

LES NEUROSCIENCES ET LE PROBLEME DE LA CONNAISSANCE DE SOI

Claire Etchegaray

IRePh, Université de Paris Ouest

À propos des neurosciences, les discussions contemporaines se concentrent sur le problème du rapport entre corps et esprit, ou encore, dans la mesure où l'esprit se caractérise par une capacité à avoir au moins en apparence une expérience en première personne, celui de la conscience. La difficulté est de savoir comment des propriétés avancées dans des énoncés à la troisième personne (« il y a eu une activation des neurones du cortex visuel V4 »), soumises à des lois physiques, peuvent expliquer des propriétés et des comportements psychologiques, qui, même s'ils peuvent être rapportés à la troisième personne (« elle a vu du rouge »), s'énoncent dans des déclarations à la première personne (« j'ai vu du rouge ») ou s'analysent dans des expériences à la première personne (« c'était rouge » étant équivalent à « cela m'est apparu rouge »). C'est le fameux *explanatory gap*¹. Le présent article se concentrera sur un problème différent. Car quelle que soit la façon dont le problème de la conscience est posé, résolu ou dissout, il est présumé intéressant parce qu'il touche à la connaissance de soi entendue comme une capacité à se comprendre en tant qu'être ayant une expérience en première personne, ou tout au moins en tant qu'être *croyant* avoir une expérience en première personne². Le discours des neurosciences accroît notre intelligence du monde parce qu'il est informatif, mais, que nous soyons dualistes, éliminativistes ou illusionnistes, son intérêt n'est-il pas également de mieux nous comprendre ?³ Inversement, un certain dédain pour les neurosciences s'accompagne sans doute de la conviction qu'elles ne nous parlent pas véritablement de nous et ne satisfont pas notre désir de mieux nous comprendre. Et si nous préférons adopter une troisième position, qui confesse que la connaissance de la conscience est inaccessible à notre entendement, nous sommes conduits à penser qu'une partie de la connaissance de soi reste un mystère. Il y a donc un problème différent de

¹ Levine J., « Materialism and qualia : the explanatory gap », *Pacific Philosophical Quarterly*, vol. 64, 1983, p. 354-361.

² Le problème de la connaissance de soi tel que nous l'entendons dans ce qui suit ne recouvre pas cet autre problème, dont nous ne traiterons pas, qui est celui de l'identité personnelle.

³ Un exemple de dualisme, en l'espèce un dualisme des propriétés, se trouve dans *L'esprit conscient* de David Chalmers (1996, tr. fr. par S. Dunant, Paris, Ithaque, 2010, p. 183). Un indice parmi d'autres de sa prétention à procurer une compréhension de soi était, à l'époque, son opposition à une théorie quantique de la conscience (celle d'Everett) motivée par l'intuition qu'elle était intellectuellement rigoureuse mais trop « étrange » (*sic*). Quant à l'illusionnisme, il prétend non seulement expliquer que la conscience est une « illusion d'utilisateur » (Dennett D., *La conscience expliquée* (1991, tr. fr., Paris, Odile Jacob, 1993 et « Why and How Does Consciousness Seem the Way it Seems ? », in T. Metzinger et J.M. Windt (éd.), *Open MIND*, 10, Janvier 2015, doi:10.15502/9783958570245) mais aussi, plus récemment avec Keith Frankish et François Kammerer, expliquer les représentations illusoire d'avoir des *expériences en première personne* (Frankish K., « Illusionism as a Theory of Consciousness », *Journal of Consciousness Studies*, vol. 23, n°11-12, 2016, p. 11-39 et Kammerer F., *Conscience et matière. Une solution matérialiste au problème de l'expérience consciente*, Paris, Éditions matériologiques, 2019). Cette nouvelle forme d'illusionnisme prétend donc procurer une meilleure compréhension de nous-mêmes que l'éliminativisme : non seulement elle explique qu'il n'y a pas de processus cognitif qui corresponde vraiment à l'expérience phénoménale, mais elle prétend comprendre d'où viennent nos croyances en une telle expérience. Dans l'analyse proposée par Kammerer on peut même dire que nos concepts phénoménaux sont produits par une connaissance de soi naïve (décrivant et prévoyant de manière robuste nos comportements) et que l'analyse cognitive propose une connaissance de soi plus complète, laquelle explique l'illusion que la conscience procure une soi-disant connaissance phénoménale.

celui de la conscience ou du rapport corps/esprit mais qui n'est pas sans lien avec eux : le problème de la connaissance de soi. Il les croise en vertu de la supposition suivante : si l'on ne comprend pas comment le discours neurophysiologique parle de l'expérience en première personne, si l'on ne comprend pas comment le discours de l'expérience en première personne donne du sens aux observations neurophysiologiques, ou tout simplement si l'on ne comprend pas comment ces discours peuvent se référer l'un à l'autre, alors l'on ne comprend pas comment les neurosciences permettraient de mieux nous comprendre. L'étude qui suit envisage ces différentes manières de traiter le problème de la connaissance de soi.

On a souvent considéré que les neurosciences soulevaient une question sémantique : les mots qui ont un sens apparemment psychologique comme « douleur » ou « rouge » (pour autant qu'il désigne le rouge perçu) auraient-ils en réalité une signification physique ?⁴ Si ce que nous venons de dire est juste, y répondre n'est pas une affaire de définition, c'est une tâche pour ainsi dire « herméneutique » : l'enjeu est d'élucider non seulement l'intelligibilité des neurosciences, mais également le type de connaissance de soi qu'elles peuvent procurer⁵. En quoi pouvons-nous dire qu'elles parlent de nous, nous renseignent sur les êtres que nous sommes et permettent de mieux nous comprendre ?

Une première réponse consiste à admettre que la question sémantique est problématique et, en partie, sans réponse, mais que nos difficultés à y répondre sont précisément révélatrices de notre condition. Elle puise dans le scepticisme existentiel de Thomas Nagel dans *Questions mortelles*. C'est ce que nous expliquerons dans un premier temps. Ensuite, nous montrerons en quoi différents traitements de la question sémantique, chez Nagel, Kripke et Putnam, font varier la tâche herméneutique qui se présente, depuis la première réponse, comme un véritable défi. Nous en viendrons finalement à envisager un renversement à même de rendre justice du sens que peuvent avoir les recherches croisées en neurosciences et en sciences cognitives. Dans la troisième partie en effet, nous proposerons de partir du fait que les discours de ces sciences sont compréhensibles, et nous examinerons leurs conditions d'intelligibilité en suivant une ligne de lecture pragmatiste⁶. Le renversement que nous proposerons a donc pour enjeu une défense de l'approche scientifique en tant qu'elle participe à la compréhension de soi.

⁴ Il faudra préciser dans l'article ce qu'il faut entendre par « signification physique ». Une première hypothèse, provisoire, consistera à la comprendre en termes de *référence* à des objets, à des événements ou encore à des propriétés physiques.

⁵ Par « herméneutique » on peut entendre en général plusieurs choses, qui ne s'excluent pas nécessairement. Ce terme connote une façon de donner du sens-1, en particulier à un texte ou à un discours (*e.g.* l'herméneutique biblique). Il connote aussi une façon de *comprendre* quel sens-2 ce sens-1 peut avoir dans la situation du locuteur ou du lecteur. C'est ce que Schleiermacher a réalisé méthodiquement sous le titre d'« herméneutique ». Enfin, peut être dite « herméneutique » la compréhension que ce sens-2 est toujours déjà une manière pour ce locuteur ou ce lecteur *de se comprendre* et ainsi de se connaître soi-même. Dans ce dernier usage, que l'on retrouve particulièrement chez H. Gadamer ou P. Ricœur, la compréhension, la compréhension de la compréhension et la compréhension de soi sont intimement liées. Sans exporter au sens strict cette triple connotation, nous l'utilisons de manière souple pour indiquer que la question sémantique soulevée par les neurosciences répond à un intérêt « herméneutique », c'est-à-dire à un désir de connaissance de soi (troisième connotation), qui cherche à se réaliser en comprenant le sens (seconde connotation) que peut avoir pour nous l'explication neurophysiologique. Pour ce faire, on peut tenter une sorte d'herméneutique mentaliste du discours neurophysiologique (première connotation). C'est ce que nous considérerons d'abord. L'impasse nous conduira à envisager une autre voie, qui examine plutôt le sens que peut avoir ce discours pour nous du fait de notre forme de vie.

⁶ Par « intelligibilité » nous entendons le fait qu'un texte puisse être compréhensible (donc avoir du sens-1). Qu'il soit *compris* renvoie au fait qu'il ait du sens pour nous et peut donc ensuite se prêter aux enquêtes herméneutiques (décrites dans la note précédentes).

1. Le scepticisme existentiel de Thomas Nagel

Dans son article de 1974 repris dans le chapitre 12 de *Questions mortelles*, « What Is It Like To Be A Bat ? », Nagel suit une ligne argumentative que l'on peut résumer de la manière suivante : de la même façon que je ne peux pas savoir l'effet que cela fait d'être une chauve-souris quand bien même j'aurais des connaissances précises et exhaustives sur le fonctionnement du système perceptif propre à son espèce, de la même façon, je ne pourrais pas connaître l'expérience humaine de la perception par les descriptions objectives du fonctionnement de mon système perceptif. Les connaissances objectives, neurocérébrales particulièrement, sur le fonctionnement de nos systèmes de perception ne pourraient donc rien nous dire sur l'expérience vécue de la perception en première personne (*what it is like to perceive*).

Cet argument peut sembler *existentiel*, à première vue, en un sens très lâche : l'existence des événements et des lois physiologiques ne garantit pas l'existence de l'expérience perceptive, notre existence. Il y aurait même, comme le dit Gordon Bearn, « un combat existentiel » : soit la physiologie (ou disons plutôt ce qu'elle décrit) existe, soit *nous* existons⁷. Comme on va le voir toutefois, ce n'est pas exactement en ce sens que Nagel y a vu un défi existentiel.

