire, par exemple, par la tendance à croiser les jambes et claquer des doigts, ou à se concentrer sur théorèmes mathématiques plutôt que sur le monde qui l'entoure, etc. Le paradigme béhavioriste échoue à rendre compte de la douleur d'un fou, puisqu'il identifie la douleur à ses manifestations comportementales. Il est donc logiquement concevable de dissocier état mental et comportement, ce qui suffit à montrer l'insuffisance du concept béhavioriste d' « état mental ». Mais le paradigme éliminativiste n'est pas en meilleure posture. Pour les éliminativistes, un état mental doit être conçu comme l'état d'un organisme déterminé par sa constitution neurophysiologique. Or, là encore, il est logiquement possible de concevoir deux individus possédant une constitution physiologique différente (par exemple un Martien et un être humain) mais vivant le même état mental (par exemple éprouvant de la douleur) : le martien pourrait même être dépourvu de cerveau et posséder à la place un organe fait de cavités susceptibles de gonfler, mais il est plausible de maintenir qu'il pourrait néanmoins faire l'expérience de la douleur. Bref, Lewis soutient que ni le béhaviorisme ni le matérialisme ne peuvent servir de base à une théorie de l'esprit parfaitement satisfaisante, parce qu'ils échouent à penser un état mental qui serait, *au moins en un certain sens*, commun à un individu sain d'esprit et à un fou, ou à un terrien et à un martien :

1.

If I want a credible theory of mind, I need a theory that does not deny the possibility of mad pain. I needn't mind conceding that perhaps the madman is not in pain in *quite* the same sense that the rest of us are, but	Si je veux disposer d'une théorie de l'esprit crédible, j'ai besoin d'une théorie qui ne nie pas que la douleur du fou soit possible. Je n'ai pas d'objection à concéder que peut-être le fou n'a pas mal *tout à fait* au même sens

there had better be some straightforward sense in which he and we are both in pain. […] A credible theory of mind had better not deny the possibility of Martian pain. I needn't mind concedind that perhaps the Martian is not in pain in *quite* the same sense that we Earthlings are, but there had better be some straightforward sense in which he and we are both in pain (Lewis 1980, p. 216-217).

que nous, mais il doit y avoir un sens passablement clair où lui et nous avons mal […]. Une théorie de l'esprit crédible ferait mieux de ne pas nier que la douleur du Martien soit possible. Je n'ai pas d'objection à concéder que peut-être le Martien n'a pas mal *tout à fait* au même sens que nous Terriens, mais il doit y avoir un sens passablement clair où lui et nous avons mal (trad. fr., p. 290).

3 / Ces difficultés ont conduit une bonne partie des philosophes de l'esprit à écarter les précédentes approches au profit d'un troisième paradigme, le *fonctionalisme*, dont les principaux représentants sont Hilary Putnam (pour sa version computationnelle), David Lewis (pour sa version causaliste), et Jerry Fodor (pour sa version représentationnelle). Les fonctionalistes considèrent que l'esprit est une manifestation de niveau supérieur qui repose sur un soubassement neuronal de niveau inférieur. Ce soubassement est suffisant mais non nécessaire pour produire un état mental : métaphysiquement, il n'est pas nécessaire d'admetttre un autre genre d'entité que le soubassement physique, mais ce soubassement physique est contingent, ce qui veut dire qu'il aurait pu être différent. Peu importe que le support physique soit un système nerveux dans un corps, un cerveau dans une cuve (selon l'exemple célèbre de Putnam) ou un circuit informatique. Pour les fonctionnalistes, un état mental se définit comme un rôle. On voit ce qui distingue cette position du matérialisme, d'après lequel un état mental est, non pas un rôle (indifférent au support physique), mais justement « l'état physique ou biologique qui implémente ce rôle » (*cf.* Block 2003).

Le fonctionnalisme est une approche qui a pu sembler séduisante. Mais il y a un aspect que le fonctionalisme ne parvient pas à prendre en compte : cet aspect, c'est précisément ce que l'on appelle la conscience phénoménale. Mais d'abord, que faut-il entendre, au juste, par conscience phenoménale ?

2. La conscience phénoménale : clarifications préalables

On parle de conscience phénoménale pour désigner l'expérience telle qu'elle est vécue par le sujet (l'expérience subjective), ou encore les aspects qualitatifs de l'expérience (plus simplement : les *qualia*). Depuis l'article fameux publié par le philosophe américain Thomas Nagel en 1974, intitulé « What is it like to be a bat ? », on parle aussi couramment de « l'effet que cela fait » de vivre telle ou telle expérience. On retient en effet la définition (ou l'équivalence) suivante :

2.

An organism has conscious mental states if and only if there is something that it is like to *be* that organism – something it is like *for* the organism. We may call this the subjective character of experience (Nagel 1974, p. 436).	Un organisme a des états mentaux conscients si cela lui fait un certain effet d'*être* cet organisme – un certain effet *pour* l'organisme. Nous pouvons appeler cela le caractère subjectif de l'expérience (trad. fr., p. 392).

Pour y voir plus clair, il est d'usage de distinguer le concept de conscience phénoménale d'autres concepts de conscience. Dans l'ensemble, il semble qu'il faille admettre au moins trois concepts distincts : 1. le concept psychologique de conscience (le fait d'être conscient *de* qqch) ; 2. la conscience phénoménale (*l'effet que cela fait* d'être conscient de qqch) ; 3. la conscience réflexive ou introspection (le fait d'opérer un acte de réflexion qui prend pour objet l'acte irréfléchi). On peut donc se représenter les choses de la façon suivante :

(1)		(2)		(3)
Conscience intentionnelle	≠	Conscience phénoménale	≠	Conscience réflexive

Il est important de préciser d'emblée que ces distinctions sont conceptuelles, et qu'il y a plusieurs manières de penser l'articulation réelle de ces différents types de conscience. La question de savoir s'ils peuvent exister isolément, ou bien au contraire s'ils s'impliquent les uns les autres, est une question de grande ampleur que je laisserai ici de côté, du moins provisoirement. (Je reviendrai, dans un exposé ultérieur, sur la thèse de l'intentionnalité phénoménale, d'après laquelle il y a une étroite connexion entre conscience intentionnelle et conscience phénoménale).

Pour nombre de philosophes de l'esprit, un support physique (un ordinateur) peut gérer de l'information, opérer des raisonnements, etc., sans être doté d'une conscience phénoménale. Une machine de Turing peut produire de l'information à propos de quelque chose, un zombi philosophique peut se trouver dans un état mental intentionnel, sans savoir l'effet que cela fait d'être dans cet état mental. Dans le cas contraire, on dira que l'état mental intentionnel s'enrichit d'une propriété particulière : il possède un aspect qualitatif qui est l'objet de la conscience phénoménale. C'est en ce sens qu'un état mental est dit conscient *ssi* il y a un certain effet que cela fait d'être dans cet état mental (un état mental n'est pas conscient si l'individu qui se trouve dans cet état mental ne sait pas l'effet que cela fait d'être précisément dans *cet* état mental et non dans tel autre). Mais d'un autre côté, la conscience phénoménale ne doit pas être confondue avec la conscience réflexive : pour ceux qui rejettent les théories d'ordre supérieur, la conscience réflexive n'est pas intrinsèquement liée à la conscience phénoménale, mais constitue encore un stade ultérieur de la vie de la

conscience. Ils soutiennent que la conscience phénoménale peut être pré-réflexive (pour employer un terme plutôt sartrien) au sens où, pour vivre un état mental à propos de quelque chose (conscience au sens n°1) et éprouver l'effet que cela fait d'être dans cet état mental (conscience au sens n°2), je n'ai nullement besoin d'effectuer un acte de réflexion qui aurait la conscience pour objet. Encore une fois, l'articulation des trois niveaux est controversée et je ne développerai pas cet aspect des choses pour l'instant.

