UNIVERSITE DE BOURGOGNE

UFR Lettres et Philosophie

THÈSE

Pour obtenir le grade de

Docteur de l'Université de Bourgogne

Discipline : Philosophie

par

Anna CIAUNICA

le 24 septembre 2011

PHYSICALISME ET *QUALIA*

Limites de la rationalité scientifique au XXe siècle

Directeur de thèse
Alain BOUTOT

Jury

M. BITBOL Michel – Professeur, Directeur de Recherche au CNRS et au Centre de Recherches en Epistémologie Appliquée (CREA/Ecole Polytechnique Paris)
M. BUEKENS Filip – Professeur, Tilburg Center for Logic and Philosophy of Science
M. ROY Jean-Michel – Professeur, École Normale Supérieure de Lyon
M. SIMONS Peter – Professeur, Trinity College Dublin

© Anna Ciaunica

Pour F. et M., mes lumières…

Remerciements

Ce travail de recherche n'aurait pas pu voir le jour sans l'aide de mon directeur de thèse, Prof. Alain Boutot, dont le livre sur l'invention des formes a constitué un véritable déclic m'orientant vers l'étude de la complexité de la matière.

Je remercie surtout Michel Bitbol dont les écrits sur l'esprit et la mécanique quantique ont été pour moi plus que des guides, de véritables bouffées d'air frais, m'indiquant la voie à suivre dans le labyrinthe glissant des recherches sur l'esprit et la matière.

Je remercie Peter Simons d'avoir accepté de poser son regard critique sur ce travail de recherche dont les sinuosités rappellent les mésaventures de Don Giovanni.

Des fragments de ce travail ont été déjà présentés à de différentes conférences – VAF IV (Louvain), BPS-CEP (Oxford), BLT (Lausanne), Expression and the Inner Workshop (Oviedo), ESPP (Bochum, ENS (Lyon)). Je remercie aux participants à ces échanges, et plus particulièrement : Filip Buekens, John Bickle, Mike Beaton, Carlos Castrodeza, Mario de Caro, Robert Lockie, Manuel de Pinedo, Murali Ramachadran, Jean-Michel Roy, Barry Stroud, Susan Stuart, Pepa Toribio, Max Velmans, Stephen L. White – dont les commentaires fort utiles m'ont permis d'épurer et de clarifier la trame de mes arguments.

Enfin, je n'aurais jamais pu accomplir ce chemin sans l'aide patiente, de l'ombre vers la lumière, de Gilles, mon mari. Je lui en suis reconnaissante pour toujours.

Résumé

Réduit à sa plus simple expression, ce travail de recherche met face à face deux acteurs s'affrontant dans un duel philosophique : *le physicalisme* et *l'argument de la connaissance de Franck Jackson*. La question clé autour de laquelle s'agencera notre discussion ici est :

(1) Est-il vrai que « Tout est physique » ?

Le cœur de la thèse que nous défendons peut être résumé ainsi :

(T) Le clivage post-platonicien (tacite) entre les formes *objectivées* et les formes *subjectivées* de la pensée, provoque des fossés explicatifs (explicites), qui sont franchis (très souvent) *via* des sauts conceptuels (mystérieux), intercalés entre les étapes d'une argumentation.

Il sera argumenté ici que la distribution actuelle des débats sur l'axe états physiques (objectifs) / états qualitatifs (subjectifs) subit une pression souterraine post-platonicienne. Conséquemment, une bonne partie de l'énergie des penseurs actuels est dépensée à la recherche du « bon saut » ou « crochet » conceptuel permettant d'attacher les rives du mental à celui du physique. Nous soutiendrons que le cœur du problème ne consiste pas à résoudre le différend entre les physicalistes et les avocats des *qualia*, mais plutôt à trouver la réponse à la question : pourquoi en sommes-nous arrivés là, i.e., à ce clivage sur l'axe phénoménal/physique ?

Mots clé : physicalisme, argument de la connaissance, *qualia*, conscience, perception.

Abstract

"Everything is physical" physicalists claim. "Everything except *qualia*" reply the defenders of the Knowledge Argument. This thesis argues that both parties to this debate are committed to a traditional picture according to which one can tacitly adopt the standpoint of an off-stage narrator, capable of distinguishing *ab initio* between the different items in this conceptual scenario. The main claim here is that every time we artificially introduce a sharp conceptual distinction separating these two items or levels (mental and physical), we must automatically make a sophisticated conceptual leap in order to link the first-person qualitative perspective with the external physical world. Thus the physicalism/*qualia* dispute is only a by-product of an extended theoretical conception of the mind/world link which entails two distinct kinds of problems: (i) *structural problems* (to define and determine conceptually dual items like thought/matter, reason/senses, subject/object, etc.). Such problems require us to question "how are these items supposed to work together?" and this leads us to the second group of problems: (ii) *mediation problems*. The first group of problems inevitably leads to explanatory gaps; the second ends up needing to appeal to conceptual leaps in order to ensure the necessary link between two separate items.

This exerts a powerful influence over the cast of our thoughts: seen from this angle, all philosophical enterprise comes down to the question of where to place the three following parameters: the world as it is, the conceptual picture which aims to represent the world as it is and the theorist-painter gifted with the ability to capture the world picture as it is. We also face the problem of finding the right mediators to assure coherence among the members of this trio, and the problem of setting the valid criteria guaranteeing the theorist-painter that what is captured on his conceptual canvas does indeed correspond to the world as it is, i.e. that he is not laboring under the delusions of an evil genie. In this thesis I argue *against* this way of putting things.

Key words: physicalism, *qualia*, knowledge argument, perception, representation, consciousness.

1 Table des matières

Remerciements 9

Résumé 11

Abstract 13

Introduction 21

Première partie Le Physicalisme 49

1 L'hégémonie du physicalisme aujourd'hui 51

1.1 De phusis à la physique et de la physique au physicalisme 56

1.2 Unité de la science et stratification du monde 63

1.2.1 Réduction et niveaux du réel 66

1.2.2 La complétude de la physique et l'argument causal en faveur du physicalisme 72

1.3 Définir le slogan physicaliste standard : « Tout est physique » 79

1.4 Physicalisme et Nécessitation 85

1.4.1 Propriétés physiques instanciées et mondes possibles 86

1.4.2 Nécessitation, contingence et ectoplasme 89

1.5 Deux conceptions du 'physique' 93

1.5.1 L'approche basée sur l'Objet Physique 94

1.5.2 L'approche basée sur la Théorie Physique 98

1.5.3 Le dilemme d'Hempel 103

1.5.4 Quelques conclusions provisoires 107

2 Du sujet logique au sujet psychologique 113

2.1 L'antipsychologisme frégéen : quelques conséquences 115

2.2 Un premier exemple de « saut conceptuel » 121

2.3 Carnap et les contenus d'expérience vécue 127

2.4 La perception du « rouge » dans le physicalisme sémantique 134

3 Le mental et le physique 145

3.1 L'impasse dualiste 147

3.2 La douleur en héritage : du behaviorisme au matérialisme réductionniste 153

3.2.1 La Théorie de l'Identité Psychophysique 157

3.2.2 Identité-Type et Identité-Token 164

3.3 Physicalisme et survenance 169

3.3.1 Le Fonctionnalisme en trois questions 172

4 Physicalisme et *qualia* 179

4.1 Définir les qualia 181
4.2 L'argument du fossé dans l'explication 186
4.3 Autres arguments antimatérialistes 193
4.3.1 *Qualia* absents, *qualia* inversés, et autres curiosités 193
4.3.2 L'argument de la 'concevabilité' 199
4.4 Comment franchir le fossé explicatif ? 205
4.4.1 Matérialisme de type-A ou le réductionnisme radical 206
4.4.2 Matérialisme de type-B 210
4.4.3 Matérialisme de type-C 214
4.5 Variantes non-réductionnistes 220
4.5.1 Dualisme de type-D ou l'interactionnisme cartésien 222
4.5.2 Dualisme de type-E ou l'épiphénoménalisme 224
4.5.3 Monisme de type-F 226
Deuxième partie L'argument de la connaissance 231
5 Mary, la prisonnière 233
5.1 L'archange, le super-martien, la chauve-souris, et le savant sourd 234
5.2 La structure de l'argument de Jackson et ses retranscriptions 238
5.2.1 (V1) Version Faible ou épistémique du KA 243
5.2.2 (V2) Version Forte ou ontologique du KA 243
5.2.3 Un dispositif tripartite 245
5.3 L'architecture de l'argument en sept questions 247
5.4 Connaissance non-propositionnelle : « savoir que » vs « savoir faire » 248
5.4.1 L'hypothèse de l'habilité 249
5.4.2 L'hypothèse de la connaissance par accointance 256
5.5 La stratégie des concepts phénoménaux 259
5.5.1 Concepts physiques et concepts phénoménaux : M. Tye 261
5.5.2 Le Représentationalisme de W. Lycan 267
5.5.3 Concepts à la première et à la troisième personne 268
5.5.4 L'approche indexicale : J. Perry 270
5.6 Objections contre la stratégie des concepts phénoménaux 283
5.6.1 L'Argument des Propriétés Distinctes : St. White 283
5.6.2 Le cas de Marianna 286
5.7 Rejeter l'intuition : doutes sur le bien-fondé de l'argument 290
5.7.1 La banane bleue 291
5.8 Jackson change d'avis 292
5.8.1 La Théorie Représentationnelle de la Conscience Phénoménale 293
5.8.2 Objections anti-représentationalistes 298

5.8.3 Le 'Pack Jackson' *vs* le 'Pack Physicaliste Non-Réductionniste' 306
5.8.4 Quelques conclusions provisoires 308
6 La thèse du savoir complet 313
6.1 Faits Subjectifs et Faits Objectifs 313
6.2 Omniscience divine et omniscience terrestre 318
6.2.1 Omniscience physique *simpliciter* 320
6.2.2 L'argument de la possession du concept 323
6.2.3 Information physique théoriquement communicable 325
6.3 Propriétés fondamentales et « humilité Ramseyenne » 328
6.4 Ce que Dieu nous apprend sur Mary 332
6.5 Le physicalisme nonthéorique de Nagasawa 334
6.6 Arbre récapitulatif synoptique 337
7 Physicalisme *a priori* et philosophie en fauteuil : un vieux ménage 343
7.1 Le physicalisme *a priori* 345
7.1.1 Dérivabilité a priori et a posteriori 346
7.1.2 Le principe du passage *a priori* 351
7.2 Contester le principe du passage *a priori* 359
7.2.1 L'objection de Block et Stalnaker 360
7.2.2 L'objection de Blackburn ou l'argument de l'inventeur du zip 362
7.3 Analyse conceptuelle et sauts conceptuels 365
7.3.1 Le bon plaisir de Dieu et l'*a priori* 365
7.3.2 Le tandem fossés explicatifs/sauts conceptuels 367
7.3.3 Faites entrer l'accusé : pourquoi Descartes n'est pas le seul 'coupable' 372
7.3.4 Ecce intrusus ! 380
7.3.5 Retour à une vieille querelle: Carnap, Quine, et la question de l'être 383
7.4 Naturaliser le physicalisme? 392
7.4.1 Le Physicalisme : portrait robot final 393
7.4.2 Deux dogmes du physicalisme 397
7.4.3 Naturaliser les *qualia* ? 405
Conclusion 409
8 Un argument pour concilier Mary avec le physicalisme 411
8.1 Retour au clivage Subjectif/Objectif 412
8.2 Divine Mary 417
8.3 La nuit, tous les chats sont gris 421
8.4 Le mental, le physique et la lampe d'Aladin 424
Bibliographie 429

Index lexical 461

Index des noms 473

2 Table des Figures

Figure 1: Variations sur le thème du dualisme 32
Figure 2: L'unité de la science et la stratification du monde 65
Figure 3: Les trois mondes de Frege 119
Figure 4: Un premier exemple de saut conceptuel 123
Figure 5: Le pôle sujet/objet dans les énoncés protocolaires 137
Figure 6: Un deuxième exemple de saut conceptuel 141
Figure 7: Le mental et le physique 145
Figure 8: Réduction des états mentaux 160
Figure 9: Les arguments des *qualia* absents et du spectre inversé 204
Figure 10: Le KA - Arbre récapitulatif : La Thèse de l'Apprentissage 309
Figure 11: Le KA - Arbre synoptique général 338
Figure 12: Tandem fossés explicatifs / sauts conceptuels 370

Introduction

> *"We class sensations along with emotions, and volitions, and thoughts, under the common head of states of consciousness. But what consciousness is, we know not; and how it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of the Djin when Aladdin rubbed his lamp, or any other ultimate fact of nature".* (T. Huxley 1866)

Mary, la prisonnière

Mary est une neurophysiologiste très douée qui a toujours vécu dans un environnement fermé ne contenant que des objets en noir et blanc. Imaginons qu'elle ait assimilé, à l'aide des livres et de la télévision, toujours en noir et blanc, le contenu de *tous* les traités neurophysiologiques sur la perception des couleurs, ces informations lui étant présentées sous la forme des textes et des diagrammes en noir et blanc. On peut affirmer, dans ces conditions, qu'elle connaît tous les *faits* et *les informations physiques* à propos de la perception des couleurs. A présent, supposons qu'une âme charitable vienne libérer Mary de sa prison en noir et blanc. Pour la première fois de sa vie, la savante Mary est confrontée au monde coloré qui s'étend au-delà de son environnement confiné. A l'instant même où elle franchit le seuil qui sépare sa prison d'un champ de blé parsemé de coquelicots, disons, elle *sait* quelque chose qu'elle ne savait pas auparavant : elle sait désormais *ce que c'est* de voir des couleurs. Puisqu'elle a appris quelque chose, il faut admettre que l'ensemble de connaissances dont elle disposait dans sa prison était incomplet. Mais si elle possédait déjà toute l'information physique, il s'ensuit qu'il y a davantage à obtenir que cela, et, par conséquent, le slogan physicaliste selon lequel « tout est physique » est faux. C'est le défi qu'a lancé F. Jackson[1] aux défenseurs du physicalisme en philosophie de l'esprit.

[1] F. Jackson, (1982), "Epiphenomenal *Qualia*", *Philosophical Quarterly* 32: 127–136. F. Jackson, (1986), "What Mary Didn't Know", *Journal of Philosophy* 83: 291–295. Réédités avec un *Postscript* in P. Ludlow, Y. Nagasawa & D. Stoljar, (éds.), (2004), *There is something about Mary: essays on*

Réduite à sa plus simple expression, l'enquête que nous nous apprêtons à mener ici suit un scénario mettant face à face deux acteurs s'affrontant dans un duel philosophique : (1) un courant de pensée dominant la philosophie de l'esprit contemporaine – i.e. *le physicalisme*, et (2) un argument visant à le réfuter – i.e. *l'argument de la connaissance de Franck Jackson*. Les enjeux de cet affrontement sont majeurs : il s'agit de déterminer la place de l'esprit dans un monde régi par des lois physiques, et de plaider pour une explication à la fois naturaliste et non-réductive de la conscience expérientielle et de sa corrélation avec les phénomènes neurophysiologiques.

La controverse que nous allons examiner ici se prête, *prima facie*, à une distribution bipolaire assez simple : d'un côté nous avons le camp des défenseurs du physicalisme, tandis qu'au coin opposé du ring il y a ceux qui, convaincus par la force de l'argument de Jackson, essayent de faire une place aux *qualia* parmi les propriétés fondamentales du réel. Mais ce serait se livrer à une grossière exagération que de décrire ce combat selon le modèle d'un match de boxe au terme duquel on proclame un vainqueur, un vaincu, et la recette des paris. Tout au long de notre investigation nous espérons montrer qu'en réalité la situation est beaucoup plus compliquée. Pour prendre la mesure de cette complexité, il suffit de rappeler que l'auteur même de l'argument de la connaissance, F. Jackson[2], a récemment changé de camp, et est devenu un des défenseurs les plus influents du physicalisme, tandis que d'autres penseurs ont commencé leur carrière en défendant le physicalisme, et ont fini par se «convertir » aux *qualia*. Quoiqu'il en soit, partisans et détracteurs du physicalisme partagent le même vecteur commun, à savoir l'ambition de résoudre ce que D. Chalmers[3] appelle « le problème difficile » que pose la conscience, et cela *depuis l'intérieur du naturalisme*, autrement dit, en respectant, pour paraphraser une célèbre expression de Quine, les décisions du tribunal suprême des sciences naturelles.

La question clé autour de laquelle s'agencera notre discussion ici est :

(1) Est-il vrai que (*p*) « Tout est physique » ?

---

phenomenal consciousness and Frank Jackson's knowledge argument, Cambridge, MA: MIT Press. (Dans ce qui suit, nous faisons référence à cette dernière édition.)

[2] F. Jackson, (1998), *From Metaphysics to Ethics: A Defense of Conceptual Analysis*, Oxford: Clarendon.

[3] D. Chalmers, (1996), *The Conscious Mind*. Oxford: Oxford University Press.

Cette question se ramifie à son tour en deux volets-corolaires:

(1_A) : Si (*p*) est vrai, quelles sont les bases argumentatives soutenant *p*, et avec quelles implications pour la place du mental dans le réel physique ?

(1_B) : Si (*p*) est faux, quelles sont les bases argumentatives récusant *p*, et avec quelles implications pour la place du mental dans le réel physique ?

Conséquemment, notre enquête sera divisée en deux : une première partie traitera le slogan « tout est physique », la deuxième partie étant dédiée aux implications de l'argument antiphysicaliste du « premier » Jackson. Tout au long de cette investigation, nous allons soumettre les acteurs de ce débat à passer *le test du naturalisme*. En quoi consiste-t-il, ce test? Pour répondre à cette question, il nous faut fixer quelques éléments d'arrière-plan.

Comme l'a relevé B. Stroud[4], le terme « naturalisme » est fort ambigu. Une bonne méthode de l'approcher serait de remplacer la question « qu'est-ce que le naturalisme ? » par la question « quels sont les enjeux contenus dans l'affaire du naturalisme » ? Qu'y-a-t-il à perdre ou à gagner en adhérant au naturalisme ? D'abord, tout le monde s'accorde à dire que le naturalisme se définit avant tout en tant qu'ennemi déclaré du « surnaturalisme ». Une autre possible manière de l'approcher serait de faire appel à des exemples. Prenons le cas d'un philosophe comme Hume[5] : bien que ce dernier ne s'est jamais ouvertement placé sous l'étiquette de « naturaliste », Hume est souvent cité comme un exemple typique de penseur naturaliste. Pourquoi ? Parce qu'il voulait comprendre la nature humaine sans faire appel à des éléments transcendant l'expérience humaine. Ses efforts d'introduire une méthode de raisonnement expérimentale visaient clairement à s'en tenir uniquement à ce que l'entendement humain puisse atteindre, i.e. sans aller au-delà des limites du monde naturel. Cette dernière exigence, chacun le sait, n'a pas été respectée par ses grands prédécesseurs, comme Descartes, Berkeley, Malebranche, Locke, car tous ces penseurs ont tôt ou tard admis dans leurs modèles

[4] B. Stroud, (1996), 'The Charm of Naturalism' in *Proceedings of the American Philosophical Association* 70, pp. 43-55. Réédité *in* M. de Caro & D. Macarthur (eds.) *Naturalism in question*, Cambridge Massachusetts, Harvard University Press, 2004, pp. 21-35. (Nous faisons reference ici à cette dernière édition). Voir aussi B. Stroud, (2009),'The Transparency of 'Naturalism'', The Romanell Lecture, *Proceedings and Addresses of the American Philosophical Association*, pp. 157-169.

[5] D. Hume, (1740), *A Treatise of Human Nature*, édité par David Fate Norton & Mary J. Norton, Oxford/New York: Oxford University Press, 2000. Voir aussi D. Hume, (1784), *An Enquiry concerning Human Understanding*, édité par Tom L. Beauchamp, Oxford/New York: Oxford University Press, 1999

explicatifs de la perception et de la pensée, des éléments « supérieurs » ou « divins » de l'au-delà. A l'opposé, la pensée de Hume se veut avant tout une philosophie expérimentale. Que cela plaise ou non aux philosophes, nous sommes tous obligés de commencer par notre position *dans* le monde, i.e. avec nos expériences et capacités, essayant de comprendre et d'expliquer mes mécanismes effectifs par le truchement desquels nous entretenons des croyances et nous assertons des énoncés sur le réel. Bref, il n'y a aucune autorité « supérieure » transcendant épistémiquement notre position dans le monde naturel. Le naturalisme se définit ainsi avant tout comme une résistance aux chants de sirène du surnaturalisme. C'est ainsi qu'au XXe siècle que les penseurs du Cercle de Vienne ont instauré en philosophie ce qu'on peut appeler le consensus « antisurnaturaliste ». On appelle donc « surnaturaliste » toute stratégie explicative qui consiste : (i) soit à faire appel à des éléments explicatifs qui transcendent l'univers familier et naturel de tous les jours ; (ii) soit à invoquer un agent dont les actions ne seraient pas comprises comme un membre à part entière du monde naturel ; (iii) soit les deux.

C'est précisément ce consensus « antisurnaturaliste » qui a soudé les empiristes logiques dans leur croisade contre les métaphysiques traditionnelles. Le test du naturalisme consiste donc à passer au crible les éléments argumentatifs d'une théorie afin de détecter si ceux-ci ne font pas appel illicitement, à un moment ou à un autre dans leur trame argumentative, à des éléments surnaturels dans le sens défini plus haut par Stroud. L'objectif ici est de vérifier si dans le duel physicalisme/*qualia*, ses protagonistes respectent *la contrainte naturaliste*. Une des thèses que nous défendons ici sera de dire que dans la controverse opposant les défenseurs du physicalisme aux avocats de l'argument de la connaissance, les deux protagonistes « trichent », pour ainsi dire, et insèrent à un moment donné de leur schéma argumentatif, une prémisse qui tombe sous le couperet de l'objection antisurnaturaliste.

Plus spécifiquement, nous verrons que le physicalisme contemporain dans sa version réductionniste *a priori* s'enlise dans le bourbier des questions métaphysiques traditionnelles, ressuscitant ainsi les démons du passé autour de la question de l'être. Quant à l'argument de la connaissance de Jackson, on notera qu'en s'appuyant sur l'idée que Mary connaît *toute* l'information physique durant sa captivité, il ouvre la boîte de Pandore des notions démiurgiques comme celle d'« omniscience ».

Notre investigation commencera par la présentation du premier membre du duel qui nous concerne ici, i.e., le physicalisme. Pour mesurer la véritable portée et l'influence de ce courant de pensée, il suffit de rappeler la remarque, un rien ironique de N. Chomsky[6] selon laquelle, de nos jours, à peine se met-on à comprendre quelque chose, qu'on le qualifie de *physique*. En tant que science qui a "réussi", la physique est souvent mise en avant comme un élève modèle.

Dans les sections 1.1. et 1.2. du premier chapitre, nous entamerons un travail d'éclaircissement des termes « physique » et « physicalisme ». Nous verrons ainsi que vigoureusement appuyé par *l'argument de l'incroyable succès de la science*[7], le physicalisme se cristallise autour d'une thèse métaphysique centrale, souvent exprimée dans une formule dont l'élégante simplicité peut facilement séduire : « Le monde contient exclusivement des entités reconnues par la physique »[8]. Cette définition, apparemment très sobre, n'est cependant pas dénuée de sérieuses difficultés. Car bien qu'étayée par le succès grandissant des sciences de la nature, l'adoption d'une ontologie matérialiste a fini par provoquer ce que D.M. Armstrong appelle une véritable « crise en philosophie »[9]. Au cœur de la stratégie physicaliste, nous retrouvons un principe philosophique que l'on peut appeler *le principe de la complétude causale, nomologique et explicative du domaine des états physiques* : pour tous les états physiques *p*, dans la mesure où un état physique *p* a des causes, est soumis à des lois et permet une explication, l'état physique *p* a des causes physiques complètes, est soumis à des lois physiques complètes, et possède une explication physique complète.

Ontologiquement, le physicalisme assume un *monisme matérialiste*, épistémologiquement renforcé par deux principes régulateurs: (i) le principe de l'unité de la science[10], et (ii) le principe de la clôture causale de la physique. Leur

---

[6] Cité par J. Searle, (1992) *The Rediscovery of the Mind*, MIT Press, Cambridge, trad. de l'anglais par C. Tiercelin, *La redécouverte de l'esprit*, Paris, Gallimard, 1995. p. 50. (Référence ici à l'édition française.)

[7] M. de Caro & D. Macarthur (eds.), 2004, Préface in *Naturalism in question*, Cambridge Massachusetts, Harvard University Press, p. 5.

[8] D.M. Armstrong, (1980) 'Naturalism, Materialism, and First Philosophy' in *The Nature of Mind and Other Essays*, St. Lucia, University of Queensland Press, p. 156. Repris *in* P. Moser & J. Trout, (éds.) 1995, *Contemporary Materialism*, London: Routledge. (Référence à cette dernière édition).

[9] D.M. Armstrong, (1981) 'The Causal Theory of the Mind' in *The Nature of Mind and Other Essays*, Cornell University Press, réédité in W.G. Lycan (ed.) *Mind and Cognition – a Reader*, Blackwell, Oxford & Cambridge, 1990. (Référence à cette dernière édition).

[10] P. Oppenheim & H. Putnam, (1958), 'The unity of science as a working hypothesis', *in* H. Feigl et al., eds., *Minnesota Studies in the Philosophy of Science*, vol. 2, Minneapolis: Minnesota University Press.

rôle est celui d'organiser, sous un seul et même sceptre, les différents regards que les scientifiques posent sur le réel.

Comme l'a relevé H. Putnam[11], le principal attrait du monisme matérialiste réside dans son ambition de fonder une métaphysique *naturelle*, bâtie avec les moyens et *depuis l'intérieur* de la science, contrastant ainsi avec la métaphysique *supranaturelle* des traditions idéalistes. Le matérialisme satisfait à la fois les ambitions métaphysiques traditionnelles - légiférer sur ce que le monde *est* - tout en restant sur un terrain prudent des pratiques expérimentales qui nous renseignent sur la façon dont le monde *est connu*, et limitant ainsi son encadrement théorique aux seules entités qui se laissent appréhender empiriquement. Or, la seule entité que la science peut installer effectivement sous sa loupe grossissante, est la *matière*, distribuée dans l'espace-temps, et gouvernée par les lois de la physique. Traditionnellement, les théories physiques ciblent explicitement la découverte des lois fondamentales qui gouvernent la structure et la dynamique des objets macroscopiques, allant jusqu'à postuler la possibilité d'une description physique *complète* de tout événement, fait ou propriété qui composent le monde naturel.

Réduit à sa plus simple expression, le credo physicaliste affirme donc que *tout est physique*. Ce précepte s'applique aussi bien au monde actuel dans lequel nous vivons, qu'à tous les autres mondes possibles. Le physicalisme ainsi défini affiche clairement ses ambitions métaphysiques. Les mondes (actuels et possibles) sont constitués d'objets qui instancient des propriétés fondamentales qui, à leur tour, instancient des relations fondamentales. Selon la définition qu'en donne D. Lewis[12], les « véritables » propriétés fondamentales, très peu nombreuses d'ailleurs, sont celles que l'on peut appeler « propriétés parfaitement naturelles » et dont l'instanciation est intrinsèque, c'est-à-dire non-divisible ou mixte : la masse ou la charge des particules élémentaires sont d'excellents exemples de telles propriétés « parfaitement naturelles ». C'est la mission de la physique de fournir un inventaire *complet* de toutes les propriétés et relations fondamentales qui sous-tendent le réel, ainsi que de découvrir les lois fondamentales qui gèrent leurs actions.

---

[11] H. Putnam, (1983), 'Why There isn't a Ready Made World', in *Realism and Reason*, vol.3, Cambridge: Cambridge University Press.

[12] D. Lewis, (1983), 'Mad Pain and Martian Pain' and 'Postscript'. In D. Lewis *Philosophical Papers, Vol.* 1, Oxford, Oxford University Press.

Par conséquent, les sections allant de 1.3. à 1.5. seront consacrées à l'interprétation du crédo physicaliste « tout est physique ». Dans cette entreprise périlleuse, nous allons nous appuyer notamment sur les récents travaux de D. Stoljar[13] dédiés au thème du physicalisme. Trois axes articuleront cette enquête. (1) Le premier concerne ce que Stoljar appelle la *question de l'interprétation*, autrement dit la réponse à la question : « quelle est la signification de la thèse 'tout est physique' » ? (2) Le deuxième axe concerne la *question de la vérité*, c'est-à-dire la réponse à la question : « Est-ce que le physicalisme est vrai ? » (3) Enfin, le troisième point, plus général, vise les *enjeux philosophiques* du physicalisme, autrement dit, sa candidature au rôle du meilleur 'Weltanschauung' de l'époque moderne. Seule la réponse à la question de l'interprétation nous occupera dans le premier chapitre. Les deux autres interrogations seront étayées et développées au fur et à mesure de notre enquête.

A la fin du premier chapitre, nous arriverons à la conclusion que non seulement la question de définir le physicalisme est loin d'être close[14], mais, qui plus est, cela nous obligera à ouvrir le coffre poussiéreux du grenier abritant le physicalisme de « première génération ». Conséquemment, dans le deuxième chapitre, nous allons reculer provisoirement d'un pas, et réexaminer le thème du physicalisme en partant cette fois-ci de ses origines. Plus spécifiquement, nous allons amorcer notre enquête avec l'antipsychologisme frégéen. Pourquoi était-il si important, aux yeux de Frege, d'éliminer le psychologisme, c'est-à-dire d'attaquer l'idée que les règles de la pensée sont fondées sur les régularités de la dynamique mentale ? La réponse à cette question – parce que la logique est dotée d'un statut solide la plaçant au-delà des errements humains – nous amènera au plus près du cœur de l'hypothèse que nous défendons ici. Nous verrons ainsi que la principale raison qui nous a encouragés à regarder ce problème à travers les lunettes déformantes, certes, mais fort révélatrices, des premiers « analytiques », est

[13] D. Stoljar, (2010), *Physicalism*, London & New York: Routledge.

[14] Ce point a été éminemment capturé par le bien connu dilemme d'Hempel : si le concept de « physique » est caractérisé en termes puisant dans le vocabulaire de la physique contemporaine, quelqu'un pourrait objecter en soulevant la difficulté suivante : étant donné que la physique contemporaine est loin d'être complète, il est faux d'affirmer que le prédicat « est physique » s'applique à *toutes* les entités du réel. D'autre part, en supposant que l'on arrive à imposer une théorie physique complète, alors le prédicat « est physique » risque de devenir trivial, à moins que ne l'on apporte des précisions supplémentaires quant aux conditions qui rendent une théorie « physique ». C. Hempel, (1980), 'Comments on Goodman's *Ways of Worldmaking'*, *Synthese*, 45: 139–199.

stratégique. Car si l'on insiste ici sur les efforts des premiers philosophes analytiques d'éliminer toute référence au mental et à la subjectivité phénoménale, c'est parce que nous sommes persuadés que la plupart des raisons qui ont incité ces auteurs à amputer le discours philosophique de son versant « subjectif », émanent d'un ensemble d'hypothèses tacites, héritées *via* le modèle logico-déductif de la connaissance privilégié depuis Platon, et glissées subrepticement au cœur même des controverses qui animent la philosophie analytique contemporaine. Décoder et rendre visible l'influence souterraine de ces thèses d'arrière-plan sur la philosophie de l'esprit actuelle s'avère, à notre sens, d'une indiscutable urgence.

Prenons un exemple : nombreux sont aujourd'hui ceux qui imputent à Descartes les conséquences désastreuses de son dualisme des substances. Pour constater cela, il suffit de survoler brièvement la table des matières de n'importe lequel 'Companion to the Mind' paru au sein du courant analytique à partir de la deuxième moitié du XXe siècle. Presque sans exception, la critique du dualisme cartésien ouvre le bal, pour ainsi dire, et prépare le terrain pour l'introduction triomphale du physicalisme et/ou du matérialisme moniste. La démolition du dualisme des substances demeure un fort confortable point de départ pour les partisans du physicalisme devenant, en peu de temps, l'adversaire idéal, étape destructive incontournable précédant la construction d'une théorie matérialiste de l'esprit. Bien que nous fassions partie du camp de ceux qui admettent le caractère démodé de ce type de dualisme radical, nous aimerions néanmoins exprimer, dès le départ, quelques réserves concernant la manière dont la tradition analytique s'est emparée de cet ennemi afin d'en construire, sur ses ruines, le château imposant du physicalisme actuel.

Nous avancerons ici l'hypothèse selon laquelle non seulement le dualisme cartésien, mais aussi son principal adversaire, le matérialiste contemporain, véhiculent, sous une forme plus sophistiquée, une tendance dichotomique d'arrière-plan beaucoup plus ancienne et tenace, dont les origines semblent se perdre dans la nuit des temps, et qui postule la distinction entre un monde immuable et objectif d'une part, et un monde immanent et subjectif d'autre part. Le premier versant

correspond à ce que S. Blackburn[15] appelait récemment le royaume des « vérités fortunées », i.e. objectives dans tous les mondes possibles, tandis que le deuxième abrite les « vérités infortunées » validées par la vie de tous les jours ici et maintenant.

De manière générale, l'histoire de la pensée philosophique peut être interprétée comme un mouvement oscillant entre les pôles de ce que l'on pourrait appeler l'équation archétypale pensée (rationnelle) / monde (empirique). Dans les courants d'influence empirique, c'est l'objectivité du monde extérieur, matériel, qui garantit l'objectivité de la connaissance. Dans les courants d'influence rationaliste, ce sont les règles de la pensée rationnelle et logique qui garantissent l'objectivité de la connaissance. Mais dans tous les cas, et c'est ici le point sur lequel nous allons insister, le vecteur commun qui caractérise à la fois empiristes et rationalistes est sans doute *le rejet de la latitude subjective*, responsable des errements et des fluctuations glissantes.

Ce qui va sans dire, c'est la facilité avec laquelle on succombe à cette bipartition sur l'axe subjectif/objectif. Par contre, ce qui pose réellement problème, ce sont les problèmes de *médiation*, de jointure et d'articulation que ce clivage provoque.

Situé à mi-chemin entre ces deux mondes, l'être humain, *rationnel* de par l'esprit, *empirique* de par le corps, provoque une profonde perplexité. En effet, quelle place pour l'homme dans cette représentation bipartite du monde? La solution platonicienne semble la plus élégante : l'être humain appartient de par l'esprit rationnel au monde objectif et « fortuné» des entités immuables, tandis qu'il appartient de par le corps matériel au monde « infortuné » de tous les jours. Suite à ce clivage quasi-mythologique, l'homme se retrouve ainsi contraint d'occuper, pour parodier la célèbre formule cartésienne, le rôle ingrat d'une glande pinéale « grandeur nature », articulant deux niveaux ontologiques et épistémiques distincts : matériel/immatériel et objectif/subjectif. Pour mesurer la portée et la profondeur des empreintes laissées par ces clivages dans la philosophie occidentale il suffit de suivre de près le sillage conceptuel que laisse derrière lui le clivage mythologique

[15] S. Blackburn, (2009), 'Analysis, Description, and the *A priori*?' In Ian Ravenscroft (ed.), *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*, Oxford: Clarendon Press, p. 24.

«pensée/monde », véritable « Big Bang Gouffre » archétypal, si l'on peut appeler ainsi, provoquant une spectaculaire cascade de « gouffres collatéraux ». Cristallisés sur l'axe matériel/immatériel, ces gouffres seront remis au goût du jour par les différents courants de pensée. Bref, à chaque époque, sa dualité. (Cf. Fig. 1)

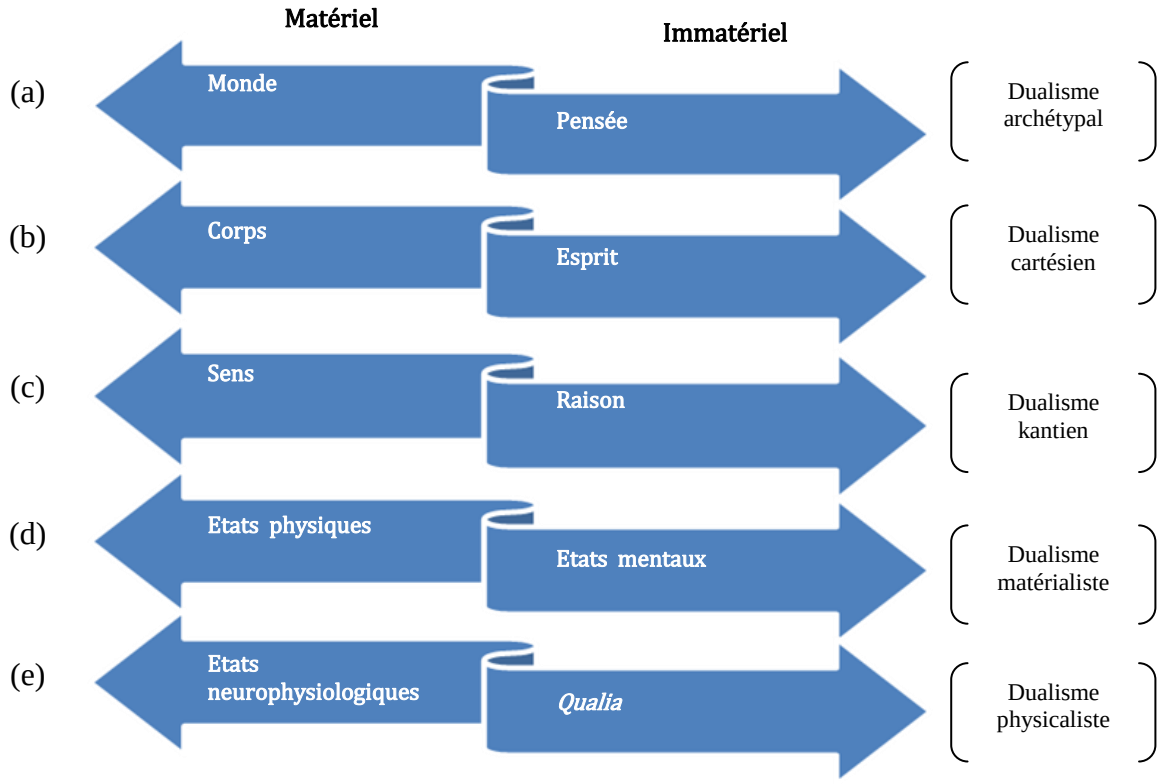


Figure 1: Variations sur le thème du dualisme

Tout ceci a pour but de mettre en relief le constat suivant : bien que le thème que nous allons investiguer ici concerne uniquement les niveaux (d) et (e) du tableau ci-dessus, ce serait se livrer à une considérable erreur que d'entamer le débat autour du duel entre le physicalisme et les défenseurs des *qualia* sans mentionner et sans prendre en considération les racines profondes qui le rattachent à ses prédécesseurs. Nous défendrons l'idée selon laquelle le dualisme cartésien, véritable « bouc émissaire » du matérialisme contemporain, ne doit pas comparaître seul devant le tribunal qui condamne actuellement le schisme mental/physique avec tant de fermeté : car d'autres schèmes conceptuels impliqués et responsables dans cette distribution sont encore en liberté, pour ainsi dire, et continuent de faire une belle carrière, influençant souterrainement la pensée contemporaine sous des formes plus ou moins déguisées, et provoquant de nombreux pseudo-problèmes. Or, si l'on

scrute les raisons profondes qui ont incité Descartes à introduire sa célèbre dichotomie des substances, on s'apercevra qu'en réalité la séparation corps/esprit, et plus tard son équivalent physicaliste états neurophysiologiques/états mentaux, n'est qu'un dommage collatéral, si l'on peut appeler ainsi, d'un modèle d'intelligibilité beaucoup plus vaste, et dont la généalogie remonte jusqu'aux penseurs grecs, et plus particulièrement à Platon. De fait, la pensée cartésienne ne fait que prolonger une dichotomie d'arrière-plan beaucoup plus ancienne, quasi-mythologique, et qui sépare deux ordres ontologiques distincts : (i) l'ordre de l'*objectivité* qui correspond *grosso modo* au profil du monde platonicien des Idées ; et (ii) l'ordre de la *subjectivité* qui abrite les événements de « tous les jours ». Lorsque Platon fait le choix de ranger les Idées dans un ordre ontologique détaché de l'ordre familier dans lequel évolue l'homme de la rue, il réalise un double exploit : il assure, d'une part, un logement confortable aux entités immuables qui servent de repère sécurisant dressé comme un rempart contre la périssabilité dangereuse du quotidien. Après tout, tous les humains (y compris Platon) ont connu au moins une fois dans leur vie la peur de la mort. D'autre part, il prend le soin de laisser une porte entrouverte permettant à l'individu d'accéder, par le truchement de l'esprit rationnel, à la lumière de ces vérités idéales, capables, d'une manière ou d'une autre, de projeter l'esprit dans les bras de l'éternité. De ce point de vue, A.N. Whitehead avait certainement raison de caractériser la philosophie occidentale comme une série d'annotations aux préceptes de la métaphysique platonicienne[16].

L'objectif ici c'est de montrer que l'ontologie platonicienne, et ses diverses variations remises au gout du jour, prêtent le flanc à l'objection antisurnaturaliste évoquée plus haut, et y succombe.

Par conséquent, bon nombre de problèmes qui agitent la philosophie de l'esprit d'aujourd'hui *sont devenus* des problèmes suite à l'implantation sous-jacente d'un corps parasite de clivages post-platoniciens qui agissent comme un courant souterrain, et qui finissent par conduire la réflexion philosophique à des impasses et des difficultés insurmontables. Notre enquête ici est animée par l'intuition selon laquelle une difficulté de même type pourrait être à l'origine des controverses que

[16] "The safest general characterization of the European philosophical tradition is that is consists of a series of footnotes to Plato." A. N. Whitehead, (1929), *Process and Reality*, Cambridge University Press, p. 53.

suscite le cas de Mary dans l'affaire du physicalisme. Par bien des aspects, suivre la trace de ces filiations suspectes peut s'avérer vitale pour tout programme d'analyse opérant avec des concepts placés sous haute tension tels les concepts d'« esprit », de « corps » ou encore d'« expérience phénoménale ».

Dans ce qui suit, nous montrerons que ce type de conceptualisation dichotomisant repose sur une vision mécanique et rigide des rapports entre la pensée et le monde, et qui mène inévitablement à des fossés explicatifs. Le concept de « fossé explicatif » auquel nous faisons allusion ici respecte la définition classique, initiale, que lui a donnée J. Levine[17], avec une réserve importante cependant: à la suite de R. Van Gulick[18], nous soutiendrons l'idée selon laquelle la pratique des fossés explicatifs est visible à l'œuvre dans la pensée philosophique *bien avant* que Levine introduise le label dans son fameux article paru en 1983.

Sans vouloir insister ici sur un sujet qui sera développé et détaillé plus loin au cours de notre enquête, énonçons dès à présent le cœur de la thèse que nous voulons défendre ici : la philosophie post-platonicienne est contaminée, d'une manière plus ou moins explicite, par les effets secondaires de ce que nous avons appelé le « Big Bang Gouffre » séparant la pensée et le monde. La pensée rationnelle revêt l'aspect d'une critique intersubjective garante de l'objectivité, dont le but déclaré est de chercher et de trouver le Saint Graal : la Vérité. Cet accent mis sur la rationalité a pour effet collatéral l'amputation du versant « subjectif » et « psychologisant » du sujet connaissant. Les états mentaux, flous et fluctuants, sont relégués en second plan, tandis que la logique rationnelle est tenue pour solide. Rien de ce que nous venons d'énoncer n'est nouveau ou inhabituel. Or, dans ce qui suit, nous soutiendrons que ce « dommage collatéral » du post-platonisme est accompagné d'un phénomène théorique que nous pourrions baptiser le « ***tandem fossés explicatifs/ sauts conceptuels*** ».

Prenons un exemple, pour plus de clarté : la recherche frégéenne des formes objectivantes de la connaissance inhibe le versant subjectif du mental. Or, plus un schéma conceptuel insiste, dans ses trames argumentatives, sur le clivage objectif/subjectif, plus il provoque le phénomène de fossé dans l'explication. Dans le

[17] J. Levine, (1983), 'Materialism and *qualia*: the explanatory gap', *Pacific Philosophical Quarterly*, 64: 354-361.
[18] R. Van Gulick, (2003), 'Maps, gaps and traps', *in* A. Jokic & *Q. Smith* (eds). *Consciousness: New Philosophical Perspectives.* Oxford: Oxford University Press.

cas de Frege, un gouffre sépare les "pensées objectives" des "contenus d'expérience vécue". Cette objectivité « pure », logée dans la 'tête-esprit' du sujet connaissant, déclenche le problème vertical de la médiation, et du transfert entre les différents éléments s'emboîtant dans l'esprit. Pour assurer la cohésion de l'ensemble, Frege est obligé d'intercaler un « saut mystérieux » entre les deux versants, subjectif et objectif, *via* le processus de « saisie d'une pensée ». En clair, le cœur de notre thèse peut être résumé ainsi :

> (T) Le clivage post-platonicien (tacite) entre les formes *objectivées* et les formes *subjectivées* de la pensée, provoque des fossés explicatifs (explicites), qui sont franchis (très souvent) *via* des sauts conceptuels (mystérieux), intercalés entre les étapes d'une argumentation.

Si notre hypothèse de départ est correcte, il s'ensuit que la distribution actuelle des débats sur l'axe états physiques (objectifs) / états qualitatifs (subjectifs) subit la même pression souterraine post-platonicienne, et qu'une bonne partie de l'énergie des penseurs actuels est dépensée à la recherche du « bon saut » ou « crochet » conceptuel permettant d'attacher les rives du mental à celui du physique. Or, à y regarder de plus près, on s'aperçoit que cette distribution éminemment statique a pour conséquence l'introduction d'une interface artificielle séparant soigneusement le domaine des expériences subjectives (privées, à la première personne) du domaine des objets physiques (publics, à la troisième personne). Comme le remarque fort bien Putnam[19], quelques postulats d'autant plus confortables, car tacitement adoptés, viendront renforcer au fur et à mesure ce tableau dichotomisant:

(1) il existe un domaine autonome des expériences ou des phénomènes mentaux ;
(2) ces phénomènes mentaux se déroulent dans l'esprit/cerveau humain ;
(3) le *locus* où se déroule cette intrigue est la tête ;
(4) enfin, la perception présuppose une relation cognitive spéciale connectant les expériences « internes » aux objets « externes ».

Une fois cette distribution acquise, il devient relativement facile d'orienter toute l'énergie des recherches vers la quête d'un point de départ suffisamment solide à partir duquel le sujet connaissant puisse «lever » la connaissance vraie au-dessus des sables mouvants de la subjectivité, autrement dit, établir un fondement stable à

[19] H. Putnam, (1999), *The Threefold Cord – Mind, Body, and the World*, New York: Columbia University Press, p. 30.

l' «édifice chancelant de notre savoir » pour reprendre ici l'expression de M. Schlick[20]. Or, toute enquête de ce type repose sur une hypothèse hautement suspecte selon laquelle le théoricien-philosophe serait en principe capable de *s'extraire* du tableau, ou du cadre du monde afin de mieux l'examiner, endossant en quelque sorte le point de vue d'un démiurge survolant le « nid de coucous » et photographiant, depuis le fauteuil confortable qu'offre l'analyse conceptuelle et la rationalité, la carte du réel. Le point sur lequel nous aimerions insister est le suivant : il se peut que le rêve platonicien de quitter la « caverne » de la matérialité afin d'accéder à la lumière de l'immatérialité rationnelle, contamine, bien plus que nous ne le pensons, la manière dont la pensée philosophique aborde la question épineuse des rapports entre le corps et l'esprit, et d'ici plus loin l'affaire des relations entre les *qualia* et les états neurophysiologiques du cerveau.

Il s'agit donc de montrer que, de manière générale, la solution-type aux difficultés verticales de *médiation* entre le subjectif et l'objectif, consiste à produire *in extremis* des concepts-hybrides, des concepts-relais ayant le don magique d'opérer des « sauts conceptuels » entre les pôles des dichotomies. Le terme de « saut conceptuel » nécessite un éclaircissement préalable plus substantiel. D'emblée, il est commode d'interpréter un saut conceptuel en s'inspirant des cas répertoriés de « sauts quantiques » dans la physique actuelle. Bien que métaphoriquement suggestive, cette stratégie risque de nous embrouiller plus que nous édifier sur l'affaire, car à l'heure même où nous écrivons ces lignes, la question de l'interprétation de la mécanique quantique est largement ouverte. Pour cette raison même, nous allons caractériser la notion de « saut conceptuel » telle que nous l'entendons ici, en faisant appel modestement aux outils conceptuels spéculatifs.

Selon l'interprétation que nous lui donnons ici, un saut conceptuel est une manœuvre explicative illicite et inséré *ad hoc*, censée assurer *a posteriori* la cohérence, l'assemblage ou le raccordement entre les rives d'un clivage conceptuel brut, i.e. introduit primitivement tel quel. Pour formuler les choses d'une manière romancée, disons que les sauts conceptuels sont l'envers nocturne des fossés

[20] M. Schlick, (1934), 'Über das Fundament der Erkenntnis', *Erkenntnis* 4, trad.fr. *in* Ch. Bonnet & P. Wagner (éds.), *L'âge d'or de l'empirisme logique*, Paris, Gallimard, 2006, p. 415. (Référence faite ici à cette dernière édition).

explicatifs, les deux phénomènes se présentant ainsi sous la forme d'une et seule monnaie conceptuelle.

L'essentiel pour nous est de mettre en évidence le lien qui se tisse entre les efforts d'analyse et de délimitation conceptuelle entraînant la migration des tensions philosophiques vers les lisières qui séparent les différents membres des dichotomies : dans le cas qui nous intéresse ici, les deux membres en question étant (i) les propriétés qualitatives de l'expérience phénoménale d'une part, et (ii) les propriétés physiques fondamentales d'autre part. La discussion de ces points fixera le canevas sur lequel nous allons broder par la suite la thèse qui sera à défendre ici, ainsi que les arguments qui sont à considérer en sa faveur.

Le chapitre 2 aura donc pour difficile tâche d'introduire, à titre d'exemple, deux cas concrets de « sauts conceptuels » interceptés dans un premier temps chez Frege et sa théorie des pensées objectives, et ensuite chez Carnap et son projet physicaliste de reconstruction rationnelle. Ces penseurs tentent de se débarrasser du pôle subjectivisant de la vie mentale. Mais la subjectivité, chassée par la porte de la logique, revient sur la porte de l'ontologie, et se « venge » en quelque sorte, obligeant les deux philosophes à se livrer à de véritables acrobaties conceptuelles afin de justifier la jointure entre la pensée et le monde. L'examen de ces points fixera la trame sur laquelle se déroulera notre enquête.

Le chapitre 3 esquisse la toile d'arrière plan qui installe le matérialisme au cœur de la métaphysique contemporaine de l'esprit. Il nous sera utile, dans un souci à la fois de cohérence et de concision, d'entamer notre enquête par une analyse topographique visant à esquisser un bref état des lieux fixant le problème avant de s'attaquer à son noyau central. Nous allons donc passer brièvement en revue les principaux points forts de cette évolution, en commençant avec le behaviorisme, en passant par la théorie de l'identité psychophysique, et allant jusqu'au fonctionnalisme. Nous verrons ainsi que le fonctionnalisme apporte une flexibilité conceptuelle qui allège considérablement la rigide et lourde ontologie du matérialisme de type réductionniste. Affirmer que la caractéristique fondamentale d'un état mental est celle d'être un système organisé d'une certaine manière, dispense les matérialistes de chercher l'équivalent d'une glande pinéale à pointer du doigt à chaque fois que le mental interagit causalement avec le physique. Par conséquent, l'abandon progressif de la théorie de l'identité psychophysique à partir

des années 70 ne provoqua pas le déclin du monisme matérialiste comme les dualistes pouvaient l'espérer. Au contraire, le maintien du physicalisme en tant que toile de fond unificatrice permit que la transition vers le fonctionnalisme se fasse, pour ainsi dire, sans douleur et sans regrets.

Le chapitre 4 (sections 4.1. et 4.2.) entame l'analyse du deuxième protagoniste de ce duel – les *qualia* – et prend comme point de départ le célèbre argument des fossés explicatifs de Levine. Considérons l'analogie suivante: aucun obstacle intuitif ne nous empêche de considérer que le liquide inodore et incolore nommé « eau » peut être expliqué en termes de structure moléculaire H_2O. Lorsqu'un scientifique nous explique les diverses connexions, conditions et structures qui font que l'eau bout à telle température, et s'évapore à une telle autre, rien de ce qu'il puisse nous expliquer ne nous semble arbitraire. Au contraire, lorsque nous voulons appliquer la même procédure de réduction aux processus phénoménaux, comme par exemple, l'expérience visuelle des couleurs, quelque chose, semble-t-il, est omis, laissé inexpliqué. Certes, les neurosciences arrivent à corréler certains stimuli visuels à certains processus neuronaux localisables dans certaines zones du cerveau. Mais ces corrélations sont loin d'offrir une définition satisfaisante, en termes physiques, de l'aspect qualitatif de l'expérience phénoménale. Comme le relève Levine, la connexion entre la description neurologique d'un événement et la conception « interne », à la première personne, que nous avons sur ce même événement, semble complètement « arbitraire »[21]. C'est précisément ce contraste, souligne Levine, entre l'arbitraire qui règne entre ces dernières corrélations, s'opposant vivement à l'intelligibilité que dégagent les autres classes de réduction de type « eau-H_2O », qui constitue le cœur du « fossé explicatif »[22], et qui constitue le point de départ des attaques antimatérialistes.

Toute tentative physicaliste d'annihiler le « fossé » - insiste Levine - devra respecter la « Contrainte Matérialiste[23] ». Autrement dit, la solution au problème du

[21] "The connection between the neurological description and our first-person conception of what it's like seems totally arbitrary. One feels that this neurological configuration could just as easily have gone with a bluish visual experience as a reddish one. In fact, for all we can tell, it could just as easily have gone with a state that was like nothing at all for the subject." J. Levine, (2007), 'Phenomenal Concepts and the Materialist Constraint', *in* T. Alter & S. Walter (éds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism*, Oxford: Oxford University Press, p. 145.

[22] (J. Levine, 1983)

[23] "The Materialist Constraint: namely, that no appeal be made in the explanation to any mental property or relation that is basic." (J. Levine, 2007: 150)

« fossé » ne doit faire appel à aucune entité ou relation mentale considérées comme étant fondamentales, ou primitives. La contrainte matérialiste de Levine peut être considérée comme un corolaire de la *contrainte naturaliste* que nous avons énoncée au début de cette introduction. Conséquemment, toute explication ayant pour finalité de dissoudre ou de franchir le gouffre explicatif, et qui fait appel à des items explicatifs non-physiques dans une architecture argumentative physicaliste, sera automatiquement disqualifiée. Dès lors, les tenants du physicalisme doivent se limiter à expliquer en termes *exclusivement physiques* la manière dont l'expérience qualitative phénoménale est implémentée et incorporée dans des systèmes et mécanismes physiques.

La section 4.3. présente les autres arguments antimatérialistes avec lesquels Mary partage la tête d'affiche, pour ainsi dire. Les sections 4.4. et 4.5. répertorient, en suivant de près la taxinomie proposée par Chalmers, les différentes tentatives de relever le défi posé par l'argument du « fossé dans l'explication ». Le but du quatrième chapitre sera donc de situer sur la carte des controverses concernant le problème des rapports entre la conscience phénoménale et l'information physique, les coordonnées officielles, si l'on peut les appeler ainsi, du matérialisme actuel. C'est seulement après ce survol nous permettant une vue panoramique sur les cadres du débat, que nous introduirons enfin, dans la deuxième partie de cet ouvrage, la vedette de notre histoire, Mary la prisonnière.

Le chapitre 5 procède à un examen détaillé des différentes variantes de l'argument de la connaissance, ainsi que des ripostes qu'il a suscitées. L'histoire de Mary est devenue, en quelque sorte, l'échiquier sur lequel les diverses subdivisions de la philosophie analytique actuelle croisent l'épée et ravivent de vieille querelles philosophiques sans âge. La force de cette désormais célèbre expérience de pensée de Jackson tient au fait qu'elle tente de transformer, *via* une alchimie mystérieuse, « le métal ordinaire des prémisses épistémiques en l'or noble des conclusions ontologiques »[24]. Inutile de dire que nous ne prétendrons pas résoudre dans cet espace limité, une question aussi complexe que celle de savoir si l'argument de Jackson a réussi à destituer le physicalisme de sa fonction de 'Weltanschauung' des temps modernes. Mais nous voudrions, en toute simplicité, jeter un faisceau de

[24] Y. Nagasawa, (2008), *God and Phenomenal Consciousness – A Novel Approach to Knowledge Arguments*, Cambridge: Cambridge University Press, p. 3.

lumière sur les vifs débats qui animent certaines controverses autour de la question épineuse des *qualia*. Pour des raisons qui seront étayées au fur et à mesure de notre investigation, nous pensons que l'impact de l'argument de la connaissance de Jackson dans la philosophie de l'esprit contemporaine constitue un excellent prisme à travers lequel le débat mettant face à face les défenseurs du physicalisme et ses détracteurs, peut être remis en question.

L'examen de l'argument de la connaissance sera divisé en deux étapes. La première étape occupe tout le chapitre 5 et concerne *la thèse de l'apprentissage* de Mary. Autrement dit, il s'agit de répondre à la question : qu'est-ce que Mary *apprend* lors de sa sortie de prison ?

La deuxième étape occupera le chapitre 6, et traitera *la thèse de la complétude* de la connaissance physique que Mary possède durant sa captivité. A la suite de Nagasawa, nous allons proposer un léger changement de perspective, et nous allons traiter la question de l'omniscience de Mary en lien étroit avec la thèse de l'omniscience dans la philosophie de la religion, et plus particulièrement, en dressant une analogie entre l'argument de Jackson d'une part, et l'argument de la possession du concept, d'autre part. L'examen de ces points jettera un éclaircissement fort intéressant, à notre sens, sur l'affaire de Mary et des éventuelles implications sur le physicalisme.

Enfin, le chapitre 7 constitue la synthèse vers laquelle converge le travail agencé dans cet essai. Il bâtit, avec les éléments que les autres chapitres lui ont mis à sa disposition, l'armature de la thèse que nous défendons ici. D'abord, les sections 7.1. et 7.2. présentent les conséquences de la thèse de la déductibilité *a priori*, les ambitions du physicalisme *a priori*, ainsi que les différentes objections qu'il doit affronter. A partir de la section 7.3. nous adresserons notre propre critique au principe du passage *a priori* du « deuxième » Jackson, en suggérant que ce type de physicalisme radical prête le flanc aux mêmes difficultés que son prédécesseur, le physicalisme sémantique de Carnap. Nous intercepterons ainsi un saut conceptuel au sein même des tentatives acrobatiques de Jackson de relier, *via* l'analyse conceptuelle, les vérités analytiques « fortunées » aux vérités factuelles « infortunées ». L'analyse de ces points nous amènera à soutenir, à la suite de S. Blackburn, que malgré l'appareillage conceptuel sophistiqué de la philosophie actuelle opérant avec « le contingent *a priori* », « les nécessités *a posteriori* », « les

mondes possibles », « la survenance », « les énoncés Ramsey », « les propriétés *de se* », « la logique modale bidimensionnelle », etc., lorsque la machinerie d'une argumentation bute sur la difficulté de la médiation entre le versant subjectif et celui objectif, il semblerait que les options qui nous restent ne soient pas bien différentes de celles signalées par des prédécesseurs comme Locke et Leibniz. Autrement dit, soit on fait appel au « bon plaisir de Dieu », mais dans ce cas on tombe dans le « mystérianisme » ; soit on défend les bienfaits de l'analyticité et de la rationalité *a priori*, mais dans ce cas, on se livre à de véritables acrobaties conceptuelles afin de maintenir la cohérence de l'ensemble.

Démasquer la présence d'un « saut conceptuel » au cœur même de la stratégie physicaliste *a priori* nous permettra de refermer la boucle d'avec l'échec du physicalisme sémantique de Carnap. Rappelons-nous que selon Carnap, c'est en manipulant des « symboles formels » - ces derniers se détachant en quelque sorte sur l'arrière-plan hérité des « Idées » platoniciennes - qu'on arrive à architecturer le sens de la réalité. Dès lors, on prend conscience de la possibilité d'une reconstruction rationnelle des fondements de la connaissance en s'appuyant sur la syntaxe logique du langage. Certes, comme on l'a déjà remarqué, Carnap n'est « ni le seul, ni le premier à concevoir la possibilité d'appliquer les moyens logiques des *Principia Mathematica* à des champs de la connaissance qui dépassent le cadre des mathématiques »[25]. Cependant, son projet de réduction physicaliste a été mené dans un esprit de rigueur et d'ascèse remarquables, qui ont fini par laisser une empreinte indélébile sur la pensée analytique. Comme l'a relevé F. Rivenc, il s'agit de prendre note que "le mode contentuel du discours, qui croyait parler de choses, d'objets, de concepts, de nombres, est banni au profit du mode purement formel, qui ne connaît que signes, enchaînements de signes, enchaînements de ces enchaînements suivant des règles de construction et de transformation"[26]. Or, dans son examen critique des dogmes empiristes, Quine[27] montrera de façon très convaincante que le champ d'investigation récupéré par le discours formel ne peut pas couvrir, en parfait

[25] P. Wagner, (2002), 'Le contexte logique de l'Aufbau', in P. Wagner, (ed.), *Les philosophes et la science*. Paris: Gallimard, p. 29.
[26] F. Rivenc, (1993). *Recherches sur l'universalisme logique*. Paris: Payot, p. xi.
[27] W.V. Quine, (1951a), 'Two Dogmas of Empiricism', *Philosophical Review* 60, pp. 20-43, réédité in W.V. Quine *From a Logical Point of View*, Cambridge, Mass.: Harvard University Press, 1953, trad.fr. par P. Jacob *Deux dogmes de l'empirisme*, *in* P. Jacob, (ed). (1980), *De Vienne à Cambridge - l'héritage du positivisme logique*. Paris: Gallimard, p 107. (Dans ce qui suit, nous allons nous référer à cette édition).

"copier-coller", si l'on peut dire ainsi, la verticalité de la sémantique. Une vieille querelle sera ainsi ravivée, entre Carnap et Quine, au sujet du statut des entités abstraites (section 7.3.5.).

Dans la section 7.3., nous argumenterons l'idée que l'organisation même des débats autour du problème des rapports entre les états mentaux et les états physiques, est tributaire de l'adoption tacite d'une hypothèse-intrus, qui « contamine » la distribution des débats sur cet axe. Une fois cet intrus démasqué et neutralisé, nous verrons qu'il n'y a vraiment aucun problème « difficile » de la localisation de l'esprit dans le monde physique à résoudre. Il suffit, à l'instar de Van Gulick[28], de montrer que les faits subjectifs « infortunés » ne sont pas incompatibles avec le physicalisme. Mais cette solution a un prix : d'abord, il faut rendre le physicalisme à nouveau « humble », i.e. empirique, et respectueux des découvertes que les laboratoires lui mettent à sa disposition à la sueur de leur front. Pour ce faire, nous avons besoin de « naturaliser » à la fois la métaphysique en fauteuil et le physicalisme *a priori*. Cette naturalisation présuppose inévitablement de dénoncer les dogmes sur lesquels le physicalisme actuel bâtit ses châteaux de cartes abstraites. Le traitement de ces points hautement controversés sera réalisé dans la section 7.4. Nous conclurons que, paradoxalement, le physicalisme en tant que 'Weltanschauung' de l'époque moderne, *ne* passe *pas* le test du naturalisme tel qu'évoqué plus haut. La question qui surgira alors sera la suivante : comment expliquer ce formidable paradoxe qui place le physicalisme en conflit avec les sciences physiques, celles-là mêmes qui le justifient dans son ambition de doctrine ontologique unificatrice ?

Enfin, nous terminerons notre enquête ici en livrant notre propre « solution » au problème de Mary (chapitre 8). Plus spécifiquement, il s'agit de reprendre l'argument de la connaissance et de lui faire passer, à lui aussi, le test du naturalisme. Si l'argument ne passe pas le test, autrement dit, s'il fait appel dans ses trames argumentatives à un élément qui transcende l'ordre naturel « d'ici bas », alors, il sera considéré comme irrecevable.

A la lumière de ces considérations, il devient évident que la philosophie actuelle doit entamer un grand chantier de « ménage » conceptuel destiné à balayer

[28] R. Van Gulick, (1985), 'Physicalism and the subjectivity of the mental'. *Philosophical Topics*, 13: 51-70.

tous les présupposés tacites qui la poussent à tourner en rond, dans une spirale infernale de « fossés explicatifs / sauts conceptuels ». Il s'agit donc de penser une vision alternative non-schismatique, i.e. une stratégie insistant sur l'impératif de redéfinir les rapports entre le mental et le physique en s'appuyant sur une conception plus souple[29] et plus ouverte du réel, misant, pourquoi pas, sur le mariage inattendu entre la mécanique quantique et la phénoménologie.

Il est certainement très tentant, pour le philosophe, de s'installer dans un fauteuil abstrait, et de tracer un cercle symbolique autour, capable de le protéger, à la manière des cercles des sorcières dans les contes de fées, de sables mouvants des présupposés subjectifs « de tous les jours ». A l'intérieur de ce cercle, l'illusion de parfaite accessibilité de la rationalité semble à portée de main. Mais la poursuite de cet idéal épistémique de parfaite accessibilité provoque – comme l'a fort bien relevé T. Williamson[30] – un étrange paradoxe : car plus on durcit les conditions de définition de la vérité et de l'évidence, moins c'est « évident » de saisir comment réaliser le passage depuis la vérité logique vers les énoncés du monde familier.

Essayons de synthétiser, en guise de conclusion, les principales étapes du travail d'investigation que nous allons entamer ici. D'abord, une question de fond:

Pourquoi avoir choisi le thème du duel entre le physicalisme et les avocats de l'argument de la connaissance ?

Le choix de notre périmètre de recherche a été motivé par l'évidente ascension et influence du physicalisme dans la philosophie de l'esprit contemporaine. Le point de départ de notre enquête constitue les rapports sous haute tension entre le physicalisme conçu comme une théorie explicative exhaustive et complète du réel, d'une part, et le problème « difficile »[31] de la conscience phénoménale d'autre part. L'argument de la connaissance de Jackson constitue un excellent carrefour diffuseur d'interrogations clé autour desquelles se déroule, à l'heure même où nous écrivons

[29] M. Bitbol, (2010), *De l'intérieur du monde: pour une philosophie et une science des relations*, Paris, Flammarion.

[30] "If we imagine that some candidate criterion of rationality is perfectly accessible, then we are always likely to prefer that criterion; but once we recognize that *the perfect accessibility is quite generally an unattainable ideal*, we can learn to live with an imperfectly accessible criterion. We have nothing else to live with". T. Williamson, (2000), *Knowledge and its limits*, Oxford: Oxford University Press, p. 16. (Nous soulignons)

[31] (D. Chalmers, 1996)

ces lignes, le combat pour l'instauration ou la destitution de la suprématie physicaliste.

Quelle est la thèse que nous voulons défendre ?

Nous défendrons ici l'idée selon laquelle le problème de l'expérience qualitative phénoménale *est devenu un problème* suite à l'adoption d'un modèle d'intelligibilité rigide et dichotomisant, dont l'influence plus ou moins souterraine frappe de plein fouet la philosophie contemporaine de l'esprit. Démasquer les véritables mécanismes qui sous-tendent cette influence souterraine est devenu, à notre sens, une urgence pour la tradition analytique. Bien que les efforts de conceptualisation convergent actuellement vers l'élimination définitive des résidus idéalistes et supranaturels d'une pensée nourrie plutôt des mythes que du réel, nous argumenterons ici que l'*antisurnaturalisme*[32] prédominant dans la philosophie actuelle est loin de constituer une condition suffisante pour l'éradication pure et simple de toute mythologie prête à saper les fondations scientifiques du 'Weltanschauung' physicaliste.

Bien plus, nous serons surpris de constater que beaucoup de défenseurs acharnés du matérialisme contemporain non seulement luttent contre le mauvais ennemi, mais dans le feu de l'action, ils emploient les mêmes outils et les mêmes stratégies argumentatives que leurs prédécesseurs idéalistes, ou leurs ennemis jurés, les dualistes. Cet étrange état de choses est visible à l'œuvre, pensons-nous, dans la manière où les philosophes contemporains gèrent l'affaire du physicalisme et des niveaux de réalité censés accueillir les énigmatiques *qualia*. Dans ce qui suit, nous proposons de changer de perspective, et d'aborder le débat, si possible, en « survolant le nid de coucous », s'il nous est permis d'emprunter cette métaphore au cinéma de Milos Forman. Après tout, si le physicalisme ambitionne d'offrir une explication de ce que les choses *sont* ultimement, il sera difficile de s'en tenir uniquement aux aspects épistémiques de l'affaire. Le véritable enjeu dans l'affaire du physicalisme, comme nous allons le voir, dépasse les cadres d'une conception scientifique, et par conséquent falsifiable, du réel, et touche de près à la question métaphysique de l'universalité sur laquelle tout 'Weltanschauung' fonde ses aspirations de complétude et d'exhaustivité. Avant d'entamer notre périple, il nous faut également prendre certaines précautions et préciser qu'il ne s'agit là nullement

[32] J. Dupré, (2004), 'The Miracle of Monism', in M. de Caro & D. McArthur, (éds.) op. cit., p. 36.

de fournir une recherche historique ou une monographie du physicalisme. Plus modestement, notre intérêt se limitera au défi lancé aux partisans du physicalisme par les défenseurs de l'argument de la connaissance, et s'attachera aux réponses et repliements que cette attaque a provoqués parmi ses artisans.

L'histoire qui nous intéresse ici est le récit d'un argument à succès (l'argument de la connaissance de Jackson) qui a obligé, comme tout bon argument doit le faire, les adeptes de la partie adverse à se replier, redéfinir et redistribuer leurs pions argumentatifs en positions de plus en plus claires, et de plus en plus offensives. Suivre les sinuosités dialectiques véhiculant le contenu des arguments et contrarguments forgés tout au long de ce débat (qui se poursuit d'ailleurs à l'heure où nous écrivons ces lignes) constitue, à notre sens, un excellent exercice argumentatif, qui nous permettra à la fois de diagnostiquer et de prendre le pouls du cœur du dilemme que pose l'existence de l'esprit dans une représentation physicaliste du monde.

Quelle sera la stratégie adoptée ?

La stratégie que nous allons adopter ici est quelque peu hétérodoxe, car elle prend comme hypothèse de départ la prémisse inhabituelle selon laquelle le physicalisme et son adversaire direct, le dualisme, ne sont, à la base, que les visages de Janus d'un seul et même courant de pensée souterrain dont l'adoption tacite provoque l'apparition, en spirale, de ce que nous allons appeler, le tandem « fossés explicatifs/sauts conceptuels ».

Quelle conclusion pour la fin?

Le duel entre le credo physicaliste « Tout est physique » et la riposte antiphysicaliste « Tout, sauf les *qualia* ! » constitue le cœur de notre enquête ici. Mais tout bon scénariste doit se garder de dévoiler le dénouement de l'intrigue avant la fin du film. Disons seulement ceci : nous allons faire passer aux protagonistes du duel que nous poursuivons ici le *test du naturalisme*. Sauront-ils s'affronter sans « tricher » pour ainsi dire, c'est-à-dire sans faire appel à des éléments surnaturels ? Allons voir à présent s'ils passent ce test…

Première partie

Le Physicalisme

1 L'hégémonie du physicalisme aujourd'hui

> « *Mais quand on nous aura tout expliqué, quand tout sera mis en ordre et fixé à l'avance (ce qui est fort possible, car il est ridicule, car il est stupide de croire que certaines lois de la nature demeureront indéchiffrables), alors, évidemment, il n'y aura plus place pour ce qu'on nomme les désirs* ».
>
> F.M. Dostoïevski, *Le sous-sol*

C'est le mois de juin à Venise. Imaginez-vous somnolant au soleil, allongé sur un transat sur la plage du Lido. Pendant que vous regardez d'un œil indolent le bleu de l'Adriatique, vous ressentez soudainement le désir de manger des fraises. A cet instant précis, quelqu'un s'approche de vous et, d'une voix sûre, vous prie de lui communiquer la *masse* de votre *désir* de manger des fruits, ou les *dimensions* de votre *croyance* selon laquelle manger des fraises apaisera votre *envie* de fruits frais. Quelle serait votre première réaction ? Admettez qu'une telle requête vous paraîtrait parfaitement absurde, et une fois l'effet de surprise passé, vous songeriez sérieusement à changer de transat, de plage, de ville, ou même de pays afin d'éviter l'étrange individu vous importunant avec ce genre de questions extravagantes. En revanche, si l'inconnu en cause vous interroge sur la masse, la forme, et les dimensions des fraises, ou du transat dans lequel vous somnolez, ou même de la mer Adriatique, vous serez probablement enclin à croire qu'il s'agit d'un canular, d'un sondage, ou d'un reportage, etc. Vous vous sentirez probablement irrité, vexé, ou au contraire, flatté d'avoir été choisi, mais ces questions en elles-mêmes n'ont rien d'absurde d'après vous. Nous avons ici la première et la plus intuitive distinction entre les états mentaux et les états physiques : à la différence des chats, des transats, et des fraises, les états mentaux comme les désirs, les croyances, ne semblent pas pouvoir faire l'objet d'une caractérisation spatiale.

De Venise, changeons de décor et rendons-nous en Bourgogne : supposons à présent que vous êtes à Nuits-Saint-Georges, à une dégustation de vin entre amis, en train de savourer un exquis verre de vin rouge. Poussons l'expérience de pensée un peu plus loin et supposons qu'au même moment un savant fou scrute, à l'aide d'une machinerie sophistiquée, les zones de votre cerveau activées durant cette dégustation.

Supposons également que vous êtes suffisamment courageux pour regarder en direct, projetées sur un écran, les images de votre crâne ouvert sous la machinerie. Force est de constater alors que rien dans l'aspect gris et visqueux de la substance que vous percevez sur l'écran ne laisse présager l'expérience qualitative particulière que vous ressentez lorsque vous délectez votre palais avec un verre de grand cru.

A première vue, l'existence de ces aspects différenciant le domaine du mental de celui du physique semble étayer l'hypothèse dualiste traditionnelle selon laquelle les substances /propriétés/ prédicats mentaux sont (i) des éléments essentiels pour une description complète du monde, et (ii) irréductibles à des substances / propriétés / prédicats physiques. Certains penseurs font appel au principe leibnizien de l'indiscernabilité des identiques[1] pour appuyer l'hypothèse dualiste : si les états mentaux *X* sont identiques aux états physiques *y*, alors une propriété de *X* est une propriété de *y* et vice-versa. Or, étant donné que la masse, par exemple, n'est pas un attribut de *X*, il s'ensuit que *X* n'est pas identique avec *y*. Par conséquent, les états mentaux ne sont pas *identiques* ni *réductibles* aux états physiques[2].

Physique ou non-physique, il faut se décider. Bref, comme le relève P. Pettit[3], aucun philosophe contemporain de l'esprit ne peut se permettre le luxe d'ignorer le physicalisme : entre ceux qui le rejettent, ceux qui le défendent et, au milieu, ceux qui essaient de le redéfinir, en espérant qu'un changement de perspective résoudra le mystère d'un coup de baguette aussi surprenant que définitif, la valse des arguments et des contre-arguments n'a pas fini d'être jouée. Sans doute, le physicalisme est-il un visage incontournable de la philosophie contemporaine. Mais il n'en demeure pas moins qu'aucun physicaliste ne peut se permettre le luxe d'ignorer le problème que pose la « localisation »[4] de l'esprit dans le réel physique. Comme le relève T. Crane[5], le problème des rapports entre le mental et le physique comporte deux volets, à savoir (a) le problème de la causalité mentale, et (b) le problème de la conscience, volets qui

[1] Très sommairement décrit, le principe leibnizien affirme que si deux choses sont identiques, alors toutes leurs propriétés sont identiques (pour tout *X* et tout *y*, si *X* est identique avec *y*, alors pour toute propriété *F*, *X* a *F* et si et seulement si *y* a *F*).

[2] Certains auteurs qui visent s'inscrire en faux contre la position dualiste, affirment qu'il est possible d'imaginer que les savants du futur découvriront qu'un état mental comme la douleur, par exemple, *possède* une masse, ou que les désirs exercent une force d'attraction gravitationnelle, etc. Voir D. Braddon-Mitchell & F. Jackson, (1996), *Philosophy of Mind and Cognition*, Oxford: Blackwell, p. 5.

[3] P. Pettit, (2008), 'Consciousness and the Frustrations of Physicalism', in I. Ravenscroft (ed.), *Minds, Ethics, and Conditionals Themes from the Philosophy of Frank Jackson*, op. cit., p. 164.

[4] F. Jackson, (1998), *From Metaphysics to Ethics: A Defense of Conceptual Analysis*, Oxford: Clarendon.

[5] T. Crane (2000), 'The Origins of *Qualia*', in T. Crane & S. A. Patterson (eds.), *The History of the Mind-Body Problem*, Routledge.

peuvent être regardés comme les deux cornes d'un dilemme. D'une part, l'interaction causale entre le mental et le physique exige que tous les phénomènes mentaux causalement efficaces soient physiques. D'autre part, la présence du phénomène de la conscience semble indiquer que pas tous les phénomènes sont physiques.

Selon D. Chalmers[6], ce qui rend le problème de l'expérience phénoménale aussi « difficile », est le fait que malgré une connaissance approfondie et minutieuse des substrats neuronaux qui sous-tendent les fonctions cognitives que l'on situe habituellement à la lisière de la conscience – comme, par exemple, la capacité de discrimination, d'intégration, de contrôle, d'accès, etc. – la réponse à l'interrogation « pourquoi la mise en œuvre de toutes ces fonctions est-elle accompagnée de l'expérience phénoménale ? » nous laisse sinon perplexes, du moins suspendus au-dessus d'un véritable vide explicatif. C'est la persistance obstinée avec laquelle cette difficulté revient comme un boomerang sur le devant de la scène, après avoir été écartée par les premiers analytiques, qui confère au problème de l'expérience le titre édifiant de « problème ardu ». Certes, au vu des progrès scientifiques considérables accomplis ce dernier siècle, il serait pour le moins légitime d'espérer que nous accéderions un jour à une explication du phénomène de la conscience en termes purement physicalistes. Mais en attendant, souligne Chalmers, il serait hasardeux de bâtir des explications uniquement sur les bases spéculatives que laisse entrevoir un tel espoir messianique. Pour l'instant, insiste-t-il, la seule certitude à laquelle nous sommes confrontées concerne cette « résistance » têtue qu'oppose le phénomène de la conscience à toute tentative de réduction hâtive à un niveau purement physique.

Dans la vie de tous les jours, nos attributions d'états mentaux étayent un éventail très large d'expressions. Si une guêpe me pique le nez, je m'écrie « Aïe ! Ça fait *mal* ! » Mais je peux également m'écrier lors d'un concert de musique classique : « Ça me fait *mal* d'entendre cette magnifique sonate de Scarlatti jouée si péniblement ! » L'état-douleur provoqué par la guêpe n'a pas le même *sentir qualitatif* que celui provoqué par le pianiste malhabile. On a tôt fait de remarquer qu'aucun phénomène mental n'est plus central à la compréhension adéquate de l'esprit que la conscience. En clair, quoi qu'on puisse affirmer de vrai des états mentaux, il n'en demeure pas moins qu'on ne considérerait pas du tout un état comme mental s'il n'avait pas la propriété de phénoménalité.

[6] (D. Chalmers, 1996)

Tel que nous le connaissons aujourd'hui, le problème de la place de la conscience dans le monde naturel fut explicitement abordé dans un ouvrage publié en 1925 par C. D. Broad et qui porte le titre significatif de *The Mind and its Place in Nature*[7]. Bien que le problème de la « localisation » de l'esprit dans le monde physique puisse être situé historiquement, la séparation esprit/matière, elle, n'a pas d'histoire. Difficile de spéculer sur les origines de ce clivage. On pourrait suggérer, à la suite de W. James[8], que le fait le plus primitif et le plus concret auquel nous sommes confrontés en tant que sujets de l'expérience, est la prise de conscience du fait même que nous sommes les sujets de cette expérience interne, privée.

Dire que la marque distinctive du mental est la phénoménalité nous oblige à avancer un pas plus loin et à s'interroger : qu'est-ce que la *conscience phénoménale* ? D'emblée, la réponse à cette question comporte deux difficultés considérables.

La première tient à l'ineffabilité même du phénomène de la conscience qui semble se dérober à toute manœuvre de type « arrêt sur image », i.e. à toute tentative de le figer sur un cadre immobile, un peu comme on épingle un papillon sur du papier de soie. La conscience nous glisse entre les doigts, comme par magie. Ensuite, en supposant que l'on parvienne un jour à la fixer sur une toile de fond stable, la question qui se posera alors sera celle de savoir quelle est la place de la phénoménalité dans le monde physique des transats et des fraises. Commençons par décrire avec plus de précision le concept de « monde physique ».

Sans doute, la définition à la fois la plus simple et la plus provocatrice que l'on pourrait donner au physicalisme serait de dire que si Dieu[9] était le véritable créateur de l'univers, alors tout ce qu'il aurait eu à faire, lors de la création, aurait été de créer les faits et les lois physiques, le reste découlant nécessairement et naturellement de celles-ci. Certains auteurs, comme T. Burge[10], ont poussé l'analogie encore plus loin,

---

[7] C.D. Broad, (1925), *The Mind and its Place in Nature*, New York: The Humanities Press Inc, London: Routledge et Kegan Paul LTD.

[8] "The first and foremost concrete fact which everyone will affirm to belong to his inner experience is the fact that *consciousness of some sort goes on.*" *W.* James, (1910), *Psychology*, New York: Henry Holt and Co. Fragments repris in N. Block, O. Flanagan, O. et al. (eds.) (1997), *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press. (Nous faisons ici référence à cette dernière édition).

[9] T. Crane (1991), '*All God has to do*', *Analysis* 51: 235-44.

[10] "The flood of projects over the last two decades that attempt to fit mental causation or mental ontology into a 'naturalistic picture of the world' strike me as having more in common with political or religious ideology that with a philosophy that maintains perspective on the difference between what is known and what is speculated Materialism is not established, or even deeply supported, by science". T. Burge, (1993), 'Mind-Body Causation and Explanatory Practice', *in* J. Heil & A. Mele (eds.) *Mental Causation*, Oxford: Clarendon Press, p. 117.

assimilant l'engouement actuel pour le physicalisme à la grande, dangereuse et puissante famille des idéologies religieuses et politiques.

Le physicalisme a entamé la belle carrière qu'on lui connaît actuellement en s'épanouissant sous la haute protection des progrès scientifiques qui ont changé le visage de la modernité au cours du XXe siècle. Ce spectaculaire succès a encouragé les philosophes matérialistes à resserrer leurs rangs et à systématiser leurs doctrines autour du slogan « *tout est physique* ». Ce slogan agira d'ailleurs, nous allons le voir, en véritable boîte de Pandore libérant les vieux démons de la métaphysique traditionnelle. Dire que « tout est physique » ne résout évidemment pas d'un simple coup de baguette magique tous les problèmes philosophiques qu'il subsume. D'importants désaccords surgissent dès qu'il s'agit de préciser la nature et le rôle des états mentaux, autrement dit, de regarder de plus près la question épineuse des entités telles que les croyances, les désirs, les sensations, les perceptions, les qualités phénoménales, etc., bref, ces phénomènes très familiers qui accompagnent l'individu dans son quotidien. C'est autour de ce « noyau dur »[11] que s'agenceront par la suite les controverses les plus ardues et les plus fécondes qui agitent le cœur des débats actuels en philosophie de l'esprit. Fondamentalement, ce qu'on reproche aux théories physicalistes de l'esprit c'est d'omettre l'aspect qualitatif ou conscient de la vie mentale de l'individu.

Comme le note B. Loewer[12], plusieurs thèses accompagnent et soutiennent, telle une armature souterraine, l'hégémonie du physicalisme à notre époque. En premier lieu, la physique est un modèle paradigmatique de la connaissance scientifique : toutes les autres sciences (et vérités) sont réductibles à la physique. En deuxième lieu, les seules propriétés, événements et objets authentiques sont ceux étudiés par la physique fondamentale. Troisièmement, les seules lois authentiquement fondamentales sont les lois de la physique. Enfin, la seule relation de causation réelle et effective est celle impliquée par des événements physiques. Pour saisir l'ampleur et l'enthousiasme engendrés par la promesse d'universalité contenue dans les lois de la physique, il suffirait de suivre le sillage que laisse derrière elle l'histoire du lien inextricable qui unit les interrogations du Philosophe à celles du Savant. Une analyse historique et

[11] C. McGinn, (1989), 'Can we solve the mind-body problem?', *Mind*, 98: 349-66. Réédité in R. Warner et T. Szubka, (eds.), (1994), *The Mind-Body Problem - a Guide to the Current Debate*, Cambridge Massachusetts: Blackwell, p. 99. (Nous faisons référence ici à cette dernière édition)

[12] B. Loewer, (2001), 'From Physics to Physicalism', in C. Gillett et B. Loewer (eds.), *Physicalism and Its Discontents*, Cambridge: Cambridge University Press, pp. 41-42.

minutieuse de ce lien ne fera que déborder les modestes cadres de cet essai. Nous nous en tiendrons donc à certains aspects dans l'affaire du physicalisme, sans prétendre ici à une quelconque exhaustivité.

1.1 De phusis à la physique et de la physique au physicalisme

Quelques rapides remarques terminologiques pour commencer. Il n'est pas inhabituel dans la littérature analytique contemporaine, d'utiliser les termes « physicalisme » et « matérialisme » de façon interchangeable pour désigner la même doctrine qui attribue aux entités peuplant le monde (y compris les entités mentales) une nature ultimement physique ou matérielle. Certains auteurs[13] affichent ouvertement leur préférence pour l'utilisation du terme « matérialisme » au détriment du « physicalisme », en arguant que ce dernier comporte une connotation réductionniste, tandis que le « matérialisme » se veut une appellation plus souple.

D'autre part, il est vrai que le matérialisme fait référence à une époque où l'objet d'étude de la physique était *la matière*. Il est vrai également que par la suite, des recherches plus approfondies ont mis en évidence l'existence des propriétés et des entités naturelles d'un type spécial (champs, ondes, énergie, etc.) débordant les cadres classiques de la notion de « matière ». D'autre part, note D. Lewis[14], abandonner le terme « matérialisme » pour cette raison, serait une pédanterie comparable à celle des dirigeants des Chemins de Fer Britanniques lorsqu'ils décidèrent de changer l'appellation des cabines «seconde classe » en cabines « classe standard » afin de ne pas léser l'amour-propre des citoyens « middle-class ». Dans ce qui suit, nous allons nous conformer à la pratique habituelle des matérialistes comme D. Armstrong[15] et J.J. Smart, et utiliser 'physicalisme' et 'matérialisme' de manière interchangeable, mais pas

[13] Voir W. Seager (2000), *'Physicalism'* in W. Newton-Smith (ed), *A Companion to the Philosophy of Science*, Blackwell Publishing. Voir aussi M. Lockwood, (1989), *Mind, Brain and the Quantum*, Cambridge, MA: Basil Blackwell Ltd.

[14] D. Lewis, (1994), 'Reduction of Mind', *in* S. Guttenplan, (ed), *A Companion to the Philosophy of Mind*, Oxford, Blackwell, p. 413

[15] "I follow J.J. Smart (1963) and identify contemporary Materialism (or Physicalism) with the view that the world contains nothing but the entities recognized by physics. Contemporary Materialism takes a Realistic view of the theoretical entities of physics – molecule, atom, fundamental particle, and so on – and then asserts that everything there is wholly constituted by such entities, their connections and arrangements". D.M. Armstrong, (1995), 'Naturalism, Materialism, and First Philosophy' in P. Moser et J. Trout, (eds.), *Contemporary Materialism*, London: Routledge, p. 40.

avant de prendre quelques précautions, c'est-à-dire non sans clarifier certains aspects révélateurs, à notre sens, des raisons sous-jacentes qui ont motivé les philosophes analytiques dans ce choix d'interchangeabilité.