La question posée par Nagel : un défi pour la connaissance de soi

Dans l'article qu'il publie en 1974, il examine « l'hypothèse physicaliste » qui affirme que les états ou événements mentaux *sont* des états ou événements corporels. Toute la question est, selon lui, de savoir quel sens donner à ce « sont » :

« Après tout, pourrait-on dire, la signification du physicalisme est suffisamment claire : les états mentaux sont des états du corps ; les événements mentaux sont des événements physiques. Nous ne savons pas de *quels* états et événements physiques il s'agit, mais cela ne devrait pas nous empêcher de comprendre l'hypothèse en question. Quoi de plus clair que les mots « est » et « sont » ? Mais je crois que c'est précisément la clarté apparente du mot « est » qui est trompeuse (*deceptive*)⁸. »

L'hypothèse physicaliste de l'identité est donc d'emblée tenue pour l'identification d'un état ou événement mental à un état ou événement physique. Une identification, qui plus est, *apparemment claire*. C'est cette hypothèse physicaliste qui est la cible de l'argumentation développée par Nagel dans cet article devenu, après 1979, l'une de ses plus fameuses *Questions mortelles*.

Le rapport entre l'argument de la chauve-souris et l'hypothèse physicaliste de l'identité est le suivant : *pour comprendre qu'un discours neurophysiologique portant sur les événements physiques puisse nous apprendre quoi que ce soit sur notre expérience mentale*, il faudrait comprendre comment un terme désignant notre expérience mentale pourrait avoir pour référence un événement physique.

Ainsi, pour que les neurosciences soient une connaissance de soi, il faut qu'elles apportent une connaissance de l'expérience phénoménale. Or comment savoir si elles nous donnent une telle connaissance ? La réponse tient dans la proposition ϕ (où M est un événement ou état mental et P l'événement physique décrit par les neurosciences) :

⁷ Bearn G., « Bat ? Again ? William James, Consciousness and Our Insipid Existence », in *The Journal of Speculative Philosophy*, vol. 29, n°4, 2015, p. 522-546, p. 536.

⁸ Nagel T., « What Is It Like To Be A Bat ? », *The Philosophical Review*, vol. 83, n°4 oct. 1974, p. 435-450, p. 446-447, reprint in Nagel T., *Mortal Questions*, chapitre 12, Cambridge/New York/Melbourne, Cambridge University Press, 1979, p. 176, trad. fr., *Questions mortelles*, Paris, PUF, 1983, p. 195-211, p. 206.

ϕ (physicalisme) : Les neurosciences procurent une connaissance de M si « M » a pour référence un événement P.

Le défi de Nagel consiste à demander d'élucider donc la référence de « M » à P et *ipso facto* l'identité (type ou *token*) de M et P. Or selon Nagel « M » a un sens essentiellement expérientiel ou phénoménal que la référence à P ne peut apparemment pas expliquer. C'est ce que montre l'argument de la chauve-souris. Pour répondre au défi de Nagel, il ne faut donc pas seulement montrer comment nous pouvons utiliser les résultats des neurosciences concernant l'écholocation de la chauve-souris pour nous faire une idée de cette expérience. Un bel article de Kathleen Akins l'a fait en 1993 et nous y reviendrons plus loin. Mais pour lors, une telle réponse paraît insuffisante parce que le défi de Nagel suppose un réquisit sur les conditions d'intelligibilité des neurosciences, en l'espèce l'élucidation de l'identité (type ou *token*) entre M et P. Les neurosciences ne permettent de *se comprendre* (ou de comprendre une expérience phénoménale que l'on n'a pas) qu'à condition que l'on puisse comprendre *comment* ou *à quelles conditions* « M » réfère à P. C'est exactement ce défi que l'on retrouve encore dans un article de Laurence Bonjour, publié en 2013. Bonjour propose une expérience par la pensée similaire à l'argument de Nagel mais censée dépasser une faiblesse de ce dernier⁹. Car l'argument de Nagel n'est probant que si l'on précise en quoi consiste *la connaissance* qu'a la chauve-souris et dont je suis dépourvu. C'est ce qui prête le flanc aux critiques de Churchland ou Dennett : si l'on ne peut pas définir cette soi-disant connaissance étrangère au savoir objectif, elle est sujette à caution¹⁰. C'est pourquoi, Bonjour adopte un scénario inverse : accorderions-nous qu'un alien expert en neurophysiologie humaine puisse savoir quand *nous* percevons un objet vert et quand *nous* percevons un objet rouge ? Dans ce nouveau scénario le savoir perceptif est supposé connu du lecteur : l'expérience de percevoir du rouge ou du vert est partagée par Bonjour et son lecteur ; la question est de savoir si l'alien pourrait s'en faire une idée et l'instancier. Ce nouveau scénario dispense donc Bonjour de définir ce qu'est la connaissance perceptive par expérience. Il retrouve la conclusion de Nagel, mais en évitant le problème d'une définition de l'expérience perceptive : un alien, même en ayant une connaissance neurophysiologique exhaustive du système perceptif humain, ne pourrait pas savoir ce que cela me fait de percevoir. Le problème contourné, d'ailleurs, porte en lui-même la preuve de la conclusion : la neurophysiologie ne peut suffire à définir l'événement mental, ce dernier comprenant un aspect expérientiel. Bonjour remarque que l'enjeu chez Nagel est de montrer non la fausseté du physicalisme mais le fait qu'on ne peut pas savoir comment il pourrait être vrai¹¹. Dans ce même dessein, Bonjour choisit de présupposer la connaissance phénoménale que nous avons de nous-mêmes et de montrer qu'on ne peut pas comprendre comment un alien y aurait accès. Dans le scénario de Bonjour, à la différence du défi de Nagel, le discours de la conscience phénoménale est immédiatement intelligible, du fait même qu'on est l'organisme que l'on est, c'est-à-dire du fait qu'on a l'expérience que l'on a (et dès lors,

⁹ Bonjour L., « What Is It Like To Be A Human (Instead of A Bat) ? », *American Philosophical Quarterly*, vol. 50, n°4, 2013, p. 373-385.

¹⁰ Bonjour relève que le scénario bien connu de Jackson a la même faiblesse. Jackson présuppose que la connaissance acquise par Mary quand elle sort de sa chambre est différente de ses connaissances neurophysiologiques, mais pour le présupposer il faut déjà savoir en quoi elle consiste. Le scénario de Jackson en faveur des *qualia* est un argument circulaire. Voir Jackson F., « Epiphenomenal Qualia », in *Philosophical Quarterly*, vol. 32, n°127, 1982, p. 127-136 et « What Mary didn't know », *Journal of Philosophy*, vol. 83, no. 5, 1986, p. 291-295.

¹¹ Nagel T., « What Is It Like To Be A Bat ? », *Mortal Question*, p. 176, trad. fr., p. 206.

peut-on ajouter, que l'on maîtrise les règles conventionnelles de ce discours)¹². L'insuffisance des connaissances neurophysiologiques à nous faire comprendre cette conscience phénoménale est donc le signe d'une énigme : comment pourraient-elle contribuer à une connaissance de soi ?

Demandons-nous maintenant pourquoi ce défi herméneutique est existentiel. Aux yeux de Nagel, la raison en est qu'il nous plonge dans un embarras sceptique.

Le défi de Nagel en 1979 : existentiel car sceptique

L'article de 1974 devient en 1979 un chapitre de *Questions morelles*. Situé au sein de cette collection d'essais, il prend une résonance sceptique, tout comme d'ailleurs le non moins célèbre chapitre précédent sur les *Split brains*, soulevant la question de savoir combien d'esprits il y a dans les phénomènes de dissociation hémisphérique, ou encore celui découlant de la question du panpsychisme au chapitre 14. Ces chapitres qui forment un ensemble portant sur le *mind/body problem* sont à lire dans la perspective donnée au début de l'ouvrage dans le chapitre sur « L'absurde ».

Nagel y montre que le scepticisme, qui découle de notre interrogation sur la justification ultime de nos choix et de nos exigences de rationalité, révèle une forme d'absurdité existentielle. Il ne s'agit pas seulement de dire que notre quête de vérité serait *sans raison*. L'absurdité caractérise notre condition car la possibilité du scepticisme *se maintient alors même que nous sommes et restons engagés dans la vie* :

« Pourtant, quand nous adoptons ce point de vue et reconnaissons comme arbitraire ce que nous faisons, cela ne nous désengage pas de la vie, et c'est là que réside notre absurdité : non pas dans le fait que l'on puisse prendre sur nous un point de vue extérieur, mais dans le fait que nous-mêmes puissions le prendre, sans cesser d'être les personnes dont les préoccupations ultimes sont considérées aussi froidement.¹³ »

Or, le défi soulevé par les neurosciences comme la question de l'unité de la conscience illustrent parfaitement ce trait caractéristique du scepticisme que Hume avait mis en évidence dans le *Traité de la nature humaine* l.iv.7 (lequel est également cité par Nagel¹⁴). Ce défi n'est pas *réfuté* ou *rendu inepte* par l'engagement dans la vie – ici la vie scientifique où, rappelons-le, le sens du « sont » est apparemment clair d'après Nagel. Il ne conduit ni à abandonner les neurosciences ni à dire que le physicalisme est faux parce qu'ils occulteraient notre existence. Il repose sur une tâche sémantique qui nous plonge dans un embarras sceptique, lequel nous fait éprouver l'absurdité de notre condition humaine. Nagel fait explicitement référence au *Mythe de Sisyphe* de Camus¹⁵. Il prétend même aller plus loin que l'existentialisme français car, selon lui, « l'absurdité de notre situation dérive

¹² Locke et Hume faisaient des observations similaires à propos des idées simples et claires dont aucune définition ne suffit à donner la compréhension sans l'expérience. Selon Locke, aucune définition de « rouge » ne sera susceptible de faire comprendre la référence de ce terme, laquelle est uniquement constituée par l'expérience vécue de la perception. À l'inverse, il suffit d'avoir perçu du rouge pour que, si l'on maîtrise les règles d'usage de ce terme, on en ait une conception claire (Locke J., *Essai sur l'entendement humain*, III.iv.11, trad. fr. J-M Vienne, Paris, Vrin, 2006, 2 vols., p. 76). La même remarque est faite par Hume à propos de la croyance (*Enquête sur l'entendement humain*, V.ii, trad. fr. M. Malherbe in *Essais et traités*, Paris, Vrin, 2004, p. 86).

¹³ Nagel T., *Mortal Questions*, chapitre 2, p. 15, trad. fr., p. 29.

¹⁴ *Ibid.*, p. 20, trad. fr., p. 34.