Que retenir de tout cela ? Aujourd'hui, lorsqu'on parle du problème de la conscience, on a essentiellement en vue la conscience au sens (2), la conscience phénoménale. Plus exactement, on distingue habituellement, avec David Chalmers (1996), deux types de problèmes liés à la conscience : (a) les « problèmes faciles » (*easy problems*), comme rendre compte de l'état de veille (*awakeness*) et de sa différence par rapport au sommeil ; rendre compte de le la *reportability,* c'est-à-dire de notre capacité à produire des compte rendus de nos propres états mentaux, etc. Tous ces « problèmes faciles » sont des problèmes susceptibles d'être résolus d'un point de vue computationnel et neurologique. La conscience psychique ou intentionnelle, par exemple, peut en effet être expliquée au moyen d'une approche en termes de rôle causal ou de rôle fonctionnel. Par contre, la conscience phénoménale est une dimension de l'esprit que les principaux paradigmes explicatifs qui dominaient la philosophie de l'esprit jusqu'en 1980 ont échoué à intégrer. Pour simplifier, on pourrait donc dire que, ce qui a motivé le développement des *consciousness studies*, c'est l'échec des précédents paradigmes et leur incapacité à produire une théorie complète et unifiée de l'esprit. Lorsqu'on parle du « problème difficile » de la conscience, on a en vue la conscience phénoménale, ce que Chalmer préfère appeler l'expérience.

Le problème peut être formulé comme suit : comment se fait-il que, dans un univers composé de choses physiques, il y ait quelque chose comme de la conscience phénoménale ? Pourquoi est-ce que, en plus de penser, nous vivons une expérience subjective ? Cette question semble miner le projet de penser l'articulation du corps et de l'esprit dans un cadre physicaliste. C'est pourqui Thomas Nagel a pu dire que, sans la conscience, le *Mind-Body Problem* est « beaucoup moins intéressant », mais avec la conscience, « il paraît sans espoir de solution » (Nagel 1974 ; trad. fr., p. 392). Je reprends le passage que je citais il y a un instant :

2.

An organism has conscious mental states if and only if there is something that it is like to *be* that organism – something it is like *for* the organism. We may call this the subjective character of 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 termes of the causal role of experiences in relation to typical human behavior – for similar reasons (Nagel 1974, p. 436).

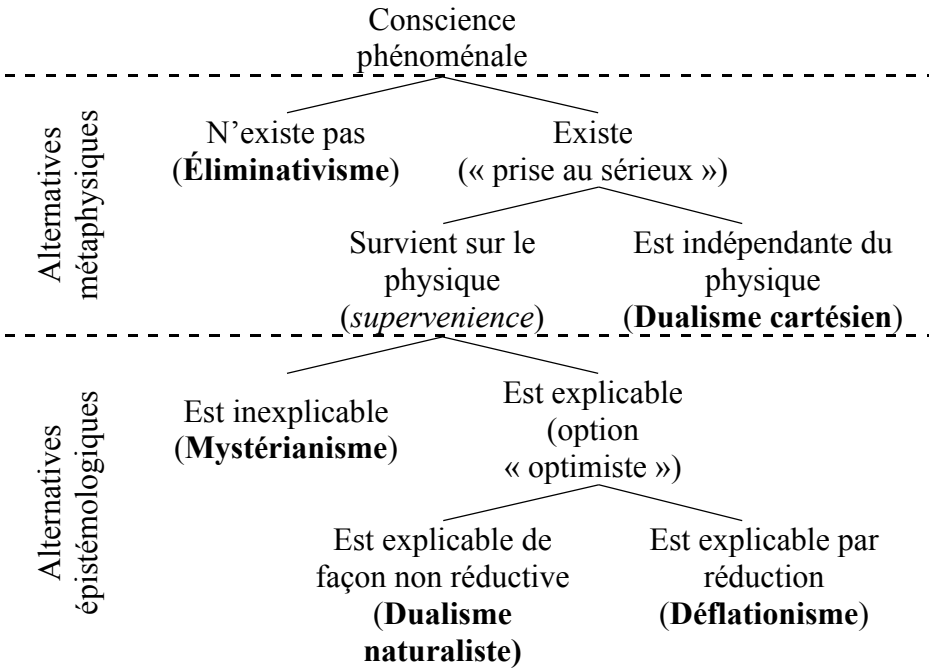
Un organisme a des états mentaux conscients si cela lui fait un certain effet d'*être* cet organisme – un certain effet *pour* l'organisme. Nous pouvons appeler cela le caractère subjectif de l'expérience. Il n'est saisi par aucune des analyses habituelles réductrices du mental conçues précédemment, car toutes sont logiquement compatibles avec son absence. Il n'est pas analysable en termes d'un quelconque système explicatif d'états fonctionnels, ou d'états intentionnels, car ceux-ci pourraient être attribués à des robots ou à des automates se comportant comme des personnes bien qu'ils n'aient aucune expérience. Il n'est pas analysable en termes de role causal d'expériences en relation à un comportement humain typique – pour des raisons similaires (trad. fr., p. 392-393).

Fondamentalement, le problème est que le fonctionnalisme est vrai même d'un être dépourvu de conscience. Un automate ou un « zombi philosophique » (un être humain en tous points semblable aux êtres humains normaux, mais dépourvu de conscience phénoménale, de ressenti qualitatif) peuvent très bien se trouver dans le même état fonctionnel qu'un humain, sans être conscients. En s'appuyant sur la théorie fonctionaliste, il est donc impossible de faire la différence entre organisme conscient et organisme qui fonctionne sans être doté de conscience. En un mot : le fonctionalisme, qui était pourtant le paradigme explicatif le plus prometteur, ne saisit pas la conscience phénoménale.

Le problème ne concerne toutefois pas le fonctionnalisme en tant que paradigme particulier. Il rejaillit en réalité sur l'ensemble de la conception physicaliste du monde qui sous-tend tout le programme d'explication du mental à partir du physique. Sans la conscience phénoménale, le monde semble explicable intégralement grâce aux sciences physiques. Mais la conscience phénoménale est un élément qui résiste à toute explication physicaliste ou naturaliste. Faut-il voir là le signe que le physicalisme est faux ? La reconnaissance de la conscience phénoménale implique-t-elle la reconnaissance d'un genre d'entités (les *qualia*) *non physiques* ou *irréductibles à des états physiques* (chimiques, biologiques, neurophysiologiques, etc.) ? La quasi-totalité des discussions concernent exclusivement le problème, à la fois métaphysique et épistémologique, d'harmoniser l'existence des *qualia* avec la vision physicaliste et naturaliste de l'univers que fournissent les sciences physiques.

3. Suite ; aperçu de quelques stratégies « classiques »

Ce qui complique passablement les choses, c'est qu'un grand nombre de stratégies sont possibles. L'existence de stratégies variées parfois radicalement divergentes explique qu'il soit si difficile, à première vue, de s'orienter dans les discussions. Il sera bon, dès lors, de disposer d'une vue d'ensemble des différentes positions qui sont apparues jusqu'ici. Je pense que l'on peut se représenter la situation, *grosso modo*, de la façon suivante :



Ce qui est très caractéristique, c'est que ces alternatives se situent d'emblée sur un terrain métaphysique et épistémologique. Le dualisme cartésien apparaît aujourd'hui comme une théorie métaphysique obsolète, car irréconciliable avec la vision du monde des sciences physiques. La conscience phénoménale serait, selon cette option, équivalente à un phénomène « magique » ou surnaturel, inexplicable par les sciences physiques parce que formant un niveau de réalité distinct et indépendant de la réalité physique. Aussi le dualisme métaphysique de style cartésien est-il l'option la plus contestée aujourd'hui. D'autre part, le poids de la conception physicaliste du monde pèse tellement sur l'analyse de la conscience phénoménale que certains auteurs vont jusqu'à nier purement et simplement qu'il existe dans le monde quelque chose comme des *qualia*. Plutôt que de chercher une solution au problème de la conscience phénoménale, ils suppriment ainsi l'énoncé même du problème, prônant du même coup un éliminativisme ou un matérialisme extrême.

Les options les plus intéressantes – ou, en tout cas, celles qui ont le plus de succès à l'heure actuelle – sont celles qui se situent entre ces deux extrêmes.