Bien que le terme proprement dit de « matérialisme » n'ait fait son apparition qu'assez tard, au cours du XVIIIe siècle, la doctrine stipulant qu'il n'est de réalité que matérielle semble avoir préoccupé depuis longtemps la pensée philosophique. Le terme est attesté en anglais dès 1668 et introduit en français par Leibniz en 1702 pour exprimer la conception selon laquelle « il n'y a que de la matière, et qu'elle suffit pour tout expliquer »[16]. La racine étymologique du terme « matière » - le latin « *mater* » - renvoie à l'idée de récipient originaire, indifférencié, prêt-à-modeler par l'intervention d'une forme extérieure. Chez les philosophes grecs, le premier à avoir conceptualisé la notion de « matière » (« hule ») semble avoir été Aristote, bien que des réflexions sur la nature des éléments composant l'univers soient décelables bien avant, chez les présocratiques par exemple, ou chez les atomistes Leucippe et Démocrite. Comme le relève B. D'Espagnat, aujourd'hui, dans la pensée du plus grand nombre, « le "mot matérialisme" évoque une vision principalement atomistique de la nature. Il s'agit d'un réalisme *objectiviste* et même d'un réalisme *proche*, dont le concept central est dérivé, par simplification et abstraction, de notre expérience des objets solides. L'atome philosophique – autrement dit la particule – y est conçu au moyen d'un passage à la limite : on songe implicitement à une suite de grains de sable ou de poussière de diamètres toujours plus petits. (…) On pose enfin que le monde tout entier – y compris nous et nos cerveaux – est constitué de tels "atomes" et que les phénomènes qui s'y produisent, pensée comprise, sont, en définitive, exclusivement les conséquences de l'existence de ces atomes et des forces qui les relient »[17].

Le terme « phusikè » quant à lui, désigne en grec « choses de la nature », et de ce fait, reste directement dépendant de l'observabilité empirique. Il est néanmoins intéressant de noter qu'étymologiquement, le terme « physique » partage la même racine que le mot grec « phusis » désignant à la fois l'*origine* et l'*effet* du déroulement d'un processus : « la nature, dans son sens primitif et fondamental, c'est l'essence des êtres qui ont, en eux-mêmes et en tant que tels, le principe de leur mouvement. La

[16] *Grand Dictionnaire de la Philosophie* (sous la direction de M. Blay), Larousse - CNRS Editions, Paris, 2003, pp. 651-2.
[17] B. d'Espagnat, (2002), *Traité de physique et de philosophie*, Paris : Fayard, p. 305. (Soulignements en original).

matière, en effet, ne prend le nom de nature que parce qu'elle est susceptible de recevoir en elle ce principe ; et le devenir et la croissance, que parce que ce sont des mouvements procédant de ce principe. La nature, en ce sens, est le principe du mouvement des êtres naturels, immanent en quelque sorte, soit en puissance, soit en entéléchie »[18].

Deux acceptions du prédicat « matériel »/« physique » ont été prédominantes dans la philosophie occidentale. (i) Dans un premier temps, on s'accorde à dire qu'une entité est « matérielle » si elle *occupe une extension dans l'espace* ; cette version explicative reste traditionnellement attachée à la célèbre distinction cartésienne entre *res extensa / res cogitans*. La principale objection que soulève habituellement ce type de définition concerne le danger de circularité : la notion même d'« extension dans l'espace » dépendant intimement de celle de « matière », l'*explanans* et l'*explanandum* risquent donc de se confondre dans un amalgame peu informatif sur la nature effective du réel.

(ii) Une deuxième définition, plus récente, se contente de puiser dans le vocabulaire des théories scientifiques ayant déjà acquis droit de cité dans le cercle restreint des sciences physiques, et ayant déjà mis en place un schéma conceptuel explicatif à succès. Considéré sous cet angle, le prédicat « est physique »/« est matériel » est tenu pour faire référence à une entité physique si et seulement s'il existe des théories qui l'englobent avec succès dans leur architecture explicative. Nous aurons l'occasion de revenir plus en détail sur ce sujet. Pour le but introductif de ce chapitre, notons seulement que selon la définition standard[19], un élément est tenu pour physique s'il appartient au champ d'investigation délimité par les sciences physiques. L'idée dominante est de confier à l'appareillage conceptuel des sciences particulières le soin de définir le contenu du terme « matériel », chacune de ces branches spécialisées étant donc libre de choisir son propre cadre définitionnel. Après tout, qui pourrait fournir un

[18] Aristote, *Métaphysique*, trad. J. Tricot, Vrin, Paris, 1974, livre Δ [1015a].

[19] "By a physical ingredient, we mean the kinds of entities, properties and relations posited in the physical and biological sciences (for short, the physical sciences). We cannot count psychology, and the social sciences in general, as physical sciences for the purpose of explaining the notion of a physical property, for then the contrast between materialism and dualism will be lost. […] Now, of course, the physical sciences are far from complete. Physics especially is far from complete. To that extent, we do not yet know exactly what the physical properties, entities and relations are according to our definition. But the hope is that the problems in current physical science are not going to call for the acknowledgement of properties, entities, and relations different in kind from those now on the scene". (D. Braddon-Mitchell & F. Jackson, 1996:19)

meilleur regard sur la nature même des objets physiques, sinon les sciences mises à l'épreuve dans leur confrontation directe avec le réel ?

Traditionnellement, on s'accorde à dire que le terme de « physique » acquiert définitivement la signification que nous lui connaissons aujourd'hui, c'est-à-dire celle d'*étude scientifique de la nature*, i.e. de la phusis, vers la fin du XVIIIe siècle, lorsqu'il remplace le concept newtonien de « philosophie naturelle ». Comme le relève A. Boutot, la physique atomistique newtonienne engendra un véritable enthousiasme, car elle exprimait « l'espoir que l'homme plaçait en la nouvelle science, quantitative, inductive et expérimentale. Newton avait eu l'immense mérite d'avoir substitué au monde le l'à-peu-près, au monde des qualités et des perceptions sensibles, l'univers du calcul et de la précision. Désormais, la science pouvait progresser d'une manière sûre et reconnue en se maintenant à l'écart des fausses querelles. Cette science conquérante se voulait ou se prétendait universelle. La connaissance intégrale du réel, de tous les événements du monde passés, présents et à venir n'était plus, au moins en principe, un rêve inaccessible. L'omniscience n'avait plus quelque chose d'absurde et semblait à la portée des hommes dans un avenir plus ou moins proche »[20].

Corrélativement, le mot « physique » véhicule deux sens intrinsèquement reliés : sur le versant épistémique, il désigne le mode d'accès objectif qui caractérise les sciences physiques ; tandis que sur le versant ontologique, il fait référence à la constitution ultime des choses. Quant au terme proprement dit de "physicalisme", il fut introduit dans les années 30 du siècle dernier par O. Neurath[21], et relayé par les membres du Cercle de Vienne. Réduite à sa plus simple expression, la thèse fondamentale du physicalisme de « première génération », appelé aussi « *physicalisme sémantique* »[22] peut être résumée ainsi : « toutes les propositions sont également traduisibles dans la langue de la physique : celles des protocoles comme celles du système de la science, dont la construction se rattache aux protocoles et prend la forme d'un système d'hypothèses. La langue de la physique est une langue universelle et, puisque aucune autre langue universelle n'est connue, la langue de la science. (…)Toute autre langue-système possible est donc traduisible dans la langue de la

[20] A. Boutot, (1993), *L'invention des formes,* Paris: Odile Jacob, p. 185.
[21] O. Neurath, (1931b), 'Physikalismus', *Scientia* 50, pp. 297-303. Repris *in* R.S. Cohen & M. Neurath (eds.), *Philosophical Papers 1913–1946*, Dordrecht: D. Reidel Publishing Company, 1983. Dans une œuvre tardive, Neurath fit appel à la métaphore musicale d'un orchestre pour illustrer le rôle dirigeant du physicalisme. Voir O. Neurath (1946), 'The Orchestration of the Sciences by the Encyclopedism of Logical Empiricism'. *Philosophy and Phenomenological Research*, 6: 496-508.
[22] D. Stoljar, (2010), *Physicalism*, London, New York: Routledge.

physique, et peut être interprétée comme une sous-langue de la langue de la physique sous un habillage différent. Du fait que la langue de la physique devient la langue de base de la science, *la science entière se fait physique*»[23].

Si l'on veut résumer rapidement la situation des rapports entre le matérialisme et les sciences de la nature, on peut dire que la physique a réussi l'exploit de fixer le matérialisme en tant qu'arrière plan ontologique stable, épousant la forme d'un canevas théorique à l'aide duquel les différentes branches des sciences particulières n'auront plus qu'à dessiner inexorablement leur spectaculaire progrès. Mais ce serait se livrer à une grossière exagération que de prétendre que l'affaire du physicalisme se résume à cette assimilation. L'ascension du physicalisme reste un processus complexe, dont les racines se mêlent à celles d'autres doctrines et courants de pensée. Par conséquent, avant d'entamer notre périple autour des différentes modèles explicatifs concernant ce thème, quelques précautions préliminaires s'imposent. Notre objectif premier n'étant pas de dresser une monographie exhaustive du physicalisme, nous allons seulement indiquer ici les étapes principales de son évolution. Il ne s'agit là nullement d'une recherche historique au sens propre du terme. En outre, la redoutable complexité et l'étendue du sujet nous oblige à privilégier une démarche plutôt synoptique qu'une approche rigoureusement chronologique.

Par conséquent, la description que nous allons donner des principales étapes ponctuant le problème des rapports entre le physicalisme et les *qualia* sera nécessairement brève et incomplète : nous aborderons certaines des questions qui semblent les plus fondamentales pour l'objectif que nous poursuivons au cours de cette enquête – le duel entre le physicalisme et l'argument de la connaissance de F. Jackson – mais nous ne serons pas exhaustifs. Il ne sera pas non plus question ici d'un certain nombre de problèmes relatifs aux autres thèmes majeurs qui accompagnent habituellement le « mind-body problem », comme le problème de la causalité mentale, l'intentionnalité, l'externalisme et la question du contenu mental, la connaissance de soi, les théories computationnelles de l'esprit, etc. Tous ces thèmes font actuellement l'objet d'un examen attentif et minutieux de la part des philosophes contemporains. A nouveau, il serait trop long, et d'ailleurs utopique, de prétendre pouvoir fournir un

[23] R. Carnap, (1932b), 'Die physikalische Sprache als Unversalsprache der Wissenschaft', *Erkenntnis* 2, pp. 432-465. Trad. fr 'La langue de la physique comme langue universelle de la science', *in* C. Bonnet et P. Wagner, (eds.) (2006), *L'âge d'or de l'empirisme logique.* Paris: Gallimard, p. 356-359. (Soulignements originaux)

compte-rendu exhaustif de cette véritable cascade de questions et de réponses qui ont façonné et continuent de le faire, à l'heure même où nous écrivons ces lignes, la physionomie des débats autour des modèles matérialistes de l'esprit. Aussi, nous voulons éviter dans ce chapitre de faire une revue excessivement historique qui nous empêcherait de dégager et de suivre l'objectif que nous poursuivons ici. Mais il est clair que quel que soit le point de vue que nous adoptions, il nous faut réunir les divers fils conducteurs qui rattachent et situent l'objet de notre investigation sur une indispensable toile d'arrière-plan.

Ce chapitre a pour but d'illustrer la complexité de cette toile de fond, bien qu'on ne puisse évidement pas prétendre à une quelconque exhaustivité. Quelques uns des thèmes mentionnés ici se verront clarifiés et développés plus avant, dans la deuxième partie de notre enquête, dédiée à l'examen détaillé de l'argument de la connaissance de F. Jackson.

Pendant une longue période, on s'accordât pour reconnaître, selon une expression qui revient souvent sous la plume des historiens de la philosophie, la supériorité de la « philosophie première » dans l'explication du réel. Or, l'avènement des méthodes d'investigation scientifiques mises en place suite à la révolution industrielle, et atteignant leur apogée au cours du XXe siècle, bouleversent cette hiérarchisation ancestrale : désormais, c'est vers la science qu'il faut se tourner si l'on veut apprendre ce que les choses *sont* ultimement. Par exemple, l'essence de l'eau, cette « chose » liquide que l'on boit tous les matins, au réveil, c'est d'avoir une structure moléculaire de type H_2O. Rien de plus, rien de moins. Cependant, pour que le paradigme d'intelligibilité des théories physiques puisse étayer et élargir son propre idéal d'objectivité à l'ensemble de la pensée du réel, usurpant ainsi le rôle de l'ancienne métaphysique, il a fallu que ce même paradigme acquît une légitimité méthodologique de premier ordre, autrement dit, s'élevât au rang de *paradigme d'intelligibilité-modèle*. C'est l'analyse de ce cheminement qui retiendra notre attention dans ce chapitre.

Vers la fin du XIXe siècle, plusieurs failles fondamentales semblent saper irrévocablement les fondements de la tradition métaphysique orthodoxe : d'une part, on enregistre une radicale remise en question de l'ontologie moniste de l'idéalisme allemand néo-hégélien, et une totale antipathie envers l'idée des « totalités organiques » qui composent et gouvernent le tout. A.J. Ayer résume admirablement cette attitude hostile envers le soi-disant privilège du philosophe: « La croyance que c'est l'affaire du

philosophe de rechercher les premiers principes, est liée à la conception familière de la philosophie considérée comme l'étude de la réalité comme totalité. (…) Si elle implique, comme cela arrive quelquefois, que le philosophe décolle en quelque sorte hors du monde et le regarde avec l'œil de l'oiseau, alors c'est simplement une conception métaphysique »[24]. Après une longue période oscillant entre Raison Spéculative et Expérience Sensible en tant que modes d'accès privilégiés au Réel, la pensée philosophique s'installe vers le début du XXe siècle, autour du pôle empirique. Le rôle catalyseur que jouera le Cercle de Vienne[25] dans cette stabilisation du centre de gravité des recherches autour des méthodes d'investigations privilégiant le modèle des sciences de la nature, est aujourd'hui indiscutable, et à ce titre, il a fait l'objet de nombreuses études critiques. C'est la raison pour laquelle nous n'insisterons pas ici. Nous partirons tout simplement de l'hypothèse, courante parmi les historiens de la philosophie analytique, qui place le Manifeste du Cercle de Vienne – acte fondateur du positivisme et de l'empirisme logique – à l'origine du changement de paradigme d'intelligibilité privilégiant la recherche de la vérité fondée sur l'observation et la vérification des énoncés.

Ceux parmi les philosophes qui affirment que le rôle de la philosophie est avant tout celui de veiller à la bonne formation et vérification des énoncés, tirent parti de l'idée que l'épistémologie est en mesure de se passer de l'ontologie dans la fondation d'une théorie vraie. Dans un premier temps, et aussi surprenant que cela puisse paraître, le dépassement de la métaphysique par l'analyse *logique* du langage, déclenchant ce qu'on appelle communément le « tournant linguistique »[26], propulse une doctrine de veine *empirique* - le physicalisme - au rang de doctrine *métaphysique* clé embrassant la vénérable question de l'être, de *ce qui est*.

---

[24] A.J. Ayer, (1936), *Language, Truth and Logic*, 2nd ed., London: Gollancz, 1946. Trad. fr. J. Ohana, *Langage, vérité et logique*, Paris, Flammarion, 1956. (Nous faisons ici référence à l'édition française.)

[25] Malgré les opinions parfois divergentes qui divisent ses membres, le Cercle de Vienne institue trois points stables définissant la conception scientifique du monde telle qu'elle est présentée dans le célèbre « Manifeste » : (1) le premier aspect est ce qu'on appelle le « positivisme », c'est-à-dire l'importance accordée au critère de l'expérience au détriment de l'introspection, considérée trop « glissante » et subjective ; (2) le physicalisme est un autre point essentiel de leur position ; (3) enfin, le rôle capital accordé à l'usage de la logique formelle complète le tableau.

[26]« C'est la logique et l'analyse linguistique – *via* la stratégie de l'escalade sémantique – qui ont été les instruments d'investigation privilégiés de la philosophie analytique dans sa réflexion critique sur les sciences ». E. Pacherie, (1993), *Naturaliser l'intentionnalité*, Paris, P.U.F., p. xiv. Concernant la célèbre formule "tournant linguistique", T. Williamson note que selon R. Rorty, qui publie en 1967 une importante anthologie intitulée *The Linguistic Turn – Recent Essays* in *Philosophical Method*, University of Chicago Press, le premier auteur à avoir employé l'expression « tournant linguistique » semble avoir été G. Bergmann dans son livre *Logic and Reality*, Madison: University of Wisconsin Press, 1964. Voir T. Williamson, (2007), *The Philosophy of Philosophy*, Blackwell, 2007, p. 10.

Dans sa formulation canonique, le physicalisme affirme que le *monde naturel* est composé de faits physiques et *rien de plus*. Autrement dit, le monde tel que nous l'expérimentons quotidiennement n'abrite aucun élément « surnaturel ». Selon la définition qu'en donne B. Stroud[27], un élément surnaturel est un agent ou une force qui se situerait d'une façon ou d'une autre à l'extérieur du monde naturel et familier auquel nous sommes habitués et dont les actions ne pourraient pas être comprises comme faisant partie de ce dernier. Par conséquent, le physicalisme moderne se définit avant tout comme un « antisurnaturalisme» [28]rejetant l'existence des entités autres que celles reconnues par les sciences naturelles.

La structure de ce chapitre sera agencée de la manière suivante : la section 1.2. traitera des mécanismes qui ont favorisé l'installation du paradigme conceptuel de la physique dans le rôle de meilleur oracle nous dévoilant ce que le monde *est* ultimement. Il s'agit de mettre en lumière l'importance de deux principes fondamentaux (le principe de l'unité de la science et le principe de la clôture causale du monde physique) sur la base desquels le physicalisme assoit la légitimité de ses ambitions d'offrir une description complète et unifiée de tous les phénomènes. Les sections de 1.3. à 1.5. résument, suivant les derniers travaux de D. Stoljar[29], les différentes versions définitionnelles du physicalisme structurées à la fois autour du slogan « tout est physique », ainsi qu'autour de la définition du terme de propriété/objet physique.

1.2 Unité de la science et stratification du monde

Au siècle dernier, la matière, placée sous la loupe grossissante des laboratoires de recherche n'apparaît plus comme le postulait l'ancien atomisme, i.e. solide, inerte, compacte et impénétrable. Ces constats obligèrent les matérialistes à se retrancher derrière la position, beaucoup plus modeste, selon laquelle le monde est composé uniquement d'*énoncés vrais* garantis par les paradigmes de recherche des sciences physiques. Mais cette manœuvre soulève à son tour deux questions supplémentaires : (1) Qu'est ce que les sciences physiques ? Autrement dit, quelles sont les disciplines qui ont le droit de faire partie du club privé des sciences physiques ? Et (2) quel est le

[27] B. Stroud, (1996), 'The Charm of Naturalism' in *Proceedings of the American Philosophical Association* 70, pp. 43-55. Réédité in M. de Caro & D. Macarthur (eds.), (2004), *Naturalism in question*, op. cit., pp. 21-35.
[28] J. Dupré, (2004), 'The Miracle of Monism', in M. de Caro & D. Macarthur (eds.) op. cit., p. 36.
[29] D. Stoljar, (2010), *Physicalism*, op. cit.

facteur qui leur confère l'autorité de légiférer sur l'ontologie fondamentale, i.e. sur ce que les choses *sont* ultimement ?

Commençons avec la première question. Les « sciences physiques » incluent tout d'abord la physique elle-même, dans sa fonction incontestable de paradigme de recherche. Viennent ensuite la chimie, la biologie moléculaire et la neurophysiologie. Pour une large majorité de philosophes contemporains, la physique remplit toutes les conditions requises pour arborer l'ambition de fixer *l'unité de la science*. Le point central consiste alors à trouver le ciment permettant aux diverses sciences particulières d'être intégrées dans une seule et grande vision unificatrice. Cela entraîne de nombreuses interrogations : de quelle nature est la relation qui permet aux différents champs théoriques de s'emboîter sous l'égide d'une architecture explicative unique ? S'agit-il d'une relation de *réduction*, de *réalisabilité*, de *survenance*, ou de *nécessitation logique* ? Et surtout, comment faut-il architecturer ces relations pour obtenir la vision unificatrice de la science ?

Un point de départ classique des discussions est l'interprétation de H. Putnam et P. Oppenheim[30]. Selon ces auteurs, l'expression « unité de la science » comprend au moins trois niveaux de réduction : (i) d'abord, il y a le niveau du vocabulaire : les termes d'une discipline peuvent être réduits aux termes d'une autre discipline : c'est la thèse de *l'unité du langage*. (ii) Vient ensuite le niveau des lois : les lois d'une discipline peuvent être réduites aux lois d'une autre : c'est la thèse de *l'unité des lois*. Enfin, (iii) l'implication la plus forte concerne *l'unité de la science*, dans le sens où les lois d'une discipline sont non seulement réductibles aux lois d'une autre, mais les lois de cette discipline sont d'une certaine manière tissées et unifiées sous le sceptre de la discipline unificatrice. Selon l'acception de Putnam et Oppenheim[31], la thèse de l'unité de la science comporte un véritable essaim de thèses adjacentes: en premier lieu figure la thèse des niveaux de réalité. Le nombre de ces niveaux doit être fini. En deuxième lieu, il doit y avoir un seul et unique niveau fondamental (un micro-niveau auquel se réduisent, du moins en principe, tous les autres niveaux). Troisièmement, toute entité, objet et phénomène doit être décomposable en éléments appartenant au niveau

[30] P. Oppenheim & H. Putnam, (1958), 'Unity of Science as a Working Hypothesis' *in* H. Feigl, M. Scrive, et G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, Minneapolis : University of Minessota Press. Réédité in R. Boyd, Ph. Gasper & J.D. Trout (eds.) (1991 /1999 : 405), *The Philosophy of Science*, Cambridge Mass. : MIT Press. (Nous utilisons ici cette dernière édition). (Soulignements en original)

[31] (P. Oppenheim & H. Putnam, 1958/1999 : 409)

immédiatement inférieur, à l'exception, bien entendu, du niveau fondamental. Enfin, aucune entité, objet, phénomène situé à un niveau déterminé ne doit posséder un élément situé à un niveau supérieur. Les niveaux doivent être choisis et établis en respectant un ordre naturel et justifié du point de vue des sciences empiriques mises à jour.

Ce tableau grandiose repose, comme l'a récemment rappelé D. Papineau[32], non seulement sur l'adoption implicite du slogan « tout est physique », mais aussi sur l'adhésion à l'idée que « tout est *micro*scopiquement déterminé ». Chacune des sciences particulières divise le réel en fonction de ses propres intérêts d'investigation et impose ses propres taxinomies. On obtient ainsi une stratification[33] du réel qui peut être figurée ainsi :

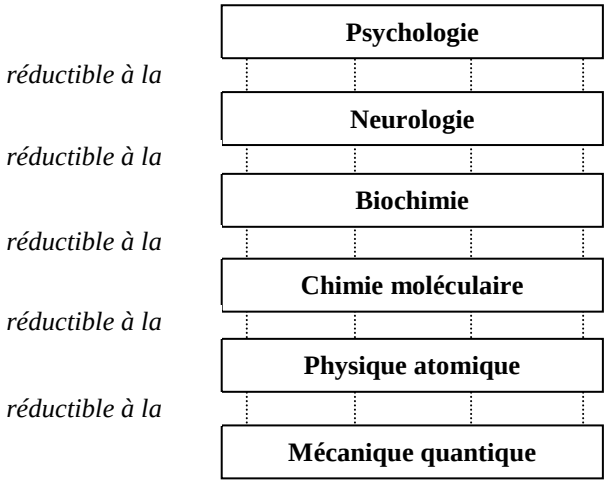


Figure 2: L'unité de la science et la stratification du monde[34]

Au niveau fondamental il y a les électrons, les quarks, et autres entités microscopiques décrites par la mécanique quantique. Ces entités possèdent un certain type d'attributs, comme le spin, la masse, l'énergie, la charge, etc. Le niveau fondamental est donc le niveau de la physique. A l'étage immédiatement supérieur on

[32] D. Papineau, (2008), 'Must a Physicalist be a Microphysicalist?' *in* J. Hohwy & J. Kallestrup (eds.), *Being Reduced: New Essays on Reduction, Explanation, and Causation*. Oxford University Press, p. 126. (Soulignement en original).

[33] La stratification originale de Putnam & Oppenheim (1958/1999 : 409) était la suivante :
6…….. Groupes sociaux
5…….. Organismes vivants (multicellulaires)
4…….. Cellules
3…….. Molécules
2…….. Atomes
1…….. Particules élémentaires

[34] P. Carruthers, (2004), *The Nature of the Mind.* New York: Routledge, p. 164.

retrouve les entités étudiées par la chimie, dont l'exemple paradigmatique est l'eau et ses propriétés de transparence, liquidité, etc. A partir de là, on monte encore un étage et l'on retrouve la biologie et l'étude du vivant, des organismes considérés comme un tout, et ayant des attributs et des propriétés spécifiques comme la digestion, la reproduction, etc. C'est à partir de ce palier que les choses se compliquent. Car la science du vivant touche le sujet sensible de la vie mentale, de la psychologie et des sciences cognitives qui étudient à leur tour des attributs typiques comme les attitudes propositionnelles, l'intentionnalité, les *qualia*, les réseaux neurophysiologiques, etc. La cohérence de cet ensemble est assurée par l'idéal régulateur d'unité qui gouverne le projet omni-englobant des sciences de la nature.

1.2.1 Réduction et niveaux du réel

Au cœur du projet d'unification physicaliste, il y a d'abord la notion de « réduction ». Comme l'a relevé S. Sturgeon[35], c'est l'adhésion à quatre thèses distinctes qui a facilité la cristallisation du procédé de « réduction » au sein du matérialisme contemporain. (i) En premier lieu, il convient de placer l'observation de sens commun selon laquelle les phénomènes/objets, qu'ils soient grands (les montagnes) ou petits (les grains de sable), partagent certaines propriétés, (la forme, disons) qui ne diffèrent qu'en grade, et non pas en nature, d'une échelle à l'autre. Bref, « grand » et « petit » sont des termes conceptuellement homogènes. (ii) Deuxièmement, les propriétés des objets complexes sont composées à partir des propriétés de leurs constituants. Par exemple, les objets familiers « de tous les jours » ont des couleurs spécifiques, mais ces dernières ne sont que la somme des couleurs des parties plus petites qui les constituent. Bref, les propriétés des objets « quotidiens » sont des propriétés réductives de type « bottom-up »[36]. (iii) Troisièmement, dans la vie quotidienne, les changements à niveau macroscopique sont expliqués par des changements au niveau microscopique, de type « bottom-up » également. (iv) Enfin, quatrièmement, si l'on part du principe selon lequel une propriété est *toujours* réalisée ou *jamais* réalisée (la réalisation étant une question de tout ou rien) il s'ensuit que les

[35] S. Sturgeon, (2001), 'The Roots of Reductionism', *in* C. Gillett & B. M. Loewer (eds.), *Physicalism and its Discontents*. Cambridge University Press, pp. 197-8.
[36] La traduction littérale de l'expression « bottom-up » est « d'en bas en (vers le) haut » Nous avons fait le choix de garder ici, pour plus de facilité et d'élégance, le terme original anglais.

propriétés que l'on trouve dans l'univers quotidien, n'apparaissent jamais comme non-réalisées. Bref, ces propriétés affichent une réalisation modalement invariante.

Ces quatre conditions, note Sturgeon[37], étayent ce qu'on appelle l'Argument de la Surdétermination en faveur de la réduction du mental au niveau de la physique fondamentale, (représentée ici par la mécanique quantique), argument qui peut être synthétisé ainsi :

SURDÉTERMINATION

(1) Les états/événements mentaux ont des effets macrospatiaux. (ex : je bouge le bras si je le veux)
(2) De tels effets sont constitués d'effets spatiaux quantiques. (ex : le comportement « tunnel » des particules)
(3) Chaque effet spatial quantique est caractérisé par une histoire entièrement et exclusivement quantique.
(4) Pour chaque événement spatial C, E, et E* : si C cause E, et E constitue E*, alors C cause E*.
(5) Les effets spatiaux des événements mentaux ne sont pas généralement surdéterminés.

Par conséquent,

(6) Les événements mentaux sont des événements quantiques.

Comme le remarque fort bien Sturgeon[38], la prémisse (4) est supposée être « évidente ». Mais il est clair que quelle qu'en soit l'analyse, (4) doit son hypothétique évidence à l'endossement tacite des trois premières thèses (i-iii) mentionnées plus haut, à savoir :

(i) Les propriétés spatiales 'grandes' et 'petites' sont conceptuellement homogènes.
(ii) Les propriétés spatiales 'grandes' sont des propriétés réductives « bottom-up »
(iii) Les changements spatiaux macroscopiques sont des changements « bottom-up ».

Sans vouloir insister sur un sujet qui a été déjà abondamment discuté dans la littérature analytique, rappelons très rapidement que dans un premier temps, la plupart des discussions autour du concept de réduction prennent comme point de départ la célèbre analyse d'E. Nagel[39]. Une théorie *T'*est réductible à une théorie *T* si et

[37] (S. Sturgeon, 2001 : 202)
[38] *Ibid.*
[39] E. Nagel, (1961), *The Structure of Science*, London: Routledge and Kegan Paul, pp. 336-66.

seulement si les lois qui régissent *T'* sont dérivables des lois de *T*, ces dernières jouant ainsi un rôle paradigmatique dans l'explication des phénomènes. E. Nagel distingue deux types de réduction : dans le plus simple des cas, le vocabulaire de la théorie réduite T′ ne contient aucun terme descriptif qui ne soit déjà une sous-classe des termes contenus dans T. Pour cette même raison, ce type de réduction est connu sous le nom de *réduction homogène*[40]. Mais les cas les plus intéressants et problématiques apparaissent lorsque T′ manipule des concepts *en plus*, différents de ceux contenus dans T, autrement dit lorsque la réduction doit gérer des niveaux *hétérogènes*. L'exemple classique que Nagel présente concerne la réduction de la thermodynamique à la mécanique statistique en s'appuyant sur l'analyse du concept de *température*.

> T2 est réductible à $T1_{def.}$ = (1) [la condition des lois-pont] pour chaque prédicat primitif *F* de T2, il y a un prédicat *G* en T1 en sorte qu'une « loi-pont » ayant la forme biconditionnelle ''F*x* ↔ G*x*'' intervient, et (2) [la condition de la dérivabilité] chaque loi de T2 est dérivable logiquement à partir des lois de T1, avec la condition des loi-pont considérée comme prémisse auxiliaire[41].

La notion Nagelienne de réduction pose un certain nombre de difficultés. Premièrement, il s'agit de situer avec plus de précision le statut et le rôle des « lois-pont »[42]sur lesquelles repose la délicate mission d'assurer la cohérence d'ensemble des stratégies réductives. Plus particulièrement, les interrogations concernent la nature de ces relations : sont-elles purement linguistiques ou impliquent-elles des découvertes empiriques, factuelles ? S'agit-il d'une relation d'identité nécessaire, analytique, ou d'une opération synthétique ? Une discussion détaillée[43] de ces points dépasserait les

---

[40] *Ibid.*, p. 339.

[41] Pour palier les éventuelles difficultés de passage d'un vocabulaire à l'autre, E. Nagel introduit deux précisions supplémentaire sous la forme de deux conditions formelles : « (1) *Connectability*. "Assumptions of some kind must be introduced which postulate suitable relations between whatever is signified by 'A' [the term to be reduced, that is, an element of the vocabulary of theory *T'*] and traits represented by theoretical terms already present in the primary [reducing] science. (…) (2) *Derivability*. "With the help of these additional assumptions, all the laws of the secondary science, including those containing the term 'A,' must be logically derivable from the theoretical premises and their associated coordinating definitions in the primary discipline." (E. Nagel, 1961: 353-354)

[42] Selon R. Batterman, pour une grande majorité de théoriciens, ces lois-pont expriment un certain type d'identité. Il cite à ce propos L. Sklar, (1967), "Types of inter-theoretic reduction", *The British Journal for the Philosophy of Science*, 18, p. 120: "the place of correlatory laws [bridge laws] is taken by empirically established *identifications* of two classes of entities. Light waves are not correlated with electromagnetic waves, for they *are* electromagnetic waves." R. Battermann, *Intertheory relations in Physics*, http://plato.stanford.edu

[43] Plus récemment, S. Sarkar a proposé de différencier quatre types de réduction, deux épistémologiques, et deux ontologiques : (a) fondamentalisme ; (b) approximation ; (c) hiérarchisation abstraite ; (d) hiérarchisation spatiale. Selon Sarkar, une réduction forte satisfait les trois derniers points, tandis qu'une

modestes cadres de notre enquête ici. Rappelons néanmoins que la réduction Nagelienne (réduire des théories en montrant comment les lois de la théorie-cible peuvent être logiquement dérivées des lois appartenant à la théorie réductive) a été rapidement désavouée[44], pour la simple bonne raison que les miraculeuses lois-pont étaient introuvables. On a préféré alors, comme nous allons le voir par la suite, remplacer ce modèle par un modèle fonctionnaliste[45].

Selon ce dernier modèle, défendu entre autres par D. Chalmers[46] et J. Kim[47], pour réduire une propriété-cible, il convient d'en fournir d'abord une retranscription fonctionnelle *a priori* en termes de rôles causaux caractéristiques. Dans un deuxième temps, on circonscrit de façon *a posteriori* l'élément qui réalise la propriété donnée dans une base réductive dont elle joue, d'une façon plus ou moins approximative, le rôle respectif. Et c'est seulement après l'accomplissement de ces manœuvres qu'il nous est loisible d'identifier la propriété en cause avec l'élément qui la réalise. Cette stratégie présente l'avantage inestimable de rendre intelligible l'explication réductive au moins du point de vue du modèle nomologico-déductif de l'explication. Pour qu'une explication soit authentiquement réductive, il faut que la prémisse explicative d'un phénomène impliquant la propriété *F* ne fasse référence à aucune propriété de type-*F*. Il est vrai que dans le modèle fonctionaliste le premier pas dans la réduction fait référence à ce type de propriétés, mais il s'agit là d'une définition, et les définitions ne sont pas considérées comme des prémisses supplémentaires s'ajoutant aux déductions explicatives. Par contre, le modèle Nagelien classique des lois-pont ne respecte pas cette contrainte essentielle relative à l'explication de type réductif[48].

Concernant la notion de « réduction », il importe de souligner que certains auteurs, comme J. Bickle[49] ont récemment dénoncé l'utilisation abusive d'une image

---

réduction faible satisfait uniquement (a). Voir S. Sarkar, (1998), *Genetics and Reductionism*, New York: Cambridge University Press.

[44] Rappelons que la réduction Nagelienne fut vivement critiquée, entre autres, par P.K. Feyerabend. Voir P.K. Feyerabend, (1962), 'Explanation, reduction, and empiricism', in H. Feigl & *G*. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, volume 3, pages 28–97. D. Reidel Publishing Company.

[45] Voir D. Lewis, (1970c), 'How to Define Theoretical Terms', *Journal of Philosophy*, 67: 427–446. Voir aussi D. Lewis, (1999a), *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press.

[46] (D. Chalmers, 1996)

[47] J. Kim, (2005), *Physicalism, Or Something Near Enough*, Princeton: Princeton University Press.

[48] J. Kim, (2008), 'Reduction and Reductive Explanation: Is One Possible Without the Other?' in J. Hohwy & J. Kallestrup (eds.), op. cit., pp. 93-114.

[49] "The scientific examples described to illustrate this account are telling. They are not even examples from the history of real science (like the ones that intertheoretic reductionists at least appeal to). Rather, they are from *elementary school science education*—like the boiling of water near sea level due to the

trop simplificatrice, voire caricaturale, du procédé de réduction tel qu'il est réellement utilisé dans la pratique scientifique, ce qui simplifierait illicitement la tâche à ses détracteurs. D'autres auteurs[50]s'inscrivent ouvertement en faux contre une interprétation trop simpliste de ce procédé. Malgré ces mises en garde, toujours est-il que le matérialisme contemporain préfère opérer avec des notions plus souples que celle de « réduction Nagelienne ».

On a tôt fait de remarquer, à la suite de T. Crane et D.H. Mellor[51], que même si certaines disciplines se prêtent à une réduction au niveau de la physique fondamentale, il se peut que cette réduction ne soit pas effectivement et universellement applicable à *toutes* les disciplines. On peut, certes, objecter à cette remarque, qu'il suffit de la considérer applicable *en principe*. Appelons, pour plus de commodité, *la thèse de la réductibilité en principe* avec l'abréviation REP. Maintenant que signifie, plus précisément, le terme « être réductible en principe » ? Selon Crane et Mellor, on ne peut pas défendre la REP sans faire appel à un raisonnement circulaire : les sciences particulières ne peuvent pas être réductibles en principe *parce qu*'elles sont déclarées physiques uniquement si la REP fixe, *ab initio*, les limites du domaine physique. Par conséquent, nous avons besoin d'un *autre* principe statuant la manière dont la REP s'applique aux sciences réductibles à la physique, autrement dit, décider quelle est la physique à laquelle la REP est applicable : la physique actuelle, ou une hypothétique physique future ? Nous reviendrons sur cette difficulté plus loin dans ce chapitre.

Une autre source de l'attrait qu'exerce la REP sur le matérialisme contemporain est contenue dans l'idée qu'à la base, le réel est composé de particules élémentaires un peu à la manière dont les briques forment un mur[52]. Appelons cette thèse « microréduction » (MR). La MR implique et soutient l'édifice de ce qu'on pourrait

---

dynamics of H_2O molecules". J. Bickle, (2008), 'Real Reduction in Real Neuroscience: Metascience, Not Philosophy of Science (and Certainly Not Metaphysics!)', in J. Hohwy & J. Kallestrup (eds.), op. cit., p. 35.

[50]Voir par exemple F. Dizadji-Bahmani, R. Frigg, S. Hartmann, (2010), 'Who's Afraid of Nagelian Reduction?', *Erkenntnis* 73 (3):393-412.

[51] T. Crane & D.H. Mellor, (1990), 'There is no question of physicalism', *Mind* 99 (394):185-206. Réédité *in* P. Moser & J. Trout, (eds.) 1995, *Contemporary Materialism*, London: Routledge. (Nous faisons ici référence à cette dernière édition)

[52] Pour une analyse critique très détaillée de cette thèse voir M. Bitbol, 2001, *De l'intérieur du monde : Pour une philosophie* et *une science des relations*, Paris, Flammarion. Voir aussi J. Ladyman et D. Ross (avec D. Spurrett et J. Collier) 2007: *Every Thing Must Go: Metaphysics Naturalised*, Oxford: Oxford University Press.

appeler le *microphysicalisme* (MP). Selon D. Papineau[53], le MP est la thèse selon laquelle :

(MP) Tous les faits surviennent métaphysiquement sur des faits microphysiques.

Ainsi défini, le MP est une conjonction de deux thèses distinctes : le 'Physicalisme' (P) et le 'Microscopisme'(M):

(P) Tous les faits surviennent métaphysiquement sur des faits physiques.

(M) Tous les faits physiques surviennent métaphysiquement sur des faits microphysiques.

Plusieurs arguments viennent soutenir et justifier le pouvoir explicatif de la microréduction. Les partisans de l'*argument inductif*[54] aiment rappeler que jusqu'à présent, tous les faits susceptibles d'être examinés scientifiquement se sont avérés réductibles, ou surviennent sur des faits microphysiques. L'exemple le plus célèbre reste, bien évidemment, l'image des deux tables d'Eddington[55]. L'étude de ces particules élémentaires constitue le noyau dur incontestable de la physique. Mais cela revient à dire que les microparticules sont les seules entités qui existent réellement. Un paradoxe surgit alors : doit-on supposer que l'existence des particules élémentaires nous interdit d'affirmer que les atomes, les molécules, les arbres, les chats, et les agents conscients *existent* aussi? Les défenseurs de la MR pourraient, bien entendu, affiner leur définition du concept de « non-existence » en sorte que l'on puisse distinguer plus facilement entre la « non-existence » des chats, et la « non-existence » des licornes, disons. Car la véritable finalité de la microréduction est de montrer que tous les faits physiques concernant les chats, par exemple, peuvent être réduits à des faits relatifs à ses constituants subatomiques. En outre, si l'on adhère à l'exigence selon laquelle il est préférable de ne pas multiplier la quantification des entités au-delà du nécessaire, il s'ensuit que les seules entités quantifiables seraient les microparticules. Mais il est loin d'être évident que la MR étaye un principe de type-REP suffisamment puissant pour qu'elle puisse couvrir la *totalité* des phénomènes naturels. Au mieux, la thèse de la

[53] D. Papineau, (2008), 'Must a Physicalist Be a Microphysicalist?' in J. Hohwy & J. Kallestrup (eds.), op. cit., pp. 127-8.
[54] (D. Papineau, 2008 :131)
[55] Du point de vue de l'homme de la rue, une table se définit selon sa forme, couleur, masse, etc., tandis que pour un scientifique, une table n'est qu'un amas de myriades de microparticules tenues ensemble par des champs d'énergie. Voir A.S. Eddington, (1928), *The Nature of the Physical World*. London: J.M. Dent et Sons.

microréduction doit être considérée comme une hypothèse de travail acceptable, même si elle n'est pas toujours vérifiable, surtout au niveau de la microphysique[56].

Sans vouloir anticiper sur un sujet qui sera détaillé plus loin au cours de notre investigation, signalons seulement que la conclusion qu'en tirent Crane et Mellor[57] stipule que la MR, même dans sa formulation modeste, c'est-à-dire en tant que simple principe méthodologique, n'est pas universellement applicable, même en physique, et que par conséquent, sa portée ne peut pas être métaphysique, légiférant sur ce que les choses *doivent être*[58].

Dans la section suivante, nous examinerons quel est le facteur qui confère à la physique l'autorité de légiférer sur l'ontologie fondamentale de ce que les choses *sont* ultimement. Autrement dit, il s'agit de répondre à la question : qu'est-ce que la physique possède *en plus* face à ses possibles rivales « naturelles », la biologie ou la chimie ?

1.2.2 La complétude de la physique et l'argument causal en faveur du physicalisme

Comme nous l'avons déjà évoqué dans les sections précédentes, deux principes soutiennent les ambitions physicalistes à l'universalité: (1) le premier est la thèse méthodologique de l'*unité de la science*[59] : si le monde est un tout ontologiquement homogène, il devient alors possible d'unifier les sciences particulières sous le sceptre d'une théorie physicaliste *complète* du monde naturel. (2) Le deuxième principe est la

[56] Prenons l'exemple, bien que controversé, de la mécanique quantique : la physique quantique orthodoxe n'est pas micro-réductive, tandis que certains domaines de la physique sont manifestement macro-réductifs. Nous pensons ici au principe de Mach selon lequel la masse inertielle même des microparticules dépend de la manière dont la matière est distribuée dans tout l'univers.

[57] (T. Crane & D.H. Mellor, 1990)

[58] Par ailleurs, l'idée même de microparticules est sujette à des révisions : à tout moment on pourrait découvrir que les particules élémentaires se divisent à leur tour en sous-particules encore plus « élémentaires ». Dès lors, il devient difficile de statuer sur la nature ultime des « briques » qui composent l'univers. Le principe de réductibilité à la physique, ou mieux dit à la microphysique, ne semble pas pouvoir assurer, à lui tout seul, les prétentions ontologiques de cette science. Au contraire, il semblerait que la réductibilité échoue parfois au sein même de la physique. Pourquoi s'étonner alors si cette réduction échoue dans le cas du mental ? Bref, si le physicalisme repose sur l'aspiration de la physique à l'universalité, alors il repose tout bonnement sur un mythe, concluent Crane et Mellor.

[59] Ce principe, bien qu'extrêmement populaire parmi les théoriciens actuels du physicalisme, prête le flanc à de sérieuses objections. Voir, pour une excellente analyse critique, l'ouvrage de J. Dupré, (1993), *The Disorder of Things. Metaphysical Foundations of the Disunity of Science*, Cambridge, MA: Harvard University Press.

thèse ontologique de *la complétude et de la clôture causale, nomologique et explicative de la physique*: pour *tout* état/événement/phénomène physique décrit par une théorie physique fondamentale, dans la mesure où cet état a des causes, il a des causes physique ; dans la mesure où cet état est soumis à des lois, il est subsumable à des lois physiques ; dans la mesure où cet état admet des explications, ce sont des explications physiques qui lui sont applicables. Ces principes se soutiennent mutuellement, leur validité étant corroborée par l'impressionnant succès prédictif des théories scientifiques.

A première vue, note D. Papineau[60], l'élément clé qui a fait pencher la balance en faveur du physicalisme doit être cherché du côté de la *thèse de complétude* de la physique. Brièvement décrit, le principe de complétude affirme que les causes physiques suffisent pleinement pour expliquer les effets physiques obéissant à des lois physiques universelles. Observons d'abord que ce principe ne suffit pas, à lui tout seul, pour imposer le physicalisme. En effet, la thèse de la complétude est muette quant aux entités non physiques, telles les croyances, l'intentionnalité, ou l'expérience phénoménale. Elle se contente de légiférer sur l'agencement et le fonctionnement des entités physiques, rien de plus. Pour que le physicalisme puisse s'affirmer pleinement en tant que conception métaphysique omni-englobante, il doit justifier une thèse beaucoup plus forte, à savoir l'idée que « *tout* est physique ». Mais pour prouver cette identité, il faut trouver une base justificative solide.

Selon Papineau[61], les arguments avancés par des philosophes matérialistes d'avant-garde comme J.J. Smart, D. Armstrong ou D. Lewis, (et sur lesquels on aura l'occasion d'insister plus avant au cours de notre enquête) respectent, malgré leurs discordances, une structure canonique stable, qui peut être résumée sous la forme d'un *argument causal* à trois volets :

[60] "All physical effects are fully determined by law by prior physical occurrences". (D. Papineau, 2001 : 8)
[61] *Ibid.*, p.9.

ARGUMENT CAUSAL

(P1) *La complétude de la physique* : Tous les effets physiques sont entièrement déterminés, *via* des lois physiques, par des occurrences et des événements physiques antérieurs.

(P2) *L'action causale* : Tous les événements mentaux peuvent causer des effets physiques.

(P3) *Surdétermination*: Les effets physiques provoqués par des causes mentales ne sont pas tous surdéterminés.

Par conséquent,

(C): Les événements mentaux doivent être identiques aux événements physiques.

L'argument présenté sous cette forme, bien que valide, prête le flanc à plusieurs objections. D'une part, il ne protège pas le physicalisme d'éventuels dégâts infligés par l'épiphénoménalisme[62] : par exemple, si l'on est prêt à adhérer à une conception de type « harmonie préétablie », ou à accepter l'idée selon laquelle les états mentaux n'ont pas de pouvoir causal sur les états physiques, alors l'argument causal présenté ci-dessus n'avance aucune preuve irréfutable contre l'épiphénoménalisme. Nous verrons que le fonctionnalisme est en principe compatible avec l'épiphénoménalisme, ce qui peut provoquer un certain embarras dans les rangs des matérialistes. Nous y reviendrons plus amplement plus loin au cours de notre enquête. Toujours est-il que la menace épiphénoménaliste met en lumière un autre point faible de l'argument causal : la prémisse (P2) nous renseigne uniquement sur les événements non-physiques qui *ont* un pouvoir causal sur les événements physiques, mais elle reste muette quant aux états mentaux qui *n*'ont *pas* un pouvoir causal sur le monde physique. Par conséquent, l'argument n'a pas d'emprise sur les niveaux putatifs de réalité situés en dehors du niveau causal stipulé dans la prémisse (P2), et qui n'ont, *a fortiori*, aucune influence causale : nous pensons ici aux mondes mathématiques, aux valeurs morales, etc. Il s'agit là d'une limitation sérieuse de l'argument, limitation qui, selon Papineau, nous incite à réviser le slogan initial du physicalisme et à adopter plutôt une version

[62] Un *épiphénomène* désigne ce qui se surajoute à un phénomène sans exercer sur lui aucune influence causale effective.

explicitement causale, remplaçant ainsi (Φ) « tout est physique » par (Φ_C) « tout ce qui interagit causalement avec le monde physique est physique»[63].

De toute évidence, aucun des arguments en faveur du matérialisme avancés par les matérialistes australiens comme Smart, Lewis ou Armstrong, n'aurait pu imposer le physicalisme en tant que doctrine métaphysique omni-englobante sans l'appui crucial et souterrain de la thèse de la complétude de la physique. Mais cette thèse non plus, et ceci est un point important sur lequel Papineau insiste particulièrement, n'aurait pas pu, à elle toute seule, imposer le physicalisme dans son rôle de 'Weltanschauung' moderne. Il aura fallu la présence d'un élément supplémentaire, catalyseur, dont l'apport décisif a fini par faire pencher irréversiblement la balance en faveur du physicalisme. Prenons un exemple, pour plus de clarté : bien que la thèse de la conservation de l'énergie soit intimement liée à la complétude de la physique, elle ne suffit pas à réfuter l'existence des causes vitales, des énergies spirituelles, ou autres forces « spéciales » transcendant le niveau physique. Le principe de la conservation de l'énergie est non-informatif quant à la *nature* des forces qui opèrent dans l'univers. Est-ce la gravité la seule force fondamentale de l'univers ? Que dire alors des forces électromagnétiques, nucléaires, etc. ? Il est clair que la thèse de la conservation de l'énergie laisse la porte entrouverte à d'autres types de forces (spirituelles, par exemple), et n'informe pas sur la nature des forces qui opèrent dans le réel. Tout ce qu'elle affirme, c'est que ces forces, quoiqu'elles soient, respectent le principe de conservation. D'ailleurs, à l'époque où le jeune Helmholtz publie son célèbre traité *Über die Erhaltung der Kraft*[64], il était parfaitement plausible, et en aucun cas répréhensible, d'expliquer le phénomène de conservation de l'énergie relatif aux organismes vivants en faisant appel à des « forces vitales ». Dans ce contexte, le défi à relever est à présent identifié : il s'agit d'examiner ce qui a radicalement changé pour que l'hypothèse vitaliste devienne, à partir des années 50, complètement saugrenue.

Selon Papineau, la faillite de l'hypothèse mentaliste des forces « vitales » a été précipitée par deux arguments : un argument abstrait, inspiré par la physique théorique, et un concret, inductif, basé sur des observations empiriques.

---

[63] (D. Papineau, 2001: 11)

[64] H. von Helmholtz, (1847), *Über die Erhaltung der Kraft: Eine physikalische Abhandlung*, Vorgetragen in der Sitzung der physikalischen Gesellschaft, 1847. Berlin: G. Reimer. Trad. de l'allemand par L. Pérard, *Mémoire sur la conservation de la force*, précédé d'un *Exposé élémentaire de la transformation des forces naturelles*, V. Masson et fils, Paris, 1869.

(1) Le premier argument, ***l'argument des forces fondamentales***, s'appuie sur la prémisse selon laquelle toutes les forces spécifiques étudiées par les diverses branches des sciences naturelles, peuvent *se réduire* à un ensemble déterminé et restreint de *forces fondamentales* dont la fonction serait celle de conserver l'énergie. Dans cette acception, lorsque nous attribuons l'apparition d'un certain effet à l'action d'une « force musculaire » ou à une « cause mentale », nous devrions être capables, en principe, de décomposer ces forces et causes mentales, afin d'en déceler leur structure fondamentale qui, elle, est entièrement physique.

Selon le célèbre modèle nomologico-déductif (ND)[65] dont Hempel[66] fut un des plus influents artisans, une explication scientifique standard est composée de deux éléments : un *explanandum*, c'est-à-dire une assertion décrivant le phénomène qui est à expliquer, et un *explanans*, i.e. la classe des propositions appelées à expliquer le phénomène ciblé. Pour que l'*explanans* puisse rendre compte de manière adéquate de l'*explanandum*, plusieurs conditions sont à respecter.

(i) D'abord, le processus d'explication doit prendre la forme d'un argument déductif, au cours duquel l'*explanandum* devient une conclusion inférée à partir des prémisses contenues dans l'*explanans*. Nous avons ici le composant « déductif » du modèle.

(ii) Deuxièmement, le versant nomologique affirme que l'*explanans* doit contenir au moins une « loi de la nature » dont l'intervention doit s'avérer primordiale dans la dérivation de la conclusion, et dont l'absence entraînerait *ipso facto* l'invalidité de l'argument.

Dans sa formulation la plus générale, le modèle-ND est censé s'appliquer tant à l'explication des « lois » ou des « régularités généralisées », qu'à la description des occurrences instanciées à des moments et des endroits déterminés. Nous voyons donc, pour résumer, que selon le modèle-ND, les lois jouent un rôle essentiel dans l'explication de la causalité. La principale tâche des théories scientifiques est de générer et de vérifier des prédictions. Selon l'orthodoxie matérialiste, les causes ont une nature déterministe. Un système est considéré comme déterministe si chacun de ses états, à chaque instant *t*, est une conséquence des états précédents soumis aux lois de la nature. La notion déterministe de « loi » n'accepte aucune exception.

---

[65] Désormais « modèle-ND » en abrégé.
[66] C. Hempel, (1942), 'The Function of General Laws History', *Journal of Philosophy* 39 (2):35-48.

(2) Le deuxième argument justifiant l'ascension du physicalisme, que Papineau[67] désigne sous le nom de ***l'argument de la physiologie***, concerne le versant inductif du problème. Cet argument affirme, tout simplement, qu'il n'existe pas de preuves directes, basées sur des observations empiriques en faveur des « forces vitales » ou des « causes mentales ». Et c'est précisément à ce niveau que se situe, insiste Papineau, l'élément catalyseur permettant le solide ancrage du physicalisme en tant qu'arrière-plan ontologique légitime. Jamais les savants des siècles précédant le XXe siècle n'ont disposé d'autant de moyens et de sources empiriques d'expérimentation que leurs collègues travaillant après la deuxième guerre mondiale. Ces circonstances historiques exceptionnelles ont permis aux scientifiques d'étendre leurs investigations et de pousser de plus en plus loin les frontières de l'observable. Or, au fur et à mesure que la science avançait sur le terrain des découvertes empiriques, on n'a pu que constater l'effondrement de l'hypothèse des forces « vitales » pour la simple et bonne raison que rien de surnaturel n'était observé au niveau des structures et du fonctionnement cellulaire.

L'ascension du physicalisme, conclut Papineau[68], est le reflet direct du développement des stratégies d'expérimentation. Il est vrai que sans la thèse clé de la complétude de la physique, le matérialisme, et avec lui la thèse de l'identité psychophysique (selon laquelle les états mentaux *sont* des états neurophysiologiques du cerveau), n'aurait pas pu se frayer un chemin dans la philosophie de l'esprit actuelle. Mais le facteur qui a contribué de façon décisive à l'établissement du physicalisme en tant que doctrine métaphysique prédominante à partir de la deuxième moitié du XXe siècle, a été, sans aucun doute, la prolifération des méthodes d'investigation expérimentale qui ont permis aux savants de scruter, avec de plus en plus de précision, les entrailles du réel. Ensemble, la conjugaison de ces deux conditions a permis au monisme matérialiste de s'installer définitivement sur la scène philosophique en tant qu'arrière-plan ontologique unificateur.

Comme nous allons le voir dans les chapitres suivants, même si les théories de l'identité stricte entre le mental et le physique vont laisser peu à peu la place à des descriptions physicalistes plus souples, jamais la recherche de la vérité du côté de la physique ne fut abandonnée. Nous verrons ainsi que la « désanctuarisation » de

[67] (D. Papineau, 2001 : 30)
[68] *Ibid.*, p.32.

l'identité rigide mental/physique a été encouragée, en outre, par l'introduction du concept de *survenance*, qui fut accueilli avec un formidable enthousiasme car il possède l'avantage de remplir le *vide métaphysique*, pour reprendre ici expression de J. Kim[69], séparant le domaine du mental de celui du physique. Dire que le mental survient sur le physique c'est donner une description *positive* de leur emboîtement. La survenance, nous allons le voir, arrange fort bien les affaires du physicalisme sur un point essentiel : elle permet une justification très claire de la prévalence des lois physiques sur le mental, et cela sans s'appuyer sur le procédé controversé de réduction Nagelienne, autrement dit, sans sacrifier l'autonomie du mental par rapport au physique. Mais avant de s'arrêter plus en détail sur ces points, une parenthèse définitionnelle s'impose.

Jusqu'ici nous avons présenté une vue synoptique, très générale, des fondements qui justifient l'ascension du physicalisme au XXe siècle. Les deux premières sections de ce chapitre ont traité la question des mécanismes qui ont poussé la pensée philosophique moderne à considérer la science physique – et son produit dérivé, le physicalisme – comme l'élève-modèle, i.e. le meilleur schéma explicatif capable de nous renseigner sur ce que le monde *est* ultimement. Le but a été de mettre en lumière l'importance de deux principes fondamentaux (le principe de l'unité de la science et le principe de la clôture causale, nomologique et explicative du monde physique) sur la base desquels le physicalisme assoit la légitimité de ses ambitions d'offrir une description complète et unifiée du réel.

Dans ce qui suit, nous allons esquisser, en prenant comme fil conducteur les travaux récents de D. Stoljar, le portrait standard du physicalisme, en passant par ses différentes variantes définitionnelles. Il nous faut, à ce stade, circonscrire avec plus de précision le premier acteur du duel que nous suivons ici, à savoir, le physicalisme.

[69] J. Kim, (1998), *Mind in a Physical World*, Cambridge: Cambridge University Press, p. 7.

1.3 Définir le slogan physicaliste standard : « Tout est physique »

Dans sa récente analyse du rôle du physicalisme dans la tradition analytique contemporaine, D. Stoljar[70] constate que toute tentative de définir le physicalisme se heurte d'emblée au dilemme suivant :

(A) D'une part, il est clair que le physicalisme est un modèle explicatif que nous avons de fortes raisons d'accepter au vu des découvertes et du succès des sciences empiriques. Le physicalisme n'étant pas une conception intrinsèquement *a priori* à la façon des doctrines mathématiques ou éthiques, il s'appuie avant tout sur un traitement *empirique* des données mises à disposition par des laboratoires de recherche hautement spécialisés. Ce qui veut dire que les penseurs qui le rejettent, bien qu'ils ne commettent aucune erreur conceptuelle ou logique, s'exposent néanmoins à l'opprobre plus ou moins virulent de la communauté scientifique.

(B) D'autre part, il est clair que malgré l'incontestable hégémonie dont le physicalisme jouit aux yeux de la communauté des savants, il y a cependant des faits dans la vie de tous les jours, des intuitions qui semblent, du moins en apparence, incompatibles avec la thèse « tout est physique ». Plus spécifiquement, le slogan physicaliste semble entrer d'emblée en conflit avec un certain nombre d'observations du sens commun. En voici une liste non-exhaustive :

(1) les individus éprouvent des sensations corporelles de toutes sortes (douleurs, crampes, nausées, démangeaisons, etc.) ;
(2) les individus ont la capacité de parler et d'échanger des réflexions quant à la nature du monde et des autres;
(3) au moins certains mots possèdent des *significations* ;
(4) les corps des individus, ainsi que les objets physiques en général, possèdent des textures, couleurs, odeurs spécifiques ;
(5) les individus agissent suivant certaines raisons, et poursuivent certains objectifs ;
(6) les individus agissent et pensent librement (du moins partiellement ou ponctuellement) ;

[70] D. Stoljar, (2010), *Physicalism*, op.cit., p. 13.

(7) les individus participent activement à des actions collectives, et prennent des décisions collectivement, dont l'impact sur les vies individuelles reste considérable ;
(8) il existe des vérités mathématiques et logiques (ex : « 5+7=12 ») accessibles *via* l'entendement[71].

Comment concilier la conception scientifique du monde (A) avec les intuitions du sens commun décrites en (B) ? Trois options s'offrent à nous : soit on abandonne le physicalisme, mais cette alternative semble peu probable si l'on tient compte du succès prédictif impressionnant des modèles explicatifs étayés par les sciences physiques. Soit on rejette les thèses du sens commun de (1) à (8), mais ce refus nous mettrait dans une position très embarrassante par rapport aux croyances et jugements qui influencent le quotidien de chacun d'entre nous. Enfin, une troisième option consiste à trouver un compromis, un juste milieu entre les deux pôles conflictuels.

Selon Stoljar, la question « Qu'est-ce que le physicalisme ? » peut être examinée en suivant trois pistes distinctes :

(i) En premier lieu, on peut s'interroger sur *l'interprétation* du slogan « tout est physique » (*la question de l'interprétation*).
(ii) Un deuxième ensemble d'interrogations concerne *la question de la vérité/fausseté* du physicalisme, ainsi que de la validité des arguments pour ou contre. En clair, il s'agit d'établir si le physicalisme est une conception vraie du monde.
(iii) Enfin, le troisième ensemble d'interrogations concerne les *enjeux philosophiques* du physicalisme, autrement dit la *signification métaphysique* de sa position omni-englobante, et de son rôle de *Weltanschauung* de l'époque moderne.

Dans cette section, nous allons nous concentrer uniquement sur (i) *la question de l'interprétation*, réservant la question de la vérité du physicalisme, ainsi que la question de sa portée métaphysique pour la deuxième partie de notre investigation. Précisons également que la question épineuse de la déductibilité *a priori* dans l'affaire du physicalisme fera l'objet d'un traitement à part, plus loin dans le chapitre 7 de cet essai, dans le cadre de l'analyse critique que nous comptons engager autour du concept de

[71] *Ibid.*, p. 14.

« physicalisme *a priori* » tel qu'il a été récemment défendu par F. Jackson[72]. Mais d'abord, il nous faut tirer au clair la signification du terme « physique » tout court.

Selon Stoljar[73], la *question de l'interprétation* du slogan « **tout est physique** » peut être approchée en mettant l'accent soit sur le terme situé à gauche de l'identité (« tout »), soit sur le terme situé à droite (« physique »). À partir de là, la question initiale se divise à son tour en deux interrogations distinctes:

(1_A) la *question de la complétude* : « Que signifie de dire que *'Tout* est physique'? »

(1_B) la *question de la condition* : « Que signifie de dire que 'Tout est *physique'* ? »

Dans le cas (1_A), on considère comme déjà acquise la signification du terme « physique », le reste du travail s'organisant autour de l'objectif de spécifier l'application du prédicat « est physique » à l'*ensemble* des états, phénomènes, et objets réels. Dans le deuxième cas (1_B) on part du principe que la condition d'applicabilité à la totalité du réel est non-problématique; en revanche, on s'interroge sur la *nature* même de cette condition : qu'est ce que cela signifie le prédicat « être physique » ? Nous examinerons plus amplement cette distinction lorsque nous aborderons les notions clé d'*omniscience* et de *nécessitation a priori* autour desquelles se construit le physicalisme contemporain. Une étape essentielle reste à franchir, qui est celle de définir adéquatement le physicalisme en partant de son cœur même, le concept de « physique ».

Prenons comme point de départ l'énoncé *p* « Tout est physique ». Compte tenu du fait que le mot « tout » est un quantificateur universel, une façon plus commode et plus rigoureuse de retranscrire *p* serait de reformuler à nouveaux frais la question sous-jacente suivante : si tout est physique, quelles sont les choses qui *composent* le « tout » physique ? Cette question nous permet, note Stoljar[74], d'avancer une toute première tentative de définir le physicalisme :

(1) Le physicalisme est vrai si et seulement si *absolument tout ce qui existe* est physique.

Si l'on examine de plus près la définition (1), on se rend compte qu'elle pèche par excès de zèle, car elle est excessivement forte. Quelqu'un pourrait objecter en prenant l'exemple du chiffre « 2 » : il serait en effet étrange de stipuler que « 2 » est non-réel

[72] F. Jackson, (1998), *From Metaphysics to Ethics: A Defense of Conceptual Analysis*, op.cit.
[73] D. Stoljar, 'Physicalism' *in* Stanford Encyclopedia of Philosophy http://plato.stanford.edu
[74] (D. Stoljar, 2010 : 29)

car non-physique. Mais si « 2 » est physique, dans quel sens l'est-il ? Un autre cas problématique serait alors celui d'un organisme juridique comme la Cour de Cassation, disons. Personne ne songe à nier l'existence de la Cour de Cassation, mais en même temps, il serait pour le moins étrange d'affirmer que la Cour de Cassation *est* physique. A partir de là, deux options s'ouvrent à nous : soit on rejette la définition (1) en concluant que le physicalisme est faux, et on s'arrête là. Soit on accorde une autre chance au physicalisme, en essayant de l'assouplir et de moduler sa définition en fonction d'autres critères. Supposons, note Stoljar[75], que vous vous prépariez à partir en vacances à la campagne et que vous remplissiez le coffre de votre voiture avec vos affaires. Votre femme vous demande : « Tout est prêt ? » Et vous répondez, peut-être avec la légère fierté du travail bien fait : « *Tout* est dans le coffre ». Bien sûr, le dernier énoncé ne veut pas dire que « *absolument tout ce qui existe* est dans le coffre », mais plutôt « toutes les affaires dont j'aurai besoin à la campagne sont dans le coffre ». De manière analogue, les physicalistes pourraient restreindre le champ d'application de l'énoncé « tout est physique » aux seules entités pertinentes, qui « comptent », pour ainsi dire, dans l'affaire. Cela dit, la nouvelle question est : lesquelles ? Quelles sont les entités pertinentes à faire rentrer dans le « coffre » du physicalisme ?

Certains auteurs ont proposé alors de restreindre le périmètre définitionnel aux seules entités particulières physiques, obtenant ainsi la définition suivante :

(2) Le physicalisme est vrai si et seulement si chaque *entité particulière concrète* est physique.

Ici le terme « entité particulière » ou « occurrence » doit être compris dans sa différenciation paradigmatique d'avec le concept de « propriété ». Les entités sont des instanciations concrètes de l'être, tandis que les propriétés sont des *modes sous lesquels* les entités se présentent à l'être. L'eau possède la propriété d'être liquide. Les fraises mûres possèdent la propriété d'être rouges, et ainsi de suite. Mais nous ne sommes toujours pas au bout de nos peines, car comme le relève Stoljar, si la définition (1) avait le défaut d'être trop forte et restrictive, la proposition (2) se montre trop lâche, dans la mesure où elle n'exclut pas le dualisme des propriétés. Une façon commode de se soustraire à cette difficulté serait de raffiner encore plus le tri, et ne garder que les propriétés. On obtient ainsi :

[75] *Ibid.*, p. 30.

(3) Le physicalisme est vrai si et seulement si chaque *propriété* est une propriété physique.

Bien que la restriction aux seules propriétés nous mette sur la bonne voie, l'énoncé (3) reste néanmoins, trop prohibitif. Pour voir cela, il suffit de considérer le cas suivant : considérons la propriété *d'être président de la France issu du Parti Socialiste en 2007*. C'est une propriété bien réelle, et d'ailleurs Ségolène Royal a failli être la personne à qui l'on aurait pu attribuer cette propriété. Mais, compte tenu du fait que les électeurs en ont décidé autrement, il résulte que la propriété *d'être président de la France issu du Parti Socialiste en 2007* se retrouve ainsi non-instanciée. Pour contourner cette impasse, certains proposent alors une définition se limitant aux seules propriétés instanciées :

(4) Le physicalisme est vrai si et seulement si chaque *propriété instanciée* est une propriété physique.

Comme le relève Stoljar, il suffit que l'on pense à l'exemple de la Cour de Cassation mentionné précédemment, pour réaliser que l'énoncé (4) s'avère lui aussi trop restrictif, bien qu'il soit, une fois de plus, sur la bonne voie. Une autre solution a été proposée par D. Lewis[76]. Selon Lewis, notre monde, ainsi que tous les autres mondes possibles, sont constitués d'entités qui instancient des propriétés et des relations *fondamentales*. Le nombre de propriétés fondamentales est fort restreint[77]. Tout ce qui existe dans ce monde – écrit Lewis – « est une vaste mosaïque d'occurrences locales de faits particuliers, rien qu'une petite chose et puis une autre, et ainsi de suite. (…) Nous avons la géométrie : un système de points avec des relations externes de distance spatio-temporelle entre eux. Peut-être des points de l'espace-temps lui-même, peut-être des points de matière (ou d'un éther ou de champs) ; peut-être les

[76] "This world, or any possible world, consists of things which instantiate fundamental properties and which, in pairs or triples or . . . , instantiate fundamental relations. Few properties are fundamental: the property of being a club or a tub or a pub, for instance, is an unnatural gerrymander, a condition satisfied by miscellaneous things in miscellaneous ways. A fundamental, or 'perfectly natural', property is the extreme opposite. Its instances share exactly some aspect of their intrinsic nature. Likewise for relations. I hold, as an *a priori* principle, that every contingent truth must be made true, somehow, by the pattern of coinstantiation of fundamental properties and relations. The whole truth about the world, including the mental part of the world, supervenes on this pattern. If two possible worlds were exactly isomorphic in their patterns of coinstantiation of fundamental properties and relations, they would thereby be exactly alike *simpliciter*." D. Lewis, (1994b), 'Reduction of Mind', *in* S. Guttenplan (ed.), *A Companion to Philosophy of Mind*, Oxford: Blackwell Publishers, p. 412. Voir aussi D. Lewis, (1986a), *On the Plurality of Worlds*, Oxford: Blackwell Publishers.

[77] Nous aurons l'occasion de revenir plus en détail sur la notion de « niveau fondamental » et de « propriété fondamentale » dans le chapitre 7.

deux à la fois. En ces points se trouvent des qualités locales : des propriétés intrinsèques parfaitement naturelles qui n'ont besoin de rien de plus grand qu'un point où être instanciées. En bref : nous avons un arrangement de qualités. Et c'est tout. Il n'y pas de différence sans différences dans l'arrangement des qualités. Tout le reste survient sur cet arrangement »[78].

On obtient ainsi une version améliorée de la définition initiale:

(5) Le physicalisme est vrai si et seulement si chaque *propriété fondamentale instanciée* est physique.

Si l'on reprend à nouveau l'exemple de la propriété d'« être une Cour de Cassation », et qui nous sert d'ailleurs de critère de vérification, on s'aperçoit que l'énoncé (5) passe facilement le test : la propriété d'être Cour de Cassation est instanciée, mais elle n'est pas ce que Lewis appelle une propriété fondamentale. Par conséquent, elle ne réfute pas le physicalisme et l'énoncé (5) a plus des chances que l'énoncé (4) d'être retenu. Mais même avec toutes ces précautions, nous ne sommes pas au bout de nos peines, car une formulation de ce type ne prend pas en compte le statut des *relations*. Il faut alors élargir la définition sous la forme suivante (6) :

(6) Le physicalisme est vrai si et seulement si chaque propriété instanciée est soit physique, soit dépend d'une certaine relation *R* à une propriété physique instanciée.

Selon Stoljar[79], la thèse impliquant des propriétés fondamentales amène avec elle un sérieux problème supplémentaire, à savoir celui de déterminer quelle est la relation qui rattache ces propriétés fondamentales aux propriétés instanciées « de tous les jours», si l'on peut les appeler ainsi. Il devient alors évident que l'on doit prendre en compte non seulement les propriétés isolables à la façon des briques idéales composant un mur abstrait, mais aussi dans leur agencement dynamique, ce qui nous conduit inévitablement à statuer sur la manière dont les propriétés instanciées s'articulent aux propriétés physiques fondamentales et en dépendent. Quel est le ciment qui assure la cohérence entre les différentes strates, bref, quelle est la nature de 'R' ? Parmi les candidats pour déterminer la nature de *R* on compte la réduction, la survenance, la réalisabilité, et la nécessitation métaphysique. Nous allons retenir ici, provisoirement, seulement cette dernière solution, à savoir la « nécessitation métaphysique », en raison

[78] (D. Lewis, 1986 : ix-x)
[79] (D. Stoljar, 2010: 36)

de ses implications ouvertement métaphysiques. La nouvelle définition se présenterait alors ainsi :

(7) Le physicalisme est vrai si et seulement si chaque propriété instanciée est, soit physique, et/ou soit elle découle par nécessité d'une autre propriété physique instanciée.

La notion de *nécessitation métaphysique* à laquelle on fait référence ici est celle proposée par S. Kripke[80], et qui stipule qu'une propriété *F* implique nécessairement une propriété *G* si l'instanciation de *F* est nécessairement accompagnée de l'instanciation de *G*. Cette idée est généralement énoncée en faisant appel à la notion de « mondes possibles ».

Dans la section suivante il sera question de recadrer les différentes stratégies définitionnelles que nous venons de passer en revue, en leur rajoutant la clause essentielle de la nécessitation métaphysique. Avant d'entamer ce périple autour de ces points, quelques précautions sont à prendre. La complexité du sujet rend illusoire toute prétention de pouvoir présenter ici, de manière exhaustive, la globalité des enjeux qui concernent le problème métaphysique des mondes possibles. Nous aborderons, avec les modestes outils de notre entreprise, certaines des questions qui entourent la notion de « nécessitation métaphysique », mais nous ne serons absolument pas exhaustifs. Il ne sera pas non plus question dans cette section d'un certain nombre de problèmes relatifs aux controverses qui concernent la *caractérisation* de propriétés physiques. On ne les mentionnera que pour les mettre de côté car ils feront l'objet d'une section à part (1.5), plus loin au cours de notre investigation. On se contentera de noter ici quelques aspects cruciaux relatifs aux diverses tentatives d'encadrer conceptuellement la notion clé de « nécessitation » dans l'affaire du physicalisme.

1.4 Physicalisme et Nécessitation

A la fin de l'ouvrage *Naming and Necessity*, Kripke lançait son célèbre défi aux défenseurs du matérialisme: « Le matérialisme doit, à mon sens, soutenir qu'une description physique du monde est une description complète, que tout fait mental est

[80] S. Kripke, (1980), *Naming and Necessity*, Harvard University Press, - parution initiale 1972; trad. fr. de P. Jacob et F. Récanati, (1982), *La logique des noms propres*, Paris, Les Ed. de Minuit. (Nous utilisons ici l'édition française).

"ontologiquement dépendant" des faits physiques (c'est-à-dire en découle par nécessité). Aucun théoricien de l'identité n'a, que je sache, présenté un argument convaincant contre le sentiment intuitif selon lequel ce n'est pas le cas »[81]. Soit « A » le nom d'une sensation de douleur particulière et soit « B » le nom de l'état du cerveau correspondant ou de l'état du cerveau avec lequel un certain théoricien de l'identité psychophysique souhaiterait identifier A. « Si A=B, insiste Kripke, alors l'identité entre A et B est nécessaire et toute propriété essentielle de l'un doit être une propriété essentielle de l'autre. Celui qui défend la thèse de l'identité ne peut pas se contenter d'admettre les intuitions cartésiennes : que A peut exister sans B, que B peut exister sans A, que la présence de telle ou telle propriété mentale n'est qu'en corrélation contingente avec l'existence de B, que la présence de telle ou telle propriété physique spécifique n'est qu'en corrélation contingente avec l'existence de A»[82]. Examinons à présent la manière dont les physicalistes entendent soulever ce défi.

1.4.1 Propriétés physiques instanciées et mondes possibles

La notion de *nécessitation métaphysique* kripkéenne ne concerne pas *certains* mondes possibles, mais *tous* les mondes possibles. Essayons, pour commencer, de reformuler à nouveaux frais la définition (7) du physicalisme, en y englobant cette fois-ci la notion de monde possible *w*[83] :

> $(1)_w$ Le physicalisme est vrai pour un monde possible *w* si et seulement si pour chaque propriété *F* instanciée en *w*, il y a une propriété physique *G* instanciée en *w* de sorte que, dans tous les mondes possibles *w**, si *G* est instanciée en *w**, alors *F* est instanciée en *w**[84].

Selon Stoljar[85], l'idée centrale derrière $(1)_w$ est la suivante : si l'on considère, d'une part les propriétés instanciées dans un monde déterminé (disons notre monde actuel), et l'on prend, d'autre part, les propriétés physiques instanciées dans ce dernier

[81] (S. Kripke, 1980/ 1982:144)
[82] *Ibid.*, pp. 136-7.
[83] Le signe '*w*' désigne ici un monde possible arbitrairement choisi. Nous avons préféré de garder ici la terminologie habituellement utilisée dans la littérature analytique anglaise *w* (« world ») pour désigner « monde possible ».
[84] (D. Stoljar, 2010:112)
[85] Stoljar signale à juste titre que dans la définition $(1)_W$ le concept de « propriété » peut être remplacé, dans certaines variantes, par les concepts d' « états de choses » ou de « particulier *X* ». (Voir D. Stoljar, 2010 : 113)

monde, alors $(1)_w$ dit que si le physicalisme est vrai pour notre monde actuel, alors pour chaque propriété appartenant au premier groupe, il y a une propriété dans le second groupe en sorte que la première propriété est nécessairement impliquée – dans le sens de nécessitation métaphysique forte – par la deuxième.

Jusqu'ici nous avons pris comme point de départ et outil conceptuel la notion kripkéenne de « nécessitation métaphysique ». Mais il est possible de penser la définition du physicalisme en s'appuyant sur une autre notion clé de l'appareillage conceptuel analytique, à savoir la notion de « survenance »[86] (sur laquelle nous allons revenir plus amplement au chapitre dédié au fonctionnalisme). On a tôt fait de remarquer que l'introduction de ce concept s'est avérée salutaire dans l'articulation des nombreux champs théoriques : on dit par exemple que la sphère du moral ou de l'esthétique *survient* sur celle du physique. On s'appuie également sur la notion de survenance pour expliquer les relations entre la sphère de l'intentionnel ou des qualités phénoménales et les particules élémentaires physiques. Pour le but introductif de cette section, il suffit de mentionner que dans la définition du physicalisme, la notion de survenance s'accompagne de l'insertion du concept de « double » ou de « duplicata » des mondes possibles. En synthétisant les formulations qu'en donnent D. Lewis[87] et F. Jackson[88], on obtient ainsi la variante suivante :

[86] Rappelons ici, pour mémoire, le passage où D. Davidson introduit la notion de "survenance": « Although the position I describe denies there are psychophysical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect. Dependence or supervenience of this kind does not entail reducibility through law or definition: if it did, we could reduce moral properties to descriptive, and this there is good reason to *believe* cannot be done; and we might be able to reduce truth in a formal system to syntactical properties, and this we *know* cannot in general be done."(D. Davidson 1970:214)

[87] D. Lewis, (1983), 'New Work for a Theory of Universals', *Australasian Journal of Philosophy*, 61(4): 343–377. Voir aussi D. Lewis, (1986), *On the Plurality of Worlds*, Oxford: Blackwell.

[88] Consider any possible world that is a minimal physical duplicate of our world. It is, that is, exactly like ours in every physical respect: it is physical individual, property and relation exactly like our world, and moreover it contains nothing extra; it contains nothing more than it has to in order to be physically exactly like our world. (We can count the necessarily existing entities, if there are any, that all worlds have in common as trivially physical for our purposes here.) Materialists who hold that materialism is a complete account of our world, or a complete account of our world as far as the mind is concerned—materialists who are, that is, not some kind of dual attribute dualist—must hold that these minimal physical duplicates are psychological duplicates of our world. They must, that is, hold the following supervenience thesis: (S) Any world that is a minimal physical duplicate of our world is a psychological duplicate of our world." F. Jackson, (1995), 'Postscript on 'What Mary didn't know'", in P. Moser & J. Trout (1995), op. cit., p. 192. Voir aussi F. Jackson, 1998.

> $(2)_W$ Le physicalisme est vrai pour un monde *w* si et seulement si pour chaque monde possible *w**, si *w** est un double physique de *w*, alors *w** est un double *simpliciter* du monde *w*.

Avant de regarder de plus près ces définitions, notons d'abord que $(1)_W$ et $(2)_W$ sont des thèses modales. Comme le résume M. Esfeld, « si nous imaginons la distribution entière des états physiques fondamentaux du monde redoublée (y compris les lois de la nature) – au sens où l'on duplique uniquement la distribution entière des états physiques fondamentaux du monde sans y ajouter quelque chose qui n'existe pas dans le monde réel -, créant ainsi un double physique *minimal* du monde réel, nous obtenons comme résultat que *tout* ce qu'il y a dans le monde – y compris les personnes et leurs états mentaux – est redoublé »[89].

Ainsi que le note Stoljar, la définition $(1)_W$ affirme que si le physicalisme est vrai, alors toute propriété instanciée est nécessairement impliquée par une propriété physique, sachant que la notion de « nécessitation » utilisée ici est celle kripkéenne, forte, définie en termes de mondes possibles. Autrement dit, si le physicalisme est vrai, alors ses thèses sont vraies dans *tous* les mondes possibles. Naturellement, cette affirmation peut soulever des difficultés de taille. La première est : sur quelle base doit-on définir le physicalisme de cette façon ambitieuse ? Autrement dit, est-ce que la partie droite de la définition $(1)_W$ est une *condition nécessaire* de la partie gauche ? Certains auteurs ont tenté d'assouplir la nécessitation métaphysique forte et de la transformer sous une forme plus modeste, nomologique :

> $(3)_W$ Le physicalisme est vrai pour un monde possible *w* si et seulement si pour chaque propriété *F* instanciée en *w*, il y a une propriété physique *G* instanciée en *w* de sorte que, dans tous les mondes *nomologiquement* possibles *w**, si *G* est instanciée en *w**, alors *F* est instanciée en *w**[90].

Le problème avec cette définition, note Stoljar, c'est qu'elle n'est pas incompatible avec certaines formes de dualisme, comme celui de D. Chalmers[91] par exemple, qui est compatible avec la nécessitation nomologique entre les propriétés mentales et celles physiques. Or, nous avons besoin d'une définition du physicalisme qui *exclue* cette possibilité, et ne laisse planer l'ombre d'aucun doute dualiste. Par

[89] M. Esfeld, (2005), *La philosophie de l'esprit : de la relation entre l'esprit et la nature*, Paris, Armand Colin, p. 44. (Soulignements originaux.)
[90] (D. Stoljar, 2010:128)
[91] (D. Chalmers, 1996)

conséquent, il semblerait que la nécessitation métaphysique soit ce dont nous avons besoin, si nous voulons définir le physicalisme de façon non-équivoque. Mais cela ne va pas sans demander quelques éclaircissements supplémentaires. D'abord, quelqu'un pourrait objecter que le physicalisme ne concerne *que* notre monde actuel : pourquoi se soucier alors d'hypothétiques mondes possibles ? A cela on peut répondre en disant : certes, l'enjeu du physicalisme vise le monde actuel, mais nous avons vu que la définition $(1)_W$ possède un contenu modal. Ce serait faire preuve de parcimonie injustifiée que d'utiliser la thèse $(1)_W$ dans un sens restreint. Nous laisserons ici de côté la question « qu'est-ce qu'un monde possible ? » qui mériterait, à elle seule, un traitement détaillé dans un ouvrage à part.

Le deuxième éclaircissement concerne le choix de préférer ici la définition $(1)_W$ et non pas celles de (1) à (7) que nous avons mentionnées plus haut dans le paragraphe précédent. Etant donné qu'il est impératif que le physicalisme soit différencié des formes rivales de dualisme, il est préférable d'employer une définition qui ne laisse aucune place aux ambiguïtés. Enfin, pour terminer par le point le plus important, même si nous sommes prêts à entamer notre enquête en nous appuyant sur la thèse $(1)_W$, nous ne sommes toujours pas au bout de nos peines, car il reste beaucoup d'autres points à clarifier.

1.4.2 Nécessitation, contingence et ectoplasme

Définir le physicalisme en termes de nécessitation pose plusieurs problèmes. Le plus frappant, sans doute, concerne la relation entre la thèse $(1)_W$ et la thèse selon laquelle le physicalisme, à la différence de certaines doctrines mathématiques ou éthiques, reste une doctrine *contingente* avant tout. Dans ce cas, comment accommoder la contingence du physicalisme avec sa définition en termes de nécessitation métaphysique ?

Une possible stratégie de réponse consiste à faire appel à la notion kripkéenne de désignateur rigide[92]. Selon Kripke, nous le savons bien, l'assertion « Jean-Michel César

[92] Kripke démontre l'insuffisance du modèle descriptif de la signification des noms selon lequel nous devrions être capables d'identifier le porteur du nom à l'aide des descriptions abrégées par ce nom. Or, le contexte modal renvoie à la question : comment la référence à l'individu se maintient dans un monde possible où les descriptions que nous associons à la référence ne valent plus. En effet, si nous désignons Aristote par la description « a écrit la *Métaphysique* » et « a été le disciple de Platon », comment

= l'empereur des Romains » est contingente, en vertu du fait que le membre de droite qui flanque le signe de l'identité n'est pas un désignateur rigide. Une expression est un désignateur rigide si et seulement si elle désigne le même objet dans tous les mondes possibles dans lesquels le monde respectif existe. Par conséquent, la définition du physicalisme en termes de nécessitation peut être contingente si l'un de ses membres constituants s'avère un désignateur non-rigide. En clair, l'assertion de nécessitation « toute propriété instanciée est nécessairement impliquée par une propriété physique donnée» est contingente en raison du fait que l'extension des expressions « toute propriété instanciée » et «propriété physique donnée » diffère d'un monde à l'autre. Par exemple, dans le monde cartésien, la thèse $(1)_W$ est fausse, tandis que dans celui imaginé par un matérialiste comme J.J.C. Smart, disons, elle est vraie.

Ces considérations peuvent conduire à l'idée selon laquelle, avec l'introduction de la distinction kripkéenne entre désignateur rigide et non-rigide, nous avons vaincu le denier dragon qui protège la citadelle de la définition $(1)_W$. Mais ce serait crier « victoire » trop tôt. Car il y a une deuxième critique que l'on peut adresser à la stratégie définitionnelle en termes de nécessitation, et qui cible un problème beaucoup plus épineux que celui que nous venons d'évoquer, problème qui s'est frayé chemin dans la littérature analytique contemporaine sous le nom exotique d'«ectoplasme épiphénoménal ». Selon la synthèse qu'en propose Stoljar[93], cette difficulté prend la forme suivante : imaginons, pour plus de commodité, que dans notre monde actuel il n'y a que 17 types d'épisodes mentaux. (C'est une simplification outrageuse, bien évidemment, mais fort utile pour illustrer la question qui nous intéresse ici.) A présent, s'il y a 17 épisodes psychologiques, il doit y avoir également une propriété spécifique, à savoir la propriété *d'être en sorte qu'il n'y a que 17 épisodes mentaux*. Tout ce qui est instancié dans ce monde possède cette propriété que nous appellerons, pour plus de facilité, P_{17}. Jusqu'ici, rien de problématique. Mais considérons l'alternative suivante : imaginons un autre monde possible, *w** qui est identique avec notre monde *w* sous tous ses propriétés physiques qui y sont instanciées, mais avec une seule différence : *w**

procédons-nous pour comprendre le nom propre « Aristote » employé pour désigner Aristote dans un monde possible où celui-ci n'aurait ni écrit la *Métaphysique*, ni été le disciple de Platon ? La solution de Kripke est d'introduire le concept de « désignateur rigide » pour permettre de désigner le même individu dans tous les mondes possibles, malgré les éventuelles modifications que nous lui apportons par imagination. Kripke pose le fondement de la distinction entre désignateurs rigides et non rigides dans un test intuitif auquel il attache beaucoup d'importance : « bien qu'il eût été possible que quelqu'un d'autre que celui qui est en fait le président des Etats-Unis en 1970 (par exemple, Humphrey aurait pu l'être), personne d'autre que Nixon n'aurait pu être Nixon ». (Voir S. Kripke, 1980/1982 : 36)

[93] (D. Stoljar, 2010:133)

contient *en plus* certaines propriétés mentales dont la nature est constituée d'*ectoplasmes* (c'est-à-dire des entités essentiellement non-physiques). Cerise sur le gâteau, ces entités non-physiques sont également *épiphénoménales*, dans le sens où elles n'ont aucun pouvoir ou efficience causale sur rien du tout. En bref, *w** est un monde qui est exactement comme notre monde actuel, sauf qu'il contient quelques ectoplasmes additionnels. Le point névralgique est le suivant : il est pour le moins évident que le tenant du physicalisme ne peut pas nier que *w** soit un monde *possible*. Le physicaliste est en droit de nier que *w** soit notre monde actuel, mais dire que *w** n'est pas *actuel* ne veut pas dire que *w** soit *impossible*. Or, *si le physicalisme est vrai* pour notre monde actuel, il s'ensuit que *w** n'est pas possible. Cependant, un monde *W** tel que nous venons de le décrire semble parfaitement concevable et possible, compatible de surcroît avec la nature contingente de la doctrine du physicalisme que nous avons évoquée plus haut, au début de cette section. Comment sortir de cette impasse ?

Deux sorties de secours ont été suggérées. La première, évoquée par F. Jackson[94] consiste à introduire *la clause de seuil minimal* : en clair, au lieu d'affirmer que les propriétés physiques impliquent nécessairement toutes les autres propriétés instanciées, on soutient que *seules* les propriétés physiques impliquent nécessairement toutes les autres propriétés instanciées. On obtient ainsi une version améliorée de la définition du physicalisme :

> $(4)_W$ Le physicalisme est vrai pour un monde possible *w* si et seulement si pour chaque propriété *F* instanciée en *w*, il y a une propriété physique *G* instanciée en *w* en sorte que, dans tous les mondes possibles *w**, si *G* est instanciée en *w**, et *rien d'autre n'est instancié en w**, alors *F* est instanciée en *w**.[95]

Cette définition écarte le problème de l'ectoplasme, mais laisse intacte la question de la signification de l'expression « rien d'autre n'est instancié ».

La deuxième suggestion nous vient de D. Chalmers qui propose, lui aussi, d'introduire une restriction supplémentaire. Plus précisément, il s'agit de restreindre la formulation du physicalisme aux seules propriétés qu'il baptise « propriétés positives ».