¹⁵ On notera, comme une confirmation de notre hypothèse que l'enjeu du problème de la conscience est bien une forme de connaissance de soi, même aux yeux des illusionnistes, que D. Dennett met encore le *Mythe de Sisyphe* en exergue de son premier chapitre de *La conscience expliquée* (*op. cit.*, p. 35).

non pas d'un affrontement entre nos attentes et le monde, mais d'un affrontement à l'intérieur de nous-mêmes¹⁶ ». D'une part notre engagement dans la vie scientifique suppose la croyance que le sens de « sont » est intelligible (« *mental states are states of the body ; mental events are physical events* » disait Nagel) ; d'autre part nous éprouvons un doute légitime mais insoluble sur cette croyance, et donc le sens de cet engagement (la clarté du sens de « est » est illusoire, admet Nagel).

En somme, la compréhension de soi est problématique mais, comme on vient de le voir, ce caractère problématique est tout de même très instructif sur notre condition.

Comme ce problème vient d'une demande sémantique (élucider la référence de « M » à P et *ipso facto* l'identité (type ou *token*) de M et P) qui paraît impossible à satisfaire (car « M » a un sens essentiellement expérientiel ou phénoménal que la référence à P ne peut expliquer), venons-en aux raisons analytiques qui justifient selon lui ce point.

2. Comprendre la référence, un défi pour la connaissance de soi

Le défi Kripke-Nagel

D'abord il faut comprendre pourquoi le défi *se maintient* puisque c'est ce maintien qui est caractéristique de la condition humaine selon Nagel. On pourrait penser que le simple fait que la référence de « M » à P soit incompréhensible invalide ϕ . Mais ce qui donne une forte plausibilité à ϕ , c'est une intuition qui est supposée dans le scénario contrefactuel interspécifique (que ce soit celui de Nagel avec les chauves-souris, ou celui de Bonjour avec les aliens) : changer de constitution physique c'est changer d'expérience mentale¹⁸. D'un autre côté, ce qui fait que ϕ est mise en question, c'est une intuition inverse, qui se maintient également, celle selon laquelle le lien d'un état mental à notre constitution est contingent. C'est une vieille intuition, bien entendu, qu'on peut faire remonter à la philosophie dualiste classique¹⁹. Mais les raisons mises en avant par Nagel ne sont pas données par une métaphysique substantialiste. Elles sont psychologiques et *analytiques*, nourries à la fois d'une critique menée contre David Armstrong et d'une lecture de Saul Kripke²⁰.

La problématique signifiante d'un point de vue existentiel consiste à *comprendre* comment ces deux intuitions du lien entre notre constitution physique et notre expérience phénoménale se maintiennent. Cela complexifie un peu la tâche sémantique du physicaliste. Non seulement il lui faut montrer que « M » a pour référence P, et donc que leur lien est nécessaire mais il lui faut également montrer aussi pourquoi ce lien paraît à tort contingent. Ou bien, abandonner.

¹⁶ Nagel T., *Mortal Questions*, chapitre 2, p. 17, trad. fr., p. 31. Voir également Millière R., « La vie est-elle absurde ? » in *Klêsis*, 2018, n°41.

¹⁸ Par « constitution » on peut entendre une structure ou un système, non tant qu'une substance, si bien que le fonctionnalisme ne change rien à cette intuition. Dans l'article « What Is It Like To Be A Bat ? », cette intuition est clairement impliquée par la thèse de la ressemblance : plus on ressemble à quelqu'un, plus on croit comprendre ce qu'il est susceptible d'éprouver.

¹⁹ Elle se retrouve chez Thomas Reid (*Recherches sur l'entendement humain d'après les principes du sens commun* (1764), trad. fr., Paris, Vrin, 2015), en lien avec le projet de faire une science objective dégagant des lois de la perception sans perdre de vue l'effet que cela fait de percevoir. Mais Reid a un avantage sur nous : il dispose d'une métaphysique théiste où le *gap* entre le subjectif et l'objectif est un *principe originel* de la nature (un fait de Dieu véral) qui peut être tenu à la fois pour une loi *naturelle* et un principe d'immédiateté épistémique.

²⁰ L'instruction fut mutuelle entre Nagel et Kripke. Dans une conférence de *Naming and Necessity* publiée en 1980 mais qui fut donnée le 29 janvier 1970, Kripke renvoie à la critique de Nagel contre Armstrong publiée dans la recension d'un numéro de la *Philosophical Review* de 1970 ; Nagel renvoie à l'article de Kripke « Naming and Necessity » publié dans *Semantics of Natural Language* en 1972.

C'est un défi que Kripke posait lui aussi au physicalisme. Dans une note importante, Nagel dit penser « comme Kripke », que « l'hypothèse selon laquelle un certain état cérébral devrait nécessairement avoir un certain caractère subjectif est incompréhensible sans autre explication », et qu'une théorie qui en rendrait compte devrait aussi expliquer pourquoi cela apparaît néanmoins comme contingent. La connaissance de soi est mise au défi suivant : *ou bien* de constater une incompréhension signe de notre condition existentielle, *ou bien* si l'on est physicaliste de s'acquitter de la tâche suivante : faire comprendre qu'il y a une apparence contingente c'est-à-dire fournir une *explication* en termes non-phénoménaux (« non-topiques ») consistant à nous faire comprendre *pourquoi* il y a une apparence contingente et *pourquoi* elle est fautive²¹. La tâche est désormais d'expliquer *pourquoi il y a un lien nécessaire entre M et P* et *pourquoi néanmoins ce lien peut paraître contingent*. Kripke pensait que c'était probablement peine perdue²². L'argument de Kripke consiste à dire que la relation entre « M » (« douleur ») et « P » (« excitation dans les fibres C ») n'est pas la même que la relation entre des termes comme « chaleur » et « énergie cinétique moléculaire ». Le scientifique identifie en effet « chaleur » et « mouvement moléculaire » parce qu'en les employant il réfère à une même chose dans tous les mondes possibles. (Ce sont, dans le vocabulaire de Kripke, deux désignateurs rigides²³). L'identification est donc *nécessaire*. On pourra certes penser que l'énergie moléculaire cinétique aurait pu exister sans qu'il y ait de chaleur mais ce que l'on voudra dire par là c'est qu'elle aurait pu exister *sans que l'on ne ressente de la chaleur*, pas sans qu'il y ait de « chaleur » physique.

Dans le cas de l'identification typique entre douleur et état neurocérébral, en revanche, l'apparence de contingence ne peut pas être expliquée de la même façon. Car cette fois on ne peut pas dire « que l'état cérébral aurait pu exister sans que l'on *ne ressente de la douleur* mais qu'il n'existe pas sans douleur ». Il se trouve que le mot « chaleur » a été nommé à partir d'une propriété *accidentelle* qui est la sensation de chaleur, mais le désignateur rigide (employé par le scientifique) ici réfère à quelque chose qui ne se définit pas essentiellement par cette sensation. Ce n'est pas le cas du mot « douleur » : dans ce cas on a retenu une propriété phénoménologique pour référer à quelque chose qui a *essentiellement* cette propriété d'être ressenti. « A la différence de la chaleur, la douleur n'est pas seulement désignée rigidement par « douleur », mais la référence du désignateur

²¹ La double exigence d'*expliquer l'illusion de contingence* et de *justifier qu'elle est fautive* est ce que David Chalmers nomme récemment le « métaproblème de la conscience » (« The meta-problem of consciousness », in *Journal of Consciousness Studies*, vol. 25, n°9-10, 2018, p. 6-61). Le métaproblème qu'il formule est en effet le suivant : soit il faut expliquer notre intuition selon laquelle la conscience ne se laisse pas décrire dans des termes non-phénoménaux (c'est-à-dire non-qualitatifs), soit il faut expliquer *pourquoi* c'est impossible. C'est un défi qui doit selon lui se poser aux théories qui voient dans la conscience expérientielle une illusion (comme celle de Dennett dans *La conscience expliquée*) ou à celles qui voient dans un mécanisme neurocérébral une base pour la conscience expérientielle. Mais Chalmers envisage ce métaproblème dans un cadre un peu différent de celui de Nagel – celui où les mécanismes neurocérébraux expliquent la production de jugement (de haut niveau) *sur* la conscience.

²² L'illusionnisme auquel nous faisons référence dans l'introduction a récemment relevé le défi de Kripke. Mais notre but n'est pas de proposer un tableau des réponses au problème de la conscience. Notre propos est de montrer comment, historiquement, il a pu paraître nécessaire de résoudre la question sémantique pour montrer l'intérêt philosophique des neurosciences.

²³ Quand je parle de « la propriété formalisée dans une équation de Maxwell et Boltzmann », je fais référence à « l'énergie cinétique moléculaire ». Mais dans un autre monde que le nôtre, cela pourrait ne pas être le cas. Ainsi « énergie cinétique moléculaire » est un désignateur rigide (comme peut l'être un nom propre) alors que « propriété formalisée dans une équation de Maxwell et Boltzmann » désigne quelque chose en le décrivant par le biais d'autres références.

est, de plus, déterminée par une propriété essentielle du référent²⁴ ». En somme, l'apparence de M (par exemple la douleur) est une propriété essentielle de ce que désigne « M » (« douleur ») et dans cette apparence, il n'y a rien de neurologique – par conséquent on ne voit pas pourquoi l'apparence de contingence serait une illusion. Il n'en va pas de même de l'apparence (sentiment) de chaleur, qui semble ne rien contenir de l'énergie cinétique moléculaire, mais qui ne doit pas nous laisser penser que le lien entre les mots scientifiques « chaleur » (température) et « énergie cinétique moléculaire » est contingent (ces mots désignent bien une seule et même chose dans tous les monde possibles).

Ainsi, la tâche sémantique est loin d'être un « jeu d'enfant ». En 1979, Nagel pense que l'alternative est la suivante : ou bien la tâche sémantique ϕ imposée au physicaliste est impossible à résoudre et cela nous révèle l'absurdité de notre condition, ou bien sa solution doit permettre de comprendre *pourquoi* intuitivement elle nous paraît problématique. Sans rappeler les détails de l'argument de Kripke, il se borne, en 1979, à donner une illustration psychologique de *l'apparence de contingence*. Quand nous cherchons à imaginer ce que nous *percevons si...* (par exemple : si nous étions une chauve-souris), nous envisageons soit l'état physique que le scientifique pourrait percevoir, soit ce que nous pourrions ressentir en étant une chauve-souris. La première imagination est *perceptuelle* et la seconde *sympathique* selon Nagel. Ainsi, si on s'imagine avoir un certain état cérébral on utilise l'imagination perceptuelle et on n'utilise pas l'imagination sympathique – c'est la preuve que ce qui est imaginé n'est pas une expérience en première personne. Et vice versa. Selon Nagel, dans l'imagination sympathique (qui se projette dans un état en première personne), l'expérience se présente sans son état cérébral associé ; dans l'imagination perceptuelle (qui conçoit un état observé en troisième personne), l'état cérébral est présenté sans l'état mental associé. Cette distinction laisse bien paraître que la relation entre l'état mental et l'état cérébral est contingente.