Elles se caractérisent par une certaine reconnaissance du dualisme, mais contestent que ce dualisme soit un dualisme de type cartésien. Une telle approche repose essentiellement sur le concept de « survenance » (*supervenience*) : un état ou une propriété qui *survient sur le physique* – comme, mettons, la propriété d'élasticité[4] – est un état que nous pouvons expliquer sans admettre d'autres ingrédients, dans notre ontologie, que des entités physiques (*e.g.* les molécules qui composent le corps élastique), mais que l'on ne peut pas interpréter en termes purement physicalistes comme le voudraient les éliminativistes. Il se pourrait donc que, métaphysiquement parlant, les états mentaux surviennent sur le physique. Dans ce cas de figure, deux nouvelles alternatives se présentent à nous : soit les états mentaux survenants sont inexplicables, soit ils sont explicables ; s'ils sont explicables, ils peuvent admettre une explication non réductive ou une explication réductive.

La thèse de l'inexplicabilité des états mentaux est défendue, notamment par Collin McGinn. Dans *The Character of Mind* (1997), McGinn commence par prendre acte de la « discontinuité » entre la conscience et les phénomènes physiques. Il remarque que, lorsque nous prenons au sérieux l'existence de la conscience, nous n'avons pas seulement affaire à une nouvelle configuration d'éléments physiques, mais à quelque chose d'entièrement nouveau. La question, dès lors, est la suivante : « Comment un tel phénomène unique peut-il se produire à partir de la matière et quelle sorte d'entité est le cerveau pour pouvoir générer ce phénomène ? » (McGinn 1997, p. 41). McGinn écarte quatre options qu'il considère comme les quatre options « standard » : (a) l'éliminativisme, qui considère que parler de conscience phénoménale, c'est embrasser une conception pré-scientifique du monde ou adhérer à une croyance analogue à la croyance aux fantômes, aux sorcières et aux ectoplasmes ; (b) le dualisme métaphysique *à la Descartes*, qui transforme la conscience en quelque chose de magique et qui fait de son apparition quelque chose de similaire au miracle de la transformation de l'eau en vin ; (c) la thèse de l'irréductibilité, qui consiste à envisager la conscience comme une dimension primitive de l'expérience comparable à l'espace et au temps en physique (qui ne peuvent être réduits à des phénomènes plus primitifs) ; (d) le déflationisme, selon lequel la conscience existe, n'est pas un miracle et peut être expliquée en termes plus familiers, par exemple en termes béhavioristes ou fonctionalistes. Selon cette dernière conception, il serait possible de « domestiquer le phénomène », c'est-à-dire de le ramener à quelque chose de moins mystérieux.

McGinn estime pour sa part qu'aucune de ces options n'est pleinement satisfaisante ; mais il estime aussi qu'elles n'épuisent pas toutes les options théoriques et qu'il est possible de dégager une cinquième option, qui a sa préférence. Cette cinquième option consiste à soutenir la conscience est un

[4] Je remercie Denis Seron pour cet exemple.

mystère que l'esprit humain n'a pas le pouvoir d'expliquer. McGinn insiste sur le fait que cette position est davantage épistémologique que métaphysique : dire que la conscience est un mystère n'est pas la même chose que dire qu'elle est un miracle. Dire que la conscience est un miracle, c'est formuler une position ontologique, traiter la conscience comme quelque chose de surnaturel. Or, pour McGinn, rien ne nous autorise à tirer une telle conclusion. Tout ce que l'on peut conclure de l'échec des paradigmes explicatifs standard, c'est une thèse épistémologique (non métaphysique) : dire que la conscience est un mystère, c'est précisément formuler une thèse *épistémologique*, c'est parler des capacités cognitives humaines. Le « mystérianisme » (l'expression a été forgée par Owen Flanagan en 1992) consiste à soutenir, contre l'éliminativisme, que la conscience existe, contre le dualisme métaphysique fort, qu'elle n'est pas surnaturelle, contre la thèse de l'irréductibilité, qu'elle a bel et bien une explication, et contre le déflationisme, que nous sommes incapables de fournir cette explication. McGinn défend donc l'idée que la conscience « nous est fermée sur le plan cognitif » (*is cognitively closed to us*), ce qui l'amène à proposer un « naturalisme transcendantal » selon lequel la conscience est un phénomène naturel mais que nous ne pouvons pas expliquer. Les conditions de possibilité d'une explication font défaut, car notre capacité à connaître est affectée par des « limitations conceptuelles » intrinsèques. En fin de compte, pour les mystérianistes, il n'y a donc pas de problème *philosophique* de la conscience, il n'y a pas de problème conceptuel, puisque la conscience échappe inévitablement à notre prise conceptuelle. La notion centrale, ici, est celle de « clôture cognitive » (*cognitive closure*) : « Essayer de forcer la connaissance là où elle ne peut l'être », écrit McGinn (1997, p. 48), cela ne peut engendrer que des « monstres intellectuels ». En fin de compte, le mystérianisme implique deux choses : d'une part, l'approche physique est incomplète, puisqu'il y a quelque chose dans le monde qu'elle ne peut expliquer ; d'autre part, la conscience a une structure « cachée » (*hidden*) accessible seulement dans l'introspection, que McGinn identifie au point de vue « en première personne ».

Dans l'ensemble, on peut sans doute considérer que l'option mystérianiste constitue une forme de pessimisme face au problème de la conscience phénoménale. Ce pessimisme est toutefois loin d'être partagé. D'autres auteurs, comme David Chalmers (*The Conscious Mind*, 1996), considèrent que le pessimisme de McGinn est prématuré. Pour Chalmers, il ne faut pas abandonner toute tentative d'explication, mais simplement abandonner toute tentative d'explication *réductive*, c'est-à-dire ne pas tenter d'expliquer la conscience à partir de quelque chose de plus simple. La conscience phénoménale, *i.e.* l'expérience subjective, est *fondamentale*. Il faut simplement jeter un « pont explicatif » (*explanatory bridge*) au-dessus de ce que l'on appelle couramment, depuis Levine (1983), le « fossé dans l'explication » (*explanatory gap*). La solution, estime Chalmers, consiste à ajouter quelque chose à notre ontologie. Si

tout, dans les théories physiques, est compatible avec l'absence de la conscience, les théories physiques permettent d'expliquer le monde mais elles ne permettent pas de faire la différence entre un monde où les processus physiques s'accompagnent d'expérience et un monde jumeau où l'expérience est absente. Pour ce faite, il est nécesaire d'ajouter un ingrédient supplémentaire, qui possède ses propres lois.

Concrètement, Chalmers propose la construction d'une théorie psycho-physique qui est une sorte de dualisme, mais un « dualisme innocent » (un « dualisme *naturaliste* »). Ce dualisme n'a pas besoin d'entrer en contradiction avec la vision scientifique du monde qui est suggérée par les sciences physiques. Il ne s'agit pas d'admettre le psychique comme une force mystérieuse ou magique, ni come une sorte d'élan vital qui agirait sur le monde physique en tant que principe causal : le monde physique peut être causalement clos sur lui-même (tout effet physique peut bien être le résultat d'une cause physique). Bref, le « psycho- » de « psychophysique » ne renvoie ici à rien de spirituel ou de mystique. Simplement, il existe des principes psycho-physiques, qui *connectent* les propriétés des processus physiques aux propriétés de l'expérience.

Pour terminer ce rapide tour d'horizon, je dirais un mot de la question suivante : où se situe, dans ce paysage, l'approche phénoménologique ? L'option la plus radicale et certainement la plus intéressante consiste à admettre que l'approche phénoménologique se situe en amont de toutes ces distinctions : en « retournant aux choses mêmes », la phénoménologie implique de mettre entre parenthèses toutes les thèses métaphysiques. C'est la conception défendue par Shaun Gallagher et Dan Zahavi dans *The Phenomenological Mind* (2008). Parce qu'elle ne s'empêtre ni dans des considérations métaphysiques ni dans des alternatives épistémologiques, l'approche phénoménologique pourrait bien représenter une approche particulièrement prometteuse pour traiter le problème de la conscience phénoménale. Comme je l'ai dit, je ne développerai toutefois pas ce point ici.