---

[94] (F. Jackson, 1998: 12 - 25)
[95] (D. Stoljar, 2010: 135) (Nous traduisons)

Qu'entend-on par « propriété positive » ? Selon Chalmers[96], une propriété positive est une propriété telle que si elle est instanciée dans un monde *w*, elle est également instanciée par l'occurrence correspondante dans tous les mondes qui contiennent *w* comme élément-part. Il est commode d'appeler ces propriétés des *propriétés c-positives*. Ayant établi ces éléments d'arrière-plan, nous sommes en mesure à présent d'avancer une autre version dans la définition du physicalisme :

> $(5)_W$ Le physicalisme est vrai pour un monde possible *w* si et seulement si pour chaque *propriété c-positive F* instanciée en *w*, il y a une propriété physique *G* instanciée en *w* en sorte que, dans tous les mondes possibles *w**, si *G* est instanciée en *w**, alors *F* est instanciée en *w**[97].

Cette définition écarte elle aussi le danger de l'ectoplasme, car il est clair, pour revenir à l'exemple évoqué plus haut, que la propriété que nous avons baptisée P_{17} *n'est pas* une propriété c-positive. Notons entre parenthèses que le problème de l'ectoplasme touche non seulement la définition du physicalisme en termes de nécessitation, mais aussi sa définition en termes de survenance. Rappelons-nous cette dernière :

> $(2)_W$ Le physicalisme est vrai pour un monde *w* si et seulement si pour chaque monde possible *w**, si *w** est un double physique de *w*, alors *w** est un double *simpliciter* du monde *w*.

Si la définition $(2)_W$ est vraie, alors elle exclue des mondes possibles qu'elle *ne devrait pas* exclure. Comme nous venons de le constater, afin de contourner cette difficulté, deux stratégies sont applicables. Celle de Jackson (1998) consiste à introduire la clause du seuil minimal, le résultat étant $(6)_W$:

> $(6)_W$ Le physicalisme est vrai pour un monde *w* si et seulement si pour chaque monde possible *w**, si *w** est un double physique *minimal* de *w*, alors *w** est un double *simpliciter* du monde *w*.

Par « double physique minimal » Jackson entend un monde possible qui est identique sous tous les aspects physiques avec le monde *w*, mais qui ne contient rien d'autre *en plus*. Bref, il ne contient pas d'éventuel ectoplasme. De son côté Chalmers propose de remplacer $(2)_W$ par $(7)_W$ ci-dessous:

---

[96] (D. Chalmers, 1996: 40)
[97] (D. Stoljar, 2010:136)

(7)$_w$ Le physicalisme est vrai pour un monde *w* si et seulement si pour chaque monde possible *w**, si *w** est un double physique de *w*, alors *w** est un *double c-positive* du monde *w*.

Le syntagme « c-positive double du monde *w* » désigne un monde possible qui instancie toutes les propriétés c-positives du *w*.

En cherchant à désamorcer la tension opposant les différentes tentatives de définir le physicalisme en termes de nécessitation et les contre-arguments de type « ectoplasme », on s'aperçoit que les tenants du physicalisme ont un sérieux défi supplémentaire à relever. Certains auteurs ont répliqué en signalant que le type de difficulté contenue dans l'histoire de l'ectoplasme ne concerne pas le physicalisme en particulier, mais plutôt *toutes* théories confondues qui cherchent à fournir une théorie explicative *complète* du monde. Nous reviendrons sur ce point dans la dernière partie de ce chapitre.

Jusqu'ici nous nous sommes intéressés principalement aux différentes versions et tentatives de définir le physicalisme. Ainsi avons-nous abordé la façon dont la notion de « nécessitation » s'applique aux diverses stratégies explicatives impliquant les relations entre le mental et le physique. Il nous reste une tâche majeure à accomplir dans ce chapitre. A la lumière des considérations faites jusqu'ici, quelqu'un pourrait objecter : essayer de définir le physicalisme sans rien dire sur la signification du terme « propriété physique », ou de préciser qu'est-ce que pour une propriété d'être physique, serait comme si on essayait de définir l'« impérialisme » sans rien dire sur la signification du terme « empire ». Dans ce qui suit, nous allons approfondir cet aspect : qu'est-ce que pour une propriété d'être physique ?

1.5 Deux conceptions du 'physique'

Récemment, le métaphysicien P. Van Inwagen[98] signalait le paradoxe suivant : si au niveau fondamental, le réel se décompose ultimement en des atomes méréologiques, i.e. des entités simples, indécomposables en des parties encore plus petites, il s'ensuit qu'en dernière instance, lorsqu'on se cogne l'orteil contre un fauteuil, disons, ce n'est pas la collision avec le fauteuil qui nous fait mal, mais la collision avec les atomes. En

[98] P. Van Inwagen, (1990), *Material Beings*, Ithaca: Cornell University Press.

voilà une autre bizarrerie des philosophes !- pourrait-on s'exclamer. Comment rendre compte de cette étrange situation ? L'exemple de l'orteil cogné contre le fauteuil/électron nous indique l'urgence, pour le physicalisme, de définir la notion d' « objet » ou de « propriété physique ». Selon Stoljar[99], on peut avancer deux pistes définitionnelles distinctes : soit on fait appel au modèle des « objets » physiques, soit à celui d'une « théorie physique ». Commençons par la première.

1.5.1 L'approche basée sur l'Objet Physique

Il est d'abord commode de représenter une *propriété* physique comme étant une propriété appartenant à un *objet* physique. Qu'est-ce qu'un objet physique ? Selon Stoljar, une possible définition synthétique aurait la forme suivante :

> $\text{Objet}_{\text{physique}}$: *X* est un objet physique si et seulement si *X* possède (suffisamment) les propriétés suivantes : il a une dimension, une forme, une extension dans l'espace, la capacité de se déplacer ou d'être déplacé dans l'espace, ainsi que la capacité de manifester des divers processus, comme par exemple celui de se briser, de brûler, etc. ; mais le caractère le plus important est qu'il possède une solidité qui oppose intrinsèquement une résistance aux actions que l'on exerce sur lui[100].

Cette définition, Stoljar l'admet volontiers, est loin d'être satisfaisante, mais faute de mieux, nous allons nous appuyer, du moins provisoirement, sur le maigre point de départ qu'elle nous offre. Car non seulement sa portée explicative est faible, mais aussi elle laisse entendre qu'il serait possible de passer subrepticement de la caractérisation d'un *objet* physique à la définition d'une *propriété* physique. Or certains auteurs comme B. Stroud[101] ont démontré avec succès que cette opération de transfert conceptuel depuis l'objet physique vers une propriété physique ne va pas de soi. A partir de là, plusieurs options s'ouvrent à nous : (i) soit on affirme qu'une propriété est

[99] D. Stoljar, (2001), 'Two Conceptions of the Physical', *Philosophy and Phenomenological Research*, 62: 253-281. Réédité in P. Ludlow, Y. Nagasawa et D. Stoljar (eds.) (2004), op. cit., p. 312. (Nous nous référons ici à cette dernière édition.)

[100] "*X* is a physical object if and only if *X* has (or has enough of) the following properties: it has size, shape, extension in space, the capacity to move and to be moved, the capacity to undergo various processes such as bending, breaking, and burning, and perhaps most importantly it has solidity or bulk – that is, it is intrinsically such that it resists or would resist pressure from other physical objects, for example, pressure from human bodies." (D. Stoljar, 2010: 52)

[101] Voir B. Stroud, (1986), 'The Physical World', *Proceedings of the Aristotelian Society*, 87: 263–277.

physique exactement dans le même sens où un objet est physique. Mais cela ne nous avance guère, car les propriétés elles-mêmes ne possèdent pas les propriétés que les objets possèdent intrinsèquement. (ii) Soit on soutient qu'une propriété est physique uniquement si elle est instanciée par un objet physique. Mais cette option n'est pas satisfaisante non plus, car les objets physiques peuvent avoir diverses propriétés, pas toutes physiques. En outre, si l'on adopte cette alternative, on se place dans l'incapacité de distinguer entre le physicalisme et le dualisme des propriétés. (iii) Enfin, soit on soutient qu'une propriété est physique si et seulement si elle est instanciée par un objet *purement* physique ; mais dans ce cas, on tombe dans une circularité qui est non seulement vicieuse, mais aussi non-informative.

Au terme de cette brève présentation de différentes tentatives de définir une propriété physique, le lecteur pourrait tout simplement baisser les bras et admettre qu'il existe une barrière logique insurmontable qui nous empêche de définir la *propriété* physique par le truchement du concept d'*objet* physique. Ainsi, le projet de définir le physicalisme semble-t-il condamné d'avance. Mais est-ce la bonne attitude à adopter ? Certains philosophes ont choisi courageusement de lever le défi et de tenter une réponse en s'appuyant sur la notion de « concepts-faisceau ». Qu'est-ce qu'un « concept-faisceau » ? Il n'est pas aisé d'en donner une définition précise, mais la manière la plus commode de le présenter c'est d'en donner un exemple.

Imaginons que l'on rencontre un jour une communauté francophone de personnes qui parlent la langue française exactement comme nous le faisons, mais qui utilisent en plus, un concept dont nous n'avons jamais eu la connaissance, le verbe 'musiner' disons. Lorsqu'on leur demande ce que 'musiner' veut dire, elles nous répondent que 'musiner' veut dire « jouer du tuba, le mercredi, exclusivement par un individu nommé Michel ». Autrement dit, *il est analytiquement nécessaire* que l'opération de « musiner » soit performée par le dénommé Michel, le mercredi, et à l'aide d'un tuba. En partant de cet exemple, on peut proposer de traiter le concept de « propriété physique » de manière analogue. Selon la « définition-faisceau » qu'en propose Stoljar[102], la définition du concept de « propriété physique » se déploie sur les niveaux d'investigation suivants :

[102] (D. Stoljar, 2010: 56)

(a) Objet	• les propriétés physiques sont des propriétés distinctives des objets intuitivement physiques
(b) Théorie	• les propriétés physiques sont des propriétés exprimées par les prédicats des théories physiques
(c) Objectivité	• les propriétés physiques sont objectives ou intersubjectives (i.e. instanciées à partir de plusieurs points de vue possibles)
(d) Méthode	• les propriétés physiques sont celles que l'on arrive à connaître par le truchement des méthodes investigatrices des sciences de la nature
(e) Contraste	• les propriétés physiques *ne* sont *pas* de propriétés distinctives des âmes, des ectoplasmes, epiphénomènes, etc.

Avec ces éléments d'arrière-plan établis, nous sommes à présent en mesure d'avancer, suivant Stoljar[103], une définition plus précise du concept de « propriété physique » :

(1) *F* est une propriété physique si et seulement si :

(a) *F* est une des propriétés distinctives des objets physiques ; et
(b) *F* est exprimée par le truchement d'un prédicat d'une théorie physique ; et
(c) *F* est objective ; et
(d) *F* est une propriété que nous arrivons à connaître par le truchement des méthodes scientifiques ; et
(e) *F* n'est pas une des propriétés distinctives des âmes, ectoplasmes, etc.

A présent, si l'on applique cette « définition-faisceau » aux différents modèles explicatifs matérialistes de l'histoire de la philosophie, on peut recueillir des acceptions très variées, voir hétéroclites, de ce qu'une propriété physique *est*. Commençons avec l'exemple classique issu de l'atomisme démocritéen selon lequel une des principales caractéristiques des atomes est leur capacité de collision les uns avec les autres. A présent, si l'on veut pousser la définition encore plus loin et s'interroger sur ce qu'une collision avec un objet physique *est*, on peut suivre (ou pas !) le conseil de Stoljar qui nous invite à nous cogner volontairement l'orteil ou la tête (au choix) contre le premier

[103] *Ibid.*, p. 57.

meuble qu'on croise sur notre chemin. La leçon que l'on doit en tirer c'est qu'au niveau infiniment petit, la collision des atomes se produit exactement de la même manière. Le bien-fondé de cette image est rudement mis à l'épreuve plus tard, lorsque la physique newtonienne introduit le concept révolutionnaire de « gravitation » qui agit, chacun le sait, sans que les objets se cognent entre eux, comme un orteil contre un fauteuil. En clair, la force de la gravitation n'est pas intuitivement physique, comme l'est par exemple une pomme ou un caillou qui vous tombe sur la tête. En outre, plusieurs chercheurs[104] ont montré que les humains entretiennent spontanément une vision naïve et préscientifique de la causalité : un objet est tenu influencer un autre objet uniquement s'il entre en contact direct, de type orteil/fauteuil, avec lui.

Si l'on suit le fil que l'histoire de la physique déroule devant nous, on s'aperçoit que les percées théoriques de la physique moderne post-newtonienne ne font que rajouter une couche supplémentaire de difficultés à celles que l'on éprouve déjà : quels sont les objets et les propriétés postulées par la physique moderne ? Des champs, des ondes, des cordes, de l'espace-temps, de l'énergie… que des objets étranges, du moins pour le sens commun. A ce point de la discussion quelqu'un pourrait observer ironiquement que l'on a affaire, du moins en principe, à une myriade de modes de description du monde physique, le tout dépendant de la perspective théorique que l'on adopte. Dans le meilleur des cas, tout ce que l'on peut dire c'est ce que les propriétés physiques maniées par la physique moderne *ne sont pas* : elles *ne* sont *pas* des propriétés distinctives des objets physiques ordinaires que l'on rencontre dans la vie de tous les jours ou à l'échelle macroscopique. Mais à partir de la prémisse selon laquelle le physicalisme peut être interprété de façon différente, l'on ne saurait tirer validement la conclusion que le physicalisme est tout simplement faux. D'autres options s'ouvrent à nous.

Jusqu'ici nous avons développé une approche dans la définition des propriétés physiques en s'appuyant sur le modèle paradigmatique de l'*Objet Physique*, modèle baptisé par Stoljar « Object-Based Conception». Mais cette option n'épuise pas le champ des possibilités. Il nous reste, note Stoljar, le modèle de la *Théorie Physique*, ou « Theory-Based Conception ».

[104] Voir E. Spelke, (1994), 'Initial Knowledge: six suggestions', *Cognition*, 50: 443-47.

1.5.2 L'approche basée sur la Théorie Physique

D'emblée, cette version est beaucoup plus libérale que celle présentée dans la section précédente, car si les contours des objets sont relativement identifiables, en revanche, les frontières qui séparent une telle théorie physique d'une telle autre sont beaucoup plus floues, et font penser au célèbre *sfumato* de Léonard de Vinci. Selon Stoljar[105], l'approche basée sur la Théorie Physique comporte deux volets. Le premier nous dit ce qu'une *propriété* physique est :

(2) *F* est une propriété physique si et seulement si *F* est exprimée par un prédicat d'une théorie physique.

Le deuxième volet nous renseigne sur ce que le *physicalisme* est :

(3) Le physicalisme est vrai si et seulement si chaque propriété instanciée découle par nécessité d'une propriété instanciée physique basée sur une théorie physique.

A ce point de la discussion, le lecteur pourrait nous arrêter, un sourire ironique au coin des lèvres, et nous accuser de vouloir escamoter le problème au risque de compliquer le système en substituant « objet » physique avec « théorie » physique. Car à présent la question qui se pose est :

(1') Qu'est-ce qu'une *théorie physique* ?

A première vue, nous sommes condamnés à suivre une des deux voies suivantes : (i) soit on considère qu'une théorie physique est une théorie qui vise un objet d'étude déterminé, comme par exemple les objets physiques ordinaires, paradigmatiques, comme les arbres, les chaises ou les pierres. (ii) Soit on choisit comme modèle-paradigme une certaine théorie physique donnée, comme par exemple la théorie médiévale de *l'impetus*, ou la mécanique quantique actuelle. On voit d'après ce qui précède que la question que l'on est inévitablement amené à se poser est celle de savoir si cette façon de présenter les choses nous aide à avancer dans l'élucidation du problème que nous poursuivons ici. Or, si l'on regarde de près, on s'aperçoit que ces deux options sont loin d'être complètement satisfaisantes. Nous le verrons dans la section suivante, cette manière de poser le débat fait l'objet d'une difficulté connue sous le nom de « dilemme d'Hempel ».

[105] (D. Stoljar, 2010:71)

Selon Stoljar[106], il existe néanmoins une troisième voie de secours : (iii) on peut définir une théorie physique comme étant une théorie élaborée par les scientifiques au cours d'un processus explicatif *continu* visant la nature et le comportement des objets physiques familiers. L'avantage le plus important de cette troisième option tient au fait qu'une théorie ainsi définie possède intrinsèquement *une fin ouverte*. On est alors en mesure d'avancer la description d'une théorie physique comme étant tout simplement une théorie physique *vraie*, abstraction faite qu'elle soit oui ou non définitivement formulée.

(4) F est une propriété physique si et seulement si F est exprimée par un prédicat d'une théorie physique vraie.

Cette troisième option, que nous allons retenir ici, peut être interprétée à son tour sous deux aspects différents. En premier lieu, on peut comprendre (4) comme faisant référence à une théorie physique vraie *pour le monde actuel*, tel que nous l'expérimentons aujourd'hui, ici et maintenant. Nous obtiendrons ainsi une version *actualiste* (5) :

(5) F est une propriété physique si et seulement si F est exprimée par une théorie physique qui est vraie pour le monde actuel.

Mais on peut interpréter (4) comme faisant référence à une théorie physique vraie pour d'autres mondes possibles. Nous obtenons ainsi une version *possibiliste* (6) :

(6) F est une propriété physique si et seulement si F est exprimée par une théorie physique qui est vraie pour d'autres mondes possibles (pas nécessairement le monde actuel).

Ces précisions peuvent-elles nous laisser l'espoir d'avoir enfin touché le Graal et trouvé *la* définition adéquate d'une propriété physique ? Hélas, nous ne sommes toujours pas au bout de nos peines, car, comme nous allons le voir d'ici peu, d'autres failles viennent mettre à l'épreuve notre entreprise de définition du physicalisme. Pour voir cela, dit Stoljar, il est plus commode d'opérer avec une variation sur le thème de la célèbre expérience de pensée de la Terre Jumelle[107] de H. Putnam.

---

[106] (D. Stoljar, 2010: 74)

[107] Pour mémoire, voici le résumé de l'argument : Putnam nous demande d'imaginer une planète lointaine, avant l'année 1750, qui est exactement comme la Terre, excepté qu'à la place de l'eau (H_2O), il y a une substance à la composition chimique identique (XYZ). Néanmoins, les macros propriétés de cette substance sont supposées être exactement les mêmes que celles de notre eau sur Terre. Elle a le même aspect : incolore, inodore et sans saveur ; elle remplit les lacs et les océans et coule dans les rivières.

Mais avant de poursuivre sur cette voie, essayons d'énoncer de manière un peu plus précise la version actualiste du physicalisme en s'appuyant sur les outils conceptuels que l'on s'est donnés plus haut. On obtient ainsi la définition suivante:

> <LE MONDE-PHYSIQUE MODERNE>[108] : ceci est un monde possible dans lequel chaque propriété instanciée découle par nécessité des propriétés distinctives des états de choses postulées par la physique moderne. Les propriétés instanciées dans ce monde dupliquent n'importe lesquelles propriétés qui sont instanciées dans le monde actuel, dans la mesure où cela est possible.

Bien que la conception <LE MONDE-PHYSIQUE MODERNE> confirme l'hypothèse selon laquelle le physicalisme est celui promu par les théories physiques applicables au monde actuel, on peut, dit Stoljar, imaginer des situations où tel n'est pas le cas. Pour voir cela, nous avons besoin de dresser un parallèle avec la célèbre expérience de pensée de Putnam. Supposons que nous adoptions la conception scientifique actuelle qui statue comme fondamentales des propriétés comme la masse, le spin, la charge, etc. Imaginons, à la suite de Putnam, l'existence d'un monde-jumeau qui serait parfaitement identique au nôtre, à l'exception des propriétés fondamentales qui elles, quoique non-mentales, sont de nature radicalement différente de celles du monde originalement considéré. Nous obtenons ainsi la formulation suivante :

> <LE MONDE-PHYSIQUE JUMEAU>[109] : ceci est un monde-jumeau possible dans lequel chaque propriété instanciée découle par nécessité des propriétés fondamentales telles que le *spin-jumeau*, la *charge-jumelle*, la *masse-jumelle*, etc.

Cependant, sur Terre ou sur Terre Jumelle, avant 1750, personne ne peut distinguer entre l'eau (H_2O) et l'eau (XYZ). Ainsi, un individu sur Terre, Oscar, qui utilise le mot « eau » se référera à H_2O et non à XYZ. Bien sur, Oscar, avant 1750, ignore que l'eau est H_2O. Néanmoins, cela ne l'empêche pas de se référer à H_2O quand il utilise le terme « eau ». De la même façon, un individu sur Terre Jumelle, Sosie-Oscar, réplique exacte, à la molécule près d'Oscar, en utilisant le même terme d' « eau » se référera à XYZ et non à H_2O. En conséquence, puisque le même mot « eau» est utilisé de la même manière dans les deux mondes et que les organismes jumeaux sont identiques dans l'ensemble de leurs aspects physiologiques, ce que veut dit Oscar sur Terre et ce que veut dire Sosie-Oscar sur terre Jumelle, lorsque chacun prononce « l'eau est sans saveur », n'est pas identique. Ce que veut montrer cette expérience, selon Putnam, c'est que la signification des mots, ce que nous voulons dire lorsque nous les utilisons, n'est pas dans la tête, mais dépend de l'environnement de celui qui les prononce. Voir H. Putnam, (1975), 'The Meaning of 'Meaning', *Philosophical Papers, Vol. II: Mind, Language, and Reality*, Cambridge: Cambridge University Press.

[108] "THE MODERN PHYSICS WORLD: this is a possible world at which every instantiated property is necessitated by the properties distinctive of the things postulated by modern physics. The properties instantiated at this world duplicate whatever properties are instantiated at the actual world, insofar as this is possible". (D. Stoljar, 2010: 76)

[109] "THE TWIN-PHYSICS WORLD: this is a possible world or twin-earth at which every property is necessitated by twin-mass, twin-charge, and twin-spin. The properties instantiated at this world duplicate whatever properties are instantiated at the actual world, insofar as this is possible". (*Ibid.*, p. 77)

> Les propriétés instanciées dans ce monde dupliquent n'importe lesquelles des propriétés qui sont instanciées dans le monde actuel, dans la mesure où cela est possible.

On reconnaît facilement ici une variation sur la célèbre expérience de pensée de Putnam. Quelles sont les conséquences de cette analogie pour le cas que nous enquêtons ici ? L'aspect le plus frappant est que, tandis que le physicalisme est clairement vrai pour le MONDE-PHYSIQUE JUMEAU (autrement dit, il est vrai que le physicalisme-jumeau est une version du physicalisme), le physicalisme dans sa version *actualiste* n'est pas vrai pour LE MONDE-PHYSIQUE JUMEAU. La raison pour laquelle le physicalisme dans sa version actualiste n'est pas vrai pour LE MONDE-PHYSIQUE JUMEAU est que ce dernier est *défini* en termes d'une théorie physique qui est *vraie* pour le monde *actuel*. Plus spécifiquement, si les théories scientifiques valides pour le monde actuel nous disent que la masse, le spin et la charge *sont* des propriétés physiques fondamentales, alors le physicalisme dans sa version actualiste sera faux à chaque fois que la masse, le spin et la charge *ne sont pas* des propriétés physiques fondamentales. Or, tel est le cas pour LE MONDE-PHYSIQUE JUMEAU[110].

Doit-on capituler devant cette impasse ? *Prima facie*, la réponse est 'non' et cela pour une très bonne raison : jusqu'ici nous nous sommes intéressés uniquement à la version *actualiste* du physicalisme défini en fonction des théories physiques. Qu'en est-il de la version *possibiliste* ? Est-elle en mesure d'offrir une définition satisfaisante du physicalisme ? Comme le relève Stoljar, le point faible de cette deuxième acception consiste en son excessive permissivité : en effet, non seulement elle est valide pour les mondes physiques possibles que nous avons déjà brièvement évoqués (atomiste, newtonien, moderne), mais elle semble aussi permettre la manifestation des situations qui ne sont pas, du moins intuitivement, assimilables au physicalisme, comme le *dualisme*, par exemple. Considérons, dit Stoljar, le cas suivant :

> <LE MONDE DUALISTE CLASSIQUE>[111] : ceci est un monde possible qui est exactement comme le monde atomiste classique, à une différence près : lorsque les atomes constituent le corps humain, ils sont associés à une âme, de sorte que

[110] (*Ibid.*, p. 78)

[111] "The classical dualist world: this is a possible world which is exactly like the classical atomist world, but with this modification: when atoms come together to form human bodies, such bodies are yoked together with a soul in such a way that the behavior of the body is explained only on the assumption that it is influenced by soul and its distinctive properties. The properties instantiated at this world duplicate whatever properties are instantiated at the actual world, insofar as this is possible." (*Ibid.*, p. 80)

le comportement du corps soit expliqué exclusivement en termes de l'influence que l'âme et ses propriétés distinctives exercent sur lui. Les propriétés instanciées dans ce monde dupliquent n'importe quelles propriétés instanciées dans le monde actuel, dans la mesure où cela est possible.

Si l'on regarde de près ce dernier cas et y applique la version possibiliste du physicalisme, on se confronte à la difficulté suivante : alors qu'il est clair que le physicalisme est faux pour le <MONDE DUALISTE CLASSIQUE>, on s'aperçoit que le physicalisme dans sa version possibiliste pourrait être vrai pour <LE MONDE DUALISTE CLASSIQUE>. Si une théorie physique est une théorie qui explique le comportement et la constitution des objets physiques ordinaires (communs), le dualiste classique peut avancer sur le même terrain et dire que la théorie en question fait référence à des entités comme les âmes, et d'autres propriétés mentales irréductibles. Mais dans ce cas, ces propriétés mentales irréductibles vont être considérées comme physiques, *parce qu*'elles sont exprimées par une théorie physique vraie pour le monde respectif. Bref, nous tournons en rond.

Récapitulons. Jusqu'ici, nous avons tenté, au risque de l'aridité, d'explorer les multiples facettes de cette mission de définir le physicalisme en s'appuyant sur la notion de *théorie physique*. Parvenus au terme de notre esquisse, on s'aperçoit que non seulement nous ne sommes pas arrivés à un consensus, mais en plus, nous débouchons sur une impasse qui prend la forme d'une variation sur le thème du bien connu dilemme d'Hempel. Cette variation, selon Stoljar, peut être exprimée sous la forme d'une métathèse stipulant qu'il n'existe aucune définition ou forme du physicalisme qui soit à la fois vraie et qui mériterait son nom. L'argument qui étaye cette métathèse, et que nous avons déjà évoqué par petits fragments depuis le début de ce chapitre, peut être assemblé et soudé dans une expression plus rigoureuse qui prend la forme de l'inférence suivante :

(F)1 En formulant le physicalisme, on doit opérer (i) soit avec une conception basée sur l'*objet* physique, (ii) soit avec une conception plus libérale, basée sur une *théorie* physique.

(F)2 D'une part, si l'on opère avec (i), il est effectivement possible d'articuler une version du physicalisme qui mérite ce nom, mais cette version s'avère fausse.

(F)3 D'autre part, si l'on opère avec (ii), il est possible d'articuler une version du physicalisme qui est vraie, mais qui ne mérite pas son nom de 'physicalisme' pour une des deux raisons suivantes :

(a) soit il est *vrai* pour des mondes possibles où aucune version du physicalisme ne devrait être vraie (version actualiste) ;

(b) soit il est *faux* pour des mondes possibles où aucune version du physicalisme ne devrait être fausse (version possibiliste).

Mais avant de poursuivre plus loin, présentons brièvement la structure du dilemme d'Hempel. S'agissant d'un sujet classique dans la littérature analytique, nous n'entendons pas en fournir ici une description exhaustive, mais seulement une ébauche nous permettant de mieux articuler par la suite le problème spécifique que pose l'argument de la connaissance de Jackson pour les tenants du physicalisme.

1.5.3 Le dilemme d'Hempel

Il importe ici, afin de mieux saisir la portée de l'argument connu sous le nom de « dilemme d'Hempel »[112], d'introduire une distinction supplémentaire en plus de celle entre version actualiste ou possibiliste du physicalisme. Si l'on tient compte du fait que la physique s'est construite comme une discipline et une activité qui dure depuis un certain temps, on peut alors interpréter la thèse (4) – « *F* est une propriété physique si et seulement si *F* est exprimée par le prédicat d'une théorie physique *vraie* » – soit en faisant référence à une théorie que nous pensons vraie *actuellement*, soit en faisant référence à une théorie *idéale*, indépendamment du fait qu'elle soit ou non identique à celle que les scientifiques créditent aujourd'hui. La première version – que Stoljar appelle « *le point de vue de la théorie actuelle* » – peut alors être exprimée ainsi :

(1) *F* est une propriété physique si et seulement si *F* est exprimée par un prédicat d'une théorie physique *vraie* au vu des développements théoriques dont on dispose *à l'heure actuelle*.

Semblablement, la deuxième interprétation, « *le point de vue de la théorie idéale* », peut être formulée ainsi :

[112] C. Hempel, (1980), 'Comments on Goodman's *Ways of Worldmaking*', *Synthese*, 45: 139–199.

(2) *F* est une propriété physique si et seulement si *F* est exprimée par un prédicat d'une théorie physique *vraie* à une limite *idéale*.

On voit, d'après ce qui précède que la contrainte de définir le physicalisme n'est pas résolue, mais seulement déplacée, car à présent la question qui s'impose est : qu'est-ce qu'une *théorie idéale* plus précisément ? Si l'on projette l'idéal dans le futur, que dire alors s'il n'y a pas de futur tout simplement ? En outre, il n'y a aucune garantie de type *a priori* pour nous assurer que l'avenir nous approchera de la limite idéale de la vérité. Il pourrait, au contraire, nous enfoncer davantage dans l'ignorance et l'obscurantisme.

Une meilleure façon de contourner le problème serait d'établir tout simplement un signe d'identité entre *théorie idéale* et *théorie vraie*. Après tout quelle meilleure caractérisation d'une théorie idéale que celle offerte par la vérité ? La formulation originale du dilemme ainsi formulé vise avant tout le physicalisme de « première génération » qu'on appelle communément « physicalisme sémantique ». Le physicalisme, observe Hempel[113], présuppose l'adoption d'un langage capable de formuler une théorie vraie englobant l'explication de *tous* les phénomènes. Mais quel langage utiliser plus précisément ? D'une part, il est évident qu'un tel langage privilégié ne peut pas être celui de la physique du XVIIe siècle, disons, car il emploie des notions désuètes comme « fluide calorique ». D'autre part, si l'on choisit de se tourner vers la physique actuelle, valeur plus « sûre », on s'aperçoit que rien ne la met à l'abri des possibles révisions futures. Quel point de repère choisir alors, pour définir le fameux langage physique ?

Transporté sur le terrain du physicalisme actuel, le dilemme d'Hempel a été reformulé par G. Hellman[114] ainsi : rien ne semble plus évident que la physique actuelle est incomplète à la fois dans son ontologie et dans ses tentatives de répertorier les lois de la nature. A partir de là, (i) soit les principes physicalistes *s'appuient* sur l'appareillage explicatif de la physique actuelle, mais dans ce cas nous avons toutes les bonnes raisons de penser que cet appareillage est ou bien incomplet, ou bien tout simplement faux. (ii) Soit les principes physicalistes *ne* s'appuient *pas* sur l'appareillage explicatif élaboré par la physique actuelle, mais dans ce cas on s'aperçoit que le physicalisme s'appuie sur *un fondement virtuel qui n'existe pas encore*.

[113] (C. Hempel, 1980: 194-5)
[114] C. Hellman, (1985), 'Determination and Logical Truth', *The Journal of Philosophy*, 82(11): 607–616.

Comment justifier alors l'élection de tel ou tel critère pour définir les notions clé de « propriété », « objet » ou « loi » physique[115] ?

Comme le relève Stoljar, il est commode de représenter la structure argumentative de ce dilemme sous la forme d'un syllogisme :

DILEMME D'HEMPEL

(P1) Si par définition, les propriétés physiques sont celles exprimées par les prédicats d'une théorie physique courante, alors le physicalisme est faux.

(P2) Si, par définition, les propriétés physiques sont celles exprimées par les prédicats d'une théorie physique idéale, alors nous ne savons pas ce que le physicalisme veut dire.

(P3) Les propriétés physiques sont exprimées soit par les prédicats d'une théorie physique courante, soit par les prédicats d'une théorie physique idéale.

Par conséquent,

(C) Le physicalisme est soit faux, soit insensé.

L'armature formelle de l'argument ainsi formulé est clairement valide. La seule manière de l'attaquer serait de remettre en question le contenu de ses prémisses. En ce qui concerne la première prémisse (P1), il convient de noter qu'elle n'affirme pas que les théories physiques actuelles *se trompent* dans leur inventaire répertoriant les propriétés physiques. Tout ce qu'elle affirme, c'est que cet inventaire est loin d'être complet. La thèse derrière P1 est celle de l'incomplétude des théories physiques actuelles. Si les propriétés physiques sont définies à l'aide de l'appareillage conceptuel des ces dernières théories, alors le physicalisme est faux. Il n'y a rien à ajouter à propos de ce volet de l'argument.

La deuxième prémisse (P2) peut être retranscrite, selon Stoljar[116], sous la forme d'une liste potentiellement infinie de conditionnels ayant la forme suivante :

[115] "(…) current physics is surely incomplete (even in its ontology) as well as inaccurate (in its laws). This poses a dilemma: either physicalist principles are based on current physics, in which case there is every reason to think they are false; or else they are not, in which case it is, at best, difficult to interpret them, since they are based on a 'physics' that does not exist – yet we lack any general criterion of 'physical object, property, or law' framed independently of physical theory". (C. Hellman, 1985: 609)

[116] (D. Stoljar, 2010: 100)

(C)1 Si les propriétés exprimées par une physique idéale sont les propriétés distinctives des *atomes classiques*, démocritéens, alors le physicalisme est la doctrine qui affirme que tout découle par nécessité des propriétés distinctives des atomes classiques.

(C)2 Si les propriétés exprimées par une physique idéale sont les propriétés distinctives des *fonctions d'onde quantiques*, alors le physicalisme est la doctrine qui affirme que tout découle par nécessité des propriétés distinctives des fonctions d'onde quantiques.

(C)3 Si les propriétés exprimées par une physique idéale sont les propriétés distinctives des *monades leibniziennes*, alors le physicalisme est la doctrine qui affirme que tout découle par nécessité des propriétés distinctives des monades leibniziennes.

(C)4 Si les propriétés exprimées par une physique idéale sont les propriétés distinctives des *citrouilles*, alors le physicalisme est la doctrine qui affirme que tout découle par nécessité des propriétés distinctives des citrouilles

(C)5 … et ainsi de suite *ad infinitum*.

Nous savons qu'au moins un des énoncés conditionnels membres de la liste entamée ci-dessus est le *vrai* candidat pour la définition du physicalisme (en s'appuyant sur la notion de théorie physique idéale). En clair, la raison qui est derrière P2 est que nous sommes dans l'impossibilité de préciser ce qu'une physique idéale est. A partir de là, si les propriétés physiques, par définition, *sont* celles exprimées par les prédicats d'une théorie physique idéale, il s'ensuit, bien évidemment, que nous ne savons pas ce que le physicalisme veut dire.

Selon Stoljar, il existe une troisième manière d'éviter les cornes du dilemme, et qui consiste à regarder de plus près la troisième prémisse (P3) qui est généralement admise sans une enquête plus approfondie. La raison pour laquelle on se dispense d'en trouver une justification est liée, note Stoljar, au fait que (P3) est facilement assimilable à une vérité logique. Or il s'agit là d'une déplorable confusion, car la troisième prémisse qui étaye le dilemme d'Hempel *n*'est *pas* une vérité logique. Pour voir cela, il suffit d'observer que (P3) stipule que les propriétés physiques doivent être définies en termes d'une des deux théories physiques spécifiées. Même si cette contrainte est vraie, cette véracité ne lui confère en rien le statut de vérité logique. Après tout, pourquoi les

propriétés physiques devraient-elles être absolument définissables en termes de *théories*[117] physiques et non pas d'*objets*[118] physiques ? En outre, même s'il s'avère que les propriétés physiques soient effectivement définissables en termes de théories physiques, pourquoi en choisir *une seule* en particulier ?

1.5.4 Quelques conclusions provisoires

On voit d'après ce qui précède qu'en abordant la question « qu'est-ce que le physicalisme ? » ou plus précisément, « qu'est-ce qu'une propriété/objet physique ? » on est loin d'obtenir un consensus absolu. A ce stade de notre enquête, le lecteur, aussi patient et courageux soit-il, sera très certainement tenté de baisser les bras en s'exclamant : à quoi bon continuer à suivre le duel entre l'argument de la connaissance et le physicalisme si l'on ne sait pas très bien de quoi on parle ? Continuer sur cette voie équivaudrait à s'obstiner à se battre, comme Don Quichotte, contre les moulins à vent. A vrai dire, pourrait-on lui en vouloir, à notre patient lecteur, de songer à quitter un navire à la dérive ?

Il y a deux manières de reprendre courage et de continuer l'aventure: une option « facile », et une autre, plus « risquée», mais davantage plausible. La plus facile consiste à rappeler que si l'on exigeait une définition aussi rigoureuse de tous les concepts qui inspirent l'aventure philosophique depuis des siècles, on serait vite amené à jeter l'éponge et se consoler avec une carrière dans la finance, le droit, et pourquoi pas, dans la politique. La deuxième réponse, plus modeste, mais plus efficace à notre sens, consiste à rappeler que les délimitations définitionnelles chirurgicales restent l'apanage de certaines disciplines privilégiées, comme la logique ou les mathématiques. La question est : avons-nous véritablement besoin de cette précision quasi-absolue ? Comme toute affaire philosophique importante, la question du physicalisme est

[117] "(…) a physical property is a property which either is the sort of the sort of property that physical theory tells us about or else is a property which metaphysically (or logically) supervenes on the sort of property that physical theory tells us about. According to this conception, for example, if physical theory tells us about the property of having mass, then having mass is a physical property." (D. Stoljar, 2001)

[118] "A physical property is a property either is the sort of property required by a complete account of the intrinsic nature of paradigmatic physical objects and their constituents or else is a property which metaphysically (or logically) supervenes on the sort of property required by a complete account of the intrinsic nature of paradigmatic physical objects and their constituents" […] According to this conception, for example, if rocks, trees, planets and so on are paradigmatic physical objects, then the property of being a rock, tree or planet is a physical property". (D. Stoljar, 2001/2004: 312)

entourée d'une sorte de « halo conceptuel » qui, quoique gravitant autour d'un noyau compact stable, possède une souplesse et une perméabilité de lisière impressionnante, lui permettant ainsi de communiquer, voir de fusionner avec d'autres problématiques philosophiques, comme par exemple avec la question métaphysique des propriétés fondamentales, ou la question ontologique des niveaux de réalité. L'important pour nous ici c'est d'avoir identifié, au milieu de cette nébuleuse, un sillon, parmi tant d'autres possibles, dans la caractérisation du physicalisme. C'est uniquement au terme de notre enquête, et au fur et à mesure que nous déploierons l'intégralité du projet que trament les différents fils de notre argumentation, que nous aurons la confirmation ou l'infirmation du fait que nous avions, ou pas, suivi le bon fil depuis le début, afin de démêler le nœud gordien que forment les *qualia* au cœur du tissu physicaliste.

Mais avant d'aller plus loin dans notre enquête, un bref retour rétrospectif s'impose. Nous venons de voir que le tableau standard du physicalisme contemporain le place dans la compétition au titre de 'Weltanschauung' de l'époque moderne. Le physicalisme affiche ainsi des ambitions métaphysiques manifestes. Or, supposons qu'un philosophe comme Neurath ou Carnap, ressuscités par un lutin farceur, lisent les définitions du physicalisme que nous venons de passer en revue. Qu'en auraient-ils conclu ? Que le bateau des philosophes est à nouveau à la dérive, sur la mer capricieuse des questions métaphysiques insensées. Rappelons-nous que la question de la réduction métaphysique du psychique au physique n'était pas au cœur de l'empirisme logique du siècle dernier. Ce fait ne devrait surprendre personne, car l'analyse néopositiviste centrée sur l'examen des propriétés formelles et sémantiques générales du discours indiquait clairement que la métaphysique ne pouvait en aucun cas se prononcer sur les composantes du réel, et qu'elle devait être rejetée comme un discours insensé lorsqu'elle prétendait le faire.

Restées orphelines, si l'on peut dire ainsi, de leur latitude ontologique, les questions métaphysiques se refugièrent dès lors au niveau de l'analyse logique du langage de la science. C'est la raison pour laquelle l'ambition des « premiers » physicalistes n'a pas été celle de « physicaliser » le réel, mais de « physicaliser » les *énoncés* portant sur le réel. Pour les positivistes logiques, la philosophie ne pouvait être qu'*a priori*, et sa seule mission l'« analyse », d'où le qualificatif de « philosophie *analytique* ». La philosophie quant à elle, tout en laissant carte blanche aux professionnels de la science dont la mission était d'expliquer ce que les choses sont, se

réservait la part la plus noble du gâteau, « noble » car pure et abstraite, à savoir, la partie formelle, la logique. Dans le chapitre suivant, nous allons suivre le fil qui relie le physicalisme « antimétaphysique » des empiristes logiques au physicalisme actuel tel que nous venons de le présenter jusqu'ici.

Le fil conducteur de notre enquête est la question de l'expérience phénoménale dans une ontologie physicaliste, mais le souci qui l'anime n'est pas d'ordre historique. Néanmoins, afin de mieux comprendre les interrogations qui ont marqué la physionomie des débats actuels autour de cette question, nous ne pourrons pas faire l'économie d'un bref survol sur les ambitions initiales du physicalisme, survol qui nous aidera à saisir la manière dont le sujet psychologique a été « évincé » au cours de l'affaire des énoncés protocolaires, pour revenir ensuite sur la scène philosophique avec la force d'un véritable boomerang.

Il est généralement admis que la transformation du physicalisme « formel » ou « sémantique » en un physicalisme métaphysique a été élaborée *via* le programme quinien[119] de *naturalisation de l'épistémologie*. Le physicalisme visait donc dans un premier temps un projet de traduction de toutes les propositions, en partant des énoncés protocolaires et jusqu'aux systèmes théoriques, dans la langue de la physique. Or, situés à mi-chemin entre le formalisme d'un langage pur et l'empiricité aléatoire du monde, les énoncés protocolaires se voient octroyer le rôle "ingrat" d'intermédiaire hybride chevauchant deux niveaux distincts : le domaine de la pensée et celui du monde. C'est l'examen de ce 'chevauchement' entre le mode *contentuel* de discours et le mode *formel*, qui retiendra notre attention dans le chapitre suivant. Cette «physicalisation facile des états d'esprit »[120], pour reprendre l'expression de Quine, nous servira de prétexte à l'introduction du terme de « saut conceptuel » avec lequel nous allons opérer tout au long de notre enquête. Trois objectifs seront à accomplir dans le chapitre suivant :

---

[119] W.V. Quine, (1969), "Epistemology Naturalized", in Quine, *Ontological Relativity and Other Essays*, New York: Columbia University Press, pp. 69-90; trad. fr. A. Fagot-Largeault, in *W*.O. Quine, *Relativité de l'ontologie* et *autres essais*, (1977), Paris : Aubier Montaigne repris in S. Laugier et P. Wagner, (eds.) (2004), *Philosophie des Sciences - Naturalismes* et *réalismes* (vol. II). Paris, Vrin. (Nous utilisons ici cette dernière édition).

[120] W. V. Quine, (1957b), 'The Scope and Language of Science' in *British Journal for the Philosophy of Science*; trad. fr. par P. Jacob, 'Le domaine et le langage de la science', in P. Jacob, (ed.). (1980). *De Vienne à Cambridge : l'héritage du positivisme logique*. Paris: Gallimard, p. 237. (Nous faisons référence ici à l'édition française.)

(i) En premier lieu, il s'agit de trouver une réponse à la question suivante : quelle est la raison pour laquelle on refuse d'accorder à la psychologie le même statut qu'aux sciences physiques, et s'il peut y avoir une psychologie sans âme, comme le voulait Brentano[121], peut-il y avoir une philosophie sans rapport aux états mentaux ?

(ii) L'examen du rejet des thèmes psychologisants nous permettra de détecter non seulement les traces d'une éventuelle filiation reliant le physicalisme « sémantique » aux versions «métaphysiques » du physicalisme actuel, mais aussi de saisir en quoi les deux doctrines partagent les mêmes ambitions d'arrière-plan, qui se révéleront, comme nous espérons le montrer par la suite, fort significatives pour le cas des *qualia*.

(iii) Enfin, l'aspect le plus important c'est d'arriver à circonscrire avec plus de précision la pratique des sauts conceptuels au sein des stratégies explicatives clivées sur l'axe archétypal subjectivité/objectivité.

Le chapitre suivant marquera une parenthèse mettant en scène une brève rétrospective autour des ambitions initiales du physicalisme de « première génération », et comporte à son tour quatre volets : les sections 2.1. et 2.2. découpent un périmètre d'enquête bien précis, à savoir l'étude des conséquences de l'antipsychologisme frégéen sur le 'mind-body problem'. Cette incursion nous permettra d'établir les premiers éléments dans la définition de la notion de « saut conceptuel » avec laquelle nous allons opérer tout au long de notre investigation. Enfin, les sections de 2.3. à 2.4. s'organiseront autour d'un deuxième exemple de « saut conceptuel » repérable au niveau de l'articulation entre les contenus d'expérience vécue dans la distinction carnapienne entre le *mode formel* et le *mode contentuel* du discours, dans l'affaire des énoncés protocolaires. Nous verrons ainsi que malgré la tombée en désuétude du physicalisme « sémantique » de « première génération », ses ambitions glissent, de manière tacite, vers l'ossature argumentative du physicalisme contemporain. Par conséquent, nous argumenterons que bien des controverses qui agitent la philosophie de l'esprit contemporaine, portent sur leurs épaules, bien plus que leurs acteurs ne veulent l'admettre, le fardeau des ambitions du physicalisme sémantique. Nous verrons ainsi

[121] F. Brentano, (1874), *Psychologie vom empirischen Standpunkt*, Leipzig. Trad. fr. M. de Gandillac, *Psychologie du point de vue empirique*, Paris : Vrin, pp. 24-25. Nous faisons référence ici à l'édition française.

que suite à l'échec du projet de « reconstruction rationnelle » du réel mené dans l'esprit du Cercle de Vienne, les philosophes analytiques seront de plus en plus nombreux à suivre les conseils de Quine les invitant à se focaliser, plus modestement, sur l'analyse de ce phénomène naturel qu'est le « sujet humain physique »[122]. En réunissant, dans la même phrase, les termes clé '*nature*', '*sujet*', et '*physique*', Quine délivre le certificat de naissance d'une philosophie du mental menée dans l'esprit d'une enquête psychologique – c'est-à-dire épistémologique, si par cette dernière on entend une science *naturelle*.

Ce détour rétrospectif vers les origines du physicalisme actuel nous permettra d'approcher un aspect essentiel, à notre sens, des controverses qui entourent la question de l'explication matérialiste non-réductive de la conscience expérientielle et de sa corrélation avec les phénomènes neurophysiologiques. Ce chapitre servira ainsi de tremplin pour nous aider à survoler les différents modèles explicatifs de l'esprit, étape indispensable à l'amorce de l'analyse proprement dite de l'argument de la connaissance.

[122](W.V. Quine, 1958/2004: 51)

2 Du sujet logique au sujet psychologique

A l'origine, le physicalisme se définit avant tout comme un 'antipsychologisme'. La plupart des historiens de la philosophie du XXe siècle placent le rejet des thèmes dits "psychologisants" à l'origine de ce mouvement de pensée responsable des transformations conceptuelles qui ont engendré, paradoxalement, l'essor de l'étude de l'esprit. Le rejet du sujet psychologique, soupçonné d'être, en quelque sorte, un cheval de Troie susceptible de laisser pénétrer les démons du psychologisme dans l'édifice objectif d'une conception scientifique du monde, est étroitement lié à l'aversion manifeste des philosophes analytiques envers les systèmes métaphysiques postkantiens.

Il sera argumenté ici qu'une bonne partie des raisons derrière de ce rejet doivent être cherchées dans l'expulsion des contenus d'expérience vécue chez les premiers analytiques comme Frege et Carnap. D'aucuns expliquent le bannissement de la psychologie hors des frontières respectables des sciences physiques par le fait que les sciences humaines en général ont produit très peu de résultats concrets, leurs prédictions étant loin d'être aussi fécondes sur le plan empirique que leurs consœurs, les lois physiques. Mais cette critique ne suffit pas à elle seule, à enlever à la psychologie l'autorité de légiférer sur ce que les états psychologiques sont, car une analyse attentive montre que beaucoup de lois psychologiques ont bel et bien été confirmées. Ce qui nous amène à suggérer, à la suite de Crane et Mellor[1], que les auteurs qui tiennent la psychologie pour épistémiquement suspecte, le font en vertu du fait que son objet d'investigation *n*'est *pas* physique, et non pas *à l'inverse*. Autrement dit, l'exclusion du mental du domaine des sciences physiques se fait sur d'autres bases que la prétendue infériorité prédictive de la psychologie face à la physique. Bref, les frontières de la respectabilité du domaine physique sont imposées *depuis l'extérieur*. Mais, comme nous le verrons par la suite, cette expulsion du mental hors du domaine privilégié du physique devra être solidement justifiée sur des bases indépendantes, et non pas être avancée comme un dogme brut.

Une autre option pour justifier l'exclusion de la psychologie serait de rappeler que non seulement les diverses branches de la physique, mais aussi les autres sciences naturelles, comme la chimie et la biologie sont en principe *réductibles* à la physique, tandis que la psychologie s'obstine, quant à elle, à garder un pied en dehors du cercle

[1] (T. Crane & D.H. Mellor, 1990/1995)

tracé par le vocabulaire et le champ d'investigation de la physique. Comme le souligne fort bien J. Heil[2], la réussite de la stratégie réductive classique a été renforcée par l'application d'une technique de type *dividere et imperare*, dont la très efficace « division du travail » a eu pour conséquence l'épanouissement de nombreuses branches spécialisées. Enfin, le positionnement de la psychologie en-dessous du seuil minimal fixé par les sciences physiques repose sur une interprétation distinctive de la notion de « causation » qui, à première vue, fait partie de l'apanage exclusif de la physique. Autrement dit, il s'agit de miser sur la force du principe de clôture causale : le monde contient uniquement des propriétés, entités, relations et faits qui ont des causes[3] et des effets physiques.

Ce chapitre peut être vu comme une mise en perspective proposant un survol rapide des thèses antipsychologistes ou antimentalistes du « premier » physicalisme. Par certains aspects, cette parenthèse peut sembler incongrue, voire même encombrante dans une recherche comme la nôtre, explicitement dédiée au problème particulier que pose l'expérience phénoménale dans une description purement physicaliste de l'esprit. Or, nous verrons qu'il est possible de discerner, dès l'affaire des propositions protocolaires, et dans sa forme embryonnaire, la silhouette des futures controverses qui s'organiseront par la suite autour de l'affrontement entre les partisans du physicalisme et les défenseurs de l'argument de la connaissance.

---

[2] J. Heil, (1998), *Philosophy of Mind.* Routledge.

[3] Cette formulation débouche, bien évidemment, sur le célèbre problème des rapports causaux entre le corps et l'esprit, ainsi que sur les interrogations afférentes concernant la manière dont le physicalisme entend traiter cette relation intuitive de cause à effet entre les faits mentaux et les faits physiques. Nous touchons ici à un sujet dont la complexité ne nous permet pas de l'aborder en détail dans le présent essai. Ce qu'on appelle, dans la littérature analytique contemporaine, *le problème de la causalité mentale* (« mental causation ») fait actuellement l'objet d'une vaste analyse dont les enjeux considérables mériteraient un traitement minutieux dans un ouvrage à part. Pour l'objectif que nous poursuivons ici il suffit de souligner qu'en reprenant le flambeau de l'ancienne « philosophie première », la physique se voit assigner une mission des plus complexes : assurer le ciment épistémique nécessaire à l'incorporation des techniques de type *dividere et imperare* qui se placent, paradoxalement, à l'extrême opposée de l'idéal régulateur d'unité alimentant souterrainement ses ambitions métaphysiques. Par conséquent, le véritable défi à relever sera celui de relier l'*Un* épistémique des lois universelles qui fondent le modèle logico-déductif de la connaissance au *Multiple* ontologique des champs de recherches des sciences particulières. Nous avons vu que la solution trouvée fut celle des *niveaux hiérarchiques* : le monde s'organise sur plusieurs niveaux de réalité. Pour une très bonne introduction dans le thème de la causalité mentale voir l'anthologie d'articles parue sous la direction de J. Heil & A. Mele (1993/2003), *Mental Causation*, op. cit.

2.1 L'antipsychologisme frégéen : quelques conséquences

Lorsque G. Frege écrit en 1918 que « ni la logique, ni les mathématiques n'ont pour tâche d'étudier les âmes ou les contenus de conscience dont l'homme individuel est le porteur », mais que leur tâche est « l'étude de l'esprit : de ***l***'esprit, non des esprits »[4], l'auteur des *Fondements de l'Arithmétique* exprime ouvertement une thèse d'inspiration platonicienne. Depuis *Menon*, le statut particulier des entités logico-mathématiques a fait rêver plus d'un lointain disciple de Platon. C'est de ce côté – pensent-ils – qu'il faut rechercher le pilier stable pour contrecarrer les sables mouvants des présupposés subjectifs. On l'a souvent répété : le XXe siècle aura été dans les pays de langue anglaise étonnamment fécond pour le courant de pensée analytique. Cependant, et bien que le sujet qui nous concerne représente un phénomène précis de la philosophie de l'esprit – à savoir l'impact de l'argument de la connaissance sur les ambitions de complétude du physicalisme – il convient d'abord de commencer par fixer la toile d'arrière-plan sur laquelle se déroulera par la suite l'intrigue philosophique que nous investiguons ici. La présente section n'ayant pas pour ambition d'illustrer d'une manière exhaustive les différentes étapes qui ont marqué les temps forts de cette recherche des « fondements stables », nous restreindrons notre enquête à un problème plus localisé. Après un bref passage en revue des difficultés qu'entraîne l'adoption pure et dure du « slogan antipsychologiste » chez Frege, nous bifurquerons ensuite vers l'exploration d'un exemple de ce que nous allons appeler ici un « *saut conceptuel* » décelable au cœur même de la mythologie[5] frégéenne des trois mondes, et cela dans le but de mieux saisir les mécanismes qui ont orienté le regard des philosophes depuis le sujet *logique*, considéré dans sa dimension immatérielle, platonicienne, vers le sujet *psychologique*, considéré dans sa dimension matérielle, ancrée dans le monde naturel.

Bien que la question des rapports entre philosophie et psychologie ait connu une longue et complexe évolution, se déroulant sur plusieurs siècles et impliquant les figures les plus marquantes de la pensée occidentale, et bien qu'elles soient indissociables jusqu'à un certain point de leur histoire commune, vers la fin du XIXe siècle ces deux domaines se retrouvent contraints d'opérer une démarcation

[4] G. Frege, (1918), *Der Gedanke. Eine logische Untersuchungarticle*, trad. fr. C. Imbert, *Écrits logiques* et *philosophiques*, Paris, Ed. du Seuil, 1971 p. 191.
[5] M. Dummett, (1988), *Ursprünge des analytischen Philosophie.* Frankfurt am Main: Suhrkamp Verlag, trad. fr par M.-A. Lescourret, *Les origines de la philosophie analytique*, Paris, Gallimard, 1991. (Nous utilisons ici l'édition française)

méthodologique plus incisive de leurs outils et champs d'investigation respectifs. Cette manœuvre entraînera, du moins provisoirement, un éloignement, voire même un « refroidissement » de leurs rapports. Ce thème ayant été approché plusieurs fois et de façons différentes, il est préférable de commencer par brosser une esquisse synthétisant le cadre général de l'analyse de l'esprit dans les écrits analytiques de « première génération ».

De manière générale, trois phases semblent accompagner et caractériser cette entreprise. (i) Dans un premier temps, on enregistre une vive préoccupation liée au soin de dissocier aussi rigoureusement que possible la science objective de la logique de celle « glissante » et subjective de la psychologie ; cela se traduit concrètement par un net rejet des thèmes psychologisants, considérés comme « perturbateurs d'objectivité ». (ii) Une deuxième étape consigne la priorité, dans l'ordre de l'explication, du langage sur la pensée, le langage devenant ainsi l'axe central du tournant analytique. Parallèlement, l'adoption du modèle physicaliste propulse le langage de la physique au rang de paradigme garantissant l'unité des sciences. (iii) Enfin, une troisième étape critique due aux objections soulevées simultanément par Quine[6] et N. Goodman[7] signent le retour au sujet psychologique *via* les programmes de naturalisation de l'esprit, contribuant ainsi à l'effondrement du physicalisme sémantique. Commençons donc avec la première étape, à savoir le programme ouvertement antipsychologiste de Frege.

Mais d'abord, qu'est-ce que le psychologisme ? Le ménage atypique que forme le couple logique/psychologie a toujours été difficile. Mais on peut dire qu'une très sérieuse dispute éclata lors du célèbre *Psychologismus –Streit* vers la fin du XIXe siècle et début du XXe. La pomme de la discorde concernait la question de savoir si la logique et l'épistémologie sont ou non des parts de la psychologie. Les plus célèbres protagonistes étaient bien évidemment Frege et Husserl[8].

---

[6] (W.V. Quine, 1969)

[7] N. Goodman, (1951), *The Structure of Appearance.* Cambridge Massachusetts: Harvard University Press. Voir aussi N. Goodman, (1955), *Fact, Fiction, and Forecast,* Cambridge Massachusetts: Harvard University Press.

[8] Nous laisserons ici de côté la discussion de l'antipsychologisme chez Husserl. Notons seulement, en guise de résumé, les trois conséquences du psychologisme : (i) si les lois logiques étaient basées sur des lois psychologiques, alors la logique aurait été aussi vague que les lois psychologiques (§ 21). (ii) Si les lois logiques étaient des lois psychologiques, alors elles n'auraient pas pu être connues *a priori* (§21). (iii) Si les lois logiques auraient été psychologiques, alors elles auraient du référer à des entités psychologiques (§23). Voir E. Husserl (1900-1901), *Logische Untersuchungen* : *Prolegomena zur reinen*

Mais avant de présenter la position antipsychologiste de Frege, donnons un exemple[9] d'argument psychologiste[10].

ARGUMENT PSYCHOLOGISTE

(P4) La psychologie est la science qui étudie tous (types de) lois de la pensée.

(P5) La logique est un domaine d'analyse qui étudie un sous-ensemble de l'ensemble des lois de la pensée.

Par conséquent,

(C) La logique est une part de la psychologie.

Les *Fondements de l'Arithmétique*[11] font partie de cette catégorie de livres qui semblent s'inscrire *naturellement*, pour ainsi dire, dans la lignée des grandes œuvres marquant l'évolution de la pensée moderne. Les trois principes que Frege énonce au début de son ouvrage – à savoir (i) il faut nettement séparer le psychologique du logique, le subjectif de l'objectif ; (ii) on doit rechercher ce que les mots veulent dire non pas isolément, mais pris dans leur contexte ; (iii) il ne faut jamais perdre de vue la différence entre concept et objet – continuent de jalonner la pensée philosophique. Comme le résume fort bien P. Jacob[12], l'univers de Frege se scinde ontologiquement en deux classes bien distinctes : les concepts et les objets. La relation référentielle qui s'établit entre le langage et la réalité extralinguistique est médiatisée par les propositions et le sens de leurs constituants. En outre, bien maîtriser l'usage d'une langue consiste à maîtriser la relation référentielle entre celle-ci et la réalité. Autrement dit, le contenu d'une proposition affirmative prise comme tout *est* une pensée. Afin de dissiper les éventuelles confusions, Frege insiste sur la distinction subjectif/objectif : « J'entends par la pensée – écrit Frege – non pas l'acte subjectif de penser, mais son contenu objectif, lequel peut être la propriété commune de plusieurs sujets »[13].

Logik, Niemeyer. Trad. fr. H. Élie, A. L. Kelkel & R. Schérer (1974) *Recherches logiques*, t. 3, Paris : P.U.F.

[9] Voir M. Kusch, *Psychologism*, http://plato.stanford.edu/entries/psychologism

[10] Selon M. Kusch, on retrouve la trame de cet argument chez T. Lipps, (1893), *Grundzüge der Logik*. Hamburg and Leipzig: Verlag von Leopold Voss. Voir M. Kusch (1995), *Psychologism: A Case Study in the Sociology of Philosophical Knowledge*. London: Routledge.

[11] G. Frege (1884), *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl, Breslau: W. Koebner*. Trad. fr. Cl. Imbert, *Les fondements de l'arithmétique*, Paris, Seuil, 1969. Trad. angl. J.L. Austin, *The Foundations of Arithmetic*, Oxford Basil Blackwell, 1959.

[12] (P. Jacob, 1980: 52)

[13] G. Frege (1892a), 'Über Sinn und Bedeutung', in *Zeitschrift für Philosophie und philosophische Kritik*, 100: 25-50. Trad. fr. par C. Imbert 'Sens et dénotation' in G. Frege, *Ecrits logiques et philosophiques.* Paris: Ed. du Seuil, (1971), p. 108. (Nous faisons référence ici à cette dernière édition)

Frege appuie particulièrement la distinction contenu de conscience/pensée : « Bien que la pensée n'appartienne pas au contenu de la conscience de l'homme qui pense, il faut bien que quelque chose dans sa conscience vise cette pensée. Quelque chose qui ne doit pas être confondu avec la pensée elle-même (…). La pensée n'appartient ni au monde intérieur en tant qu'elle serait ma représentation, ni au monde extérieur, le monde des choses perçues par les sens »[14]. Selon la synthèse qu'en propose P. Engel[15], on peut comptabiliser trois voies argumentatives distinctes que Frege utilise contre le psychologisme. En premier lieu, il y a ce que l'on peut appeler *l'argument de la variabilité* : les jugements logiques et mathématiques authentiques, ainsi que les objets sur lesquels ils portent « ne doivent pas être variables, mais invariables, objectifs et éternels ». A cela s'ajoute *l'argument de la dissolution* associé à l'idée que le psychologisme « dissout les objets et les lois logiques et mathématiques dans un magma de représentations fluctuantes, i.e. l'idée selon laquelle, faute d'objets et de lois stables qui soient seulement la référence des énoncés qui les expriment, il ne serait même pas possible de les connaître ». Enfin, *l'argument de l'utilisation* : contre l'idée selon laquelle le sens d'une notion est donné par son usage, on évoque une « erreur de catégorie » : « le psychologiste fait une erreur de catégorie en cherchant à définir des notions en termes d'opérations et de processus mentaux, en confondant une analyse génétique de la manière dont on forme psychologiquement une notion ou une description de ce qui se passe dans l'esprit de celui qui pense cette notion avec le contenu objectif de cette notion, avec son sens public »[16].

Nous touchons ici à un point crucial de notre discussion, à savoir l'expulsion des pensées hors du flux de la conscience dans le but explicite d'en préserver leur objectivité : instituer l'objectivité des pensées par rapport à la subjectivité des représentations mentales constitue une étape fondamentale à l'installation des fondements d'une théorie épurée de toute intervention psychologisante. L'attitude manifestement hostile à l'égard du psychologisme incite Frege à insister sur la distinction nette entre logique et psychologique. Cette bipartition stricte, ajoutée à la consigne d'expulsion des pensées hors du contenu de conscience le conduit

[14] (G. Frege, 1892a/1971 :191-92)
[15] (P. Engel, 1996 : 69)
[16] *Ibid*., p. 73.

inévitablement vers ce que M. Dummett[17] appelle la « mythologie du troisième monde », et qui peut être schématisée ainsi :

Monde	Mode	Référent
Monde 1 - intérieur	Subjectif	Sensations privées, contenus de conscience
Monde 2 – extérieur	Objectif	Objets matériels
Monde 3 au-delà du clivage intérieur /extérieur	Objectif	Pensées et leurs composantes sémantiques (bien qu'accessibles à tous, ces éléments immuables et immatériels n'ont pas d'effet sur les sens, mais nous pouvons cependant les *saisir*.)

Figure 3: Les trois mondes de Frege

Comme l'a fort bien relevé Dummett, cette distribution ontologique a pour conséquence immédiate le rejet du psychologisme. Car si la pensée n'appartient pas au contenu de la conscience, elle ne peut pas être décomposée, par analyse, en éléments du contenu de la conscience. C'est pourquoi la logique et les théories de la pensée et de la signification doivent être clairement distinguées de la psychologie. Pour Frege, « les pensées ne sont absolument pas irréelles, mais leur réalité est d'une *nature différente* de celle des choses »[18] dans le sens où « l'être-vrai d'une pensée est indépendant du temps »[19]. A la différence d'une représentation mentale, une pensée est objective et autonome. L'orientation de l'analyse conceptuelle dans une direction clairement antipsychologiste oblige Frege – observe Dummett – « à chercher autour de soi, quelque chose de *non mythologique mais d'objectif et d'extérieur à la conscience individuelle* (…). Et quel meilleur endroit pour cette découverte que l'institution d'un langage commun»[20] ?

Cette question laisse cependant percer une faille majeure: si la pensée reste une entité ontologiquement objective, qu'en est-il du processus de *saisie* d'une pensée ? Autrement dit, comment conçoit-on la traversée des différents niveaux depuis le

[17] (M. Dummett, 1988/1991)
[18] (G. Frege, 1892a/1971: 194) (Nous soulignons)
[19] *Ibid.*, p. 191.
[20] (M. Dummett, 1988/1991: 4) (Nous soulignons)

radicalement subjectif vers le radicalement objectif ? D'une part, il est évident que la saisie d'une pensée ne peut être portée, bien évidemment, que par un *acte* psychologique, mais c'est un acte par lequel la conscience appréhende une chose qui lui est extérieure (« extérieure » au sens où c'est quelque chose qui existe indépendamment du fait d'être saisie par tel ou tel sujet). Comment concilier alors l'indéniable versant psychologique de tout acte de *saisie* d'une pensée objective avec l'impératif frégéen de maintenir rigoureusement *séparé* le subjectif de l'objectif ? A vrai dire, il est difficile de retirer des écrits de Frege une représentation claire de ce que signifie « saisir une pensée objective » ou un « sens »[21]. « Selon Frege – écrit Dummett - quand il s'agit de caractériser le sens d'une expression ou d'une proposition, nous ne devons accorder aucune considération au processus de la saisie de sens – qui relève seulement de la psychologie. C'est pourquoi Frege prend rarement ce processus en considération »[22]. L'affaire des pensées objectives est significative dans la mesure où elle paraît conforter l'idée que *la séparation dichotomique entre le domaine du logique et du psychologique* suffit à garantir le statut *ontologiquement objectif* d'une pensée.

Le défi à relever est à présent identifié : en premier lieu, il s'agit d'établir en quoi certaines entités mentales (telles les représentations) seraient-elles nuisibles à l'objectivité d'une trame argumentative, alors que telles autres (les pensées) en constituent d'excellents composants, allant jusqu'à en postuler un espace ontologique distinct pour les accueillir? Un autre point, plus controversé encore, concerne, comme nous l'avons déjà mentionné, la *médiation* entre le radicalement subjectif (les contenus d'expérience) d'une part, et le pleinement objectif (les pensées) d'autre part. Pour reformuler la question sous un angle différent, il s'agit de trouver la « glande pinéale » permettant au sujet de fusionner les deux niveaux. « La saisie d'une pensée – concède Frege – suppose quelqu'un qui la saisisse, quelqu'un qui la pense. Ce quelqu'un est alors porteur de l'acte de penser, non de la pensée »[23]. Or, s'il y a un aspect dans tout ce qui précède qui mérite une attention toute particulière, c'est justement la question de la *jointure* entre l'*acte* de penser et la *pensée* elle-même. Par quelle mystérieuse alchimie s'opère-t-il donc le passage depuis l'acte *subjectif* vers la pensée *objective* dans la tête

[21] « Nous ne pouvons donc retirer des écrits de Frege nulle représentation claire de ce que signifie saisir un sens. » Sa mythologie fait du « sens un objet à l'existence indépendante, avec lequel la conscience entrerait de quelque façon en contact ». (*Ibid.*, p. 101)
[22] *Ibid.*, p.94.
[23] (G. Frege, 1892a /1971:192)

d'un sujet lorsque celui-ci, se cognant l'orteil contre le fauteuil, il éprouve de la douleur et pense objectivement « J'ai mal » ?

2.2 Un premier exemple de « saut conceptuel »

La fragilité de l'édifice frégéen des trois mondes tient, comme l'a relevé P. Engel[24] à la fois à l'invocation d'un monde immuable, et à la rigidité de la séparation subjectif/objectif. Cette remarque est particulièrement importante à nos yeux car elle nous permet de toucher à un point essentiel de notre enquête ici, à savoir la question de l'*articulation* entre les deux membres du clivage. Nous défendons ici l'hypothèse selon laquelle, de manière générale, les stratégies explicatives privilégiant l'utilisation des découpages conceptuels rigides entraînent automatiquement l'apparition de dichotomies favorisant la génération des fossés explicatifs[25]. Or, nous pensons qu'afin d'expliquer de manière satisfaisante un phénomène/état/processus, il ne suffit pas de figer ses coordonnées, ou de l'épingler en quelque sorte sur la carte statique et immanquablement incomplète qu'une théorie nous met à la disposition à un moment donné de son histoire. Encore faut-il rendre compte des connexions et des relations qui le placent dans l'ensemble dynamique du réel. Ce type de difficulté, que nous allons baptiser ici « difficulté de *médiation* » ou d' « *articulation* », émerge notamment en tant qu'effet indirect ou dommage collatéral, si l'on peut l'appeler ainsi, des stratégies explicatives employant des clivages conceptuels rigides. De manière générale, pour pallier à ce type de difficulté, la manœuvre la plus répandue consiste à essayer de « combler» le fossé en faisant appel plus ou moins explicitement à un subterfuge que nous allons baptiser ici « saut conceptuel ».

Selon l'interprétation que nous lui donnons ici, un saut conceptuel est une manœuvre explicative illicite et *ad hoc* censée assurer *a posteriori* la cohérence ou le raccordement entre les rives d'un clivage conceptuel brut, i.e. introduit primitivement tel quel. Pour formuler les choses d'une manière romancée, disons que les sauts conceptuels sont l'envers nocturne des fossés explicatifs, se présentant ainsi sous la forme d'un et seul schème conceptuel. Quelques précisions terminologiques pour

[24] P. Engel écrit à propos de ce sujet : « Cette coupure radicale entre le troisième et les autres (le monde des représentations et le monde physique) rend complètement mystérieux le fait qu'il puisse y avoir le moindre contact entre celui-là et ceux-ci ». (P. Engel, 1996:106)
[25] J. Levine, (1983), 'Materialism and *qualia*: the explanatory gap', op. cit.

commencer. Comme nous l'avons déjà brièvement mentionné dans la partie introductive de cet essai, le terme « saut » représente une métaphore utile et révélatrice, pensons-nous, des difficultés qu'on rencontre lorsque l'on s'efforce à figurer la médiation entre les membres d'une dichotomie. Dans l'exemple que nous avons vu ici, il s'agit de « combler » la séparation entre l'aspect logique et psychologique d'une pensée. Frege hérite de ce clivage sans se soucier du problème de la « communication » entre les deux paliers. Certes, le problème n'est jamais formulé de cette manière sous la plume de Frege, mais on peut tenir la difficulté signalée par Dummett lors de sa critique du processus de « saisie d'une pensée » comme un excellent exemple de « saut conceptuel ».

Quant au deuxième membre de l'expression « saut conceptuel », le prédicat « conceptuel » fait référence ici à l'idée selon laquelle la médiation que l'on insère, bien souvent *in extremis*, afin d'assurer l'intelligibilité d'un modèle explicatif, est une opération de nature conceptuelle. Si nous insistons ici sur cet aspect, c'est dans le but de mettre en évidence le point suivant: non seulement la *saisie d'une pensée* chez Frege constitue un excellent exemple de saut conceptuel entre les pôles subjectif/objectif, mais surtout, nous sommes persuadés que l'utilisation de ce type de polarisation provoque un nombre incalculable de pseudo-problèmes que nous retrouverons agiter le cœur des débats dans la philosophie contemporaine de l'esprit.

Notre hypothèse repose sur l'observation, contestable certes, mais pas non-envisageable, selon laquelle, les distinctions et les clivages conceptuels nets n'ont pas d'équivalent ou de correspondant ontologique direct. En bref, les clivages, si clivages il y en a, sont entièrement l'œuvre d'une opération de l'entendement. Pour revenir à l'exemple de Frege, nous soutenons que le fait même d'avoir introduit un clivage net entre le versant logique et psychologique d'une pensée, crée un « fossé explicatif » entre les deux rives. Lorsqu'il s'agit de « recoller les morceaux » afin d'assurer la cohésion d'ensemble et l'intelligibilité de la théorie explicative, c'est-à-dire lorsque l'on aperçoit l'existence de ce « gouffre », le stratagème le plus souvent employé semble être celui que nous venons d'identifier sous le nom de « saut conceptuel » : autrement dit, on insère un « concept-glu », si on peut l'appeler ainsi, dont la mission ingrate est de chevaucher les deux rives du gouffre initialement intercalé. Le principal objectif de notre travail ici c'est de suggérer qu'une erreur de même type pourrait être à l'origine des nombreuses confusions qui agitent actuellement l'affaire des rapports

entre l'expérience phénoménale et les états neurophysiologiques du cerveau au sein du physicalisme contemporain.

Car le mystérieux phénomène des « sauts conceptuels » est en réalité un fidèle compagnon de route de la pensée philosophique. Il ne convient pas d'en dresser ici une liste exhaustive. Pour les objectifs de notre investigation, il nous suffit de montrer qu'un et même clivage archétypal pourrait être à l'origine de nombreuses controverses suivant l'axe subjectivité/objectivité remis au goût du jour. Le premier exemple que nous avons donné ici concerne le transfert entre les contenus de conscience et les pensées objectives par le truchement d'un mystérieux phénomène de « saisie d'une pensée » chez Frege. Si l'on veut synthétiser, à l'aide d'un schéma, cette situation, nous obtenons la figure suivante :

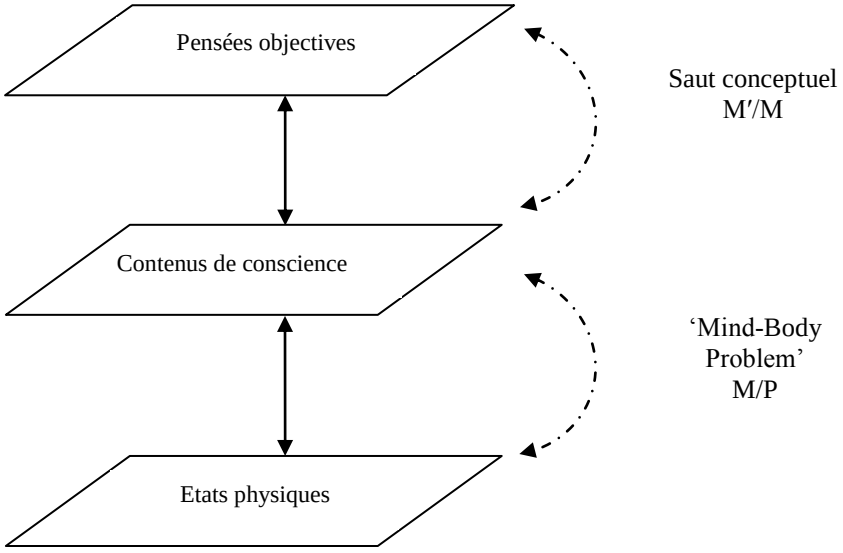


Figure 4: Un premier exemple de saut conceptuel

Mais il est possible de trouver d'autres variations sur le thème du passage mystérieux entre le subjectif et l'objectif. Comme nous allons le voir dans la section suivante (2.3), un saut conceptuel non moins mystérieux relie les modes contentuel et formel de discours chez Carnap. Et c'est à la recherche d'un mystérieux ciment entre le mental et le physique que la tradition analytique à partir du milieu du siècle dernier, emploie toute son énergie. L'hypothèse que nous essayons de dégager ici en parlant de « zones intermédiaires » où l'on rencontre habituellement les mystérieux sauts conceptuels, stipule que ces derniers, loin de faire l'objet d'une application limitée et ponctuelle (dans le cas présent dans la saisie d'une pensée chez Frege), font état d'une pratique généralisée, bien que tacite, profondément ancrée dans la philosophie post-

platonicienne. Délimiter et signaler la présence de ces « zones sensibles » et « hybrides » de jointure, cumulant des tensions philosophiques considérables, constitue, à notre sens, un des éléments clé qui nous permettront de saisir les mécanismes qui ont transformé le problème de la conscience en problème « difficile » pour le physicalisme.

Notre objectif pour le moment étant d'évaluer les dommages collatéraux que l'attitude antipsychologiste des premiers analytiques a provoqués dans l'étude moderne de l'esprit, sans entrer dans les détails d'une discussion détaillée concernant les *qualia* (qui sera étayée plus loin au cours de notre enquête), il convient seulement de noter ceci: il existe un lien très étroit entre la volonté de réduire le langage subjectif de la psychologie à celui objectif de la physique, d'une part, et la subordination des thèmes antipsychologistes à l'ontologie dérivée de l''empirisme d'autre part. Cette subordination peut facilement créer l'illusion d'une applicabilité ou d'une continuité qui glisserait depuis le domaine du langage vers celui de l'ontologie, bref, depuis les concepts vers les faits. Pour C. Hempel, par exemple, la méthode de l'analyse logique du langage scientifique mise au point par les héritiers du Cercle de Vienne abolit de manière décisive « le fossé infranchissable »[26] qui sépare les sciences de l'esprit et de la culture en général du domaine des sciences de la nature. Si bien que lorsque R. Carnap publia en 1928 son *Der logische Aufbau der Welt*[27] l'affaire du sujet psychologique semblait définitivement classée hors du périmètre respectable de l'analyse logique du langage.

Bien plus, dans la mesure où le contenu théorique d'une science réside dans ses énoncés, il suffit de convertir les énoncés de la psychologie en énoncés physiques pour conférer à la psychologie le statut d'une discipline scientifiquement respectable. Or, comme l'a relevé Dummett, la véritable raison de la thèse suivant laquelle les « pensées et leurs composantes sémantiques n'appartiennent pas au contenu de la conscience, ne se trouve pas là où Frege l'a située – c'est-à-dire dans l'objectivité de la pensée et dans la subjectivité de la conscience, mais dans *la différence catégorielle* entre les pensées et les images mentales d'une part, et les impressions sensibles de l'autre. Inventée pour la défense de cette théorie, la mythologie du "troisième monde" ne fait en réalité que la

[26] C.G. Hempel, (1935), 'Analyse logique de la psychologie', *Journal de Synthèse*, vol. 10, p. 27-42, réédité in D. Fisette et P. Poirier, (eds.). (2002). *Philosophie de l'Esprit - Psychologie du Sens Commun* et *Sciences de l'Esprit* (Vol. I). Paris, Vrin, p. 200. (Nous faisons ici référence à l'édition française).

[27] R. Carnap, (1928a), *Der logische Aufbau der Welt*, Berlin: Bernary; trad. angl. par R.A. George *The Logical Structure of the World*, Berkeley: University of California Press, 1967; trad. fr.par E. Schwarz et T. Rivain. *La Construction logique du monde*, Paris, Vrin, 2002. (Désormais « Aufbau » en abrégé).

saper : *notre saisie des pensées ne peut être médiatisée*, et ce dont nous avons une connaissance immédiate ne peut être qu'un contenu de conscience »[28]. Dès lors, un double processus d' « objectivation » s'institue : d'une part, l'accessibilité aux pensées reposera sur la possibilité de leur expression dans un langage articulé, c'est-à-dire *objectif* et *public*. D'autre part, l'objectivité des pensées et leur indépendance à l'égard des contenus subjectifs de la conscience sera définie et ordonnée, comme l'observe Dummett, par « l'accord existant dans la communauté linguistique sur les normes d'utilisation correcte, et par des critères de vérité des énoncés »[29]. Très brièvement résumé, on pourrait dire que chez les premiers analytiques, toutes les conditions sont remplies pour que l'aspect *logique* (***l***'esprit) l'emporte sur le *psychologique* (***les*** esprits) dans l'ordre du réel. Bien plus, nous verrons que l'insertion d'une différenciation séparant l'objectivité de la pensée et la subjectivité de la conscience laissera une empreinte indélébile sur la physionomie des controverses qui passionneront par la suite les protagonistes du duel opposant les '*qualiaphiles*' aux '*qualiaphobes*'.

Récapitulons. Dans cette section, on a cherché notamment à mettre en lumière un premier exemple emblématique de saut conceptuel situé au niveau de l'articulation entre le radicalement subjectif et le pleinement objectif dans le processus de « saisie d'une pensée » chez Frege. Cet exercice marque la première étape de l'examen « radiologique » que nous nous proposons d'élaborer ici, visant notamment à démasquer les présupposés tacites et corrosifs qui vicient certains aspects des débats actuels sur l'esprit. Cette stratégie nous permettra de remettre en cause le bien-fondé du monopole théorique qu'exerce sur la philosophie analytique contemporaine un certain paradigme d'intelligibilité de filière platonicienne qui privilégie une perspective que nous pouvons appeler « mécanique » et rigide dans le traitement du problème des rapports entre le monde physique et l'expérience phénoménale. Nous comptons désamorcer ainsi, étape par étape, l'action souterraine d'un véritable dogme spéculatif, hautement influent dans la philosophie actuelle, qui agit dissimulé sous une apparence extérieure des plus anodines, mais dont la rigidité explicative semble difficilement rendre compte de la complexité et de l'incroyable souplesse qui caractérise la dynamique du monde naturel.

---

[28] (M. Dummett, 1988/1991:139)
[29] *Ibid.*, p. 43.

Il nous reste deux tâches majeures à accomplir dans les deux sections suivantes. En premier lieu, il s'agit de signaler la présence d'un autre exemple de saut conceptuel situé cette fois-ci au cœur même du physicalisme sémantique carnapien. A l'instar de Frege, qui élude la question vitale, à nos yeux, de l'articulation entre l'*acte* de penser (subjectif) et la *pensée* (objective) dans l'analyse des processus mentaux, nous soutiendrons que Carnap laisse en suspens un problème appartenant à la même famille, si l'on peut appeler ainsi, difficulté décelable dans la célèbre affaire des énoncés protocolaires, et situable au niveau de l'articulation entre le *mode contentuel* et le *mode formel* d'un discours. Ces espaces hybrides, situés sous haute tension, et dont le statut d'entre-deux reste hautement suspect, sont symptomatiques, selon nous, de la présence d'une puissante tension conceptuelle, conséquence directe de l'adoption tacite du clivage subjectivité/objectivité. Enfin, il s'agit de montrer en quoi l'échec de cette version initiale et sémantique du physicalisme fait écho à l'adoption tacite d'un arrière-plan théorique beaucoup plus vaste, dont les effets sont visiblement encore à l'œuvre dans les controverses que suscite actuellement la question des *qualia* chez les défenseurs du physicalisme.

Le processus d'expulsion des pensées hors des contenus de conscience, étayé par le projet positiviste de remplacement des « faits » par des « énoncés portant sur les faits » a eu une conséquence surprenante: dans sa qualité de "véhicule public, objectif des pensées"[30] le langage devient un acteur clé dans la compréhension du fonctionnement de l'esprit. Or, à supposer que cette séparation des domaines d'application (d'un côté les *faits*, de l'autre les *énoncés portant sur les faits)* évolue sur l'arrière-plan conceptuel revendiquant l'unité de la science, la séparation faits/énoncés ne pourra aboutir qu'à une impasse : nous retrouverons ainsi, en germe, un *fossé explicatif* au sein même du physicalisme carnapien. Ce constat appuie et étaye ainsi l'hypothèse récemment avancée par R. Van Gulick[31] selon laquelle, le problème des fossés est beaucoup plus ancien que le label imposé par J. Levine[32].

Avant d'aller plus loin, il importe d'écarter une possible objection préliminaire. En effet, le lecteur pourrait soulever l'interrogation suivante : à quoi pourrait bien aboutir une mise en relation entre deux événements si éloignés, à savoir le projet

[30] P. Engel, (1994), *Introduction à la philosophie de l'esprit*, Paris: Et la Découverte, p. 7.
[31] R. Van Gulick, (2003), 'Maps, Gaps, and Traps' in Q. Smith & A. Jokic (eds.), *Consciousness – New Philosophical Perspectives*, op. cit.
[32] (J. Levine, 1983)

carnapien de reconstruction rationnelle du monde d'une part, et l'affaire des *qualia* dans le physicalisme contemporain d'autre part ? Notre réponse est la suivante : un des objectifs de cet essai est de mettre en lumière une possible analogie entre les failles du physicalisme sémantique carnapien et celles du physicalisme *a priori* contemporain. Pour ce faire, il nous faut montrer qu'une et seule armature fondamentale les sous-tend. Sans vouloir anticiper ici un sujet qui sera repris et détaillé plus loin au cours de notre enquête, rappelons seulement que parmi les thèses[33] qui traversent et consolident l'édifice théorique carnapien, la distinction analytique/synthétique figure à une place d'honneur. Considérant la distinction *vérité analytique* et *vérité de fait*, Carnap écrit: « Pour la philosophie des sciences, une distinction très nette entre analytique et synthétique est à mon avis, de la plus haute importance »[34]. La typologie analytique/synthétique a donné du fil à retordre aux philosophes postkantiens. Il s'agissait de décider si l'on doit situer, selon l'excellente formule de J. Fodor et J. Katz[35] « la charrette du formalisme devant le bœuf de l'empirisme » ou vice-versa, et avec quelles conséquences pour l'évolution de la pensée philosophique.

Or, si notre hypothèse de départ est correcte, et si nous arrivons à circonscrire un exemple de « saut conceptuel » au sein même du projet physicaliste carnapien d'unification de l'ensemble des propositions sous l'égide du langage de la physique, il nous sera plus facile de saisir, et par la suite de dénoncer, l'égarement qui le sous-tend, et qui sera décelable au cœur de certaines formes réductionnistes du physicalisme actuel.

2.3 Carnap et les contenus d'expérience vécue

Dans la mesure où pour les philosophes de Vienne représenter le monde extérieur équivaut à le reconstruire logiquement à partir des données de sens dans la plus pure

[33] P. Wagner résume en cinq points les thèses qui traversent invariablement la pensée carnapienne : (i) la distinction analytique/synthétique ; (ii) la possibilité d'énoncer un critère de la signification; (iii) le projet d'une reconstruction rationnelle de la science; (iv) la construction d'un langage logique pour la science unitaire; (v) la nécessité de remplacer la philosophie conçue comme une connaissance théorique par une activité, celle qui est à l'œuvre dans la logique de la science. Voir P. Wagner, (2002), 'Carnap et la logique de la science' *in* P. Wagner, (ed.), *Les philosophes et la science.* Paris: Gallimard, p. 295-96.

[34] R. Carnap, (1966), *Philosophical Foundations of Physics: An Introduction to the Philosophy of Science*, New York: Basic Books. 'Les énoncés analytiques dans un langage d'observation', trad. fr. par J.-M. Luccioni & A. Soulez: *Les fondements philosophiques de la physique*, Paris: Armand Colin, 1973, p. 249. (Nous faisons référence ici à l'édition française)

[35] J. Fodor & J. Katz, (1965), *The Structure of Language.* New Jersey: Prentice Hall, Inc., Englewood Cliffs, p. 7.

tradition de type Russell-Whitehead[36], il est clair que celui qui fut le plus près de venir à bout de cette entreprise fut l'auteur du *Der logische Aufbau der Welt*[37].

Ce qui suit est un petit rappel, très schématique, des tentatives du physicalisme carnapien de fonder un modèle rationnel de la connaissance épuré de toute référence mentaliste aux états psychologiques de l'agent locuteur. Nous n'entendons pas nous interroger ici sur les multiples influences d'arrière-plan qui ont conduit l'auteur de l'*Aufbau* à envisager une « construction logique du monde ». Il ne s'agit pas non plus de faire « référence vertueuse et souvent obligée à Carnap »[38] pour le critiquer ensuite sur la base de quelques préjugés superficiellement survolés. La contribution de Carnap à la reconstruction rationnelle de la connaissance est un sujet bien trop complexe pour que l'on puisse l'aborder ici dans son intégralité. Plus modestement, nous limiterons notre analyse à un seul de ses aspects, afin de mieux élucider la présence d'un *saut conceptuel* intercalé au niveau de l'articulation entre le mode formel et le mode contentuel de discours dans le processus de « physicalisation » des énoncés protocolaires.