Le défi du second Nagel : la théorie du double aspect

Toutefois, comme on sait, Nagel lui-même proposera une autre approche que celle qui est énoncée par ϕ , une approche qui met en avant la « théorie du double aspect ». Cette approche indiquée dans *Le point de vue de nulle part* et réitérée plus tard (par exemple en 1993 dans « Mais qu'est-ce que tout cela veut dire ? ») admet qu'une même chose peut avoir des propriétés de types différents, les unes physiques et les autres mentales – ou plus exactement qu'on puisse avoir sur une même chose deux perspectives, l'une objective et l'autre subjective. Elle conduit à justifier la valeur herméneutique des neurosciences un peu différemment cette fois, en adoptant la thèse χ , où « M » est le nom d'une propriété mentale (manifeste sous une certaine perspective, subjective), « P » celui d'une propriété physique (manifeste sous une autre perspective, objective) et X une réalité²⁵. Cette thèse est la suivante :

χ (*double aspect*) : Les neurosciences peuvent nous donner une connaissance de X si « M » réfère à X et « P » réfère à X.

²⁴ Kripke S., *Naming and Necessity* (1980), trad. fr. par P. Jacob et F. Recanati, *La logique des noms propres*, Paris, Les éditions de Minuit, 1982, p. 141.

²⁵ On peut distinguer le dualisme des propriétés (qui distingue les propriétés M des propriétés P, malgré un monisme substantialisme) de la théorie du double aspect (qui est un monisme ontologique et distingue seulement les points de vues ou perspectives sous lesquelles se manifestent les propriétés). Selon cette théorie, « M » renvoie à une perspective expérientielle sous laquelle se manifeste la propriété « rouge » et « P » renvoie à une perspective observationnelle sous laquelle se manifeste la propriété « excitation du cortex V4 ».

La modification ne paraît pas très importante. D'ailleurs, Nagel ne voit pas en χ une option différente de ϕ , mais simplement une manière de comprendre ϕ si on l'applique au rapport mental/physique et que l'on tient « M » et « P » pour des descriptions de propriétés. Cela conduit pourtant à modifier le défi herméneutique. La possibilité qu'on puisse penser M sans P n'est plus du tout suffisante pour réfuter la possibilité que les neurosciences fournissent une connaissance de soi. Et cette possibilité n'est même plus du tout problématique : le défi herméneutique n'est plus significatif d'une situation existentielle absurde. Dans l'esprit de Nagel, la tâche à accomplir reste indiquée par ϕ , mais cette tâche n'est plus aussi contradictoire qu'elle pouvait paraître. Il continue de penser que la tâche principale est de comprendre que les événements mentaux sont des événements physiques mais il pense que la théorie du double aspect permettra de le comprendre. Ce n'est sans doute qu'une piste à ses yeux et c'est en quoi la tâche herméneutique ne disparaît pas. Mais elle n'est pas insoluble dans son principe.

Dans la théorie du double aspect, le discours neurophysiologique et le discours de l'expérience personnelle sont certes incommensurables pour nous au sens où passer de la compréhension de l'un à celle de l'autre c'est changer de perspective et donc de cadre conceptuel. Mais cette limite qui tient à notre condition, notre incapacité de fait à nous tenir aux deux points de vue en même temps, n'empêche pas le philosophe de *penser* la possibilité d'un lien essentiel entre le mental et le physique, une possibilité dont nous n'avons pourtant aucun concept.

« Pour ce qui est des concepts ordinaires de douleur, ou d'expérience visuelle, il se pourrait qu'ils ne possèdent aucune propriété essentielle qui ne soit mentale. Mais il se peut que ces concepts ne saisissent pas toutes les propriétés essentielles du phénomène. En réalité, il est peut-être impossible pour un phénomène mental de ne pas posséder également des propriétés physiques, même si nous ne pouvons pas former une conception d'un lien nécessaire de cette sorte²⁶. »

La critique de Nagel vise Kripke mais elle porte aussi bien, croyons-nous, sur un point de son propre article sur la chauve-souris : ce n'est pas parce que je peux imaginer un état mental (par imagination sympathique) sans utiliser d'imagination perceptuelle (laquelle me permet d'imaginer des états physiques observables) qu'il est *possible* qu'en réalité un état mental ne possède aucune propriété physique²⁷. Que les neurosciences ne soient pas une connaissance qui dise *ce que cela fait d'être soi* (c'est-à-dire ce que *cela fait* d'être un organisme tel que soi) ne les empêche pas tout de même de contribuer à une forme de connaissance de soi, plus exactement de procurer une connaissance de ce qui en soi échappe à l'expérience phénoménale de soi. Ce X, c'est ce que nous sommes, et qui, quoiqu'indéterminable, est connaissable *sous certains aspects*, c'est-à-dire sous une perspective objective et sous une perspective subjective. Une articulation possible entre ces aspects reste pensable, sans doute du fait même de notre capacité à passer d'un point de vue à l'autre, mais elle est impossible à conceptualiser²⁸.

²⁶ Nagel T., *Le Point de vue de nulle part*, trad. fr. S. Kronlund, Paris, Éditions de l'éclat, 1993, p. 59.

²⁷ Rappelons que l'imagination que Nagel qualifie de « perceptuelle » se figure un état que percevrait *le scientifique*. Pour éviter de la confondre avec l'imagination de ce que l'on percevrait en première personne, il vaudrait peut-être mieux parler d'imagination « observationnelle ».

²⁸ François Loth a montré que le naturalisme de Nagel se donne pour exigence d'intégrer à la perspective objective un point de vue subjectif, comme « condition de compréhension du phénomène lui-même ». La leçon que Loth retient de Nagel est donc que l'articulation n'est pas une connexion « *a priori* » mais une connexion « *a posteriori* » entre le physique et le mental, pour laquelle plaide notamment une attention à la causalité mentale (Loth F., « L'objectif et le subjectif », in *Klêsis*, numéro spécial *Nagel*, n°41, 2018,

Est-ce à dire que l'argument des chauves-souris remettant en question la vertu des neurosciences pour la connaissance de soi ne vaut plus non plus ? Il vaut encore, pour Nagel, car il y a fondamentalement un point qui reste mis en évidence par Kripke et que nous devons garder à l'esprit selon lui : la description physique doit montrer qu'elle *réfère essentiellement* à X. Si la description physique se sert de noms qui ne désignent X que par des propriétés accidentelles, alors ce ne sera pas véritablement une *connaissance* de X. On peut donc préciser χ :

χ : *Les neurosciences peuvent nous donner une connaissance de X si « M » réfère à X et « P » réfère essentiellement à X. (Par hypothèse « M » réfère essentiellement à X).*

Le lien entre ce qui échappe à l'expérience phénoménale et cette expérience phénoménale reste mystérieux – sans qu'on ne sache s'il est mystérieux en raison de nos concepts ou de la nature (comme Nagel le disait déjà en 1974-1979). On ne sait pas pourquoi X pourrait avoir des propriétés P accompagnées de telles ou telles propriétés M, ni si ces propriétés P pourraient être accompagnées d'autres propriétés M, ni d'ailleurs si elles pourraient être accompagnées d'aucune propriété M. Mais on peut définir alors la tâche : il faudra préciser à l'avenir comment les processus mentaux et les propriétés physiques peuvent coexister²⁹. Nagel en appelle à de « nouveaux outils » pour penser l'impossible », à savoir « l'idée que l'esprit est produit par la recombinaison de la matière »³⁰. Ce n'est pas le propos de cet article que d'évaluer quelle notion d'émergence peut constituer un tel outil chez Nagel mais c'est ici qu'une telle discussion pourrait s'insérer.

La transformation du défi chez Putnam

Une autre manière de formuler le défi herméneutique peut être proposée. En effet, on peut trouver une autre proposition, proche de χ , mais qui ne présuppose pas nécessairement que la réalité décrite soit un ensemble d'objets donnés « en soi ». On peut envisager la *coréférentialité* sans réalisme métaphysique.

Énonçons alors cette nouvelle thèse :

ζ (*Coréférentialité*) : *Les neurosciences peuvent nous donner une connaissance de l'esprit si « M » et « P » ont une même référence.*

La tâche devient en quelque sorte un défi de traductibilité³¹. Il suffit de se demander pourquoi ou comment « M » pourrait être *traduit* par « P » ou inversement. Ainsi fait-on l'économie du X que χ supposait.

Pour comprendre que ζ est à la fois porteuse de promesses et liée à un cadre de pensée qu'il sera fécond de renverser, on peut s'appuyer sur trois ouvrages de Putnam manifestant l'évolution de sa pensée sur ce point.

Les arguments de *Raison, vérité et histoire*, publié en 1981, permettaient d'espérer que ζ fournisse une approche évitant les écueils du réalisme métaphysique. Comme Putnam le soutient à l'époque, il est en effet possible de soutenir ζ tout en rejetant la thèse de la vérité-correspondance ou celle d'une adéquation entre nos concepts et une réalité donnée

p. 161-179). Il nous semble, comme nous y reviendrons dans la troisième partie, que l'expérience même du passage d'un point de vue à l'autre, bref l'expérience de la compréhension, est également une évidence *a posteriori* en faveur de cette connexion, même si elle ne nous dit rien sur *la façon* dont l'identité ou la corrélation est possible.

²⁹ Nagel T., *Le Point de vue de nulle part*, p. 61.

³⁰ Nagel T., « The psychophysical nexus », in *Concealment and Exposure and Other Essays*, New York, Oxford University Press, 2002.

³¹ « En quelque sorte » seulement car un frégeén pourrait objecter que la bonne traduction doit non seulement conserver la référence mais aussi maintenir le sens.

car « donner une connaissance » peut s'entendre comme « présenter une description "rationnellement acceptable" en un sens pragmatique ». Ainsi, la proposition ζ ne prétend nullement pouvoir faire correspondre « M » ou « P » à un X déterminé. Elle suppose seulement que « M » et « P » puissent avoir une même référence. Dans cette perspective, ζ ne peut être traduit par χ car χ est pris dans les rets (les *rei*, si l'on nous pardonne ce mauvais jeu de mots) du réalisme métaphysique.