4. Ce qui pose problème avec les qualia (*excursus* historique)

Quelle que soit l'attitude que l'on adopte face au problème de la conscience phénoménale, ce problème est intrinsèquement lié à la conception physicaliste du monde. Plus simplement, on peut dire que, si la conscience phénoménale est bien un *problème*, c'est précisément parce qu'elle constitue un défi pour le physicalisme. Cela dit, la référence aux qualia n'a pas toujours eu une signification purement anti-physicaliste dans l'histoire récente de la philosophie. À vrai dire, un examen même sommaire de la littérature montre que l'on prête souvent aux qualia différentes propriétés. Avant d'en venir à l'argument de la connaissance développé par Frank Jackson – qui est, comme je l'ai dit, le

principal argument avancé contre le physicalisme –, je voudrais mettre en évidence quelques propriétés que l'on attribue couramment aux qualia. Cela sera l'occasion, en outre, de mentionner différents raisonnements qui ont parfois été considérés comme des « ancêtres » de l'argument de la connaissance.

On trouve une anticipation remarquable du problème de la conscience chez Emil Du Bois-Reymond (1818-1896), l'un des pionniers de la neurophysiologie. Dans sa célèbre conférence de Leipzig, « Sur les limites de la connaissance de la nature » (1872), Du Bois-Reymond soutient que la pensée au sens de Descartes – en ce compris les sensations primitives et es qualia (par exemple, l'effet que cela me fait de voir du rouge) – échappe au pouvoir explicatif des sciences de la nature et en constitue ainsi une *limite* infranchissable. L'argument invoqué est la totale hétérogénéité du psychique et du physique. Aucune combinaison d'éléments physiques ne saurait, par une mystérieuse alchimie, se transformer en état psychique. Du physique ne découle rien d'autre que du physique : « L'effet mécanique est absolument égal à la cause mécanique qui s'épuise à le produire » (Du Bois-Reymond, p. 42 ; trad. fr. modifiée, p. 343). De là résulte l'impossibilité de réduire les états mentaux à des états physiques. Or, cette impossibilité, insiste Du Bois-Reymond, ne tient pas à l'état actuel des sciences empiriques. Même dans l'hypothèse où nous disposerions de ce qu'il appelle une « connaissance astronomique » de l'homme, c'est-à-dire une connaissance absolument complète qui nous permettrait de calculer l'état passé et futur de la mécanique humaine avec autant de certitude que l'astronomie le fait pour les corps célestes, les phénomènes spirituels resteraient malgré tout « en dehors de la loi de causalité », et « cela suffit pour les rendre incompréhensibles » (*id.*).

3.

Ein aus irgend einem Grunde bewusstloses, z. B. ohne Traum schlafendes Gehirn enthielte, astronomisch durchschaut, kein Geheimnis mehr, und bei astronomischer Kenntnis auch des übrigen Körpers wäre so die ganze menschliche Maschine, mit ihrem Athmen, ihrem Herzschlag, ihrem Stoffwechsel, ihrer Wärme, u. s. f., bis auf das Wesen von Materie und Kraft, völlig entziffert. Der traumlos Schlafende ist begreiflich, wie die Welt, ehe es Bewusstsein gab. Wie aber mit der ersten Regung von Bewusstsein die Welt doppelt unbegreiflich ward, so wird es auch der Schläfer wieder mit dem ersten ihm dämmernden Traumbild. […] Damit ist die andere Grenze unseres Naturerkennens bezeichnet. Nicht minder als die erste ist sie eine unbedingte (Du Bois-

Un cerveau privé de conscience pour une raison quelconque, par exemple dormant sans rêver, n'aurait plus rien de caché pour qui le connaîtrait astronomiquement ; et en supposant de même le reste du corps astronomiquement connu, la machine humaine tout entière, avec sa respiration, les battements de son cœur, sa statique chimique, sa caloricité, serait de tout point comprise, hormis toujours bien entendu l'essence de la force et de la matière [= la première des deux limites admises par Du Bois-Reymond – AD]. L'homme qui dort sans rêver est compréhensible au même degré que le monde avant qu'il y eût la pensée. Tout comme avec la première sensation éprouvée, le monde devint doublement incompréhensible, ainsi le dormeur le redevient à la première lueur d'un

Reymond 1872, p. 28-29).

rêve. […] Voilà donc l'autre borne de notre philosophie naturelle. Elle n'est pas moins infranchissable que la première (trad. fr., p. 343).

À la question de savoir comment la conscience apparaît à partir des processus physiologiques ou physico-chimiques, Du Bois-Reymond répond par sa célèbre formule : *ignoramus*, nous l'ignorons, et *ignorabimus*, nous l'ignorerons à jamais (Du Bois-Reymond, p. 51 ; trad. fr., p. 345). Bref, exprimée dans les termes de la philosophie de l'esprit contemporaine, la thèse de Du Bois-Reymond est tout simplement que *l'esprit ne se laisse pas naturaliser*[5].

Un autre auteur connu pour avoir anticipé l'argument de la connaissance est Charlie Dunbar Broad (1887-1971). Dans son ouvrage pionnier, *The Mind and its Place in Nature* (1925), Broad propose l'expérience de pensée suivante : supposons qu'une théorie mécaniste de la chimie soit vraie, et supposons qu'un être – représenté sous les traits d'un archange – dispose de compétences mathématiques illimitées et soit en outre capable de percevoir la structure microscopique des atomes. Il aurait une connaissance physique parfaite (comme disait Du Bois-Reymond, une connaissance « astronomique ») de la composition chimique de diverses substances – mettons, de l'ammoniaque – et de leur comportement lorsqu'elles sont soumises à telle ou telle pression environnementale. Néanmoins, remarque Broad, il y a quelque chose que l'archange ne pourraît pas prédire : c'est *l'odeur* qu'aura une substance possédant la structure moléculaire et les propriétés chimiques de l'ammoniaque. Tout ce qu'il pourraît prédire, c'est le comportement physique de la membrane muqueuse qui nous permet de sentir et le comportement des nerfs olfactifs lorsqu'ils sont stimulés par de l'amoniaque. Mais il n'a aucun moyen de savoir que ces modifications du système nerveux s'accompagnent de l'apparition d'une odeur en général (sauf si on lui dit) ni quelle odeur particulière a l'amoniaque en particulier (sauf s'il l'a senti lui-même).

4.

[5] Cette position a donné lieu à une vaste controverse connue dans la littérature sous le nom d'*Ignorabimus-Streit* (« Querelle de l'*ignorabimus* »). Pour une présentation récente de cette querelle, voir Bayertz/Gerhard/Jaeschke (2007).

Take any ordinary statement, such as we find in chemistry books; *e.g.*, "Nitrogen and Hydrogen combine when an electric discharge is passed through a mixture of the two. The resulting compound contains three atoms of Hydrogen to one of Nitrogen; it is a gas readily soluble in water, and possessed of a pungent and characteristic smell". If the mechanistic theory be true the archangel could deduce from his knowledge of the microscopic structure of atoms all these facts but the last. He would know exactly what the microscopic structure of ammonia must be; but he would be totally unable to predict that a substance with this structure must smell as ammonia does when it gets into the human nose (Broad 1925, p. 71).

Prenez n'importe quelle assertion ordinaire, telle que nous la trouve dans un livre de chimie, par ex. : « L'azote et l'hydrogène se combinent lorsqu'un courant électrique traverse une solution qui les contient. Le composé qui en résulte contient trois atomes d'hydrogène pour un atome d'azote. C'est un gas soluble dans l'eau, et il possédait une odeur âcre et caractéristique ». Si la théorie mécaniste est vraie, l'archange pourrait déduire de sa connaissance de la structure microscopique des atomes tous ces faits à l'exception du dernier. Il saurait exactement quelle doit être la structure micriscopique de l'ammoniaque, mais il serait totalement incapable de prédire qu'une substance avec cette structure doit sentir comme sent l'ammoniaque lorsqu'il pénètre dans le nez humain.

Tout comme pour Du Bois-Reymond, l'existence de qualia, pour Broad, rend le monde incompréhensible, ce qui veut dire ici impossible à prédire. La reconnaissance de la conscience phénoménale va donc bien de pair, d'emblée, avec l'idée que les sciences physiques ne peuvent pas expliquer le monde en totalité. Il y a un « plus qualitatif » qui échappe à l'investigation physique, chimique et biologique.