Notre propos est donc de saisir les empreintes laissées par le discours physicaliste carnapien dans l'organisation des débats sur l'axe dichotomique états mentaux/états physiques. Une mise sur le banc d'essai de ce clivage est d'autant plus indispensable aujourd'hui, lorsque nous devons faire face aux conséquences dramatiques qu'entraîne l'adhésion à ce modèle dans le débat autour des *qualia*.

Aux yeux de Carnap, il n'y a pas de pire ennemi pour la philosophie que l'inexactitude. Or, seule la logique de la science, dont la mission est celle de rendre possible la construction métathéorique et l'analyse des langues logiques dans lesquelles peuvent être exprimées les propositions des théories scientifiques, permet un traitement et une formulation précise et claire des interrogations philosophiques. Il est pourtant simple, affirme Carnap, d'éviter les fâcheuses confusions entre les propositions qui

[36] W.O. Quine, (1969), "Epistemology Naturalized", in *W.O. Quine, Ontological Relativity and Other Essays*, New York: Columbia University Press; trad. fr. A. Fagot-Largeault, in *Relativité de l'ontologie* et *autres essais*, (1977), Paris : Aubier Montaigne repris in S. Laugier, et P. Wagner, (eds.). (2004). *Philosophie des Sciences - Naturalismes et réalismes* (Vol. II). Paris: Vrin, p. 41. (Nous utilisons ici cette dernière édition) Pour une discussion concernant les differents enjeux de l'*Aufbau*, voir M. Friedman, (1999), *Reconsidering Logical Positivism*, Cambridge : Cambridge University Press.

[37] R. Carnap, (1928a), *Der logische Aufbau der Welt*, Berlin: Bernary, trad. angl. par R.A. George *The Logical Structure of the World*, Berkeley: University of California Press, 1967; trad. fr. E. Schwarz et T. Rivain. *La Construction logique du monde*, Paris : J. Vrin, 2002. (Dans ce qui suit, nous allons nous référer à l'édition anglaise de 1967).

[38] S. Laugier, (2001), (ed). *Carnap et la construction logique du monde*. Paris: Vrin, p. 8.

portent sur des objets et celles qui portent sur le langage. Il suffit d'introduire une délimitation nette entre deux modes de discours : contentuel et formel. Une fois que l'on se débarrasse du mode contentuel, pour ne garder que le mode formel, on réalise que la syntaxe logique du langage consolide l'architecture de l'édifice entier des sciences. En bref, la logique de la science permet de fixer et de garantir, au niveau de l'analyse syntaxique, que toutes les propositions des sciences particulières peuvent être exprimées dans une langue unique. A supposer que cette langue unique est la langue de la physique, la pensée de Carnap marque ainsi son adhésion explicite à la thèse du *physicalisme sémantique*.

Quelques brèves remarques générales pour commencer. Lorsque Neurath introduit les termes de « physicalisme »[39] et d'« unité de la science »[40], un des problèmes majeurs que les empiristes logiques doivent affronter est celui qui concerne le statut des *énoncés protocolaires*. Il serait trop long de reprendre ici dans son intégralité la célèbre controverse autour du statut de ces énoncés au destin si particulier. Essayons néanmoins de rappeler les points essentiels du débat. Lorsque M. Schlick[41] et F. Waismann[42] affirment que les énoncés protocolaires portent sur des données de sens, ils s'exposent au danger redoutable de voir ainsi s'ouvrir la boîte de Pandore du psychologisme : en effet, faire tenir la solide et impressionnante pyramide des théories scientifiques sur la seule pointe glissante des contenus d'expérience individuels, relève d'un véritable exploit. Malgré tout, cette conception a le mérite de prendre courageusement en compte la question des deux pôles extralinguistiques qui attachent la pensée - *via* l'énoncé - au monde, à savoir le sujet psychologique et le monde extérieur, empirique.

Pour Neurath et Carnap, en revanche, les énoncés protocolaires portent sur l'observation d'objets physiques préexistants, c'est-à-dire donnés en quelque sorte « en cadeau » à la pensée du philosophe. Il est vrai que le solipsisme[43] initial de l'*Aufbau*

---

[39] O. Neurath, (1931b), "Physikalismus", *Scientia* 50, pp. 297-303, réédité in O. Neurath, (1983) *Philosophical Papers 1913-1946* (et by R.S. Cohen and M. Neurath), Dordrecht: Reidel.

[40] O. Neurath, (1935a), "Einheit der Wissenschaft als Aufgabe", *Erkenntnis* 5, pp. 16-22, réédité *in* O. Neurath, (1983), op. cit.

[41] M. Schlick, (1935a), "Facts and Propositions", *Analysis* 2, pp. 65-70. Voir aussi M. Schlick, (1936a), "Meaning and Verification", *Philosophical Review* 45, pp. 339-369; réédité in M. Schlick, (1979b), *Philosophical Papers* Vol.2 (1925-1936) (et L. Mulder and B. van de Velde-Schlick), Dordrecht: Reidel.

[42] F. Waismann, (1965), *Principles of Linguistic Philosophy* (ed. R. Harré), London: Macmillan.

[43] Selon Ayer (et nous le suivons sur ce point), l'ajout de l'épithète « méthodologique » au solipsisme « a été mis pour assainir le solipsisme, mais on peut se demander si c'était suffisant pour le but proposé ».

prend comme inévitable point de départ l'expérience subjective d'une personne à un moment déterminée dans l'espace-temps. Les vécus élémentaires sont là, présents et assimilables aux atomes de sensations. L'avantage de cette position réside, comme le note A.J. Ayer[44]dans le fait qu'il évitait beaucoup de problèmes désagréables. Mais en avait-il le droit? Schlick est un des premiers à avoir soulevé de sérieuses objections quant à l'applicabilité du critère fort de signification au cas particulier des énoncés protocolaires employant des concepts de *couleur* : "La constatation que je fais en disant « ici maintenant bleu », *n*'est *pas* la même chose que la proposition protocolaire « M. S. a perçu du bleu tel jour d'avril 1934, à tel ou tel instant, et en tel et tel lieu »; au contraire, cette dernière proposition est une hypothèse, et, en tant que telle, elle est toujours entachée d'incertitude. (...) Dans les propositions protocolaires il est *toujours* question de perceptions"[45]. Il s'agit donc de faire une place aux *expériences personnelles subjectives* latentes dans le « solipsisme méthodologique ».

Lorsque l'on définit, suivant Carnap et Neurath, les énoncés protocolaires comme des énoncés portant sur l'observation d'objets physiques, il va de soi que la neutralité de l'observation traduit l'exigence empiriste de dresser un rempart contre les dangers du psychologisme. Or, nous essayerons de montrer dans cette section que Carnap se « donne en cadeau » non seulement *le monde extérieur public*, comme le signalait à juste titre Ayer, mais aussi – et c'est ici un point important – *le monde intérieur privé*. Essayons d'énoncer d'une manière plus claire le présupposé auquel nous faisons allusion.

Lorsqu'il épouse la distinction kantienne analytique/synthétique, Carnap hérite en « trousseau », si l'on peut dire ainsi, de la délicate dentellerie des problèmes que représente le *raccordement* entre le corpus d'énoncés théoriques en tant que système formel d'une part, et les faits empiriques, bruts, d'autre part. Situé à la périphérie du système, l'énoncé d'observation « porte son propre contenu empirique sur ses épaules »[46] comme le résume Quine. Or, derrière ce contenu empirique on retrouve non seulement le vieux problème du monde extérieur, mais aussi le problème des contenus d'expérience subjective, car le sujet est lui *aussi* un membre légitime du réel empirique. Laisser le sujet en dehors de l'affaire de la connaissance, l'oublier en quelque sorte

A.J. Ayer, 'Le Cercle de Vienne', in J. Sebestik & A. Soulez, (eds.), (1986). *Le Cercle de Vienne - Doctrines et Controverses.* Paris: Meridiens Klincksieck, p. 67.

[44] (A. J. Ayer, 1986: 76)

[45] (M. Schlick, 1934/2006: p. 437-438) (Soulignement en original)

[46] (W.V. Quine, 1969/2004 : 58)

comme une malle encombrante sur le quai de la gare, mènerait inévitablement à de fâcheuses inadvertances. Comme l'a fort bien remarqué M. Bitbol[47], la dualité kantienne schème/contenu est une véritable armature formelle de la dualité épistémique sujet/objet. Le défi à relever est à présent identifié : il s'agit de saisir comment construire un *fait* objectif en partant du contenu d'expérience vécue qu'abrite intrinsèquement tout énoncé élémentaire contenant l'inévitable « je » privé et subjectif. Nous retrouverons ainsi, dans un état embryonnaire, une très intéressante variation sur le thème des rapports entre la phénoménalité de l'expérience subjective et l'objectivité des états physiques.

Maintenant que le défi est identifié, comment faut-il s'y prendre pour le relever ? Autrement dit, il nous faut tirer au clair ce que signifie d'incorporer la subjectivité de chaque protocole dans le magma objectif d'une théorie scientifique. Si notre hypothèse de départ concernant *le tandem fossés explicatifs/sauts conceptuels* est correcte, il devrait s'ensuivre que la manière dont Carnap traite le passage d'un niveau à l'autre – c'est-à-dire depuis le flux des contenus d'expérience vécue, vers l'objectivité théorique d'un système scientifique – est tout, sauf claire. Ce qui va sans dire, c'est le lien inextricable qui unit chaque sujet à son « protocole ». Ce qui ne va pas sans dire, par contre, c'est essayer de montrer « comment peut-on construire, d'étape en étape, toute la batterie des concepts nécessaires pour décrire le monde »[48] en se tenant en équilibre, comme un acrobate, sur la pointe de la subjectivité d'un protocole.

D'emblée, il importe de noter que pour Carnap, seules des résistances émotionnelles nous empêchent d'accepter la thèse de la traduction suivant laquelle toute proposition *P* d'une science empirique correspond une proposition *Q* formulée en langage physique, dans le sens où *P* et *Q* sont mutuellement déductibles à l'aide de définitions et règles de correspondance permettant de substituer au mode « classique » *contentuel* de discours (qui traite, en occurrence, des contenus et expressions psychologiques), un mode *formel* de discours, qui représente leur équivalent syntaxique. Selon la synthèse qu'en donne D. Andler[49], deux étapes guident le programme carnapien de fondation de la connaissance : l'étape de *traduction* où il faut montrer que les termes théoriques et les termes désignant des objets physiques peuvent

47 (M. Bitbol, 2010)
[48] (A.J. Ayer, 1986: 67)
[49] D Andler, A. Fagot-Largeault, et B. Saint-Sernin, (2002), *Philosophie des Sciences* (Vol. I). Paris: Gallimard, p. 233.

être entièrement définis à partir de termes désignant des données de sens. Ensuite, dans l'étape suivante, de *dérivation*, le philosophe est en principe capable de vérifier si les énoncés, une fois transformés, dérivent effectivement ou non de l'expérience.

Au point de départ, à la simplicité d'une évidence intuitive de type cartésien, on substitue donc, en accord avec les exigences empiristes, *l'expérience* de l'agent locuteur. En bon empiriste, Carnap[50] ne concède rien à l'introspectionnisme cartésien. Ce passage obligatoire à la case "expérience" est censé ancrer les racines de toute entreprise philosophique dans le terrain empirique. Chez Carnap, le "positivisme méthodologique" ou le "solipsisme méthodologique" est la conception selon laquelle « les protocoles forment la base de la construction d'ensemble de la science dans la mesure où *chaque sujet ne peut prendre pour base que son propre protocole*"[51]. Bref, le point de départ, souligne Carnap[52], doit être pris tel quel, une simple occurrence. L'individu évolue dans le « *flux (continuum) de l'expérience* »[53].Par conséquent, il est parfaitement possible, affirme Carnap[54], de construire, ou mieux dit de *re*construire, l'objectif en partant des contenus privés d'expérience vécue subjective. C'est la manœuvre que nous comptons contester ici. Pour ce faire, un bref détour rétrospectif s'impose.

Rappelons-nous que selon la version la plus radicale du vérificationnisme, que nous simplifions ici à l'extrême, les énoncés affrontent un par un le monde empirique et se prêtent à une hiérarchisation ascendante de plus en plus complexe, en commençant avec les énoncés protocolaires et allant jusqu'aux théories scientifiques. La proposition "*Monsieur X a observé ceci et cela à tel instant et en tel lieu"* peut se résumer à "certains complexes de signes faits de sons ou d'encre"[55]. Il s'agit donc d'un arrangement théorique "commode" dans lequel, comme le résume P. Engel, "la structure des propositions atomiques reflète celle du monde : aux propositions atomiques correspondent des faits, eux-mêmes arrangés en états de choses, qui sont eux-mêmes constitués d'arrangements d'objets. Les noms désignent directement des

[50] "The **sum** does not follow from the **cogito**; it does not follow from < I experience > that < I am >, but only that an experience is". (R. Carnap, 1928a/1967: 261)
[51](R. Carnap, 1932b/2006: 357) (Nous soulignons)
[52] "At the beginning of the system, the experiences must simply be taken as they occur". R. Carnap, (1928a/1967: 101)
[53] "the stream of experience" (*Ibid.*, p.102)
[54] « (...) in our constructional system, 'my experiences' are the basic elements (…) Science are essentially concerned with structure and that, therefore, there is a way to construct the objective by starting, from the individual stream of experience". (*Ibid.*, p. 104-7)
[55](M. Schlick, 1934/2006 : 418)

objets, et parce que les propositions élémentaires sont constituées des noms qui exhibent le même arrangement que l'arrangement des objets dans les états de choses élémentaires, cet isomorphisme permet de dire que les propositions atomiques « représentent » des états de choses"[56].

Dans son livre *Syntaxe logique du langage*[57], considéré aujourd'hui comme véritable chef-d'œuvre de la pensée analytique, Carnap procède à un véritable tour de force afin de "purifier" la philosophie de tous les "pseudo-énoncés" qui l'encombraient jusqu'à présent. L'effort central de son argumentation converge vers l'idée qu'un langage peut contenir et définir sa propre syntaxe sans circularité et sans contradiction[58] et à ce titre, cette thèse constitue, comme l'a relevé A. Coffa, « la première alternative sérieuse à la conception kantienne de l'*a priori* »[59]. Il s'agit non seulement de ramener la science à une forme unifiée, mais aussi, aspect très important, de remplacer la modalité de théorisation habituelle qui traite d' « objets » et d' « états de choses » par une "formulation correcte qui parle de *mots* et non d' « objets », de *propositions* et non d' « états de choses »"[60]. Bref, remplacer le mode contentuel du discours avec le mode formel, dans l'esprit d'une réforme radicale de la philosophie, qui se retrouve ainsi "purifiée" d'un bon nombre de pseudo-problèmes, transforme la philosophie en logique de la science[61]. L'intention explicite de Carnap est de montrer que toutes les questions liées à la logique de la science et aux fondements des sciences individuelles sont, à la base, des questions de *syntaxe*. Bref, la logique de la science est syntaxe, et elle peut être correctement formulée *depuis l'intérieur*. La syntaxe, pure et descriptive, n'est rien d'autre qu'une mathématique et physique du langage[62]. Examinons à présent de plus près les tentatives acrobatiques de Carnap pour faire tenir l'objectivité de la science physique sur la pointe du protocole subjectif de l'individu.

---

[56] P. Engel, *La norme du vrai - philosophie de la logique*, Paris, Gallimard, 1989, p.114.

[57] "The aim of logical syntax is to provide a system of concepts, a language, by the help of which the results of logical analysis will be exactly formulable. *Philosophy is to be replaced by the logic of science (...) for the science is nothing other than the logical syntax of the language of science"*R. Carnap, 1934/37, *Logische Syntax der Sprache*, Wien: Springer. *The logical syntax of language*, London: Routledge et Kegan Paul, 1967, pour l'édition consultée, p. xiii (Les soulignements sont par Carnap)

[58] "A language may contain its own syntax without contradiction ". (R. Carnap 1934/1967:287)

[59] A. Coffa, (1986), 'Le positivisme logique, la tradition sémantique et l'*a priori'*, in J. Sebestik & A. Soulez, (eds.), op. cit., p. 94.

[60] (R. Carnap, 1932b/2006: 325) (Nous soulignons)

[61] "Only the logic of science remains. *(...) the logic of science takes the place of the inextricable tangle of problems which is known as philosophy.*" (R. Carnap, 1934/1967: 279) (Les soulignements sont par Carnap.)

[62] "Syntax, pure and descriptive, is nothing more than the mathematics and physics of language". *Ibid.*, p. 283.

2.4 La perception du « rouge » dans le physicalisme sémantique

Soit *p* la proposition protocolaire « (Svoit) maintenant du rouge ». Soit *P* la proposition physique « Le corps S est en train de voir maintenant du rouge ». Selon Carnap, « on peut déduire de *P* la proposition protocolaire p : " (Svoit) maintenant du rouge " ; et inversement, de *p* on peut aussi déduire *P*. Autrement dit, *p* est traduisible en *P*, *p* et *P* ont le même contenu. (L'expression métalogique "ayant le même contenu" est définie par "mutuellement déductibles").

Toute proposition de la langue protocolaire de *S* est ainsi traduisible en une proposition physique qui décrit l'état corporel de *S*. Autrement dit, la langue protocolaire de *S* est coordonnée à une sous-langue tout à fait spéciale de la langue de la physique, d'une manière telle que dès qu'apparaît une proposition quelconque de la première langue dans le protocole de *S*, alors la proposition physique coordonnée a une valeur intersubjective et inversement. « Deux langues isomorphes de ce type ne se distinguent que par les mots qui figurent dans les propositions. Du fait qu'on a établi l'isomorphisme, la *langue protocolaire* est transformée en *une sous-langue de la langue de la physique*[63].

Réduite à sa plus simple expression, la version initiale du physicalisme, que l'on va appeler dorénavant, suivant Stoljar[64], **physicalisme sémantique**, désigne la conception selon laquelle *« la physique est une langue universelle et peut par conséquent servir de langue fondamentale pour la science*. (…) Toute autre langue-système possible est donc traduisible dans la langue de la physique »[65]. Bref, « la science toute entière se fait physique »[66].

A présent, si nous reformulons à l'aide des deux modes de discours l'idée selon laquelle *P* et *p* ont le même contenu, on obtient ceci :

[1] « *p* est déductible de *P* et inversement. » - } mode formel

[2] « *P* et *p* décrivent le même état de choses. » - } mode contentuel

Chez Carnap, le mode formel d'un discours [1] sera toujours préférable au mode

[63] (R. Carnap, 1932b/2006 : 353-4) (C'est Carnap qui souligne.)
[64] (D. Stoljar, 2010)
[65] (R. Carnap, 1932b/2006: 358-9) (C'est nous qui soulignons.)
[66] *Ibid*.

contentuel [2]. Dans le premier cas [1], toute référence à un fait extralinguistique (objet physique) ainsi qu'au locuteur émettant cette assertion (agent-locuteur) est cumulée et contenue dans la formalisation. Dans le deuxième cas [2], la présence des termes descriptifs rend, au contraire, plus explicite la nécessité d'un raccordement aux données factuelles. A la différence du discours contentuel, le discours formel carnapien fixe ses références non pas au sens des symboles qu'il met en scène, mais seulement aux règles qui régissent, intrinsèquement, depuis l'intérieur, la syntaxe logique du langage donné.

Selon Carnap, "la proposition protocolaire *p* « (S voit) maintenant du rouge » peut être déduite à partir d'une certaine description de l'état corporel de *S*"[67]. «Dans la description de l'état corporel – écrit Carnap – il est question avant tout de l'état du système nerveux central; et c'est ici en premier lieu de la zone corticale du cerveau qu'il s'agit ; mais *nous n'avons pas besoin d'entrer plus avant dans les détails pour notre propos* »[68]. Dès lors, il devient clair que la *possibilité même* d'une telle description ne semble pas constituer un sujet d'inquiétude pour Carnap : "On émettra peut-être la réserve qu'une déduction de ce type est utopique et ne pourrait être réellement réalisée qu'à condition d'avoir une connaissance précise de la physiologie du système nerveux central"[69].

C'est précisément au niveau de l'articulation entre la *réalisation physique* (le cortex) et la description de l'état mental, que se situe le nœud gordien des futurs débats concernant le 'mind-body problem'. Qu'à cela ne tienne, Carnap balaie d'un simple revers de main le problème, en y apportant cette réponse aussi surprenante qu'insatisfaisante : « on peut effectuer dès maintenant cette déduction, elle est même continuellement effectuée dans la vie quotidienne lorsque des êtres humains se comprennent entre eux »[70]. Autrement dit, nous pouvons effectuer le passage depuis le contenu d'expérience subjective de S vers une description d'un état corporel de S *parce que* les êtres humains le font depuis toujours !

Or, et c'est ici un point important sur lequel nous aimerions attirer l'attention du lecteur : si l'on regarde de plus près l'affaire du *passage* depuis le mode contentuel vers le mode formel du discours, on s'aperçoit que pour que la proposition protocolaire *p* puisse être déduite à partir d'une description de l'état corporel de S, nous devons tenir pour acquises les prémisses tacites suivantes :

---

[67](R. Carnap, 1932b/2006:352)
[68] *Ibid.* (Nous soulignons)
[69] *Ibid.*
[70] *Ibid.*

(1) il existe, pour tout sujet *S*, un état corporel *c* en sorte qu'en présence de tout objet rouge *R*, *c* permet à *S* d'avoir un accès direct, immédiat à *R*.

(2) il existe pour chaque *c* une description *d* telle que si S est dans l'état *c*, *d* peut être décrite objectivement par un tiers S*.

Il semble que cette solution se heurte d'emblée à une difficulté presque insurmontable. Car pour pouvoir justifier la traduction des énoncés protocolaires, Carnap ne peut pas faire l'économie de deux hypothèses supplémentaires : (a) une stipulant l'infaillibilité de l'accès de S à ses propres états mentaux ; (b) l'autre postulant l'accessibilité d'un tiers S*, i.e. le théoricien, à la description des états mentaux de S. Une autre possible alternative, signalée par Ayer, serait de « s'en remettre à la fiction d'un observateur idéal, mais ce n'est pas une notion bien précise. De quels pouvoirs allons-nous créditer un tel observateur »[71] ?

Force est donc de constater que l'équation énoncé/fait s'appuie chez Carnap sur la prémisse tacite et hautement suspecte selon laquelle le sujet serait doté non seulement du pouvoir mystérieux d'*accéder* à ses propres expériences phénoménales, mais aussi du pouvoir encore plus mystérieux de *s'en éloigner*, afin de mieux saisir, comme depuis une vue en surplomb, extraterritoriale, la carte des entités et concepts qui la peuplent. Si la description *d* ainsi obtenue passe le test de vérification, alors elle devient un « membre » légitime du mobilier du monde. Surgit alors une question qui à notre sens est cruciale : quelle est la place réservée aux états psychologiques du sujet dans ce récit à la fois réducteur et optimiste du physicalisme carnapien ?

Regardons-y de plus près : dans un mode formel de discours de type [1] « p est déductible de P et inversement », la relation qui unit *p* et *P* est logique, formelle, intrinsèque et, par conséquent, libre de toute intervention « extérieure ». En revanche, l'équivalent contentuel du même discours [2] « *P* et *p* décrivent le même état de choses » nécessite l'intervention de l'ascension verticale vers l'extérieur. Considérons, pour plus de clarté, un exemple donné par Carnap lui-même dans son article 'Vérité et confirmation'[72] : "Je vois par exemple une clef sur mon bureau et je produis l'énoncé « Une clef se trouve sur mon bureau »; sur la base de mes observations visuelles et peut-être tactiles, je tiens cet énoncé pour confirmé à un degré élevé et par conséquent,

[71] (A.J. Ayer, 1986:76)

[72] R. Carnap, (1936), "Wahrheit und Bewährung", *Actes du Congres International de Philosophie Scientifique*, Sorbonne, Paris, Facs. IV, "Induction et Probabilité", Paris, Hermann et Cie., pp. 18-23, repris in C. Bonnet et P. Wagner, (eds.), op. cit., (2006), pp. 559-566. (Nous faisons référence ici à cette dernière édition).

je l'accepte"[73]. Si nous essayons de retranscrire cet exemple sur un schéma figuratif, nous obtenons ceci :

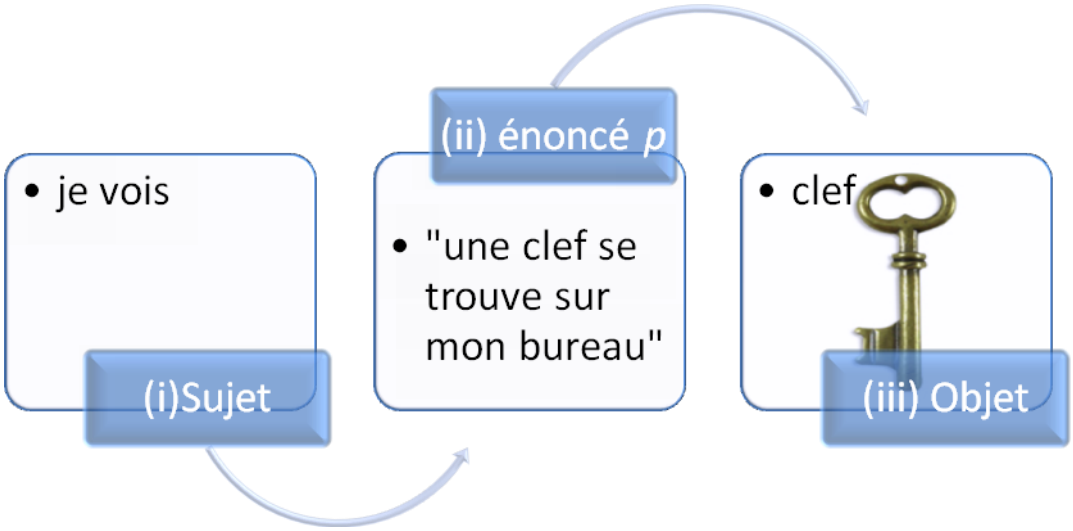


Figure 5: Le pôle sujet/objet dans les énoncés protocolaires

Dans la case (i), nous pouvons identifier ce que l'auteur de l'*Aufbau* appelait "*the stream of experience*", un contenu d'expérience vécue. Ce schéma nous permet de saisir l'idée suivant laquelle l'énoncé protocolaire *p* occupe, aux yeux de Carnap, une position équidistante et neutre, entre deux entités extralinguistiques : le *sujet connaissant* (en occurrence, Rudolf Carnap) et l'*objet* perçu (en occurrence, la clef).

Nous avons ainsi *deux* entités extralinguistiques qui rattachent l'énoncé empiriquement testable *p* « *Une clef se trouve sur mon bureau* » au monde réel: (i) le sujet-observateur (figuré ici par l'indexical « je ») et (iii) l'état des choses « clef-sur-le-bureau ». Dans la mesure où la portée ontologique des « objets » / « états de choses » est réduite à une analyse des énoncés empiriques, cette manière d'interposer les propositions entre le sujet connaissant et l'objet perçu s'avère une manœuvre salutaire censée assurer l'objectivité du discours. Or, - et c'est ici un point sur lequel nous aimerions attirer l'attention du lecteur - lorsqu'il s'agit de construire un *fait objectif* en partant du *protocole subjectif*, c'est-à-dire lorsqu'on doit relier le niveau de l'expérience phénoménale du sujet individuel à l'énoncé "*Je vois une clef*" (ou le "je" peut être analysé dans la neutralité plénière d'un nom universel) et d'ici plus loin à l'assertion publiquement testable *p* "*Une clef se trouve sur mon bureau*", il se produit un glissement imperceptible : dans la formulation qu'en donne Carnap, le lecteur est

[73] (R. Carnap, 1936/2006:562)

subrepticement incité à croire qu'une fois que l'on fixe l'énoncé (l'entité linguistique) dans un mode formel du discours, les deux autres pôles *extra*linguistiques - (i) l'agent locuteur et (iii) l'objet empirique - se dissolvent et disparaissent par miracle au profit de l'énoncé *p*. Le philosophe trace ainsi un cercle magique autour de *p*, à l'intérieur duquel il se sent libre ensuite de bâtir un véritable château formel débarrassé de toute menace de « subjectivisme » ou de « phénoménalité ». Nous allons le voir, cette technique de traçage d'un cercle magique protégeant la pensée philosophique des aléas et déboires de l'empiricité, est loin d'être éteinte, et nous allons la retrouver intacte chez les défenseurs du physicalisme *a priori* et de l'analyse conceptuelle.

Or, en admettant que les "faits" jouent ici le rôle de seuil empirique de l'analyse et qu'à l'autre bout l'on place un observateur qui émet un énoncé dérivant d'une observation qu'il a effectuée d'une manière aussi objective que possible, il est évident que l'équation tripartite « sujet/ énoncés protocolaires/objet » telle qu'elle est posée par l'empirisme logique, se découpe sur un arrière-plan amputé de tout son versant psychologique. Chez Carnap, l'antipsychologisme frégéen est recyclé, si l'on peut dire ainsi, de façon à ce que l'agent locuteur se « dissolve » *dans le produit* de son acte d'énonciation. Paradoxalement, l'émetteur de parole se voit réduit au statut de support *silencieux* et invisible de l'énoncé, ce dernier devenant ainsi, subrepticement, l' « usurpateur » ontologique du sujet.

Anticipant les éventuelles difficultés liées aux transitions reliant les entités extralinguistiques aux propositions, Carnap s'empresse de préciser : "Habituellement, on ne pose aucune règle explicite concernant la manière dont une proposition doit ou peut être formulée si l'on fait des observations déterminées (…) Les enfants apprennent à faire usage du langage courant, et ainsi à accomplir correctement les opérations décrites, par la pratique, par imitation, le plus souvent sans l'aide d'aucune règle"[74]. Autrement dit, bien qu'on dispose de règles explicites gouvernant les structures formelles à l'intérieur d'un langage, on ne dispose en revanche d'aucune règle explicite nous indiquant pourquoi, en voyant une clef, on peut former validement l'énoncé "*Je vois une clef*" et non pas un énoncé de type " *Le ciel est clef*."

On l'a souvent répété, dans l'affaire des énoncés protocolaires, « cette mise hors la loi de la sémantique » [75] a fini par vicier l'édifice purement syntaxique auquel aspirait Carnap. Certes, du point de vue technique, c'était une « performance

[74] (R. Carnap, 1936/2006: 563)
[75] (A.J. Ayer, 1986: 77)

extraordinaire, mais la tentative de rapporter à la syntaxe la tâche de la sémantique a échoué, tout comme a échoué la reconstruction des propositions philosophiques non-métaphysiques en tant qu'énoncés syntaxiques déguisés en énoncés de faits (tel était le clou de la distinction célèbre de Carnap entre le mode formel et le mode matériel du discours). Quoi qu'on pût en dire, les propositions philosophiques n'étaient certainement pas syntaxiques »[76].

Si les règles syntaxiques peuvent avoir des conséquences qui portent sur les faits empiriques, la question qui se pose est celle d'expliquer le raccordement « observations empiriques/énoncés protocolaires » autrement qu'en endossant le rôle d'un baron de Münchhausen se tirant seul du marécage analytique de la syntaxe linéaire et se déposant, sain et sauf, sur le terrain concret des énoncés testables empiriquement ou vice versa. Par ailleurs, la formalisation de la syntaxe risque, comme le souligne F. Rivenc, d'être "une tâche toujours future, à moins que quelque moyen se présente de trancher ce nœud"[77].

Si l'on conjugue la prémisse stipulant que les propositions protocolaires « n'ont pas besoin d'être elles-mêmes confirmées, mais (…) servent de fondement à toutes les autres propositions de la science »[78] avec la thèse qui sous-tend le solipsisme méthodologique - « les protocoles forment la base de la construction d'ensemble de la science, dans la mesure où chaque sujet ne peut prendre pour base que son propre protocole »[79] - il devient difficile dès lors d'expliquer comment faire tenir le solide édifice des énoncés qui entrent dans la composition d'une théorie scientifique sur la « pointe » du protocole privé de chaque sujet. Pour le dire autrement, il s'agit d'insister ici sur l'interrogation suivante (considérée comme un pseudo-problème[80] par Carnap) : « comment concilier la relation de déduction entre la langue de la physique et la langue protocolaire avec le fait que la première parle d'états de choses physiques et la seconde de contenus d'expérience vécue ?»[81]

La véritable question est : peut-on donc délimiter avec précision le seuil inférieur d'une propriété observable dans l'immédiateté de la perception? Et surtout comment

---

[76] *Ibid.*
[77] F. Rivenc, (1993), *Recherches sur l'universalisme logique*. Paris: Payot, p. 206.
[78] (R. Carnap, 1936/2006: 329)
[79] *Ibid.*, p. 357.
[80] Comme réponse à cette difficulté, Carnap fait appel à un tour de passe-passe linguistique : « ces pseudo-questions sont automatiquement éliminées par l'emploi du mode formel du discours ». (*Ibid.*, p. 350)
[81] *Ibid.*

faut-il gérer la "traversée" des niveaux cognitifs, depuis la subjectivité privée d'une perception « maintenant rouge » à la reconstruction rationnelle d'un langage physique lisible par et pour un auditoire expert?

Avec cette question, nous retrouvons, bien que sous une forme masquée, le cœur de la problématique impliquant la prisonnière Mary, héroïne de l'argument antiphysicaliste de F. Jackson. Cependant, malgré l'échec du fameux remplacement du mode contentuel du discours (qui traite de contenus d'expérience vécue) par un mode formel (qui traite de « protocoles » et d'« états de choses physiques ») cette formulation a le grand mérite de présager les impasses vers lesquelles se dirigeront les débats qui animent la philosophie de l'esprit contemporaine : « lorsqu'on use du mode contentuel du discours - écrit Carnap - on peut éviter les contradictions en l'utilisant de façon moniste : soit on ne parle que des "contenus d'expérience vécue" - au sens du solipsisme - soit on ne parle que d'"états de choses physiques" - au sens du matérialisme. Mais si on parle de façon dualiste – comme c'est presque toujours le cas en philosophie – de "contenus d'expérience vécue" et également "d'états de choses physiques" (de "corps" et d'"esprit", de "corps" et d'"âme", de "psychique" et de "physique", d' "actes de conscience" et d' "objets intentionnels"), alors les contradictions sont inévitables »[82]. Nous retrouvons ici le cœur des litiges dans l'affaire des rapports corps-esprit[83].

Résumons, encore une fois, au risque de lasser le lecteur, l'affaire des énoncés protocolaires en prenant comme fil conducteur la question que nous suivons ici, à savoir celle des « jointures mystérieuses » qui interviennent dans la médiation entre l'expérience phénoménale, subjective, et sa formulation objective, à la troisième personne.

Dans les deux dernières sections de ce chapitre nous avons tenté de dégager le point suivant : l'attitude initiale de Carnap visant à réduire ou à retranscrire tous les énoncés scientifiques dans un langage purement formel, abritait - nous semble-t-il - la

[82] *Ibid.*, p. 351.

[83] Si les considérations qui viennent d'être esquissées ont un sens - bien qu'il faille les préciser beaucoup plus que nous ne l'avons pas fait ici - on peut se demander s'il existe, après tout, une différence fondamentale entre "le vieux point de départ" ou " l'absurde" *cogito* cartésien et l'insaisissable "*expérience vécue"* carnapienne, les deux considérés dans leur position délicate de candidats au poste du meilleur "fondement ultime de tout savoir". Un parallèle pourrait être envisagé également entre <ce qui est présentement perçu, vécu>, en tant que point de départ "vierge" des empiristes, d'une part, et le *cogito*, en tant que point de départ "vierge" des rationalistes, d'autre part. Dans les deux cas, la qualité de « base » demeure la possibilité de détacher l'objectivité des énoncés scientifiques du marécage des expériences subjectives, privées.

conviction profonde, nourrie probablement tout au long de sa carrière philosophique, qu'il serait possible d'élargir l'applicabilité des privilèges typiques des propositions "pures", analytiques aux assertions édifiées sur des propositions de lisière, dites "protocolaires" qui, elles, auraient la difficile tâche de "heurter" le domaine, par excellence incertain et aléatoire, du réel. L'idée que nous aimerions retenir dans l'affaire des énoncés protocolaires est la suivante : lorsque les empiristes logiques confrontent un *énoncé* à un *fait*, ils procèdent comme si ces deux entités ontologiquement objectives seraient les acteurs d'un éventuel affrontement face-à-face et « bloc contre bloc » décisif. Si l'énoncé passe le test de vérification, alors il devient un « membre » légitime du mobilier du monde. Et le sujet connaissant dans tout ça ? Il est tacitement extrait du tableau et investi du statut privilégié de narrateur « voix-off », installé en surplomb.

Par conséquent, le tableau hérite d'un niveau supplémentaire de discours, que l'on peut appeler « métaformel » ou « métadescriptif ».

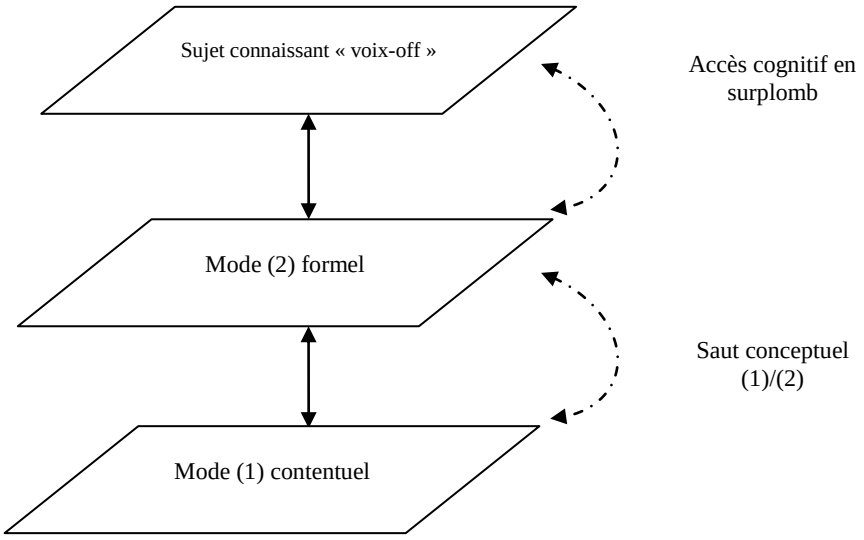


Figure 6: Un deuxième exemple de saut conceptuel

Une fois le saut conceptuel opéré depuis le mode contentuel du discours (insatisfaisant, car susceptible d'ouvrir la boîte de Pandore des expériences subjectives) vers le mode formel du discours (préférable et sécurisant, car il expulse les contenus d'expérience privée hors du cadre formel des énoncés), la tâche du philosophe – désormais débarrassé du problème encombrant des entités extralinguistiques – sera celle de garantir la validité formelle des enchaînement de propositions impliquées dans

la construction de plus en plus sophistiquée des théories scientifiques. Ce schéma montre que pour établir une opération impliquant des propositions « mutuellement déductibles » glissant du mode [1] vers le mode [2] il faut avoir déjà adopté tacitement une hypothèse auxiliaire tacite stipulant l'accès métaformel de l'observateur narrateur (en occurrence ici, Rudolf Carnap) à une *description neutre de l'état corporel de S*, c'est-à-dire endosser un point de vue situé *en dehors* des cadres initiaux. Nous sommes ici en présence d'un excellent exemple d' « illusion d'exterritorialité »[84] selon l'heureuse expression de M. Bitbol.

Récapitulons. Le présent chapitre a été organisé autour d'une brève rétrospective destinée à mettre en lumière les racines profondes qui traversent et soutiennent, telle une colonne vertébrale invisible, la pensée analytique. Dans les sections 2.1. et 2.2. nous avons vu que le rejet de toute référence mentaliste à la psychologie du sujet, jugée trop « perturbatrice » aux yeux d'un certain idéal d'objectivité hérité depuis Platon, oblige Frege à faire appel à un saut conceptuel mystérieux, intercalé entre les pensées objectives et les contenus d'expérience subjective, saut conceptuel opéré *via* le processus de « saisie d'une pensée ». Dans les sections 2.3. et 2.4., nous avons vu que la fondation des énoncés protocolaires sur l'observation idéalement neutre, en surplomb et « à la troisième personne », témoigne d'une tension qui se déploie sur l'axe traditionnel subjectivité/objectivité. Cependant – et c'est ici le point sur lequel nous aimerions insister – la distribution sur cet axe n'aurait pas pu se faire sans l'intervention tacite du « troisième œil » d'un narrateur 'voix-off' censé placer le fauteuil de l'observateur en surplomb, à la fois *à l'extérieur de ses propres états mentaux* et *à l'intérieur des états mentaux des autres*. A la question « peut-on formuler la syntaxe d'un langage, comme le voulait Carnap, depuis l'intérieur, sans être forcé de l'attacher dans une prolifération infinie de métalangages, dont la source se trouve ancrée, pour ainsi dire, dans le contenu d'expérience vécue de chaque locuteur pris individuellement ? », la réponse semble négative.

Le rejet de toute référence mentaliste à la psychologie du sujet aura des conséquences décisives sur la manière dont les philosophes héritiers du courant analytique aborderont le problème de l'esprit. Nous tendrons, en ce qui nous concerne, à conclure provisoirement que l'ambition carnapienne de refondation rationnelle de la connaissance doit être perçue dans la *continuité* des métaphysiques traditionnelles, et

[84] (M. Bitbol, 2010:43)

cela malgré les efforts soutenus de Carnap de se détacher de cette filiation. Nous aurons l'occasion de revenir plus amplement dans le chapitre 7 sur la querelle qu'a opposé Quine à Carnap au sujet du statut des entités abstraites.

Deux aspects sont à relever ici : (1) premièrement, l'adhésion à une polarité conceptuelle rigide sur l'axe subjectivité/objectivité contamine la caractérisation des rapports entre les contenus d'expérience vécue et les faits physiques. Ce clivage – et c'est ici le point sur lequel nous aimerions attirer l'attention du lecteur – repose tacitement sur l'intervention d'un « troisième œil », censé garantir à l'observateur le privilège d'investir un espace méta-narratif neutre, de type "voix-off". Ainsi scindé, le sujet sera contraint de jongler selon les besoins théoriques de l'argumentation, tiraillé entre la subjectivité de la première personne et l'objectivité de la troisième.

(2) Deuxièmement, si nous avons insisté ici sur les prétentions, rapidement désavouées par la suite, du physicalisme sémantique de fournir un schéma unificateur purement formel, i.e. objectif du réel, c'est dans le but de mettre en relief la naissance d'une véritable science dédiée à l'étude de l'esprit. D'une certaine manière, cela ne pouvait se réaliser qu'en réaction aux échecs enregistrés par la philosophie analytique dans ses premières tentatives de fournir un modèle explicatif de l'esprit amputé de toute référence aux états psychologiques de l'être pensant. Il nous semble qu'en un sens, éclairer cette filiation est une question cruciale aujourd'hui.

Ces éléments d'arrière-plan étant fixés, il est temps à présent de descendre sur le terrain et d'amorcer le travail d'investigation proprement dit autour des modèles explicatifs édifiés sur l'axe mental/physique.

3 Le mental et le physique

La distinction corps/esprit n'a pas d'histoire. Mais ce qu'on appelle communément le *'mind-body problem'* dans la tradition analytique fut explicitement énoncé dans un très influent article de H. Feigl paru en 1958, et affichant le significatif titre de "The 'mental' and the 'physical'"[1]. Beaucoup de scientifiques, note Feigl[2], essaient désespérément de renvoyer le problème des rapports entre le corps et l'esprit dans les limbes de la métaphysique spéculative. Après avoir tenté, avec plus ou moins succès, de résoudre le puzzle d'une façon ou d'une autre, ces auteurs, note Feigl, finissent habituellement par choisir une des options suivantes : soit ils abandonnent le problème aux soins des philosophes, soit ils le déclarent tout bonnement un pseudo-problème. Selon Feigl, cette stratégie de l'autruche ne fait qu'augmenter en flèche les difficultés auxquelles se voient confrontés les scientifiques dans leur travail de recherche et d'explication du réel. Pour ne donner qu'un exemple, la simple question de la définition de la psychologie donne naissance à une importante controverse : est-ce que l'objet d'étude de la psychologie est l'expérience mentale ou le comportement seul ?

Selon l'excellent résumé qu'en donne H. Feigl, presque toutes les perplexités qui naissent autour du problème corps/esprit s'organisent autour des polarités suivantes :

Mental	**Physique**
Subjectif (privé)	Objectif (public)
Non-spatial	Spatial
Qualitatif	Quantitatif
Holistique	Atomistique
Emergeant	Compositionnel
Intentionnel	« Aveugle » (non-intentionnel)
Libre arbitre	Mécanique

Figure 7: Le mental et le physique

[1] H. Feigl, (1958), 'The "mental" and the "physical"', in H. Feigl, M. Scriven et G. Maxwell, (eds.) *Concepts, Theories and the Mind-Body Problem: Minnesota Studies in the Philosophy of Science, vol. 2*, Minnesota University Press, Minneapolis, p. 370. [Republié in 1967 en tant que livre avec un *Postscript*, une préface au Postscript et une bibliographie additionnelle.]

[2] (H. Feigl, 1958)

De manière générale, la question des rapports entre le mental et le physique vient se greffer sur le tronc de l'ancienne dichotomie corps/esprit, et prend la forme d'une interrogation visant à expliciter les relations entre les *propriétés/concepts/prédicats* mentaux, d'une part, et les *propriétés/concepts/prédicats* physiques d'autre part. Selon K. Ludwig[3] ces relations peuvent prendre les formes suivantes (où *M* représente les prédicats mentaux, et *P* les prédicats physiques) :

[A] Pour tous les *X*, si *X* est *P*, alors *X* est *M*.

[B] Pour tous les *X*, si *X* est *M*, alors *X* est *P*.

[C] pour tous les *X*, *X* est *M* si et seulement si *X* est *P*.

Bien que [C] soit en fait la conjonction entre [A] et [B], il est important de pouvoir les examiner séparément, car les fondements et preuves qui sous-tendent la vérité des assertions [A]–[C] sont déterminants dans la caractérisation des relations entre le mental et le physique. Pour le dire très succinctement, les diverses tentatives de réponse s'organisent autour de possibles attitudes que l'on adopte par rapport à deux éléments clé : la réduction et la thèse du monisme matérialiste.

- une explication *réductionniste* du mental aspire à définir celui-ci en faisant appel à des termes et principes exclusivement physiques.
- une explication *matérialiste* (ou physicaliste) du mental soutient que ce dernier n'est, à la base, qu'une propriété/concept/prédicat physique.
- une explication *non-matérialiste* (ou non-physicaliste) consiste à dire que les phénomènes mentaux, bien qu'associés de près aux phénomènes physiques, sont fondamentalement non-physiques.
- une explication *non-réductionniste* adopte la thèse selon laquelle le mental en général, et le phénomène de la conscience plus particulièrement, ne peuvent pas être réduits aux propriétés physiques.

Conséquemment, la question des rapports entre les états mentaux et les états physiques divise le matérialisme en trois filières distinctes. (1) Le matérialisme *réductionniste* soutient que tout phénomène, état ou processus mental peut être (du moins en principe) réduit à une description utilisant le vocabulaire de la physique, les disciplines du niveau supérieur (« higher-level ») étant réductibles à celles de niveau

[3] K. Ludwig, (2003) '*The Mind-Body Problem: an Overview*', *in* S. Stich & T. Warfield (eds.), *The Blackwell Guide to Philosophy of Mind*, Blackwell Publishing, p. 7.

inférieur (« lower-level ») *via* les lois de la physique. (2) D'autres soutiennent une forme radicale *d'éliminativisme*[4], affirmant que le discours traditionnel employant des termes liés aux états mentaux et à la psychologie du sens commun sera peu à peu éliminé au profit d'un discours inspiré du vocabulaire des neurosciences. (3) D'autres adhèrent ouvertement à des formes plus ou moins radicales de dualisme. (4) Enfin, une quatrième voie, beaucoup plus intéressante, et qui fera l'objet de notre enquête ici, semble s'ouvrir du côté des versions *non-réductionnistes*. Ses adeptes, qui gagnent de plus en plus de terrain dans la philosophie actuelle, affirment que le discours employant des concepts mentaux est non seulement cohérent et intelligible, mais aussi parfaitement compatible avec le matérialisme, et cela sans faire appel à des amputations réductionnistes appauvrissant le domaine du mental.

Dans ce qui suit, nous allons dresser un bref état des lieux établissant les principaux repères autour desquels s'organisent les débats visant à tirer au clair la place de l'esprit dans la nature. Le compte rendu que nous allons présenter ici ne peut, bien évidemment, qu'effleurer le complexe ensemble d'interrogations, argumentations et difficultés qui accompagne la question des rapports entre le mental et le physique. La section 3.1. présente très rapidement l'impasse dualiste. Nous bifurquerons ensuite, dans la section 3.2. vers le cœur même du matérialisme, en présentant la théorie de l'identité psychophysique. Enfin, dans la section 3.3. une présentation d'ensemble de la « révolution » fonctionnaliste complétera le compte-rendu de cet état des lieux du 'mind-body problem'.

3.1 L'impasse dualiste

Selon la version la plus naïve du point de vue dualiste-interactionniste, l'esprit et le corps sont deux substances ontologiquement distinctes, possédant deux attributs spécifiques – la pensée et l'étendue – qui communiquent par le truchement de la célèbre « glande pinéale ». Le dualisme est clairement formulé et bâti sur un des arguments les plus célèbres de l'histoire de la philosophie, le *doute cartésien* :

[4] Voir P.M. Churchland, (1981), 'Eliminative Materialism and the Propositional Attitudes', *Journal of Philosophy* 78: 67-90; P. Churchland, (1989), *The Nature of Mind and the Structure of Science - a Neurocomputational Perspective.* Cambridge Massachusetts: MIT Press. Voir aussi P.S. Churchland, (1986), *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Cambridge, MA: MIT Press.

« *Je ne sais pas* – écrit Descartes de sa plume un rien provocatrice – *si je dois vous entretenir des premières méditations que j'y ai faites, car elles sont si métaphysiques et si peu communes qu'elles ne seront peut-être pas au goût de tout le monde. […] Ainsi, à cause que nos sens nous trompent quelque fois, je voulus supposer qu'il n'y avait aucune chose qui fût telle qu'ils nous la font imaginer ; et pource qu'il y a des hommes qui se méprennent en raisonnant, même touchant les plus simples matières de géométrie, et y font des paralogismes, jugeant que j'étais sujet à faillir autant qu'aucun autre, je rejetai comme fausses toutes les raisons que j'avais prises auparavant pour démonstrations ; et enfin, considérant que toutes les mêmes pensées que nous avons étant éveillés nous peuvent aussi venir quand nous dormons sans qu'il y en ait aucune pour lors que soit vraie, je me résolus de feindre que toutes les choses qui m'étaient jamais entrées en l'esprit, n'étaient plus vraies que les illusions de mes songes. Mais aussitôt après je pris garde que, pendant que je voulais ainsi penser que tout était faux, il fallait nécessairement que moi, qui le pensais, fusse quelque chose : et remarquant que cette vérité, je pense, donc je suis, était si ferme et si assurée que toutes les plus extravagantes suppositions des sceptiques n'étaient pas capables de l'ébranler, je jugeai que je pouvais la recevoir sans scrupule pour le premier principe de la philosophie que je cherchais* »[5].

Rhabillé avec des vêtements modernes, l'argument du doute cartésien a été récemment reformulé par P. Carruthers[6] ainsi:

(P1) Je peux douter du fait que mon corps/cerveau existe.

(P2) Je ne peux pas douter du fait que *moi*, *j*'existe.

DOUTE CARTESIEN

Par conséquent,

(C) Je ne suis pas mon corps/cerveau. Bref, je ne suis pas une chose physique.

[5] R. Descartes, *Discours de la méthode*, Librairie Générale Français, Paris, 1973, pp. 127-128.

[6] "(1) I can doubt whether my body and brain exist. (2) I cannot doubt whether *I* exist. (C) Therefore I am not my body and/or brain – I am a non-physical thing." (*Ibid.*, p. 39) (Nous traduisons)

Le point névralgique de ce raisonnement ne tient pas à la plausibilité des prémisses (P1) et (P2), mais à la possibilité d'inférer correctement la conclusion (C) à partir de la conjonction des énoncés (P1) et (P2). Pour saisir cela, dit Carruthers[7], il suffit de considérer la comparaison suivante avec un argument respectant une structure analogue et prenant comme exemple le raisonnement suivant mis dans la bouche d'Œdipe :

OEDIPE

P1' Je peux douter du fait que ma mère existe encore.

P2' Je ne peux pas douter du fait que Jocaste existe maintenant.

Par conséquent,

(C') Jocaste n'est pas ma mère.

Dans cet exemple, l'inconsistance de (C') devient plus facile à saisir. L'argument du doute cartésien ayant déjà fait l'objet de nombreux examens critiques, nous allons retenir ici un seul exemple, avancé par D. Braddon-Mitchell et F. Jackson[8]. Selon ces auteurs, déterminer ce qui est possible et ce qui est impossible, constitue un acte qui dépend de la manière dont l'objet du doute possible est *décrit*, autrement dit, c'est une affaire de simple description. Les partisans du matérialisme pourraient répondre à Descartes ainsi : tout ce que l'argument du doute prouve c'est que je ne peux pas douter de mon existence considérée sous la description « chose pensante », alors que je peux en douter sous la description « corps physique » ; de façon analogue, je ne peux pas douter de l'existence des doutes eux-mêmes prises sous la description « doutes », alors que je peux douter d'eux sous la description « états cérébraux ».

Une des raisons majeures de cette impopularité réside dans l'obligation d'intercaler, à une jointure cruciale de l'explication des capacités cognitives ou perceptuelles de l'être humain, une notice dont la teneur serait « ici s'accomplit un

[7] "(1) I can doubt whether my mother still exists. (2) I cannot doubt whether Jocasta now exists. (C) Therefore Jocasta is not my mother." (P. Carruthers 2004:39). (Nous traduisons)

[8] "Returning to Descartes, the point is that what it is possible and what it is not possible, to doubt depend on how the subject of the possible doubt is described Materialist can argue that all that Descartes's argument shows is that you cannot doubt your existence under the description 'thinking thing' but can under the description 'physical body', and likewise that you cannot doubt that there are doubts under the description 'doubts' but can under the description 'brain states'". D. Braddon-Mitchell & F. Jackson, (1996), *Philosophy of Mind and Cognition*, op. cit., pp. 7-8.

Grand Mystère»[9]. Cette situation est certes, troublante, et aucun philosophe ne peut songer sérieusement à abdiquer devant un tel « mystère ». Cela dit, il serait erroné de croire que réfuter le dualisme cartésien épuise le problème des rapports corps/esprit. En réalité, il ne représente que la pointe de l'iceberg. Aujourd'hui, l'accord général concernant la faillite du dualisme semble définitivement acquis. D'autre part, nous verrons que le dualisme est loin d'être le seul courant de pensée amené à intercaler de telles « *jointures mystérieuses* » reliant différents éléments dans un dispositif argumentatif. Au contraire – et c'est ici une des thèses que nous nous emploierons à prouver ici – il se peut que des « raccordements » non moins mystérieux viennent s'insérer à des différents niveaux dans l'architecture argumentative des théories matérialistes.

La critique du cartésianisme représente un fort commode point de départ de toute entreprise matérialiste : "Aujourd'hui - écrivait M. Schlick en 1934 - la logique nous a suffisamment ouvert les yeux sur le *cogito* : nous savons que c'est une simple pseudo-proposition, qui ne devient pas davantage un énoncé véritable si on l'énonce sous la forme *cogitatio* est, « les contenus d'existence existent ». Une telle proposition qui n'exprime rien par elle même ne peut en aucun sens servir de fondement à quoi que ce soit"[10].

Il serait trop long, et d'ailleurs superflu de reprendre ici en détail les disputes[11] qui ont opposé dualistes et antidualistes. On résumera seulement à gros traits quelques uns de ses aspects clé. On s'accorde communément à dire que l'ouvrage de G. Ryle[12] *The Concept of Mind* a joué un rôle crucial dans l'offensive antidualiste, établissant un

---

[9] D. Pinkas, (1995), *La matérialité de l'esprit*, Paris, Ed. la Découverte, p. 8.

[10] M. Schlick, (1934), 'Über das Fundament der Erkenntnis', *Erkenntnis* 4, pp. 79-99, trad. fr. 'Sur le fondement de la connaissance', in C. Bonnet et P. Wagner, (eds.), (2006), *L'âge d'or de l'empirisme logique*. Paris, Gallimard, p. 429. (Nous utilisons ici cette dernière édition).

[11] Parmi les objections antidualistes les plus « classiques » on peut ranger l'argument du dormeur : si l'essence de l'esprit est le fait d'être conscient, que se passe-t-il pendant le sommeil ? Il est difficile d'admettre que l'esprit *cesse d'exister* pendant le sommeil pour *renaître* au réveil. Comme on peut facilement le deviner, cet argument est incapable, à lui tout seul, de réfuter le dualisme. Selon P. Carruthers on peut comptabiliser au moins trois possibles contre-arguments : (i) premièrement, on peut choisir d'affirmer que ce qu'on appelle « sommeil » n'est qu'une forme faible d'un état conscient, si faible qu'il devient difficile d'en garder le souvenir exact ; (ii) une seconde option consiste à soutenir que ce qui est véritablement important n'est pas la vie mentale consciente, mais la vie mentale tout court, qu'elle soit consciente ou inconsciente, en s'appuyant dans ce cas sur les écrits de Freud concernant les émotions, désirs, etc. inconscients ; (iii) enfin, on peut admettre que l'esprit connaît des périodes intermittentes d'enchaînement <conscience – inconscience>. Pour que le dualisme soit recevable, il suffit d'accepter qu'un seul et même esprit survit et traverse ces enchaînements. (P. Carruthers, 2004: 50-52)

[12] G. Ryle, (1949), *The Concept of Mind*, Hutchinson Publ., London. Trad. fr. par S. Stern-Gillet, *Le concept d'esprit*, Paris, Payot, 1978.

véritable programme de démolition des « fantômes » cartésiens. En prenant comme point de départ le célèbre argument wittgensteinien du langage privé, Ryle décide de viser plus haut, et de s'attaquer au concept même d' « esprit ». Selon lui, les philosophes sont victimes d'une erreur catégorielle[13], erreur qui peut être résumée *grosso modo* ainsi : on commence habituellement avec l'idée que les esprits sont des entités distinctes, quoique semblables aux corps, mais lorsque l'on s'aperçoit qu'ils sont introuvables, pour ainsi dire, dans le monde physique, on finit par leur inventer un domaine à part, un royaume de l'immatériel. L'esprit devient ainsi, selon la célèbre expression de Ryle, un *fantôme dans la machine*[14].

Venant de la part d'un des personnages les plus influents de la philosophie britannique du milieu du siècle dernier, la critique de Ryle a eu des effets considérables sur la façon dont la tradition analytique entendra s'emparer du problème corps-esprit. Il est vrai que dans sa croisade contre le mythe du « fantôme dans la machine » Ryle s'attaque plutôt au « fantôme » qu'à la « machine », et cela avec des conséquences inestimables pour la philosophie de l'esprit du XXe siècle. En dépit de cela, *The Concept of Mind* apporte un message philosophique essentiel dans la mesure où il parvient à canaliser l'attention des philosophes sur un problème que Ryle prétendait, paradoxalement, effacer de la carte des enquêtes légitimes de la philosophie, à savoir le 'mind-body problem'.

Le dualisme interactionniste souffre d'emblée d'un défaut majeur : il déclenche l'inflation de ce que W. Lycan appelle « ontological excrescences »[15] (« excroissances ontologiques »), provoquant ainsi une dangereuse prolifération des niveaux de réalité. Par contraste, l'ontologie matérialiste se révèle nettement plus économe. Mais le véritable talon d'Achille du dualisme se trouve dans le problème que pose *l'interaction causale psychophysique*[16]. Comme le résume fort bien J. Fodor[17], si l'esprit est non-physique, il s'ensuit qu'il n'occupe aucune étendue ou position dans l'espace. Mais si tel est le cas, la question qui se pose est la suivante : comment un état non-physique

---

[13] "It represents the facts of mental life as if they belonged to one logical type or category… when actually they belong to another" (*Ibid.*, p.16).

[14] "The theoretically interesting category mistakes are those made by people who are perfectly competent to apply concepts, at least in the situations with which they are familiar, but are still liable in their abstract thinking to allocate those concepts to logical types to which they do not belong." (*Ibid.*, p. 17)

[15] W. Lycan, (1987), *Consciousness*, Cambridge Massachusetts: MIT Press, p. 2.

[16] *Ibid*.

[17] J. Fodor, (1981), 'The Mind-Body Problem', *Scientific American* 244. Réédité *in* J. Heil (ed.) (2004), *Philosophy of Mind - a Guide and Anthology*, Oxford University Press, p. 169. (Nous faisons référence ici à cette dernière édition)

pourrait-il interagir causalement avec un état physique *sans* violer les lois de la nature concernant la conservation de la masse, de l'énergie, etc. ?

Autrement dit, comment réconcilier le célèbre principe de clôture physique avec l'observation familière et de bon sens selon laquelle c'est précisément mon *désir* de manger des fraises qui provoque, *via* un enchaînement de processus chimiques et neurophysiologiques sophistiqués, le mouvement de mon bras vers le panier de fruits qui se trouve sur la table ?

Les débats qui se cristallisent autour du problème de l'interaction causale psychophysique ont fini par révéler une importante brèche dans l'édifice dualiste. Sans vouloir insister sur un sujet qui a été déjà abondamment discuté dans la littérature analytique, signalons brièvement que le camp des dualistes a riposté en avançant différentes solutions. En premier lieu, une façon commode de répondre aux attaques antidualistes est de renvoyer en quelque sorte la balle dans le camp des matérialistes, et de rappeler qu'un examen attentif de l'histoire des sciences qu'on appelle aujourd'hui « exactes » ou « dures », nous permet d'identifier l'utilisation de plusieurs éléments inexplicables et mystérieux, insérés à des étapes clé dans l'élaboration d'une théorie. Bien souvent, les savants se contentent de prendre certains éléments pour des faits bruts. La conclusion naturelle qui semble s'imposer dans ce cas c'est que le problème de l'interaction causale de type mental/physique n'est pas plus obscur que celui de l'interaction physique/physique. Après tout, il est loin d'être clair comment un événement physique (disons, la force de la gravitation) agit et cause un autre événement physique (la chute des graves). A cela, les matérialistes répondent que pour établir une relation causale entre deux événements *e* et *f* il suffit de montrer que *e* et *f* appartiennent à deux *types* d'événements *E* et *F* dont l'interaction causale est gouvernée par une loi universelle de type : « Tout événement *E* est accompagné d'un événement *F* ». Or, cette universalité dans l'application des lois physiques semble faire défaut dans le cas de l'interaction psychophysique. En effet, comme D. Davidson[18] l'a déjà montré avec son monisme anomal, il n'existe pas de lois exactes gouvernant la causalité mentale, mais seulement des généralisations de type *ceteris paribus.* Enfin, le dualisme semble incompatible avec les méthodes et les résultats obtenus par la psychologie

[18] D. Davidson, (1970), 'Mental Events', in D. Davidson, *Essays on Actions and Events*, Oxford: Oxford University Press, 207–223.

expérimentale[19]. Pour clore cette brève section dédiée au dualisme, on peut dire qu'après un long règne qui remonte au moins jusqu'au XVIIe siècle, le dualisme entre en « collision », selon l'heureuse expression de *W.* Lycan[20], avec le vérificationnisme triomphant du milieu du siècle dernier

Nous allons donc passer en revue les principales versions du matérialisme contemporain. Cette mise en perspective nous permettra de réunir les divers fils conducteurs qui nous mèneront à regarder de plus près le rôle de l'argument de la connaissance sur le matérialisme contemporain. La question que l'on sera inévitablement amené à se poser concerne les stratégies réductives du mental au physique. Nous verrons ainsi que le matérialisme, du moins dans sa version standard, manque sa cible principale qu'est la réduction du mental au physique. Cette faillite sera amplement discutée dans la deuxième partie de notre enquête, dédiée au célèbre argument de la connaissance de Franck Jackson.

3.2 La douleur en héritage : du behaviorisme au matérialisme réductionniste

Imaginez-vous à la campagne : vous êtes dans un jardin, au bord de la Méditerranée, à l'ombre d'un mûrier. Les cigales remplissent vos oreilles de l'exquis bruit de l'été. Vous êtes en train de diner avec des amis lorsque soudain vous ressentez une forte douleur dans l'oreille droite : vue de l'extérieur, votre expérience « interne » demeure inobservable aux yeux de vos voisins de table qui continuent de manger et de boire allégrement. En revanche, si vous vous mettez à grimacer, à vous couvrir l'oreille avec la main, à gémir, etc., tous ces éléments qui caractérisent votre comportement constituent tout autant d'indices qui signalent à vos convives la présence d'un événement particulièrement désagréable qui a lieu à l'intérieur de votre corps. Si vous accompagnez vos grimaces par des grognements suivis de l'énoncé « Aïe! J'ai mal à l'oreille! », votre voisin de table, s'il est behavioriste, peut dériver en toute confiance, la conclusion suivante : "Michel a mal à l'oreille droite", corroborant ainsi la célèbre

[19] "If mental processes were different in kind from physical processes, there would be no reason to expect these methods to work in the realm of the mental. In order to justify their experimental methods many psychologists urgently sought an alternative to dualism". (J. Fodor, 1981/2004: 169)
[20] (W. Lycan 1987:1)

formule de J.B. Watson[21] selon laquelle ce que les psychologues appelaient à une époque « pensée » n'est en réalité rien d'autre qu'un langage ou un dialogue entre moi et moi-même.

Selon le résumé qu'en donne G. Graham[22], le behaviorisme se bâtit sur trois thèses principales : (1) d'abord il y a l'idée que la psychologie est la science du comportement, et non pas la science de l'esprit. (2) Deuxièmement, le comportement peut être décrit sans faire aucune référence aux états mentaux ou à tout autre processus psychologique interne. Les sources du comportement sont externes (dans l'environnement) et non pas internes (dans l'esprit). (3) Enfin, si une théorie de la psychologie développementale utilise des termes mentaux, alors ou bien (a) ces termes ou concepts doivent être remplacés ou éliminés par des concepts behavioristes, ou bien (b) ils peuvent être traduits en termes comportementaux.

En prenant appui sur ces trois thèses, on peut dès à présent esquisser le portrait-robot de l'argument en faveur du behaviorisme qui peut être résumé, selon Pinkas, en cinq étapes : (i) la fertilité de la science consiste sa capacité prédictive ; (ii) pour établir qu'un événement prédit a eu lieu, il faut qu'il soit publiquement observable ; (iii) les seules propriétés ou événements humains qui soient publiquement observables sont les comportements ; (iv) donc, dans le cas des êtres humains, les seules prédictions possibles concernent les propriétés ou événements comportementaux ; (v) les prédictions seront d'autant meilleures qu'elles seront déduites de lois reliant les circonstances publiquement observables au comportement subséquent. Par conséquent, le but central de la psychologie est de découvrir de telles lois.[23]

On s'accorde généralement à distinguer trois types de béhaviorisme[24] : méthodologique, psychologique et analytique (ou philosophique). Considérées indépendamment, chacune des thèses (1)-(3) mentionnées plus haut sert d'appui à une forme particulière de behaviorisme. Le behaviorisme« méthodologique » repose sur la vérité de (1), le behaviorisme « psychologique » suppose la vérité de (2), tandis que le

[21] "What the psychologist have hitherto called thought is in short nothing but talking to ourselves". J.B. Watson, (1930), *Behaviorism*, Norton: New York. Réédité *in* W. Lycan (ed.) *Mind and Cognition – A Reader*: Blackwell Publishing 1990, p. 14. (Nous faisons référence ici à cette dernière édition)
[22] G. Graham, *Behaviorism*: http://plato.stanford.edu/entries/behaviorism
[23] (D. Pinkas, 1995 : 21)
[24] Pour une classification détaillée des diverses formes du behaviorisme voir G. Rey, (1997), *Contemporary Philosophy of Mind: A Contentiously Classical Approach*. Oxford: Blackwell.

behaviorisme « analytique » ou « logique » repose sur la thèse (3). Pour G. Ryle[25], un des principaux inspirateurs et artisans du behaviorisme, être dans un certain état psychologique ne veut pas dire avoir un certain « état interne » non-observable et immatériel, mais plutôt être disposé à se *comporter* de telle ou telle manière. Lorsque nous attribuons à nos interlocuteurs une croyance ou un désir, nous ne soutenons pas qu'ils se trouvent dans un tel ou tel état « interne » ou psychologique inobservable : nous voulons dire tout simplement qu'ils sont *disposés à se comporter* de telle ou telle façon en fonction de certains conditionnements environnementaux. Rien de plus, rien de moins.

De nos jours, le behaviorisme logique compte très peu d'adeptes, et on peut dire sans exagération qu'il ne partage plus le cap d'affiche avec les autres versions du matérialisme[26]. Nous l'avons compris, si l'exigence de vérifiabilité imposée par le positivisme logique et l'attitude antimentaliste du behaviorisme sont étroitement articulées, c'est que la cible est commune : l'élimination de l' « âme » du domaine des études scientifiquement respectables. Cependant, un lecteur attentif de l'ouvrage *Le Concept de l'Esprit* de Ryle (1949), aurait certainement noté que le philosophe britannique ne va pas, ou mieux dit, *ne* peut *pas* aller jusqu'au bout de sa thèse réductionniste radicale qui découle naturellement de son programme. Plus spécifiquement, il finit par concéder trop d'exceptions et d'expressions comme « conversations silencieuses », « soliloques silencieux », « préparer des remarques dans la solitude », « se dire des choses à soi-même *sotto voce* », c'est-à-dire, tout autant de métaphores destinées à cacher tant bien que mal l'indispensabilité des concepts mentaux. Cette omission l'amène également à ignorer, ou du moins à sous-estimer, l'asymétrie entre la connaissance à la première et à la troisième personne.

Enfin, le coup d'arrêt qui ébranla définitivement la position behavioriste fut porté par l'argument des « super-spartiates », i.e., une expérience de pensée ingénieuse et désormais classique, étayée par H. Putnam[27] dans un article où il dresse une critique

[25] G. Ryle, (1949), *The Concept of Mind*, Hutchinson Publ., London. Trad. fr. par S. Stern-Gillet, Paris, Payot, 1978.

[26] La plupart des accusations anti-behavioristes portent sur la stérilité des analyses en termes de vérifiabilité du comportement. La singularité de cette situation a été captée dans une célèbre plaisanterie qui circulait parmi les philosophes à l'époque : deux behavioristes se rencontrent, et l'un d'entre eux interpelle l'autre: « Tu vas bien. Et moi, comment je vais ? » (« You feel fine, how do I feel?»)

[27] H. Putnam, (1965), 'Brains and Behavior' in R. J. Butler, (ed.) *Analytical Philosophy*, Oxford: Blackwell. Réédité in J. Heil, (ed.)(2004), op. cit., p. 97. (Nous nous référons ici à cette dernière édition)

acerbe du behaviorisme logique. Putnam prend comme cible une des thèses behavioristes clé selon laquelle les événements mentaux sont des constructions logiques découlant des événements comportementaux, exactement à la façon dont les nombres sont des constructions logiques issues des classes abstraites. Reconsidérons notre exemple de départ : le fait que Michel crie « aïe ! » lorsqu'il ressent une douleur dans oreille ne nous apprend *rien* d'intéressant sur la nature ou les causes *internes* qui ont provoqué son mal. Aucune relation de nécessité logique forte ne relie le fait « crier 'aïe !'» à l'état « oreille douloureuse ». Les causes (la douleur) ne sont pas de constructions logiques issues de leurs effets (le comportement). Après tout, insiste Putnam[28], on peut fort bien imaginer une société « super-superspartiate » où l'on apprend à ses citoyens de maîtriser dès leur plus jeune âge, les signes extérieurs de douleur. Si le behavioriste logique a raison, cela voudrait tout simplement dire que les enfants « super-superspartiates », une fois devenu adultes, n'éprouvent pas de la douleur lorsqu'une guêpe leur pique l'oreille, par exemple. Chose hautement improbable. La conclusion de Putnam[29] est sans appel : le behaviorisme logique est tout simplement faux. Un état mental comme la douleur n'est pas réductible aux simples critères extérieurs et symptômes observables de la douleur. Bien plus, comme l'ont relevé N. Block et J. Fodor, le behaviorisme ne peut être vrai que si la corrélation mental/comportement est vraie. Or tel n'est pas le cas.[30] En outre, le comportement qu'un organisme est disposé d'adopter à moment *t* est déterminé par un réseau extrêmement complexe impliquant préalablement tout un système de croyances, désirs et connaissances qui agissent comme un ensemble indivisible.

De manière générale, l'échec des stratégies behavioristes a mis en évidence l'urgence dans l'élaboration d'un modèle matérialiste de l'esprit suffisamment flexible et sophistiqué pour qu'il puisse rendre compte du versant « interne » et « invisible » des états mentaux. Il devint alors nécessaire d'édifier une caractérisation de l'esprit capable de concilier les exigences d'une ontologie matérialiste avec l'observation de bon sens selon laquelle la douleur *ne se résume pas* à crier 'aïe !' ou à grimacer, ou à chercher de

[28] "Causes (pains) are not logical constructions out of their effects (behaviour)". H. Putnam, *Ibid.*, p. 101

[29] "Logical behaviourism is a simply mistake. (…) From the statement '*X* has a pain' by itself *no* behavioural statement follows – not even a behavioural statement with a 'normally' or a 'probably' in it ». (H. Putnam 1965/2004: 104)

[30] « Behaviorism cannot be true unless mind/behavior correlationism is true, and mind/behavior correlationism is not true ». N. Block et J. A. Fodor, (1972), 'What Psychological States Are Not', Philosophical Review 81:159-81. Repris in N. Block, (2007): *Consciousness, Function, and Representation: Collected Papers, Volume I*: Bradford Books, p. 45. (Nous nous référons ici à cette dernière édition).

la consolation auprès de ses pairs. Cet échec permit aux philosophes antidualistes de s'engager dans une lutte à la fois contre (i) l'utilisation exclusive de la troisième personne dans l'analyse des états mentaux, et pour (ii) le projet de faire une place à ceux-ci dans une ontologie moniste matérialiste.

3.2.1 La Théorie de l'Identité Psychophysique

Si la douleur n'est pas une simple façade comportementale, qu'est-ce que la douleur *sinon*? La parution en 1956 de l'article programmatique d'U.T. Place 'Is Consciousness a Brain Process?'[31] apporte une réponse à ce problème et marque l'acte de naissance de la Théorie de l'Identité Psychophysique[32], parfois connue sous le nom plus informel de « matérialisme australien », du fait que ses principaux artisans sont originaires du pays des kangourous. Selon Place, l'idée selon laquelle il existerait une classe séparée d'événements, les événements mentaux, qui serait non-énonçable en termes ou concepts utilisés par les sciences physiques, est tout simplement fausse[33]. Réduite à sa plus naïve expression, la Théorie de l'Identité Psychophysique (désormais TIP en abrégé) affirme qu'un état mental comme la douleur, pour reprendre l'exemple devenu paradigmatique, est à identifier avec *l'activation des fibres C* insérées à leur emplacement anatomo-physiologique normal, c'est-à-dire dans le réseau neuronal d'un individu vivant et éveillé. Conséquemment, l'ambition déclarée des partisans de la TIP est de cibler le juste milieu entre les « fantômes » des dualistes et les « machines » comportementales des behavioristes. Nous avons vu que le problème de la « localisation » consiste à trouver littéralement une *place* à l'esprit dans le monde tel qu'il nous est décrit par les sciences physiques.

---

[31] U.T. Place, (1956): 'Is Consciousness a Brain Process?', *British Journal of Psychology*, 47, 44-50. Repris in *W.* Lycan (ed.), *Mind and Cognition – a Reader*, Blackwell Publishers Ltd, 1990, pp. 29-36. Nous utilisons ici cette dernière édition.

[32] Selon les indications de Place, suivi par J.J. Smart, il semblerait que la provenance de l'expression 'théorie de l'identité' n'est pas très éloignée du sens original du terme 'matérialiste' qui désignait un apothicaire (*materia medica*), puisque le syntagme 'théorie de l'identité' fut d'abord employée par le psychologue expérimental E.G. Boring en 1933. Voir J.J. Smart, http://plato.stanford.edu/entries/mind-identity

[33] U.T. Place, (1956/1990: 29)

La TIP propose justement une solution ingénieuse à l'impasse dualiste/behavioriste. Le très influent article de Place[34] étaye l'idée selon laquelle la conscience serait assimilable à un processus neurophysiologique du cerveau. Cette hypothèse doit être traitée et mise à l'épreuve de la vérification comme n'importe quelle autre hypothèse scientifiquement testable. Selon Place, ce qui nous empêche de réaliser cette identification-corrélation, c'est une erreur de jugement qu'il baptise l'«illusion phénoménologique» ("the phenomenological fallacy")[35], et qui consiste à stipuler que lorsqu'un sujet décrit ses propres expériences, ou lorsqu'il décrit la façon dont les choses lui apparaissent dans le toucher, l'odorat, le sentir, etc., il ne fait que décrire littéralement les propriétés des objets projetées en quelque sorte sur un écran interne, à l'aide duquel le sujet perçoit, comme au cinéma, sa propre vie mentale. Or, selon Place, le sens "littéral" d'une description ne peut être donné légitimement que par l'ensemble adéquat de propositions d'une théorie scientifique. Que se passe-t-il quand je vois une tache rouge? Nous croyons connaître les propriétés réelles des objets à travers leurs propriétés phénoménales, alors que c'est en fait le contraire qui est vrai. Une fois l'erreur phénoménologique écartée, on réalise que rien de ce qu'un individu puisse décrire au sujet de ses expériences conscientes n'est incompatible avec ce que les physiologistes[36]ont à dire concernant le même sujet.

Un autre modèle de l'identité psychophysique a été proposé par J.J.C. Smart[37]. Dans l'article 'Sensations and Brain Processes', Smart défend une version connue sous le nom de 'Central State Identity Theory ' (CSIT). Un individu n'est rien d'autre qu'un vaste et sophistiqué arrangement de particules physiques et aucun état additionnel de type « sensation » ou « conscience » n'est à ajouter à cette ossature fondamentale. Bref, rien n'existe au-delà des frontières régies par les sciences physiques. Pour Smart, la douleur est identique à l'état cérébral d'une stimulation des fibres-C exactement à la façon dont l'éclair est identique à une décharge électrique[38]. Les sensations, pour revenir au sujet qui nous préoccupe ici, n'ont pas d'existence propre; il y a des transats,

[34] *Ibid*.
[35] "The mistake of supposing that when the subject describes his experience, when he describes how things look, sound, smell, taste, or feel to him, he is describing the literal properties of objects and events on a peculiar sort of internal cinema or television screen, usually referred to in the modern psychological literature as the 'phenomenal field'. (U.T. Place, 1956/1990: 35)
[36] *Ibid.*, pp. 35-6.
[37] J.J.C. Smart, (1959), 'Sensations and Brain Processes', *Philosophical Review*, 68, 141-156. Repris in J. Heil (ed.) *Philosophy of Mind – a Guide and Anthology*, Oxford: Oxford University Press. (Nous faisons référence ici à cette dernière édition)
[38] "When I say that a sensation is a brain process or that lightning is an electric discharge, I am using 'is' in the sense of strict identity". (J.J.C. Smart, 1959/2004:119)

des chats, des fraises, bref, des arrangements complexes de particules physiques, mais pas d'autre niveau ajouté à celui-ci qui serait celui des états de conscience. En revanche, on peut dépeindre le comportement de ces vastes mécanismes, y compris pour dire qu'ils ont une disposition ou tendance à rapporter l'existence d'objets. En d'autres mots, face à une tache de couleur orangée au mur, le sujet connaissant est « disposé » à dire « J'ai une image différée de couleur jaune et orange ». Mais décrire une image différée, ce n'est pas rendre compte d'un état interne, pas plus que dire « je te méprise », c'est décrire un état intérieur. Il s'agit plutôt d'un comportement qui inclut l'*exercice* de certaines dispositions. Etant donné que la cible fondamentale de la TIP c'est de rendre compte dans un langage non-mentaliste l'information que le sujet rapporte lorsqu'il décrit la sensation de jaune ou de jaune orangé, il devient vital de pouvoir montrer que décrire un état mental tel qu'une « image différée », c'est émettre une proposition du type « Il y a quelque chose qui se passe en moi comme ce qui se passe quand un certain type de stimulus physique est présent » . Et si l'on objecte en soulevant l'exigence de préciser à *quel* événement physique l'événement mental est censé ressembler, Smart répond en introduisant sa célèbre clause de flexibilité, qu'il appellera plus tard la dimension « *topic neutral*»[39] de la TIP. Il s'agit donc de savoir garder une certaine souplesse à l'égard du contenu empirique concret, qui lui permet de s'adapter à la réalité scientifique qui elle, change constamment.

Pour étayer la même hypothèse de l'identité psychophysique, le philosophe australien D.M. Armstrong[40] va jusqu'à dresser un parallèle entre les états mentaux d'un être humain et la description des états d'un thermomètre. Non seulement les états mentaux sont *déterminés* par les états neurophysiologiques du cerveau, mais ils sont réellement *identiques* à ceux-ci :

[39] L'introduction par Smart du terme "topic neutral" est en lien direct avec l'objection suivante: certains antimatérialistes ont noté que même si les sensations sont identiques aux processus neurophysiologiques, les propriétés introspectives des sensations sont elles, non-physiques. A cela Smart répond en affirmant que les propriétés de l'expérience sont 'topic neutral'. Il s'inspire des travaux de G. Ryle, qui utilisait cette notion pour caractériser des termes comme « si », « ou », « et », « parce que ». Si vous lisez une phrase ne contenant que ces mots, vous seriez dans l'impossibilité de déterminer si le contenu se réfère aux mathématiques, à l'histoire, à la géologie, etc. Dans le même sens, affirme Smart, les propriétés introspectives de l'expérience sensorielle sont « topic neutral » entre le physicalisme et le dualisme. Le point central est que selon Smart, le terme « moi » est 'topic neutral', dans la mesure où il désigne l'énonciateur de la proposition en cause. (Voir J.J. Smart, 1959)

[40] D.M. Armstrong, (1968a), *A Materialist Theory of Mind*, London: Routledge and Kegan Paul.

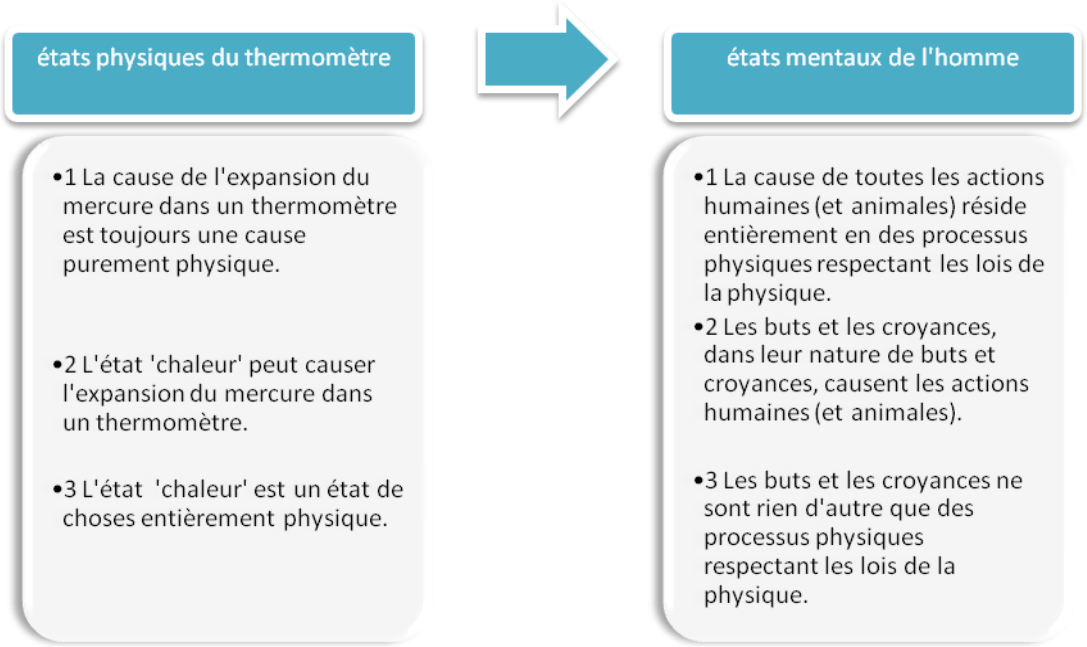


Figure 8: Réduction des états mentaux

Le matérialisme défendu par Armstrong[41] est radical : le philosophe australien s'appuie sur un raisonnement à deux volets, dont la première partie (A) est logique – et inspirée par le béhaviorisme ryléen – et la seconde partie (B) est empirique, correspondant à la CSIT : le premier volet (A) consiste à dire qu'être un état mental, par définition, *c'est occuper un rôle causal*. Le deuxième volet (B) affirme qu'un état mental est un état physico-chimique du cerveau[42].

Selon Armstrong[43], la TIP se présente ainsi comme une vue de l'esprit cohérente, puisqu'elle satisfait cinq critères essentiels: (i) elle rend compte de l'unité du corps et de l'esprit; (ii) elle peut expliquer la différence « numérique » entre différents esprits (le problème de l'individuation); (iii) elle explique l'interaction du corps et de l'esprit; (iv) elle peut rendre compte de l'émergence de l'esprit, par exemple en termes évolutionnistes; enfin, (v) elle peut rendre compte des états internes. Nous aurons l'occasion de revenir plus amplement sur les points faibles de cette position. Pour l'instant, il nous faut retenir l'importance des fondements physicalistes soutenant

[41] "Reality consists of nothing but a single all-embracing spatio-temporal system". (…) "The spatio-temporal system certainly exists. Whether anything else exists is controversial". D.M. Armstrong, (1980) 'Naturalism, Materialism, and First Philosophy' in *The Nature of Mind and Other Essays*, St. Lucia, University of Queensland Press. Réédité in P. Moser et J. Trout, (eds.) (1995), *Contemporary Materialism*, op. cit., p. 35. (Nous nous référons ici à cette dernière édition)
[42] (D.M. Armstrong, 1968a : 90-91, 356)
[43] *Ibid*., p. 75-6.

l'ossature de l'identité psychophysique. En prenant appui sur l'idée suivant laquelle la nature, c'est-à-dire le système spatio-temporel, est un système causalement clos[44], on conclut que le mental doit être ultimement physique. Selon Smart[45], qui reprend ainsi à son compte la célèbre distinction frégéenne entre sens et dénotation, les termes douleur et respectivement « activation des fibres-C » ont le même référent sans avoir le même sens. Le fait que l'identité « douleur »/« activation des fibres-C » ne soit pas de type *a priori*, ne veut pas dire qu'elle soit fausse.

Vers la fin des années 60, l'affaire de l'identité psychophysique a connu un nouveau rebondissement suite notamment à la parution en 1966 d'un très influent article signé par D. Lewis : 'An Argument for the Identity Theory'[46]. Comme son titre le laisse deviner, Lewis défend également la thèse de l'identité psychophysique en s'appuyant, lui aussi, sur un argument à deux volets. Une première prémisse stipule que la caractéristique fondamentale de l'expérience en tant que telle réside dans son rôle causal, rôle qui est à son tour traductible dans un ensemble fini de conditions spécifiant les effets typiques que l'on obtient sous différentes conjonctures. Selon Lewis[47], il est analytiquement nécessaire que ces conditions relatives à l'expérience soient à la fois vraies et distinctes de l'expérience elle-même.

La deuxième prémisse de l'argument de Lewis en faveur de la TIP repose sur le principe de l'unité de la science[48]. Selon Lewis[49], non seulement l'attribut déterminant de toute expérience en tant que telle est son rôle causal, mais surtout ces rôles causaux qui sont tenus d'appartenir, de façon analytiquement nécessaire, à l'expérience, appartiennent en réalité à des états physiques déterminés. Par conséquent, si la douleur est un état qui occupe une certaine « niche » causale, et si les théories scientifiques

---

[44] "The development of the natural sciences very strongly suggests that Nature, the spatio-temporal system, is a causally self-enclosed system". (D. M. Armstrong, 1980/1995: 38)

[45] J.J.C. Smart, (1963), 'Materialism', *Journal of Philosophy*, 60: 651-662.