Ainsi, l'hypothèse envisagée par Putnam dans le chapitre 4 de *Raison, vérité et histoire* est que les neurosciences puissent nous donner une description physique rationnellement acceptable de ce qui pourrait également être décrit de façon rationnellement acceptable en des termes mentaux. S'il y a identité, c'est une identité entre deux descriptions qui ne font pas référence à un universel *donné* ou à la propriété *définie* d'un événement métaphysique individuel³². Ces deux descriptions, toutes deux rationnellement acceptables selon Putnam, font référence à ce que nous tenons pour un fait *d'après notre usage du monde* et qu'elles soient toutes deux rationnellement acceptables ne nous renseigne pas tant sur ce qu'est le monde en soi que sur notre conception de la réalité – en l'espèce non tant sur ce qu'est l'esprit en soi que sur notre conception de l'esprit.

Le défi herméneutique se modifie donc un peu à la lumière de *Raison, vérité et histoire* : que nous puissions ou non rendre raison du fait que ces descriptions aient une même référence, le fait même que nous les acceptions pour rationnelles (voire que nous acceptions rationnellement qu'elles aient une même référence) nous renseigne sur notre conception de l'esprit.

Reste qu'il faut s'acquitter de la tâche sémantique³³. Or, l'un des acquis de cette réflexion en 1981 est d'écarter une manière de le faire : Putnam établit que la référence de « P » ne peut pas renvoyer à une chose donnée qui aurait M pour corrélat, et que la référence de « M » ne peut renvoyer à une chose donnée qui aurait P pour corrélat. Il développe deux argumentations, l'une montrant que la propriété cérébrale correspondante n'est pas discernable par elle-même (pas « autoidentifiante ») et l'autre montrant qu'un *quale* n'est pas non plus bien défini par lui-même. Nous nous attacherons à la première³⁴. Elle met en effet en jeu, comme on va le voir, un argument d'indiscernabilité qui a une portée qui nous intéresse immédiatement ici puisque l'une de ses conséquences est de brouiller la différence anthropologique entre l'homme et les autres animaux, et donc de faire peser le doute sur la vieille délimitation impliquée par la superposition entre connaissance de soi et connaissance de l'homme.

L'argumentation est la suivante³⁵. En raison des phénomènes de vicariance, de plasticité et de réalisabilité multiple, la propriété ou le corrélat physique de ce à quoi réfère un sujet

³² Putnam H., *Reason, Truth and History*, Cambridge/New York /Melbourne, Cambridge University Press, 1981, p. 101, trad. fr. A. Gerschenfeld, Paris, Les éditions de Minuit, 1984, p. 115-116.

³³ Dans *Raison, vérité et histoire*, Putnam soutient que l'identité entre la description « avoir la température T » et « avoir l'énergie cinétique moléculaire E » est non point synonymique mais synthétique car les physiciens ne pourraient pas utiliser le second prédicat dans tous les contextes où le premier prédicat est employé et inversement, il y a manifestement des conditions d'usage différentes qui rendent pertinent chacun de ces prédicats. La question est alors de savoir si « M » et « P » peuvent être *synthétiquement* identiques. Ce que vise à montrer Putnam en 1981, c'est qu'on ferait une grave erreur si on cherchait à répondre à cette question en continuant à souscrire à un réalisme métaphysique.

³⁴ Quant à la remise en question du fait que le *quale* soit déterminé en lui-même, Putnam en donne trois principaux arguments dans *Raison, vérité et histoire*. *Primo* la description de notre expérience phénoménale est elle-même dépendante de conceptualisation. *Secundo* on ne peut pas dire que le *quale* déforme plus ou moins les faits physiques, ou inversement que les faits physiques soient une version artificielle des *qualia*. *Tertio* on ne peut pas savoir si la variation des *qualia*, par exemple quand on regarde une nappe rouge à deux moments de la journée, indique qu'il y a un seul *quale* qui change ou qu'il y a différents *qualia*.

³⁵ Putnam H., *Reason, Truth and History*, p. 85-92, trad. fr., *Raison, vérité, histoire*, p. 99-106.

qui dit « je vois une carte rouge » doit nécessairement être décrite dans une disjonction. Ainsi, même un scénario de dissociation hémisphérique du type *Split Brain* ne nous permettrait pas d'isoler le corrélat. Mais c'est là qu'intervient l'argument qui nous intéresse, cette fois sur *les théories* à avancer. Selon Putnam, même si l'on admet que la propriété cérébrale est disjonctive, pour discerner une telle propriété il faudrait déjà avoir résolu un problème portant sur la possibilité de trancher entre deux théories concernant la discernabilité de la propriété. C'est ce point qui retient notre attention. Pour le montrer, il ajoute une étape de science-fiction à l'expérience par la pensée inspirée du *Split brain* : on recoud les deux hémisphères et on demande au sujet *s'il percevait* la carte rouge quand les hémisphères étaient dissociés. Que conclure s'il répond affirmativement ? Dans ce cas, deux théories peuvent s'affronter. Appelons « Théorie A » celle qui prétend identifier la propriété, et « Théorie B » la théorie sceptique.

Un partisan de la Théorie A dira : « le sujet dit qu'il voyait du rouge quand son cerveau était coupé en deux, il voyait donc du rouge ; par conséquent la sensation correspond à une propriété du lobe droit (séparé du lobe gauche responsable du langage) ».

Un partisan de la Théorie B lui répondra « non, on ne peut pas savoir car la réponse présente du patient dont le cerveau est recousu est une activité cérébrale qui ne nous informe en rien sur ce qu'il *voyait* quand son cerveau était coupé en deux ».

Pour départager ces deux théories, donc pour établir *si on peut ou non savoir* si le sujet voyait du rouge, on se demande si le lobe droit (sans le langage) *suffit à avoir une expérience consciente*. On peut par exemple – et c'est ce qui retient notre attention – demander si le cerveau des animaux qui ne parlent pas suffit à leur donner une expérience consciente³⁶.

À l'appui de la Théorie A, qui identifie le fonctionnement du lobe droit avec la sensation de rouge, on peut donc convoquer une thèse sur les animaux : les animaux ont un cerveau sans langage et pourtant ils ont bel et bien une sensation correspondant au fonctionnement de ce cerveau.

À l'appui de la Théorie B sceptique, on peut trouver une thèse inverse sur la différence anthropologique : ce fonctionnement du lobe droit (ou fonctionnement du cerveau animal), ce n'est pas de la conscience. Une version moins « chauvine » (c'est-à-dire admettant que l'homme n'est pas le seul être doué de conscience) dira plutôt : à tout le moins, ce n'est pas la même conscience *que nous*³⁷.

Quoi qu'il en soit, ces deux théories sont selon Putnam « observationnellement indiscernables ». L'on ne peut pas sortir de notre conscience³⁸ pour voir si un animal voit comme nous, pour voir si la sensation correspond à une propriété du seul lobe droit, ou pour voir maintenant qu'on est recousu (!) si tout à l'heure (quand notre cerveau était coupé en deux) on voyait du rouge.

Pour décider entre ces théories, Putnam pense à l'époque où il écrit *Raison, vérité et histoire*, que la seule issue est de « légiférer », c'est-à-dire, par convention, fixer l'extension de la conscience³⁹. Ainsi un matérialisme fonctionnaliste permettra de satisfaire la demande de coréférentialité. Mais son évolution ultérieure va l'amener à changer sa position. À partir de la fin des années 80, en approfondissant les problèmes de la

³⁶ Ainsi, l'argumentation de Putnam articule trois lignes sceptiques tracées par Nagel dans ses chapitres de 1979 sur les *Split Brain*, la chauve-souris et le panpsychisme (c'est-à-dire l'étendue des phénomènes conscients dans la nature).

³⁷ Un argument en la faveur de cette théorie non-chauviniste fera valoir que le fonctionnement du cerveau animal est *intégré* : chaque fonction se fait *en fonction* aussi de *l'ensemble*, de l'organisme total ; par conséquent elle n'est en aucun cas semblable au fonctionnement du seul hémisphère droit.

³⁸ Le Putnam de cette époque entend par là qu'on ne peut pas sortir du fonctionnement de notre cerveau.

³⁹ Putnam H., *Raison, vérité et histoire*, p. 106.

coréférentialité déjà présents dans *Raison, vérité et histoire*, Putnam donne en effet plusieurs assauts contre ζ ⁴⁰. Sa critique de la *corrélation* culmine dans *La triple corde*. Il y rejette non tant la thèse de la corrélation psychophysique que « l'idée que la question ait un sens »⁴¹. Car selon lui cette question suppose une image faussée de nos traits psychologiques, à savoir : « cette image de nos caractéristiques psychologiques comme "états internes" qui, en tant qu'états internes, doivent être ou bien "corrélés" ou bien "non corrélés" avec ce qui se produit à l'intérieur de nos corps (au sens littéral d'intérieur)⁴². »

Cette image est trompeuse. La question même à laquelle répondent le réductionnisme ou l'antiréductionnisme manque d'intelligibilité parce qu'elle suppose l'idée du mental comme ensemble d'états psychologiques internes plutôt que du mental comme ensemble de *capacités*.

Cette critique de la notion d'état psychologique interne conduit à une critique des tentatives d'*individuation* de ce à quoi réfère « M » par des critères « internes ». Selon Putnam « nos caractéristiques psychologiques sont individuées de manières qui sont sensibles au contexte et extrêmement complexes, impliquant des facteurs externes (la nature des objets que nous percevons, auxquels nous pensons, et sur lesquels nous agissons), des facteurs sociaux ainsi que les projections que nous trouvons naturel ou pas de faire⁴³ ». Parallèlement, il est impossible d'identifier un corrélat neurocérébral ou un processus minimal unique dans le cerveau qui serait responsable d'un comportement compris comme témoignant d'une croyance du sujet⁴⁵.

Poser l'idée d'état psychologique interne c'est poser une idée qui *part* de la décorrélation. Il n'est pas étonnant selon Putnam qu'ensuite cette idée ouvre la voie à une certaine forme de scepticisme (que Putnam qualifie de berkeleyen), le scepticisme qui soutient « qu'il est impossible de comprendre comment les expériences perceptives pourraient être dans une relation de corrélation unique avec des objets qui sont "à l'extérieur"⁴⁶ ». Cet effet sceptique (dont Putnam dit en toute honnêteté qu'il est imputable aussi à son application de l'argument modèle-théorique) va donc de pair avec *le fait même de poser la question* de savoir si « M » réfère à P, ou si « M » et « P » ont une même référence – dès lors que l'on admet que « M » a une signification psychologique interne, c'est-à-dire qu'elle nomme mon expérience phénoménale présente comme « douleur » nomme une apparence qui, présente, ne peut pas manquer d'être réelle, si l'on se souvient de Kripke⁴⁸. Putnam soupçonne même que « la métaphore de l'émergence est mauvaise » tant qu'elle vient répondre à un faux-problème.