La question qui se pose, dès lors, est de savoir que faire de cet excédent. Une attitude possible consiste à nier que les expériences décrites soient des connaissances. C'est l'attitude adoptée par le fondateur du Cercle de Vienne, Moritz Schlick. Dans les années 1910 et 1920, Schlick développe l'idée que le « contenu » ou la « qualité » d'une expérience subjective est inexprimable et donc inconnaissable : il peut seulement être « vécu » (*erlebt*) et non « connu » (*erkennt*). Héritant de la notion russellienne de « connaissance par contact direct » (*knowledge by acquaintance*), Schlick s'en démarque en effet immédiatement en refusant de considérer l'expérience vécue (le contact direct) comme un genre de connaissance. La seule connaissance, aux yeux de Schlick, est la connaissance par description (connaître, c'est substituer une expression co-référentielle à une autre). La conception de Schlick est instructive, car elle montre que l'on peut prendre en compte l'existence des qualia dans une optique censément compatible avec le physicalisme. Avoir des qualia, c'est vivre une expérience subjective, et ce n'est pas acquérir une connaissance de l'objet (contre l'intuitionisme, Schlick soutient que voir n'est pas connaître). On se trouve donc, ici, aux antipodes de l'argument de la connaissance, qui consiste précisément à faire de l'expérience vécue un genre de connaissance particulier :

5.

Wenn ich eine rote Fläche anschaue, so kann ich niemandem sagen, wie das Erlebnis des Rot beschaffen ist. Der Blindgeborene kann durch keine Beschreibung eine Vorstellung von dem Inhalt eines Farbenerlebnisses bekommen. Wer nie Lust gefühlt hätte, würde durch keine Erkenntnis davon unterrichtet werden können, was man erlebt, wenn man Lust erlebt. Und wer es einmal erlebt und dann vergessen hätte und nie wieder zu fühlen imstande wäre, dem könnten es auch etwaige eigene Aufzeichnungen niemals sagen. Und das Gleiche gilt, wie jeder sofort zugibt, von allen Qualitäten, die als Inhalte des Bewußtseinsstromes auftreten. Sie werden nur durch unmittelbares Erleben bekannt (Schlick 1926, p. 146 ; rééd. [1]1938, [2]1969, p. 183).

Quand je vois une surface rouge, je ne peux dire à personne en quoi consiste l'expérience vécue du rouge. Aucune description ne peut fournir à l'aveugle de naissance une représentation du contenu d'un vécu de couleur. À qui n'a jamais ressenti de plaisir, aucune connaissance n'enseignera jamais ce dont on a le vécu quand on a l'expérience vécue du plaisir. Et qui l'a une fois vécu, puis oublié sans être en mesure de le ressentir à nouveau, les notes que lui-même aura pu prendre ne le lui diront jamais. Il en va de même, tous l'accordent immédiatement, de toutes les qualités qui figurent comme contenu du courant de conscience. Elles ne sont connues qu'à travers une expérience vécue immédiate (trad. fr., p. 175).

La même thèse se retrouve explicitement dans le manifeste du Cercle de Vienne :

6.

Die subjektiv erlebten Qualitäten – die Röte, die Lust – sind als solche eben nur Erlebnise, nicht Erkenntnisse ; in die physikalische Optik geht nur da sein, was auch dem Blinden grundsätzlich verständlich ist ([Anonyme] 1929, p. 119-120).

Les qualités vécues subjectivement – le rouge ou le plaisir – sont en tant que telles seulement des expériences vécues, non des connaissances. Dans l'optique physique entre seulement ce que même un aveugle peut en principe comprendre (trad. fr., p. 115).

La thèse selon laquelle les qualia ne sont accessibles que dans l'expérience vécue immédiate, soit « en première personne », figurera encore au centre de l'article célèbre de Thomas Nagel, « What is it like to be a Bat ? » (1974). Toutefois, Nagel a donné une inflexion très particulière au problème des qualia. Contrairement à Schlick, Nagel soutient en effet que l'expérience vécue « en première personne » n'est pas *ipso facto* une expérience privée et incommunicable. L'argumentation générale de Nagel est bien connue : même si nous disposions d'une connaissance objective complète de la manière dont les chauve-souris se déplacent (à savoir par écholocation), nous ne saurions pas « l'effet que cela fait d'être une chauve souris ». Nous n'avons aucune idée de l'effet que ça fait de s'orienter par écholocation – car c'est quelque chose qui ne peut être expérimenté qu'en première personne, subjectivement, et l'expérience

subjective des chauves-souris est trop éloignée de la nôtre pour que nous puissions nous représenter l'effet que cela fait de se déplacer par écholocation.

Cela étant, au cours de son argumentation, Nagel met en évidence le point suivant : un individu qui apprendrait l'effet que cela fait de vivre telle ou telle expérience ne gagnerait pas seulement une information sur lui-même, mais sur les *autres* individus vivant la même expérience. Le fait que l'accès aux qualia soit conditionné par une expérience vécue en première personne ne signifie pas que les qualia soient privés : je n'ai pas l'expérience de *mes* propres qualia, mais bien l'expérience des qualia que ressent tout individu lorsqu'il est en train de vivre la même expérience. L'expérience subjective est donc, en un sens, source d'informations objectives, qui ne me concernent pas moi, mais aussi les autres individus. Il y a, si l'on veut, une *objectivité des qualia* :

7.

Whatever may be the status of facts about what it is like to be a human being, or a bat, or a Martian, these appear to be facts that embody a particular point of view. I am not adverting here to the alleged privacy of experience to its possessor. The point of view in question is not one accessible only to a single individual. Rather it is a *type*. It is often possible to take up a point of view other than one's own, so the comprehension of such facts is not limited to one's own case. There is a sense in which phenomenological facts are perfectly objective: one person can know or say of another what the quality of other's experience is. They are subjective, however, in the sense that even this objective ascription of experience is possible only for someone sufficiently similar to the object of ascription in the first person as well as in the third, so to speak (Nagel 1974, p. 441-442).

Quel que puisse être le statut de faits concernant l'effet que cela fait d'être un être humain, ou une chauve-souris, ou un Martien, ces faits semblent envelopper la présence d'un certain point de vue. Je ne fais pas allusion au caractère soi-disant privé de l'expérience pour celui qui la possède. Le point de vue en question n'en est pas un qui soit accessible seulement à un individu unique. C'est plutôt un *type*. Il est souvent possible d'envisager un autre point de vue que le sien propre, en sorte que la compréhension de faits de ce genre ne soit pas limitée à notre propre cas. En un certain sens les faits phénoménologiques sont parfaitement objectifs : une personne peut savoir ou dire ce qu'est l'expérience de l'autre qualitativement. Ils sont subjectifs, cependant, au sens où même cette attribution objective d'expérience est possible seulement pour quelqu'un qui soit suffisamment semblable à l'objet de l'attribution pour être en mesure d'adopter son point de vue – pour comprendre l'attribution aussi bien à la première qu'à la troisième personne, pour ainsi dire (trad. fr., p. 397).

Je pense que cette précision est d'une importance capitale. Aussi longtemps que l'on admet l'équation *conscience phénoménale = expérience subjective = expérience privée*, on situe non seulement la conscience phénoménale en dehors du champ des sciences physiques, mais aussi – plus radicalement – en dehors du

champ de la connaissance en général. Une telle conception barre radicalement la voie au projet d'édifier une théorie (objective) de la conscience phénoménale. Elle équivaut à rejeter purement et simplement la description des qualia du côté de la poésie. Si l'on admet la position de Nagel, en revanche, la situation est toute différente : certes, l'édification d'une théorie de la conscience phénoménale pose un problème de méthode, mais elle n'en demeure pas moins possible par principe. Si les qualia qui accompagnent une expérience quelconque ne sont pas seulement privés, mais sont des données *objectives* (accessibles à tout individu vivant la même expérience), alors cela a un sens d'entreprendre de les décrire et d'en faire une théorie. Nous verrons que Frank Jackson se souviendra de cette idée au moment de répondre à certaines objections avancées contre l'argument de la connaissance (cf. *infra*, section 6.).