[46] D. Lewis, (1966), 'An Argument for the Identity Theory', *Journal of Philosophy* 63. Réédité in J. Heil (ed.) *Philosophy of Mind – a Guide and Anthology,* Oxford University Press, 2004, pp. 150-157. (Nous faisons reference ici à cette dernière edition)

[47] "(…) the definitive characteristic of any experience as such is its causal role. The definitive causal role of an experience is expressible by a finite ensemble of conditions that specify its typical effects under various circumstances. By analytic necessity these conditions are true of the experience and jointly distinctive of it". (D. Lewis, 1966/2004: 153)

[48] "My second premise is the plausible hypothesis that there is some unified body of scientific theories, of the sort we now accept, which together provide true and exhaustive account of all physical phenomena (i.e. all phenomena describable in physical terms)". (*Ibid*., p. 155)

[49] "The definitive characteristic of any (sort of) experience as such is its causal role, its syndrome of most typical causes and effects. But we materialists believe that these causal roles which belong by analytic necessity to experiences belong in fact to certain physical states". (*Ibid*., p. 150)

révèlent que ces niches ou rôles sont occupés par un tel ou tel état neurophysiologique, alors il s'ensuit, grâce à la transitivité de la relation d'identité, que la douleur *est* un état neurophysiologique du cerveau. On voit, d'après ce qui précède, que le nouveau défi du physicalisme, et par ricochet, des partisans de la TIP, n'est plus celui de trouver les constituants physiques qui composent le mental à la manière dont les briques forment un mur, mais de répondre à la question : comment intégrer le rôle manifestement causal des états mentaux dans un agencement purement physique du réel ?

Incontestablement, la théorie de l'identité psychophysique a influencé de manière décisive le destin du physicalisme en tant que toile de fond du *'mind-body problem'*. D'abord, l'importance accordée au rôle causal dans l'explication du mental renforce l'idée que les états mentaux font effectivement partie de l'appareillage du monde et qu'à ce titre, ils peuvent faire l'objet d'une étude scientifique respectable. Ensuite, l'avantage de la TIP c'est qu'elle respecte de manière déterminante le principe physicaliste de clôture causale : tout événement physique doit avoir une cause explicative physique.

On ne s'étendra pas ici sur les objections, abondamment discutées ailleurs, élevées au fil des débats, par les adversaires de la TIP. Il suffit, pour le but rétrospectif de ce chapitre de passer en revue quelques unes des critiques les plus influentes. Une des failles majeures de la thèse de l'identité psychophysique fut signalée par S. Kripke[50] dont le célèbre argument des vérités nécessaires *a posteriori* est aujourd'hui largement adopté dans la littérature analytique. L'argument de Kripke affirme, *grosso modo*, que l'analogie douleur/activation des fibres-C d'une part, et eau/H_2O n'est pas valable car l'identification eau/ H_2O, bien que découverte *a posteriori*, est néanmoins nécessaire, et non pas contingente. Par contraste, si l'état mental de douleur est identique à l'état physique « activation des fibres-C », il s'ensuit que dans tous les mondes possibles, n'importe quel état mental qui a les traits phénoménaux de l'expérience d'une douleur est identique à un état cérébral du type stimulation des fibres-C.

Comparons l'énoncé d'identité (1) comme on en trouve tant en science, avec l'énoncé d'identité psychophysique (2) :

(1) Eau=H_2O

(2) Douleur = Activation des fibres-C

[50] (S. Kripke, 1980)

On peut raisonnablement concevoir chacun de ces énoncés comme étant faux, ce qui veut dire que leur vérité n'est pas *a priori*. Le problème est que, s'ils sont vrais, ils le sont nécessairement, et dans ce cas comment réconcilie-t-on leur contingence avec leur nature nécessaire ? Selon Kripke, il est relativement facile de sortir de cette impasse dans le cas de l'énoncé (1), car il est possible d'imaginer une situation dans laquelle l'est *n*'est *pas* de l'H_2O, c'est-à-dire où la substance en cause se comporte en surface comme de l'eau, mais qui en réalité n'est pas de l'H_2O. Que se passe-t-il si l'on se transpose, de façon analogue, dans la situation décrite par l'énoncé (2) ? Force est d'admettre alors qu'une explication similaire ne pourra pas rendre compte de l'aspect de contingence de (2), puisque imaginer une situation dans laquelle un individu S ressent un état qui ressemble en surface à de la douleur, ce n'est rien d'autre qu'imaginer une situation dans laquelle cette personne *ressent* effectivement une douleur.

Une autre importante objection contre la TIP fut avancée par Max Black (signalée par Smart dans son article de 1962), et développée ultérieurement par S. White[51]. En s'appuyant sur les travaux de Frege (1892), Black, et plus tard White, observent que la seule manière dont deux termes ayant des sens différents peuvent dénoter la même chose, c'est d'exprimer deux propriétés distinctes, ou « modes de présentation » de l'état en question. Mais cela entraîne une conséquence surprenante, car si des termes comme « douleur », « pensée » ou « désir » *ne* sont *pas* équivalents en sens à des descriptions en termes physicalistes, alors il s'ensuit qu'ils peuvent dénoter des états physiques uniquement en exprimant leurs propriétés mentales *irréductibles* à des descriptions physiques. Cet argument, connu sous le nom de « l'argument des propriétés distinctes » a fait l'objet de nombreux commentaires et nous aurons l'occasion d'y revenir lorsque nous enquêterons sur l'affaire de Mary dans l'argument de la connaissance de Jackson.

[51] Voir S. White, (1986), 'Curse of the *Qualia*', *Synthese* 68: 333–68. Réédité comme chap. 3 in S. White (1991) *The Unity of the Self*, Cambridge: MIT Press; Voir aussi S. White, (1999c), 'Why the Property Dualism Argument Will Not Go Away'. Texte non publié, présenté à New York University Language and Mind Colloquium, le 4 avril 1999. Pour une réponse voir N. Block, 'Max Black's Objection to Mind-Body Identity' in T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. New York: Oxford University Press.

On a également reproché aux adeptes de l'identité psychophysique ce que N. Block[52] appelle leur « chauvinisme » : si les états conscients sont littéralement identiques aux états neurophysiologiques du cerveau humain, alors il devient logiquement impossible pour des créatures autres que les humains – martiens, robots, etc., d'avoir des états conscients. Mais l'argument décisif qui ébranla sérieusement l'édifice de la TIP vint d'ailleurs : dans une série d'articles publiées dans les années 60, H. Putnam[53] reformule à nouveaux frais la question « est-ce que la douleur est identique avec un état cérébral ? » et y répond, de manière convaincante, par la négative, mettant ainsi les bases d'un nouveau courant au sein des théories matérialistes de l'esprit : le fonctionnalisme.

3.2.2 Identité-Type et Identité-Token

L'accent mis par Lewis sur la fonction causale dans la défense de la TIP a fini par provoquer un surprenant rebondissement dans l'affaire de l'articulation entre le mental et le physique. Bien que largement anticipé dans les écrits des certains défenseurs de la *théorie causale de l'identité*, comme Lewis et Armstrong, le fonctionnalisme tel que nous le connaissons aujourd'hui reste traditionnellement associé au nom de H. Putnam et fut explicitement formulé et articulé à partir de la deuxième moitié du XXe siècle. Selon Putnam, pour que la TIP prenne toute sa valeur, ses partisans doivent assumer *tous* les risques qu'entraîne l'adhésion explicite à la thèse de l'identité psychophysique forte, affirmant que la douleur *est* un état physico-chimique du cerveau à un certain moment *t*, à la façon dont l'eau *est* une combinaison moléculaire d'oxygène et d'hydrogène. Bien plus, il est évident que cette identification doit rendre compte des connexions physico-chimiques déterminantes dans l'apparition du phénomène de douleur non seulement dans le cerveau d'un mammifère, mais aussi dans la totalité du règne animal.

---

[52] N. Block, (1980a), 'Troubles with Functionalism', in N. Block (ed.) *Readings in the Philosophy of Psychology*, vol. 1, Cambridge, MA : Harvard University Press, 268-305.

[53] H. Putnam, (1967), 'Psychological Predicates', in W.H. Capitan et D.D. Merrill (eds.) *Art, Mind and Religion*, Pittsburg: University of Pittsburg Press. Réédité sous le titre 'The Nature of Mental States' in H. Putnam *Mind Language and Reality: Philosophical Papers Vol. 2.* Cambridge: Cambridge University Press. Trad. fr. in D. Fisette & P. Poirier, (2002), *Philosophie de l'Esprit - Psychologie du Sens Commun* et *Sciences de l'Esprit*, vol. I, p. 276. (Nous faisons réference ici à cette dernière édition.)

Stipuler que la douleur *est* l'activation des fibres-C telle qu'elle se manifeste dans le cerveau d'un mammifère, disons, exclut *ipso facto* l'accès des *autres* classes d'organismes (reptiles, mollusques, etc.) à la douleur. Or, le bon sens nous dit qu'un serpent ou un escargot de Bourgogne peuvent, *eux aussi*, ressentir de la douleur. La solution proposée par Putnam pour sortir de cette impasse est à la fois ingénieuse et innovante, jetant ainsi les bases du futur édifice fonctionnaliste : « La douleur – écrit Putnam – n'est pas un état cérébral au sens d'un état physico-chimique du cerveau (ou même de l'ensemble du système nerveux), mais un état d'une tout autre espèce. Je propose l'hypothèse que la douleur, ou l'état d'éprouver de la douleur, est un état fonctionnel de l'organisme tout entier»[54]. Réduit à sa plus simple expression, le fonctionnalisme défend la thèse selon laquelle ce qui définit un état mental (la douleur, pour reprendre l'exemple paradigmatique) n'est pas la structure brute ou la constitution d'un état, mais plutôt à la façon dont les différents éléments *fonctionnent* ensemble, c'est-à-dire, à la manière dont l'état mental concerné est causalement intégré au sein d'un système déterminé.

Selon Putnam, la thèse d'influence carnapienne selon laquelle les concepts (les intensions) ***sont*** la même chose que les propriétés, véhicule une idée risquée, puisqu'il est évident qu'un énoncé comme « la température *est* l'énergie cinétique moléculaire moyenne » est un parfait exemple d'énoncé d'identité de deux propriétés qui soit vrai, tandis que la proposition « le concept de température *est* le même que le concept d'énergie cinétique moléculaire moyenne » est tout simplement fausse[55].

Par conséquent, conclut-il, "le concept de douleur n'est pas le même concept que celui d'être dans l'état cérébral S"[56].

Qu'est-ce que la douleur *sinon* ? Est-elle un état cérébral ? Selon Putnam, « puisque je n'examine pas le problème de savoir en quoi consiste le concept de douleur, mais qu'est-ce que la douleur, en un sens d'être qui fait appel à la construction d'une théorie empirique (ou du moins à de la spéculation empirique), je n'ai aucune raison de m'excuser de répondre en avançant une hypothèse empirique"[57]. En conclusion, la douleur n'pas un état cérébral au sens d'état physico-chimique du

[54] (H. Putnam, 1967/2002: 276)
[55] *Ibid.*, p. 272.
[56] *Ibid.*
[57] *Ibid.*, p. 276.

cerveau, mais un état d'une tout autre espèce, à savoir un état *fonctionnel* de l'organisme tout entier. La valeur des arguments de Putnam ayant été amplement discutée dans la littérature analytique, on se contentera ici de quelques remarques synthétiques, en insistant particulièrement sur l'impact de l'hypothèse de la *multiple réalisabilité* sur l'évolution du physicalisme.

Réduite à sa plus simple expression, l'hypothèse de la réalisabilité multiple stipule que tout état fonctionnel (tout état occupant un certain rôle) peut être réalisé par et dans différents systèmes, soit par des systèmes neuronaux organiques, soit par des systèmes artificiellement construits. Bref, les types d'états mentaux peuvent être réalisés de manières multiples. Selon le résumé qu'en donne J. Bickle[58], le schéma général de l'argument qui vise s'inscrire en faux contre la thèse de l'identité psychophysique peut être résumé ainsi:

MULTIPLE RÉALISABILITÉ

(P1) Toutes les entités mentales sont multiple-réalisables par des entités physiques distinctes.

(P2) Si une entité mentale donnée est multiple-réalisable par des entités physiques distinctes, alors elle ne peut pas être identifiée à une entité physique spécifique donnée.

Par conséquent,

(C) Aucune entité mentale n'est identique avec une entité physique particulière.

Il n'est sans doute pas excessif d'affirmer que l'argument de la réalisabilité multiple a permis de réviser entièrement l'ontologie matérialiste sous-tendant la thèse de l'identité psychophysique. On a tôt fait de remarquer, à la suite de J. Fodor[59], que l'argument de la multiple réalisabilité montre seulement que la propriété d'éprouver de la douleur n'est pas identique avec une propriété physique universellement applicable ; cela ne veut pas dire, en revanche, que les occurrences des propriétés/états mentaux particuliers ne soient pas identiques avec des exemplaires, i.e., des occurrences des

[58] J. Bickle, *Multiple Realizability*: http://plato.stanford.edu/entries/multiple-realizability

[59] J. Fodor, (1974), 'Special Sciences: Or the Disunity of Science as a Working Hypothesis', *Synthese*, 28: 97-115.

propriétés/états physiques particuliers. Une distinction importante se crée alors entre ce qu'on appelle la thèse de « l'identité-type » et « l'identité-token ».

Quelques brèves remarques terminologiques d'abord sur la signification des termes "type" et " token"[60]. Un télégramme ayant la forme « *bises* et *bises* et *bises*… » contient seulement deux mots "types", mais cinq mots "tokens"[61]. La distinction type/token doit son origine, comme le rappelle Smart[62], à une analogie empruntée à la linguistique. Les 'tokens' sont donc à considérer comme des occurrences, des exemplaires particuliers instanciant un certain *type* de phénomène, événement, objet ou processus. Avec l'introduction de cette distinction, la théorie de l'identité psychophysique se ramifie en deux nouvelles branches : d'une part il y a ceux qui adhèrent à l'idée d'une identité psychophysique forte ou 'Identité Type/Type' selon laquelle pour chaque type de processus, état, événement mental il existe un processus neurophysiologique-type le sous-tendant. D'autre part, il y a les partisans d'une identité psychophysique plus souple ou 'Identité Token/Token', suivant laquelle même si la douleur *n'est pas* identique à un état physico-chimique d'un type universel présent chez *tous* les agents susceptibles de ressentir de la douleur, il n'en reste pas moins que l'état mental que nous appelons familièrement « douleur » survient toujours en présence d'un *certain* événement physique particulier (un 'token') instancié par le truchement d'un système physique donné. De manière générale, les fonctionalistes sont nombreux à adhérer au modèle de l'identité-token, qui présente, entre autres, l'inestimable avantage d'être compatible avec la thèse de la multiple réalisabilité.

Avant de passer en revue les différentes tentatives de concilier l'argument de la multiple réalisabilité avec la théorie de l'identité psychophysique, notons d'abord que certains auteurs comme L. Shapiro[63] et J. Bickle[64]ont récemment remis en cause le bien-fondé empirique de la multiple réalisabilité. Cette thèse, arguent-ils, n'est pas aussi facilement applicable aux cas concrets réels que ses défenseurs le laissent entendre. Autrement dit, si l'on compare deux organismes qui partagent une même

[60] Le choix de garder ici la transcription littérale de l'anglais 'token' au détriment de la traduction française 'occurrence', 'échantillon', ou 'exemplaire', n'est pas arbitraire. En effet, les termes français se prêtent facilement à de fâcheuses ambigüités.
[61] On peut exprimer la même idée en faisant appel à la distinction « Chat » (avec majuscule) et « chat » (avec minuscule), où ce dernier représente un exemplaire particulier, souvent signalé et accompagné d'un démonstratif (« *ce* chat »), exemplaire appartenant à une classe-type (« *les* chats »).
[62] J.J.C. Smart, *The Identity Theory of Mind* in http://plato.stanford.edu/entries/mind-identity
[63] L. Shapiro, (2000), 'Multiple Realizations', *Journal of Philosophy*, 97: 635-654.
[64] Voir J. Bickle, *Multiple Realizability*: http://plato.stanford.edu/entries/multiple-realizability

propriété mentale sans partager la même propriété physique, on s'aperçoit que (i) soit l'assemblage physique des deux organismes n'est pas très différent, (ii) soit que leur vie mentale n'est pas similaire, et que, par conséquent, la comparaison ne tient pas.

D'autres auteurs, comme J. Kim[65] ont argué que la thèse de la multiple réalisabilité ne montre pas que le concept d'être « joyeux », par exemple, pourrait ne pas s'appliquer à une créature seulement s'il instancie une propriété physique. A. Melnyk[66] soutient par exemple, qu'en dépit du fait que la thèse de la multiple réalisabilité montre que la propriété d'éprouver de la douleur ne soit pas identique avec une propriété-type physique de premier ordre, une propriété mentale pourrait être néanmoins identique avec une propriété physique de *second-ordre*. Bien que, de façon générale, on traite les propriétés de second-ordre comme des les propriétés d'une autre propriété, dans le cas de la multiple réalisabilité on parle d'une propriété qui implique un quantificateur sur une autre propriété. Pour donner un exemple : la propriété d'être *absorbant* est une propriété de second ordre d'une autre propriété qui dispose un objet (une serviette disons) à absorber l'eau. Pour voir le lien avec la thèse de la multiple réalisabilité, considérons l'exemple suivant : Michel et Jean-Michel ont acheté chacun une serviette de bain différente : celle de Michel absorbe l'eau grâce au tissage des fils en coton, alors que celle de Jean-Michel absorbe l'eau grâce à la texture des fils du tissu en lin, disons. Les deux serviettes sont absorbantes, car les deux ont la propriété *a* d'avoir une certaine propriété (a_{tissage} et a_{texture}) qui les disposent à absorber l'eau. Mais il est clair que les propriétés a_{tissage} et a_{texture} sont deux propriétés différentes. Car tandis que la propriété d'absorption de la serviette de Michel est réalisée par le truchement d'une propriété de tissage, dans le cas de Jean-Michel, la propriété est réalisée par le truchement de la texture du fil.

Laissant provisoirement de côté ces critiques, et dans le but de rendre possible une enquête sur ces questions qui ne soit faussée par une prise de parti subreptice, nous allons dresser, dans le chapitre suivant, un compte-rendu synthétisant les différentes

[65] (J. Kim, 1998) Toujours Kim, mais dans un autre livre, soutient que s'il est vrai que la thèse de la multiple réalisabilité montre que la propriété d'être joyeux n'est pas identique à une propriété physique unique, type, cela ne veut pas dire que la propriété mentale en cause ne soit pas identique avec une disjonction de propriétés physiques. Voir J. Kim, 1993, *Mind and Supervenience*, Cambridge: Cambridge University Press.

[66] Voir A. Melnyk, (2003), *A Physicalist Manifesto: Thoroughly Modern Materialism*, Cambridge: Cambridge University Press.

réactions et réajustements de la théorie de l'identité psychophysique provoqués par l'introduction de l'argument antiréductionniste de la multiple réalisabilité.

La section suivante, qui sera dédiée aux thèmes clé du fonctionnalisme, est la dernière de la série introductive avec laquelle nous avons débuté la mise en perspective de notre enquête. L'entrée du fonctionnalisme sur la scène des débats autour des rapports entre le mental et le physique nous servira à la fois de fermeture à l'esquisse synoptique des principaux courants de pensée qui ont vu naître ce thème durant le XXe siècle, ainsi que d'ouverture vers les arguments et critiques auxquels le physicalisme se verra confronté. Une présentation succincte de la manière dont les fonctionalistes ont abordé le problème des rapports corps-esprit constitue, à notre sens, une excellente introduction au problème des rapports compliqués qu'entretient le physicalisme avec les récalcitrants *qualia*.

3.3 Physicalisme et survenance

Bien que le terme « fonctionnalisme » couvre, comme le relève Ned Block[67], des champs disciplinaires aussi disparates que l'anthropologie, la critique littéraire ou la psychologie, le fonctionnalisme en philosophie de l'esprit se regroupe autour de trois filières distinctes: le *fonctionnalisme analytique*, le *psycho-fonctionnalisme*, et le *fonctionnalisme-machine ou computationnel*. On s'accorde généralement à considérer le « fonctionnalisme-machine » de Putnam et le monisme anomal[68] de D. Davidson comme les héritiers directs de la version causale de la TIP. Une analyse détaillée des implications du fonctionnalisme sur les modèles explicatifs du mental dépasserait les modestes objectifs de cette section. Nous allons rappeler néanmoins, en quelques points, les étapes clé qui ont ponctué son évolution, en espérant que ce bref survol nous donnera accès à une vue panoramique sur la manière dont le problème des *qualia* a surgi dans l'affaire du physicalisme. Mais d'abord quelques précisions concernant le concept clé de « survenance ».

[67] N. Block, (1980a), 'What is Functionalism' in *Readings in the Philosophy of Psychology, vol.1. (*Cambridge, MA: Harvard University Press. Réédité *in* J. Heil (ed.), 2004. (Nous utilisons ici cette dernière édition)

[68] Désormais (MA) in abrégé.

Dans l'article 'Mental Events', D. Davidson[69] tente la difficile mission d'investiguer sur les mécanismes responsables de l'interaction causale psychophysique, *sans* sacrifier pour autant l'autonomie du mental au déterminisme des lois physiques. Selon Davidson, les événements mentaux ont un statut ontologique spécial, dans le sens où ils ne subissent pas la même pression ontologique stricte que l'on retrouve au sein du réseau causal et nomologique d'une théorie physique. Le dilemme philosophique des rapports entre le mental et le physique repose, selon Davidson, sur un malentendu relatif à l'interprétation des trois principes suivants :

(i) le *Principe d'Interaction Causale* selon lequel certains événements mentaux interagissent causalement avec des événements physiques.

(ii) le *Principe du Caractère Nomologique de la Causalité* affirmant que là où il y a causalité, il doit y avoir une loi stricte sous laquelle tombent les événements concernés.

(iii) enfin, le point central de la théorie davidsonienne, c'est le *Principe de l'Anomie du Mental*, selon lequel "il n'y a pas de lois déterministes strictes à partir desquelles on puisse prédire et expliquer la nature exacte des événements mentaux"[70].

Selon Davidson, ces trois thèses sont vraies conjointement, et leur contradiction est seulement apparente. Il défend ainsi une version 'token' du matérialisme. Autrement dit, le MA valide la réduction ontologique, mais rejette et évite la réduction conceptuelle du mental au physique, évitant ainsi la contradiction interne[71]. Le MA affirme donc que les événements mentaux *sont* des événements physiques, mais rejette la thèse selon laquelle il existerait des lois strictes raccordant le physique au mental. C'est justement pour préciser la nature de ce « raccordement », que le philosophe américain introduit la notion-clé de *survenance* avec le succès retentissant qu'on lui connaît aujourd'hui.

---

[69] D. Davidson, (1970b), 'Mental Events', *in* L. Foster & J. *W.* Swanson (eds.), *Experience and Theory*, London: Duckworth. Réédité *in* D. Davidson 1984, *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press. Trad. fr. par P. Engel *in* D. Fisette & P. Poirier (eds.) 2002, op. cit. Nous utilisons ici cette dernière édition.

[70] (D. Davidson 1970b/ 2002 : 240)

[71] « Mental entities (particular time – and space-bound objects and events) are physical entities, but that mental concepts are not reducible by definition or natural laws to physical concepts". D. Davidson, (1993), 'Thinking Causes' in J. Heil & A. Mele (eds.), *Mental Causation*, op. cit., p. 3.

Le concept de « survenance », déjà attesté et utilisé par certains philosophes émergentistes britanniques comme Lloyd Morgan[72], fut incontestablement lancé et popularisé dans le passage suivant du célèbre article de D. Davidson[73] : « Bien que la position que je décris nie qu'il y ait des lois psychophysiques, elle est compatible avec la thèse selon laquelle les caractéristiques mentales sont en un certain sens dépendantes des caractéristiques physiques, ou *survenantes* par rapport à elles. On peut interpréter cette survenance comme signifiant qu'il ne peut pas y avoir deux événements qui soient semblables sous tous leurs aspects physiques mais qui diffèrent sous un aspect mental quelconque »[74].

Comme le relève fort bien J. Kim[75], il n'est sans doute pas excessif de considérer ce bref passage de 'Mental Events' comme établissant le certificat de naissance d'un concept destiné à faire une longue et brillante carrière au sein du physicalisme. Nous l'avons compris, l'enjeu crucial dans l'affaire du MA consiste à concilier l'autonomie du mental avec la thèse métaphysique du monisme matérialiste[76]. En outre, la survenance vient remplir le *vide métaphysique*[77] qui séparait le rive du mental de celui du physique. Par conséquent, les applications de ces deux concepts centraux – survenance et multiple réalisabilité – ont permis l'instauration, vers le milieu des années 70, de ce que N. Block appelle le « consensus antiréductionniste » [78]. Dans la section suivante, nous rappelons comment le fonctionnalisme permet d'élaborer un

[72] L. Morgan (1923), *Emergent Evolution*, Williams and Norgate.

[73] Lorsqu'il publie en 1970 son désormais classique 'Mental Events', Davidson pense à tort, comme il l'avouera plus tard, que ce fut G.E. Moore qui avait utilisé avant lui le terme de « survenance » pour décrire la relation entre les notions évaluatives comme « bien » et celles descriptives comme « pointu », ou « carré ». En réalité, dit-il, le terme 'survenance' lui fut probablement inspiré d'un livre de R.M. Hare. (Voir D. Davidson, 1993:4)

[74] (D. Davidson, 1970b/2002:248) (Nous soulignons) Plus tard, Davidson reformulera cette définition dans un format plus "économique": "A predicat *p* is supervenient on a ensemble of predicates *S* if and only if *p* does not distinguish any entities that cannot be distinguished by *S*". (D. Davidson, 1993:4)

[75] (J. Kim, 1998)

[76] Dans l'article 'The Myth of Non-Reductive Materialism' J. Kim dresse une critique du monisme anomal et soutient que ce dernier n'est qu'une forme sophistiquée d'éliminativisme. Bien plus, le matérialisme non-réductionniste qui le sous-tend n'offre pas une solution viable au vieux problème des rapports psychophysiques. Selon Kim, le MA propose seulement une description *négative* des rapports entre le mental et le physique, et nous informe uniquement sur ce que les états mentaux *ne sont pas*. Or nous sommes dans le devoir de bâtir une théorie informative, *positive* sur la manière dont le mental et le physique se raccordent *:* l'appel aux lois psychophysiques entraîne automatiquement la présence, ou du moins la promesse d'une réduction du mental au physique. A cela Davidson répond en précisant que si les lois en question ne sont pas strictes, mais de type *ceteris paribus*, alors la menace de la réduction disparaît. Pour une discussion détaillée de ces points voir J. Kim, (1989b), 'The myth of non-reductive physicalism' in *Proceedings and Addresses of the American Philosophical Association, 63:* 31-47. Pour une contre-réplique voir D. Davidson (1993).

[77] (J. Kim, 1998 :7)

[78] N. Block, (1997), *'Anti-Reductionism Slaps Back', in* J.E. Tomberlin (ed.) *Philosophical Perspectives*, 11, Oxford: Blackwell.

modèle de l'esprit compatible avec la thèse de la réalisation multiple des états mentaux. Nous verrons ainsi que l'avantage du fonctionnalisme est qu'il peut s'accommoder d'une large palette des vues de l'esprit.

3.3.1 Le Fonctionnalisme en trois questions

Comme nous l'avons brièvement évoqué au début de cette section, on s'accorde généralement à considérer les trois branches du fonctionnalisme, à savoir : (i) le fonctionnalisme « machine » ou computationnel, (ii) le « psycho-fonctionnalisme », et (iii) le fonctionnalisme analytique, comme étant les héritiers respectifs des théories computationnelles, du behaviorisme empirique et du behaviorisme logique. En dépit de leurs divergences, tous les fonctionnalistes adhèrent, sans exception, à une thèse fondamentale : les états mentaux *sont* des états fonctionnels du cerveau. Différencier un état mental d'un état non-mental n'est pas une affaire de trouver les « bonnes briques » d'un présumé substrat ultime, mais plutôt d'établir la manière dont ces éléments s'organisent dans un système d'ensemble. Bref, pour le fonctionnaliste, le seul déterminant de la nature d'un état psychique réside dans le rôle fonctionnel qu'il occupe. Tout état (neurologique, électronique, computationnel, etc.) capable d'occuper le rôle approprié *est* un état psychique.

Les fonctionnalistes souscrivent ouvertement à une ontologie physicaliste, autrement dit, ils considèrent l'esprit humain comme un élément légitime et à part entière du monde naturel, s'inscrivant ainsi en faux contre la position que B. Stroud[79] a dépeint sous le nom de « surnaturalisme ».

Mise à part cette entente concernant la définition des états mentaux comme des états fonctionnels et holistiques, l'affaire du fonctionnalisme couve d'importants désaccords. Selon R. Van Gulick[80], les controverses peuvent être organisées et divisées en plusieurs volets, compte tenant compte des trois questions suivantes.

(1) Première question : *quels sont états mentaux qui sont visés par les définitions fonctionnalistes* ? A première vue, les états concernés sont ceux recensés par la psychologie du sens commun, à savoir les attitudes propositionnelles : croyances, désirs, intentions, etc. D'autre part, les thèses fonctionnalistes peuvent viser également

[79] (B. Stroud, 1996)
[80] R. Van Gulick, (2007), 'Functionalism and *Qualia*', *in* M. Velmans et S. Schneider (eds.), *The Blackwell Companion to Consciousness*, Blackwell Publishing, p. 393.

l'étude empirique des états psychologiques englobés au sein d'une certaine théorie, qu'elle soit présente, future ou idéale. Ce type d'états théoriques peuvent ou pas correspondre à ceux que l'on trouve au sein d'une psychologie du sens commun, et dans la mesure où ils ne correspondent pas à ces derniers, leur statut d'états fonctionnels devient un sujet d'investigation en soi. Mais au-delà de la question de savoir si les états mentaux sont théoriques ou extra-théoriques, il convient de souligner que les interrogations visant l'aspect qualitatif de l'expérience phénoménale ne peuvent pas être éludées par aucun modèle fonctionnaliste de l'esprit.

(2) La deuxième question est : *comment doit-on caractériser les relations fonctionnelles*? Bref, qu'est-ce qu'une *fonction* ? Les diverses tentatives de réponse à cette question divise les fonctionnalistes en deux camps : d'une part il y a ceux qui définissent la notion de fonction en s'appuyant sur une interprétation *téléologique*. D'autre part, il y a ceux qui la rejettent, et préfèrent une interprétation *non-téléologique*, optant pour un sens strictement mathématique, et insistant sur l'importance des opérations de codification ou de traitement des données selon le modèle computationnel de type input/output.

Commençons avec ce dernier modèle. Le *fonctionnalisme-machine* ou *computationnel* mise sur une explication de l'esprit en termes de machine de Turing[81] et se construit autour du modèle de l'ordinateur, fournissant une explication de l'esprit en termes de programme (software) implémenté dans un système matériel (hardware), en occurrence, le cerveau humain. Cette version est une dérivation des théories et méthodes employées par les sciences cognitives. Dans son analyse de l'histoire de la révolution cognitive, H. Gardner[82] a signalé l'importance des conditions sociales qui ont favorisé l'essor de ce type d'approche dans la philosophie analytique.

Dans le contexte historique[83] exceptionnel d'entre deux guerres mondiales survenues presque coup sur coup, les années de la fin du XIXe siècle et du début du

---

[81] Une "machine de Turing" est un outil computationnel abstrait conçu par A. Turing dans un célèbre article publié en 1937, 'Computability and λ-Definability', *The Journal of Symbolic Logic*, 2: 153–163. Voir pour une discussion plus détaillé H. Putnam, (1960), 'Minds and Machines', *in* S. Hook (ed.) *Dimensions of Mind*, New York: New York University Press, pp. 148-180. Réédité *in* H. Putnam, (1975b), *Mind, Language and Reality* (1975), pp. 362-385.

[82] H. Gardner, (1985), *The Mind's New Science - A History of the Cognitive Revolution.* Basic Books. Trad. fr. J.-L. Peytavin *Histoire de la révolution cognitive*, Paris, Payot, 1993. Nous faisons référence à la traduction française.

[83] Selon Gardner, il existe un consensus sur une date de naissance des sciences cognitive: 11 septembre 1956, date d'un symposium sur la théorie de l'information (MIT) auquel participèrent, parmi tant d'autres, les auteurs suivants : Bruner, Goodnow, Austin - *Study of Thinking* du côté de la psychologie; Chomsky - *Three models of language* (du côté de la linguistique); George Miller - *The magical number*

XXe furent d'une importance décisive pour l'essor des recherches computationnelles. Peu à peu, l'idée selon laquelle l'ordinateur et l'esprit humain peuvent être considérés comme des "systèmes symboliques", c'est-à-dire des entités physiques qui traitent, transforment, élaborent, et, en général, manipulent divers types de symboles, s'est frayé chemin dans la philosophie analytique. Par un étrange renversement de perspective, l'invention de l'ordinateur et de l'Intelligence Artificielle sert de preuve d'existence de l'esprit humain : "si on peut dire d'une machine construite par l'homme qu'elle raisonne, qu'elle a des objectifs, qu'elle corrige son comportement, qu'elle transforme une information, et a d'autres activités de ce genre, les êtres humains méritent d'être caractérisés de la même façon"[84]. Suite aux développements de l'Intelligence Artificielle[85], il est devenu donc légitime d'établir un niveau d'analyse distinct, généralement appelé « niveau de représentation ».

Pour les fins de notre discussion dans cet essai, il convient de noter que l'émergence du paradigme computationnel de l'esprit a permis d'envisager la construction d'un modèle de l'esprit abstrait, i.e. situé en dehors du contexte dans lequel des actions ou pensées particulières se produisent, c'est-à-dire à l'abri de l'influence nocive des facteurs affectifs, émotionnels, ou des facteurs historiques et culturels. L'esprit devient ainsi un « automate probabiliste » à l'intérieur duquel un programme détermine pour chaque état et ensemble d'inputs, la probabilité avec laquelle la machine (le corps) va produire un certain output. Conséquemment, l'idée selon laquelle les états mentaux, entièrement descriptibles en termes de relations entre les données d'un système input/output d'une part, et le réseau formé par les autres états mentaux d'autre part, peuvent être subsumés à des lois, et donc donner lieu à des prédictions, marque une étape cruciale dans l'évolution des modèles explicatifs de l'esprit.

---

seven (représentant la psychologie également); ainsi qu'Allen Newel, Herbert Simon - *Logical Theory Machine* (ambassadeurs de la nouvelle discipline, l'informatique). Comme le résume fort bien Gardner, « ce qui avait germé dans les années 50 germa rapidement dans la décennie qui suivit. Les sources gouvernementales et privées apportèrent un soutien financier notable."(H. Gardner, 1985/1993 : 46)

[84] *Ibid.*, p 55.

[85] Avec le recul nécessaire, on considère aujourd'hui que les bases symboliques des sciences cognitives telles que nous les connaissons aujourd'hui ont été posées durant le symposium Hixon, tenu en septembre 1948 au *California Institute of Technology*[85]. Le thème proposé constituait un véritable défi lancé au behaviorisme dominant à l'époque: comment le système nerveux contrôle-t-il le comportement? Comme on pouvait s'y attendre, les behavioristes ont évité soigneusement, dans leurs réponses, d'aborder des sujets tels que l' « esprit », la « pensée » ou d'employer des concepts tels que « croyances », « désirs » ou « intentions ». Or, ce furent les discussions autour du thème de « contrôle du comportement » qui suscitèrent chez de nombreux chercheurs la conviction grandissante que l'adhésion aux critères béhavioristes rendait difficile, voire impossible l'étude scientifique de l'esprit.

Dans le camp des théories fonctionnalistes *non-téléologiques*[86], on retrouve également les versions qui interprètent la notion de « fonction » en termes de relations causales, comme par exemple les dispositions de causer ou de freiner l'apparition de certains effets. A leur tour, les théories fonctionnalistes de type causal se différencient en fonction des ressources conceptuelles qu'elles utilisent dans la définition des rôles causaux. Certains se contentent de suivre une voie épurée, aride, voire mécanistique, tandis que d'autres ouvrent la porte à des ressources plus souples, tels la rationalité ou l'intentionnalité. Bien évidemment, nous allons le voir, la question de savoir *lesquels* sont les concepts les plus appropriés pour définir les fonctions causales, aura une incidence directe sur le problème des *qualia*.

Du côté des versions *téléologiques*[87] du fonctionnalisme, nous retrouvons des interprétations du concept de « fonction » en termes de « finalités » analogues aux fonctions biologiques caractérisant un organisme vivant. Un état mental déterminé est caractérisé par un rôle causal, *parce que* sa nature consiste à jouer précisément ce rôle spécifique. A la différence des stratégies behavioristes qui affirment que dans la composition des lois de la psychologie on retrouve uniquement des *dispositions* à comportement, les psychologues cognitivistes sont prêts à admettre que les meilleures théories explicatives du comportement mettent en scène le résultat d'un complexe réseau interactif d'états mentaux et processus instanciés en termes de rôles causaux, participant activement à la production du comportement. Nous verrons plus loin au cours de notre investigation qu'au vu de cette description, le défi qui sera lancé aux fonctionnalistes téléologiques concernera la représentation d'une description adéquate de l'expérience phénoménale en s'appuyant sur le concept de fonction téléologique. Bien évidemment, comme le souligne Van Gulick[88], une importante partie de l'affaire repose sur la question de savoir si dans la description des fonctions il est permis ou pas d'utiliser des termes mentaux ou phénoménaux.

(3) Enfin, la troisième question est : *quelle est la signification de la thèse selon laquelle les états mentaux doivent être compris et expliqués en termes de leurs rôles fonctionnels* ?

---

[86] Voir D. Lewis, (1972), 'Psychophysical and theoretical identifications', *Australasian Journal of Philosophy* 50: 249-58.
[87] Voir R. Van Gulick, (1980) 'Functionalism, information and content', *Nature and System* 2, 139–62. Voir aussi W. Lycan (1987)
[88] (R. Van Gulick, 2007 : 393)

Pour certains penseurs[89], cette thèse mène inévitablement à un réductionnisme rigide : les propriétés mentales *sont* des propriétés fonctionnelles tout court. La réduction du mental au fonctionnel est, dans ce cas, la marque d'une identité stricte ou forte. Pour d'autres[90], le rôle fonctionnel ne fait que spécifier les descriptions ou les conditions nécessaires à l'identification et à la détermination du référent d'un terme mental, postulant ainsi que l' « essence » ou le « substrat » qui caractérise la propriété respective peut ne pas être, quant à lui, déterminé ou spécifié en termes fonctionnels. Pour reprendre l'exemple paradigmatique de la douleur, même si le concept de « douleur » peut être *défini* en termes de rôles fonctionnels, il n'en demeure pas moins que son référent est *réalisé* par un état neurologique localisable dans le cerveau. Une variante encore plus faible de ce type d'approche propose une interprétation méthodologique[91]des rapports entre le mental et le fonctionnel, dans le sens où l'équation états mentaux/états fonctionnels doit être conçue comme un outil conceptuel efficace, certes, dans la description de certaines particularités des états mentaux, mais incapable de rendre compte de leur essence ultime, de leur nature intrinsèque.

Pour résumer, disons que le fonctionnalisme, comme toute doctrine philosophique influente, a divisé les débats concernant la nature de l'esprit en général, et de l'expérience phénoménale en particulier, en deux camps : d'un côté se rangent ceux qui adhèrent pleinement au *credo* fonctionnaliste[92], tandis que dans l'autre coin du ring se rangent ceux qui reprochent au modèle fonctionnaliste standard de manquer l'essentiel, c'est-à-dire de laisser inexpliquée la question cruciale des *qualia*[93].Pour le dire de manière succincte, les controverses concernant les rapports entre le fonctionnalisme et le problème de l'expérience phénoménale s'organisent autour d'une question (Q_F) qui synthétise l'élément déclencheur de notre enquête :

(Q_F) *Le fonctionnalisme est-il en mesure de fournir une explication adéquate, c'est-à-dire* complète, *de la nature des états mentaux en général et des qualia en particulier ?*

[89] (D. Lewis, 1972)
[90] (N. Block, 1980a)
[91] (Voir D. Chalmers, 1996)
[92] J. Fodor, (1968), *Psychological Explanation*. New York: Random House. Voir aussi N. Block & J. Fodor, (1972), 'What Psychological States Are Not', *Philosophical Review*, 81: 159–181.
[93] Voir N. Block, (1980a), J. Searle (1992) et D. Chalmers (1996).

Si l'on s'en tient à l'affirmation wittgensteinienne selon laquelle, en philosophie, une question se traite avant tout comme une maladie, on peut suggérer qu'une bonne méthode de s'attaquer au problème serait de commencer par diagnostiquer et clarifier ses symptômes apparents, i.e. les concepts-protagonistes intervenant dans le débat, à savoir (1) le fonctionnalisme, (2) les *qualia*, et (3) le verbe-liant « expliquer ». Par conséquent, trois sont les niveaux interprétatifs autours desquels on peut construire les interrogations autour des rapports entre le fonctionnalisme et les *qualia*.

(F1) Comment doit-on interpréter le terme « fonctionnalisme » ?

(F2) Quelles sont réellement les caractéristiques des *qualia* qui ont besoin d'être expliquées ?

(F3) Quels sont les éléments qui pèsent dans une théorie explicative à succès ?[94]

Cela va sans dire, la façon dont on aborde (F1) à (F3) influence de manière décisive le traitement de la question initiale (Q_F). (i) D'une part on peut s'interroger sur la légitimité de vouloir établir un ensemble de fonctions exclusivement computationnelles dont l'implémentation soit logiquement suffisante pour en déduire une explication adéquate de l'expérience qualitative. Alternativement, on peut s'interroger jusqu'à quel point le concept de *qualia* tel qu'il est utilisé par le sens commun peut être explicité en termes mentaux jouant un rôle causal précis et interagissant avec d'autres états mentaux, tels les croyances, les intentions, les désirs, etc. (ii) D'autre part, du côté du fonctionnalisme, il est clair que l'ingrédient qui pèse le plus lourd dans cette équation est étroitement lié à la place que l'on est disposé à accorder à la thèse réductionniste, autrement dit, à la façon dont on compte répondre à la question : dans quel sens un état mental est tenu pour être *réductible* à un état fonctionnel ? Bien que parfaitement légitimes, ces questions ne représentent que le bout de l'iceberg que pose le problème de l'expérience phénoménale pour le physicalisme. Il est temps à présent de revenir à notre problème initial et de se pencher du côté des *qualia*.

Dans ce chapitre, nous avons dressé un très succinct état des lieux du problème des rapports entre le mental et le physique dans la tradition analytique contemporaine. Nous avons pu constater ainsi que plusieurs étapes ponctuent l'évolution de ces rapports : d'abord il y a l'étape fondatrice du rejet du dualisme interactionniste cartésien, suivi de très près d'un mouvement d'enthousiasme béhavioriste. Ce dernier

[94] (R. Van Gulick, 2007: 382)

laisse vite sa place à la théorie de l'identité psychophysique qui, à sont tour, et suite à l'argument antiréductionniste fondé sur la thèse de la réalisabilité multiple, donne naissance au fonctionnalisme. Ce dernier résout le problème des rapports entre le mental et le physique en stipulant que tout état (neurologique, computationnel, électronique, extraterrestre, etc.) capable d'occuper le rôle approprié, est un état mental.

Le modèle fonctionnaliste de l'esprit fut accueilli avec beaucoup d'enthousiasme par la communauté philosophique, car il rendait compte de manière plus élégante que la TIP de la complexité et de la plasticité de la vie mentale. Ce tableau idyllique fut cependant troublé par un nuage récalcitrant : l'aspect qualitatif de l'expérience phénoménale. Véritable 'trouble-fête', l'expérience qualitative trouve difficilement sa place dans le monde fonctionnaliste. Dans le chapitre suivant, nous allons esquisser le portrait du deuxième acteur dans le duel que nous examinons ici : les *qualia*.

4 Physicalisme et *qualia*

Il n'est pas forcément nécessaire d'être Marcel Proust ou d'aimer les madeleines pour comprendre la notion d' « effet-que-cela-fait » de sentir un phénomène, ou d'éprouver une sensation qualitative particulière. Pensez à l'effet que cela fait d'entendre un air d'opéra mozartien, ou de déguster un bon verre de vin rouge, ou de marcher pieds nus dans le sable brûlant de l'été. Toutes ces facettes « privées » et qualitatives de notre vie mentale correspondent à ce qu'on appelle la « phénoménalité » de la conscience.

L'idée selon laquelle la conscience est la marque essentielle du mental n'est pas nouvelle. Elle remonte très loin dans l'histoire de la pensée philosophique. Bien que l'histoire des rapports qui relient les entités, faits, processus *immatériels* à des entités, faits et processus *matériels* semble avoir été une préoccupation constante de la réflexion philosophique, la question de la nature de la conscience et de sa place dans une conception scientifique, i.e. matérialiste du monde, a pris une ampleur sans précédent à partir des années 60 du siècle dernier, pour devenir aujourd'hui, selon l'expression de D. Chalmers[1], le « problème difficile » que tout philosophe de l'esprit respectable se doit un jour affronter. Jamais l'invitation quinienne d'étudier le *sujet humain physique*[2] n'a été aussi fidèlement et assidûment suivie qu'aujourd'hui. Mais à la différence du « premier » physicalisme, le physicalisme moderne se voit imposer un tout autre défi : ce n'est plus le langage, mais l'esprit en général, et les *qualia* en particulier, qu'il faut « physicaliser » si l'on veut expliquer la nature de la pensée en respectant le principe fondamental de clôture causale, nomologique et explicative du physique. Une épaisse et fascinante littérature traitant de ces aspects a vu le jour ces dernières décennies. La redoutable complexité de ce thème, ainsi que l'hétérogénéité des facettes sous lesquelles il a été récemment abordé, nous interdit de prétendre fournir ici une analyse exhaustive et complète des interrogations qui concernent la place de la conscience dans une description matérialiste du monde. Plus modestement, nous nous contenterons d'en esquisser schématiquement ses contours d'ensemble. Bien que l'objectif principal de cette incursion soit celui de situer avec plus de précision

[1] (D. Chalmers, 1996)
[2] (W.V. Quine, 1958/2004: 51)

l'argument de la connaissance de Jackson sur l'orbite des interrogations qui gravitent autour du physicalisme, on ne pourra faire l'économie d'un bref détour par la présentation de l'arrière-plan conceptuel qui entoure la question des *qualia* en dehors du cas spécifique de Mary.

Selon l'heureuse expression de J. Levine[3], le problème des *qualia* dans la littérature analytique contemporaine peut être figuré sous la forme d'un pendule oscillant entre deux pôles : d'une part il y a ceux qui traitent les *qualia* comme des propriétés *intrinsèques* de l'expérience; à l'opposée il y a ceux qui les conçoivent comme des entités plutôt *relationnelles*. Avant même d'aller plus loin, on voit déjà se profiler une évidente difficulté définitionnelle, car cette bipartition nous oblige non seulement à préciser qu'est-ce que pour une propriété d'être *intrinsèque* ou *relationnelle*, mais surtout à tracer avec netteté la frontière qui sépare, dans la caractérisation d'un objet ou d'un état, la classe de ses propriétés intrinsèques (de premier ordre) de celle des propriétés relationnelles (de second ordre). Nous voyons donc qu'au-delà de leur statut indéniable de cible récalcitrante du réductionnisme radical, les *qualia* constituent avant tout un problème philosophique en soi. Qu'est ce que les *qualia* ? S'agit-il des entités réelles, des membres légitimes de l'ameublement du monde ? Et si la réponse est affirmative, lesquels de leurs traits doit-on traduire et expliquer en termes physicalistes si l'on veut respecter le principe de clôture causale du monde physique et d'éviter ainsi les écueils des dualismes ?

Il est naturellement difficile, voire impossible de dresser ici un compte-rendu complet épuisant les différentes variantes définitionnelles élaborées au fur et à mesure que la situation philosophique des *qualia* est devenue officiellement une épine dans le pied de la tradition physicaliste. Aussi, nous voulons éviter dans ce chapitre de faire une revue excessivement historique de ces points. Mais il est clair que quel que soit l'angle d'analyse choisi, une mise en perspective du thème est tout simplement vitale. Dans cette section, nous aborderons donc certaines des questions liées aux rapports entre le physicalisme et *qualia*, mais nous ne serons absolument pas exhaustifs.

Ce chapitre sera organisé de la manière suivante : dans la section 4.1., nous commencerons par survoler les principales acceptions du concept de « *qualia* », pour

[3] "It seems to me that a lot of the literature about *qualia* over the past two decades can be seen as a pendulum, with various proposals bouncing back and forth between treating *qualia* as intrinsic and treating them as relational, but none overcoming the basic structure of this dilemma: *qualia* as intrinsic properties can't be integrated into naturalistic framework, but no proposal to treat them as relational seems at all compelling". J. Levine, (2001), *Purple Haze: The Puzzle of Conscious Experience*. Cambridge, MA: MIT Press, p. 93.

bifurquer ensuite vers la description des différentes facettes sous lesquelles ce problème surgit au sein des différentes versions du matérialisme contemporain. Pour ce faire, nous allons passer en revue, dans les sections allant de 4.2. à 4.5., et à la suite des classifications proposées récemment par D. Chalmers[4], les principaux arguments antiphysicalistes, avant de s'arrêter à l'argument de la connaissance, dans la deuxième partie de cet essai.

4.1 Définir les qualia

De toute évidence, nous sommes conscients lorsque nous sommes éveillés. Chaque état mental est un mode sous lequel se manifeste la conscience. Selon l'acception la plus courante, la conscience est un continuum d'expériences vécues dotées de qualités phénoménales spécifiques. Pour décrire l'aspect vécu ou phénoménal de ces expériences, on utilise le terme distinctif de « *qualia* ». Réduits à leur plus simple expression, les *qualia* sont des qualités sensorielles (sensibles) que l'on associe habituellement à l'expérience phénoménale, autrement dit, à des modes privés d'expérience sous lesquels le monde se présente à nous. Pour reprendre la désormais classique définition de Th. Nagel[5], un organisme est conscient uniquement s'il existe « *un-effet-que-cela-fait* » d'être l'organisme en question. Sans insister ici sur un sujet qui sera développé plus avant au cours de notre enquête, rappelons seulement que traditionnellement, les philosophes analytiques utilisent le terme « *qualia* » pour désigner les propriétés intrinsèques des états mentaux ou des objets auxquels nous avons un accès non-inférrentiel et direct, et dont nous en faisons l'expérience sous le mode d'une connaissance immédiate. Ce type de *qualia* privés et ineffables sont habituellement évoqués pour désigner le caractère distinctif des expériences subjectives de type « l'effet-que-cela-fait », mais il serait une erreur grossière de les réifier. Les *qualia* font partie, certes, de l'ameublement du monde, mais non pas à la manière des transats, des chats ou des découverts bancaires.

Traditionnellement, l'étude des *qualia* est étroitement liée aux théories représentationnelles classiques de la perception, dont l'origine remonte au moins

[4] D. Chalmers, (2007), 'The Hard Problem of Consciousness', *in* M. Velmans & S. Schneider (eds.), op. cit.

[5] T. Nagel, (1974), 'What is it like to be a bat?' *Philosophical Review*, 83: 435-456. Repris in T. Nagel *Mortal Questions*. Cambridge: Cambridge University Press: 165–80. Trad.fr 'Quel effet cela fait d'être une chauve-souris ?' in Th. Nagel, *Questions mortelles*, PUF, Paris, 1983.

jusqu'à J. Locke[6] (1688) et à sa théorie des idées. Selon Locke, l'être humain ne jouit pas d'un accès direct, dans la perception, des objets externes. La perception du monde extérieur est rendue possible *via* les idées et les images sensibles qui, elles, et elles seulement, ont la capacité et le pouvoir de représenter les choses extérieures. Dans cette conception, les *qualia* ne sont rien d'autre que les propriétés que ces idées sensibles projettent sur l'écran intérieur de l'esprit. Pour donner un exemple : lorsque je perçois un rubis, l'idée sensible du rouge se présente à mon esprit sous la forme d'une photo-image, les *qualia* étant dans ce cas les propriétés de *cette* photo-image interne, et non pas de l'objet extérieur proprement dit, en occurrence, du rubis. Vu de cet angle, l'expression de "mental paint" (peinture mentale) proposée par N. Block[7]capte fidèlement la manière dont ce type de *qualia* projette sur l'écran intérieur qu'est l'esprit humain, les objets du monde extérieur. De nos jours, peu de philosophes sont enclins à souscrire à cette conception.

Encouragés par la faillite de cette acception traditionnelle, certains auteurs comme D. Dennett[8], ont vu là une bonne occasion pour proposer l'abandon pur et simple des *qualia*, anciens vestiges d'une conception démodée de l'expérience phénoménale. P. Pettit[9] par exemple, dresse une liste de raisons pour lesquelles les *qualia* seraient « suspects » : (a) d'abord les *qualia* pourraient s'avérer inintelligibles ; (b) ensuite, postuler l'existence des *qualia* n'est pas une manœuvre ontologiquement économique ; (c) enfin, les *qualia* échouent au test de la vérification empirique.

Certains auteurs, comme S. Shoemaker[10], proposent de continuer à créditer la pertinence des *qualia* en tant que sujet philosophique, à condition qu'on s'accorde à les examiner sous un angle permettant plus de précision. De manière générale, deux sont les conditions fondamentales qui accompagnent une caractérisation adéquate des *qualia*: (i) la possibilité d'un accès introspectif aux états perceptuels qui les sous-tendent et (ii) la possibilité de caractérisation de ces états en termes qualitatifs de type l'« effet-que-cela-fait ».

---

[6] J. Locke, (1688) *An Essay on Human Understanding*.
[7] N. Block, (1996), 'Mental paint and mental latex', in E. Villanueva, (ed.) *Perception*. Atascadero, CA: Ridgeview.
[8] Voir D. Dennett, (1990), 'Quining *qualia*', in W. Lycan (ed.) (1990), *Mind and Cognition – A Reader*, Oxford: Blackwell, 519-548. Et aussi D. Dennett, (1991), *Consciousness Explained*. Boston: Little, Brown and Company.
[9] P. Pettit, (2004), 'Motion Blindness and the Knowledge Argument', *in* Ludlow et al., op. cit
[10] Voir S. Shoemaker, (1975), 'Functionalism and *Qualia*', *Philosophical Studies*, 27, 291-315. Voir aussi S. Shoemaker, (1990), 'Qualities and *qualia*: what's in the mind', *Philosophy and Phenomenological Research*, Supplement, 50: 109-131.

Quoiqu'il en soit, les approches modernes du problème des *qualia* reposent sur un remaniement plus ou moins radical dans l'ossature de base de l'acception classique. La plupart des auteurs sont ainsi prêts à remettre en cause l'hypothèse de l'infaillibilité de la thèse de la connaissance *immédiate* qui est censée d'assurer la transparence des *qualia*. D'autres, comme N. Nelkin[11] et D. Rosenthal[12], vont encore plus loin et sacrifient l'hypothèse selon laquelle les *qualia* seraient intrinsèquement liés à des états *conscients*. Dans ce cas, les *qualia* sont traités comme des propriétés des états perceptuels qui peuvent se manifester *même en absence* d'un accès conscient explicite. Par exemple, les qualités sensibles des couleurs rouge et vert peuvent être considérées comme des simples propriétés des nos états perceptuels nous aidant à percevoir une tomate mûre comme étant plus proche de la couleur d'une orange sanguine, disons, que de celle d'une aubergine. Dans le même ordre d'idées, on peut imaginer le cas d'un individu qui s'arrête au passage-piétons affichant rouge, mais qui continue au vert, sans qu'il ait pour autant un méta-accès explicitement conscient et dirigé vers les qualités sensibles qui sous-tendent ses expériences de rouge et respectivement, de vert.

Les partisans de la théorie des *qualia* non-conscients sont prêts à concéder, bien évidemment, que les contenus d'expérience vécue peuvent surgir *aussi* par le truchement des expériences conscientes. Mais cela ne veut pas dire qu'il existerait un événement ou un état particulier de type « l'effet-que-cela-fait », pour reprendre l'expression de Th. Nagel, qui caractériserait uniquement le phénomène de la perception consciente du rouge ou du vert. Dans la version traditionnelle, le fait que la 'peinture' mentale qu'un individu projette sur son écran intérieur provoque, à la vue d'une tomate, une nuance spécifique de rouge, doit être considéré comme un fait caractérisant l'image interne elle-même, et non pas la façon dont elle est connectée aux choses extérieures, en occurrence, aux tomates. C'est dans ce sens que doit être comprise la propriété d'être « intrinsèques » des *qualia*.

Récemment, certains philosophes comme M. Tye[13] et *W.* Lycan[14], pour ne citer que ces deux noms, ont choisi de définir les *qualia* sous leur aspect de propriétés *relationnelles*, au détriment des propriétés intrinsèques, identifiant les qualités sensibles d'un état perceptuel à leur *contenu représentationnel*. Les défenseurs de la théorie

[11] Voir N. Nelkin, (1989), 'Unconscious sensations', *Philosophical Psychology*, 2: 129-41.
[12] Voir D. Rosenthal, (1991), 'The independence of consciousness and sensory quality', in E. Villanueva, (ed), *Consciousness.* Atascadero, CA: Ridgeview Publishing.
[13] M. Tye, (1995), *Ten Problems of Consciousness*, Cambridge, MA: MIT Press.
[14] W. Lycan, (1996), *Consciousness and Experience*, Cambridge, MA: MIT Press.

représentationnelle de la conscience affirment que les propriétés mentales d'un état perceptuel sont une affaire de comment ces états *représentent* le monde. Si ces états n'ont pas d'autres propriétés mentales mis à part leurs propriétés représentationnelles, alors deux états perceptuels ayant le même contenu ne peuvent pas être différents sous aucune autre description mentale. Bref, pour le représentationaliste, le contenu intentionnel d'un état perceptuel se résume à sa nature mentale. Selon Van Gulick[15], on peut exprimer la même idée à l'aide de la proposition suivante :

> (R) Il ne peut pas y avoir de modifications dans le mental sans provoquer des modifications dans le contenu représentationnel afférent. Les modifications dans le mental sont covariantes avec les modifications dans le représentationnel.

En clair, le fonctionaliste définit les entités mentales en termes de rôles causaux que ces derniers jouent au sein d'un réseau complexe d'inputs et d'outputs. Par conséquent, les entités avec lesquelles il opère sont essentiellement des entités *relationnelles*. Maintenant, la question qui se pose est : peut-on assimiler ce type d'entités mentales aux *qualia*, qui sont censés d'être des propriétés *intrinsèques* des états mentaux ? Si, comme l'a relevé Kim[16], ce qui est intrinsèque est par définition non-relationnel, toute description fonctionnaliste des *qualia* semble couver une contradiction interne.

Selon la synthèse qu'en propose Van Gulick[17], le fonctionnaliste a à sa disposition deux voies échappatoires pour éviter ce dilemme.

(a) Dans un premier temps, il peut abandonner l'idée selon laquelle les *qualia* seraient des propriétés intrinsèques. Pour un représentationaliste comme M. Tye[18], par exemple, il suffit d'interpréter *qualia* en termes de contenus représentationnels pour esquiver le conflit entre les niveaux intrinsèques/extrinsèques au sein du fonctionnalisme. Les *qualia* seraient assimilables dès lors aux *contenus* de nos états perceptuels, qui déterminent à eux seuls l'effet-que-cela-fait d'être dans un tel ou tel état particulier. C'est seulement à ces contenus que nous avons un accès introspectif. Cependant, cela ne manque pas de poser un problème quand on affirme que le contenu d'un état perceptuel est déterminé sinon intégralement, du moins en large mesure, par le tissu des relations causales qui sous-tendent des processus, états et faits du monde réel.

---

[15] "For the representationalist, a perceptual state's intentional content exhausts its mental nature. To put it in a slogan, there can be no mental difference without a difference in representational content". (R. Van Gulick, 2007: 387)
[16] J. Kim, (2005), *Physicalism or Something near Enough*. Princeton, NJ: Princeton University Press.
[17] (R. Van Gulick, 2007)
[18] (M. Tye, 1995)

Ce qui veut dire que le contenu d'un état est une propriété relationnelle et non pas intrinsèque.

(b) Une deuxième option serait de reformuler la distinction entre les notions d' « intrinsèque »/« extrinsèque » en insistant sur la façon dont ces termes s'appliquent aux systèmes complexes, multi-stratifiés (« many leveled systems »). Le but est de trouver le procédé et les dispositifs par le truchement desquels certaines propriétés intrinsèques localisables à un niveau spécifique de l'architecture organisationnelle d'un système peuvent émerger ou être réalisées par un système subjacent de relations ou structures. Dans cette acception, les *qualia* peuvent être regardés comme des propriétés à la fois intrinsèques et fonctionnelles, chevauchant deux niveaux distincts: un niveau manifeste, où ils représentent des états mentaux sur le mode dynamique de l'organisation d'un système ; et un niveau sous-jacent, composé d'éléments fonctionnellement structurés, et dont le rôle est celui de réaliser les modes occupant le niveau immédiatement supérieur.

Certains auteurs comme Dennett[19] considèrent que cette option est suffisamment convaincante pour réfuter l'existence même des *qualia*. D'autres, comme F. Dretske[20] et M. Tye[21] en tirent une morale fort différente. Les représentationalistes rejettent en bloc la négation de (R). Autrement dit, selon eux il ne peut pas y avoir de modifications des états mentaux indépendamment des variations du contenu représentationnel. Rien ne nous oblige à rejeter l'existence des *qualia*, mais seulement l'idée selon laquelle les *qualia* seraient des qualités *intrinsèques* des états perceptuels. Ces auteurs identifient *qualia* avec les contenus représentationnels, en s'appuyant sur l'idée que cette identification est la seule capable de satisfaire les conditions (i) et (ii) que nous avons mentionnées plus haut, nécessaires pour une caractérisation adéquate des *qualia*.

Grace à leur transparence, les états introspectifs nous permettent de saisir le monde tel qu'il est « re-présenté » à la perception, c'est-à-dire directement, et non pas à travers le voile des *propriétés* des représentations. Les partisans du représentationalisme rejettent ainsi vigoureusement l'idée selon laquelle l'être humain aurait en quelque sorte une prise *directe* sur les images mentales localisées sur l'écran interne de l'esprit. Seuls les contenus représentationnels remplissent les deux conditions (i) et (ii).

---

[19] (D. Dennett, 1990)
[20] F. Dretske, (1995), *Naturalizing the Mind*. Cambridge, Mass: The MIT Press, Bradford Books.
[21] M. Tye, (2000), *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.

A la lumière de ces considérations, il est évident qu'un traitement adéquat de la question qui nous préoccupe dans ce chapitre, à savoir (Q_F): « Le fonctionnalisme est-il en mesure d'expliquer le phénomène des *qualia* ? » dépend plus que jamais non seulement de l'espace conceptuel que l'on est disposé à accorder au terme « *qualia* », mais aussi de la manière dont on entend décrire l'ontologie des états physiques qui les sous-tendent.

4.2 L'argument du fossé dans l'explication

De manière générale, on s'accorde à dire qu'une des caractéristiques fondamentales des états mentaux est celle d'être des états conscients. Malgré les diverses tentatives de « caser » le phénomène de la conscience dans une niche respectable de l'étude du mental, force est de constater que le terme « conscience » reste éperdument ambigu. Cette ambigüité est imputable à la complexité et à la diversité des phénomènes qui sont traditionnellement associés au problème de la conscience.

Selon Chalmers[22], les problèmes qui gravitent autour du thème de la conscience peuvent être rangés en deux classes distinctes : problèmes faciles et problèmes difficiles. La classe des problèmes «faciles » concerne plus spécifiquement les tâches que doit exécuter un système cognitif générateur de comportements intelligents, et comprend habituellement les phénomènes suivants :

(1) la capacité de discrimination, catégorisation et réaction aux stimuli extérieurs ;
(2) l'intégration de l'information par un système cognitif ;
(3) la capacité de rapporter les états mentaux ;
(4) la capacité d'un système d'accéder à ses propres états internes ;
(5) la focalisation de l'attention ;
(6) le contrôle du comportement ;
(7) la distinction entre l'état de veille et l'état de sommeil[23].

L'étiquette de « facile » utilisée pour désigner l'investigation de ces aspects ne veut pas insinuer, bien évidemment, qu'elle ne traite pas de sujets importants concernant le fonctionnement ou le comportement des organismes conscients.

[22] (D. Chalmers, 1996)
[23] (D. Chalmers, 2007)

Seulement ces interrogations sont à traiter plutôt comme des « casse-têtes » que des véritables énigmes. Selon Chalmers, le véritable « côté difficile» du mystère de la conscience réside dans le problème de l'expérience phénoménale. Il convient de commencer par ce que nous savons avec certitude : les découvertes récentes des neurosciences nous permettent de situer le *locus* des processus mentaux dans un organe matériel hautement sophistiqué: le cerveau.

Comme nous l'avons déjà brièvement évoqué, le physicalisme est d'abord une doctrine ontologique à propos de ce que le monde *est*. C'est pourquoi remettre en question le physicalisme revient à rejeter l'ontologie qui le sous-tend. Par conséquent, le versant *ontologique* du dilemme de la conscience peut être résumé ainsi : par quelle alchimie s'opère-t-elle cette métamorphose magique qui permet de transformer la matière grise des cellules neuronales en l'or noble de la conscience ?

Les défis qui engagent le matérialisme dans la bataille contre ce mystère ne sont pas liés exclusivement aux fondements métaphysiques. Dans un article devenu désormais classique, J. Levine[24] observe que l'affaire du matérialisme nous confronte à un dilemme situé d'abord sur le versant *épistémique*. L'étude scientifique du mental nous indique, sans l'ombre d'un doute, que les états mentaux conscients sont physiquement réalisés par des états neurophysiologiques localisables dans le cerveau. Or, malgré cette confortable certitude, il se trouve, note Levine, que nous n'avons aucune idée précise de la manière dont les états mentaux sont implémentés dans l'organisation fonctionnelle du système nerveux. Il existe donc un *fossé explicatif*[25] séparant la connaissance du phénoménal de celle du physique.

Selon le résumé qu'en donne Van Gulick[26], l'épure de l'argument du fossé explicatif peut être retranscrite ainsi :

[24] J. Levine, (1983), 'Materialism and *Qualia*: The Explanatory Gap', *Pacific Philosophical Quarterly*, op. cit.

[25] Levine précise que l'idée de ce gouffre dans l'explication a été capturée bien avant lui par Locke dans sa discussion des relations entre les qualités primaire et les qualités secondaires, lorsqu'il s'agit de rendre intelligible l'emboîtement entre les deux niveaux. (J. Levine, 1983: 359)

[26] (R. Van Gulick, 1993/2008: 668)

(P1) Les *qualia* sont des entités primitivement simples.

(P2) Si les *qualia* sont des entités primitivement simples, alors tout rapport entre les *qualia* et la structure organisationnelle du substrat neuronal qui les sous-tend doit être arbitraire.

FOSSÉ EXPLICATIF

Par conséquent,

(C) Les rapports entre les *qualia* et leur substrat neuronal sont inintelligibles et nous conduisent à un fossé explicatif.

Une comparaison reprenant l'exemple paradigmatique de l'identité « eau=H_2O »[27] nous permettra de mieux saisir la présence de ce fameux gouffre. Considérons d'abord le cas explicitant le phénomène d'ébullition de l'eau: qu'explique-t-on au moyen de la théorie voulant que l'eau soit de l'H_2O ? Prenons le point d'ébullition de l'eau au niveau de la mer comme exemple d'une propriété macroscopique s'expliquant par la réduction de l'eau à de l'H_2O. Cette explication ressemblerait à ceci. Les molécules d'H_2O se déplacent à des vitesses variées. Certaines molécules plus rapprochées de la surface du liquide et se déplaçant plus rapidement que les autres, possèdent suffisamment d'énergie cinétique pour échapper aux forces d'attraction intermoléculaire de l'H_2O. Ces molécules entrent ainsi dans l'atmosphère : c'est l'évaporation. La valeur précise des forces d'attraction intermoléculaire détermine la pression de la vapeur d'une masse d'eau liquide, i.e. la pression exercée par les molécules cherchant à s'échapper vers un air saturé d'eau. La pression de la vapeur augmente proportionnellement à l'énergie cinétique moyenne des molécules. Lorsque la pression de la vapeur atteint le point où elle est égale à la pression atmosphérique, de grosses bulles se forment dans le liquide, et jaillissent à sa surface : l'eau bout.

« Etant donné cette explication, ou une élaboration suffisamment riche de celle-ci, j'affirme, écrit Levine, qu'il n'est pas possible de concevoir que l' H_2O ne bouille pas à 100°C au niveau de la mer (en présumant, encore, la constance du monde chimique)»[28].

[27] J. Levine, (1999), 'Conceivability, Identity, and the Explanatory Gap' in S.R. Hameroff, A. Kaszniak, et D. Chalmers, (Eds.) (1999), *Toward a Science of Consciousness III - The Third Tucson Discussion and Debates.* Cambridge Massachusetts: MIT Press, pp. 4-5.

[28] J. Levine, (1993), 'On leaving out what it's like", *in* M. Davies & G. Humphreys (eds.), *Consciousness, Psychological et Philosophical Essays*, 1993, Blackwell. Trad. fr. in D. Fisette et P.

Si nous essayons de transcrire ce raisonnement sous une forme plus rigoureuse, on obtient l'inférence suivante, que nous pouvons baptiser ici, pour plus de simplicité, [EAU] :

[EAU]

(1) Les molécules H_2O exercent une pression-vapeur P à l'énergie cinétique E.

(2) Au niveau de la mer, l'action de la pression-vapeur P a pour effet l'évaporation rapide des molécules dans l'air.

(3) La rapide évaporation dans l'air *est* l'ébullition.

(4) 212° F. est l'énergie cinétique E.

(5) L'eau est H_2O.

(6) *Ergo*, l'eau bout à 212° F. au niveau de la mer.

Supposons qu'un appareil sophistiqué permet d'indiquer que l'activation des fibres-C accompagne l'expérience qualitative spécifique qu'un individu S possède à la vue d'un rubis rouge. Nous obtenons alors l'argument analogue suivant, baptisé cette fois-ci [*QUALIA*] :

[*QUALIA*]

(7) S occupe un état cérébral C.

(8) Occuper l'état cérébral C c'est éprouver l'expérience du *quale* R-rougeur.

(9) *Ergo*, S est en train d'éprouver le *quale* R-rougeur.

Cette comparaison nous renseigne sur trois points essentiels.

(i) D'abord, il y a l'observation que, dans les deux cas, l'*explanans* (les énoncés qui font le travail explicatif) fonctionne comme une prémisse à partir de laquelle l'*explanandum* (les énoncés qui représentent le phénomène à expliquer) est déductivement dérivé.

(ii) Ensuite, les deux arguments font appel à des *prémisses-gonds* - l'énoncé (5) dans le cas [EAU] et (8) dans le cas [*QUALIA*] – qui identifient le phénomène à

Poirier, (2003), *Philosophie de l'esprit*, vol. II, Vrin, Paris, pp. 209-210. (Référence faite à l'édition française.)

expliquer à un autre phénomène descriptible dans un vocabulaire du niveau microphysique immédiatement inférieur.

(iii) Enfin, et c'est ici le point essentiel soulevé par Levine, à la lecture de ces deux arguments, nous avons l'étrange impression que l'argument [*QUALIA*] *n'explique pas* la nature du *quale* R-rougeur. On reste, pour ainsi dire, sur sa faim quant au pouvoir explicatif réel de l'identité « état cérébral-C = *quale* R-rougeur ».

A partir de là, on peut, suivant Chalmers[29] schématiser l'affaire du fossé explicatif sous la forme d'un argument à deux prémisses avec des conséquences dévastatrices pour l'ontologie moniste matérialiste:

(P1) Il y a un fossé épistémique entre les vérités physiques et les vérités phénoménales.

(P2) S'il y a un fossé épistémique entre les vérités physiques et les vérités phénoménales, alors il y a un fossé ontologique entre elles.

FOSSÉ ONTOLOGIQUE

Par conséquent,

(C) Le matérialisme est faux.

Les théoriciens physicalistes de la conscience réagissent de manière différente devant cette difficulté. D'un côté, celui des défenseurs du réductionnisme radical, se manifeste une volonté d'éliminer les *qualia* et avec eux, le fossé explicatif tout court. De l'autre côté, celui des partisans du non-réductionnisme et du dualisme, la tendance est au contraire à trouver des explications à ce problème sans sacrifier la conscience sur l'autel de la matérialité. De manière générale, ces réactions peuvent être rangées en trois groupes distincts :

(a) Certains rejettent le bien-fondé de la prémisse (1), niant la présence du fossé épistémique tout court.

(b) D'autres remettent en cause seulement la prémisse (2), autrement dit, ils acceptent le caractère insurmontable du « gouffre » épistémique qui sépare le

[29] D. Chalmers, (2003a), 'Consciousness and its place in Nature', *in* S. Stich, et T. Warfield (eds.) *The Blackwell Guide to Philosophy of Mind*, Blackwell Publishing, p. 108.

mental et le physique, mais rejettent la dérivabilité d'un fossé ontologique à partir d'un fossé épistémique.

(c) Enfin, d'autres acceptent le bien-fondé du fossé épistémique, mais ils préfèrent adopter une attitude quasi-messianique, projetés dans l'espoir qu'un jour les avancées des neurosciences permettront de dissoudre le fossé (rejet de la conclusion (3)).

La solution préférée par Levine[30], l'artisan même de l'argument du fossé explicatif, est de montrer que le clivage entre le phénoménal et le physique est seulement d'ordre épistémique, et non pas ontologique. Il y a – insiste Levine – une asymétrie fondamentale entre le mental et le physique. A cela, le matérialiste peut riposter ainsi : il est vrai que le défi de l'argument vise l'impossibilité de combler le fossé. Mais c'est précisément ce que le partisan matérialiste affirme, à savoir l'impossibilité de jeter un pont entre les deux rives d'un putatif gouffre. Il ne faut donc pas s'étonner, lorsqu'on demande l'impossible, de ne pas l'obtenir. A cela Levine réplique en exhortant le matérialiste à offrir une explication non pas du fossé ontologique séparant les propriétés phénoménales de leur substrat physique (en supposant qu'un tel fossé existe), mais du fossé dans l'explication. Car paradoxalement, note Levine, le matérialiste prétend pouvoir expliquer *pourquoi* il *ne* peut *pas* expliquer le mental en termes de propriétés physiques!

En clair, selon Levine[31], les états ou les concepts phénoménaux peuvent effectivement être réalisés par des systèmes physiques. Le problème n'est pas là. Le problème c'est que même si tel était le cas, nous n'arrivons toujours pas à expliquer la métamorphose du substrat matériel, neuronal, en une expérience phénoménale qualitative selon le modèle de l'explication de l'ébullition de l'eau *via* le comportement des molécules de l'H_2O, Et c'est précisément cette inintelligibilité qui accentue le caractère troublant du problème des rapports entre le phénoménal et le physique.

Pour beaucoup d'auteurs, la solution à ce problème passe par une révision fondamentale des concepts impliqués dans le puzzle, que ce soit au niveau de la

[30] J. Levine, (1999) 'Conceivability, Identity, and the Explanatory Gap' in S.R. Hameroff, A.W. Kaszniak, et D. Chalmers, (eds.) (1999), *Toward a Science of Consciousness III - The Third Tucson Discussion and Debates.* Cambridge Massachusetts: MIT Press.

[31] J. Levine, (2007) 'Anti-materialist arguments and influential replies', in Max Velmans & Susan Schneider (eds.), *The Blackwell Companion to Consciousness*, op. cit.

redéfinition du concept de « phénoménalité » comme chez R. Van Gulick[32], ou du repositionnement du terme « physique » comme chez G. Strawson[33] qui propose, par exemple, d'élargir le contenu du concept de « physique » afin de pouvoir laisser une place à l'expérience phénoménale dans l'ameublement physicaliste. La présentation précédente du problème du fossé suffit toutefois à montrer que plusieurs voies de résolution du conflit sont *a priori* possibles selon qu'on remet en cause l'une ou l'autre des hypothèses qui l'engendrent.

En conclusion, il est clair que le fonctionnalisme se heurte à une foule de problèmes que leurs partisans s'affaireront à résoudre. Maintenant, la leçon que l'on doit tirer de l'affaire du fossé explicatif est que plusieurs niveaux d'organisation fonctionnelle sont en jeu. Un organisme peut, par exemple, manifester un comportement extérieur en parfait accord avec les généralisations de la psychologie du sens commun, et pourtant avoir un ressort psychologique ou motivationnel radicalement différent de celui d'un individu « normal ». Seule une recherche empirique minutieuse et patiente pourrait nous révéler l'existence de cette profonde différenciation. Donc, lorsqu'il s'agit d'établir si deux organismes partagent le même type d'état conscient, ou du moins d'établir si un robot, par exemple, a ou non accès à la conscience (est conscient tout court), la question qui se pose naturellement est : quel est le niveau d'organisation qui détermine la présence ou l'absence des états conscients ? Comme le relève Levine[34], on peut connaître de manière exhaustive toute l'information physique relative à la structure interne d'un organisme, et pourtant ignorer si cette créature est oui ou non consciente, si elle éprouve une expérience qualitative de type « l'effet-que-cela-fait ». Comment cela est-il possible? Quel est le fait qui nous échappe ici ? S'agit-il d'une limitation due à notre ignorance, ou d'une barrière infranchissable intrinsèque au monde extérieur ? Nous aurons l'occasion de revenir plus amplement sur le concept d' « *ignorance* » lorsque nous examinerons le concept d'humilité Ramseyenne » tel qu'il fut développé par D. Lewis[35]. Ainsi qu'on va commencer à le voir dans la section suivante, l'affaire du fossé explicatif provoque

[32] R. Van Gulick, (1999), 'Conceiving Beyond Our Means: The Limits of Thought Experiments' in S.R. Hameroff, A.W. Kaszniak, et D. Chalmers, (eds.) op. cit.
[33] G. Strawson, (1994), *Mental Reality*, MIT Press.
[34] (J. Levine, 2007: 377)
[35] D. Lewis, (2009), 'Ramseyan Humility' in D. Braddon-Mitchell & R. Nola (eds.), *Conceptual Analysis and Philosophical Naturalism*, Cambridge Mass.: MIT Press.

une fissure non-négligeable dans l'armature de l'édifice physicaliste, avec des conséquences majeures sur l'évolution du problème de la conscience phénoménale.

4.3 Autres arguments antimatérialistes

Incontestablement, la question de la conscience fascine et ne laisse personne indifférent. Rien d'étonnant à cela. Récemment, plusieurs arguments ont été élaborés afin de marquer les véritables enjeux de cette « résistance » qu'oppose la conscience aux diverses tentatives de réduction matérialistes, freinant ainsi les ambitions de complétude et d'universalité du physicalisme. A nouveau, une épaisse et fascinante littérature traitant de ces aspects a vu le jour ces dernières décennies. La plupart des analyses critiques qui remettent en question la capacité du matérialisme d'offrir une explication complète du mental reposent sur des arguments de type *a priori*, des expériences de pensée qui tentent de prouver que les *qualia* échappent au réseau explicatif physicaliste mis en place.

Dans ce qui suit, nous allons brièvement présenter la structure de base de ces arguments, ainsi que les conséquences qu'ils entraînent sur la physionomie de débats actuels dans la philosophie de l'esprit, sans entrer néanmoins dans les sinuosités d'une discussion qui s'allongerait facilement, et qui mériterait, à elle seule, un traitement minutieux dans un essai à part. Comme nous l'avons déjà mentionné, un seul de ces arguments antimatérialistes sera la 'vedette' de notre enquête ici. Cependant, et afin d'approfondir la discussion sur l'argument de la connaissance de Jackson, on ne pourra faire l'économie d'un bref détour par la présentation des hypothèses d'arrière-plan qui caractérisent et accompagnent la discussion des autres arguments avec lesquels Mary partage l'affiche de la rébellion antimatérialiste.

4.3.1 *Qualia* absents, *qualia* inversés, et autres curiosités

La plupart des objections antifonctionnalistes mettent en place des expériences de pensée stipulant l'existence d'organismes fonctionnellement semblables à nous, mais dont les expériences ne possèdent pas les mêmes aspects qualitatifs qui accompagnent nos propres expériences. On imagine également des organismes fonctionnellement semblables à nous, mais ne possédant *aucune* expérience qualitative que ce soit.

Ces arguments antiphysicalistes s'agencent donc autour de la prémisse selon laquelle l'explication du mental en termes de structures physiques spatio-temporelles et de rôles causaux omet d'expliquer un aspect essentiel de l'expérience consciente, à savoir la nature intrinsèquement qualitative de l'expérience vécue. Si l'on veut tracer l'épure de ce type d'argument, nous obtenons ceci :

ARGUMENT ANTIFONCTIONNALISTE

(1) Les explications en termes physiques rendent compte des structures et fonctions, tout au plus.

(2) Expliquer les structures et les fonctions omet d'expliquer l'aspect qualitatif, phénoménal de la conscience.

Par conséquent,

(3) Aucune explication en termes exclusivement physiques ne peut expliquer le phénomène de la conscience.

De manière générale, les plus influents arguments antifonctionnalistes mettent en scène des expériences de pensée construites autour des certaines situations imaginaires. Les exemples les plus connus sont l'expérience des *qualia inversés* et des *qualia absents*. Une discussion minutieuse de ces arguments déborderait du cadre de cet essai. Nous nous limiterons d'en fournir un simple compte-rendu avant de s'arrêter plus longuement sur l'argument qui constitue l'objet de notre enquête ici, à savoir l'argument de la connaissance de F. Jackson.

Commençons avec le cas du *spectre inversé*[36]. Le point de départ de cette expérience de pensée est une intuition dont l'origine remonte au moins jusqu'à Locke[37]. Imaginons le cas d'un individu Y dont l'architecture interne fonctionnelle est analogue à celle d'un individu normal *X*, à la différence que chez le premier les *qualia* qui jouent certaines fonctions sont en réalité le revers des *qualia* qui occupent le même rôle chez l'individu « normal ». Exemple : lorsque *X* et Y regardent simultanément un rubis, les deux le décrivent comme étant « rouge », les deux se trouvent dans l'état fonctionnel-physique qui sous-tend et relie la propriété de « rougeur » correspondant à l'objet externe donné. Et pourtant, dit l'expérience de pensée, le *quale* interne qui se produit dans l'esprit de *X* lorsque celui-ci perçoit un rubis, est en réalité le même qui se produit

[36] Voir N. Block & J. Fodor, (1972), 'What Psychological States Are Not', *Philosophical Review*, 81: 159-81.

[37] "The same Object should produce in several Men's Minds different Ideas at the same time; v.g. the Idea, that a Violet produces in one Man's Mind by his Eyes, were the same that a Marigold produced in another Man's and *vice versa*". (Voir J. Locke, 1689, livre II, ch. 32, section 15)

chez Y lorsque ce dernier perçoit une émeraude, et *vice versa*. Pour les défenseurs de cet argument, il existe une intuition très forte selon laquelle, malgré la conformité de l'organisation fonctionnelle et du comportement identique des deux protagonistes, il existe une différence essentielle entre le *quale* de *X* et celui de Y lorsqu'ils perçoivent simultanément une tomate bien mûre, disons.

L'argument des *qualia* absents[38], plus radical encore, nous invite à imaginer le cas d'un personnage, appelons-le Z, qui possède une organisation fonctionnelle interne identique à celle des individus normaux, à cette différence près : Z n'éprouve *aucune* expérience qualitative d'aucune sorte. Par exemple, lorsque Z perçoit un rubis, il se trouve dans l'état fonctionnel interne identique à celui d'un individu normal *X* à la vue d'un rubis. Bref, Z est ce que Chalmers[39] appelle un « zombie ». Selon Van Gulick[40], la structure de l'argument peut être retranscrite ainsi :

QUALIA ABSENTS

P1 La structure fonctionnelle du mental doit pouvoir être réalisée, du moins en principe, par des systèmes dépourvus de *qualia*.

P2 Un modèle explicatif M d'une propriété donnée F explique F (ou ce que F *est*) uniquement si rien ne peut être une réalisation de M sans être en même temps une réalisation de F.

Par conséquent,

(C) Le fonctionnalisme ne peut pas expliquer ce que l'expérience phénoménale *est*.

Les ripostes fonctionnalistes s'organisent selon plusieurs stratégies. Certains[41] remettent en cause la cohérence et la validité de ce type d'expérience de pensée, rejetant l'idée selon laquelle ces cas imaginaires font état de possibilités réelles. D'autres[42] préfèrent remettre en cause la validité des conclusions auxquelles amènent ces expériences de pensée. Chacune de ces deux options ont connu des nombreuses variations, versions, et réponses critiques.

---

[38] (N. Block, 1980a). Voir aussi N. Block, (1980b), 'Are absent *qualia* impossible?', *Philosophical Review*, 89/2: 257-74.
[39] (D. Chalmers, 1996)
[40] (R. Van Gulick, 1993/2008 : 670)
[41] (Voir D. Dennett, 1991)
[42] Voir S. Shoemaker, (1990), 'Qualities and *qualia*: what's in the mind', *Philosophy and Phenomenological Research*, Supplement, 50: 109-131.

Par exemple, on a émis des réserves quant à la plausibilité même du projet d'établir des distinctions *intersubjectives*, capturées du point de vue de la troisième personne, entre des propriétés qui par définition restent essentiellement privées, associées au point de vue de la première personne. Après tout, si l'on stipule que les individus « normaux » et les « inversés » (ou les « zombies ») exhibent un comportement identique, et possèdent la même structure fonctionnelle physique, par quel moyen serait-il possible alors de vérifier qu'il y ait ou pas une différenciation entre le *quale* de *X* et celui de Y, pour reprendre l'exemple donné plus tôt?

La difficulté principale concerne la possibilité de justifier un jugement dans le cas des comparaisons intersubjectives. Car s'il est relativement facile de justifier les comparaisons intra-subjectives – Marcel Proust inspecte et compare ses propres *qualia* à des différents moments de sa journée de lassitude à Combray, disons – il est nettement moins aisé d'établir un critère neutre pour justifier une comparaison intersubjective. Même l'utilisation de certains procédés controversés comme la télépathie ou l'empathie ne peuvent pas nous être de grand secours ici, car ils comportent systématiquement un inéliminable résidu d'incertitude quant à la véritable fiabilité de leurs résultats.

A nouveau, une discussion[43] détaillée de ces points nous éloignerait de l'objectif que nous poursuivons ici. Retenons seulement que l'argument des *qualia* inversés concerne uniquement le cas où les *qualia* sont considérés comme des propriétés intrinsèques, et non pas relationnelles, des états perceptuels. Par conséquent, ceux qui identifient *qualia* à des contenus représentationnels[44], rejettent *ab initio* la possibilité des *qualia* absents. Pour les représentationalistes, le contenu d'un état mental est fixé par les relations qui le connectent à un contexte extérieur, autrement dit aux états réels qu'il compte représenter de manière fiable. Par conséquent, si toutes ces relations sont maintenues inchangées, alors leurs contenus demeurent inchangés aussi, et les inversions des *qualia* deviennent dès lors empiriquement impossibles. D'autre part, si l'on est adepte du fonctionnalisme computationnel, alors la question de savoir s'il existe des robots partageant pleinement avec l'être humain la même organisation fonctionnelle et cognitive, tout en étant dépourvus de vie mentale, devient d'intérêt central. De façon analogue, si l'on définit les fonctions en termes de rôles causaux, la

[43] Pour une discussion plus détaillée voir S. Shoemaker (1982), 'The inverted spectrum', *Journal of Philosophy* 79, 357–81; C. Hardin, (1986), *Color for Philosophers*. Indianapolis: Hackett; St. White (1986).
[44] (F. Dretske, 1995)

question qui se pose alors est celle de savoir si un système peut avoir un état D-douleur qui correspondrait aux rôles causaux afférents à l'état-D « Douleur », mais qui néanmoins resterait dénué de toute expérience qualitative.

Comme l'a relevé Van Gulick[45], si le cas des *qualia* absents s'avère possible, alors le fonctionnaliste pourrait se voir dans l'impossibilité de spécifier les conditions qui désignent indexicalement un *quale* particulier. Il pourrait, tout au plus, identifier une classe équivalente de *qualia*. Mais si les conditions visent certains organismes ou systèmes caractérisés par un *certain* type (spécifique) de *qualia*, alors ces caractérisations peuvent être considérées comme satisfaisantes pour une explication de la conscience phénoménale et de ses états de type « l'effet-que-cela-fait ». Dans ce cas, une division spécialisée du travail devrait laisser au neurophysiologiste le soin d'identifier avec précision le *quale* qui accompagne un état fonctionnel particulier. Considéré sous cet angle, le cas des *qualia absents* ne constituerait plus une menace pour le fonctionnaliste.

Regardons de plus près la définition « naïve » standard d'un état D-douleur[46]: de manière générale, les états-D d'un système sont le produit des endommagements ou des menaces de nuire au système, endommagements qui à leur tour causent un comportement de repli et d'évitement, suite auquel les états-D agiront comme des déclics de renforcement négatif, et dont la tâche future sera d'inhiber les éventuels faits ou comportements qui avaient conduit le système vers la situation-D dans le passé. Les états-D produisent, bien entendu, la motivation et le désir de mettre terme à cet état. Ce désir se trouve ainsi projeté directement au sommet de l'hiérarchie de préférences du système endommagé. Les états-D interviennent promptement et brouillent la capacité du système de se focaliser sur d'autres processus ayant lieu au même moment (en même temps), ce qui a pour conséquence leur stockage dans une niche particulière de la mémoire en vue d'une utilisation future.