En suivant l'évolution de Putnam, il nous faut donc redéfinir le défi herméneutique. Il pouvait initialement être résumé ainsi : *pour que les neurosciences soient connaissance de*

⁴⁰ Dans *Représentation et réalité*, Putnam considère en particulier la relation d'équivalence que cherche à établir un fonctionnaliste pour traduire « M » en langage computationnel. Les problèmes dus à la réalisabilité multiple et à l'indiscernabilité d'un corrélat typique se reposent. Il n'y a aucune raison de penser qu'il n'y a qu'un seul et même état physique et computationnel entre différents organismes croyant qu'il y a un chat dans le voisinage. Mais l'équivalence entre *token* mentaux et physiques est également mise à mal, car il n'y pas davantage de raison de penser qu'il n'y a qu'un seul et même état physique ou computationnel chez un seul organisme ou un seul être humain à des moments différents (H. Putnam, *Représentation et réalité* (1988), trad. fr. C. Engel-Tiercelin, Paris, Gallimard, p. 168).

⁴¹ Putnam H., *La Triple corde* (1999), trad. fr. P. Fasula et al., Paris, Vrin, 2017, p. 183.

⁴² *Ibid.*

⁴³ *Ibid.*, p. 184.

⁴⁵ *Ibid.*, p. 152-153

⁴⁶ *Ibid.*, p. 228.

⁴⁸ De cette critique, découle donc naturellement la remise en cause de la thèse selon laquelle la douleur est un état psychologique interne.

soi, il faut qu'elles nous donnent une connaissance de l'expérience phénoménale et pour cela il faut que notre description ou nos termes phénoménaux aient une référence physique ou, tout au moins, une référence à ce à quoi réfère les termes des neurosciences. Désormais ces deux voies sont coupées. Quelle autre voie est envisageable ?

3. Les conditions d'intelligibilité des neurosciences

Une troisième voie permettrait de définir un défi herméneutique un peu différent⁴⁹. Le discours neuroscientifique est *compréhensible*, c'est un fait. Partons de ce fait, et demandons-nous plutôt : *quelles sont les conditions d'intelligibilité* du discours neuroscientifique ? Il a certainement un sens par sa cohérence et sa référentialité propre. C'est ce sens que les « aliens » sont parfois supposés comprendre dans les scénarios fictifs où, dépourvus de notre expérience phénoménale (ou de notre croyance en cette expérience phénoménale), ils prendraient connaissance de notre science. C'est ce sens que nous comprenons en lisant des études sur la physiologie des chauves-souris et des macaques. On peut d'ailleurs étudier les conditions d'intelligibilité de ce sens de différentes manières (en prêtant attention aux jeux de langage, etc.). Mais de fait, les études de neurophysiologie humaine se comprennent comme parlant *de nous*. D'aucuns diront qu'elles le font partiellement, mal ou de façon insatisfaisante. Certes, mais alors qu'exige-t-on des neurosciences pour qu'elles nous parlent de nous ? Faut-il d'abord résoudre la question sémantique pour comprendre qu'elles parlent de nous ? On a remarqué au début de notre enquête que traiter le problème de la conscience, même lorsque l'on est illusionniste, permet de comprendre que des descriptions neuroscientifiques procurent une intelligibilité de notre expérience (entendue comme *ce que notre cerveau nous représente comme expérience*, si l'on est illusionniste). Le réquisit indispensable n'est donc pas de résoudre au préalable la question sémantique, mais d'exiger que la compréhension de ce que disent les descriptions neuroscientifiques procure une intelligibilité notre expérience. Or cette exigence peut être remplie. Du simple fait que l'on *comprend que* les descriptions neuroscientifiques *réfèrent* à notre expérience, elles nous apprennent quelque chose sur nous. On se souvient que Bonjour mettait déjà en évidence ce fait dans l'acte de publication et de lecture. On a également rappelé que *le changement de point de vue*, dans la théorie du double aspect, *nous est de fait accessible*. Dès lors il ne faut pas d'abord tenter de résoudre la question sémantique pour prouver ensuite que le discours des neurosciences peut avoir du sens pour nous. Il faut partir du *fait* qu'il est possible de *se comprendre* mieux au travers de ces descriptions pour s'interroger sur les conditions de ce (formidable) fait. Il faut interroger les conditions d'intelligibilité des neurosciences, en constatant que cette intelligibilité est un fait, non seulement *parce qu'on comprend ce qu'elles disent*, mais parce qu'on comprend *que leur discours* peut signifier quelque chose pour nous, sur nous.

Sans doute ne peut-on pas se contenter de renverser ϕ en ψ et attendre de trouver un moyen de faire « P » référer à des états psychologiques internes :

ψ : *Les neurosciences donnent une connaissance de M si « P » réfère à M.*

Ce serait retomber dans les pièges dénoncés par Putnam. Non seulement il faut renoncer à identifier terme à terme le mental et le physique, mais il faut renoncer à suspendre à une telle identification terme à terme la contribution des neurosciences à une connaissance de soi. La solution au problème esprit/corps n'est pas le fondement sans

⁴⁹ Il nous semble que c'est une voie empruntée par exemple par Denis Forest dans son livre *Neurosepticisme*, quand il analyse les expériences de Spuznar sur la mémoire (*Neurosepticisme. Les sciences du cerveau sous le scalpel de l'épistémologue*, Paris, Ithaque, 2014, p. 123).

lequel il resterait douteux que les neurosciences nous éclairent sur notre condition de vie humaine. C'est au contraire en constatant que d'aucuns explorent ce qu'elles peuvent nous apprendre sur nous, que le problème du rapport esprit/corps s'éclaire (et ce, comme on va le voir, dans la mesure où ces explorations, par leurs gains théoriques et thérapeutiques, renforcent l'articulation entre neurosciences et psychologie).

Le nœud entre la compréhension sémantique et la compréhension de soi reposerait donc sur une conviction évidente mais fondamentale pour comprendre qu'il y ait des gens qui s'engagent dans l'activité scientifique : les descriptions des neurosciences nous apprennent quelque chose sur le fait de vivre comme nous le faisons, et sur le fait de vivre comme le font les chauves-souris. Suivons donc ceux qui tentent l'aventure.

Akins

Kathleen Akins a précisément tenté de comprendre ce que la science peut nous dire de *what it is like to be a bat*. Son article s'ouvre sur une déclaration qui semble relever le défi de Nagel, mais qui, selon nous, en change profondément les termes : « Les faits scientifiques, semble-t-il, peuvent nous dire quelque chose à propos de l'expérience consciente, bien qu'il ne soit pas clair de savoir exactement ce que la science peut dire⁵⁰. » « Semble-t-il » exprime ici une conviction *prima facie*. Akins ne répond pas au défi ϕ (trouver comment « M » peut référer à P), mais se donne pour tâche de comprendre en quoi les descriptions neuroscientifiques procurent une intelligence de notre expérience. Elle propose de laisser de côté la question de savoir comment une conscience physiquement incarnée est possible (donc le problème de l'*explanatory gap*) et bien plutôt de suivre l'intuition selon laquelle « la science a quelque chose à dire » et « simplement se demander ce que cela pourrait être »⁵¹. Assurément il ne s'agit pas pour la science de donner l'expérience de ce que cela fait d'être une chauve-souris, mais de procurer un savoir sur à *quoi cela ressemble (what it is like) de vivre en chauve-souris*.

Un point de départ assez naturel est de chercher les types d'informations sensorielles qui sont nécessaires aux chauves-souris pour leur survie. Puisque l'écholocation semble être analogue à la vision dans la mesure où elle donne des informations spatiales, il semble naturel de comparer la fonction du système auditif des chauves-souris à celui du système visuel humain. C'est ce qui permettra de mettre en évidence les conditions du traitement d'information chez les chauves-souris. En réalité, Akins va remettre en question l'idée que l'écholocation soit un véritable analogue du système visuel car elle pense que l'écholocation ne fournit pas une véritable représentation (image ou conception) d'un monde d'objets.

Notre propos n'est pas de discuter cette thèse. Un des problèmes est d'ailleurs de savoir s'il peut y avoir une expérience phénoménale sans représentation *de*, conception *de* ou intentionnalité. Nous nous bornerons à relever les conditions qui permettent selon elle d'utiliser les données scientifiques pour comprendre qu'elles éclairent l'expérience phénoménale possible de la chauve-souris. En l'occurrence, elle *part* d'une analogie que nous faisons naturellement quand nous cherchons à comprendre les descriptions du système perceptif par écholocation : la chauve-souris perçoit l'espace par un système auditif comme nous le percevons par la vue. La comparaison des conditions de perceptions spatiales par la réception de la lumière avec celles d'une perception par le son permettent de mettre en évidence un certain nombre de problèmes auxquels ce dernier doit faire face. Par exemple, les sons émanent de multiples directions (et pas

⁵⁰ Akins K., « What Is It Like To Be Boring and Myopic ? », in *Dennett and His Critics*, Oxford, Blackwell, 1993, p. 125, nous traduisons. Nous remercions Denis Forest d'avoir attiré notre attention sur cet article.

⁵¹ *Ibid.*, p. 127.

seulement d'une seule source comme le soleil) et c'est pourquoi un organisme qui utilise le son pour acquérir des informations spatiales devra plutôt produire son propre son, et attendre qu'il revienne – mais alors il *doit faire face à des problèmes d'interférences*. Akins fait une revue de la littérature éthologique et neuroscientifique disponible et en infère quelques caractéristiques du champ perceptif d'une espèce de chauve-souris particulière, le murin à moustaches. Par exemple, le champ « auditif » de cette chauve-souris est un analogue de la fovéa rétinienne (cette aire centrale de la rétine qui est densément tapissée de récepteurs produisant une image du monde en haute résolution). Mais à la différence du champ visuel humain qui s'affaiblit à la périphérie, le champ perceptif du murin à moustaches aura un grand nombre de régions concentriques d'intensité peu à peu déclinantes (des régions concentriques très délimitées). Plus le murin s'approche de sa cible, plus la profondeur de champ se rétrécit. C'est pourquoi Akins parle, par analogie, de « myopie »⁵². Mais elle indique aussi les limites de ces analogies en affirmant par exemple que les chauves-souris n'ont pas de « point de vue », objectivant, *sur* le monde.