5. L'argument de la connaissance (*Knowledge Argument*)

Dans la section précédente, j'ai passé brièvement en revue quelques propriétés couramment attribuées aux qualia. Ceux-ci sont tour à tour décrits comme :

(1) incompréhensibles
(2) impossibles à prédire
(3) incommunicables et privés (dépourvus de statut épistémique)
(4) liés au « point de vue en première personne » mais « objectifs »

Il y aurait certainement beaucoup à dire sur ces propriétés, qui ne sont naturellement pas interchangeables. Elles traduisent, à nouveau, différentes attitudes face aux qualia. Laissant ce point de côté, je voudrais à présent présenter la version classique de l'argument de la connaissance, telle qu'elle a été formulée par Frank Jackson en 1982.

Le fait est que les précédents arguments en faveur de l'irréductibilité des qualia ne satisfont pas Jackson. L'argument de Nagel, en particulier, porte sur ce qu'il est ou non dans le pouvoir de l'esprit humain d'accomplir (il m'est très difficile de me transposer en imagination dans la peau d'une chauve-souris) ; or, remarque Jackson, le physicalisme n'affirme rien concernant ce que l'esprit humain peut ou non concevoir. Ce ne sont pas les capacités imaginatives de l'être humain qui sont en cause lorsqu'on fait valoir l'existence de qualia. C'est pourquoi Jackson cherche un autre argument qui serait plus neutre sur ce point et donc, aussi, éventuellement plus convaincant car plus ciblé.

Comment se présente l'argument de Jackson ? On peut le reconstruire en trois étapes. D'abord (1), il commence par baptiser « information physique » toutes les informations que les sciences physiques, chimiques et biologiques nous procurent à propos du monde. Par exemple, lorsqu'un médecin m'explique les processus qui ont lieu dans mon système nerveux, on peut dire qu'il me

transmet de l'information physique. Ensuite (2), Jackson définit le physicalisme comme la conception selon laquelle « toute information (correcte) est une information physique ». Enfin (3), il rejette la thèse physicaliste au moyen de l'idée suivante : si quelqu'un me dit tout ce qu'il se passe dans un cerveau vivant, s'il me transmettait toutes les informations physiques possibles à propos de l'état de mon système nerveux, le rôle fonctionnel de mes états mentaux, etc., il y a quelque chose qu'il ne m'aurait pas dit. Il ne m'aurait encore rien dit de la souffrance provoquée par des douleurs physiques, de l'état suscité par la jalousie ; il ne m'aurait rien dit à propos de l'expérience consistant à goûter un citron ou à sentir une rose. Bref, aucune information physique ne capture l'odeur de la rose, or l'odeur de la rose est pourtant une information nouvelle, donc le physicalisme est faux (puisque selon le physicalisme, toute information théoriquement pertinente sur le monde est une information physique).

Jackson illustre son argumentation au moyen de deux scénarios célèbres. Le premier scénario met en scène un individu anormal nommé Fred. Fred présente une particularité remarquable, il souffre d'un « daltonisme inversé » : il a une plus grande sensibilité à la couleur que n'importe quel autre être humain. Cette capacité de discrimination lui permet notamment de distinguer deux sortes de rouge, ROUGE 1 et ROUGE 2, là où un être humain normal ne voit qu'une seul sorte de rouge. Supposons par exemple que, mis en présence de tomates mûres, nous voyons une seule couleur alors que Fred en voit deux. ROUGE 1 et ROUGE 2 sont des couleurs différentes pour lui exactement comme le bleu et le jaune sont des couleurs différentes pour nous ; il peut donc aisément séparer les tomates qui sont ROUGES 1 de celles qui sont ROUGES 2. En bref, il voit une couleur que nous ne voyons pas. On aura beau rassembler toutes les informations physiques que l'on veut à propos du cerveau de Fred et de son nerf optique, nous ne saurons jamais à quoi ressemble l'expérience qu'il a de la couleur inconnue de nous, nous ne saurons jamais à quoi ressemble cette nouvelle couleur ou ces nouvelles couleurs. Donc, il y a quelque chose que nous ne savons pas. Mais, par hypothèse, nous savons tout ce qu'il y a à savoir sur le corps de Fred, sur son comportement et sa physiologie interne, bref : nous avons toutes les informations physiques dont nous puissions diposer. Il reste que nous n'avons pas pour autant toutes les informations. Jackson en conclut que le physicalisme est faux. Supposons par ailleurs que Fred meure et donne son corps à la science, on pourra transplanter son nerf optique dans quelqu'un d'autre : cette personne apprendrait probablement quelque chose, elle aurait dès lors un accès aux informations qui demeuraient inaccessibles auapravant. Or, justement, si l'on admet que nous en saurions *plus* après l'opération, c'est qu'avant il y avait un défaut d'information. Donc le physicalisme est *incomplet*.

Le second scénario se déroule de manière similaire mais avec un individu normal dans des conditions de vision normales (sans aucune capacité extraordinaire à percevoir des couleurs que personne ne perçoit). C'est le

scénario le plus connu. Il met en scène une neurophysiologiste du nom de Mary. Élevée dans une pièce en noir et blanc, Mary a pu, grâce à un téléviseur noir et blanc, acquérir toutes les informations possibles sur les processus neurophysiologiques qui sous-tendent la perception des couleurs. Elle sait parfaitement ce qui se passe dans notre système nerveux lorsque notre œil est stimulé par le bleu du ciel et lorsque ce stimulus nous amène à dire que « le ciel est bleu ». La question de Jackson est la suivante : que se passera-t-il lorsqu'on libérera Mary ou lorsqu'on lui donnera un téléviseur couleur? Est-ce qu'elle apprendra quelque chose? Manifestement oui : elle apprendra l'effet que cela fait de voir une rose rouge, c'est-à-dire qu'elle acquiérera une nouvelle information qui n'était pas encore en sa possession. Or, elle avait toutes les informations physiques qu'il était possible d'avoir. Donc, à nouveau, le physicalisme est faux :

8.

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 specialises in the neurophysiology of vision and acquires […] all the physical information there is to obtain about what goes on whe we see ripe tomatoes, or the sky, and use terms like 'red', 'blue', and so on. […] What will happen when Mary is released from her black and white room or is given a colour 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, p. 130).

Mary est une brillante scientifique qui, pour une raison quelconque, est forcée d'effectuer ses recherches recherches par l'entremise d'un téléviseur noir et blanc depuis une pièce noire et blanche. Elle se spécialise en neurophysiologie de la vision et acquiert […] toute l'information physique qu'il est possible d'acquérir concernant les phénomènes nerveux qui se produisent en nous lorsque nous voyons des tomates mûres, ou lorsque nous voyons le ciel, et utilisons des termes comme « rouge », « bleu », et ainsi de suite. […] Qu'adviendra-t-il si on libère Mary de la pièce noire et blanche ou si on lui donne un téléviseur couleurs ? *Apprendra*-t-elle quelque chose ou non ? Il semble tout à fait évident qu'elle apprendra quelque chose au sujet du monde tout comme de notre expérience visuelle du monde. On doit alors conclure que ses connaissances précédentes étaient incomplètes. Mais nous avons posé au départ qu'elle possédait *toutes* les connaissances physiques. Il s'ensuit que les connaissances physiques n'épuisent pas l'ensemble des connaissances et que le physicalisme est faux[6].

[6] Trad. fr. du passage par P. Poirrier dans *Philosophie de l'esprit*, t. II, Paris, Vrin, 2003, p. 201-202.

6. Objections soulevées contre l'argument de la connaissance

On a adressé une série d'objections à l'argument de la connaissance. Ces objections visent toutes à contester que Mary ait appris une nouvelle information ou une information d'un nouveau genre (non physique) : soit parce qu'elle n'aurait rien appris ; soit parce qu'elle aurait appris autre chose. Dans tous les cas, ces objections n'ont qu'un seul but : souver l'équation *connaissance = connaissance physique*.

1 / Une première objection consiste à dire ceci : Mary n'a pas connaissance de tous les faits physiques aussi longtemps qu'elle ne sait pas ce que ça fait de voir une rose rouge. Daniel Dennet défend cet argument et maintient que, si Mary sait vraiment tout à propos des phénomènes physiques, alors elle devrait aussi savoir ce que ça fait de voir une rose rouge. Bref, cette objection revient à balayer l'argument de la connaissance d'un revers de la main en niant que Mary ait appris quelque information nouvelle que ce soit.