L'intuition qui est au cœur de l'affaire des *qualia* absents affirme que les états d'un système donné peuvent satisfaire toutes les conditions susmentionnées et que, malgré tout, ces systèmes ne possèdent pas d'expériences qualitatives de type « l'effet-que-cela-fait d'éprouver une douleur ». Ce système serait ce qu'on appelle un

[45] R. Van Gulick, (1993), 'Understanding the phenomenal mind: Are we all just armadillos?' in M. Davies et G. Humphreys, (eds)., *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell. Réédité in W. Lycan & J. Prinz (eds.), (1990/2008), *Mind and Cognition – an Anthology*, Blackwell Publishing. Nous faisons référence ici à cette dernière édition.

[46] (R. Van Gulick, 2007: 389-90)

« zombie ». Il est néanmoins possible, comme le note Van Gulick, de mettre à l'épreuve cette intuition. Pour voir cela, il suffit de se rappeler qu'habituellement, en présence d'un phénomène particulier comme la douleur, un individu est confronté à la *croyance* qu'il éprouve de la douleur. Il est difficile, dans ce cas, de faire abstraction de cette croyance, à moins qu'elle ne soit pas provoquée par des états réels de douleur. La plupart des fonctionnalistes adhèrent à l'idée selon laquelle le contenu d'un état représentationnel comme la croyance est au moins en partie déterminé par des propriétés du monde extérieur. Si tel est le cas, alors il devient impossible d'avoir des croyances de type « *X* éprouve de la douleur » sauf si ces croyances sont effectivement causées par les états-D réels de *X*, chose impossible dans le cas des systèmes de type-zombie, car ces derniers n'éprouvent pas réellement de la douleur. Par conséquent, si le fonctionnaliste s'autorise de faire appel aux croyances insérées dans la chaîne des rôles causaux que l'on associe habituellement au phénomène de douleur, il semblerait que les systèmes non-*qualia* se retrouvent dans l'incapacité de satisfaire les conditions susmentionnées.

A cela, les antifonctionnalistes peuvent riposter, à juste titre, en rappelant qu'une telle manœuvre repose sur un raisonnement circulaire, dans la mesure où l'on fait appel à des états mentaux, à savoir les croyances qui, de par leur nature, *sont* des états qualitatifs. La question de savoir si le fonctionnaliste triche en faisant appel aux croyances dans l'explication des *qualia*, dépend, bien évidemment, des concepts que l'on s'autorise à invoquer dans la spécification des rôles causaux agissant à l'intérieur même de cette théorie. Pour dire les choses brièvement, plus le fonctionnalisme mise sur l'emploi des notions « quasi-qualitatives » pour spécifier les rôles causaux impliqués, plus le coût de cette opération est important, car moins ils arrivent à expliquer les *qualia* en utilisant des termes non-qualitatifs.

Une version similaire[47] de cette stratégie consiste à expliquer *qualia* en termes de dispositions à produire des croyances fondées sur la perception (« perceptual-based beliefs ») et l'analyse des similarités et des ressemblances objectives. Ainsi compris, les *qualia* seraient des propriétés de nos états perceptuels qui comportent une structure agencée autour des relations de ressemblances établissant des similarités entre les états, et dont le rôle serait celui de produire des croyances concernant leurs éventuels états, faits, et processus équivalents que l'on retrouve dans le réel, et vers lesquels les états

[47]Voir D. Rosenthal, (1991), 'The independence of consciousness and sensory quality', in E. Villanueva, (ed). *Consciousness*, Atascadero, CA: Ridgeview Publishing.

perceptuels sont orientés. Ce type de modèle, observe Van Gulick[48], est satisfaisant, à condition qu'il y subsume l'hypothèse des *qualia* non-conscients. Or, beaucoup de penseurs restent sceptiques devant cette option. On n'ira pas plus loin que ces quelques remarques en ce qui concerne les arguments des *qualia* absents et du spectre inversé. Tournons-nous à présent vers un autre argument antimatérialiste très influent dans la littérature philosophique contemporaine.

4.3.2 L'argument de la 'concevabilité'

Il s'agit d'un argument devenu célèbre dans la littérature sous l'insolite titre de « l'affaire des zombies ». Selon Chalmers[49], son principal artisan, il est parfaitement possible de concevoir un système physiquement identique à celui d'un être conscient, mais dépourvu d'états conscients, bref de concevoir un « zombie » Z. Un tel système serait, du point de vue de sa constitution physique, parfaitement identique à une personne consciente « normale »*X*. A contempler la paire *X*/Z depuis l'extérieur, du point de vue de la troisième personne, il nous serait impossible de distinguer lequel d'entre eux est un zombie. C'est seulement du point de vue de la première personne, que l'on pourrait différencier entre l'expérience qualitative accompagnant, disons, l'effet-que-cela-fait-d'être-David-Chalmers, et le vide expérientiel qui caractérise le zombie-Chalmers. Bref, il n'existe pas un « effet-que-cela-fait-d'être-un-zombie-David-Chalmers », pour la simple bonne raison que ce dernier ne possède pas d'états expérientiels. Certes, on peut évoquer un tas de bonnes raisons de croire que des créatures de ce type n'existent pas effectivement dans le monde. Mais l'objectif de l'argument est de montrer qu'il nous est relativement facile de *concevoir* un monde possible peuplé de créatures qui seraient nos parfaites sosies physiques, mais qui seraient dépourvues de toute expérience « interne » consciente. Après tout, Dieu lui-même, dans sa toute-puissance, aurait pu choisir de créer un tel « zombie-land ».

L'épure de *l'argument de la concevabilité*[50] peut être tracée ainsi :

---

[48] (R. Van Gulick, 2007)
[49] (D. Chalmers, 1996)
[50] D'autres versions de l'argument de la concevabilité ont été avancées par J. Campbell (1970), *Body and Mind*. London: Doubleday; S. Kripke (1980); G. Bealer, 'Mental Properties', *Journal of Philosophy*, 91: 185–208.

ZOMBIE 1

(1) L'existence des zombies est concevable.

(2) Si l'existence des zombies est concevable, alors leur existence est métaphysiquement possible.

(3) Si l'existence des zombies est métaphysiquement possible, alors la conscience est non-physique.

Par conséquent,

(4) La conscience est non-physique.

Pour plus de précision, l'argument est parfois élaboré de la façon suivante[51] :

ZOMBIE 2

(1) Il est concevable que $P \wedge \neg Q$.

(2) S'il est concevable que $P \wedge \neg Q$, alors il est métaphysiquement possible que $P \wedge \neg Q$.

(3) S'il est métaphysiquement possible que $P \wedge \neg Q$, alors le matérialisme est faux.

Par conséquent,

(4) Le matérialisme est faux.

L'argument de la concevabilité présente un incontestable air de famille avec l'argument des *qualia* absents, à une différence près : il ne se limite pas aux seuls cas des doubles fonctionnels. Plusieurs points sont à signaler ici. Premièrement, une partie importante de la structure de l'argument repose sur l'étendue de la plage des significations que l'on est prêts à accorder au terme de « concevabilité ». D'une part, il est évident que l'existence des zombies semble une hypothèse étrange, ou du moins hautement improbable, au vu des données empiriques que l'on possède actuellement sur l'activité cérébrale d'un individu normal. Mais l'argument de Chalmers contourne cette intuition, car il fait appel à une acception plus « technique » du terme concevabilité, dans le sens où il s'agit d'établir si le concept de « zombie » n'implique aucune contradiction *conceptuelle* intrinsèque interne. L'intuition selon laquelle l'existence des zombies semblent *prima facie* hautement improbable, ne constitue évidement pas une raison suffisante pour réfuter leur *possible* existence. D'autre part, la question de l'intelligibilité d'une situation n'est pas toujours énonçable en termes d'observation empirique ou de bon sens. Il existe des situations qui, bien qu'à première

[51] Le signe "$\wedge$" représente "et", le signe "$\neg$" représente « non ». (D. Chalmers, 2003a)

vue concevables et plausibles, s'avèrent, à un examen plus attentif, profondément incohérentes. Par exemple, dans le cas de certains théorèmes mathématiques complexes, avant même d'arriver à une conclusion, le faux du théorème en cause semble parfaitement concevable, et pourtant, une fois la preuve apportée, on s'aperçoit qu'on a failli se faire piéger, et que le théorème est bel et bien valide. Bref, la frontière qui sépare le concevable de l'inconcevable est loin d'être fixée avec netteté.

Admettons qu'on accepte la prémisse (1) « les zombies sont concevables ». Ce qui distingue le matérialiste de son rival le dualiste n'est pas l'hypothèse d'un lien entre le mental et le physique, mais plutôt la *nature* de ce lien. Là où le dualiste des propriétés voit un rapport causal reliant deux propriétés ontologiquement distinctes, le matérialiste voit un lien constitutif entre deux propriétés dont une (le mental) est réductible à l'autre (le physique). Un moyen commode de caractériser la nature de ce lien intrinsèque est de faire appel à la notion métaphysique de survenance que nous avons déjà rencontrée. Rappelons-nous qu'un ensemble de propriétés P_1 survient sur un ensemble de propriétés P_2 si et seulement si une différenciation ou une variation dans P_1 entraîne nécessairement (du point de vue logique et/ou métaphysique) une différenciation ou une variation dans P_2. Par conséquent, si chaque monde possible qui est un double physique minimal du monde réel, est en même temps un double *simpliciter* du monde réel, alors la thèse de la survenance peut être appliquée avec succès pour décrire la dépendance métaphysiquement nécessaire du mental au physique.

La concevabilité des zombies lance un véritable défi au matérialisme. Comment expliquer le fait qu'une situation qui semble *empiriquement* impossible peut être *logiquement* ou *conceptuellement* concevable ? Car si un cas est concevable, alors il n'est pas *a priori* impossible, et le matérialisme est mis en danger.

Une manière commode de répondre à ce défi consiste à rappeler que la simple concevabilité d'un fait ne garantit pas *ipso facto* sa possibilité. Imaginons l'exemple d'un monde possible *w* où il existe de l'eau, mais non pas de l'H_2O. Cette description ne semble aucunement contradictoire ou incohérente en soi, et pourtant, si l'eau *est* H_2O, l'existence de *w* devient tout simplement impossible. Les énoncés d'identité, insistent les partisans du matérialisme, expriment des vérités nécessaires, mais ils ne sont pas toujours formellement ou conceptuellement nécessaires. Par conséquent, si les états mentaux conscients sont identiques avec les états neurophysiologiques ou fonctionnels du cerveau, alors il est impossible d'en avoir deux doubles physiques dont

un serait conscient et l'autre pas. La simple *concevabilité* conceptuelle ne garantit pas la *possibilité* métaphysique.

La riposte des antimatérialistes comme Kripke[52] et Chalmers[53] ne se laissera pas attendre. Vrai ! – disent-ils – pas tout ce qui est concevable est *ipso facto* possible. Parfois les propositions d'identité peuvent exprimer des vérités nécessaires en dépit du fait que l'on puisse douter de la cohérence ou la validité des énoncés eux-mêmes. Mais lorsque nous sommes confrontés à des situations de ce type, il existe une explication plausible de cette apparente discordance. Repenchons-nous sur le cas de l'identité paradigmatique eau/H_2O. S'il s'avère que l'eau *est* réellement H_2O, alors, par voie de loi de l'identité logique, l'eau *ne peut pas être* quelque chose d'autre. Et pourtant, quelqu'un pourrait remettre en cause de la vérité de cette identité sans pour autant commettre une erreur logique, ou faire preuve d'incohérence conceptuelle. Comment expliquer alors la possibilité de ce glissement entre la concevabilité et la possibilité dans ce cas précis?

Dans l'exemple qui précède, la substance unique de type eau/H_2O étale une palette de propriétés distinctes, comme « être liquide » ou « avoir une certaine structure moléculaire », ou « trouvable dans les rivières ». Il est évident, néanmoins, que des entités autres que l'eau/H_2O peuvent arborer à la surface les mêmes propriétés extérieures qui nous servent habituellement à caractériser l'eau. Or, lorsque nous désignons la substance « eau/ H_2O » à l'aide du mot « eau », nous faisons référence implicitement aux propriétés familières qui l'accompagnent d'ordinaire dans la vie de tous les jours ; et lorsque nous y référons *via* le concept H_2O, nous désignons implicitement les propriétés spécifiques qui caractérisent sa structure moléculaire fondamentale.

Puisqu'il est réellement possible qu'une entité affichant certaines propriétés de surface (par exemple la propriété de « trouvable dans les rivières ») n'ait pas la structure moléculaire spécifique H_2O – cet exemple nous permettant par ailleurs de mieux saisir le contenu de l'idée que l'eau pourrait *ne pas être* H_2O – il n'y a aucune incohérence intrinsèque dans la pensée « eau *n*'est *pas* H_2O », malgré la thèse que *si l'eau est* H_2O*, alors elle l'est nécessairement*.

Cet exemple nous aide à saisir le fossé qui peut surgir entre la *concevabilité* et la *possibilité* dans certaines situations. Cependant, lorsqu'on essaie d'appliquer la même

[52] (S. Kripke, 1980)
[53] (D. Chalmers, 1996)

stratégie explicative pour décrypter le cas des rapports entre le mental et le physique, on heurte d'emblée un très sérieux obstacle. Reprenons l'exemple devenu standard de l'identification de la douleur avec un état neurophysiologique spécifique, N. Si la douleur est « être dans l'état N », alors étant donné que, par définition, chaque entité est ce qu'elle est et non pas quelque chose d'autre, il est impossible pour un organisme d'éprouver de la douleur sans être de par là même dans l'état N. Et pourtant, il est bien évidemment concevable qu'un organisme puisse ressentir de la douleur sans être dans l'état N. Comment expliquer cela ? Essayons d'appliquer la même procédure que celle que venons d'utiliser dans le cas eau/ H_2O.

Il existe plusieurs propriétés par le truchement desquelles nous déterminons l'identité douleur/état N, ainsi que ses attributs neurophysiologiques. La question est : quelle est la propriété de la douleur qui est saisie par le concept à l'aide duquel nous désignons ordinairement le mot « douleur » ? N'est-il pas vrai que cette propriété *est* le caractère qualitatif qui accompagne l'effet-que-cela-fait d'éprouver de la douleur ? Mais si tel est le cas, et si l'on admet que cette propriété phénoménale peut être attribuée à des individus qui ne se trouvent pas dans l'état N, alors il s'ensuit que les propriétés mentales – qui constituent, rappelons-nous, le cœur et l'enjeu de l'affaire – ne sont ni identiques, ni réductibles aux propriétés physiques. Bref, l'argument de la concevabilité représente un sérieux défi pour les défenseurs du matérialisme, car si les zombies sont réellement concevables, alors l'échec du fonctionnalisme dans l'explication complète du mental est immédiat et sans appel. A l'heure même où nous écrivons ces lignes, la question de savoir si les zombies sont oui ou non concevables fait couler beaucoup d'encre dans le cercle des philosophes analytiques.

Récapitulons, pour finir, le tableau comparatif des divers arguments antiphysicalistes mentionnés jusqu'ici. Selon P. Carruthers[54] le schéma représentant le cœur de l'affaire des *qualia* inversés et des *qualia* absents peut être résumé ainsi :

[54] (P. Carruthers, 2004:9)

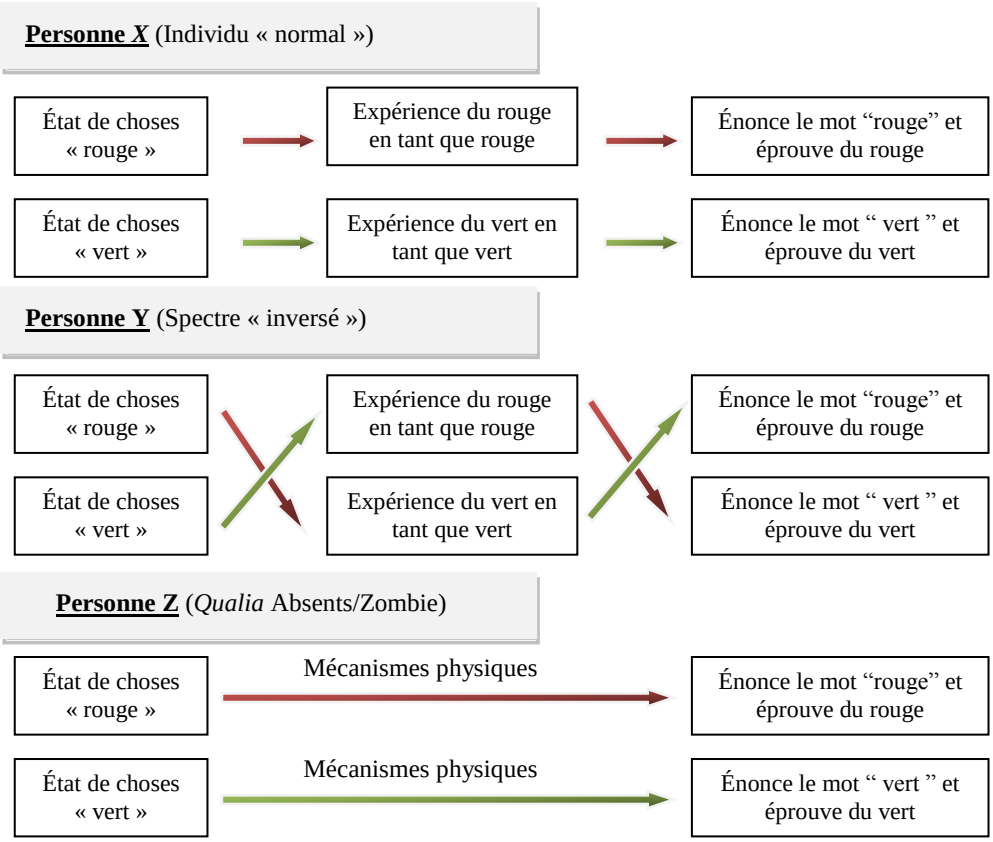


Figure 9: Les arguments des *qualia* absents et du spectre inversé

Bien que les arguments impliquant les *qualia* absents ou inversés soient les plus influents dans la philosophie de l'esprit d'aujourd'hui, ils ne sont pas les seuls à mettre à l'épreuve les théories matérialistes de l'esprit. D'autres arguments antifonctionnalistes ont été également avancés, comme par exemple les arguments qui mettent en scène les conséquences surprenantes de ce qu'on peut appeler des 'réalisations étranges', comme l'argument de la population chinoise de N. Block[55].

Pour terminer, rappelons seulement que les interrogations qui tournent autour de la question des expériences qualitatives mettent en évidence le dilemme auquel se confronte tout fonctionnaliste dès qu'il s'agit de spécifier les concepts qu'il s'autorise à

[55] Block nous propose d'imaginer l'expérience de pensée suivante : supposons qu'on arrive à connecter la totalité du peuple chinois à l'aide d'un réseau téléphonique géant organisé de telle manière qu'il puisse implémenter la totalité des rôles causaux qui caractérisent les états neurologiques d'un cerveau humain. Par exemple, lorsqu'un certain input déclenche l'activation d'un ensemble de dix neurones, disons, alors la personne recevant un coup de fil (l'équivalent de l'input) doit appeler les dix personnes afférentes qui se trouvent sur sa liste causale. Une question surgit alors : si chaque personne est ainsi connectée au réseau global, est-ce que la totalité du peuple chinois peut être considérée comme étant un *Suj*et conscient, à la manière dont un cerveau humain l'est ? Est-ce qu'un tel système peut éprouver de la douleur, ou avoir des sensations visuelles ? Block répond par la négative. Intuitivement, il nous est difficile, voire impossible, de définir un tel système comme un système « conscient ». Par conséquent, insiste Block, être conscient ne veut pas dire 'être identique avec la possession d'une certaine organisation fonctionnelle'. (Voir N. Block, 1980a :268-305).

employer dans la description des rôles fonctionnels. Au-delà de la question d'établir quelle est la *nature* de ces fonctions (computationnelle, causale, téléologique, etc.), l'aspect fondamental du problème que nous comptons aborder ici concerne, comme nous allons le voir par la suite, la portée explicative des *concepts phénoménaux* ou qualitatifs dans la spécification des relations et des interactions causales autour desquelles s'agence la vie mentale du sujet.

Jusqu'ici nous avons esquissé la trame des principaux arguments antimatérialistes, en insistant sur le caractère fondateur, si l'on peut appeler ainsi, de l'argument du fossé dans l'explication chez Levine. A partir de là, il existe, selon Chalmers, une manière commode de classer et de visualiser les différentes variantes du matérialisme actuel en prenant comme point de repère la solution que ses défenseurs envisagent offrir au problème du fossé explicatif. Le fait de déployer, dans un passage en revue synoptique ces différentes formes du matérialisme, nous permettra de mieux envelopper du regard le contexte qui a vu naître et se développer l'argument de la connaissance de Jackson, ainsi que les nombreuses critiques dont il a été l'objet.

4.4 Comment franchir le fossé explicatif ?

Nous avons vu que l'affaire du fossé explicatif oblige les matérialistes à attaquer de front le problème de l'expérience phénoménale. Selon une taxinomie défendue par Chalmers[56], les conceptions les plus influentes concernant la question métaphysique de la place de la conscience dans la nature peuvent être classées en six catégories, allant de « type-A » jusqu'au « type-F ». Les trois premières concernent les attitudes réductionnistes (visant à réduire le mental au physique), tandis que les trois dernières proposent des descriptions plus souples, et non-réductives. Pour dire les choses autrement, les positions qui vont de A à C misent sur l'injonction de redéfinir le terme « conscience » afin qu'il puisse être honorablement « casé » au sein d'une ontologie exclusivement matérialiste, tandis que pour les doctrines allant de D à F, l'urgence reste celle de recadrer, voire d'élargir la signification du terme « physique » afin d'y accueillir des entités non-physiques, telles que les croyances, les désirs, ou, plus près de notre intérêt ici, l'expérience phénoménale.

[56] (D. Chalmers, 2003a)

Pour parvenir à une clarification adéquate de la considérable portée de l'argument de la connaissance dans la philosophie de l'esprit contemporaine, on commencera par synthétiser les principales formes d'expression qu'a provoquées l'affaire du fossé explicatif au sein du matérialisme. Sans vouloir entrer ici dans les sinuosités d'un débat qui s'allongerait facilement, nous allons néanmoins essayer de tracer en quelques points, le schéma synoptique des principales pistes de réflexion qui ont été élaborées à partir de la date de parution du très influent article de Levine[57]. Notre propos dans cette section étant essentiellement celui de fixer la toile de fond sur laquelle se déroulera l'affrontement entre le physicalisme et les défenseurs de l'argument de la connaissance, la présentation des diverses types de matérialisme que nous nous apprêtons d'afficher ici serait nécessairement brève. Des limitations d'espace nous obligent donc à découper un périmètre de recherche restreint autour des objectifs de notre enquête.

4.4.1 Matérialisme de type-A ou le réductionnisme radical

Selon la taxinomie défendue par Chalmers, les adeptes du matérialisme de type-A[58]sont les auteurs qui maintiennent qu'il n'existe aucun fossé épistémique entre le mental et le physique, ou si fossé il y en a, il peut être facilement franchi. Impossible, disent-ils, de concevoir des créatures étant nos parfaites sosies physiques, mais qui seraient en même temps dépourvues de conscience. En bref, le matérialisme de type-A nie l'existence du côté « difficile » dans l'explication de la conscience, et prend la forme soit d'un éliminativisme radical, soit d'un behaviorisme logique (ou fonctionnalisme analytique) affirmant que le concept de « conscience » peut être défini en termes exclusivement fonctionnels ou behavioristes.

Selon Chalmers, la différence entre l'éliminativisme et le behaviorisme logique est d'ordre terminologique seulement, les deux positions optant pour une définition de la conscience en termes de fonctions d'accès et de contrôle. Beaucoup de philosophes qui n'adhèrent pas ouvertement à l'éliminativisme, par exemple, peuvent être néanmoins rangés dans la catégorie « matérialiste de type-A », à partir du moment où ils affirment que tous les phénomènes qui caractérisent ou accompagnent le phénomène de la conscience peuvent être expliqués en termes de fonctions extrêmement

[57] (J. Levine, 1993)

[58] Selon Chalmers, les auteurs qu'on peut ranger dans la catégorie « matérialistes de type-A » sont : G. Ryle (1949), D. Lewis (1988), D. Dennett (1991), F. Dretske (1995), G. Rey (1995), G. Harman (1990).

complexes, impliquant des facultés hautement sophistiquées conjuguant autocontrôle, éveil, introspection, rapport de données, mémoire, etc., et interagissant en permanence avec l'environnement. En bref, *selon le matérialisme de type-A expliquer les fonctions suffit pour expliquer tout le reste*, y compris le phénomène de la conscience.

Ce qui frappe tout d'abord dans cette variante, observe Chalmers, c'est le fait qu'elle semble nier une évidence du sens commun. D'une part, personne ne songe à contester sérieusement l'idée selon laquelle les multiples et sophistiquées *fonctions* qui caractérisent l'organisation de l'expérience consciente constituent autant de pertinents *explananda*, c'est-à-dire des phénomènes dignes d'une explication. D'un autre côté, le phénomène de la conscience en lui-même, l'expérience qualitative phénoménale tout court, *est* un *explanandum in extenso*. Fermer les yeux sur celui-ci reviendrait tout simplement à amputer le mental d'une partie essentielle de sa vie. Une attitude aussi contre-intuitive doit être appuyée et étayée par des arguments puissants. Quels sont les arguments convaincants en faveur de la thèse selon laquelle *expliquer les fonctions explique tout le reste*, y compris le phénomène de la conscience ?

La stratégie favorite des matérialistes de type-A est de recourir à des analogies empruntées aux sciences naturelles, où il n'est pas inhabituel de considérer qu'expliquer un ensemble de fonctions finit tôt ou tard par englober dans sa toile explicative, un phénomène adjacent ou récalcitrant. A cela on peut objecter que dans ce cas, l'explication de l'ensemble des fonctions *est* le seul phénomène à expliquer, tandis que le cas du phénomène de la conscience est radicalement différent, car il est accompagné d'un phénomène distinct – l'aspect qualitatif, phénoménal – qui nécessite à lui tout seul, une explication à part. Par conséquent, l'analogie doit rendre compte de l'existence de cette asymétrie radicale. Les matérialistes peuvent répondre alors en avançant la stratégie suivante : il est possible de modifier l'analogie en rappelant que dans le passé, la pensée scientifique a été confrontée à beaucoup de ces « fossés explicatifs» qui se sont avérés par la suite parfaitement franchissables et explicables en termes physiques. D. Dennett[59] par exemple, rappelle qu'au XIXe siècle, un adepte du vitalisme aurait pu soutenir, de façon analogue, qu'il existe un problème « difficile» de l'explication de la vie en termes d'évolution darwinienne *en plus* de la simple explication des fonctions biologiques qui caractérisent les structures et les comportements des organismes vivants.

[59] D. Dennett, (1996), 'Facing Backward on the Problem of Consciousness', *Journal of Consciousness Studies*, 3: 4–6.

Une analyse attentive de ces analogies nous permet toutefois de saisir en quoi leur application est inefficace pour le cas qui nous intéresse ici. Prenons l'exemple choisi par Dennett. Les vitalistes acceptaient habituellement, de manière plus ou moins explicite, l'idée que les fonctions biologiques visées *épuisent* le domaine de l'*explanandum*. Ils espéraient seulement que ces fonctions ne soient pas explicables en termes purement physiques. Mais leur situation est radicalement différente de celle que nous affrontons aujourd'hui, car les fonctions typiquement associées à l'organisation de la conscience *n'épuisent pas* l'*explanandum*, et c'est précisément là que réside le cœur du problème, à savoir dans l'existence de ce quelque chose *en plus*, qui reste à élucider. Pour voir cela, il suffit d'évoquer le cas de Broad, un vitaliste qui adhérait explicitement à l'idée selon laquelle les fonctions impliquées dans l'organisation des organismes vivants doivent recevoir une explication non-mécaniste. Mais en même temps, il soutenait que dans l'explication des fonctions de la vie, à la différence des phénomènes de la conscience, les seules preuves dont nous disposons, sont comportementales. A l'époque, très peu de vitalistes auraient pris le risque de maintenir qu'il y aurait quelque chose *en plus* à expliquer ici, *excepté* les fonctions biologiques elles-mêmes. Par conséquent, l'analogie invoquée par Dennett ne tient pas, et les matérialistes de type-A doivent affronter *nolens volens* le problème que pose l'existence de cet *explanandum* supplémentaire et récalcitrant.

Une autre stratégie de réponse adoptée par les défenseurs du matérialisme de type-A serait d'emprunter un chemin biaisé, en commençant par signaler les conséquences fâcheuses qu'entraînerait le rejet de matérialisme, et qui pourraient mener au dualisme et à tous les problèmes embarrassants que ce dernier entraîne avec lui. Toujours est-il que ce type de raisonnement en biais n'exonère pas les matérialistes de l'obligation de fournir une explication *positive* au problème que pose l'existence de cet *explanandum* supplémentaire qu'est l'expérience phénoménale. Curieusement, note Chalmers, ce genre d'arguments positifs sont difficilement trouvables, et beaucoup d'entre eux sont circulaires[60]. Une possible solution serait d'introduire le concept d'« hétérophénoménologie » (rapport verbal) de Dennett[61] pour caractériser les données primitives sur lesquelles repose un système explicatif, et qui, à la différence des

[60] Par exemple, G. Rey soutient que postuler l'existence des *qualia* est inutile, car on n'a pas besoin des *qualia* pour expliquer le comportement. A cela on peut répondre en insistant sur le fait que les *qualia sont* un *explanandum* à part entière. Voir Rey, G. (1995), 'Toward a Projectivist Account of Conscious Experience', in T. Metzinger (ed.), *Conscious Experience*, Paderborn: Ferdinand Schöningh.
[61] (D. Dennett, 1991)

dispositions et des réactions comportementales, ne semblent pas avoir besoin d'être expliquées à son tour.

Une autre possible façon de s'y prendre, note Chalmers, serait d'argumenter qu'il existe un intermédiaire *X* tel que (i) expliquer les fonctions suffit pour expliquer *X*, et (ii) expliquer *X* suffit pour expliquer le phénomène de la conscience. Un possible candidat pour jouer le rôle médiateur de *X* serait la notion de *représentation*. Souvent, on identifie les états conscients aux états représentationnels désignant des choses dans le monde extérieur. Si l'on arrive à expliquer les états représentationnels en termes fonctionnels, nous obtiendrons ainsi une explication de la conscience respectant les exigences fonctionnalistes. Selon Chalmers, cet argument repose sur une ambigüité relative à la notion de « représentation ». Il faut distinguer entre : (i) le concept de *représentation fonctionnelle* selon lequel P est « représenté » lorsqu'un système répond à P et/ou produit du comportement approprié pour P. Dans cette acception, expliquer la fonction revient à expliquer la représentation, mais expliquer la représentation ne revient pas à expliquer le phénomène de la conscience.

Il existe cependant un deuxième sens du terme « représentation », à savoir (ii) la représentation phénoménale selon laquelle P est considéré comme « représenté » lorsqu'un système a une expérience consciente de P. Dans ce sens, expliquer la représentation peut expliquer la conscience, mais expliquer les fonctions n'explique pas la représentation. Et on revient au point de départ.

Enfin, une autre stratégie adoptée par les matérialistes de type-A, probablement la plus intéressante, consiste à soutenir qu'il serait possible de fournir une explication physique des nos *croyances* relatives à la conscience, comme par exemple, la croyance que nous sommes conscients, ou la croyance que la conscience est un *explanandum* supplémentaire, ou la croyance que la conscience est non-physique. Expliquer la nature de ces croyances revient à expliquer, de manière assez honorable, le phénomène de la conscience. Une possible manière d'objecter à cela consiste à signaler que cette stratégie fonctionne uniquement si les croyances en elles-mêmes sont analysables en termes de fonctions. Or, comme l'a relevé Chalmers[62], nous avons de bonnes raison de rejeter cette hypothèse.

Récapitulons. Les discussions opposant les partisans du matérialisme de type-A à leurs adversaires, tournent autour d'une intuition centrale, à savoir l'intuition selon

[62] Voir D. Chalmers, (2002), 'Content and Epistemology of Phenomenal Belief', in *Q. Smith* et A. Jokic (eds.), *Consciousness: New Philosophical Essays*, Oxford University Press.

laquelle expliquer le phénomène de la conscience (considéré sous un angle non-fonctionnel), présuppose quelque chose *en plus* de la simple explication des mécanismes et fonctions qui la sous-tendent. La force de cette intuition est tellement importante que les matérialistes de type-A doivent s'armer d'arguments extrêmement convaincants afin d'arriver à la démolir. C'est probablement pour cette raison même que les adeptes du matérialisme de type-A représentent aujourd'hui une minorité dans la communauté philosophique.

4.4.2 Matérialisme de type-B

Les adeptes du matérialisme de type-B[63] reconnaissent l'existence d'un gouffre séparant le domaine du physique du domaine de l'expérience phénoménale, mais ils affirment que la nature de ce fossé est épistémique, non pas ontologique. Cela revient à dire, entre autres, que l'existence des zombies, bien qu'épistémiquement concevable, reste métaphysiquement impossible. Selon les partisans de ce type de matérialisme, l'identité états phénoménaux/états physiques est analogue à celle qui assure l'identité eau/H_2O, autrement dit il s'agit d'une identité établie empiriquement, et non pas *a priori*, *via* l'analyse conceptuelle. En s'appuyant sur la célèbre distinction frégéenne entre sens et référence, les matérialistes de type-B soutiennent que bien que le concept « eau » ait un sens différent du concept « H_2O », les deux font référence à une seule et même entité physique. En déplaçant l'analogie sur le terrain de l'expérience phénoménale, les matérialistes de type-B affirment que bien que le concept de « conscience » ait un sens différent de tel ou tel concept fonctionnel afférent, il se peut, dans un avenir plus ou moins proche, que l'on découvre empiriquement qu'ils réfèrent à une seule et même entité physique. Cette distinction conceptuelle expliquerait l'existence du fossé épistémique séparant le domaine du physique de celui de l'expérience phénoménale, tout en niant le fossé ontologique.

Malgré l'attrait évident que présente cette alternative, elle soulève néanmoins, observe Chalmers, quelques difficultés notables. Premièrement, les matérialistes de type-B ne semblent pas prendre en compte le fait que le fossé épistémique qui concerne le problème de la conscience n'est pas de même nature que les fossés épistémiques qui

[63] Selon Chalmers, les auteurs que l'on peut ranger dans la catégorie « matérialistes de type-B sont : Levine (1983), Loar (1990/1997), Papineau (1993), Tye (1995), Lycan (1996), Block et Stalnaker (1999), Perry (2001).

surgissent dans d'autres domaines. Par exemple, dans le cas paradigmatique de l'identité eau/H_2O, expliquer la signification de l'« eau » revient à expliquer pourquoi une substance possède une certaine structure ou comportement. Admettons qu'un super-scientifique S serait en possession d'une description complète du monde physique ; il s'ensuit dès lors que si S arrive à identifier les structures et les fonctions du système, il serait *ipso facto* dans la position d'en déduire *toutes* les vérités concernant la nature de l'eau. Dans ce cas, la distinction épistémique entre l'ensemble complet de vérités physiques concernant le réel en général, et les vérités relatives à la nature de l'eau en particulier, n'est pas comparable avec le fossé que nous devons affronter dans le cas « difficile » de la conscience.

Le point névralgique du matérialisme de type-B réside dans l'impératif de classer l'identification entre les phénomènes conscients et les phénomènes fonctionnels au rang d'identités épistémiquement primitives, dans le sens où l'identité en cause *ne* serait *pas* déductible de la totalité des vérités physiques concernant le réel. Si tel était le cas, autrement dit, si cette identité serait déductible de l'ensemble complet des vérités physiques, alors le matérialisme de type-B deviendrait tout simplement un matérialisme de type-A. Par conséquent, l'identité entre les états physiques et la conscience doit être traitée comme une sorte de principe primitif ou d'axiome brut, afin d'éviter que le matérialisme de type-B ne devienne pas de type-A. Toujours est-il que cette condition de primitivité peut se heurter à plusieurs difficultés : d'abord, on pourra objecter que le seul fondement légitime pour situer ce type de primitivité est celui des lois de la physique, les seules à ne pas être déduites des lois encore plus « primitives ». Autrement dit, les défenseurs du matérialisme de type-B reposent sur un principe dont le statut épistémique est celui d'une loi fondamentale, mais qui finit par légiférer sur le statut *ontologique* d'une autre identité, en occurrence celle entre états phénoménaux et états physiques. Or, note Chalmers, cette tactique est comparable à celle d'un voleur qui changerait d'habit selon les besoins de ses missions-braquages.

Scrutons de plus près le « micmac» auquel fait allusion Chalmers: les énoncés d'identité sont basés sur des explications préliminaires, tandis que les principes primitifs sont reconnus comme lois fondamentales. Naturellement, on pourrait s'y attendre à la même application dans ce cas-ci : si l'on admet que la relation épistémiquement primitive entre états physiques et états phénoménaux est une loi fondamentale de la nature, il s'ensuit que la conscience se révélera distincte de toute autre propriété physique, car les lois fondamentales relient toujours des propriétés

distinctes, *sui generis*. Ce qui nous ramène à une forme de matérialisme non-réductionniste. En revanche, le matérialiste de type-B repose sur la prémisse d'une relation *primitive* entre le mental et le phénoménal, inexplicable en termes plus ultimes, et suggère que cette relation est à la base une *identité*. Mais, insiste Chalmers, à défaut d'en fournir des éclaircissements et des précisions plus approfondies sur la nature de cette identité, celle-ci risque de passer pour une solution *ad hoc*, dont la nature mystérieuse sert de paravent dissimulant des incohérences.

A cela, les partisans du matérialisme de type-B peuvent répondre, suivant D. Papineau[64], que les identités n'ont pas besoin d'être expliquées, et dans un certain sens, elles sont toujours primitives. Mais nous avons vu que dans d'autres disciplines ou domaines, les identités peuvent au moins être déduites à partir des vérités encore plus primitives, et que donc elles ne sont pas 'primitives' au sens strict du terme. Certains auteurs évoquent l'idée que même des vérités concernant les identités indiscutables de type eau/H_2O ne peuvent pas être réduites à des vérités physiques qui les sous-tendent. Une discussion détaillée de ces querelles, bien qu'incontestablement fascinante, déborderait du cadre de cet essai. Rappelons seulement, en suivant Chalmers et Jackson[65] que les arguments épistémiques que nous avons esquissés au début de cette section s'appuient sur le contraste qui particularise la situation du problème « difficile » de la conscience par analogie avec les autres cas de « fossés ».

Une autre possible réponse serait de suivre B. Loar[66] et d'admettre que les identités impliquant les phénomènes de conscience diffèrent des autres types d'identités de par leur nature épistémiquement primitive, tout en essayant d'expliquer cette singularité en faisant appel à des particularités du concept même de « conscience ». D'autres auteurs préfèrent arguer qu'en fin du compte, une identité états physiques/états phénoménaux n'est pas une hypothèse dont nous avons nécessairement besoin pour bâtir une théorie matérialiste de l'esprit. Tout ce dont nous avons besoin, disent-ils, c'est d'admettre que les états physiques impliquent les états phénoménaux dans le sens où il est métaphysiquement impossible d'avoir certains états physiques sans avoir en même temps les états phénoménaux afférents. En d'autres mots, il s'agit de stipuler que

[64] D. Papineau, (1993), 'Physicalism, Consciousness, and the Antipathetic Fallacy', *Australasian Journal of Philosophy* 71: 169–83

[65] Voir D. Chalmers & F. Jackson, (2001), 'Conceptual analysis and reductive explanation', *Philosophical Review*, 110/3: 315-60.

[66] B. Loar, (1990/1997), 'Phenomenal States', *Philosophical Perspectives*, 4: 81–108. Réédité in Ludlow et al. (2004), op. cit. (Nous utilisons ici cette dernière édition)

l'implication P⊃*Q* doit être nécessaire, où P représente l'ensemble complet ou la totalité des vérités physiques et *Q* est une vérité phénoménale arbitraire.

A partir de là, il est très tentant de faire appel, à nouveau !, aux célèbres travaux de Kripke sur le nécessaire *a posteriori*. Rappelons-nous que selon Kripke[67], une identité de type eau/H_2O est nécessairement vraie, sans être pour autant connaissable *a priori*. De la même manière, affirment les partisans du matérialisme de type-B, une identité de type eau/H_2O est nécessaire *a posteriori*. Dans ce cas, on devrait se heurter à un fossé épistémique seulement, et non pas à un fossé ontologique. A cela on peut répondre de plusieurs façons : d'abord, on peut insister à nouveau sur le fait que dans les autres domaines, les nécessitations ne sont pas épistémiquement primitives. Même si la relation H_2O/eau s'avère nécessaire *a posteriori*, elle reste néanmoins *déductible* d'une description complète et exhaustive des vérités physiques. Le même constat s'applique au nécessaire *a posteriori* kripkéen. En revanche, les matérialistes de type-B sont contraints de stipuler que la relation états physiques/états phénoménaux doit être épistémiquement primitive en soi, i.e. *non-déductible* d'un ensemble complet de vérités physiques. Par conséquent, le caractère mystérieux et *ad hoc* de cet axiome de « jointure » ne semble pas indiquer la sortie de l'impasse. Car stipuler que de tels « ponts » et principes explicatifs primitifs faisant le liant entre les rives du gouffre *sont* des principes fondamentaux, revient à sacrifier considérablement l'esprit et la lettre du matérialisme.

Nous voyons donc, pour résumer, que le matérialiste de type-B est contraint d'adhérer à ce que Chalmers[68] appelle une « *nécessitation métaphysique forte* », et de maintenir que le cas de la conscience est spécial, dans le sens où la relation de nécessité *a posteriori* qui raccorde les états physiques aux états phénoménaux est beaucoup plus puissante que dans les autres domaines où elle apparaît.

Selon Chalmers, une des réponses les plus intéressantes des matérialistes de type-B consistent à admettre que cette « force » est une singularité qui s'applique au cas de la conscience, tout en essayant de l'élucider par le truchement d'une analyse détaillée des particularités de nos systèmes conceptuels. Selon B. Loar[69], pour ne donner ici qu'un bref exemple, les concepts phénoménaux sont des concepts « discriminatifs » dépourvus de modes de présentation contingente. D'autre part, il est généralement

[67] (S. Kripke, 1980)
[68] L'expression originale en anglais est « strong metaphysical necessity » (D. Chalmers, 1996)
[69] (B. Loar, 1990/1997)

admis, depuis les conclusions révolutionnaires de Kripke[70] et plus tard de Putnam[71] avec sa célèbre expérience de pensée de la Terre Jumelle, que le contenu interne ne peut pas déterminer, à lui tout seul, l'extension. Mais la question de savoir si ce contenu joue ou pas un rôle constitutif et fondamental dans le processus de détermination de l'extension reste largement ouverte[72].

A nouveau, dresser ici un compte rendu minutieux des arguments pro et contre les thèses du matérialisme de type-B[73] dépasserait les modestes cadres de nos objectifs dans ce chapitre. Tournons-nous à présent sur la dernière version du matérialisme de type réductionniste.

4.4.3 Matérialisme de type-C

Le matérialisme de type-C adhère également à la thèse du fossé épistémique, mais soutient que le gouffre est, du moins en principe, franchissable. Ce qui revient à dire que même si le problème « difficile » de la conscience ne peut pas être résolu pour le moment, il reste néanmoins solvable *en principe*. Comme le note Chalmers, cette doctrine semble à première vue très attractive, car elle permet à la fois de concilier l'existence manifeste d'un fossé explicatif, tout en suggérant que sa présence est imputable à nos propres limites cognitives. Plusieurs versions de cette conception ont vu le jour. Par exemple, Nagel (1974) suggère que notre incompréhension de la manière dont les états phénoménaux *sont* des états physiques est analogue à celle des philosophes présocratiques qui étaient dans l'impossibilité de comprendre que la matière *est* de l'énergie. De son côté, P. Churchland[74] affirme également que ces limitations dues à notre incompréhension seront dépassées par le développement des

[70] (S. Kripke, 1980)
[71] (H. Putnam, 1975)
[72] Selon J. Levine, les concepts phénoménaux se caractérisent justement par un mode de présentation « non-attributive »: "According to the non-ascriptivist, it is not part of the meaning of, say, "cat" that cats are animals. Of course it may be metaphysically necessary that cats are animals, but the crucial point is that it's not *a priori*. Mere competence with the terms "cat" does not yield such knowledge". (J. Levine, 2001:53)
[73] Pour une discussion critique détaillée de ces querelles voir D. Papineau (2007), 'Phenomenal and perceptual concepts," in T. Alter & S. Walter (eds.), op. cit., p. 136-142.
[74] Voir P. Churchland, (1989), 'Knowing *Qualia*. A Reply to Jackson', in *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA: MIT. Réédité avec un Postscript in Ludlow, et al., (2004). Nous utilisons ici cette dernière édition. Voir aussi P.S. Churchland, (1996), 'The hornswoggle problem', *Journal of Consciousness Studies*, 3: 402-8.

théories scientifiques futures, tandis que Van Gulick[75] suggère que l'argument de la concevabilité, un des piliers du combat antimatérialiste, repose sur un vice de circularité, car une explication adéquate de la conscience rendrait l'existence de zombies tout bonnement inconcevable dès le départ. Une autre façon de considérer le problème serait de dire que les zombies sont à première vue concevables du point de vue des êtres humains dotés des capacités cognitives particulières, mais qu'ils ne sont pas concevables si l'on adopte le point de vue d'une *rationalité idéale*. Ou de dire que les vérités phénoménales sont déductibles en principe des vérités physiques, mais que la déductibilité en question serait proche de celle que l'on retrouve dans le cas de certaines vérités mathématiques complexes. Autrement dit, elles sont déductibles en principe, mais inaccessibles à nous *pour l'instant*, car il nous manque, pour ainsi dire, quelques éléments clé du puzzle afin de saisir, dans un panorama révélateur, le tableau d'ensemble.

Malgré ses attraits indéniables, le matérialisme de type-C, note Chalmers, finit tôt ou tard par collapser dans les autres deux types du matérialisme, et par conséquent se heurter aux mêmes difficultés que ses confrères. Une manière d'exprimer l'idée que le fossé épistémique peut être surmonté dans l'absolu, serait de dire que *dans l'absolu*, expliquer les fonctions explique par là même tout le reste, et que par conséquent, il n'y a plus d'*explanandum* à élucider. Il est tout à fait possible, bien entendu, de concevoir qu'à l'heure actuelle les philosophes sont les victimes inévitables d'une certaine confusion conceptuelle qui provoque une faille dans le tissu argumentatif déjà mis en place, et qui les pousse à croire (peut-être à tort) qu'il resterait un *explanandum* résiduel à élucider. Or, selon Chalmers, cette façon de voir les choses n'est qu'une variation de la position de type-A, dont la trame centrale tourne autour des possibles méthodes d'éliminer la pomme de la discorde initiale : l'*explanandum* récalcitrant, les *qualia*.

Réduite à sa plus simple expression, la stratégie utilisée peut être synthétisée ainsi : (1) dans un premier temps on part de la prémisse selon laquelle la description du monde en termes physiques est basée sur des descriptions en termes de structures et de dynamiques. (2) Ensuite, on stipule qu'à partir des prémisses et *vérités structurales-dynamiques*, seules peuvent être inférées des conclusions et des vérités structurales-dynamiques. Enfin, (3) on déduit que les vérités relatives au phénomène de la conscience ne sont pas des vérités structurales et dynamiques.

[75] (R. Van Gulick, 1993)

Reprenons, en suivant Chalmers, ces trois étapes, et essayons de les expliciter pas par pas.

4.4.3.1 Pas n°1 : la description microphysique du monde

Du point de vue de la physique fondamentale, la description du monde se fait en termes de distribution de particules, champs, et ondes dans l'espace-temps. Ces systèmes de base sont caractérisés à leur tour par des propriétés spatio-temporelles, et des propriétés distinctives comme la masse, la charge, et les fonctions d'onde. Ces dernières sont à leur tour définies en termes d'états d'une structure abstraite (exemple : les espaces de Hilbert), et ont une influence causale sur les autres états. On peut ranger ce type de descriptions dans la rubrique des descriptions *structurales*. L'état de ces systèmes peut varier dans le temps en fonction de la dynamique des principes impliquant les propriétés concernées. Le résultat obtenu est une description du monde en termes de structures espace-temps et de structures formelles sous-jacentes, à laquelle s'ajoute la carte interactive de leur dynamique consignant les différentes étapes de leur évolution. Or, étant donné que nous ne sommes pas encore en possession d'une physique complète, et que, par conséquent, nous ignorons les faits, états, processus que cette physique pourrait un jour décrypter, il s'ensuit que le fossé explicatif n'est en quelque sorte qu'un dommage collatéral de ce manque provisoire d'informations.

Nous touchons ici à l'une des difficultés majeures de la position matérialiste, que nous avons déjà identifié sous le nom de « dilemme d'Hempel ». Selon Chalmers, l'idée d'une future physique complète c'est une hypothèse dont on n'a pas vraiment besoin pour réfuter le matérialisme. Tout ce dont nous avons besoin c'est l'idée que les descriptions physiques se réalisent exclusivement en termes de structures et dynamiques. Il est vrai qu'à tout moment, des théories physiques plus puissantes (comme la théorie de la relativité ou la mécanique quantique) peuvent introduire des structures inédites, ainsi que des nouvelles dynamiques sur ces structures, mais le fossé qui les sépare du phénomène de la conscience reste intact.

A la suite de ces considérations, un matérialiste de type-C pourrait répondre en suggérant que les futures théories physiques pourraient tout simplement aller au-delà des descriptions classiques en termes de structures et de dynamiques. Mais étant donné la nature de l'explication physique, il est difficile d'imaginer quelle serait la nature de cette future mystérieuse et quasi-mystique nouvelle physique. L'ajout d'une propriété

physique nouvelle a pour objectif habituel de remédier ou améliorer l'explication d'un phénomène physique récalcitrant. Mais l'histoire des sciences nous a appris que lorsque cela arrive, une explication en termes de structures et de dynamiques est largement suffisante.

Une autre option possible serait de postuler non pas l'introduction ou l'ajout des nouvelles propriétés, mais d'intégrer le phénomène de la conscience *depuis l'intérieur* des théories physiques, à la manière dont certaines théories quantiques[76] le font actuellement. C'est une possibilité, admet Chalmers[77], mais il s'agit d'une option *non-réductive*, et par conséquent, elle doit être classée dans ce qu'il appellera la conception de type-D ou F. Il existe cependant un sens où l'expression « description physique complète » semble très prometteuse, souligne Chalmers. En effet, certains auteurs pensent que même si la physique actuelle caractérise ses propriétés fondamentales (masse, charge, etc.) en termes de structures abstraites et de relations, elle laisse néanmoins ouverte la question de leur nature intrinsèque. Vue de cette perspective, une théorie physique complète doit rendre compte également des propriétés intrinsèques qui fondent, depuis l'intérieur, ces structures et dynamiques. Avec l'invocation des « propriétés intrinsèques », la physique se réserve le droit d'aller au-delà des simples structures et dynamiques, de telle sorte que des faits concernant la conscience en découleront nécessairement. Même si pour l'instant, la nature de ces propriétés intrinsèques nous est cachée, elle demeure connaissable *en principe*. Selon Chalmers, cette option semble très plausible, mais, à nouveau, il s'agit d'une position non-réductive, qu'il baptisera le monisme de type-F, et qui sera discutée plus loin, dans la section 4.5.3. Il s'agit, par ailleurs, d'une solution que l'auteur de ces lignes est enclin à adopter, et qui sera défendue dans le dernier chapitre de cet essai.

[76] Voir pour une discussion détaillée M. Velmans, (2002), 'How could conscious experiences affect brains?' *Journal of Consciousness Studies 9*(11), 3-29. D. Bohm, (1990), 'A new theory of the relationship of mind and matter', *Philosophical Psychology 3*, 271-286. B. d'Espagnat, (1999), 'Concepts of reality', in *On Quanta, Mind, and Matter*, H. Atmanspacher, U. Müller-Herold, & A. Amann (eds.), Kluwer, Dordrecht, pp. 249-270. H.P. Stapp, (1993), 'A quantum theory of the mind-brain interface', in *Mind, Matter, and Quantum Mechanics*, Springer, Berlin, pp. 145-172. S.R. Hameroff, (1994), 'Quantum coherence in microtubules: a neural basis for emergent consciousness?' *Journal of Consciousness Studies* 1, 91–118. Voir aussi P. Van Loocke (ed.), (2001), *The Physical Nature of Consciousness*, Amsterdam: John Benjamins Publishing.

[77] (D. Chalmers, 2003a : 121)

4.4.3.2 Pas n°2 : De la microphysique à la physique du visible

Que peut-on inférer en partant de ce type de description en termes de structures et dynamiques ? C'est un fait bien connu qu'une description microphysique de niveau élémentaire (« lower-level ») peut rendre compte des effets surprenants dans le monde des propriétés macroscopiques (« higher level »). Mais dans tous ces cas, les propriétés macroscopiques qui en découlent sont à leur tour descriptibles en termes de structures et dynamiques. Par conséquent, cette hypothèse ne fait que renforcer la thèse (2), selon laquelle à partir des prémisses et vérités de nature structural-dynamique, seules peuvent être déduites des conclusions et vérités structurales-dynamiques. Un matérialiste de type-C pourrait se replier derrière la thèse selon laquelle il existe certains faits et vérités, comme par exemple, ceux concernant les *représentations* ou les *croyances* qui, bien qu'ils ne soient pas intrinsèquement structurels ou dynamiques, ils entrent néanmoins dans la composition des descriptions structural-dynamiques. Mais comme nous l'avons mentionné plus haut, Chalmers observe que dans tous les cas où ces vérités découlent à partir des descriptions structurales-dynamiques, elles reposent implicitement sur l'utilisation d'un sens fonctionnel des concepts de représentation et de croyance. Si des énoncés impliquant ces derniers s'avèrent vrais *en vertu du* système des descriptions structural-dynamiques, alors les notions impliquées doivent être structural-dynamiques elles-mêmes, du moins jusqu'à un certain niveau.

Selon Chalmers, une possible tentative de sortir de l'impasse serait recourir à nouveau au stratagème des étapes intermédiaires : stipulons l'existence d'une notion intermédiaire *X*, en sorte que les vérités concernant *X* soient valides en raison des descriptions structural-dynamiques, tandis que les vérités concernant la conscience soient valides grâce à *X*. Mais cela nous renvoie au cas du matérialisme de type-A : soit *X* est analysable du point de vue fonctionnel (au sens large), et dans ce cas le deuxième pas échoue, soit *X* n'est pas analysable en termes fonctionnels, et dans ce cas c'est le premier pas qui échoue. Prenons le cas emblématique de la notion de « représentation »: en ce qui concerne la notion de « représentation fonctionnelle » c'est le premier pas qui échoue, alors que pour la notion de « représentation phénoménale » c'est le deuxième.

4.4.3.3 Pas n°3 : le mystère de la conscience

La troisième étape s'agence autour de l'interrogation suivante : les explications en termes de systèmes structural-dynamiques sont-elles suffisantes, *per se*, pour expliquer le phénomène de la conscience ? Selon Chalmers, ainsi que pour les défenseurs de l'argument de la connaissance, la réponse est « non ». Nous examinerons plus en détails les raisons de cette prise de position dans le chapitre suivant. Pour le but descriptif du présent paragraphe, il suffit de rappeler, avec Chalmers, que les interprétations en termes structural-dynamiques se limitent à expliciter seulement la partie facile et mesurable de l'iceberg des problèmes qui concernent la vie mentale. Quant au problème difficile de la conscience, il reste à nouveau intact.

Selon Chalmers, la principale faille de la stratégie matérialiste de type-C réside dans le fait que l'implication épistémique qui va de la prémisse P[78] vers la conclusion *Q* présuppose d'intercaler une sorte de « crochet conceptuel » à l'aide duquel les conditions décrites en P peuvent satisfaire les contraintes conceptuelles pour la vérité de *Q*. Or, dans le cas de la conscience, cette attache conceptuelle est difficile à trouver.

Pour résumer, disons que le matérialisme de type-C s'évertue à trouver une place à soi dans le paysage des modèles explicatifs du mental. Mais une fois que l'on adhère à l'idée selon laquelle le concept de conscience n'est pas à la base un concept fonctionnel, et que les descriptions physiques du monde sont de nature structural-dynamiques, on s'aperçoit qu'il ne reste plus de place pour que le phénomène de la conscience soit conceptuellement déduit à partir des descriptions physiques du réel. Par conséquent, on se retrouve devant le choix suivant :

(i) soit on revient au matérialisme de type-A en soutenant que la conscience est un concept fonctionnel malgré tout ;

(ii) soit on admet que la description du monde physique ne se résume pas uniquement à des termes structural-dynamiques et qu'elle présuppose donc quelque chose *en plus* (ce qui nous renvoie au dualisme de type-D ou au monisme de type-F) ;

(iii) soit on affirme que la vérité du matérialisme ne requiert pas la validité de la déductibilité des vérités relatives à la conscience à partir des vérités physiques, et l'on revient ainsi au matérialisme de type-B.

---

[78] Pour mémoire : P représente l'ensemble complet ou la totalité des vérités physiques et *Q* est une vérité phénoménale arbitraire. Cf. infra.

4.5 Variantes non-réductionnistes

Selon Chalmers, il n'y a que deux options tenables pour un matérialiste[79], à savoir celles classées sous l'étiquette « type-A » et « type-B ». Les autres, soit s'avèrent infondées, soit collapsent et se dissolvent dans une de ces deux variantes. Or, le prix à payer si l'on adopte une de ces deux alternatives matérialistes « authentiques » – nier l'existence manifeste de l'*explanandum* dans le cas du matérialisme de type-A, ou embrasser la thèse des identités primitives ou des nécessités métaphysiques fortes dans le cas du matérialisme de type-B – nous incite à sérieusement reconsidérer leur bien-fondé, et à chercher des pistes de réflexion hors des sentiers A et B.

A quoi ressembleraient-elles ces *autres* alternatives ? Si le phénomène de la conscience n'est pas nécessairement déductible des faits et des vérités physiques, alors quelque chose d'ontologiquement vierge doit entrer en scène. Mais quoi, plus exactement ? Nous avons vu que traditionnellement, on s'accorde à considérer comme fondamentales les propriétés du réel qui sont ontologiquement caractérisables en termes physiques tels que l'espace-temps, la masse, la charge, etc. Or, si les arguments antimatérialistes s'avèrent justes, alors la question qui se pose est : doit-on envisager sérieusement d'élargir le 'Jockey Club' des propriétés fondamentales ?

Deux possibilités sont à retenir ici : soit l'on admet que la conscience *est* une propriété fondamentale du réel au même titre que l'espace-temps et la masse en physique, les propriétés phénoménales devenant ainsi des propriétés primitives. Soit l'on maintient que la conscience n'est pas fondamentale en elle-même, mais qu'elle est, en quelque sorte, nécessairement déductible d'un ensemble de propriétés fondamentales *X*, qui, elles, ne découlent pas des faits et vérités physiques. Ces mystérieux attributs, que l'on peut baptiser, dit Chalmers, propriétés *protophénoménales*[80], seraient quant à elles, des traits primitifs du réel. Mais dans tous les cas, la solution au problème de la

[79] Le matérialiste peut, en dernier ressort, rejeter le bien-fondé même des taxinomies et distinctions sur lesquelles repose le tableau que nous venons d'esquisser. Certains disciples de Quine (1951) rejettent les distinctions traditionnelles entre vérités conceptuelles/vérités empiriques, entre *a priori* et *a posteriori*, ou entre contingent et nécessaire. Chalmers propose de baptiser cette position matérialisme de type-*Q*. Malgré tout, cette position affronte les mêmes difficultés que ses consœurs. Dennett, par exemple, est partisan de l'idée selon laquelle expliquer les fonctions revient (suffit) à expliquer tout le reste. Mais cela revient à ressusciter les problèmes traditionnels du matérialisme de type-A. D'autres pensent qu'il est possible de postuler l'identité entre les états physiques et les états conscients en vertu des évidents liens isomorphes qui les caractérisent et que l'on trouve dans le règne naturel. (Voir P. Churchland) Mais cela remet en avant les vieux problèmes du matérialisme de type-B. Enfin, d'autres auteurs semblent miser sur la découverte de nouvelles formes d'explication, mais dans ce cas, le matérialisme de type-C ressuscite à sont tour. (Voir D. Chalmers, 2003a: 123).

[80] *Ibid*.

conscience présuppose l'introduction d'un nouveau membre au niveau de la réalité fondamentale. Reste à savoir comment arrivera-t-on à intégrer cet invité surprise dans le tissu explicatif déjà mis en place par les théories physiques « classiques ». Bref, comment concilier les primitifs non-physiques avec les primitifs physiques déjà existants ?

L'existence des propriétés fondamentales présuppose l'existence des lois fondamentales. Par conséquent, on devrait découvrir des lois psychophysiques qui régissent les rapports entre les propriétés physiques et celles phénoménales. Ces nouvelles lois doivent être fondamentales à leur tour, c'est-à-dire irréductibles aux lois physiques, plus « primitives ». Plusieurs interrogations surgissent alors: comment aménager l'existence des lois psychophysiques avec le principe de clôture causale du monde physique ? Comment intégrer les propriétés phénoménales dans un monde physiquement et causalement clos ? Ainsi qu'on va commencer à le voir au prochain chapitre, ces interrogations nous conduisent sur le terrain glissant de la métaphysique, et nous invitent à jongler avec deux concepts hautement explosifs, à savoir celui d'*omniscience* et d'*ignorance*.

Sans vouloir anticiper sur un sujet qui sera traité plus minutieusement dans la deuxième partie de notre enquête, signalons seulement que si l'on rejette le réductionnisme rigide du matérialisme de première génération, on s'aperçoit que trois options non-réductionnistes s'ouvrent à nous :

(i) Soit on rejette le principe de clôture causale de la physique, et l'on accepte l'existence des « fossés », des discontinuités qui court-circuitent les chaînes causales du réseau microphysique, et dont la nature intrinsèque est phénoménale (c'est la thèse défendue par le dualisme de type-D).

(ii) Soit on accepte la thèse de la clôture causale de l'univers microphysique, mais l'on rejette la prémisse selon laquelle les propriétés phénoménales auraient un rôle causal effectif (c'est la position que Chalmers a baptisé le « dualisme de type-E » et qui est connue surtout sous le nom d'épiphénoménalisme).

(iii) Soit on adhère au principe de clôture causale, mais on insiste sur l'idée que les propriétés phénoménales y sont pleinement intégrées, et que par conséquent, les *qualia* ont une efficience causale réelle sur le monde physique, en vertu de leur appartenance intrinsèque au physique (c'est le monisme de type-F).

Avant d'exposer brièvement les points forts et faibles de ces trois positions non-réductionnistes, on doit prendre, à la suite de Chalmers, certaines précautions et préciser que la question de savoir laquelle de ces options est la plus proche de la vérité demeure, à l'heure où nous écrivons ces lignes, une question ouverte et âprement débattue.

4.5.1 Dualisme de type-D ou l'interactionnisme cartésien

Pour les partisans[81] du dualisme de type-D, les théories microphysiques ne sont pas causalement closes, et les propriétés phénoménales ont une influence causale effective sur le monde physique. Selon cette doctrine, connue également sous le nom d'*interactionnisme*, les états physiques peuvent causer des états phénoménaux et les états phénoménaux peuvent causer des états physiques. En conséquence, l'évolution des états microphysiques n'est pas exclusivement déterminée par l'intervention des lois physiques, mais tombe sous l'incidence directe des lois psychophysiques également. L'exemple le plus classique de ce type de dualisme reste, bien évidemment, le dualisme des substances de Descartes. Mais il est compatible également avec un dualisme plus sophistiqué, à savoir le dualisme des propriétés, selon lequel le réel est une seule et unique entité dotée de deux types de propriétés fondamentales : physiques et phénoménales.

Contre cette conception, on a parfois objecté que deux états disjoints – un mental et un physique – ne peuvent pas interagir, car il n'y a aucun lien causal pour les interconnecter. Mais à une lecture plus attentive, quelqu'un pourrait riposter, à la façon d'un Hume, et dire que cela s'applique même dans le cas plus « orthodoxe » des interactions physique/physique. Après tout, la physique newtonienne ne nous apprend rien sur le tissu causal à l'aide duquel la gravitation fonctionne. Tout ce qu'elle nous dit c'est que les lois qui régissent la gravitation sont universellement applicables.

[81] Selon Chalmers, les auteurs qu'on peut ranger dans la catégorie « dualistes de type-D » sont : K. Popper & J. Eccles (1977), *The Self and Its Brain: An Argument for Interactionism* ; *W.* Sellars, (1981), 'Is Consciousness Physical?', *The Monist*, 64: 66–90 ; R. Swinburne, (1986), *The Evolution of the Soul*. Oxford: Oxford University Press; J. Foster, (1991), *The Immaterial Self: A Defence of the Cartesian Dualist Conception of the Mind*. Oxford: Oxford University Press; H. Stapp, (1993), *Mind, Matter, and Quantum Mechanics*. Berlin: Springer-Verlag.

Une autre critique, beaucoup plus puissante cette fois-ci, contre l'interactionnisme vient de l'intérieur de la physique elle-même. Les recherches scientifiques nous apprennent que le monde physique est un domaine causalement clos, et que par conséquent, il n'y a plus de place, ontologiquement parlant, pour des entités ou des états supplémentaires, mentaux ou phénoménaux. Les interactionnistes peuvent répliquer que même si aucune étude expérimentale n'a réussi de nos jours à prouver l'existence des états non-physiques, leur existence n'a pas été expérimentalement réfutée non plus, du moins jusqu'à présent.

Une autre possible réplique, beaucoup plus influente, consiste à montrer que loin d'infirmer l'interactionnisme, la physique actuelle, et plus précisément la mécanique quantique[82], l'encourage effectivement. Le monde physique quantique est traduisible par une fonction d'onde particulière, selon laquelle les entités physiques sont souvent décrites comme occupant deux états superposés, bien que ces superpositions ne soient jamais directement observables. Les fonctions d'onde peuvent évoluer de deux manières : (i) de façon linéaire suivant la forme exprimée par l'équation de Schrödinger; (ii) ou de façon non-linéaire, constituée d'états qui collapsent depuis des états superposés vers des états non-superposés. L'évolution linéaire est déterministe, alors que les collapses ne le sont pas. L'évolution de type-Schrödinger est un continuum perpétuel, un flux ininterrompu, tandis que les « catastrophes », comme on les appelle parfois, surgissent uniquement à l'occasion des opérations de mesure.

Cette théorie des « catastrophes » laisse la porte ouverte à des interprétations interactionnistes. Les « collapses » sont censés se manifester lors des opérations de mesure. Bien qu'il n'existe, à l'heure actuelle, aucun consensus définitif concernant la définition de ce concept clé, tout le monde semble d'accord pour rattacher l'opération de mesure à l'intervention d'un sujet observateur *conscient*. Un prolongement intéressant des remarques précédentes est l'observation selon laquelle il semble difficile, voire impossible, de trouver un critère purement *physique* pour circonscrire le concept de « mesure », puisque nous savons que les systèmes physiques sont régies par la dynamique linéaire de Schrödinger. A la lumière de ces considérations, il devient naturel de suggérer que l'opération de mesure *est* précisément une observation

[82] E. Schrödinger, *L'esprit et la Matière* précédé de M. Bitbol, L'élision, Seuil, 1990. Voir aussi M. Bitbol, (2000), *Physique et philosophie de l'esprit.* Paris: Flammarion. Voir aussi J. Evans & A. S. Thorndike, (2006), *Quantum Mechanics at the Crossroads, New perspectives from History, Philosophy and Physics*, Springer. Voir aussi M.A. Bedau & P. Humphreys (eds.), (2008), *Emergence-Contemporary Readings in the Philosophy and Science*, Cambridge Mass.: MIT Press.

consciente, et que, par conséquent, c'est le phénomène de la conscience qui est à l'origine du collapse.

Cette interprétation reste néanmoins à prendre avec précaution, car la mécanique quantique ne garantit aucunement qu'il s'agisse là de la *seule* interprétation possible. D'autres options semblent valables, qui ne font pas appel au concept mystérieux de « collapse », et pour lesquelles les opérations de mesure ne jouent aucun rôle spécial ou décisif. Cela dit, il serait tout simplement faux de stipuler que la mécanique quantique *exclut* l'interactionnisme. Il est cependant étonnant, remarque alors à juste titre D. Chalmers[83], que les philosophes rejettent l'interactionnisme sur la base des théories *physiques*, alors que les physiciens rejettent l'interprétation interactionniste de la mécanique quantique sur des bases *philosophiques*. La question de la place de la mécanique quantique dans la construction d'un modèle explicatif de l'esprit constitue un sujet fascinant qui mériterait un traitement plus approfondi dans un essai à part. Retenons pour l'instant qu'il s'agit là d'un point hautement controversé.

Pour résumer, bien qu'il puisse paraître à première vue comme une option démodée, l'interactionnisme ne devrait pas être congédié trop hâtivement.

4.5.2 Dualisme de type-E ou l'épiphénoménalisme

Le dualisme de type-E, plus communément connu sous le nom d'*épiphénoménalisme*, affirme que les propriétés phénoménales sont ontologiquement distinctes des propriétés physiques, mais que les états phénoménaux n'ont aucune efficience causale sur le physique, alors que les états physiques influencent causalement le domaine du phénoménal. Dans cette acception, les lois psychophysiques fonctionnent uniquement sur le versant physique/mental, mais non pas sur le versant mental/physique.

A l'instar du dualisme-D, l'épiphénoménalisme est compatible avec les deux versions du dualisme : des substances et des propriétés. De plus, il semble respecter à la fois les arguments antimatérialistes des défenseurs de la conscience, et le principe de clôture causale si cher aux physicalistes. Le point faible du dualisme de type-E demeure sa nature profondément contre-intuitive. Si l'épiphénoménalisme est correct, alors il s'ensuit que les états mentaux n'ont aucune efficience causale sur nos actions : par

[83] (D. Chalmers, 2003a :126)

exemple, la sensation de douleur n'a aucune fonction causale dans l'action qui consiste à retirer ma main de la flamme. Mon désir de contempler la beauté de l'architecture maure n'a aucune influence dans mon action de prendre l'avion, aller à Grenade, et visiter le palais de l'Alhambra. La vue d'un rubis et l'expérience qualitative de rouge qui l'accompagne ne joue aucun rôle dans l'énonciation de la proposition « Je suis en train de voir la couleur rouge. »

Pour beaucoup de penseurs, ces conséquences sont tout simplement intenables, et représentent un prix trop cher à payer. Les épiphénoménalistes peuvent répliquer en insistant sur le fait qu'aucune évidence directe ne contredit leur position. Suivant l'exemple de Hume, ils insistent en rappelant que tout ce que nos observations révèlent, ce sont des connexions et covariations régulières entre certains états phénoménaux *e* et des actions *a* qui adviennent dans le monde réel, en sorte que certains modes d'expérience sont accompagnés par certains modes d'action. L'exposition à ce type de covariations peut nous induire en erreur, produisant en nous la croyance fausse selon laquelle l'expérience *e cause* l'action *a*.

Selon Chalmers[84], une autre objection, cette fois-ci beaucoup plus intéressante, contre l'épiphénoménalisme, concerne la relation entre la conscience et les représentations de la conscience. Il est pour le moins étrange d'affirmer que la conscience ne joue aucun rôle causal dans l'énonciation de la proposition « Je suis conscient ». Pour certains auteurs, cela mène à une conclusion beaucoup plus forte, selon laquelle expliquer la conscience serait tout simplement au-delà de la portée de la compréhension humaine.

Comme réponse, l'épiphénoménaliste peut choisir de rejeter l'idée que la connaissance présuppose toujours une connexion causale, et dire qu'il existe une relation plus forte entre la conscience et les croyances sur la conscience, dans le sens où la conscience joue un rôle crucial dans la constitution des croyances et des concepts phénoménaux. L'expérience du rouge joue un rôle constitutif dans la formation de la croyance « *X* est en train de voir du rouge ». Bref, il n'y a aucune distance causale entre l'expérience et la croyance. Par conséquent, cette immédiateté qui caractérise le couple expérience/croyance a pour effet la justification de la croyance. Si tel est le cas, alors l'épiphénoménalisme ne pose aucun obstacle à la compréhension du phénomène de la conscience.

[84](D. Chalmers, 2003a)

L'étrangeté de l'épiphénoménalisme augmente lorsqu'on réalise que les rapports qui lient la conscience aux assertions sur la conscience semblent être le fruit d'une heureuse coïncidence. Après tout, remarque alors à juste titre Chalmers, si les lois psychophysiques sont indépendantes de l'évolution du monde physique, il s'ensuit qu'il peut y avoir d'autres mondes possibles dans lesquels l'évolution physique serait la même que la nôtre, tandis que les lois psychophysiques seraient radicalement différentes, provoquant ainsi un « dépareillage », une disjonction entre l'expérience et son compte-rendu. Il semblerait alors que nous soyons extrêmement chanceux de vivre dans un monde où tout concorde dans une harmonie leibnizienne psychophysique totale et quasi-miraculeuse.

En conclusion, même si l'épiphénoménalisme[85] s'avère une position cohérente sans défaillances fatales, il reste néanmoins l'option la moins élégante des toutes. Inélégance et contre-intuitivité sont donc les deux défauts majeurs de l'épiphénoménalisme.

4.5.3 Monisme de type-F

Le monisme[86] de type-F affirme que la conscience est une propriété intrinsèque des éléments physiques fondamentaux. Dans cette acception, les propriétés phénoménales *sont* des catégories fondamentales des structures physiques elles-mêmes. En clair, les propriétés phénoménales ou protophénoménales sont situées au niveau physique fondamental de la réalité, et dans un sens, sous-tendent la réalité physique elle-même. Un des inspirateurs de cette vue est B. Russell. Dans *The Analysis of Matter*[87], le philosophe britannique observe que la science physique distingue les entités physiques en fonction des relations qu'elles entretiennent entre elles, ainsi qu'en fonction rapports qui les connectent à nous, êtres humains. Un quark, par exemple, est caractérisé par les relations qu'il entretient avec d'autres particules physiques, tandis

[85] Selon Chalmers, les auteurs qu'on peut ranger dans la catégorie « dualisme de type-E » sont : T. Huxley, (1874), 'On the Hypothesis that Animals are Automata, and its History', *Fortnightly Review*, 95: 555–80. Réédité in *Collected Essays*, London, 1893; F. Jackson (1982); W.S. Robinson (1988), *Brains and People: An Essay on Mentality and its Causal Conditions*, Philadelphia: Temple University Press.

[86] Les défenseurs d'une version du monisme de type-F, selon Chalmers, sont : B. Russell (1927), *The Analysis of Matter*, London: Kegan Paul ; M. Lockwood, (1989), *Mind, Brain, and the Quantum*, Oxford: Oxford University Press; G. Strawson, (2000), 'Realistic Materialist Monism', in S. Hameroff, A. Kaszniak, & D. Chalmers (eds.), *Toward a Science of Consciousness III*. Cambridge, MA: MIT Press; et D. Stoljar, 2001.

[87] (B. Russell, 1927)

qu'une propriété primitive comme la masse est caractérisée par une disposition afférente, comme par exemple la disposition de résister à l'accélération.

Et pourtant la science physique ne nous dit rien sur la nature *intrinsèque* de ces entités et propriétés. Or, en règle générale, les relations et les dispositions sont constituées par des propriétés subjacentes et intrinsèques qui fondent, depuis l'intérieur, les entités concernées *via* ces relations et dispositions. A partir de là, nous sommes confrontés à un véritable problème métaphysique : quelles sont les propriétés intrinsèques des systèmes physiques fondamentaux ? Et comme un ennui ne vient jamais seul, une deuxième interrogation métaphysique vient s'ajouter à la première : comment les propriétés phénoménales intrinsèques s'assimilent-elles à un monde essentiellement physique ? Les propriétés phénoménales, en tant que propriétés intrinsèques, semblent difficilement s'accorder avec une description purement physique du monde. Russell suggère alors une solution qui pourrait faire d'une pierre deux coups, pour ainsi dire : il est possible, selon lui, que les propriétés intrinsèques du monde physique soient à la base de nature phénoménale ; ou peut-être ces propriétés intrinsèques du monde physique ne sont pas phénoménales, mais elles constituent malgré tout les propriétés phénoménales, i.e. des propriétés *protophénoménales.* Si tel est le cas, conscience et réalité physique seraient profondément et inextricablement enchevêtrées. Etats phénoménaux et états physiques deviennent alors les membres légitimes d'un tout naturel indivisible. Le mot « nature » doit être compris ici comme un ensemble constitué d'entités dont les qualités protophénoménales intrinsèques participent activement dans le rouage causal physique.

L'avantage de ce type de monisme – qui est, par ailleurs, parfaitement compatible avec le principe de clôture causale, ainsi qu'avec l'applicabilité des lois de la physique – est qu'il parvient à greffer l'hypothèse d'une nature intrinsèque des propriétés phénoménales sur le tronc des théories physiques déjà existantes, les propriétés protophénoménales servant alors de base primitive pour toute causation physique. Protéiforme, le monisme de type-F s'inspire à la fois du matérialisme et du dualisme. D'une part, si les termes physiques font référence non pas à des propriétés dispositionnelles, mais à des propriétés intrinsèques sous-jacentes, alors les propriétés protophénoménales peuvent être considérées comme étant physiques, et dans ce cas, le matérialisme est préservé. D'autre part, si les propriétés phénoménales et protophénoménales sont ontologiquement primitives, fondamentales, alors il en résulte une certaine dualité séparant les propriétés de type structural-dynamiques (qui tombent

sous la juridiction des théories physiques) des propriétés protophénoménales (qui interviennent dans le phénomène de la conscience).

Cette position est parfois présentée sous la forme d'un *monisme neutre* : il existe des propriétés sous-jacentes *X* (les propriétés protophénoménales) en sorte que *X* participent simultanément à la constitution du domaine physique (*via* le réseau causal), ainsi qu'à la constitution du domaine phénoménal (grâce à leur nature globale intrinsèque).

Ce type de monisme peut parfois prendre la forme d'un *idéalisme* : les propriétés mentales constituent les propriétés physiques, bien que ces dernières ne soient pas obligatoirement situées dans l'esprit d'un sujet. Une forme plus radicale du monisme de type-F concerne le panpsychisme, ou le 'protopanpsychisme', mais la discussion de ces points dépasserait les cadres du thème que nous enquêtons ici. Il nous suffit de préciser qu'à l'instar de l'épiphénoménalisme, le monisme de type-F semble compatible *prima facie* avec toutes les données dont nous disposons. Bien que ce type de monisme ne fasse l'objet d'aucune critique décisive, d'aucun argument '*knock-down'* pour ainsi dire, plusieurs possibles objections sont néanmoins à signaler à son encontre.

En premier lieu, si cette position est juste, alors les choses seraient essentiellement et littéralement identiques à leurs apparences. Or, l'évolution de la science nous a appris à nous méfier des apparences sur lesquelles reposent habituellement les explications du sens commun. Une autre objection concerne la nature pour le moins mystérieuse de ces propriétés protophénoménales, mais la réplique la plus courante dans ce cas serait de mettre cet aspect mystérieux sur le compte de notre ignorance ou des limitations de nos capacités cognitives. En tout cas, aucune de ces objections ne se montre fatale pour la vérité du monisme de type-F.

Il existe cependant une faille dangereuse dans l'édifice de ce type de modèle, note Chalmers[88]. De toute évidence, notre phénoménologie possède une structure spécifique et complexe : elle est unitaire, encadrée et différenciée sous des multiples facettes, mais dotée d'une homogénéité sous-jacente qui est étroitement liée au fait qu'elle possède un seul et unique angle d'expérience, le moi, le sujet conscient. Il est difficile de concevoir alors par quel moyen l'existence et la distribution d'un grand nombre de systèmes individuels microphysiques, chacun doté de ses propres propriétés protophénoménales, pourrait se greffer en quelque sorte à cette structure complexe et spécifique qu'est la

[88] (D. Chalmers, 2003a)

phénoménologie d'un moi individuel. Ne devrait-on s'attendre à une structure plus désordonné, semblable à un conglomérat, ou à un patchwork bariolé ?

Ce problème est parfois présenté comme une version de ce qu'on appelle parfois le *problème de l'assemblage*[89], ou *le problème de la discordance (dissonance) structurale*[90] (« the structural mismatch »). Pour les contrecarrer, nous avons besoin d'une meilleure compréhension des principes compositionnels de la phénoménologie, c'est-à-dire des principes par le truchement desquels les propriétés phénoménales peuvent être composées ou constituées à partir des propriétés protophénoménales sous-jacentes. Or, bien que l'on ait une idée assez claire des principes qui gèrent les structures et la composition du niveau physique, nous sommes loin d'en avoir l'équivalent concernant la compréhension des principes qui régissent le niveau phénoménal. Par conséquent, on peut dire, sans crainte d'exagération, que le problème de la discordance structurale pose un sérieux défi au monisme de type-F.

Notons de manière générale, suivant Chalmers, que le monisme de type-F offre une description à la fois sophistiquée et élégante de la nature de la conscience. Bien qu'aucune théorie détaillée de ce type n'ait été solidement développée jusqu'à présent, il semblerait que les objections qu'elle affronte ne soient pas assez fortes pour enterrer le monisme de type-F comme une doctrine définitivement fausse. Nous sommes par conséquent en présence d'un chantier très prometteur et en pleine construction.

Pour résumer, nous avons tenté de tirer un sillon, parmi tant d'autres possibles[91], dans le labyrinthe que forment les divers modèles explicatifs matérialistes contemporains[92]. Mais ce serait se livrer à une grossière exagération que de décrire le

---

[89] W. Seager, (1995), 'Consciousness, information, and panpsychism', *Journal of Consciousness Studies*, 2: 272-88.

[90] D. Stoljar, (2001), 'Two conceptions of the physical', *Philosophy and Phenomenological Research*, 62: 253-81. Réédité in Ludlow et al., (2004) (Nous utilisons ici cette dernière édition.)

[91] Quelles sont les options qui restent aux non-réductionnistes ? Deux positions demeurent en dehors du classement de A à F proposé par Chalmers. D'une part, il y a ceux qui affirment que les propriétés phénoménales sont ontologiquement distinctes des propriétés physiques. Bien que la microphysique soit un domaine causalement clos, les propriétés phénoménales *ont* une efficience causale sur le physique. On pourrait appeler cette position la thèse de la « surdétermination causale » (« causal overdetermination ») : les états physiques peuvent déterminer causalement le comportement, et, simultanément, les états phénoménaux peuvent avoir une efficience causale sur le comportement. Une autre option serait la médiation causale (« causal mediation ») : dans certains cas, la causation microphysique allant de phénomène A vers le phénomène B présuppose un rapport causal effectif de A vers l'esprit de B, en sorte que l'esprit pénètre dans le réseau causal sans altérer, pour ainsi dire, la structure physique même du réseau. De leur côté, certains non-matérialistes sont idéalistes, à la manière de Berkeley. Mais ces options restent minoritaires.

[92] Selon Chalmers, les versions les plus plausibles restent le dualisme de type-D, type-E, et le monisme de type-F, mais la discussion de ces points est, à l'heure où nous écrivons ces lignes, en pleine effervescence.

destin du matérialisme et des *qualia* comme étant déjà tracé et épuisé par la description de ces variantes. Une chose est sûre : la leçon fondamentale à retenir à la suite de ces présentations est que le phénomène conscience occupe bel et bien une place fondamentale dans l'étude du réel.

Dans cette section nous avons tenté d'esquisser, à titre d'étape préliminaire et synoptique, la physionomie des principaux arguments qui ont façonné, et qui continuer de le faire à l'heure même où nous écrivons ces lignes, l'arrière-plan conceptuel sur lequel évolue le physicalisme. Si nous avons brièvement dressé le croquis de ces arguments, c'est que nous croyons que leur explicitation nous permettra de mieux saisir les mécanismes qui sous-tendent l'ascension du physicalisme au rang du 'Weltanschauung' de notre siècle. Mais avant d'en venir là, un long chemin est à parcourir. Un chemin qui passe par la mise en examen d'un des plus influents arguments antiphysicalistes, l'argument de la connaissance de F. Jackson.

Deuxième partie

L’argument de la connaissance

5 Mary, la prisonnière

Dans une formule mi-sérieuse mi-provocatrice, W. Lycan[1] notait récemment qu'un jour viendra où l'on n'écrira plus d'articles sur l'argument de la connaissance. Reste à savoir, note si ce jour coïncidera avec la fin du monde ou pas. On s'accorde à dire qu'aucun philosophe contemporain de l'esprit ne peut se permettre le luxe d'ignorer le physicalisme. D'autre part, on dit souvent qu'aucun physicaliste, à son tour, ne peut se permettre le luxe d'ignorer l'argument de la connaissance. Dans la première partie de cet essai, nous avons examiné le versant physicaliste du duel qui nous intéresse ici. Il est temps à présent de se tourner vers le deuxième protagoniste de cette intrigue philosophique, à savoir l'argument de la connaissance de F. Jackson.

Un argument de la connaissance est avant tout une tentative de transformer, *via* une alchimie mystérieuse, « le métal ordinaire des prémisses épistémiques en l'or noble des conclusions ontologiques » [2]. Autrement dit, en partant de ce qu'un sujet *connaît*, on légifère sur ce que les choses *sont*. Ainsi formulé, il est clair que l'argument impliquant la désormais célèbre neuroscientifique Mary n'est pas, strictement parlant, le seul argument de la connaissance dans l'histoire de la philosophie. D'autres philosophes, à des époques différentes, ont nourri l'ambition de moissonner l'or noble des thèses ontologiques en s'appuyant sur la terre noire des prémisses épistémiques.

Ce chapitre nous allons procéder de la façon suivante : dans la section 5.1. nous allons jeter un bref regard rétrospectif sur les « précurseurs » de Mary. Cette parenthèse nous permettra de mieux saisir l'épure des arguments de la connaissance en général. Ensuite, nous allons présenter très brièvement l'argument tel qu'il fut façonné par F. Jackson. Les sections allant de 5.2. à 5.8. seront entièrement dédiées à l'analyse détaillée des enjeux contenus dans l'argument de la connaissance de Jackson. Nous verrons que l'argument s'appuie sur une *intuition* de départ qui comporte deux volets distincts : (i) d'abord il y a l'idée que Mary, avant sa libération, sait absolument *tout ce qui est à savoir* sur la nature *physique* du monde. (ii) Ensuite, il y a l'idée que Mary, en sortant de sa prison, *apprend quelque chose*. Seul ce dernier volet, appelé

[1] "Someday there will be no more articles written about the "Knowledge Argument"… That is beyond dispute. What is less certain is how much sooner that day will come than the heat death of the universe." W. Lycan (2003), 'Perspectival Representation and the Knowledge Argument', *in* Q. Smith & A. Jokic (eds.), *Consciousness: New Philosophical Essays*, Oxford University Press, p. 384.
[2] (Y. Nagasawa, 2008:3)

communément « *la thèse de l'apprentissage* », sera examiné dans ce chapitre. Le premier volet, i.e. *la thèse de l'information physique complète*, fera l'objet d'un examen à part dans le chapitre 6. Une comparaison sera faite, inspirée des travaux d'Y. Nagasawa, entre *l'argument de la connaissance*, d'une part, et *l'argument de la possession du concept* dans la philosophie de la religion, d'autre part. La discussion de ces points autour des concepts clé d' « omniscience » et d' « ignorance »[3] nous aidera à la fois de cerner les véritables mécanismes qui placent le physicalisme au cœur de la métaphysique contemporaine de l'esprit, et nous servira de tremplin pour atteindre et développer, dans le chapitre 7, le cœur de la thèse que nous défendons ici. L'objectif y sera de dresser un examen critique du physicalisme *a priori* du 'deuxième' Jackson (1998). Nous verrons ainsi que des motivations profondes relient le physicalisme de « première génération » à son ambitieux successeur, le physicalisme *a priori*. Mais avant, il nous reste un long chemin à faire, passant par la présentation détaillée d'un des plus redoutables adversaires du physicalisme : l'argument de la connaissance de F. Jackson.

5.1 L'archange, le super-martien, la chauve-souris, et le savant sourd

L'argument de la connaissance de F. Jackson (KA)[4] va ainsi :

« Mary est une brillante scientifique qui […] doit effectuer ses 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 couleur ? *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 avions posé au départ qu'elle possédait *toutes* les connaissances physiques. Il

[3]Voir D. Stoljar, (2006), *Ignorance and Imagination – The Epistemic Origin of the Problem of Consciousness*. Oxford: Oxford University Press.