En refermant l'article de Akins, a-t-on progressé dans la question de savoir *ce que cela fait d'être une chauve-souris* ? Assurément on n'a pas du tout traité des questions telles que *pourquoi il se fait que certains événements neuronaux comme les perceptions visuelles sont présents à la conscience alors que d'autres ne le sont pas*. On ne peut pas savoir si le champ perceptif acquis par l'écholocation d'une chauve-souris particulière est suffisant pour qu'elle ait une expérience phénoménale. Mais les résultats de la recherche supposent d'admettre que les neurosciences procurent une forme d'*intelligence* à partir d'une compréhension de soi : « We could estimate, for example, how far the bat could possibly "see" at different points in the hunt, the spatial resolution of the acoustic field, and how "things" would "disappear" from view under certain conditions⁵³. »

En lisant les descriptions neuroscientifiques, Akins leur donne une signification à partir de ses propres aptitudes humaines. Elle manifeste cette capacité de lecture en utilisant, en retour, une écriture qui croise les registres sensoriels et qui joue du langage sans non-sens. Une chauve-souris ne « voit » pas des « choses qui apparaissent ou disparaissent à sa vue ». Mais nous comprenons qu'elle fait quelque chose d'analogue à cela.

En somme, la condition de la lecture intelligente d'un article neuroscientifique est que *l'on sait qu'il y a quelque chose à chercher* (pour parler comme Moore cité par James⁵⁴). Mais comment sait-on qu'il y a quelque chose à chercher, que ce soit dans les descriptions neuroscientifiques portant sur les êtres humains ou sur les autres espèces ? Il semble que la réponse tient, par-delà une *imagination* d'expériences phénoménales possibles, dans *le fait de la compréhension*. Après tout, notre imagination est débordante et peu exigeante. Au contraire, le fait de la compréhension rend impérieux de mieux comprendre. Ainsi, peut-on tenir les comportements des autres animaux pour compréhensibles ; même s'ils ne sont pas encore *bien* compris, ou si on ne voit pas bien leur sens, on les tient pour *compréhensibles* et l'on cherche à comprendre comment les descriptions neuroscientifiques peuvent les éclairer.

⁵² Akins conteste par ailleurs l'idée que la chauve-souris ait une *représentation*. Selon elle, il n'est besoin d'aucune étape *représentationnelle* entre l'information auditive du cortex primaire et l'activité motrice. Il nous semble que sur ce point l'écholocation du murin à moustaches pourrait être une sorte d'analogue de la vision sans perception mise en évidence par Milner A.D. et Goodale M.A., *The Visual Brain In Action*, Oxford, Oxford Psychology Series, 2006.

⁵³ Akins K., « What Is It Like To Be Boring and Myopic ? », p. 152.

⁵⁴ James W., « La conscience existe-t-elle » (1904) trad. fr. in *Essais d'empirisme radical*, Paris, Champs Flammarion, 2005, p. 38 (commenté par G. Bearn, *op. cit.*).

Une autre compréhension de la corrélation entre neuroscience et psychologie

La lecture de Putnam permet de définir deux contraintes pesant sur la manière de comprendre le sens herméneutique que peuvent avoir les neurosciences. Il faut éviter de présupposer un état mental interne déterminé pour tenter seulement ensuite de le corrélérer avec des événements décrits par les neurosciences⁵⁵. Il faut également tenir compte de la plasticité cérébrale et de la réalisabilité multiple. Comment alors comprendre les travaux en psychologie cognitive qui font varier les déclarations, les témoignages et les comportements d'un sujet et observent en parallèle l'activité cérébrale au travers de l'IRMf ? Faut-il, pour accepter qu'ils aient un sens, revenir à une thèse sémantique de type ϕ , ψ ou ζ ? Peut-on corrélérer les descriptions neurophysiologiques et les descriptions psychologiques en se passant d'une corrélation psychophysique ?

Il nous semble que dans ce cadre la thèse d'un *corrélât neuronal de la conscience*, avancée de différentes manières par Christopher Koch, Ned Block ou Stanislas Dehaene, doit retenir notre attention. À quelle condition peut-elle nous apprendre quelque chose sur nous ? Comment, compte-tenu des critiques précédentes, éviter d'en faire un non-sens ? Notre but ne sera pas de discuter les arguments *scientifiques* en faveur ou contre l'existence d'un tel corrélât. Nous nous demandons plutôt quelles conditions d'intelligibilité peuvent avoir les discours récents qui parlent d'un tel corrélât.

On peut constater que, sous la plume de ces auteurs, la thèse du corrélât neuronal de la conscience prétend bel et bien contribuer à une connaissance de soi. Koch, par exemple, prétend nous parler de *nous*, de ce qui se passe à l'état de veille ou dans des états végétatifs, ou dans une anesthésie. Une telle prétention présuppose-t-elle encore ϕ , ψ ou ζ ? Koch s'en tient à dire que l'activation neuronale et l'état subjectif sont *signes* l'un de l'autre⁵⁶. Mais comment penser cette signification en évitant le scepticisme du premier Nagel et les écueils pointés par le dernier Putnam ? Il en va de la corrélation entre le discours physiologique, interprétant notamment les images de l'IRMf, et le discours psychologique, décrivant un comportement ou rapportant les propos d'un sujet.

Notre hypothèse est que la corrélation psychophysique n'est pas *ce qui doit être établi* au préalable pour que la compréhension du discours neuroscientifique soit possible, mais ce qui devient plus intelligible au travers de l'activité scientifique qui manifeste cette compréhension.

Un article publié par Ned Block en 2007 nous indique des pistes pour éclairer cette question. Ned Block est comme l'on sait convaincu de la pertinence de la notion de *qualia*. Il est également fervent défenseur de la distinction entre une conscience *phénoménale* et une conscience *d'accès*. Sur ces deux points il semble que sa position nous reconduise à *présupposer* un état subjectif déterminé pour le corrélérer à un état physiologique. Pourtant, dans cet article intitulé « Consciousness, accessibility, and the mesh between psychology and neuroscience », c'est un autre argument qui nous semble crucial, un argument qui dit que *tout ce qui vient à l'appui d'une corrélation entre neurologie et psychologie* peut faire l'objet d'une inférence à la meilleure explication⁵⁷. Il nous semble qu'il y a là un

⁵⁵ Putnam s'appuie sur les descriptions neuroscientifiques qui mettent les phénomènes cérébraux en continuité avec les phénomènes organiques et les phénomènes physiques extérieurs, que ce soit en amont (de sorte qu'une théorie de la perception visuelle par exemple s'inscrit dans la continuité d'une théorie de la lumière) ou que ce soit en aval (de sorte qu'une théorie du comportement ne fait pas abstraction de l'empirie dans laquelle il intervient). Il nous semble qu'une psychologie articulant particulièrement bien ces deux versants se trouve dans la théorie écologique de la perception inspirée par James J. Gibson.

⁵⁶ Koch C., *The Quest for Consciousness: A Neurobiological Approach*, Englewood, Roberts & Company Publishers, 2004, p. 17.

⁵⁷ Block N., « Consciousness, accessibility, and the mesh between psychology and neuroscience », in *Behavioural, and Brain Sciences* (2007), 30, p. 481-548 (commentaires compris).

renversement méthodologique en faveur d'une approche plus pragmatique. La méthode ne consiste pas à *établir* une corrélation entre M et P pour ensuite être autorisé à affirmer que les neurosciences contribuent à une connaissance de soi. Elle consiste, *dans la pratique* neuroscientifique qui croise naturellement les observations neurophysiologiques et les expériences psychologiques dans des buts de recherches thérapeutiques, à retenir toutes les hypothèses qui *renforcent* et éclairent l'articulation entre neurophysiologie et psychologie.

Dans cet article, Block soulève une énigme sur un cas psychopathologique : le syndrome de l'extinction spatiale visuelle. Un patient atteint de ce syndrome déclare voir un objet (par exemple un visage) quand on le lui présente à sa droite ou à sa gauche. Mais lorsqu'on lui présente *deux* objets (deux visages), un de chaque côté, il déclare voir seulement celui qui est à sa droite. Ainsi, lorsqu'il y a compétition entre la vision de droite et la vision de gauche, il prétend ne pas voir l'objet de gauche. Pourtant, même dans ce second cas, l'IRMf révèle une activité dans l'aire occipito-temporale (dite « aire fusiforme » des visages) presque aussi grande que dans le premier cas (lorsqu'il disait voir le visage à gauche et que rien n'était présenté à droite). L'énigme est donc : ce patient a-t-il ou non l'expérience visuelle du visage à gauche qu'il déclare ne pas voir ?⁵⁸ Block pense que sa distinction entre conscience phénoménale et conscience d'accès permet de résoudre l'énigme. Nous ne considérerons pas ici le bien-fondé de cette thèse. Ce qui nous intéresse est *la justification épistémologique* qu'il en donne. L'hypothèse soutenue par Block est que le patient n'a pas accès à sa conscience phénoménale, raison pour laquelle il ne déclare pas voir un visage à gauche, mais que cette conscience phénoménale est bien corrélée à l'activité attestée par l'IRMf ; en réalité, selon Block, le corrélat neuronal de l'accès cognitif (conscience d'accès) n'est pas le même que le corrélat neuronal de l'expérience phénoménale (conscience phénoménale). C'est exactement sur ce dernier point que Block est en discussion avec Stanislas Dehaene et ses collègues. Le point remarquable pour nous est la façon dont Block justifie son hypothèse. Selon lui, non seulement elle éclaire notre expérience de la conscience, mais elle se donne les moyens de penser comment la vie des personnes en état végétatif persistant peut avoir une richesse phénoménale ou expérientielle⁵⁹. En outre, selon lui un crédit doit être accordé à toute thèse qui permet de mieux articuler psychologie et neurologie. Son explication n'a pas de preuve expérimentale directe : on ne peut pas observer sur le patient la corrélation entre l'expérience phénoménale et son corrélat neurocérébral, car lorsqu'il dit voir, il a par définition aussi accès et lorsqu'il dit ne pas voir, le scientifique observe uniquement l'activation neurocérébrale. La corrélation de la pure conscience phénoménale et de son

⁵⁸ Block montre qu'il y a deux réponses corrélationnistes simples qui sont également insatisfaisantes. Appelons-les « CorrA » et « CorrB ». CorrA soutient que dans la seconde situation, le patient n'a pas l'expérience visuelle, ce qui implique que l'activité attestée par l'IRMf n'a aucun rôle dans l'expérience visuelle. CorrB prétend (comme la Théorie B sceptique de Putnam que nous présentions plus haut) qu'on ne peut pas savoir. Corr A suppose qu'on a mal identifié la base neuronale et CorrB suppose qu'on ne peut pas savoir si le patient a une expérience du visage et que c'est scientifiquement indétectable. Block identifie également les figures de ces deux réponses : du côté de la première réponse on trouverait F. Crick et C. Koch ; du côté de la seconde on trouverait les auteurs développant l'argument d'indiscernabilité, en l'occurrence le Putnam de *Raison, vérité et histoire* et D. Dennett. Pour qu'une solution soit trouvée, qui permette de corréler physiologie et psychologie, il faut, selon Block, distinguer conscience phénoménale et conscience d'accès.