2 / Une autre objection a été avancée par Churchland. Celui-ci interprète l'apprentissage de Mary comme l'obtention d'une connaissance par contact direct. Ce n'est donc pas une nouvelle connaissance par description, car Mary connaît toutes les propositions vraies qu'il y a à connaître à propos de la vision des couleurs. Elle acquiert seulement une information privée, qui ne concerne pas les couleurs mais qui la concerne elle (sa manière à elle de percevoir le rouge, l'effet que ça lui fait de voir du rouge). – Jackson a répondu à cette objection en accentuant le caractère « objectif » de l'expérience subjective : ce que Mary apprend, ce n'est pas quelque chose à propos de sa propre expérience (privée), mais quelque chose à propos de l'expérience des autres. Elle sait enfin l'effet que cela fait aux autres êtres humains de percevoir du rouge. La réponse de Jackson consiste donc à mettre en évidence la dimension « objective » ou « non privée » de la conscience phénoménale, comme Nagel (1974) l'avait fait. Pour reprendre la terminologie de ce dernier, on pourrait dire que Mary acquiert la connaissance d'un *type* d'expérience (voir du rouge ou voir une rose rouge) et non d'une expérience isolée (l'individu *y* voyant du rouge).

3 / Une autre objection influente est l'objection basée sur la « capacité » ou sur l' « aptitude ». Elle a été avancée par Laurence Nemirow, dans sa thèse de doctorat à l'Université de Stanford en 1979, *Functionalism and the Subjective Quality of Experience*, avant d'être reprise, de façon condensée, dans sa discussion de Nagel, *Mortal Questions*. On la retrouve également sous la plume de David Lewis, dans la postface (1983) à l'article « Douleur de fou et Douleur de Martien » et dans « What experience teaches » (1988). Selon cette objection, ce que Mary gagne en sortant de la pièce est une capacité et non une connaissance. Elle ne sait pas quelque chose de nouveau, mais elle a développé une nouvelle compétence : elle est capable de se souvenir, de reconnaître et/ou d'imaginer une expérience visuelle du rouge. Je cite David Lewis : « Savoir

l'effet que cela fait, ce n'est pas être en possession de quelque information que ce soit. Ce n'est pas éliminer des possibles qui nous étaient ouverts jusque-là. En fait, savoir ce que cela fait, c'est plutôt avoir fait l'acquisition d'aptitudes : des aptitudes à reconnaître, à imaginer » (postface de 1983 à Lewis 1980 ; trad. fr., p. 305). Selon cette conception, Mary disposait déjà de tous les renseignements possibles sur la perception du rouge, mais elle gagne seulement une nouvelle aptitude. – Cette approche, à nouveau, n'est pas sans soulever certaines difficultés. Au moment où Mary voit la rose rouge pour la première fois, cela n'a pas de sens de dire qu'elle gagne une capacité à se remémorer, puisque le moment n'est pas passé. Admettons par ailleurs qu'elle ai une imagination pauvre : dans ce cas, elle ne gagne pas non plus une capacité à reconnaître ou à imaginer l'expérience de voir du rouge.

4 / L'objection la plus répandue est la suivante : ce que Mary gagne, c'est seulement une nouvelle manière de se représenter le même fait qu'elle connaissait déjà auparavant. Cette objection est due à Terry Horgan (1984). L'idée de Horgan est simple : la notion d'information physique est ambiguë ou équivoque. Il y a deux sens dans lesquels ont peut parler d'information physique : l'information physique au sens n°1 désigne une information « dans le *language* physique » (*explicit physical information*) ; au sens n°2, on peut entendre par là une information à propos d'un *fait* physique (*ontollogically physical information*). Toutes les entités/propriétés auxquelles on se réfère au moyen d'information physique seraient alors des entités/propriétés physiques. Si l'on admet cette distinction, il en résulte deux version de l'argument de Jackson (*cf.* Nida-Rümelin 2009) :

a. Version faible:

   i. Mary, avant sa libération, a une *connaissance physique* complète des faits relatifs à la vision humaine des couleurs.
   ii. Mais il y a une *sorte de connaissance* concernant les faits relatifs à la vision humaine des couleurs qu'elle ne possède pas avant sa libération.
   iii. Donc, il y a une sorte de connaissance concernant les faits relatifs à la vision humaine des couleurs qui est une *connaissance non physique*.

La conclusion de ce raisonnement est une affirmation épistémologique qui est compatible avec le physicalisme.

b. Version forte:

   i. Mary, avant sa libération, connaît tous les *faits physiques* concernant la vision humaine de la couleur.

ii. Mais il y a *certains faits* concernant la vision humaine de la couleur qu'elle ne connaît pas avant sa libération.
iii. Donc il y a des *faits non physiques* concernant la vision humaine de la couleur.

Cette deuxième version conduit à une affirmation ontologique qui est incompatible avec le physicalisme (c'est probablement cette version que Jackson a en tête) : il existe des entités dont on ne peut rendre compte dans la théorie physicaliste.

Selon Horgan, le scénario de Mary montre qu'elle acquiert bien quelque chose en sortant de la pièce, mais ce n'est pas de l'information non physique au sens d'une information qui ne concernerait pas des faits physiques (Mary ne découvre pas de nouveaux faits, non physiques – plus exactement : rien ne permet d'affirmer qu'elle découvre des faits non physiques). Ce que Mary acquiert, c'est de l'information non physique au sens où elle apprend à formuler une connaissance dans des termes qui ne sont pas ceux des sciences physiques. Mais cette information *non physique* au sens 1 peut très bien être de l'information *physique* au sens 2 : ce peut très bien être de l'information (dans un langage non physique) à propos de faits physiques. Il suffit d'admettre qu'il y a plusieurs modes d'accès à la réalité physique. Bref, Mary gagnerait seulement une *nouvelle connaissance d'un ancien fait*, un nouveau concept (un concept phénoménal) d'une ancienne propriété (ancienne car déjà connue : être rouge).

Cette stratégie – la stratégie du *New Knowledge/Old Fact* – est celle adoptée par la majorité des physicalistes. Selon ce scénario, « la distinction entre subjectif et objectif, la distinction entre la première personne et la troisième personne sont des distinctions entre des genres de concepts, non des genres de propriétés » (Block 2003).

Joseph Levine a repris et développé cette stratégie en 1993, dans « Omettre l'effet que cela fait ». Selon Levine, les arguments anti-physicalistes habituels n'ont qu'une portée épistémologique et sont dépourvus de portée métaphysique. La faiblesse des théories physicalistes n'est pas d'omettre certains faits, mais d'omettre une certaine sorte de connaissance à propos des faits physiques. L'argumentation de Jackson est « contaminée » par un glissement du registre épistémologique au registre métaphysique : Jackson tente de tirer une conclusion métaphysique de prémisses épistémologiques. Or, comme le note Levine, « Nul ne peut conclure à la variété des faits à partir de l'existence d'une variété de modes d'accès aux faits » (1993, trad. fr., p. 203). L'argument de Jackson échoue à prouver la fausseté du physicalisme, puisqu'il ne fournit aucune raison de soutenir que Mary, après sa libération comme avant, ne se référait pas au même fait.

Cette ligne argumentative entraîne de nouvelles difficultés. Je me contenterai, pour terminer, de mentionner un point discuté par Levine. Si l'on

admet qu'il est possible d'accéder au même fait selon deux modalités d'accès (que Levine nomme : « à la première personne », Mary *après* sa libération, et « à la troisième personne » : Mary *avant* sa libération), le problème est de savoir comment il est possible qu'un même fait prête ainsi le flanc à deux approches foncièrement distinctes. N'est-on pas obligé d'admettre que le fait possède deux propriétés distinctes, une propriété à laquelle nous accédons en troisième personne (le rôle fonctionnel) et une propriété à laquelle nous accédons en première personne (l'aspect qualitatif) ? Si l'on répond par l'affirmative, la stratégie du *New Knowledge/Old Fact* s'effondre et l'argument de la connaissance aurait bel et bien des conséquences métaphysiques, au sens où il nous engagerait à un dualisme des propriétés. Pour sauver le physicalisme, il faudrait alors montrer que l'aspect qualitatif d'un état n'est rien d'autre qu'une propriété physique. Pour conclure à la fausseté du physicalisme, il faudrait en revanche montrer que l'aspect qualitatif d'une expérience est *irréductible* à une propriété physique, que l'on a affaire à deux types de propriétés hétérogènes. Or, selon Levine, l'argument de la connaissance ne permet pas non plus de tirer une telle conclusion métaphysique concernant l'hétérogénéité des propriétés. Tout ce qu'il permet de conclure, c'est qu'il existe ce que Levine nomme un « fossé dans l'explication » : à la question de savoir pourquoi, lorsque nous nous trouvons dans tel état physico-fonctionnel, notre expérience présente tel aspect qualitatif et non tel autre, les sciences physiques sont incapables d'approter une réponse (Levine 1993, trad. fr., p. 207).