[4] Dans la littérature analytique, l'argument de la connaissance de F. Jackson est souvent désigné sous l'abréviation « KA » (« the Knowledge Argument »). Nous avons choisi de garder ici cette abréviation afin de suivre le plus fidèlement possible son utilisation dans les textes originaux.

s'ensuit que les connaissances physiques n'épuisent pas l'ensemble des connaissances et que le physicalisme est faux »[5].

Comme nous pouvons le constater, le succès de l'argument tient non seulement à la force de l'intuition qui le fonde, mais surtout à la simplicité désarmante de l'expérience de pensée mettant en scène la scientifique prisonnière. Mais avant d'examiner de plus près sa structure, il est indispensable de procéder à un nouveau détour, et de passer brièvement en revue les « prédécesseurs » de Mary.

Dès 1925, C.D. Broad[6] discute, dans un chapitre dédié aux implications philosophiques du concept de « mécanisme », le cas d'un « archange mathématicien » doté de pouvoirs extraordinaires, et qui connaît *toutes* les vérités logiques et mécanistiques concernant les composants chimiques du monde. Prenons, dit Broad, l'exemple d'un énoncé banal que l'on trouve dans n'importe quel manuel de chimie : « lors d'une décharge électrique, les molécules d'azote et d'hydrogène se combinent et donnent naissance à un composé chimique nouveau qui contient trois atomes d'hydrogène et un d'azote. Dans les conditions ambiantes, il se présente sous la forme d'un gaz incolore, produisant une fumée de condensation à haute concentration, et dégage une très désagréable odeur putride ». Selon Broad, il est évident qu'en dépit de son omniscience concernant ces faits « mécanistiques », l'archange sera dans l'impossibilité de prédire, ou de reconnaître lorsqu'il la rencontre, l'odeur si particulière, putride, que dégage l'ammoniac.

Dans un passage célèbre, B. Russell[7] rappelle qu'un individu doté d'une vue normale peut voir des choses qu'un aveugle ne peut pas voir. Or, étant donné que l'aveugle peut connaître, en principe, *toutes* les vérités physiques, il s'ensuit que la connaissance que peut avoir l'individu « normal » n'est pas une part de la physique.

H. Feigl[8] évoque à son tour une expérience de pensée impliquant un martien savant auquel on lui aura appris *tous* les faits physiques concernant l'organisation et le comportement des organismes vivants. Malgré sa connaissance complète de ces informations traduisibles en faits physiques, le martien sera dans l'impossibilité

---

[5] (F. Jackson 1982 :130) Trad. fr. par P. Poirier, in D. Fisette et P. Poirier (eds.), *Philosophie de l'esprit*, op. cit., pp. 201-2.
[6] (C.D. Broad, 1925)
[7] "It is obvious that a man who can see knows things which a blind man cannot know; but a blind man can know the whole of physics. Thus the knowledge which other men have and he has not is not a part of physics". B. Russell, (1927), *The Analysis of Matter*, London: George Allen and Unwin, p.389.
[8] (H. Feigl, 1958)

d'éprouver l'expérience qualitative qui accompagne la vue des couleurs, le son de la musique, ou le parfum d'une rose.

Plus tard, dans un ouvrage paru en 1982, H. Robinson[9] évoque une version différente de l'argument de la connaissance, impliquant cette fois-ci le cas d'un savant sourd. Privé dès sa naissance du sens de l'ouïe, l'infortuné savant parvient néanmoins à devenir un véritable expert dans la neurologie de l'audition. Or, malgré son incroyable expertise dans l'étude de l'audition, nous gardons la très puissante intuition selon laquelle un savant sourd manque un fait essentiel, à savoir l'effet-que-cela-fait d'entendre une sonate de Scarlatti[10], disons.

La chauve-souris de Th. Nagel[11] est aussi célèbre que la prisonnière Mary : en principe, dit Nagel, un sujet *X* pourrait connaître toute l'information physique concernant les chauves-souris, et pourtant *X* serait toujours dans l'impossibilité de connaître *l'effet-que-cela-fait* d'être une chauve-souris. Selon la synthèse qu'en propose *W*. Lycan[12], l'argument de Nagel peut être retranscrit ainsi :

1. Il existe un effet-que-cela-fait d'être une chauve-souris ayant des sensations de type-sonar S.
2. *X* est en possession d'une connaissance complète et exhaustive des faits caractérisant la physiologie d'une chauve-souris.

CHAUVE SOURIS

Mais

3. *X* ne connaît pas l'effet-que-cela-fait d'être une chauve-souris en train d'éprouver S.
4. Si F1 = F2, alors tout individu qui connaît F1 connaît F2 également.

[9] H. Robinson, (1982), *Matter and Sense: A Critique of Contemporary Materialism*. Cambridge: Cambridge University Press.

[10] Des expériences de pensée analogues ont été avancées dans des ouvrages non-philosophiques, comme par exemple dans les livres d'E. O. Wilson, *Consilience*, ou par D. Lodge (*Thinks* . . .), mais aussi dans une série de la télévision britannique, *Brainspotting*. Voir T. Alter, (2007), 'The Knowledge Argument', *in* M. Velmans et S. Schneider, op. cit., p. 396

[11] (Th. Nagel, 1974)

[12] (W. Lycan, 1987:76)

Donc

5. Il existe un fait $F_{chauve\text{-}souris}$ – à savoir le fait de connaître l'*effet-que-cela-fait-d'être-une-chauve-souris* - qui est distinct de tous les faits physiologiques et fonctionnels caractérisant la chauve-souris.
6. Si le matérialisme est vrai, alors tout fait mental relatif à la chauve-souris est identique ou traduisible dans un fait physiologique ou fonctionnel**Erreur ! Signet non défini.**.

Ergo,

7. Il s'ensuit (de 5 et 6) que le matérialisme est faux.

On ne s'arrêtera pas ici sur les critiques, abondamment discutées ailleurs[14], élevées au fil des ans contre la célèbre expérience de pensée de Nagel. Notons seulement que pour les défenseurs du matérialisme, il existe plusieurs interprétations possibles de la prémisse (3), pas toutes incompatibles avec le matérialisme. L'énoncé (3) «*X* ne connaît pas l'effet-que-cela-fait-d'être-une-chauve-souris en train d'éprouver S » peut être compris sous une des deux formes suivantes : (i) nous, les humains, *n*'avons *pas* les sensations sonar S, et donc nous ne pouvons pas les éprouver par nous-mêmes S ; (ii) nous, les humains, n'avons pas le même *accès cognitif* à S que la chauve-souris. Mais ces deux thèses (i) et (ii) sont non seulement trivialement vraies, mais aussi acceptées sans réticence par les partisans du matérialisme.

Tous ces personnages connaissent des mésaventures semblables à celles de la captive Mary, et pourtant c'est le cas particulier de cette dernière qui a déclenché un véritable flot montant d'interrogations concernant l'expérience phénoménale. Comment expliquer cela ?

[13] NB : les « faits » doivent être compris ici comme des « occurrences » (i.e. « tokens ») et non pas des « types ».
[14] Voir D. Dennett & D. Hofstadter, (1981), *The Mind's I*, Basic Books.

5.2 La structure de l'argument de Jackson et ses retranscriptions

En lisant l'article que Jackson publie en 1982, on s'aperçoit que notre héroïne Mary est devenue célèbre presque par accident. Car initialement, c'est *Fred* qui tenait la tête d'affiche, pour ainsi dire. Qui est Fred ? Fred est un personnage qui a une capacité discriminative chromatique beaucoup plus importante qu'un individu normal. Il est capable de voir, de réellement voir, une couleur rouge$_1$ qui est différente de rouge$_2$, disons. Une question surgit alors : quel est l'effet-que-cela-fait de voir ces nouvelles nuances de rouge ? Tout individu doté d'une capacité discriminative normale aimerait pouvoir identifier ce nouveau *quale$_1$*, mais hélas, il ne peut pas. Et même si un savant fou arrive à scruter le cerveau de Fred afin d'en soustraire toutes les informations physiques disponibles, il est clair que notre connaissance sera à jamais incomplète, car il nous manque la connaissance de l'effet-que-cela-fait de voir la couleur rouge$_1$. L'histoire de Fred présente un certain air de famille avec l'expérience de la chauve-souris de Nagel. Mais Jackson[15] insiste sur un point essentiel: la difficulté qu'il veut souligner concerne non pas l'ignorance de l'effet-que-cela-fait d'*être* Fred, mais un tout autre aspect, à savoir l'*ignorance* d'un fait *à propos de* l'expérience de Fred, bref d'une *propriété* le concernant.

Comme le note Jackson, la faiblesse de l'argument de Nagel tient au fait que l'appareil cognitif des chauves-souris est trop différent de celui de l'homme pour que l'impact de la comparaison puisse ébranler sérieusement le physicalisme. Après tout, le physicalisme ne repose pas sur la discutable prémisse suivant laquelle un être humain pourrait ou devrait pouvoir imaginer l'effet-que-cela-fait d'être une chauve-souris. En revanche, le cas de Fred est beaucoup plus concluant : car si l'on connaît la totalité des faits physiques concernant Fred, on devrait connaître aussi l'effet-que-cela-fait, pour Fred, d'avoir l'expérience de rouge$_1$. Or, il n'en est rien.

Le cas de Mary est encore plus révélateur, car notre prisonnière ne possède aucun extra-pouvoir à l'aide duquel elle pourrait « tricher » en quelque sorte, et éprouver une expérience qualitative additionnelle, située hors de la portée des individus normaux,

[15] "When I complained that all physical knowledge about Fred was not enough to tell us what his special colour experience was like, I was not complaining that we weren't finding out what it is like to *be* Fret I was complaining that there is something *about* his experience, a property of it, of which we were left ignorant. And if and when we come to know what this property is we still will not know what it is like to *be* Fred, but we will know more about him. No amount of knowledge *about* Fred, be it physical or not, amounts to knowledge "from the inside" concerning Fred". (F. Jackson, 1982/2004: 44)

comme c'est le cas de Fred et de la couleur rouge$_1$. Conscient de ces failles, Jackson revient judicieusement sur son expérience de pensée et, quatre années plus tard, il publie un article[16] devenu le point de départ des controverses autour du statut métaphysique du physicalisme mis sur le banc d'essai dans l'affaire des *qualia*. Jackson reformule alors son argument ainsi :

(P1) Mary (avant sa libération) connaît toute l'information physique concernant les objets et les autres sujets.

(P2) Mary (avant sa libération) ne connaît pas toute l'information physique concernant les objets et les autres sujets (car elle *apprend* quelque chose de nouveau lors de sa libération, à sa sortie de prison).

KA

Par conséquent,

(P3) Il existe des propositions vraies (des vérités) à propos des objets et des autres sujets qui échappent au récit physicaliste.

Comme le résume fort bien D. Stoljar et Y. Nagasawa[17], l'argument de Jackson s'appuie sur une *intuition* de départ qui comporte deux volets distincts : (i) Mary, avant sa libération, sait absolument *tout ce qui est à savoir* sur la nature *physique* du monde. (ii) Mary, en sortant de sa prison, *apprend quelque chose*. Si l'on adhère à ces deux thèses, l'argument se montre très puissant, et le physicalisme est sévèrement menacé. Comme nous le verrons par la suite, les défenseurs du physicalisme sont loin de se contenter de baisser les bras. Les échanges durant ce duel philosophique obligeront ses protagonistes à replacer leurs pions sur l'échiquier initial du débat.

En plus du travail de clarification entrepris par Jackson dans son article de 1986, l'argument a été transcrit sous diverses autres variantes. En premier lieu, comme le relève Stoljar[18], l'argument se prête à une reformulation à l'aide de quantificateurs logiques :

---

[16] F. Jackson, (1986), 'What Mary didn't know', *Journal of Philosophy*, 83: 291-5. Repris in Ludlow et al., (2004), op. cit. (Nous utilisons ici cette dernière édition.)

[17] D. Stoljar & Y. Nagasawa (2004), *Introduction* in *There is something about Mary: essays on phenomenal consciousness and Frank Jackson's knowledge argument*, op. cit.

[18] "(P1) Every physical truth is such that Mary (before her release) knows that truth. (P2) It is not the case that every truth is such that Mary (before her release) knows that truth. (C) Therefore, there is at least one truth that is non physical". (D. Stoljar, 2004: 14)

KA LOGIQUE

P1 Chaque vérité physique est telle que Mary (avant sa libération) soit en possession de cette vérité.

P2 Ce n'est pas le cas que chaque vérité physique soit telle que Mary (avant sa libération) soit en possession de cette vérité.

Par conséquent,

P3 Il existe au moins une vérité qui est non-physique.

Ainsi formulé, l'argument est indéniablement valide. Si j'examine, par exemple, toutes les *pommes* qui se trouvent dans une boîte, mais sans examiner tous les *fruits* qui se trouvent dans la même boîte, il est clair la conclusion « il existe des fruits dans la boîte qui ne sont pas des pommes » peut être vraie. Cependant, l'inférence qui va de (P4) à (P6) ne suffit pas, à elle toute seule à destituer le physicalisme : en effet, il est possible de montrer que la conclusion (P6) n'est pas assez forte pour inquiéter les intérêts des physicalistes. Ces derniers peuvent maintenir que le physicalisme est compatible avec l'existence des vérités non-physiques. Pourquoi pas, après tout ? Il suffit, selon Stoljar, de faire appel à la notion clé de « survenance » pour contourner cette difficulté : supposons qu'il serait possible de réunir toutes les informations physiques à propos du monde dans une seule « super-vérité » *P*, et d'autre part toutes les propositions vraies concernant les états psychologiques dans une « super-vérité » *Q*. Considérons alors l'énoncé conditionnel suivant : « Si *P* alors *Q* » et baptisons-le, à la suite de Stoljar, le « *conditionnel psychophysique* » (CP)[19]. Dans ce cas, dire que le mental survient sur le physique revient à dire que l'énoncé *conditionnel psychophysique est nécessairement vrai*. Un prolongement intéressant des remarques précédentes est que si le physicalisme est vrai, alors le CP est nécessairement vrai.

A partir de là, la question de savoir si l'énoncé conclusif (P6) entraîne ou pas la faillite du physicalisme, dépend en grande mesure de la compatibilité entre le caractère nécessaire du conditionnel psychophysique, d'une part, et l'existence des vérités non-physiques d'autre part. Si l'on s'appuie sur la prémisse selon laquelle les vérités nécessaires *a posteriori* sont possibles[20], on s'aperçoit alors que le KA dans ses deux formulations de 1982 et 1986, repose non seulement sur la thèse (A) :

[19] *Ibid.*, p. 15. (Désormais CP en abrégé).

[20] Dans *La logique des noms propres*, Kripke opère une distinction importante entre la nécessité épistémique et la nécessité métaphysique qui mène à replacer sur deux plans indépendants les

(A) Si le physicalisme est vrai, alors le CP est *nécessaire.*

Mais également sur la thèse (B) :

(B) Si le physicalisme est vrai, alors le CP est *a priori*[21].

Bien que la distinction kripkéenne entre « nécessaire *a priori* »/ « nécessaire *a posteriori* » est aujourd'hui largement adoptée parmi les philosophes contemporains, il est clair que pour Jackson, le CP n'est pas une vérité nécessaire *a posteriori* : car si la visée ultime de physicalisme est celle d'offrir une explication complète et exhaustive du monde, et si le physicalisme est tenu pour vrai, alors le CP est non seulement nécessaire, mais aussi *a priori*.

A partir de là, l'argument de Mary revêt une forme *a priori* que récemment M. Tye[22] a retranscrit ainsi :

KA A PRIORI

P4 Mary, dans sa prison, connaît P, et elle est un être rationnel *a priori* idéal, mais elle ne sait pas qu'elle est en posture de connaître que *Q*.

P5 Si (P7) est vraie, alors *Q* ne découle pas de façon *a priori* de P.

P6 P! *Q* n'est pas *a priori*, alors le physicalisme est faux

Ergo,

P7 Le physicalisme est faux.

L'avantage considérable de cette retranscription, c'est qu'elle nous permet de saisir toute la puissance argumentative de l'expérience de pensée de Jackson contre le

oppositions *a priori*/*a posteriori* et nécessaire/contingent. Prenons l'énoncé : « Hesperus est Phosphorus ». Selon Kripke, cet énoncé est bel et bien un énoncé dont le statut épistémique est *a posteriori* parce que l'identité qu'il contient n'est pas de nature synonymique, mais relève plutôt d'une observation du monde. Toutefois, ce caractère *a posteriori* se marierait avec la nécessité de l'identité, car, selon Kripke, tout énoncé d'identité entre deux désignateurs rigides ayant le même référent est nécessairement vrai. Pourquoi? Simplement parce que si le désignateur rigide conserve toujours le même référent à travers les mondes possibles – ici Vénus – il ne peut se trouver un monde pour lequel le nom Hesperus ne renvoie pas à Venus et il en va de même pour Phosphorus. Ainsi, l'identité entre les deux noms, bien qu'*a posteriori* possède, selon Kripke, une nécessité métaphysique. (Voir S. Kripke, 1980)

[21] (D. Stoljar, 2005:15)

[22] "(5) Mary in her room knows that P and she is an ideal *a priori* reasoner, but still she isn't in a position to know that *Q*. (6) If (5) is true, then *Q* does not follow *a priori* from P. (7) If P ! *Q* is not *a priori*, then physicalism is false. Therefore, (8) Physicalism is false". Voir M. Tye, (2009), *Consciousness Revisited - Materialism without Phenomenal Concepts: A New Perspective on the Major Puzzles of Consciousness*, Cambridge, Mass: The MIT Press, Bradford Books, p. 124.

physicalisme. Propulsé ainsi en tête d'affiche dans l'affaire de l'expérience phénoménale, le cas de Mary a reçu de nombreux commentaires. Très vite, certains critiques ont attiré l'attention sur la nécessité de préciser la signification des termes protagonistes de l'argument, afin d'éviter toute confusion. T. Horgan[23] note, par exemple, que la notion d'« information physique » reste floue dans la formulation de Jackson, et peut dès lors engendrer des fâcheux malentendus. En effet, l'expression « information physique » peut être interprétée:

(a) soit en tant qu'« information physique explicite ». Une proposition S exprime une information physique *explicite* concernant des processus déterminés si et seulement si S appartient ou dérive d'une explication théorique adéquate de ces mêmes processus.

(b) soit comme une « donnée ontologiquement physique » : une proposition S exprime une information ontologiquement *physique* seulement dans le cas où (*b1*) toutes les entités auxquelles S réfère (ou qui sont quantifiées en S) sont des entités physiques ; et (*b2*) toutes les propriétés et les relations exprimées par les prédicats en S sont des propriétés et relations physiques[24].

Cette distinction introduite par Horgan nous permet, comme le relève fort bien M. Nida-Rümelin[25], de saisir avec plus de précision les deux versants, épistémique et ontologique, du problème de Mary. Il est possible, en effet, de remplacer la locution « posséder toute l'information physique explicite sur *X* » par la formule « avoir une *connaissance* physique complète de *X* ». De manière analogue, on peut convertir l'expression « avoir toutes les informations ontologiques physiques sur *X* » dans la formule « connaître tous les *faits* physiques à propos de *X* ». Conséquemment, l'argument de la connaissance de Jackson (KA) peut être retranscrit sous deux formes distinctes, selon qu'il emprunte le versant épistémique ou ontologique.

[23] T. Horgan, (1984), 'Jackson on Physical Information and *Qualia*', *Philosophical Quarterly* 32, 127–136. Repris in Ludlow et al., op. cit., 2004 (Nous utilisons ici cette dernière édition).
[24] *Ibid.*, p. 304.
[25] M. Nida-Rümellin, '*Qualia* : The Knowledge Argument' in http://plato.stanford.edu/entries/*qualia*-knowledge

5.2.1 (V1) Version Faible ou épistémique du KA

KA ÉPISTÉMIQUE

(1a) Mary (avant sa libération) possède une *connaissance physique complète* des faits concernant la vision des couleurs chez l'être humain.

(2a) Il existe néanmoins un certain *type de connaissance* concernant la vision des couleurs chez les humains que Mary ne possède pas durant sa captivité.

Par conséquent,

(3a) Il y a un certain *type de connaissance* des faits concernant la vision des couleurs chez les humains qui est une connaissance non-physique[26].

Tournons-nous à présent vers la formulation ontologique du KA.

5.2.2 (V2) Version Forte ou ontologique du KA

KA ONTOLOGIQUE

(1b) Mary (avant sa libération) connaît tous les *faits physiques* concernant la vision des couleurs chez l'être humain.

(2b) Il y a cependant *certains faits* concernant la vision des couleurs chez l'être humain que Mary (avant sa libération) ne connaît pas.

Par conséquent,

(3b) Il y a des faits non-physiques concernant la vision des couleurs chez l'être humain[26].

[26] M. Nida-Rümellin, '*Qualia* : The Knowledge Argument' in http://plato.stanford.edu/entries/*qualia*-knowledge
[27] *Ibid.*

Bien que la formulation initiale du KA se prête aux deux possibles interprétations (V1) et (V2), les intentions de Jackson se rattachent clairement à la version forte de l'argument. Tenant compte de ces précisions, il est à présent clair qu'afin de garantir la portée ontologique du physicalisme, son défenseur doit fermement rejeter (3b). Nombre d'auteurs ont attiré l'attention sur le fait que la vérité de la thèse (3a) n'implique pas nécessairement la vérité de (3b). Pour voir cela, note M. Nida-Rümelin, il suffit de penser à un exemple lié à un type de connaissance particulier, que l'on appelle la connaissance *de se*.

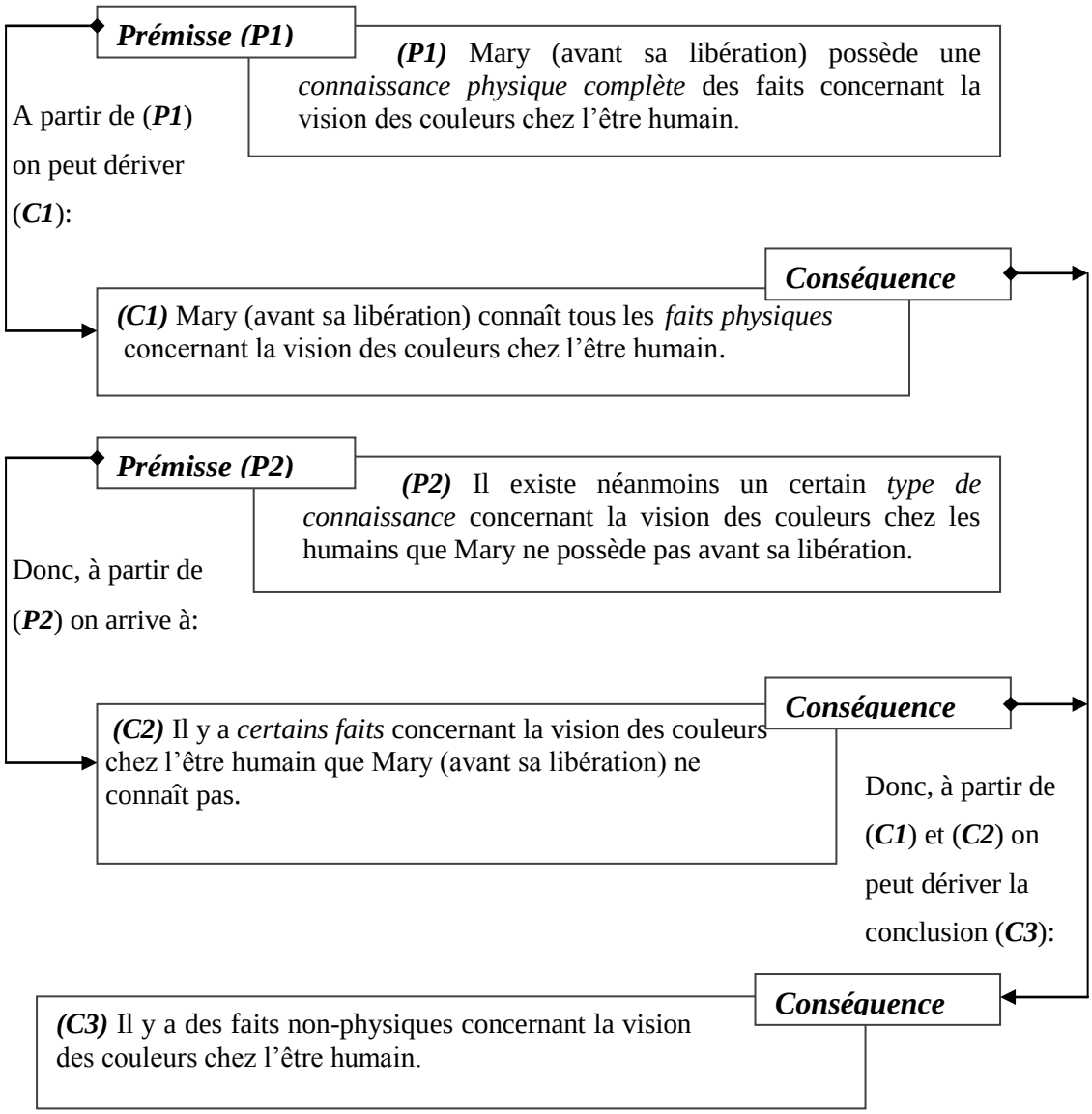
Prenons le cas de Filip qui se trouve au temps *t* à Amsterdam, mais qui *ne sait pas* qu'il se trouve à *t* à Amsterdam. (Disons qu'il pense, à tort, qu'il se trouve à Venise). Ce qui manque à Filip c'est un fragment de connaissance concernant son emplacement spatial. Or, ce manque n'entraîne pas automatiquement la conclusion "Filip ne sait pas que Filip se trouve à Amsterdam ». Il se peut en effet que Filip sache que « Filip se trouve à Amsterdam », mais que par un malheureux concours de circonstances, notre héros ait oublié qu'il *est* Filip. Lorsque Filip apprend enfin qu'il *est* Filip, *il n'apprend pas par là un fait nouveau*, mais seulement il gagne une perspective nouvelle sur un fait qu'il connaissait déjà auparavant.

Si l'on applique cette analogie au cas particulier de la perception des couleurs, on note que certains faits physiques liés à la perception chromatique peuvent nous parvenir par deux voies distinctes : (i) sous une forme physique (utilisant donc des concepts physiques) ou (ii) sous une forme non-physique (utilisant des concepts non-physiques). Tenant compte de cette distinction, il est possible d'acquérir une connaissance nouvelle (non-physique) d'un fait physique, sans acquérir pour autant la connaissance d'un *fait nouveau* (physique). Par conséquent, beaucoup de critiques sont prêts à accepter les conclusions qui découlent de la version faible du KA : ils acceptent donc l'idée que Mary gagne une connaissance propositionnelle génuine, mais ils rejettent l'idée qu'elle parvient à connaître des faits nouveaux (qu'elle ne connaissait pas auparavant). Nous reviendrons sur ce point lorsque nous passerons en revue les différentes réponses que l'argument a suscitées de la part de ses critiques. Le point essentiel à retenir pour l'instant concerne les ambitions manifestement ontologiques de l'expérience de pensée de Jackson : son but déclaré est de réfuter le physicalisme.

5.2.3 Un dispositif tripartite

Afin de mieux saisir les enjeux de l'affaire de Mary nous allons examiner à présent, à la suite de M. Nida-Rümelin[28], la structure détaillée de l'argument dans sa forme ontologique :

(V3) Formulation explicite de l'argument de la connaissance (version forte)



Comme le relève fort bien Nida-Rümelin, une fois que l'on accepte (C1) et (C2), il n'est plus possible d'éviter (C3). En outre, il semble difficile de rejeter l'idée selon

[28] M. Nida-Rümellin, '*Qualia*: The Knowledge Argument' in http://plato.stanford.edu/entries/*qualia*-knowledge. Voir aussi M. Nida-Rümelin, (1996), 'What Mary couldn't know', *in* Th. Metzinger (ed.), *Phenomenal Consciousness*, Schoenigh: Paderborn.

laquelle il serait possible, *en principe*, d'avoir une connaissance physique *complète* à propos de la vision des couleurs chez les humains. Par conséquent, (P1) semble acceptable et soutient solidement l'expérience de pensée de Jackson.

De manière générale, les matérialistes ont à leur disposition plusieurs stratégies pour éviter la conclusion néfaste (C3). Ils peuvent :

a) soit objecter contre l'inférence allant de (P1) à (C1)[29] ;

b) soit éviter la conclusion (C2) en rejetant la prémisse (P2) : ce sera la stratégie qu'adopteront les tenants de l'hypothèse de l'habilité, parfois appelée la stratégie « *savoir que versus savoir faire* ») ;

c) soit bloquer l'inférence allant de la prémisse (P2) à la conclusion (C2) : ce sera la stratégie choisie par une grande majorité de physicalistes, connue également sous le nom de « stratégie des concepts phénoménaux ».

Selon la synthèse qu'en propose T. Alter[30], l'intuition initiale du KA se prête à une structuration suivant trois axes distincts :

(M)1 ***La thèse de la connaissance complète*** : avant de quitter sa prison, Mary connaît *toute l'information physique* relative à la perception visuelle des couleurs.

(M)2 ***La thèse de l'apprentissage*** : lorsqu'elle quitte sa prison, Mary *apprend* quelque chose de nouveau.

(M)3 ***La thèse de la non-déductibilité*** : si les thèses (M1) et (M2) sont vraies, alors ce que Mary apprend lorsqu'elle quitte la prison *ne* peut *pas* être déduit de manière *a priori* (à l'aide des raisonnements seulement, sans apport empirique) de l'ensemble complet de vérités physiques.

---

[29] Il s'agit d'une stratégie peu adoptée parmi les philosophes contemporains. La plupart d'entre eux considèrent le passage depuis (P1) vers (C1) comme non problématique : si tous les faits physiques peuvent être connus sous une conceptualisation physique déterminée, il s'ensuit qu'une personne qui a une connaissance physique complète d'un sujet déterminé, il connaît en même temps tous les faits physiques pertinents relatifs au sujet en question. Mais il y a certaines voix qui se sont levées contre ce passage implicite de (P1) vers (C1). O. Flanagan par exemple, distingue entre le physicalisme *métaphysique* qui vise la thèse ontologique selon laquelle il n'y a pas d'individuels, propriétés ou relations non-physiques ; et le physicalisme *linguistique* selon lequel toute propriété, relation physique doit pouvoir être exprimée ou capturée par le langage des sciences physiques. Selon Flanagan, le KA constitue une menace uniquement pour ce dernier type de physicalisme, mais non pas pour le premier. A cela on peut répondre qu'il devient difficile alors de comprendre ce que veut dire, pour une propriété, d'être une propriété physique, si l'on abandonne l'idée selon laquelle les propriétés physiques sont celles exprimables en termes physiques. Voir O. Flanagan, (1992), *Consciousness Reconsidered*, Cambridge: M.I.T. Press.

[30] T. Alter, (2007), 'Does Representationalism Undermine the Knowledge Argument?' in T. Alter & S. Walter (eds.) op. cit.

A la lumière de ces considérations, on saisit à présent toute la force du KA. Car si les physicalistes rejettent l'intuition de départ, ils doivent assumer, par ricochet, le rejet des thèses (M1) à (M3), ce qui n'était pas le cas ni avec l'archange de Broad, ni avec la chauve-souris de Nagel.

On l'a souvent répété : l'attrait du KA tient en grande partie à son exceptionnelle limpidité. La puissance de l'intuition qui en constitue le cœur, ainsi que le défi qu'il lance aux partisans du physicalisme n'ont pas tardé à susciter de nombreux échanges, réactions critiques et remaniements. Dans ce qui suit, nous allons brièvement passer en revue les plus influentes objections, avant de retenir pour une analyse plus détaillée la prémisse qui représente à notre sens, à la fois le point fort et le point faible du KA, à savoir la thèse de l'omniscience de Mary durant sa captivité.

5.3 L'architecture de l'argument en sept questions

Récemment, R. Van Gulick[31] a proposé une analyse synthétique organisant le dossier "Mary" en sept volets distincts, agencés en fonction des possibles réponses qui gravitent autour des questions suivantes :

(Q)1 Est-ce que Mary apprend un fait *nouveau* ?

(Q)2 Est-ce que cette nouvelle acquisition est d'ordre *factuel* ou *non-factuel* ?

(Q)3 Est-ce qu'elle apprend seulement une connaissance pratique de type « *savoir faire* » ?

(Q)4 Acquiert-elle une simple connaissance par *accointance* ?

(Q)5 Dans l'hypothèse où elle acquiert effectivement un savoir factuel, s'agit-il d'un fait *radicalement nouveau*, inédit, ou d'un fait ancien déjà connu, déguisé seulement avec des habits nouveaux ?

(Q)6 Mary, dans sa prison, est-elle véritablement en possession d'un savoir *complet* et *exhaustif* de *tous* les faits physiques ?

(Q)7 Est-ce que le physicalisme est faux ?

L'examen détaillé de ces points sera organisé de la façon suivante : dans un premier temps nous allons passer en revue d'une manière très schématique les trames

[31] R. Van Gulick, (2004), 'So Many Ways of Saying No to Mary", in Ludlow et al., op. cit., pp. 365-404.

argumentatives des différentes stratégies qui ont été étayées autour de la thèse de l'apprentissage (M2), i.e. les questions allant de (Q1) à (Q5), pour s'arrêter ensuite plus amplement sur la thèse de la connaissance complète (M1), i.e. sur les questions allant de (Q6) à (Q7), et qui constituent, à nos yeux, le noyau dur et le point culminant dans le duel opposant les physicalistes à leurs détracteurs. La discussion de ces derniers points s'articulera dans le chapitre 6. autour des notions centrales d'« omniscience », de « propriétés fondamentales » et de « physicalisme *a priori* ». Mais avant d'aborder ce versant manifestement métaphysique des implications du KA, il convient d'établir un bref état des lieux des principales réponses que l'affaire "Mary" a suscitées de la part des philosophes physicalistes sur le versant (M2), i.e., la thèse de l'apprentissage.

5.4 Connaissance non-propositionnelle : « savoir que » vs « savoir faire »

Tout le monde s'accorde pour dire qu'à sa sortie de prison il se passe *quelque chose* avec Mary. Hélas, d'importants désaccords commencent à se manifester dès qu'il s'agit de préciser la nature de ce « quelque chose ». S'agit-il de l'apprentissage d'une connaissance de type propositionnel, c'est-à-dire se prêtant à une vérification et à une analyse en termes de conditions de vérité, ou s'agit-il plutôt de l'acquisition d'une simple *habilité*, d'une *aptitude*, d'un « *savoir-faire* », qui se grefferait en quelque sorte sur le noyau dur des connaissances propositionnelles déjà acquises qui, elles, restent inchangées ? Dans ce qui suit, nous allons esquisser les principaux arguments qui ont été avancés en faveur de la thèse selon laquelle le savoir que Mary acquiert à sa sortie de prison ne rentre pas dans la catégorie « connaissance propositionnelle authentique », mais plutôt dans une catégorie cognitive beaucoup plus modeste, à savoir l'acquisition de compétences de type « savoir faire ».

5.4.1 L'hypothèse de l'habilité

Défendue initialement par L. Nemirow[32] et relayée par D. Lewis[33], *l'hypothèse de l'habilité* affirme qu'à sa sortie de prison, Mary acquiert un ensemble d'*aptitudes*[34], à savoir l'habilité d'imaginer, de mémoriser, de reconnaître les couleurs, etc., et dont la maîtrise tient plutôt du domaine de la pratique, ou du maniement de certaines informations déjà existantes dans le système cognitif initial. Il est vrai, concède Lewis[35], qu'il existe un certain type de connaissance – la connaissance *de se* - dont la nature demeure irréductiblement subjective, c'est-à-dire capturée du point de vue de la première personne. Mais l'existence de ce type d'accès, insiste Lewis, ne remet pas en cause le monde extérieur, car il s'agit seulement d'un *mode* sur lequel le sujet connaissant se présente à lui-même. A ce titre, ce type de connaissance ne mérite pas l'appellation de « savoir que », mais seulement celle, beaucoup plus modeste, de « savoir faire »[36].

Pour appuyer son point de vue, Lewis s'inscrit en faux contre l'hypothèse rivale, qu'il baptise 'l'Hypothèse de l'Information Phénoménale'[37] (HIP). Selon cette dernière, il existerait, en plus des informations physiques qui nous parviennent du monde extérieur, un deuxième domaine distinct, autonome, et irréductible au physique, à savoir le domaine de l'expérience phénoménale. Or, selon Lewis, une analyse attentive de la HIP devrait nous convaincre que le postulat stipulant l'existence d'une couche supplémentaire de faits, non seulement nous complique inutilement la tâche, mais surtout s'oppose à une conception scientifique du monde. Il convient, en conséquence, de rejeter la HIP et de s'en tenir à sa rivale, « l'hypothèse de l'habilité »[38]. De son côté,

---

[32] L. Nemirow, (1990), 'Physicalism and the Cognitive Role of Acquaintance', *in W.* Lycan (ed.) *Mind and Cognition: A Reader*, Oxford: Blackwell. Voir aussi L. Nemirow, (2007), 'So This Is What it's Like: a Defense of the Ability Hypothesis', *in* T. Alter & S. Walter (eds.), op. cit., pp. 32–51.

[33] D. Lewis, (1988), 'What Experience Teaches', *Proceedings of the Russellian Society*, University of Sidney. Repris *in* P. Ludlow et al. op. cit. (2004) (Nous utilisons ici cette dernière édition).

[34] Cet ensemble de propriétés porte d'ailleurs, dans la littérature analytique, le nom de « Lewis's Abilities » (« habilités Lewisiennes »).

[35] (D. Lewis, 1988/2004 : 82)

[36] "Knowing what an experience is like just *is* the possession of these abilities to remember, imagine, and recognize. … *It isn't knowing-that. It's knowing-how*". (D. Lewis, 1990: 516) (Nous soulignons)

[37] (D. Lewis, 1988/2004: 84) Désormais (HIP) en abrégé.

[38] Une autre version de ce type de réponse a été récemment défendue par P. Pettit qui propose de considérer le cas d'Eva, une « cousine » de Mary, si l'on peut appeler ainsi. Eva subit une mésaventure analogue à celle de Mary, sauf que, dans son cas, il ne s'agit pas d'une privation de la vision des couleurs, mais de celle du *mouvement*. Selon Pettit, l'analyse du cas « Eva » rend beaucoup plus explicite le fait que Mary acquiert non pas une autre connaissance, mais plutôt un nouveau *mode de connaître*,

Nemirow affirme que connaître l'effet-que-cela-fait de voir un rubis, disons, est un savoir analogue à celui dont on dispose lorsqu'on *imagine* cette même expérience.

L'hypothèse de l'aptitude a fait l'objet de plusieurs critiques, dont les plus connues portent le nom d'« objections des trois L », du nom de leurs défenseurs : J. Levin, B. Loar, et W. Lycan.

5.4.1.1 L'objection de l'inférence des faits nouveaux

J. Levin[39] note que lorsqu'on regarde pour la première fois un objet d'une couleur non familière, on obtient des informations importantes concernant les ressemblances et les compatibilités avec d'autres couleurs, ainsi que des informations sur les effets que la nouvelle couleur provoque en nous. On obtient ainsi des informations nouvelles sur nos propres états mentaux. Nous avons donc de fortes raisons d'*inférer* que lorsqu'on nous montre pour la première fois une couleur non familière, on acquiert par là même des *faits* authentiquement nouveaux concernant non seulement la nouvelle couleur, mais aussi l'expérience que l'on en éprouve. A cela Nemirow répond, sans surprise, que l'acquisition de certaines habilités *favorise* l'acquisition d'une connaissance de type propositionnel, mais que l'habilité en elle même *n'est pas* un fait authentiquement nouveau. Par exemple, si je suis un bon danseur de valse je peux, à partir d'une certaine séquence de musique, prédire de combien de pas j'avancerai vers la gauche. Mais cela ne veut pas dire que la virtuosité en matière de valse est autre chose qu'un simple savoir-faire[40].

---

autrement dit, une aptitude pratique plutôt qu'un savoir théorique : « What happens when she leaves the room is not that she gains new knowledge but that she changes in some other knowledge-relates way: in a Broad sense she develops a new mode of knowing. (…) Eva comes to believe or know in what I describe as a practical, nonintellectual mode facts that she not previously believed or knew only in a purely intellectual way." Voir P. Pettit, (2004), 'Motion Blindness and the Knowledge Argument', *in* Ludlow et al., op. cit., pp. 106-7.

[39] J. Levin, (1986), 'Could Love Be Like a Heat Wave? Physicalism and the Subjective Character of Experience', *Philosophical Studies,* 49: 245–261. Voir aussi J. Levin, (2007), 'What is a phenomenal concept?', *in* T. Alter & S. Walter, op. cit., pp. 87–110.

[40] "If I have expertise in dancing the waltz, I can predict the steps (or possible steps) within the dance from the beginning of a dance sequence; this does not mean that my dancing expertise is other than know-how." Voir L. Nemirow, (2007), 'So This is What is Like: a Defense of the Ability Hypothesis' in T. Alter & S. Walter, op. cit., p. 38.

5.4.1.2 L'objection des « conditionnels enchâssés »[41]

Cette deuxième objection, liée habituellement au nom de B. Loar[42] conteste la légitimité de la stratégie « savoir-faire », en signalant que l'hypothèse de l'habilité est incapable d'expliquer comment une expérience particulière, par exemple l'expérience accompagnant l'effet-que-cela-fait de boire, disons, du lait de coco, peut être emboîtée dans un énoncé conditionnel de type (C) : « Si le lait de coco n'avait pas eu ce goût, alors *Q* ». A cela Nemirow répond en rappelant que l'hypothèse de l'habilité propose de remplacer la formule « savoir que l'effet-que-cela-fait de goûter du lait de coco ressemble à *ça* » par une formule de type « être capable de reconnaître, se rappeler, et imaginer *cette* expérience comme étant celle du goût du lait de coco ».

Dans sa réplique à ce type d'objection, B. Loar montre de façon très convaincante que l'emboîtement d'une occurrence ayant la forme « cela fait tel-et-tel-effet » dans une proposition de type (C) : « *si* la douleur fait tel-et-tel effet, alors *Q* », peut être expliqué seulement si l'on accepte l'idée que la connaissance que Mary acquiert à sa sortie de prison *est* véritablement d'ordre propositionnel, et non pas seulement d'ordre pratique. Dans la vie de tous les jours, chaque individu est capable de discerner, *via* l'introspection, l'effet-que-cela-fait de ressentir telle ou telle expérience spécifique. Ces *qualités* phénoménales sont théoriquement représentées par ce qu'on appelle des « *concepts phénoménaux* » dont la principale particularité tient au fait qu'ils sont toujours formés depuis le point de vue de la première personne. Selon Loar, ces concepts sont « type-démonstratifs »[43] dans le sens où ils dérivent leur référence en prenant comme point d'appui la première personne. Par exemple : « *ce* type de sensation » ou « *cet* aspect de l'expérience visuelle ».

En s'appuyant sur la distinction « concepts phénoménaux »/« propriétés phénoménales », Loar argumente en faveur de l'autonomie des concepts phénoménaux en insistant sur les points suivants : (i) en premier lieu, un sujet peut entretenir des pensées non seulement de type (P1) « Les noix de coco ont *ce* goût », mais aussi de la forme (P2) « Si les noix de coco n'avaient pas eu *ce* goût, alors *Q* ». On peut concéder, à la rigueur, que (P1) exprime un savoir-faire, mais il est clair que dans (P2)

[41] Littéralement: "the objection from embedded conditionals"
[42] B. Loar, (1990), 'Phenomenal states', in *Philosophical Perspectives*, 4: 81-108. Repris (en version révisée) in Ludlow et al., (2004), op. cit. (Nous faisons référence ici à cette dernière édition).
[43] (B. Loar, 1990/2004: 219)

l'occurrence « les noix de coco ont *ce* goût » joue le rôle d'un prédicat doté d'un contenu déterminé. (ii) Deuxièmement, étant donné qu'il nous est possible d'émettre des pensées concernant les états phénoménaux des autres sujets, il s'ensuit que ceux-ci possèdent indéniablement un contenu prédicatif distinct. (iii) Enfin, beaucoup de qualités phénoménales ne sont pas traduisibles à l'aide d'une description de type « savoir-faire », pour la simple bonne raison qu'*il n'y a pas* de candidat convenable pour occuper ce rôle. Loar[44] donne comme exemple la sensation particulière qu'il ressent dans son genou gauche lorsqu'il fait son jogging de matin, et au sentir qualitatif duquel il est difficile d'appliquer une recognition conceptuelle précise et établie d'un savoir-faire publiquement accessible.

5.4.1.3 L'objection du contenu mental

Certains auteurs ont attaqué l'hypothèse de l'habilité en soulevant l'objection suivante : la lisière qui est censée séparer la connaissance authentique de type « savoir que » de sa cousine plus modeste « savoir faire » est beaucoup plus floue que ses défenseurs le laissent entendre, et par ricochet, beaucoup plus inefficace. Pour *W.* Lycan[45], une connaissance de type « connaître *ce que*… » est intimement apparentée à une connaissance de type « savoir *que* ». Par exemple, la proposition « *Je sais ce que* Gilles fait » est vraie en vertu du fait que moi, Anna, *sait que* Gilles est dans un tel endroit, occupé à faire telle ou telle chose. De façon analogue, Lycan propose d'interpréter une proposition de type "*S sait ce que* c'est de voir du bleu" par une expression de type "*S* sait que voir du bleu c'est comme éprouver *Q* »[46] où *Q* représente le « *quale* » correspondant à la perception visuelle du bleu. Plus spécifiquement, ce que Lycan reproche aux adeptes de l'hypothèse de l'habilité c'est l'incapacité d'expliquer un aspect essentiel concernant le fonctionnement de l'imagination : clairement, il existe une manière de s'imaginer ou de représenter *correctement* ou *incorrectement* de « l'effet-que-cela-fait » de voir du bleu, et cela prouve, dit Lycan[47], que l'acquisition de ce type de savoir concerne des vérités propositionnelles, et non pas une simple aptitude d'ordre pratique.

---

[44] *Ibid.*, pp. 236-37.
[45] W. Lycan, (1996), *Consciousness and Experience*. Cambridge, MA: MIT Press.
[46] S knows that it is like *Q* to see blue".
[47] "There is such a thing as getting 'what it's like' right, representing truly rather than falsely, from which it seems to follow that knowing 'what it's like' is knowing a truth". (W. Lycan, 1996: 99)

Rappelons également que l'*hypothèse de l'imagination* avancée par Nemirow selon laquelle connaître l'effet-que-cela-fait de voir du rouge est un savoir analogue à celui dont on dispose lorsqu'on *imagine* cette même expérience, a suscité à son tour plusieurs critiques. Plus spécifiquement, on a signalé que la simple faculté d'*imaginer* une expérience en particulier ne constitue ni une condition nécessaire, ni suffisante pour *connaître* l'effet-que-cela-fait de subir l'expérience en question. Dans un article où il examine de près la notion de « connaissance phénoménale », E. Conee[48] propose d'effectuer une légère modification à l'argument initial de Jackson. Supposons que Mary est dépourvue de toute capacité d'imaginer quoi que ce soit. Ce manque, observe Conee, ne l'empêchera nullement d'avoir une réaction de surprise, et de s'exclamer « Ah ! C'est donc *cela* que de voir du rouge ! » lorsqu'elle quitte enfin sa prison. Le fait qu'elle apprend *quelque chose* lorsqu'elle sort de sa captivité, malgré sa déficience innée qui l'empêche d'imaginer un fait *avant* de l'éprouver, indique, selon Conee, que la mémoire et l'imagination ne sont pas nécessaires dans ce cas précis.

Pour étayer sa thèse, Conee propose d'opérer une légère modification dans l'expérience de pensée de Jackson: prenons le cas de Martha, une personne qui est très douée pour détecter les nuances intermédiaires entre des paires de couleurs données, et qui, par conséquent, est familiarisée avec toutes les nuances *sauf*, disons, la nuance de rouge qu'on appelle familièrement « rouge cerise ». A présent, supposons que nous informons Martha que le « rouge cerise » est une nuance qui se trouve au niveau intermédiaire entre le « rouge bordeaux » et le « rouge vif ». Armée de cette information, et grâce à son extraordinaire capacité d'imaginer et de détecter les nuances, Martha possède à présent l'*aptitude* d'imaginer le « rouge cerise ». Mais tant qu'elle ne se trouvera pas dans la position d'*exercer* cette aptitude, elle ne saura jamais l'effet-que-cela-fait de voir le « rouge cerise ».

5.4.1.4 L'approche Chomsky-Alter

Les tenants de l'hypothèse de l'habilité ont à affronter, entre autres, un défi lancé par N. Chomsky[49]qui argue, de façon très convaincante, que le savoir comment parler ou comprendre un langage, ne peut pas être réduit à un simple ensemble d'habilités. Il

[48] E. Conee, (1994), 'Phenomenal Knowledge', *Australasian Journal of Philosophy* 72: 136–150. Réédité *in* Ludlow et al., op. cit. (Nous utilisons ici cette dernière édition).
[49] N. Chomsky, (1988), *Language and Problems of Knowledge: The Managua Lectures.* Cambridge: MIT Press.

propose à son tour une autre expérience de pensée : supposons que Juan, un natif parlant l'espagnol, souffre d'une sévère aphasie après avoir subi une contusion à la tête entraînant une perte totale de la capacité de parler et/ou comprendre l'espagnol. Qu'a perdu Juan ? Chomsky maintient que l'infortuné Juan n'a pas perdu sa *connaissance* de l'espagnol, puisqu'il pourrait récupérer son aptitude à parler et à comprendre l'espagnol au fur et à mesure que les effets de sa blessure s'estompent. Malgré le choc qui a entraîné la perte de ses aptitudes à parler et à comprendre une langue, *quelque chose d'autre* perdure, à savoir un système de connaissances, un système cognitif niché dans l'esprit/cerveau. [50]

L'exemple de Juan est analogue au cas évoqué par T. Alter[51] qui introduit une distinction auxiliaire entre « *savoir comment* imaginer le fait d'avoir une expérience » et « l'*aptitude* à imaginer la même expérience ». Imaginons le cas de Hank, un individu qui, tout comme Juan, reçoit un sévère coup à la tête, suite auquel il perd l'habileté d'imaginer à quoi ressemble l'expérience de goûter une glace au chocolat. Au fur et à mesure que le cerveau de Hank récupère ses connexions, le patient retrouve son habilité perdue. Alter conclue en affirmant que malgré la perte temporaire d'une aptitude particulière, le cerveau de Hank conserve quelque chose de constant, plus spécifiquement, *le savoir-l'effet-que-cela-fait* de goûter une glace au chocolat, et ce savoir, souligne Alter[52], ne peut pas être réduit à un ensemble d'habilités.

Nemirow[53] répond aux reproches de Chomsky/Alter et rejette l'idée selon laquelle Juan/Hank auraient complètement ou effectivement perdu l'habilité de parler espagnol, et respectivement, imaginer le goût de la glace au chocolat, uniquement parce qu'ils sont dans l'incapacité *immédiate* de mettre en œuvre leur habilité. Il est possible, après tout, de traiter la capacité de parler l'espagnol, par exemple, non pas comme une habilité unique, mais plutôt comme un *faisceau d'habilités*, dont certaines sont perdues suite au choc, d'autres conservées dans le cerveau d'une manière ou d'une autre. L'existence de ces multiples facettes expliquerait le fait que Juan/Hank perdent non pas l'*habilité* de parler l'espagnol ou d'imaginer le goût de la glace au chocolat, mais plutôt l'*habilité d'exercer l'habilité* de parler l'espagnol ou d'imaginer le goût de la glace au chocolat.

---

[50] "Plainly, something was retained while the ability to speak and to understand was lost. What was retained was a system of knowledge, a *cognitive* system of the mind/brain." (N. Chomsky, 1988 : 10)
[51] T. Alter, (2001), 'Know-How, Ability, and the Ability Hypothesis', *Theoria*, 67: 229–39.
[52] (T. Alter, 2001: 234)
[53] (L. Nemirow, 2007)

Certains auteurs, comme M. Tye[54], explorent la possibilité d'assouplir l'hypothèse de l'aptitude en lui rajoutant une clause supplémentaire. Supposons, dit Tye[55], que connaître l'effet-que-cela-fait d'éprouver une expérience devrait être identifié non pas avec un faisceau d'aptitudes, comme le soutiennent Lewis et Nemirow, mais avec une aptitude plus *primitive*, à savoir *l'aptitude d'appliquer un concept indexical*, *via* l'introspection, au caractère phénoménal de sa propre expérience. Cependant, l'ajout de cette clause additionnelle ne signera pas la fin des soucis pour les adeptes de l'hypothèse de l'habilité. Pour voir cela, observe Tye, il suffit d'imaginer la situation suivante : supposons que notre héroïne Mary, au moment même où elle quitte la pièce, est distraite, bref, elle a la tête ailleurs, et ne se concentre pas sur l'expérience nouvelle qu'elle vient d'avoir. Elle se trouve donc dans l'*incapacité* d'appliquer indexicalement le *concept* de « rouge » pour définir son expérience. Dans ce cas, même si elle possède, en principe, la faculté (l'habilité) d'appliquer un concept indexical à son expérience courante, elle ne l'exerce pas (étant distraite).

Selon Tye, cette version révisée et assouplie de l'hypothèse de l'habilité n'est pas suffisamment puissante pour récuser l'argument de la connaissance. Pour voir cela, il suffit de se rappeler que lorsque Mary quitte sa prison, et qu'elle *n*'est *pas* distraite, elle acquiert une connaissance nouvelle : plus spécifiquement, elle apprend que *cela* est une expérience du rouge (où « cela » fait référence directe à son expérience en cours, immédiate, captée *via* l'introspection).

Selon Tye[56], cette nouvelle connaissance est une connaissance propositionnelle authentique de type « savoir que ». Lorsque Mary sort de sa prison, la tête ailleurs, elle *sait* comment diriger mentalement, *via* l'introspection, son attention vers le caractère phénoménal de *son* expérience, mais étant distraite, elle ne le fait pas, autrement dit, elle n'exerce pas et n'applique pas son savoir-faire. Mais lorsqu'elle le fait enfin, son *savoir faire* se transforme en *savoir que*, même si ce « savoir que » est introspectif ou obtenu par le truchement de l'introspection. En conclusion, Tye propose une définition

[54] M. Tye, (2000), 'Knowing What it is Like: The Ability Hypothesis and the Knowledge Argument', in M. Tye, *Consciousness, Color, and Content*, Cambridge, MA: MIT Press. Repris in Ludlow et al., (2004), op. cit. Nous utilisons ici cette dernière édition.

[55] "Perhaps knowing what it is like should be identified not with the cluster of abilities Lewis cites – for they may all be lacking while knowing what it is like is present – but rather with the more basic ability *to apply an indexical concept to the phenomenal character of her via introspection*." (M. Tye, 2000/2004: 153)

[56] *Ibid.*, p. 155.

de la connaissance de type phénoménal tenant compte de la disjonction entre le « savoir que » introspectif et le « savoir faire », qu'il baptise la *thèse disjonctive de la connaissance phénoménale* :

Connaître-l'effet-que-cela-fait$_{\text{def.}}$ = S *connaît* l'effet-que-cela-fait d'éprouver l'expérience *E* si et seulement s'il remplit les deux conditions suivantes : (i) soit S est en train d'éprouver *E*, et S possède un *savoir-que* du caractère phénoménal de *E* obtenu *via* son introspection en cours ; (ii) soit S possède l'ensemble d'« aptitudes Lewisiennes » afférentes à *E*[57].

Bien que fort intéressante, la discussion[58] détaillée de ces points prolongerait inutilement les modestes objectifs de cette section dont la fonction principale reste celle d'établir un état des lieux des réactions critiques les plus influentes provoquées par les mésaventures de Mary. Signalons néanmoins que certains auteurs, comme Tye, utilisent dans leur défense du physicalisme, comme l'observe fort bien Nida-Rümelin[59], une combinaison de *deux* stratégies pour contrecarrer les effets du KA : Tye applique, d'une part, la stratégie de la *Connaissance Non-Propositionnelle* que nous avons évoquée jusqu'ici pour traiter le cas d'une expérience phénoménale décrite par le deuxième terme disjoint (ii) de la définition susmentionnée. Mais il emploie une stratégie distincte, connue sous l'étiquette de *« Stratégie des Concepts Phénoménaux »*[60] lorsqu'il traite le terme (i) de la disjonction. Nous aurons l'occasion de revenir plus amplement sur cet autre type de stratégie une section plus loin (5.5.). Nous allons examiner à présent une autre variation sur le thème de la connaissance non-propositionnelle, à savoir l'hypothèse de la connaissance par accointance.

5.4.2 L'hypothèse de la connaissance par accointance

Nous avons vu, dans la section précédente, que pour certains défenseurs du physicalisme, le « nouveau » savoir de Mary n'est pas un « savoir que » mais plutôt un

[57] "Knowing what it is like"$_{\text{df}}$ = « S knows what it is like to undergo experience *E* = df Either S is now undergoing *E*, and S has knowledge-that with respect to the phenomenal character of *E* obtained *via* current introspection, or S has the Lewis abilities with respect to *E*." (M. Tye, 2000/2004:155)

[58] Pour une critique très convaincante de l'hypothèse de l'aptitude voir également J. Stanley et T. Williamson, (2001), 'Knowing how', *Journal of Philosophy* 98: 1–40.

[59] M. Nida-Rümellin, '*Qualia* : The Knowledge Argument' in http://plato.stanford.edu/entries/*qualia*-knowledge

[60] Stratégie connue également sous le label « the *New Knowledge/Old Fact*-strategy »

« savoir faire ». Certaines voix se sont levées pour défendre une troisième option, une voie du milieu, pour ainsi dire, située entre le « savoir que » et le « savoir faire ». Selon cette conception, défendue entre autres par P. Churchland[61] et E. Conee[62], Mary acquiert à sa sortie de prison, un type de connaissance qui n'est *ni* réductible au savoir de type propositionnel, *ni* assimilable à un simple « savoir-faire ». Plus spécifiquement, Mary acquiert ce que l'on peut appeler *« une connaissance par accointance »*. Selon la définition qu'en donne Conee, pour qu'un sujet obtienne une connaissance par accointance, il est nécessaire qu'il entre en *contact direct* avec l'objet en question. Etant donné que Mary est en position de se familiariser avec ce type de relation à un objet rouge uniquement *après* sa libération, il s'ensuit que la prisonnière gagne une connaissance par accointance avec les *qualia* accompagnant l'expérience qualitative du rouge uniquement *après* sa libération. L'insertion de cette « troisième » voie permet à Conee de répondre à Jackson de la manière suivante : admettons que les *qualia* soient des propriétés *physiques* de l'expérience (tandis que l'expérience est elle-même un processus physique). Appelons ces propriétés, pour plus de simplicité, *Q*. Selon Conee, même si Mary reste confinée à sa prison, elle *possède* néanmoins *une connaissance complète* et *exhaustive de Q*, elle sait même que l'expérience « voir un rubis » provoque un certain effet phénoménal *Q*. Lorsqu'elle regagne enfin l'espace extérieur, Mary se familiarise avec *Q*, et obtient une connaissance directe, par accointance. Mais ce type de « familiarisation » n'ajoute aucune information nouvelle et authentique au savoir propositionnel qu'elle possédait déjà dans sa prison.

Une version récente de l'hypothèse de la connaissance par accointance a été examinée par M. Tye[63].Selon Tye, une personne peut connaître un fait non-conceptuellement, par accointance, i.e. sans qu'elle le subsume à aucun concept, même pas à un simple concept démonstratif de type « c'est *cela* » ou « c'est ceci » de voir du rouge. Pour reprendre le cas de la perception, le fait que l'expérience vécue par le sujet favorise et rend possible l'élaboration d'une réponse conceptuelle et propositionnelle, constitue seulement une condition suffisante pour ce type de connaissance par accointance directe. Autrement dit, il n'est pas nécessaire qu'une réponse conceptuelle

[61] P. M. Churchland (1985), 'Reduction, *Qualia* and the Direct Introspection of Brain States', *Journal of Philosophy* 82 (January):8-28.
[62] E. Conee, (1994), 'Phenomenal Knowledge', *Australasian Journal of Philosophy* 72: 136–150. Repris in Ludlow, et al., (2004), op. cit. (Nous utilisons ici cette dernière édition).
[63] M. Tye, (2009), *Consciousness Revisited-Materialism without Phenomenal Concepts:* Cambridge, MA: MIT Press.

accompagne l'immédiateté de l'expérience directe, car l'état qui est en jeu ici est un *état nonconceptuel*.

On pourrait reprocher également à l'hypothèse de la connaissance par accointance de faire reposer ses argumentations sur l'utilisation d'un concept ambigu. Qu'est-ce que l'accointance[64], après tout ? Selon l'acception la plus répandue, une personne *P* est accointée avec un objet *X* si et seulement si *P* possède l'aptitude d'avoir des états mentaux *de re*[65] à propos de *X*. Pour Tye, une personne peut être accointée avec un objet/état/processus, sans avoir obligatoirement une connaissance propositionnelle des vérités relatives à l'objet/état/processus en question. Mais il y a un seuil minimal qui doit être respecté pour que le niveau de la connaissance par accointance soit atteint. Car si l'objet/état/processus visé dans l'expérience est perçu dans une lumière très faible, par exemple, ou dénaturé par un miroir déformant, peut-on continuer à dire qu'il y a connaissance par accointance de l'objet/état/processus respectif ? Il semble que non, observe Tye. La véritable connaissance par accointance repose sur le déroulement d'une expérience en conditions normales, non-dénaturées. Nous rencontrons la même exigence dans le cas d'accointance non-visuelle. Tye[66] illustre son point de vue à l'aide de l'exemple suivant. Supposons, dit-il, que je suis accointé avec la statue de David exposée dans un musée à Florence, mais je ne réalise pas que j'*ai* cette connaissance tacite. Mieux encore, je ne m'aperçois pas que la statue que je regarde est bel et bien celle de David. Puis-je affirmer alors que je *connais* la statue de David ? Selon Tye, il y

---

[64] Traditionnellement, l'accointance désigne le vécu immédiat, l'expérience directe elle-même, la relation intime du sujet à ses propres états internes.

[65] La distinction entre interprétation *de re* et interprétation *de dicto* remonte à la philosophie scolastique. *De re* signifie en latin 'à propos de la chose', et *de dicto* 'à propos de ce qui est dit'. A l'origine, cette distinction a été établie pour rendre compte de l'ambiguïté de propositions dans lesquelles un groupe nominal apparaît dans la portée d'une expression modale. Considérons la proposition (1) :

(1) Anna croit que Gilles est un espion.

On peut interpréter (1) de deux façons différentes :

(2) Il existe un individu qui s'appelle Gilles et Anna croit que cet individu est un espion.

(3) Anna croit que l'individu qu'elle identifie comme Gilles est un espion.

L'énoncé (2) correspond à la lecture *de re* : la modalité porte sur une proposition qui concerne la chose en elle-même (le personnage qui est un espion) indépendamment de son mode de présentation (le nom propre Gilles). Le nom propre n'est pas sous la portée de la modalité, (2) est vérifié même si Anna ne sait pas que l'individu qu'elle croit être un espion s'appelle Gilles. En (3) en revanche, le nom propre est sous la portée de la modalité, la croyance d'Anna est double : elle porte à la fois sur la manière d'identifier l'individu en question (l'individu appelé Gilles) et sur la propriété qui lui est attribuée (être un espion). L'exemple suivant, dû à Quine (1956), montre bien l'importance de la distinction : le pronom indéfini *quelqu'un* est interprété *de re* en (b) et *de dicto* en (c) : (a) Ralph croit que quelqu'un est un espion. (b) Ralph croit qu'il y a des espions. (c) Il y a quelqu'un dont Ralph croit que c'est un espion. Pour une discussion plus détaillée de ces aspects voir T. Burge, (1977), 'Belief De Re', *The Journal of Philosophy*, 74 : 338-362. Voir aussi G. Evans, (1982), *The Varieties of Reference*, Oxford: Oxford University Press.

[66] (M. Tye, 2009:101)

a un sens dans lequel je peux dire que je *connais* la statue de David : je la connais par accointance. Bien sûr, je ne sais pas que je connais la statue, et corrélativement, je ne sais pas le fait que la statue qui se trouve devant moi est celle de David, mais il n'empêche, il y a un sens dans lequel je continue de *connaître* la statue de Michelangelo.

Pour revenir au cas de Mary, les défenseurs de l'argument de Jackson sont prêts à concéder aux partisans de l'hypothèse de la connaissance par accointance l'idée que ce type spécial de connaissance directe joue un rôle essentiel dans la description des propriétés phénoménales *Q*. Mais, insistent-ils, posséder ce type de contact cognitif direct avec *Q* est non seulement une caractéristique importante, mais aussi une *condition nécessaire* du fait que l'expérience possède les propriétés *Q* en question. Investiguer sur la nature de ce type de propriétés constitue un véritable défi pour la métaphysique contemporaine[67].

Dans cette section nous avons passé en revue les stratégies de réponse selon lesquelles Mary acquiert à sa sortie de prison une connaissance de type *non-propositionnel*, autrement dit, un savoir de type pratique. Dans la section suivante, nous allons explorer une autre voie d'approche : en effet, certains défenseurs du physicalisme se déclarent prêts à concéder que le savoir que Mary acquiert *est véritablement de type propositionnel*, à condition – insistent-ils – qu'on admette qu'il ne s'agit pas de l'acquisition d'un fait propositionnel *nouveau*, authentique, mais seulement de la présentation d'un *fait ancien* sous un *visage nouveau*.

5.5 La stratégie des concepts phénoménaux

Pour un physicaliste prêt à concéder que l'état cognitif de Mary subit effectivement une modification, et que, par conséquent, elle *apprend quelque chose* en quittant sa prison, une manière élégante de relever le défi de Jackson serait d'affirmer que lors de sa libération, Mary acquiert la *connaissance nouvelle d'un fait ancien*, c'est-

[67] Comme nous l'avons brièvement mentionné dans le chapitre précédent, deux courants de pensée se disputent actuellement la suprématie dans le paysage de la métaphysique : le monisme matérialiste et le dualisme des propriétés. Nous aurons l'occasion d'y revenir. Pour le but de la présente section, il est important de noter que certains auteurs s'appuient sur le dualisme des propriétés pour signaler que l'existence des *qualia* semble parfaitement compatible avec ce type de dualisme, tandis que le physicalisme échoue dans toute tentative de fournir une description adéquate de l'expérience phénoménale. Voir B. Gertler, (1999), 'A Defense of the Knowledge Argument', *Philosophical Studies*, 93: 317–336.

à-dire d'un fait qu'elle connaissait déjà, se manifestant cette fois-ci sous une apparence inédite[68]. Ce qui veut dire qu'une autre possible solution au problème de Mary serait de rejeter l'idée selon laquelle la captive apprend quelque chose de *nouveau* à sa libération. Selon la synthèse qu'en propose M. Nida-Rümellin[69], il existe plusieurs traits caractéristiques communs sous lesquels on pourrait ranger les différentes réponses appartenant à *la stratégie des concepts phénoménaux*[70]. De manière générale, les adeptes de cette position s'appuient sur une ou une combinaison des thèses suivantes :

(K)1- Le caractère *phénoménal* d'une expérience (par exemple la « rougeur » éclatante d'un rubis) est une propriété *physique* des expériences.

(K)2- Afin d'obtenir la connaissance de l'effet-que-cela-fait d'avoir l'expérience d'une certaine qualité phénoménale, il faut être en possession des *concepts phénoménaux* correspondant à l'attribut phénoménal en cause.

(K)3- La description des processus à travers lesquels un organisme acquiert et possède un concept phénoménal peut être réalisée en termes physiques au sens large.

(K)4- Un sujet connaissant peut acquérir ou posséder des concepts phénoménaux uniquement s'il a réellement vécues ou éprouvées les expériences phénoménales afférentes.

(K)5- Après sa libération, Mary acquiert des connaissances relatives au caractère phénoménal de l'expérience *via* les concepts phénoménaux.

Le point essential à retenir ici est que les faits qui garantissent la véridicité de ces nouvelles données ou « fragments » de savoir *sont* ultimement des faits *physiques* auxquels Mary a été déjà initiée dans sa prison, seulement sous un autre *mode* de conceptualisation. Selon le résumé qu'en dresse Nida-Rümellin, il existe trois pôles autour desquels se tissent les relations et les divergences au sein de cette variante, connue également sous le label « Savoir nouveau/Faits anciens »: (a) la nature phénoménale de l'expérience, (b) les concepts phénoménaux et les qualités

[68]Comme le note M. Nida-Rümellin, cette stratégie attire des nombreux adeptes. Parmi eux: Horgan 1984, Churchland 1985; Tye 1986, 1995; Bigelow & Pargetter 1990; Loar 1990; Lycan 1990, 1995; Pereboom 1994; Perry 2001; Van Gulick 2005; Byrne 2002; Levin 2007; Papineau 2002, 2007. Voir M. Nida-Rümellin, '*Qualia* : The Knowledge Argument' in http://plato.stanford.edu/entries/*qualia*-knowledge

[69] *Ibid.*.

[70] L'appellation de "stratégie des concepts phénoménaux" a été proposée par D. Stoljar, (2005), 'Physicalism and Phenomenal Concepts', *Mind and Language*, Vol.20, No.5, 469-494, et fut reprise, entre autres, par D. Chalmers, (2007), 'Phenomenal Concepts and the Explanatory Gap', *in* T. Alter & S. Walter (eds.), *op. cit.*

phénoménales, (c) la relation entre les qualités phénoménales et les concepts phénoménaux afférents.

5.5.1 Concepts physiques et concepts phénoménaux : M. Tye

Une grande majorité des défenseurs de la stratégie 'Savoir Nouveau/Fait Ancien' s'accordent pour dire que les concepts physiques et les concepts phénoménaux sont cognitivement indépendants, autonomes. Autrement dit, il est impossible d'établir, de façon *a priori*, un rapport nécessaire entre, d'une part, le *concept physique* C_P qui décrit un état/objet/processus donné, et qui correspond à une qualité phénoménale *Q*, et d'autre part le *concept phénoménal* C_Q qui décrit le même état/objet/processus. C'est précisément cette indépendance qui garantit et rend possible le fait que Mary puisse détenir une connaissance physique complète concernant la « rougeur » du rubis (c'est-à-dire disposer d'un savoir complet du caractère phénoménal de la « rougeur » considérée sous une conceptualisation physique), sans pour autant posséder le *concept phénoménal* de « rougeur », et par conséquent, sans connaître aucun de ces faits (qu'elle connaissait déjà sous une description *physique)* sous leur description *phénoménale*[71].

Les défenseurs de cette stratégie dénoncent la présence d'une erreur au cœur même de l'argument de Jackson. Il est vrai, arguent-ils, que la distinction entre le concept physique et phénoménal de « bleuité », par exemple, peut encourager les défenseurs du KA de raisonner *comme si* l'on pourrait imaginer une situation dans laquelle Mary, bien qu'elle possède, *avant* sa libération, une connaissance exhaustive des faits physiques, arrive à acquérir des faits phénoménaux authentiquement nouveaux *après* sa libération. Or, selon les adeptes de l'hypothèse « savoir nouveau/fait ancien», cette situation repose sur une illusion d'optique, ou une erreur de perspective : car ce que Mary apprend à la sortie de sa geôle ce sont des faits *rendus vrais* par des faits qu'elle connaissait *déjà* dans sa captivité.

[71] Certains auteurs ont affirmé que la conceptualisation de type phénoménal n'est purement pas exprimable dans un langage. Voir A. Byrne, (2002), 'Something About Mary', *Grazer Philosophische Studien*, 63: 123-40. Et aussi B. Hellie, (2004), 'Inexpressible Truth and the Allure of the Knowledge Argument', *in* Ludlow et al. (2004), op. cit.

Dans l'article 'Jackson on Physical Information and *Qualia*', T. Horgan [72] défend une version similaire de cette stratégie. Lorsque Mary fait l'expérience du « bleu » – note Horgan - elle ne fait que se *familiariser* avec la propriété *phénoménale* de « bleuité », qui n'est en réalité qu'une propriété *physique* de l'expérience. Autrement dit, en endossant une perspective expérientielle, elle engage « une perspective à la première personne, de type ostensif, sur la propriété phénoménale de « bleuité ». Elle peut dorénavant s'y référer en employant l'expression indexicale « ce type-*là* de propriété », en même temps qu'elle se rappelle, imagine, ou accède à l'expérience du bleu du ciel, disons. Par conséquent, Mary acquiert un *nouveau concept* : le *concept phénoménal* de « bleuité ». En utilisant ce nouveau concept, elle peut articuler des nouvelles croyances, acquérir des nouvelles connaissances sur la *propriété* phénoménale de « bleuité ». A première vue, comme l'observe judicieusement Nida-Rümelin[73], cette conception se rapproche considérablement de l'hypothèse de la connaissance par accointance défendue par Conee, car les deux auteurs insistent sur le fait que les progrès épistémiques de Mary sont caractérisés d'abord par le fait qu'elle se *familiarise* avec la propriété phénoménale de « bleuité » en s'engageant cognitivement du point de vue de la première personne. Mais à la différence de Conee, pour les défenseurs de l'hypothèse « savoir nouveau/fait ancien», cette familiarisation qui se réalise depuis l'intérieur, avec la propriété phénoménale de « bleuité », permet au sujet de former et de manipuler un nouveau concept, i.e. le *concept phénoménal* de « bleuité », et lui confère par la même occasion, la possibilité d'accéder à de nouvelles croyances. En clair, ce que Mary acquiert à la fin de sa captivité n'est pas un nouveau fait non-physique, mais un nouveau *concept* : le concept phénoménal de « bleu ».

Dans l'article 'Mary Mary, Quite Contrary'[74] , G. Graham et T. Horgan s'attaquent également à l'intuition fondamentale sur laquelle repose l'argument de Jackson. Ils commencent par distinguer entre deux types de matérialisme : (i) matérialisme 'mince' ou 'faible' et (ii) le matérialisme 'substantiel', ou 'fort'. Les défenseurs du matérialisme faible rejettent l'idée selon laquelle Mary apprend *quelque chose* lors de sa libération, tandis que les partisans du matérialisme substantiel tentent de concilier les thèses fondamentales du matérialisme avec l'idée selon laquelle Mary, à

[72] T. Horgan, (1984), 'Jackson on Physical Information and *Qualia*', *Philosophical Quarterly* 32: 127–136. Repris in Ludlow et al., (2004) op. cit. (Nous faisons référence ici à cette dernière édition)
[73] M. Nida-Rümelin, (2007), 'Grasping phenomenal properties', *in* T. Alter & S. Walter, op. cit., pp. 307–349.
[74] G. Graham & T. Horgan, (2000), "Mary Mary, Quite Contrary", *Philosophical Studies* 99: 59–87.

la fin de sa captivité monochrome, est sinon sidérée, du moins surprise, étonnée, interloquée, par sa nouvelle expérience. Selon Graham et Horgan, le meilleur exemple de matérialisme 'mince' est la position de Dennett, tandis que le matérialisme 'substantiel' est illustré par M. Tye avec son modèle 'PANIC' (« Poised Abstract Nonconceptual Intentional Content »)[75].Bien qu'il nous soit impossible de développer ici, faute d'espace, l'intégralité des débats autour du modèle proposé par Tye, nous allons tenter, cependant, d'en fournir un compte-rendu aussi précis que succinct afin de mieux saisir le tableau général des subtilités dans les controverses que suscite l'histoire de Mary.

Selon le modèle 'PANIC' défendu par Tye[76], les qualités phénoménales sont des types de contenu *nonconceptuel* et représentationnel. Le point essentiel souligné par Tye repose sur l'observation que l'expérience des couleurs possède une nature subjective beaucoup plus variée et sophistiquée que les catégorisations conceptuelles censées les subsumer. Par exemple, l'expérience d'une nuance déterminée de rouge, disons $rouge_{29}$ est qualitativement différente de celle du $rouge_{28}$; mais nous ne disposons pas d'un concept spécifique pour *chaque* nuance de rouge. Bref, mes *jugements* de couleur ordinaires sont beaucoup moins discriminants que mes *expériences* de couleur. Cela prouve, insiste Tye, que le contenu représentationnel propice à l'identification avec la qualité phénoménale doit être *nonconceptuel*, dans le sens où le sujet d'une expérience qualitative *n'a pas besoin* de posséder aucun des concepts que les théoriciens utilisent habituellement pour faire référence et circonscrire les conditions de validité concernant le contenu représentationnel respectif. Etant nonconceptuel, le contenu en cause est également *abstrait*, dans le sens où il n'est pas composé d'objets et d'occurrences particulières. Mais la caractéristique la plus

[75] « Contenu Intentionnel Nonconceptuel Abstrait en Equilibre » (M. Tye, 1995: 137) Nous traduisons.

[76] "My first condition, then, on the representational content suitable for identification with phenomenal character, is that it be *nonconceptual*, where to say that a mental content is nonconceptual is to say that its subject need not possess any of the concepts that we, as theorists, exercise when we state the correctness conditions for that content. Another condition is that the relevant content be *abstract*, that is, that it be content into which no particular concrete objects or surfaces enter. (…) A third condition is that the content be suitably *poised*. This condition is essentially a functional role one. The key idea is that experiences and feelings, *qua* bearers of phenomenal character, play a certain distinctive functional role. They arise at the interface of the nonconceptual and conceptual domains, and they stand ready and available to make a direct impact on beliefs and/or desires. For example, how things phenomenally look typically causes certain cognitive responses—in particular, beliefs as to how they are if attention is properly focuset Feeling hungry likewise has an immediate cognitive effect, namely, the desire to eat. In the case of feeling pain, the typical cognitive effect is the desire to protect the body, to move away from what is perceived to be producing pain. And so on. (…)Experiences and feelings, then, have poised, abstract, nonconceptual contents. Hence the acronym PANIC: phenomenal character is one and the same as Poised, Abstract, Nonconceptual, Intentional Content. " (M. Tye, 2000:62)

intéressante concerne son état d'*équilibre*. Selon Tye, il s'agit d'un rôle fondamentalement fonctionnel. L'idée centrale est que les expériences vécues, en tant que substrats des qualités phénoménales, occupent un rôle fonctionnel distinctif et déterminé. Ces expériences surgissent *à l'interface entre le conceptuel* et *le nonconceptuel*, et elles sont en permanente alerte afin de produire un impact direct sur le domaine des croyances et désirs. Pour résumer, note Tye, les expériences ont des contenus nonconceptuels, abstraits et en permanente alerte, situés en équilibre à l'interface entre le conceptuel et le nonconceptuel. D'où l'appellation de « Contenu Intentionnel Abstrait Nonconceptuel en Equilibre ».

Tye[77] défend d'abord un modèle représentationaliste de l'expérience phénoménale. Deux sortes de concepts phénoménaux sont à répertorier : des *concepts indexicaux* (utilisés lorsqu'on indique une nuance particulière de rouge, par exemple), et *concepts prédicatifs* (employés afin de mettre en place certaines capacités discriminatives). Il essaye ainsi d'expliciter la force de l'intuition initiale contenue dans l'argument de Jackson selon laquelle Mary ne comprend pas entièrement la nature du *quale*-rougeur *avant* d'en avoir fait l'expérience effective. Etant donné que la qualité phénoménale de « rouge » - que nous allons appeler ici, pour plus de simplicité le $quale_R$ - est censée avoir une nature physique, on pourrait s'attendre à ce que sa nature physique soit entièrement compréhensible en termes physiques, et donc intégralement traduisible dans une conceptualisation physique. La solution de Tye à cet impasse est surprenante : selon Tye, bien que le $quale_R$ ait ultimement une nature physique, un sujet peut *comprendre* entièrement la nature du $quale_R$, uniquement s'il le saisit sous une conceptualisation phénoménale, autrement dit, uniquement s'il entre en possession du concept phénoménal afférant.

Plus récemment, Tye[78] a reformulé sa position en s'appuyant, comme nous l'avons déjà brièvement évoqué, sur une version assouplie de l'hypothèse de la connaissance par accointance. De manière générale, les défenseurs du physicalisme peuvent attaquer le KA en démasquant la faiblesse d'une des assomptions implicites sur lesquelles repose l'expérience de pensée de Jackson, à savoir l'hypothèse selon laquelle l'acquisition d'une *connaissance* nouvelle coïncide obligatoirement avec la découverte d'un nouveau *fait*. Autrement dit, le savoir authentique contient seulement des éléments

[77] (M. Tye, 1995)
[78] (M. Tye, 2009)

de type « savoir que ». Selon Tye[79] il faut distinguer entre connaissance par accointance et connaissance factuelle. Dans sa prison, Mary connaît bel et bien tous les *faits physiques* à propos de l'aspect qualitatif et subjectif de l'expérience perceptuelle du rouge. Ce dont elle manque pour que son savoir soit complet c'est une connaissance *directe*, par accointance, du rouge. Pour reprendre l'exemple de Tye, je peux savoir beaucoup de choses sur le pape : je sais qu'il habite à Vatican, qu'il s'appelle Benoît XVI, et qu'il s'habille en Prada – mais il y a un sens fondamental dans lequel je *ne connais pas* le pape. De manière analogue, lorsque Mary quitte enfin sa geôle, et se familiarise directement avec la perception du rouge, elle obtient une connaissance appartenant à la catégorie logique distincte de sa précédente connaissance factuelle. Bref, durant sa captivité Mary était dépourvue de la connaissance du caractère *phénoménal* de l'expérience du rouge.

Selon Tye, cette solution au problème posé par l'argument de la connaissance est tellement évidente et naturelle, qu'on peut effectivement se demander pourquoi n'est-elle pas venue plus tôt à l'esprit des philosophes, car c'est la seule solution qui rend justice à l'intuition initiale selon laquelle, après sa libération, Mary apprend quelque chose de nouveau à propos de l'expérience phénoménale du rouge. L'acquisition de la connaissance phénoménale de l'effet-que-cela-fait de voir le rouge *se greffe* sur/et *complète* la connaissance propositionnelle. Autrement dit, un savoir complet est un *savoir mixte*, *hybride* impliquant à la fois le savoir propositionnel et le savoir que Tye appelle « objectuel », *de re*. Pour le dire très succinctement, selon Tye[80], en connaissant l'effet-que-cela-fait de voir la couleur rouge, disons, Mary entretient à la fois une relation de type « savoir que » avec la proposition particulière *p* « voir la couleur rouge fait cet effet-*ci* », tout comme elle est position d'entretenir la proposition *p* sous une description phénoménale, *via* son accointance et sa familiarisation avec la couleur rouge.

---

[79] "Well, the physicalist can begin by pointing to an implicit assumption of the argument: that Mary cannot come to know something new unless she discovers a new fact. The argument thus assumes that all worldly knowledge is knowledge that. I now think that this is where the knowledge argument crucially goes wrong. As we saw in chapter 5, there is knowledge by acquaintance and there is fact knowledge. Mary in her room knows all the physical facts about the subjective character of the experience of red But there is a perfectly ordinary sense of 'know' under which she does not know the thing that is the subjective character of the experience of red She is not acquainted with that thing. When she leaves the room and becomes acquainted with the phenomenal or subjective character of the experience of red, thereby she knows it. This is genuinely new knowledge, logically distinct from her earlier factual knowledge. It is important to grasp that Mary, in acquiring this knowledge, is not just acquiring a new way of knowing something she knew by other means before. Mary in her room did not know the phenomenal character to the experience of red"(*Ibid*., p. 131)

[80] *Ibid.*, p. 133.

Il est important, souligne Tye, de distinguer le sens du concept d'« accointance » utilisé ici, du concept tel qu'il utilisé par Conee[81] selon lequel connaître par accointance l'effet-que-cela-fait d'avoir une expérience *X*, c'est d'observer *consciemment* que *X* est en train de se dérouler. Selon Tye[82], l'acception de Conee repose sur une utilisation fallacieuse du terme « observer », car l'observation comporte une nature intrinsèquement conceptuelle. Tye insiste sur le fait qu'un sujet peut avoir une connaissance par accointance d'un objet sans pour autant le subsumer nécessairement (obligatoirement) à un concept quelconque.

Le but du présent chapitre étant d'attirer l'attention sur un éventail de problématiques plutôt que d'en faire l'examen exhaustif, nous nous en tiendrons ici à ces quelques remarques concernant la nouvelle prise de position de Tye dans l'affaire « Mary ». Pour terminer, essayons de résumer, au risque de nous répéter, les points fondamentaux sur lesquels repose la solution de Tye au problème de l'expérience phénoménale. Il existe un type de connaissance à propos de la couleur rouge qui n'est pas compris dans la connaissance exhaustive de tous les faits physiques relatifs au rouge. *La connaissance par accointance est une catégorie du savoir logiquement indépendante du savoir propositionnel et factuel.* Par conséquent, conclue Tye[83], connaître tous les faits physiques ne garantit pas un savoir complet. Il y a des faits authentiquement réels qui manquent dans le tableau physicaliste. Il y a des choses, des facettes qui nous échappent lorsqu'on connaît des faits physiques.

[81] "The acquaintance hypothesis under consideration is the view that coming to know what an experience is like requires only notice the experience as it is undergone". (E. Conee 1994:143)
[82] (M. Tye, 2009: 136)
[83] "There is a kind of knowledge of the color red that is not given to us by our knowledge of all the relevant physical facts. The knowledge we get by acquaintance with red is logically independent of our knowledge of truths. (…) Thus, knowing the relevant facts does not in itself enable us to know red rather than green. Of course, if physicalism is true, the relevant physical facts with respect to my own experiences of red-involving facts necessitate that I know red (although they do not necessitate this *a priori*). The point is that it is not my reflecting on the physical facts and thus my knowing them that accounts for my knowing red (in one important sense of 'knowing red'). Knowledge of the latter sort is not part and parcel of knowledge of the former sort. This is why I have the sense that I do not know everything there is to know as I ponder the physical story. There really are things I do not know in knowing the physical facts." (M. Tye, 2009:139)

5.5.2 Le Représentationalisme de W. Lycan

Une autre version du représentationalisme combinée avec la stratégie Savoir Nouveau/Fait Ancien est défendue par W. Lycan[84]. S'inspirant des travaux de D.H. Armstrong[85] et G. Pitcher[86], Lycan[87] affirme que non seulement le rôle de l'expérience perceptuelle est celui de représenter, mais surtout que la nature même de *toutes* les sensations est celle de représenter. Le terme « représentation » doit être compris ici comme un synonyme du « sens » frégéen. Par conséquent, différentes représentations diffèrent fonctionnellement, même si elles sont des représentations d'un et seul objet, ou état des choses. Par exemple, je sais que je pèse 53 kilos, tandis que vous savez qu'Anna pèse 53 kilos. Selon Lycan, deux schémas d'interprétation entrent alors en jeux: (i) une interprétation fonctionnelle ou computationnelle qui différencie la connaissance que j'ai de mon propre poids, de votre connaissance de mon poids (sachant que ces fragments de connaissance ont des rôles causaux radicalement différents pour vous et pour moi) ; et (ii) une interprétation référentielle ou « condition-de-vérité » (« truth-conditional ») qui ne les distingue en rien. A partir de là, connaître l'effet-que-cela-fait d'avoir telle ou telle sensation est assimilable à un rôle fonctionnel plutôt qu'à une opération concrète de fixation de la référence. En clair, conclue Lycan, il n'y a pas de faits intrinsèquement subjectifs, à la première personne (« perspectival facts »), ou des faits de type « l'effet-que-cela-fait ».

La position de Lycan peut être résumée, *grosso modo*, ainsi : c'est uniquement *après* sa libération que Mary peut formuler des *représentations introspectives de second ordre* relatives à ses propres expériences chromatiques. Pour définir les *représentations introspectives de second ordre*, Lycan fait appel à la notion d' « occurrences » qui s'implantent dans un des multiples langages de la pensée du sujet connaissant. Lorsque Mary éprouve enfin l'expérience phénoménale de rouge, elle sélectionne et fixe l'occurrence mentale d'un mot sémantiquement primitif correspondant à l'état de premier ordre qu'elle éprouve intérieurement. Dans le langage de la pensée de Mary, ce mot joue un rôle inférrentiel et conceptuel *unique*, privé, dans le sens où aucun autre sujet ne pourrait employer une représentation fonctionnellement

[84] W. Lycan, (1990), 'What is the Subjectivity of the Mental', *Philosophical Perspectives* 4 (Action Theory and Philosophy of Mind): 109–130. (Voir aussi W. Lycan, 1996)
[85] D.M. Armstrong, (1961), *Perception and the Physical World*, London: Routledge.
[86] G. Pitcher, (1971), *A Theory of Perception*, Princeton, N.J., Princeton University Press.
[87] (W. Lycan, 1996: 80)

exactement similaire. Par conséquent, insiste Lycan[88], les mots que l'on utilise « intérieurement » ne sont assimilables à *aucun* mot de la langue française que l'on utiliserait pour communiquer publiquement nos expériences qualitatives.