⁵⁹ Block convoque certes aussi des évidences expérimentales en faveur d'un *débordement* de la conscience phénoménale sur la conscience d'accès (les expériences de George Sperling (1960) et de Landman et al. (2003)). Mais en réalité ces évidences expérimentales sont discutées (notamment par Dehaene, mais également par un ensemble de neuroscientifiques et de philosophes qui ont publié leurs commentaires à la suite de l'article et auxquels Block répond à la fin).

seul corrélat ne peut être observée. Mais aux yeux de Block, cette explication se justifie *par abduction* : non seulement elle suffit à expliquer les comportements du patient, mais elle offre un gain théorique (explication d'autres phénomènes), épistémologique (explication de l'articulation entre psychologie et neurologie) et philosophique (explication de la vie mentale, ayant des conséquences éthiques comme le respect pour les personnes en état végétatif). Le gain épistémologique touche à l'intelligibilité de la démarche neuroscientifique. L'argument implicite qui justifie l'emploi de cette inférence à la meilleure explication nous paraît être le suivant : *on comprend que la démarche neuroscientifique est significative à l'égard de notre expérience psychologique, sinon on ne s'y engagerait tout simplement pas. Si tel est le cas, alors toute explication qui vient à l'appui d'une corrélation entre neurologie et psychologie a une probabilité supérieure aux autres.*

Ce genre d'utilisation, par Ned Block, de l'inférence à la meilleure explication a été saluée par Putnam à la fin de sa vie⁶⁰. Cela pourrait paraître étrange si l'on se souvient de la méfiance exprimée par ce dernier à l'encontre de l'idée de *qualia* déterminés. Putnam enjoignait de ne pas y voir des états subjectifs internes bien définis qui attendraient d'être identifiés ou corrélés avec des états physiques déterminés. Mais il nous semble que la démarche abductive indiquée par Block, celle qui privilégie tout ce qui renforce la corrélation entre neuroscience et psychologie, peut être compatible avec une approche pragmatiste inspirée de *La triple corde*.

Putnam y développait l'argument suivant. Parfois les épistémologues considèrent que voir *ce qui est observable* et voir *des choses trop petites pour être observées* sont deux activités différentes, répondant donc à des règles très différentes. Pourtant le sens commun, souligne Putnam, ne fait pas la différence. Putnam emprunte à Cora Diamond une analogie : c'est un peu comme voir dans deux portraits le même visage (*the same face*)⁶¹. Notre suggestion sera donc la suivante : et si *voir* dans les descriptions neuro-cérébrales et les descriptions psychologiques un « même visage » était précisément le trait herméneutique de la science contemporaine de l'esprit, « le sens que ces pratiques revêtent pour nous » comme dirait Putnam ? Cette capacité à *voir un même visage* dans des descriptions scientifiques élaborées n'est certainement pas vulgaire ou triviale, mais c'est pourtant ce qui fait que l'activité scientifique est une compréhension de soi – c'est-à-dire de soi par tout un chacun. C'est une capacité rendue possible par le « sens commun » entendu, comme Putnam le dit dans *La triple corde*, comme une capacité à reconnaître, au travers de jeux de langages différents, un même *visage (face)* de la cognition – une capacité qui tient à l'exercice même de la cognition. Si est première la capacité à *se* comprendre dans le langage des neurosciences comme dans le langage phénoménal, alors l'effort pour élucider d'une corrélation psychophysique en *découpe*, corrélation qui peut-être expérimentalement testée tout en prêtant attention à une indétermination (un « vague ») irréductible des termes⁶².

⁶⁰ Putnam H., « Comments on Ned Block », commentaire sur l'article de Ned Block, « Wittgenstein and qualia », *Reading Putnam*, in M. Baghramian (éd.), Oxon & New York, Routledge, 2013.

⁶¹ C'est d'ailleurs par là que Putnam s'oppose à toute idée d'« *interface* » dans la perception pour défendre un véritable réalisme naturel de sens commun.

⁶² Ainsi la discussion autour de la thèse sémantique de David Papineau a pu se concentrer sur un exemple qui nous semble mettre en évidence la priorité de la compréhension sur l'élucidation sémantique. Selon Papineau, le concept phénoménal a seulement une phénoménologie différente du concept physique mais a une même référence ; pour simplifier, « rouge » et « telle activation en V4 » ont une même référence, mais l'usage de ces termes ne fait pas le même effet quand j'en parle parce qu'utiliser « rouge » c'est réactiver une expérience actuelle ou imaginaire de rouge (Papineau D., « Phenomenal and Perceptual Concepts », in *Phenomenal Concepts and Phenomenal Knowledge : New Essays on Consciousness and Physicalism*, in T. Alter et S. Walter (éd.), Oxford, Oxford University Press, 2007). Mettons maintenant que ce ne soit pas moi mais

Conclusion : connaissance de soi et connaissance de l'homme

Pour conclure, demandons-nous en quoi *se comprendre en tant qu'on est l'organisme que l'on est* serait *se comprendre en tant qu'homme*. Le premier Nagel pourrait répondre que les difficultés même de cette compréhension révèlent l'absurdité existentielle de notre condition humaine. Quant à Putnam, il a montré qu'on ne peut pas savoir s'il s'agit bel et bien de se comprendre en tant qu'homme par différence avec les autres espèces (argument d'indiscernabilité du corrélat en raison d'une indécidabilité entre théories sur la spécificité anthropologique de la conscience).

À tout le moins, pour mettre en évidence la façon dont les neurosciences permettent de se comprendre en tant qu'on est l'organisme qu'on est, il n'est pas nécessaire de postuler une définition essentialiste ni de l'homme ni de l'esprit. Il n'est pas non plus nécessaire d'établir en préalable une identité terme à terme du physique et du mental. Le défi herméneutique des neurosciences ne suppose pas de faire une exégèse mentaliste de leurs propos. Prenons acte de l'intelligibilité *prima facie* du discours neuroscientifique sans la fonder sur une interprétation mentaliste de ses termes. Cette intelligibilité est pour nous, de fait, indissociable des conditions de vie humaines dans lesquelles le discours prend un sens pour nous⁶³. C'est ce qui lui donne sa portée « herméneutique »⁶⁴. Ainsi l'articulation entre neuroscience et psychologie n'est pas une prémisse d'où déduire le sens du discours neuroscientifique, mais bien plutôt la conséquence de notre compréhension du sens que ce discours peut avoir *pour nous*, compte-tenu de l'organisme que nous sommes, et des pratiques scientifiques et thérapeutiques dans lesquelles nous nous engageons.

Ainsi, l'intelligibilité des neurosciences vient d'un usage qui articule les descriptions neurophysiologiques et cognitives notamment pour des raisons pragmatiques (inférence à la meilleure explication ; efficacité clinique)⁶⁵. En outre, la différence spécifique entre organismes n'empêche pas qu'on comprenne, à partir de notre propre expérience de cette intelligibilité, qu'elles puissent aussi éclairer à nos yeux l'expérience des autres espèces⁶⁶.

mon frère qui ait l'expérience de rouge. Ai-je besoin de savoir ce que fait à mon frère l'expérience de rouge, en réactivant ou imaginant son expérience, pour comprendre que « l'expérience de mon frère est une expérience de rouge » ? Cette question reçoit diverses réponses examinées par F. Kammerer (*Conscience et matière, op. cit.*, p. 272-284). Ce que nous remarquons ici c'est qu'une voie pragmatiste inspirée de Wittgenstein et Cavell pourrait faire valoir que la compréhension de l'expérience d'autrui ne requiert aucune fondation épistémologique préalable : c'est seulement à *partir du fait de cette compréhension* que l'analyse sémantique (et toute cette discussion) a un sens. De la même façon, il nous semble que c'est à partir du fait que nous comprenons que les neurosciences ont une corrélation avec la description psychologique de notre expérience, que notre tentative pour élucider une corrélation psychophysique (contextuelle et peut-être toujours imparfaitement déterminable) a un sens.

⁶³ C'est un sens qu'on peut qualifier d'« herméneutique » conformément à la deuxième connotation dégagée plus haut. L'élucidation d'un tel sens, qui excède l'objectif du présent article, pourrait suivre une ligne wittgensteinienne. (Je laisse ici ouverte la question de savoir si le second Wittgenstein est frégéen).

⁶⁴ Cette portée n'est autre qu'une forme de compréhension de soi (troisième connotation).

⁶⁵ La prise en compte de cette intelligibilité comme fait premier n'empêche absolument pas que l'on en vienne ensuite à s'interroger sur l'*ontologie* expliquant cette intelligibilité même. Il ne s'agit donc pas d'écarter les problèmes de métaphysique comme s'ils étaient de faux problèmes. Leur pertinence *découle* (entre autres choses) de ce phénomène qu'est l'intelligibilité fondamentale de la science. Car l'intérêt même que nous prenons aux études neurophysiologiques justifie de se demander quels sont les rapports envisageables entre l'esprit et le corps. Le présent article se borne à défendre que cet intérêt n'est pas suspendu à la résolution de ces problèmes métaphysiques et qu'au contraire, notre intérêt pour les questions ontologiques est suscité par ce fait, appartenant à la vie humaine, de la compréhension scientifique.

⁶⁶ Cet article, d'abord présenté à la journée organisée par A. Charrier et H. De Vilaine « L'humain dans les neurosciences » (IRePh), a bénéficié des échanges avec Denis Bonnay, Denis Forest et Jean-Michel Salanskis, ainsi que des remarques d'un relecteur anonyme. Ses faiblesses restent miennes.