7. Remarque conclusive

Dans ce qui précède, j'ai tâché de clarifier le problème de la conscience phénoménale et de présenter – certes très succinctement – un échantillon représentatif des débats consacrés à cette question. Comme on le voit, ces débats se déploient très largement sur un terrain métaphysique et/ou épistémologique. La question de savoir si les qualia sont des *entités non physiques* et/ou si la connaissance des qualia est une *connaissance non physique*, a constitué une préoccupation constante des philosophes de l'esprit. L'un des enjeux, en outre, est de cerner la portée métaphysique et/ou épistémologique de l'argument de la connaissance, qui a souvent été considéré comme l'argument anti-physicaliste par excellence.

Toutefois, il y a une autre dimension de l'étude de la conscience phénoménale que les querelles métaphysiques et épistémologiques ont largement recouvert : la dimension phénoménologique. C'est sans aucun doute chez Thomas Nagel que cette dimension s'exprime le plus clairement. Ce que Nagel retient du problème des qualia, ce n'est pas la réfutation d'une position métaphysique ou épistémologique (le physicalisme), mais la nécessité de

construire une phénoménologie objective, qui nous évite de recourir à l'imagination. Ce dont nous avons besoin, en l'occurrence, c'est de pouvoir édifier une théorie de l'expérience subjective (vécue en première personne) qui soit aussi objective et aussi scientifique que possible :

4.

At present we are completely unequipped to think about the subjective character of experience without relying on the imagination – without taking up the point of view of the experiential subject. This should be regarded as a challenge to form new concepts and devise a new method – an objective phenomenology not dependent on empathy or the imagination (Nagel 1974, p. 449).	Dans l'état actuel des choses, nous sommes totalement dépourvus de moyens pour penser le caractère subjectif de l'expérience sans avoir recours à l'imagination – sans adopter le point de vue du sujet de l'expérience. Ceci devrait apparaître comme un défi à la formation de concepts nouveaux et à la recherche d'une nouvelle méthode – une phénoménologie objective qui ne dépendrait pas de l'empathie ou de l'imagination (trad. fr., p. 403).

Dans quelle mesure l'approche phénoménologique – telle qu'elle s'est développée dans ce que l'on appelle traditionnellement le « mouvement phénoménologique » (représenté par Brentano, Husserl, Heidegger, Sartre, etc.) – offre-t-elle des ressourcs pour construire une telle phénoménologie objective (une théorie scientifique de l'expérience subjective) ? C'est cette question qu'il nous faudra examiner en détail par la suite.

BIBLIOGRAPHIE

[Anonyme] (1929), *Wissenschaftliche Weltauffassung, der Wiener Kreis*, Wien, Wolf ; trad. fr. A. Soulez *et alii*, « La Conception scientifique du monde. Le Cercle de Vienne », dans R. Carnap *et alii*, *Manifeste du Cercle de Vienne et autres écrits*, A. Soulez dir., 2e éd., Paris, Vrin, 2010, p. 104-146.

Armstrong David M (1968), *A Materialist Theory of the Mind*, London, Routledge ; 2e éd. révisée, avec une nouvelle préface, 1993 (réimprimé en 2002).

Bayertz Kurt, Gerhard Myriam et Jaeschke Walter (éds.) (2007), *Weltanschauung, Philosophie und Naturwissenschaft im 19. Jahrhundert*, Bd. 3 : *Der Ignorabimus-Streit*, Hamburg, Meiner.

Block Ned (2003), « Philosophical Issues About Consciousness », dans L. Nadel (éd.), *Encyclopedia of Cognitive Science*, London, MacMillan.

Broad Charlie Dunbar (1925), *The Mind and its Place in Nature*, New York, Harcourt, Brace & Company.

Chalmers David (1996), *The Conscious Mind: In Search of a Fundamental Theory*, Oxford, OUP ; trad. fr. S. Dunand, *L'Esprit conscient. À la recherche d'une théorie fondamentale*, Paris, Ithaque, 2010.

Du Bois-Reymond Emil (1872), *Über die Grenzen des Naturerkennens* [*Sur les limites de la connaissance de la nature*], Leipzig, Veit & Comp, [1-2]1872, [3]1873, [4]1876, [5]1882, [6]1884, [7]1886, [8]1891 (plus de multiples rééditions posthumes) ; trad. fr.

(anonyme), « Les bornes de la philosophie naturelle », dans *La Revue scientifique de la France et de l'étranger*, 2e série, XV (1874), p. 337-345.
Gallagher Schaun et Zahavi Dan (2008), *The Phenomenological Mind. An Introduction to Philosophy of Mind and Cognitive Science*, London-New York, Routledge.
Horgan Terry (1984), « Jackson on Physical Information and Qualia », dans *Philosophical Quarterly* 34, p. 147-152.
Jackson Frank (1982), « Epiphenomenal Qualia », dans *The Philosophical Quarterly* 32/127, p. 127-136.
— (1986), « What Mary Didn't Know », dans *The Journal of Philosophy* 83/5, p. 291-295.
Levine Joseph (1983), « Materialism and Qualia: The Explanatory Gap », dans *Pacific Philosophical Quarterly* 64, p. 354-361.
— (1993) « On Leaving Out What It's Like », dans M. Davies et G. W. Humphreys (éds.), *Consciousness: Psychological and Philosophical Essays*, Oxford, Blackwell, p. 121-136 ; trad. fr. P. Poirrier, « Omettre l'effet que cela fait », dans D. Fisette et P. Poirrier (éds.), Philosophie *de l'esprit*, t. II : *Problèmes et perspectives*, Paris, Vrin, 2003, p. 195-221.
Lewis David (1980), « Mad Pain and Martian Pain », dans N. Block (éd.), *Readings in the Philosophy of Psychology*, vol. I, Harvard, Harvard University Press, p. 216-222 ; trad. fr. D. Boucher, « Douleur de fou et douleur de martien » (avec la postface de 1983), dans D. Fisette et P. Poirrier (éds.), *Philosophie de l'esprit*, t. I : *Psychologie du sens commun et sciences de l'esprit*, Paris, Vrin, 2002, p. 289-306.
McGinn Colin (1997), *The Character of Mind. An Introduction to the Philosophy of Mind*, Oxford, OUP, 2e éd. (1re éd. 1982).
Nagel Thomas (1974), « What is it like to be a bat ? », dans *The Philosophical Review* 83/4, p. 435-450 ; trad. fr. P. Engel, « Quel effet cela fait, d'être une chauve-souris ? », dans Th. Nagel, *Questions mortelles*, Paris, PUF, 1983, p. 193-209 ; rééd. dans D. Hofstadter et D. Dennett, *Vues de l'esprit*, Paris, InterÉditions, 1987, p. 391-404.
Nida-Rümelin Martine (2009), « Qualia: The Knowledge Argument », dans *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/qualia-knowledge/.
Schlick Morritz (1926), « Erleben, Erkennen, Metaphysik », dans *Kant-Studien* 31, p. 146-158 ; rééd. dans *Id.*, *Gesammelte Aufsätze 1926-1936*, Fr. Waismann éd., Hildesheim, Olms, [1]1938, [2]1969 ; trad. fr. B. Cassin et A. Guitard, « Le vécu, la connaissance, la métaphysique », dans R. Carnap *et alii*, *Manifeste du Cercle de Vienne et autres écrits*, A. Soulez dir., 2e éd., Paris, Vrin, 2010, p. 175-188.
Watson John (1913), « Psychology as the behaviorist views it », dans *Psychological Review* 20, p. 158-177.