5.5.3 Concepts à la première et à la troisième personne

Une autre version de ce type d'approche a été défendue par D. Papineau[89] qui tente, quant à lui, de jongler entre le dualisme et le physicalisme. Plus spécifiquement, Papineau se déclare physicaliste quant aux *propriétés* phénoménales, mais dualiste quant aux *concepts* employés pour référer à ces propriétés. Il distingue ainsi entre les pensées à la première et à la troisième personne relatives à l'expérience subjective. Il est crucial, insiste Papineau[90], pour la défense et la survie du physicalisme, que ses adeptes réalisent un aspect essentiel relatif à l'expérience consciente : même si les propriétés phénoménales *sont* des propriétés physiques localisables dans le cerveau, on peut faire référence aux propriétés phénoménales en utilisant deux chemins conceptuels distincts : celui des concepts phénoménaux, et celui des concepts « à la troisième personne ». Cette distinction, note Papineau[91], s'apparente à la distinction que fait Chalmers entre les concepts « phénoménaux » et les concepts « psychologiques », et qui peut être rangée d'ailleurs sous le label « matérialisme de type-B » que nous avons déjà mentionné dans la section précédente. Rappelons-nous qu'à la différence des physicalistes de type-A (pour lesquels les concepts désignant les états conscients étaient des concepts fonctionnels tout court), les matérialistes de type-B reconnaissent l'existence d'un type de concepts non-fonctionnels, i.e. les concepts phénoménaux. Mais il serait abusif, insistent-ils, d'inférer à partir de l'existence d'un dualisme des concepts, vers un dualisme ontologique pur et dur. Autrement dit, le dualisme

---

[88] "The introspective word would certainly not be synonymous with any primitive or composite expression of public English." (W. Lycan, 1996: 101)

[89] D. Papineau, (1995), 'The antipathetic fallacy and the boundaries of consciousness', *in* T. Metzinger, (ed.) *Conscious Experience*, Paderborn: Ferdinand Schöningh.

[90] "While I am physicalist about determinate conscious *properties*, I am a sort of dualist about the *concepts* we use to refer to these properties. I think that we have two quite different ways of thinking about determinate conscious properties. Moreover, I think that is crucially important for physicalist to realize that, even if conscious properties are just physical properties, they can be referred to these in two different ways. (…) I shall call these two kinds of concepts 'phenomenal' concepts and 'third-person' concepts." D. Papineau, (2003), 'Theories of Consciousness' *in* A. Jokic & *Q. Smith* (eds). *Consciousness: New Philosophical Perspectives*, Oxford University Press, p.355.

[91] *Ibid*.

épistémique séparant le phénoménal de physique n'est pas assez fort pour entraîner un schisme *ontologique* entre le phénoménal et le physique.

A ce stade de la discussion, on pourrait, à juste titre, s'inquiéter de savoir quelle est, finalement, la *nature* des concepts phénoménaux. Presque immédiatement, des problèmes et des divergences surgissent. Certains, comme Tye[92] et Chalmers[93], attribuent aux concepts phénoménaux une nature *sui generis*, tandis que d'autres, comme B. Loar[94], leur confèrent une nature recognitionnelle et discriminative. Papineau lui-même a oscillé entre plusieurs alternatives, les définissant tantôt comme démonstratifs[95], tantôt indexicaux[96], pour s'arrêter récemment[97] à la position suivante : les concepts phénoménaux sont des cas particuliers de concepts perceptuels. Nous avons vu que pour les adeptes du matérialisme de type-B il existe une possible interprétation de l'affirmation « Mary ne connaît pas l'effet-que-cela-fait de voir du rouge » qui n'entre pas en conflit avec les thèses fondamentales de l'ontologie matérialiste. Plus spécifiquement, ce que Mary obtient à sa sortie de prison est un nouveau *mode de réflexion* et *de conceptualisation* de l'expérience du rouge. En clair, durant sa captivité, Mary est dépourvue du concept phénoménal de l'expérience qualitative du rouge. Elle pouvait, bien évidemment, réfléchir à son expérience en utilisant des concepts physiques, mais le versant phénoménal lui était caché, un peu comme le visage de la lune est caché aux terriens. C'est la raison pour laquelle Mary ignore le contenu de l'identité « voir du rouge fait *cet effet là* ». Selon Papineau, l'attribut crucial des concepts phénoménaux est le fait qu'ils sont *dépendants de l'expérience* : l'acquisition d'un concept phénoménal dépend intrinsèquement du fait que le sujet doit préalablement vivre, subir effectivement l'expérience en cause. Autrement dit, Mary a besoin de *voir* avec ses propres yeux un rubis afin d'acquérir la faculté de penser l'identité « voir du rouge = cela fait *cet effet là* ».

Dans le cas de Mary, lorsqu'elle est enfin libérée, et lorsque, après un certain temps, elle est suffisamment familiarisée avec l'expérience chromatique du rouge, elle peut « reproduire », dans son esprit, à l'aide de l'imagination, la couleur « rouge ». Ces

[92] (M. Tye, 2009)
[93] D. Chalmers, (2003), 'The content and epistemology of phenomenal belief', *in* A. Jokic & *Q. Smith* (eds.), op. cit.
[94] (B. Loar, 1990/1997)
[95] (D. Papineau, 1993a)
[96] (D. Papineau, 2002)
[97] "My current view is that phenomenal concepts are simply special cases of perceptual concepts". D. Papineau, (2007), 'Phenomenal and perceptual concepts', *in* T. Alter & S. Walter (eds.), op. cit., p.126.

images mentales intérieures reproduisant des expériences d'une nature spécifique, peuvent être utilisées pour référer à des expériences qualitatives chromatiques. Par exemple, il est évident que Mary n'aurait pas pu entretenir des pensées *à la première personne* à propos des couleurs, autrement dit, elle n'aurait pas pu utiliser les images mentales intérieures correspondantes à l'expérience du rouge *avant* d'en avoir fait effectivement l'expérience d'une tomate mûre, disons. Une fois libérée de sa captivité, Mary acquiert un *type* nouveau de croyances, à savoir des croyances à la première personne à propos de l'expérience du rouge. Mais, chose importante, pour chaque *nouvelle* croyance à la première personne qu'elle acquiert, il y aura toujours une *ancienne* croyance à la troisième personne correspondante, qui fait référence au *même* état de choses, comportant ainsi le *même* contenu factuel. Pour dire les choses autrement, l'expérience à la première personne n'est qu'un autre visage, un autre mode de présentation de l'expérience à la troisième personne. Aucun fait nouveau n'est à enregistrer, mais il s'agit seulement d'un fait ancien déguisé sous un masque différent, à savoir le *masque de la première personne*.

5.5.4 L'approche indexicale : J. Perry

Une autre variante de la position selon laquelle l'acquisition de Mary à sa sortie de prison n'est qu'un nouveau *mode* de présentation d'un fait *déjà* connu, a été défendue par J. Perry[98]. Selon Perry, la caractérisation de la nature du savoir « nouveau » que Mary acquiert lors de sa libération ne devrait pas poser de sérieux problèmes au physicalisme, en tout cas pas plus que la caractérisation des pensées indexicales de type « Je suis un philosophe ». Reprenons l'histoire de Mary depuis le début : la captive quitte sa prison. Elle voit un coquelicot rouge. Elle sait à présent que certains coquelicots sont rouges. Elle est en train de faire l'expérience d'un objet rouge pour la première fois de sa vie. Elle apprend quelque chose. Mais la question est : qu'apprend-elle plus précisément ?

Selon Perry, la « découverte » que fait Mary à sa sortie de captivité peut être exprimée à l'aide d'un des plusieurs énoncés suivants :

(1) C'est *cela* l'effet-que-cela-fait d'avoir mon expérience courante.

[98] J. Perry, (2001), *Knowledge, Possibility and Consciousness*, Cambridge, MA: MIT Press.

(2) C'est *cela* l'effet-que-cela-fait pour moi de voir du rouge maintenant.

(3) C'est *cela* l'effet-que-cela-fait de voir du rouge.[99]

Notons que l'énoncé (3) est un abrégé résumant : « c'est cela l'effet-que-cela-fait pour moi de voir du rouge maintenant, et ce qui aurait été pour moi de voir du rouge auparavant, et ce qui a été et le sera toujours pour les autres sujets dotés de facultés visuelles normales et dans des conditions extérieures normales, de voir du rouge ». Regardons l'énoncé (1) de plus près. Pour Perry, il est assez clair que (1) indique la manifestation d'un fait que Mary ne connaissait pas auparavant, mais l'énoncé n'en dit pas plus sur le *contenu* de ce fait, puisque dans notre qualité de spectateurs extérieurs à l'expérience de Mary, nous sommes dans l'impossibilité de connaître avec exactitude le caractère subjectif de l'expérience phénoménale de l'ex-captive. En clair, l'énoncé (1) fait référence, souligne Perry, à un événement inconnu par Mary pour la simple bonne raison que l'occurrence de cet événement ne s'est pas manifestée dans sa cellule durant sa captivité. Mais cette conclusion ne met pas le physicalisme en danger, car le véritable enjeu de l'argument de Jackson concerne la connaissance que Mary a des expériences qualitatives des *autres* sujets, et non pas de *ses propres* expériences.

Venons-en à présent au deuxième énoncé. Selon Perry, l'énoncé (2) ne pose pas de problème significatif non plus, car, durant sa captivité, Mary savait que les rubis, les tomates mûres et parfois les coquelicots, sont rouges, et elle savait également à quoi ces objets ressemblent. Par conséquent, lorsqu'elle franchit le seuil de sa prison et voit enfin un champ de fleurs, ou une parure de bijoux, elle sait y identifier les coquelicots et les rubis, et inférer que l'expérience visuelle qui les accompagne *est* l'expérience phénoménale du rouge. Si les énoncés (1) et (2) sont sans problèmes pour le physicalisme, il faut s'orienter alors vers l'assertion (3), car cette dernière caractérise le genre de connaissance que Mary aurait dû posséder dans sa prison, si son savoir physique avait été réellement complet. L'énoncé (3) nous informe que, dans des conditions normales, l'expérience qualitative visuelle du rouge fait *cet* effet-*là*. Si (3) exprime un fait physique, il aurait dû être transcrit dans les textes censés d'apprendre à Mary, dans sa prison, toute l'information physique relative à la perception visuelle des couleurs. Manifestement, tel n'est pas le cas, et, par conséquent, l'assertion (3) n'exprime pas un fait physique. C'est précisément ici que se trouve le point névralgique

[99] *Ibid*., p. 95.

auquel le physicalisme doit faire face. L'ennui vient du fait que lorsque Mary aperçoit pour la première fois un rubis, elle réalise à quel point sa connaissance de la vie mentale des autres était appauvrie et incomplète. Elle s'aperçoit ainsi que malgré ses recherches patientes et laborieuses de la neurophysiologie des états fonctionnels qui sous-tendent la perception visuelle des couleurs chez les autres humains, une information essentielle concernant la vie mentale des autres sujets lui était cachée.

Perry[100] nous propose alors de changer de perspective et de considérer la situation suivante : supposons que Mary apprend dans sa prison que le *quale*$_R$-rougeur *QR* est le caractère subjectif de l'expérience phénoménale de rouge des individus normaux lorsque ceux-ci perçoivent une tomate mûre dans des conditions normales, à la lumière du jour. Les manuels que consulte Mary ne précisent pas si le *QR* est une propriété physique, ou non-physique. Tout ce qu'elle sait c'est que le *QR* est impliqué dans la vision du rouge dans des circonstances environnementales normales. Avec ces précisons établies, revenons à la source, examinons à nouveaux frais la situation initiale : Mary voit un rubis. Elle sait que les rubis sont rouges. Elle sait que sa vision est normale, et en déduit que l'expérience chromatique qu'elle vient d'avoir est celle du rouge. A partir de là, elle peut inférer que sa propre expérience qualitative doit être plus ou moins similaire aux expériences qualitatives des autres sujets à la vue d'un rubis ou d'une tomate mûre. Mais cette conclusion, note Perry, repose sur l'hypothèse tacite selon laquelle il existerait un effet-que-cela-fait de voir du rouge, et que les individus en dehors de la prison possèdent le *quale*$_R$-rougeur respectif. Si Mary n'était pas en possession tacite de ces informations, il deviendrait difficile dès lors d'expliquer sa réaction de surprise à sa sortie de captivité, ainsi que l'inférence de (1) vers (3).

Selon Perry, le véritable nœud du problème que le physicaliste doit affronter ici n'est qu'une illustration d'un autre type de problème déjà signalé par Frege en 1892, et qui peut être formulé ainsi: dans quel sens les identités peuvent-elles être informatives ? Autrement dit, comment une identité pourrait apporter un contenu informatif *positif*? Revenant au cas de Mary, reposons la prémisse de départ : dans sa prison, Mary apprend que :

(4) La qualité subjective *QR* est la qualité subjective de la vision du rouge.

En s'appuyant sur l'énoncé (4), Mary peut inférer que :

(5) *QR* est *cette* qualité subjective.

[100] *Ibid.*, p.100.

De toute évidence, l'énoncé (5) est radicalement différent de (4): dans sa prison Mary connaissait (4) et ignorait (5). Si l'énoncé (5) représente un savoir authentiquement nouveau, il doit avoir un contenu distinct. Par conséquent, les conditions de vérité pour (4) et (5) doivent être différentes. Quelle est donc ce fameux nouveau savoir que Mary acquiert a sa sortie de prison ? Selon Perry[101], la réponse doit être cherchée du côté des conditions de vérités réflexives des croyances de Mary.

Lorsque nous sommes en train d'éprouver le caractère subjectif d'une expérience phénoménale, et que nous aimerions communiquer aux autres ce que nous ressentons, nous utilisons, observe Perry[102], les démonstratifs flexibles de type : « C'est *cela* la sensation que je viens d'avoir ». Baptisons ce type de démonstratif le « démonstratif interne » *cela*$_i$. A partir de là, Mary aurait pu utiliser les propositions suivantes pour exprimer ce qu'elle savait :

Avant de quitter sa prison :

(1) *QR* est l'effet-que-cela fait de voir du rouge.

Et *après* sa libération :

(2) C'est *cela*$_i$ l'effet-que-cela-fait de voir du rouge

(3) *QR* est *cela*$_i$, *cette* propriété phénoménale-*là*[103].

A présent, considérons les croyances contenues en (1), (2) et (3), et appelons-les pour plus de commodité *b*1, *b*2, et respectivement *b*3. Lorsque Mary quitte sa prison, elle est en mesure de formuler des nouvelles croyances vraies *b*2 et *b*3 qu'elle peut exprimer à l'aide des énoncés (2) et (3).

Selon Perry[104], *les conditions de vérité* de *b*1, c'est-à-dire de la croyance que Mary avait durant sa captivité et qui est exprimée en (1), sont :

---

[101] *Ibid.*, p.145

[102] "When we are attending to a subjective character in the subjective way and wish to communicate what we are feeling or noticing, we use our flexible demonstrative, "this," as in "This feeling is the one I've been having." Let's label this use of "this" as an inner demonstrative: "this*i*." Mary could use the following statement to express what she knew before leaving the Jackson room, on the basis of her reading: (1) *QR* is what it's like to see red and these statements to express what she learned upon seeing the ripe tomato: (2) This*i* is what it is like to see red. (3) *QR* is this*i* subjective character." (*Ibid.*, p. 146)

[103] Let's label this use of "this" as an inner demonstrative: "this*i*." Mary could use the following statement to express what she knew before leaving the Jackson room, on the basis of her reading: *QR* is what it's like to see red and these statements to express what she learned upon seeing the ripe tomato: (2) This*i* is what it is like to see red (3) *QR* is this*i* subjective character." (*Ibid.*)

(C)1 *b*1 est vraie si et seulement si la source du concept de *QR*, i.e. le concept impliqué en *b*1, *est* la qualité subjective de l'expérience de voir du rouge.

On voit, d'après ce qui précède, que *b*1, la croyance obtenue suite à la lecture d'un manuel, disons, est une croyance *détachée*, isolée, qui n'a pas été encore raccordée (attachée) à une action de Mary susceptible de lui fournir l'*expérience* qualitative en cause. Perry[105] illustre ce point à l'aide d'un exemple : supposons, dit-il, que je n'ai jamais rencontré Fred Dretske, mais que je sais qui il est. Mieux encore, supposons que je connais tous les faits relatifs aux livres que Dretske a écrits : *Seeing and Knowing*, *Knowledge and the Flow of Information*, *Explaining Behavior*, et *Naturalizing the Mind*. Appelons ces faits « les faits-Dretske ». Par conséquent, je peux dès lors avancer l'énoncé suivant :

(1′) Dretske a écrit *Knowledge and the Flow of Information*.

Or, voici qu'un beau jour je suis invité à une fête, et je me retrouve à côté de quelqu'un. De fil en aiguille, on commence à discuter, et l'inconnu commence à dire des choses intéressantes sur les notions de « connaissance » et d'« information ». Je réponds en l'invitant à lire un ouvrage très intéressant sur le sujet, à savoir *Knowledge and the Flow of Information*. A cela mon interlocuteur réplique « Heu… En fait… C'est moi qui a écrit *Knowledge and the Flow of Information* ». A cet instant précis, j'apprends un *fait nouveau* que je peux exprimer de deux manières distinctes :

Soit je m'adresse à mon interlocuteur en disant :

(2′) C'est toi qui as écrit *Knowledge and the Flow of Information.*

Soit je pointe du doigt vers Dretske en m'exclamant :

(3′) Cet homme-là a écrit *Knowledge and the Flow of Information*.

En quoi le contenu de mes croyances change-t-il lorsque j'acquiers ce nouveau savoir ? La réponse, dit Perry, est la suivante : au début de l'histoire, mes croyances sur

---

[104] "The reflexive truth-conditions of *b*1, the belief Mary had in the Jackson room and expressed with (1), are:
*b*1 is true iff the origin of Mary's *QR* concept, the concept involved in *b*1, is the subjective character of the experience of seeing red *b*1 was a detached belief when Mary formed it from reading a book; it never was connected to an act of attending to a subjective character. It is analogous to my first belief about Dretske, which existed for years before I had the opportunity to perceive Dretske himself and which was connected to Dretske though a chain of communicative links. So too Mary's concept is the end of a chain of communicative links;
she formed the concept reading about *QR* in a book; the chain goes back to those who introduced the term, some of whom will have done so on the basis of being subjectively aware of the sensation of red" (*Ibid.*, p147)

[105] (J. Perry, 2001 :119) (Nous traduisons)

Dretske étaient *détachées* de ma perception courante de Dretske. Une fois que ce dernier m'a avoué qui il est, mes croyances deviennent *attachées* à ma perception du philosophe Dretske.

De façon analogue, *le concept QR* dans le cas de Mary est situé au bout d'un enchaînement de liens informatifs et communicatifs. La captive obtient ce concept en lisant un manuel hypersophistiqué qui traite de *QR*. La chaîne va ainsi jusqu'à ceux qui ont introduit initialement le terme, des personnages qui l'ont introduit sur la base de leur expérience subjective directe du rouge. Mais le concept que Mary se forge dans sa prison est *détaché* de la perception du rouge, tandis que le concept introduit par ceux qui ont rédigé les livres que consulte la prisonnière, était attaché à la perception *directe* du rouge.

La croyance *b*2 est analogue à la croyance que Perry[106] entretient une fois que Dretske s'est introduit lui-même, dévoilant son identité. La croyance *b2* est donc attachée à un acte d'attention et de focalisation qui est accompagné d'un certain *quale*. Les conditions de vérité de (2) sont exactement les mêmes que celles de (1). En revanche, les conditions-réflexives de *b*2 sont différentes :

(C)2 *b*2 est vraie si et seulement si l'acte d'attention intérieure auquel la croyance est attachée, vise la nature et la qualité subjective de l'expérience de voir du rouge.

On arrive enfin à *b*3, le point névralgique de l'affaire, dans l'argument de Jackson (1982/1986). L'énoncé (3) exprime une croyance relative aux expériences *en général* des individus qui éprouvent la perception visuelle du rouge, bref une croyance que Mary aurait dû avoir si elle était réellement en possession de toute l'information physique relative à la perception chromatique. L'état de choses extérieur dont il est question dans l'énoncé (3) est le même que celui des énoncés (1) et (2), mais, et c'est ici le point essentiel, leur contenu réflexif, interne, *n*'est *pas* le même :

(C)3 *b3* est vraie si et seulement si l'acte d'attention interne qui lui est attaché *est* la source du concept *QR* de Mary.

[106] The belief *b*2 is analogous to my belief after Dretske introduced himself. That belief was attached to a perception of mine, which was of Dretske. Mary's *b*2 is attached to an act of attention, which is of a certain subjective character. The subject truth-conditions of (2) are exactly the same as those of (1). The reflexive truth-conditions of *b*2 are different: *b*2 is true iff the act of inner attention to which it is attached is of the subjective character of the experience of seeing red". (Perry 2001: 147-8)

Il s'agit d'appliquer ici une nouvelle condition-de-vérité relative aux croyances de Mary, condition-de-vérité qui résulte du changement qui s'est produit dans sa situation lorsqu'elle a aperçu pour la première fois un coquelicot rouge, disons. A nouveau, il est important de saisir, insiste Perry[107], que la modification que subit la croyance de Mary n'est pas due aux conditions de vérité qui caractérisent ses croyances relatives à l'état de choses externes (qui restent, quant à elles, invariables : une tomate mûre *est* rouge) ; en revanche, la même modification amène une transformation des conditions de vérité de ses croyances qui étaient *détachées* de leur référent : dans le cas présent, la condition de vérité en question est constituée par le caractère subjectif d'une expérience phénoménale particulière (la perception visuelle du rouge) qui, dans un premier temps a été présentée à Mary au bout d'un enchaînement informatif (livres, télévision, etc.), et auquel, par la suite, après sa libération, elle a pu avoir un accès direct, court-circuitant, si l'on peut dire ainsi, l'enchaînement informatif-communicatif initial.

En conclusion, résume Perry, l'expérience de pensée de Jackson, parfois familièrement appelé « la chambre de Jackson », comporte une fissure, presque insaisissable au premier regard, mais qui est décisive quant aux implications finales de l'argument. La remarque à première vue inoffensive selon laquelle si un fait est *connu*, il peut être rédigé et transcrit sur une feuille de papier ou sur un écran, dans une chambre noire et blanche constitue le véritable nœud problématique dans l'affaire de Mary.

5.5.4.1 L'argument de la parité des raisons

Imaginons la variation suivante sur le thème de la « chambre de Jackson » : supposons que durant sa captivité, Mary ait reçu, par le truchement d'un téléviseur noir et blanc, une série complète de leçons à propos des *qualia* de la part d'un professeur partisan déclaré du dualisme. Cette modification dans le scénario initialement imaginé

[107] "Finally we come to *b*3. This is the belief Mary expresses with (3), and it is the belief that Jackson found problematic. It is a belief about what people in general experience when they see red things, and it seems like the sort of thing she should have known in the Jackson room, if she really knew all of the physical facts about color and color perception. The subject matter content of (3) is the same as that of (1) and (2). But the reflexive content differs: *b*3 is true iff the act of inner attention to which it is attached is of the origin of Mary's *QR* concept. This is the new truth-condition on Mary's beliefs that results from the change that occurred when she saw the tomato and learned what it was like to see red As in the cases of Larry, Gary, and Terry, the change in Mary's beliefs does not result in any new conditions on the truth of her beliefs *given* what they refer to. But it does impose new conditions on the truth of her beliefs *abstracting* from what they refer to: the condition that the subjective character that is the origin of her old concept is the very one to which she is attending. That's my account of what Mary learned" (*Ibid*., p. 148)

par Jackson n'affecte en rien la plausibilité de l'intuition selon laquelle, au moment de sa libération, Mary apprend quelque chose sur le monde et les autres sujets connaissants. Par conséquent, note P.M. Churchland[108], un des principaux artisans de l'argument de la parité des raisons, si l'argument fonctionne contre le physicalisme, les partisans du dualisme ne doivent pas s'en réjouir trop vite, car la chambre de Jackson fonctionne contre le dualisme *aussi*. Regardons-y de plus près.

Supposons que le dualisme soit vrai. Après tout, il n'y a aucune raison pour que Mary ne soit pas informée, durant sa captivité, de la vérité d'une doctrine comme le dualisme, par exemple. Similairement, baptisons le caractère subjectif de la perception du rouge "*QR*". Il n'y a aucune raison non plus de considérer que l'information concernant *QR* ne soit pas accessible ou fournie sur du papier, ou par le truchement d'un téléviseur. On peut donc imaginer une variation sur le thème de Mary, et remplacer la neutralité initiale oscillant entre le physicalisme et dualisme, par une prise de parti ouvertement dualiste. Bref, on lui enseigne, à notre prisonnière, que *QR* n'est pas un état neurophysiologique du cerveau mais, un état de *nature non-physique*. A présent, lorsque Mary en quittant la prison voit pour la première fois une tomate mûre, l'intuition forte selon laquelle elle apprend quelque chose demeure intacte. Le fossé qui sépare l'information dont dispose la captive dans sa prison de l'expérience qu'elle éprouve en regardant une tomate mûre, est toujours là.

A cela Jackson[109] répond qu'il n'y a aucune raison de croire que Mary aurait pu apprendre, dans sa cellule achrome, toute l'information concernant le caractère subjectif de l'expérience phénoménale : pour obtenir un argument viable contre le dualisme, la première prémisse de l'argument initial selon laquelle Mary détient toute l'information physique concernant la perception visuelle des couleurs, doit être remplacée par une prémisse selon laquelle l'héroïne possède toute l'information possible selon les exigences du dualisme. Or, insiste Jackson la première prémisse est plausible, tandis que la deuxième non. Par conséquent, il n'y a aucun problème de « parité des raisons» dans ce cas précis.

---

[108] Voir P.M. Churchland, (1985a), 'Reduction, *Qualia*, and the Direct Introspection of Brain States', *Journal of Philosophy* 82, pp. 8-28. P.M. Churchland, (1985b), 'Review of Robinson's *Matter and Sense*', *Philosophical Review* 94, pp. 117-120. Pour une réponse critique, voir Y. Nagasawa, (2002), 'The Knowledge Argument against Dualism', *Theoria* LXVIII, pp. 205-223.

[109] (F. Jackson, 1986)

Selon Perry[110], qui vise s'inscrire en faux contre la réponse de Jackson, la solution qu'offre ce dernier comporte une faille majeure. Considérons l'état neurophysiologique qui sous-tend la perception visuelle du rouge chez des individus normaux, dans des conditions d'éclairage normal, et baptisons-le *S*. Supposons que *QR* est un attribut de *S*. Appelons le fait que *S* possède comme attribut *QR* « *ce-S-est-QR* ». Si *QR* est un aspect non-physique du cerveau, il s'ensuit que « *ce-S-est-QR*» est un fait dualiste. Imaginez à présent que vous avez un livre ouvert devant vous, où vous y lisez le fragment suivant :

« Ce qui caractérise avant tout les états conscients, est le fait que le sujet connaissant en fait l'expérience directe et immédiate. Ce trait fondamental des états neurophysiologiques du cerveau est appelé communément « subjectivité ». Il s'agit d'un aspect à la fois très intéressant et très important de la vie mentale d'un individu. Un des attributs les plus fascinants de la subjectivité représente le caractère subjectif de l'expérience qu'éprouvent les individus normaux lorsqu'ils regardent une tomate mûre, en pleine lumière du jour ».

Admettons que ce fragment est écrit dans un traité que Mary *aussi* est en train de le consulter. A présent, vous et elle (Mary) y lisez la suite, le paragraphe (P) : « *Durant une longue période, on ne savait pas avec certitude si QR était un aspect physique on non-physique du cerveau. Mais à présent on sait que….* »

Indiscutablement, il est possible de remplir l'espace laissé libre par les trois points de suspense soit par une description physicaliste, soit par une dualiste. Une fois que Mary lit la mystérieuse phrase au complet, elle *sait* que *QR* est le caractère subjectif de l'expérience visuelle du rouge. Autrement dit, elle sait que *ce-S-est-QR*. Dans les deux cas, même si elle sait que *ce-S-est-QR*, l'intuition selon laquelle elle apprend quelque chose à sa sortie de captivité reste tout aussi frappante que dans l'expérience de pensée originale de Jackson. Même si Mary examine très attentivement le paragraphe (P), il y a, semble-t-il, un fossé infranchissable entre les expériences de Mary du « dedans » et du « dehors ». Par conséquent, note Perry[111], l'existence de ce fossé n'a rien à avoir avec le physicalisme et Churchland a eu raison. Mais si tel est le cas, il s'ensuit que nous nous sommes égarés dans une mauvaise direction depuis le début de l'affaire de l'argument de la connaissance. En effet, il se peut que le problème

[110] (J. Perry, 2001)
[111] *Ibid.*, p. 166.

soit effectivement ailleurs, dans l'insertion d'un présupposé en apparence innocent, habilement introduit par Jackson dans la description de l'état de captivité de Mary, à savoir dans l'hypothèse selon laquelle si un fait est connu, il peut être en principe, écrit sur du papier noir et blanc, et mis à la disposition de Mary par le truchement d'un écran. C'est précisément à ce point névralgique que Jackson ne se montre pas « équitable » envers les physicalistes, car selon lui, toute l'information physique peut être couchée sur du papier noir et blanc, tandis que toute l'information dualiste ne peut pas l'être.

A ce stade de notre enquête, il est clair, dit Perry, que le véritable moteur de l'argument de la connaissance est à la fois son point fort et son talon d'Achille. Plus spécifiquement, on part de la présupposition tacite que le physicaliste adhère à une interprétation du concept d' « objectivité » qui lui interdit par la suite d'admettre que Mary acquiert un savoir nouveau lorsqu'elle quitte sa geôle, savoir qu'elle peut obtenir uniquement en éprouvant l'expérience respective de manière directe et immédiate.

Selon Perry, le physicaliste est parfaitement en droit d'adhérer à l'idée selon laquelle il existe des voies subjectives pour connaître des faits physiques, objectifs. Une personne normale perçoit des objets du monde extérieur sous un mode auquel une personne aveugle n'aura jamais accès. Est-ce que cela mettrait en péril le physicalisme ? Non, répond Perry[112], et nous le suivons sur ce point. Tout ce que cette thèse prouve c'est que l'argument de Jackson repose sur une interprétation erronée de la connaissance en général, à savoir sur l'idée selon laquelle il existerait un certain type de connaissance qui serait accessible à partir du point de vue de nulle part. Autrement dit, selon ce point de vue, en surplomb, la connaissance consisterait à circonscrire avec précision un sujet d'enquête déterminé, un fait. Peu importe l'angle sous lequel nous y avons accès : perspective en première ou troisième personne, langage, états sensoriels, schème conceptuel, etc. Tous ces aspects représentent en quelque sorte une extension « naturelle » du sujet d'enquête. Ce qui est connu n'a aucun pouvoir contraignant le sujet connaissant d'utiliser un type particulier de représentation. Mais il s'agit là, insiste Perry, d'une distorsion grossière de la compréhension du concept de « connaissance ».

[112] "A sighted person can know what it is like to see objects in a way that a person who has never seen cannot. Is there anything about this that violates the spirit of physicalism? I do not see that there is. All that is violated is a false picture of knowledge. This is the view that there is some kind of knowledge that involves grasping a fact not from any point of view—a view from nowhere. Knowledge consists of the subject matter fact known, and the way it is known is irrelevant: the perspective, the language, the sensory states, the particular system of ideas and notions don't matter." (J. Perry, 2001: 166)

Un système de représentation objective est un système au service de la complétion de la connaissance, et non pas un tout absolu, le corps entier de la connaissance.

En conclusion, note Perry, le physicalisme n'est pas moins compatible avec les *qualia* qu'il ne l'est avec l'existence des rendez-vous galants, disons. En clair, ce que l'expérience de pensée de Jackson montre c'est qu'une personne ayant un accès visuel à la perception du rouge connaît cette couleur sous un mode *différent* de celui d'une personne qui n'a jamais eu un accès visuel direct à un objet rouge. Or, insiste-t-il, on voit mal en quoi cette dernière phrase représenterait une menace pour le physicalisme. Tout ce qu'elle remet en cause c'est une idée erronée que l'on s'est formée de la connaissance, à savoir l'idée selon laquelle il existerait un certain type de connaissance susceptible d'être capturé du *point de vue de nulle part*.

Il serait trop long, et d'ailleurs superflu pour l'objectif qui nous concerne ici, de décrire en détail les prises de position qui s'inscrivent en faux contre le point de vue défendu par Perry. Il nous suffit de résumer à gros traits les principales objections. Dans un premier temps on a signalé que dans les cas habituels de référence démonstrative, l'objet en question est en quelque sorte, *donné à l'accès épistémique du sujet*. Lorsque je pointe vers une table et je m'y réfère en disant « cette table », l'objet respectif peut être décrit comme étant « la table qui est située à ma gauche ». Mais dans le cas de Mary on voit mal comment l'expérience phénoménale de rouge pourrait être subsumée à un concept démonstratif de type « cette sorte d'expérience-*là* » analogue à l'expression précédente « cette table-*là* ». Qui pourrait bien être l'énoncé-candidat idéal pour ce démonstratif ? Admettons que l'heureux élu serait la formule « *ce* type d'expérience que je suis en train d'éprouver ». Comme le relèvent Chalmers[113] et Levine[114], cette option pose quelques problèmes, car le mode phénoménal, privé, sous lequel le sujet semble être accointé avec l'objet-référent ne peut pas être adéquatement capturé par une expression de type neutre, démonstratif.

Examinons de plus près les deux propositions suivantes :

(1) Dretske a écrit *Knowledge and the Flow of Information*.

(2) *Cet* homme a écrit *Knowledge and the Flow of Information*.

---

[113] D. Chalmers, (2004), 'Imagination, Indexicability, and Intensions', *Philosophy and Phenomenological Research* 68:182-90 (symposium autour du livre de J. Perry (2001), *Knowledge, Possibility, and Consciousness*) Voir aussi (D. Chalmers 2003 : 220-271)

[114] J. Levine, (2007), 'Phenomenal Concepts and the Materialist Constraint', *in* T. Alter & S. Walter (eds.) (2007): 145–166.

Comme le relève fort bien Levine[115], Perry connaissait (1) avant de rencontrer effectivement Dretske à une soirée, tandis que l'énoncé (2) contient une connaissance qu'il a apprise uniquement *après* que son interlocuteur l'ai informé qu'il *est* Dretske. Mais on ne détecte aucun possible fossé explicatif entre les expressions « cet homme » et « Dretske » pour la simple bonne raison qu'elles font référence à un seul objet : le philosophe Fred Dretske, né en 1932, etc.

En y regardant de plus près, cependant, on voit que les expressions démonstratives sont presque toujours accompagnées, d'une façon ou d'une autre, d'un contenu perceptuel. Je désigne quelque chose que je vois, j'entends, ou je savoure, avec l'expression ou la pensée « *ce F* », où *F* représente le vécu subjectif afférent. Mais même si je n'emploie aucune expression de ce type, il est clair, insiste Levine, que le démonstratif reste ancré dans une représentation perceptuelle donnée. Bien que l'idée de désigner ma propre expérience subjective de manière démonstrative soit parfaitement intelligible, encore faut-il trouver un modèle détaillant le contenu matériel auquel l'expérience est ancrée, car le démonstratif en soi est incapable de remplir ce rôle. Par conséquent, note Levine, les locutions démonstratives nécessitent la présence accompagnatrice d'un certain contenu matériel afin de pouvoir s'ancrer aux objets visés (peu importe si ces objets sont des propriétés, des universels ou des particuliers). On se retrouve ainsi avec le bon vieux concept de connaissance par « accointance » sur les bras.

En ce qui concerne la notion clé d'«accointance », ou de « présence cognitive », autrement dit le concept d'une « immédiateté» primitive et privée qui caractériserait la relation entre les concepts phénoménaux et leurs objets, il est clair, note Levine[116], qu'il s'agit là d'une relation cognitive, épistémique. A partir de là, il n'est pas du tout clair pourquoi et surtout comment la présence physique se transforme en une présence cognitive. Pour dire les choses autrement : en quoi la présence physique explique-t-elle la présence cognitive ? Dans la plupart des cas, lorsqu'il s'agit de circonscrire le concept d'accointance, on fait appel à des métaphores de type « l'objet qui se présente à moi *ici* et *maintenant* ». Mais nous avons besoin de quelque chose de plus précis. Il faut expliquer comment la présence d'un état implémenté à l'intérieur d'un système

[115] *Ibid.*, pp. 154-5.

[116] "Acquaintance, or cognitive presence, or whatever it is that is supposed to constitute the especially immediate and intimate cognitive relation between phenomenal concepts and their objects, is just that: a *cognitive* relation. It is not at all clear why, or how, *physical* presence translates into cognitive presence. (J. Levine, 2007: 162)

physique (le cerveau) *devient* quelque chose d'autre, un *autre* état, dont nous sommes *conscients*[117]. Rappelons-nous, insiste Levine, qu'à l'heure actuelle, aucun mécanisme causal purement physique n'est en mesure de fournir une explication adéquate de la nature de cette métamorphose magique qui s'opère entre la présence-à-soi physique et celle épistémique. Selon Levine[118], la difficulté principale vient du fait que la connaissance par accointance se contente de *présenter* le réel, et non pas de le *représenter*, et nous avons du mal à intégrer quelque chose qui se *présente* seulement, dans un système explicatif qui aspire à *ré-présenter* le monde.

Un prolongement intéressant des remarques précédentes est l'affirmation selon laquelle le matérialisme[119] pourrait s'avérer faux non pas parce que les propriétés phénoménales elles-mêmes seraient non-physiques (car elles pourraient s'avérer physiques après tout), mais parce que les *relations* qui nous connectent à ces propriétés sont en quelque sorte primitives, brutes, et irréductibles aux relations physiques. Mais une question surgit alors: est-ce que la connaissance par accointance peut être *non-physique*, alors que les propriétés phénoménales seraient, elles, *physiques* ?

Les implications et les interrogations que soulève cette relation entre la simple *présentation* et la conceptualisation représentationnelle, font écho à un possible approchement entre l'approche phénoménologique[120] et la tradition analytique, et à ce titre, mériterait un traitement approfondi dans un essai à part. Limitons-nous à présent, pour simplifier, à passer en revue les réactions critiques que la stratégie des concepts phénoménaux a provoquées.

---

[117] *Ibid.*, p.163

[118] "We are acquainted with our experience, and as acquaintance *presents* properties, not merely represents them, we find it difficult to integrate what is presented with what is only represented in a way that allows the latter to explain the former." (*Ibid.*, p. 165)

[119] "It could turn out, then, that materialism is false not because phenomenal properties themselves are not physical—they may yet be for all we know. Rather, it would be false, on this view, because somehow we embody a relation to them that is itself brute and irreducible to physical relations. Is this a coherent position? Could phenomenal properties be physical while acquaintance is not? I don't know, but the question, to my mind, deserves exploration." (*Ibid.*)

[120] Voir J. Petitot, F. Varela, B. Pachoud, B. & J-M. Roy, (eds.), (1999), *Naturalizing Phenomenology: Issues in Contemporary Phenmenology and Cognitive Science*. Stanford, California: Stanford University Press. Voir aussi E. Thompson, (2007), *Mind in Life*: *Biology, Phenomenology, and the Sciences of Mind*, Harvard University Press.

5.6 Objections contre la stratégie des concepts phénoménaux

La stratégie dont nous venons de faire un succinct croquis prête le flanc à plusieurs critiques. On a d'abord signalé que l'immédiateté de cette référence directe ne suffit pas à éclairer la signification de la connaissance par accointance, c'est-à-dire du mode à travers lequel les qualités phénoménales se présentent à l'esprit du sujet lorsque celui-ci emploie des concepts phénoménaux. Une des réponses les plus influentes visent la thèse hautement problématique des *modes de présentation*.

5.6.1 L'Argument des Propriétés Distinctes : St. White

Dans l'article 'The Curse of the *Qualia*'[121] St. White livre une analyse très pertinente des conséquences que l'adoption de cette stratégie entraîne dans l'affaire du physicalisme. Selon White[122], les concepts phénoménaux tels qu'ils apparaissent dans les approches de type « stratégie des concepts phénoménaux » doivent satisfaire les trois conditions suivantes[123] :

(1) Les concepts phénoménaux ne sont pas équivalents aux concepts physiques-fonctionnels.

Par conséquent, en dépit de ce que soutenaient les fonctionnalistes analytiques, les énoncés d'identité de type "la douleur est identique avec la stimulation des fibres-C", sont *a posteriori*. La deuxième condition est :

(2) Les concepts phénoménaux désignent directement leurs référents.

La relation du concept "douleur" au phénomène de douleur n'est pas médiate à travers un mode de présentation de la douleur. Cette condition la rend ainsi différente de la relation référentielle habituelle entre "Hesperus"/"Phosphorus" et la planète Vénus. Par exemple, on considère habituellement que la référence du mot "Hesperus" à Vénus se réalise par le truchement d'une description de type « la première étoile visiblement scintillante du soir », et, c'est ici un point important, c'est en vertu de cette

[121] S. White, (1986), 'Curse of the *Qualia*', *Synthese*, 68: 333–368.
[122] S. White, (2007), 'Property Dualism, Phenomenal Concepts, and the Semantic Premise', in T. Alter et S. Walter, op. cit., pp. 210–248.Pour une défense de l'argument dans la même veine voir D. Chalmers (1996, 2002a)
[123] (S. White, 2007:211)

description que le concept « Hesperus » indique Vénus. En revanche, dire que la relation référentielle du concept « douleur » à l'état factuel douleur est immédiate, c'est dire qu'elle ne fait appel à aucune autre description intermédiaire. Or, comme le souligne fort bien White[124], cela ne nous informe en rien sur la manière dont les concepts phénoménaux désignent et s'attachent à leurs référents. Deux solutions ont été ainsi avancées : dans un premier temps, et à la suite de Loar, on a misé sur une interprétation recognitionnelle[125] des concepts phénoménaux. D'autres penseurs, tels Block[126] préfèrent miser sur une interprétation décitationnelle, dans le sens où les concepts phénoménaux s'attachent référentiellement aux propriétés du réel en vertu de leur emboîtement dans une instanciation des propriétés en question.

Enfin la troisième condition est la suivante:

(3) Les concepts phénoménaux ne sont pas "aveugles".

Selon l'approche recognitionnelle, il existe un concept phénoménal (de second ordre) couvrant ce que toutes les propriétés phénoménales ont en commun (premier ordre) : parfois, ce niveau est appelé « phénoménalité ». Donc, il existe un concept d'un fait qui sert de critère en vertu duquel les états phénoménaux diffèrent de ceux désignés (indiqués) par des simples concepts recognitionnels « self-orientés ». Le problème qui se pose à présent est celui de savoir si les concepts phénoménaux fournissent une explication adéquate du caractère *a posteriori* de l'identité mental/physique. A première

[124] "The ordinary assumption is that the reference, for example, of "Hesperus" to Venus is mediated by a description such as "the first heavenly body visible in the evening." And it is assumed that "Hesperus" picks out Venus *by virtue of* the latter's having the property expressed by the predicate contained in such a description—the property, in this case, of being the first heavenly body visible in the evening. To say that the referential relation of "pain" to pain is unmediated by a mode of presentation (in any ordinary sense) is not, of course, to say *how* phenomenal concepts *do* pick out their referents. On this score there are two distinct suggestions." (*Ibid*.)

[125] Une autre voix très influente dans les controverses qui gravitent autour des concepts phénoménaux est celle de Loar. Selon ce dernier, les concepts phénoménaux appartiennent à une classe de concepts plus vaste, que l'on pourrait appeler concepts « *type-démonstratifs* », et qui comportent habituellement la forme suivante : « *X* est un élément de *ce* type-*là* ». Les concepts « type-démonstratifs » sont fondés, à leur tour, sur des dispositions à classifier, par le truchement des distinctions perceptuelles, certains objets, faits, états, processus, etc. La nature des concepts phénoménaux est fondamentalement *recognitionnelle* : une personne *X* possède le concept phénoménal de « bleuité » si et seulement si *X* est en mesure de reconnaître l'expérience de bleuité lorsque celle-ci se produit. La spécificité distinctive du concept phénoménal de bleuité est de pouvoir référer *directement* à l'objet de l'expérience, autrement dit sans impliquer une propriété supplémentaire dans le processus de fixation de la référence. (Loar 1990/1997: 603–04)

[126] (Voir N. Block 2002: 396–98; N. Block, 2007)

vue, observe White[127], il y a une forte tentation de répondre par l'affirmative, étant donné que le cœur même de la théorie des concepts phénoménaux repose sur l'idée que ceux-ci ne sont pas équivalents avec les concepts physiques-fonctionnels que Mary a pu apprendre dans sa prison. Ou, pour formuler la même idée autrement, les deux types de concepts – phénoménal et fonctionnel – occupent des rôles inférrentiels différents, et avant la découverte des identités *a posteriori* mental/physique, ils sont extraits de deux *types* différents d'expériences. Par exemple, l'expérience qui est généralement décrite par l'énoncé « J'ai mal aux dents » est tirée de l'expérience de sens commun qu'une personne éprouve dans la vie de tous les jours, lorsqu'elle a un abcès dentaire, par exemple. En revanche, l'expérience décrite dans l'énoncé « Les fibres-C de mon cerveau sont stimulés », bien qu'elle réfère (ou du moins, est censée se référer) au même état de choses (ma douleur), est extraite d'une expérience distincte obtenue, disons, suite à un examen neurologique ou dentaire sophistiqué.

Or, comme le note White, dès que l'on s'enquiert de la nature de l'identité entre le mental et le physique – qui est, rappelons-nous, la cible fondamentale et le noyau dur du physicalisme – on s'aperçoit que la simple affirmation selon laquelle les concepts impliqués dans la relation d'identité aient des rôles conceptuels différents, ne suffit pas à expliquer positivement la nature *a posteriori* de l'équation états phénoménaux /états physiques. Autrement dit, ce dont on a besoin, c'est une explication adéquate de la manière dont un sujet qui affirme sincèrement ne pas croire dans une telle identité, voit le monde. En effet, rappelons-nous que le caractère *a posteriori* de ces identités entraîne un effet surprenant suivant : une personne qui refuse ou n'arrive pas à croire ces identités, reste, néanmoins, une personne rationnelle, autrement dit, elle ne peut pas être accusée d'irrationalité. Par conséquent, on doit pouvoir fournir une approche, une image explicative satisfaisante de ce que le monde serait aux yeux de ces sujets ignorants ou désinformés, mais toujours rationnels. Supposons qu'une telle explication ne sera jamais trouvée, ou qu'elle finit par engendrer des incohérences. Dans ce cas, la vérité de certaines identités mental/physique serait, du moins, en partie, de nature *a priori*, et dans ce cas, le défenseur de la stratégie des concepts phénoménaux doit trouver une réponse adéquate pour expliquer leur supposé caractère *a postériori*.

[127] "The question now is whether phenomenal concepts supply an adequate explanation of the *a posteriori* character of mental-physical identities. There is an obvious temptation to say yes, since it is a basic tenet of the theory of phenomenal concepts that they are not equivalent to any concepts in the physical-functional domain. This point could be spelled out further by saying that the two sorts of concepts have different conceptual or inferential roles". (S. White, 2007: 212)

En conclusion, note White[128], l'introduction de différentes *modes de présentation* – dans le cas des croyances relatives aux états phénoménaux – entraîne automatiquement l'introduction des différentes *propriétés* fixatrices de références, ce qui annule les ambitions initiales de la stratégie des concepts phénoménaux. Car si cette dernière approche fait appel à *deux* modes de présentation de ce type, alors cette manœuvre ne peut pas être utile dans la défense du physicalisme, car le problème des dualités n'est pas résolu, mais seulement déplacé à un niveau plus élevé, au niveau des *propriétés* : le même état phénoménal devrait être décrit en faisant référence à une propriété physique, dans le cas où la croyance du sujet vise l'état en question sous un mode de présentation physique ; et parallèlement, il devrait être décrit en faisant référence à une propriété phénoménale, dans la mesure où la croyance du sujet s'y réfère sous un mode de présentation phénoménale.

5.6.2 Le cas de Marianna

Une autre prise de position qui vise à s'inscrire en faux contre la stratégie phénoménale a été proposée récemment par M. Nida-Rümelin[129]. Cette dernière développe un argument en faveur du dualisme des propriétés qui remet en cause le bien-fondé de la stratégie phénoménale. Plus spécifiquement, Nida-Rümelin[130] introduit une distinction entre (i) le fait d'*avoir* le concept d'une propriété et (ii) le fait de *saisir* une propriété *via* un concept.

Exemple : si l'on saisit une propriété *p* par le truchement d'un concept C, alors C est un concept de *p*. Mais l'inverse n'est pas valable : on peut *avoir* le concept d'une

[128] Pour une réponse qui vise s'inscrire en faux contre l'argument de S. White voir N. Block, (2007), 'Max Black's Objection to Mind-Body Identity', *in* T. Alter & S. Walter (2007), op. cit., pp. 249–306. Block introduit une distinction supplémentaire entre modes présentation *cognitives* et *métaphysiques.*

[129] M. Nida-Rümelin, (1996), 'What Mary couldn't know', in Th. Metzinger (ed.), *Phenomenal Consciousness*, Schoenigh: Paderborn. Réédité *in* Ludlow et al., (2004), op. cit. (Nous faisons référence ici à cette dernière édition) M. Nida-Rümelin, (1998), 'On Belief About Experiences: An Epistemological Distinction Applied to the Knowledge Argument', *Philosophy and Phenomenological Research* 58 (1): 51–73.

[130] "Here I present an argument for property dualism. The argument employs a distinction between *having a concept of a property* and *grasping a property via a concept.* If you grasp a property *P via* a concept *C*, then *C* is a concept of *P*. But the reverse does not hold: you may have a concept of a property without grasping that property *via* any concept. If you *grasp* a property, then your cognitive relation to that property is more intimate than if you just have some concept or other of that property. To grasp a property is to understand what having that property essentially consists in. To have a concept of a property is to have a concept one can use to attribute the property to something." (M. Nida-Rümelin, 2007: 307)

propriété *sans* néanmoins saisir cette même propriété *via* un concept quel qu'il soit. Lorsqu'on saisit une propriété, notre relation cognitive à cette propriété-là est plus intime, plus privée, que lorsqu'on possède simplement le concept de cette propriété. *Saisir* une propriété revient à comprendre essentiellement ce que veut dire *avoir* cette propriété. En revanche, avoir le concept d'une propriété c'est avoir un concept prêt-à-l'emploi et attribuable à un état de choses. Si l'on possède le concept d'« eau », alors on peut l'utiliser pour l'attribuer la propriété d'« être eau» aux liquides, et dans ce cas, nous possédons le concept de la propriété d'« être eau ». Mais on peut imaginer facilement des cas où un individu possède le concept d' «eau», et en même temps il ignore que la composition moléculaire H_2O est essentielle pour qu'une substance soit considérée comme étant de l'eau. Selon Nida-Rümelin[131], il est relativement facile de trouver des exemples des cas où un individu *ne* saisit *pas* la propriété en cause. Mais il est plus difficile (mais pas impossible) de trouver des cas où l'on arrive à saisir des propriétés *via* les concepts. Un possible exemple nous est donné par notre compréhension des propriétés phénoménales *via* les concepts phénoménaux.

Nida-Rümelin[132]montre d'une manière très convaincante que l'expression Nagelienne traditionnellement utilisée pour définir l'aspect qualitatif de l'expérience phénoménale – connaître l'effet-que-cela-fait[133] - se prête facilement à des confusions qui peuvent avoir des conséquences importantes sur les enjeux contenus dans le KA. Plus spécifiquement, l'emploi de cette formule incite à confondre entre :

(a) la simple familiarisation par accointance avec diverses expériences visuelles chromatiques ;

(b) la connaissance relative aux types d'expériences chromatiques que les autres sujets possèdent dans une situation donnée.

Cette confusion pourrait nous empêcher de saisir ce que Nida-Rümelin appelle les *deux étapes distinctes dans le progrès épistémique* que réalise Mary une fois sortie de sa captivité. Nida-Rümelin illustre son point de vue à l'aide d'une expérience de pensée impliquant le personnage Marianna, une cousine, disons, de la célèbre Mary de Jackson.

---

[131]"I will propose an account of grasping properties. It is quite easy to find examples where we do *not* grasp the property at issue. But it might be less obvious that we sometimes do grasp properties *via* concepts. I think that we sometimes do and that a clear case is provided by our understanding of phenomenal properties *via* phenomenal concepts." (M. Nida-Rümelin, 2007: 307)

[132] (M. Nida-Rümelin, 1996)

[133] (Th. Nagel, 1974)

Comme sa cousine Mary, Marianna vit emprisonnée, dans un premier temps (t_1) dans un environnement noir et blanc. A la sortie de sa prison (t_2), et à la différence de Mary, Marianna est introduite d'abord dans une autre pièce fermée, où elle est confrontée à divers *objets-artefacts* arbitrairement colorés : tables rouges, peintures abstraites, tasses jaunes, etc., (et non pas à des bananes jaunes, fraises rouges, paysages à ciel bleu, etc., dont la couleur leur est naturellement et intrinsèquement attribuée). Dans cette situation, Marianna est de toute évidence incapable de relier les types des expériences chromatiques qu'elle vient d'avoir avec les informations dont elle disposait dans sa prison. Supposons qu'au temps t_2 on présente à Marianna quatre jetons de couleurs différentes (rouge, bleu, vert, et jaune) et on lui demande d'indiquer lequel représente la couleur du ciel telle qu'elle apparaît aux individus dotés d'une vue normale. D'une certaine manière, notre héroïne *sait*, au temps t_2, l'effet que cela fait de faire l'expérience visuelle des couleurs rouge, bleu, etc. Mais elle ignore toujours certains éléments essentiels concernant les expériences chromatiques des autres humains. Autrement dit, elle ignore toujours que le ciel apparaît *bleu* aux individus dotés d'une vue normale, et à ce titre elle peut entretenir la croyance fausse selon laquelle la couleur du ciel, aux yeux des autres individus, serait la même que celle du jeton *rouge* qui se trouve devant elle au moment t_2. C'est uniquement à t_3, lorsque Marianna est enfin libérée dans la nature et voit la couleur des fleurs, des arbres, du ciel, que la question qui se pose est : gagne-t-elle un nouveau savoir, soit la connaissance que le ciel est *bleu* de *cette façon-là* ?

Selon Nida-Rümelin, cette expérience de pensée nous permet de mieux saisir les deux étapes du progrès épistémique de Marianna : au moment t_2, lorsqu'elle voit des couleurs arbitrairement distribuées aux objets-artefacts, elle peut former des nouveaux concepts, du genre de ceux qu'on appelle des « concepts phénoménaux » à propos des différents types d'expériences chromatiques. L'acquisition de ces concepts lui permet de formuler des nouvelles questions, ainsi que de formuler des croyances (éventuellement fausses) sur la perception visuelles des autres individus. C'est uniquement au moment t_3 qu'elle acquiert le type de connaissance qui est visée par l'argument initial de Jackson, c'est-à-dire celui concernant l'expérience chromatique des autres sujets.

Une fois que l'expérience de Marianna est détaillée ainsi, il devient évident que l'expression Nagelienne « connaître l'effet-que-cela-fait » n'arrive pas à capturer le

progrès épistémique que réalise notre héroïne à la fin de sa captivité, au temps t_3, lorsque Mary et Marianna acquièrent un certain type de croyance, plus précisément la croyance *que* le ciel apparaît bleu aux autres individus dotés d'une vue normale. Nous sommes donc en présence de ce que Nida-Rümelin[134] appelle une *croyance phénoménale* qui est à son tour attachée à un concept phénoménal, et qui se distingue des croyances *non-phénoménales*. Revenons au temps t_2 dans l'expérience de Marianna, et supposons que lorsqu'on lui montre les quatre jetons et qu'on lui demande d'indiquer quelle est la couleur du ciel telle qu'elle apparaît aux autres individus dotés d'une vue normale, Marianna nous indique le jeton rouge. Dans ce cas, sa situation cognitive peut être décrite ainsi :

(1)Marianna croit *non-phénoménalement*($_{np}$) que le ciel apparait *bleu* aux autres humains dotés d'une vue normale. (Car elle l'a appris en lisant des traités et des manuels.)

(2)Marianna croit *phénoménalement*($_p$) que le ciel apparaît *rouge* aux autres humains dotés d'une vue normale. (Car, une fois qu'elle a été confrontée aux jetons colorés, elle en a acquis l'expérience phénoménale des couleurs, mais elle entretient la croyance fausse selon laquelle la couleur du ciel telle qu'elle apparait aux autres sujets est la même que celle du jeton rouge, jeton qu'elle appelle, à tort, « bleu »).

Pour formuler la même idée autrement, Marianna *ne sait pas*, au moment t_2 que le ciel apparaît bleu$_p$ (c'est-à-dire bleu phénoménalement) aux individus dotés d'une vue normale. Mais il y a également un sens dans lequel Marianna *sait* à $t_2$que le ciel apparaît bleu$_{np}$ (c'est-à-dire bleu non-phénoménalement) aux individus dotés d'une vision normale.

Lorsqu'enfin, au moment t_3 Marianna est libérée à la campagne, au milieu d'un paysage, elle acquiert la connaissance selon laquelle le ciel apparaît bleu$_p$ aux yeux des personnes dotées d'une vue normale. Selon Nida-Rümelin[135], cette distinction nous permet de tirer une conclusion importante quant aux implications du physicalisme: il y a un certain type de connaissance à propos des états phénoménaux qui est accessible uniquement à un sujet connaissant qui « subit » l'état phénoménal en question de manière personnelle, autrement dit uniquement s'il y en a fait l'expérience directe,

[134] (M. Nida-Rümelin, 1995/2004 : 244-5)
[135] *Ibid.*, p. 260

immédiate. Par conséquent, conclue-t-elle, une description d'un être conscient sur la base d'une terminologie purement physicaliste se révèle *épistémiquement incomplète* : car un sujet rationnel capable de comprendre une description physicaliste à n'importe quel niveau de complexité, peut, néanmoins, omettre certains éléments essentiels concernant la connaissance qu'en éprouvent d'autres sujets rationnels conscients, éléments qui, en principe, ne peuvent pas être communiqués au sujet en question par *aucune* description physicaliste possible, et cela en dépit de la complétude et l'exhaustivité de l'information physique mise à la disposition du sujet connaissant.

Nous voyons donc, pour résumer, que les controverses suscitées autour de la stratégie des concepts phénoménaux ont orienté le débat dans une direction fort intéressante. Signalons seulement, pour terminer cette section dédiée à l'examen des interrogations de (Q1) à (Q6)[136], que d'autres voies d'approche critiques ont été avancées comme réponse au défi lancé par l'expérience de pensée de Jackson. Dans le paragraphe suivant, nous allons donc résumer un des plus influentes positions critiques qui attaquent le bien-fondé de l'argument lui-même.

5.7 Rejeter l'intuition : doutes sur le bien-fondé de l'argument

En début de la section 5.3. nous avons évoqué, à la suite de Van Gulick, une liste[137] de possibles interrogations tournant autour de l'argument de Jackson. Certains auteurs, cependant, ont soulevé des critiques qui contournent cette liste, pour ainsi dire. En effet, il est possible d'attaquer l'argument de Jackson à ses racines mêmes, et de remettre ainsi en cause l'intuition même sur laquelle repose l'expérience de pensée de Mary. Un des plus influents arguments de cette sorte a été présenté par D. Dennett, argument généralement connu sous le nom de « *la ruse de la banane bleue* ».

[136] Pour mémoire, rappelons ici les questions en cause :
(Q)1 « Est-ce que Mary apprend un fait *nouveau* ?
(Q)2 Est-ce que cette nouvelle acquisition est-elle d'ordre *factuel* ou *non-factuel* ?
(Q)3 Est-ce qu'elle apprend seulement une connaissance pratique de type « *savoir faire* » ?
(Q)4 Acquiert-elle une simple connaissance par *accointance* ?
(Q)5 Dans l'hypothèse où elle acquiert effectivement un savoir factuel, s'agit-il d'un fait *radicalement nouveau*, inédite, ou d'un fait déjà connu, déguisé seulement en habits nouveaux ? » (Voir R. Van Gulick, 2004)

[137] Cf. infra.

5.7.1 La banane bleue

D. Dennett[138] entame son offensive contre l'argument de Jackson par l'observation que celui-ci tire sa force d'une interprétation naïve et erronée de l'intuition de base qui sert de point de départ à sa croisade antiphysicaliste. Selon Dennett, si la captive Mary était *réellement* en possession d'une connaissance exhaustive et complète de toutes les informations physiques concernant la perception visuelle des couleurs, alors il serait faux d'affirmer qu'elle sera radicalement « surprise » lorsqu'elle sera confrontée pour la première fois à un rubis, par exemple. L'argument de Jackson, dit Dennett, repose sur un tour de passe-passe que l'on peut balayer d'un simple revers de main, une fois qu'on l'a démasqué et épinglé comme une simple naïveté du sens commun. Pour illustrer sa position, Dennett met en avant une expérience de pensée alternative, destinée à démasquer la facilité avec laquelle on peut se laisser entraînés et abusés par le soi disant caractère évident de l'intuition sur laquelle repose le KA.

« Supposons, écrit Dennett, qu'un beau jour, les geôliers de Mary ont décidé qu'il était temps qu'elle voie des couleurs. Pour lui faire un tour, ils ont préparé une banane bleu clair pour la présenter comme sa toute première expérience de couleur. Mary a jeté un coup d'œil et a dit : ''Hé là ! Vous avez essayé de me jouer un tour ! Les bananes sont jaunes, or celle-ci est bleue ! '' Ses geôliers sont restés médusés. Comment a-t-elle fait ? ''C'est bien simple, a-t-elle répondu. Vous devez vous rappeler que je connais tout – absolument tout – ce qui peut être connu sur les causes et effets physiques de la vision chromatique. Et donc bien sûr avant que vous n'apportiez la banane, j'avais déjà écrit, en détaillant par le menu, quelle impression physique ferait exactement un objet jaune ou un objet bleu (ou un objet vert, etc.) sur mon système nerveux. Je savais donc déjà exactement quelles pensées j'aurais (parce que, après tout, la 'simple disposition' à penser ceci ou cela n'est pas l'un de vos fameux *qualia*, si ?). Je n'ai pas été le moins du monde surprise par mon expérience de bleu (ce qui m'a surprise c'est que vous m'ayez joué un tour aussi médiocre). Je me rends compte que c'est difficile pour vous d'imaginer que je puisse en savoir tant sur mes dispositions réactives que la manière dont le bleu m'a affectée n'ait pas été pour moi une surprise. Bien sûr, c'est difficile

[138] D. Dennett, D. (1991), *Consciousness Explained*, Boston: Little, Brown and Company. Trad. fr. P. Engel, *La Conscience expliquée*,(1993), Paris, Odile Jacob.

pour vous à imaginer. C'est difficile pour n'importe qui d'imaginer les conséquences que tirera quelqu'un qui connaît absolument tout sur le physique de quelque chose que ce soit ! »[139]

Dans ce passage, Dennett montre que si Mary possédait effectivement *toutes* les informations sur les couleurs, elle serait en mesure de les identifier correctement dès sa sortie de prison, même si l'on tente de l'induire en erreur. Ce qui veut dire que le cas 'Mary' est scruté d'un mauvais angle dès le début, car la facilité avec laquelle nous succombons devant l'intuition que Mary *apprend* quelque chose d'authentiquement nouveau, ne prouve *stricto senso* rien, excepté notre grande faiblesse. Le caractère manifestement flagrant de l'intuition sur laquelle repose l'argument de Jackson, insiste Dennett[140], ne doit pas nous séduire trop facilement, car nous ne disposons d'aucune preuve solide à l'appui de la thèse selon laquelle Mary acquiert la connaissance d'un nouveau fait propositionnel authentique. Par conséquent, il faut abandonner l'intuition naïve selon laquelle il existerait une propriété phénoménale intrinsèque de l'expérience.

Dennett[141] reconnaît à son tour le caractère hautement contre-intuitif de sa position. Mais le fait même d'aller à contre-courant, insiste-t-il, prouve que sa propre conception est sur la bonne voie. L'exemple paradigmatique de Copernic en est une excellente illustration, insiste Dennett. C'est à la philosophie « pure » que l'on doit imputer la méfiance envers les recherches scientifiques qui contredisent, parfois d'une manière violente, les préjugés du sens commun.

5.8 Jackson change d'avis

Vers la fin des années 90, dans un ouvrage publié en 1998[142], Jackson a surpris la communauté philosophique en faisant volte-face, rendant ainsi publique sa conversion

[139] (D. Dennett, 1991/1993: 495-6)
[140] Plus tard, dans l'article 'What RoboMary Knows'[140], Dennett réitèrera sa position, en conduisant une expérience de pensée analogue à celle de Jackson, mais impliquant cette fois-ci un robot sophistiquée – RoboMary. Voir D. Dennett, (2007), 'What RoboMary Knows' *in* T. Alter & S. Walter (eds.). (2007), op. cit. Pour une contre-réplique, voir M. Beaton, (2005), 'What RoboDennett Still Doesn't Know', *Journal of Consciousness Studies*, 12(12): 3-25.
[141] (D. Dennett, 2007:30)
[142] F. Jackson, (1998), *From Metaphysics to Ethics: A Defense of Conceptual Analysis*, Oxford: Clarendon. Voir aussi F. Jackson (1998), 'Postscript on *qualia*', *in* F. Jackson, *Mind, Method and Conditionals*, London: Routledge.

au physicalisme. Quelques années plus tard, en 2003[143], Jackson développe avec plus de précision sa nouvelle prise de position. Rappelons-nous que l'objectif du KA est de réfuter le physicalisme. L'argument établit dans un premier temps l'existence de certains faits phénoménaux non-déductibles *a priori* de l'ensemble complet de vérités physiques, pour inférer, à partir de cette base, l'échec du physicalisme. Selon le « deuxième » Jackson, la conclusion de l'argument initial de Mary est erronée, et l'argument ne constitue pas une menace décisive pour le physicalisme. Dans ce qui suit, nous allons brièvement résumer la position qu'il a développée *après* son changement d'avis, tout en signalant que la discussion concernant le versant ontologique du physicalisme *a priori* tel qu'il a été défendu par Jackson dans son livre de 1998 sera examinée séparément et plus amplement dans le chapitre 7.

5.8.1 La Théorie Représentationnelle de la Conscience Phénoménale

Le rejet de Jackson de son propre argument se greffe sur son adhésion à la Théorie Représentationnelle de la Conscience, parfois appelé Théorie de l'Expérience Sensorielle Consciente. Avec ce revirement de situation, le « deuxième » Jackson mise, comme le note Van Gulick[144], sur trois éléments : (1) le rejet de l'argument de la connaissance ; (2) l'adoption de la théorie représentationnelle de la conscience ; et (3) l'idée selon laquelle seule celle-ci serait capable de miner sérieusement le KA[145]. Le point névralgique de l'affaire se situe de toute évidence au niveau du troisième point (3), qui se ramifie à son tour en deux questions sous-jacentes : (3_A): est-ce que la théorie représentationnelle de la conscience offre une justification solide pour réfuter le KA ? et (3_B) : y-a-t-il d'autres voies de réfutation de l'argument de la connaissance qui ne reposent par sur la théorie représentationnelle de la conscience?

Jackson le « physicaliste converti» propose de répondre par « oui » à la première question et par « non » à la deuxième, mais comme l'observe judicieusement Van Gulick, les deux réponses sont discutables. Beaucoup d'objections contre le KA n'ont

[143] F. Jackson (2003), 'Mind and Illusion', *in* A. O'Hear (ed.), *Minds and Persons: Royal Institute of Philosophy Supplement* 53: 251–71. Cambridge: Cambridge University Press. Réédité *in* Ludlow et al., (2004), op. cit. (Nous utilisons ici cette dernière édition)

[144] R. Van Gulick, (2008), 'Jackson's Change of Mind: Representationalism, a priorism and the Knowledge Argument' in I. Ravenscroft (ed.), *Minds, Ethics, and Conditionals: Themes form the Philosophy of F. Jackson*, op. cit.

[145] (R. Van Gulick, 2008 : 189)

pas besoin de faire appel à une théorie représentationnelle de la conscience. Bien que le représentationalisme soit une position habilement défendue dans le paysage philosophique contemporain, il est loin d'être inconditionnellement accepté. Son talon d'Achille est la thèse selon laquelle les états sensoriels conscients n'ont pas de propriétés mentales autres que leurs propriétés intentionnelles et représentationnelles. Sur quoi se base-t-elle l'alliance entre l'argument de la connaissance et le représentationalisme ?

Le cœur du raisonnement du « deuxième » Jackson est constitué d'une conjonction des thèses suivantes: (i) *si* tous les faits représentationnels et intentionnels sont entièrement déductibles à partir des faits microphysiques, et (ii) *si* les expériences sensorielles, comme par exemple « voir du rouge », n'ont pas d'autres propriétés mentales (à part leurs propriétés représentationnelles et intentionnelles), alors il s'ensuit automatiquement que l'on peut déduire tous les faits mentaux et toutes les propositions concernant l'expérience du rouge à partir des faits microphysiques. La théorie représentationnelle s'appuie sur la thèse (ii), mais il est clair que l'argument nécessite des appuis supplémentaires pour la thèse (i).

Selon la synthèse qu'en fournit T. Alter[146], la position de Jackson le « physicaliste » peut être retranscrite en une inférence à deux volets, (J1) et (J2). Commençons avec le premier volet :

(J1) Tous les faits concernant le caractère phénoménal de l'expérience des couleurs sont des faits concernant le caractère représentationnel de ces mêmes expériences.

En faisant appel aux explications naturalistes avancées dans la littérature de spécialité contemporaine, et qui définissent le contenu intentionnel respectivement en termes de covariance, liens causaux ou fonctions téléologiques, Jackson[147] présuppose que l'analyse mise à notre disposition par ce type de théories véhicule des thèses qui

[146] T. Alter, (2007), 'Does representationalism undermine the knowledge argument?' in T. Alter & S. Walter (eds.), op. cit., pp. 65–67.

[147] "Seeing red is being in a certain kind of representational state on this account. The project of finding an analysis of representation is not an easy one—to put it mildly. But it is a standard item on the philosophical agenda and the answers that have been, or are likely to be, canvassed are all answers that would allow the fact of representation to follow *a priori* from the physical account of what our world is like. They are accounts that talk
of co-variation, causal connections of various kinds, selectional histories, and the like, and accounts made from these kinds of ingredients are ones that might be determined *a priori* by the purely physical." (F. Jackson, 2003/2004: 432)

peuvent être déduites (du moins *en principe*) à partir de faits microphysiques. Bref, ce type d'analyse naturaliste assure le pont inférrentiel nécessaire à la déduction des faits intentionnels en partant des faits physiques. Par conséquent, cette adhésion au physicalisme lui permet d'avancer une explication en termes physicalistes du caractère représentationnel de l'expérience, ce qui lui permet d'énoncer le deuxième volet de son argument :

(J2) Sans quitter sa prison noire et blanche, et étant en possession de toute l'information physique concernant l'expérience perceptuelle des couleurs, Mary peut déduire tous les faits relatifs au caractère représentationnel de l'expérience des couleurs.

Conjointement, les prémisses (J1) et (J2) mènent à la conclusion selon laquelle Mary peut déduire tous les faits relatifs au caractère phénoménal de l'expérience qui accompagne la perception visuelle des couleurs à partir des faits microphysiques. Si la prisonnière est capable de cette déduction, alors il s'ensuit que l'argument de la connaissance faillit. Le raisonnement que nous venons d'esquisser porte le nom de « l'argument du représentationalisme ». Chacune des prémisses (J1) et (J2) est étayée à son tour par un argument distinct : l'appui de (J1), Jackson fait appel à la notion d'« ineffabilité »[148], habituellement présentée sous la forme d'une thèse concernant l'introspection et l'attention introspective qui s'y attache, c'est-à-dire à l'idée selon laquelle on ne peut pas accéder au caractère phénoménal de l'expérience de quelqu'un sans accéder inévitablement à ce que ces mêmes expériences *représentent*[149]. Jackson prend comme point de départ la définition de la « translucidité » de l'ineffable chez G. E. Moore[150] selon laquelle le caractère qualitatif de l'expérience *est* le caractère de l'objet putatif de l'expérience. La « rougeur » des sensations de "rouge" *est* la rougeur présumée de l'objet que l'on voit.

Selon Jackson, la thèse de la « translucidité » ou de l'ineffabilité peut être interprétée sous deux modes distincts :

---

[148] « The Diaphanousness »

[149] Pour une présentation détaillée de ce type d'introspection voir M. Tye, (2000), 'Knowing What it is Like: The Ability Hypothesis and the Knowledge Argument', *in* M. Tye, *Consciousness, Color, and Content*, Cambridge, MA: MIT Press.

[150] "The moment we try to fix our attention upon consciousness and to see *what*, distinctly, it is, it seems to vanish: it seems as if we had before us a mere emptiness . . . as if it were diaphanous" (G. E. Moore, 1903: 25)

(i) en premier lieu, il s'agit de remodeler une idée déjà implémentée dans la philosophie traditionnelle sous la forme de la théorie des *sense data* : l'expérience est constituée d'un acte d'attention éveillée dirigée vers un objet ou vers un *sense datum* qui supporte et sous-tend les qualités en cause. Par exemple, les défenseurs de l'argument de la connaissance affirment que lorsque la captive Mary quitte enfin sa prison et voit pour la première fois une rose rouge, elle apprend quelque chose sur la propriété phénoménale de « rougeur », c'est-à-dire une propriété qui détermine le caractère phénoménal de l'acte de *voir* du rouge. A cela, les partisans de la théorie des « *sense data* » ajoutent une clause supplémentaire, à savoir l'idée selon laquelle la propriété phénoménale de « rougeur » *est* la propriété instanciée d'un certain type d'objet mental ou de « donné » sensoriel vers lequel l'expérience est dirigée. Cela revient à dire qu'il existe un fait de type « l'expérience E est phénoménalement rouge ».

(ii) Une deuxième option, préférée par Jackson, consiste à dire que voir du rouge n'est qu'un état représentationnel. Etant donné que l'acte de voir du rouge ne possède pas une propriété phénoménale distincte de « rougeur », il est faux de soutenir qu'il existerait des faits de type « l'expérience *E* est phénoménalement rouge». Selon Jackson, l'attribut représentationnel détermine, *in extenso*, la nature de l'expérience des couleurs. Une fois ces bases établies, il est facile d'en inférer la thèse (J1).

Passons à présent au raisonnement qui mène à la thèse (J2). Rappelons-nous que selon cette dernière, Mary, durant sa captivité, peut déduire tous les faits relatifs au caractère représentationnel de l'expérience des couleurs en partant des faits exclusivement physiques. A l'appui de (J2) Jackson fait appel à l'intervention providentielle des théories scientifiques actuelles. Mais, comme le souligne fort bien T. Alter, ces théories explicatives couvrent un terrain beaucoup plus vaste que celui visé initialement par le représentationalisme de Jackson, car elles se réfèrent à la nature de la représentation mentale en général, visant à expliquer, par exemple, comment les états mentaux tels les croyances arrivent à représenter le monde comme étant tel ou tel, autrement dit, comment un état mental peut avoir tel ou tel contenu représentationnel déterminé. En revanche, dans le cas particulier de l'expérience des couleurs, le caractère représentationnel comporte non seulement un contenu, mais aussi un certain « ressenti » ou un « sentir » d'ordre phénoménal. Par conséquent, pour que l'argumentation en faveur de (J2) soit complète et satisfaisante, il faudrait prouver que

tous les faits concernant la représentation phénoménale seraient déductibles *a priori* des faits physiques.

Dans son article 'Mind and Illusion', Jackson[151] identifie cinq attributs caractérisant les situations où l'expérience sensorielle représente les choses comme étant 'ainsi-et-ainsi' (« such and such ») : (a) premièrement, ce type de représentation s'avère très riche, car l'expérience visuelle nous permet d'accéder aux choses telles qu'elles sont ici et maintenant non seulement sous le mode de la couleur, mais aussi sous celui des formes, de l'orientation, de la localisation, etc.; (b) deuxièmement, elle n'est pas seulement riche, mais aussi inextricablement riche, dans le sens où il est impossible de séparer ces aspects les uns des autres. (c) Troisièmement, la représentation est immédiate. (d) Quatrièmement, le contenu semble incorporer un élément causal : la perception, par exemple, nous présente le monde en train d'interagir avec nous. (e) Enfin, l'expérience sensorielle joue un rôle fonctionnel déterminé.

Selon Jackson, en cumulant ces cinq caractéristiques, non seulement on parvient à expliquer la nature de la représentation phénoménale, mais surtout, l'on obtient la phénoménologie *gratis*, en prime[152]. Par ailleurs, insiste Jackson, tous les faits concernant ces caractéristiques sont déductibles *a priori* des faits physiques. Pour résumer, disons que l'inférence qui mène à (J2) comporte trois étapes stipulant que:

(1) la représentation mentale en général est explicable en termes physiques : tous les faits relatifs à la représentation en général sont déductibles *a priori* des faits physiques, *via* les théories scientifiques actuelles.

(2) l'analyse en cinq points susmentionnée constitue une explication de la représentation phénoménale : la façon dont l'expérience des couleurs apparaît phénoménalement, le « ressenti » qui l'accompagne, tiennent de l'ordre de l'immédiateté, à laquelle s'ajoute l'élément causal présent dans le contenu.

(3) toutes les cinq caractéristiques susmentionnées sont explicables en termes physiques. De façon analogue, tous les faits relatifs à ces cinq attributs sont déductibles *a priori* des faits physiques.

Jackson soutient que si le physicalisme est vrai, alors tous les faits relatifs à la conscience doivent être déductibles de façon *a priori* des faits physiques. Mais ce type

[151] (F. Jackson 2003/2004: 436-8)
[152] "We get the phenomenology for free". (F. Jackson 2003/2004)

de déductibilité doit affronter le défi contenu dans l'intuition manifeste selon laquelle Mary apprend effectivement des faits nouveaux lorsqu'elle quitte sa prison. A la différence du « premier » Jackson, le deuxième « Jackson » accepte la thèse de la déductibilité *a priori*, rejette l'intuition selon laquelle la captive apprend des faits nouveaux, et soutient que cette intuition repose sur une conception erronée de l'expérience sensorielle qui devrait être remplacée par une théorie représentationaliste, i.e. la position selon laquelle les états phénoménaux ne sont *que* des états représentationnels.

5.8.2 Objections anti-représentationalistes

On a tôt fait de remarquer que l'édifice représentationaliste présente plusieurs failles. En premier lieu, il n'est pas du tout évident de dire en quelle mesure *tous* les faits déterminants ou significatifs seraient déductibles des faits microphysiques[153]. A la différence des faits causaux et des faits covariants, le cas des faits téléologiques supporte mal la thèse de la déductibilité à partir des faits microphysiques. En deuxième lieu, les théories naturalistes que l'on retrouve dans les débats contemporains sont loin d'être aussi algorithmiquement précises que le présuppose le raisonnement de Jackson : il est rare que l'on trouve des formules véhiculant une mécanique computationnelle rigoureuse des contenus intentionnels. Par exemple, on peut identifier, à la limite, une trame à l'intérieur de laquelle une interprétation puisse s'épanouir. Mais il serait tout simplement faux de présumer que cette interprétation soit une simple affaire de computation ou de déduction. Jackson peut se réfugier, en dernier recours, derrière l'hypothèse selon laquelle ce manque n'est en réalité qu'une étape provisoire de l'évolution des sciences naturelles. Même si tel était le cas, note Van Gulick[154], on ne doit pas cautionner avec trop de facilité la portée de la confiance dans une telle théorie idéale.

Enfin, en troisième lieu – et c'est ici un point important, insiste Van Gulick – il se peut que les théories naturalistes contemporaines omettent dans leurs explications un certain type de faits représentationnels dont l'importance reste cruciale aux yeux des partisans du KA, à savoir les faits concernant la représentation phénoménale. Jackson

[153] (D. Papineau, 2008)
[154] (R. Van Gulick, 2008 :209)

admet qu'il existe une différence importante entre représenter un contenu dans la *pensée*, et représenter phénoménalement un contenu dans une *expérience visuelle* : *penser qu*'un rubis est rouge n'est pas la même chose que *voir qu*'un rubis est rouge. La structure et l'organisation de la conscience phénoménale joue un rôle déterminant dans la caractérisation du contenu intentionnel des états conscients, et pourtant aucun des courants naturalistes dominants de l'époque contemporaine n'est en mesure d'offrir une explication satisfaisante.

Selon Jackson, s'il s'avère que certains faits/processus/entités de niveau fondamental *sont* des conditions suffisantes pour causer l'apparition des entités/faits/processus à un niveau supérieur, il s'ensuit qu'il doit y avoir (du moins en principe) une relation de déductibilité *a priori* du niveau supérieur à partir du niveau immédiatement inférieur. Comme le note judicieusement Van Gulick[155], chez Jackson, le lien métaphysique de réalisabilité entre les niveaux de réalité est en lui-même une garantie du lien épistémologique de la déductibilité *a priori*. Nous aurons l'occasion de traiter plus amplement, dans le chapitre 7, le sujet délicat que représente la question de la jointure entre la nécessité métaphysique et l'apriorité épistémique dans le physicalisme *a priori* défendu par le « deuxième » Jackson. Nous aurons ainsi l'occasion d'y repérer un nouveau et surprenant « saut conceptuel » complétant ainsi la famille de ceux que nous avons déjà signalés au sein du physicalisme sémantique. Pour le but de cette section, il suffit de signaler que beaucoup de physicalistes contemporains considèrent la thèse de la déductibilité *a priori* de Jackson comme excessive. Or, Jackson considère que sans elle, il devient difficile de défendre la réalité des entités d'ordre supérieur. Car si leur nature ou existence ne peut pas être déduite de la totalité des faits microphysiques, alors le physicaliste sera amené tôt ou tard à leur refuser la réalité tout court.

Plus tard, Jackson[156]définira le physicalisme, ou mieux dit ce qu'il appelle le « physicalisme primitif », ou le « physicalisme dépouillé » (« bare physicalism »), la thèse selon laquelle le monde est la somme exacte de ce qui est requis afin d'en rendre sa description physique vraie dans chaque détail, et *rien de plus*, c'est-à-dire : les entités qui ne peuvent pas être dérivées *a priori* à partir d'une description physique ne peuvent pas être considérées comme vraies. Pour dire les choses autrement, la

[155] *Ibid.*
[156] (F. Jackson, 2003)

déductibilité des faits intentionnels et représentationnels à partir des faits microphysiques est juste un cas spécial d'une condition plus générale relative à une clause d'intelligibilité du physicalisme en tant que doctrine métaphysique selon laquelle le réel est ce qui est physiquement réalisé.

Pour utiliser un slogan, « pas de réalisation sans déduction *a priori* » ou « pas de détermination métaphysique sans dérivation logique »[157]. Nous touchons ici à un sujet fort complexe dont nous aurons l'occasion de nous entretenir plus en détail dans le chapitre 7, lorsque nous aurons en main toutes les cartes nécessaires afin d'examiner le mariage conceptuel entre le versant ontologique du physicalisme et le versant logique de l'aprioricité. Il est important néanmoins de résumer ici les conséquences de cette prise de position sur la portée du KA dans l'affaire du physicalisme.

Le fait que Mary apprend *quelque chose* lors de sa libération constitue une menace pour le physicalisme uniquement si la vérité de ce dernier implique l'existence d'une relation de déductibilité *a priori* entre le niveau microphysique et le niveau des entités/états/processus physiquement réalisables. Mais si l'on pense le physicalisme en termes d'un modèle réductionniste greffé sur le tronc empiriste de l'unité des sciences utilisant des lois-ponts conjugués à une dérivabilité théorique, alors la vérité du physicalisme ne permettrait plus à Mary d'inférer la totalité des faits physiquement réalisés. Or, beaucoup de physicalistes, et tout spécialement les physicalistes non-réductionnistes, rejettent ce type d'approche. Bien que Jackson insiste sur le fait que la déductibilité doit pouvoir opérer *en principe*, il est relativement facile de tourner le dos à cette position et revendiquer à se focaliser sur la pratique réelle des sujets humains.

Jackson traite le physicalisme en termes de relations logiques, mais l'argument de Mary présuppose un traitement épistémologique. Après tout, c'est la prémisse concernant ce que Mary *ne* peut *pas* connaître durant sa captivité qui constitue le point fondamental dans le rejet du physicalisme. La véritable ambigüité réside dans l'emploi du concept d'*a priori* dans la caractérisation des rapports entre le mental et le physique, car l'apriorité est, comme le remarque à juste titre Van Gulick[158], un attribut épistémique, et non pas logique. On peut, à la limite, remplacer le syntagme «implication *a priori* » par la locution « implication logique *connue* de façon *a priori* », mais ce faisant, on prend le risque de faire un amalgame entre les versants logiques et

[157] (R. Van Gulick, 2008: 210)
[158] *Ibid*.

épistémiques de l'affaire. Par exemple, si un certain ensemble de propositions découle logiquement d'un autre ensemble de propositions, il ne s'ensuit pas automatiquement que n'importe quel agent cognitif soit capable d'effectuer la déduction logique respective, et Jackson lui-même est loin de soutenir une telle affirmation. Mais si l'on abandonne le transfert automatique entre le niveau logique vers celui épistémique, on comprend mal en quoi le manque de savoir dont souffre Mary dans sa prison devrait concerner les affaires du physicalisme. L'intuition sur laquelle repose l'argument de la connaissance n'a pas besoin d'aller si loin, c'est-à-dire vers la dépendance logique du mental à partir de physique.

Par conséquent, le point clé du conflit qui sous-tend le cas 'Mary' pourrait ne pas se traduire en un débat opposant défenseurs et détracteurs du représentationalisme, mais plutôt en un débat opposant les défenseurs du physicalisme *a priori* aux adeptes du physicalisme non-réductionniste. Car c'est l'adhésion à la thèse de la déductibilité *a priori* qui sert de ciment à la thèse selon laquelle durant sa captivité Mary ne passe à côté d'aucune connaissance propositionnelle relative à la perception visuelle des couleurs. Selon Van Gulick, on peut pousser l'analyse encore plus loin et observer que les tenants du physicalisme *a priori* n'ont même pas besoin d'adhérer au représentationalisme.

Les défenseurs de l'argument de la connaissance peuvent riposter en s'attaquant soit à la définition même que Jackson donne du représentationalisme, soit en remettant en cause l'argument qui le soutient. Selon Alter[159], il est possible de saper une des thèses en faveur de (J2), à savoir l'idée que Mary pourrait déduire de façon *a priori* la nature représentationnelle de l'expérience des couleurs en partant des faits exclusivement physiques, et sans quitter sa prison. Le problème, souligne Alter, vient de la façon dont Jackson[160] analyse les cinq caractéristiques de la représentation évoquées plus haut. Admettons que la représentation mentale en général, ainsi que ses cinq attributs soient déductibles *a priori* à partir des faits physiques. Supposons, conséquemment, que Mary accomplit la déduction durant sa captivité. Si l'explication correcte de la manière dont une croyance, par exemple, représente un objet, implique un facteur de nature causale, alors elle connaît tous les faits relatifs concernant non seulement la causalité, mais aussi la façon dont la théorie *s'applique* aux faits

[159] (T. Alter, 2007)
[160] (F. Jackson, 2003/2004: 436-8)

respectifs. Donc, elle connaît tous les faits relatifs à l'inextricable richesse, aux rôles fonctionnels, etc., qui accompagnent l'acte perceptuel de voir du rouge. Mais lorsque Mary quitte la pièce noire et blanche et voit pour la première fois une tomate mûre, la question est : apprend-elle quelque chose de nouveau sur la représentation phénoménale de la couleur « rouge » ?

Il semblerait, observe Alter, que l'intuition selon laquelle on est tenté de répondre par l'affirmative, garde intacte sa force initiale. Comment expliquer cela ? Le savoir complet impliquant la maîtrise parfaite des cinq caractéristiques susmentionnées ne semble pas ébranler, pas plus que son savoir exhaustif des faits physiques et neurobiologiques, la force de l'intuition sur laquelle repose le KA. Cette intuition, dit Alter, demeure tout aussi puissante qu'avant l'introduction du représentationalisme. Bien évidemment, on ne peut pas décrire le gain épistémique que Mary obtient à sa sortie de prison en termes de familiarisation avec certaines propriétés phénoménales si ces dernières ne sont pas instanciées dans son expérience effective. Mais nous avons vu, avec Van Gulick, que l'on dispose d'autres solutions hormis celle-là.

Par exemple, une autre possibilité consiste à adhérer à la position de D. Chalmers[161] lorsqu'il soutient que l'apprentissage de Mary concerne l'acquisition de nouveaux modes de représentations. Une autre option serait de soutenir que voir un coquelicot, par exemple, permet à la captive Mary d'*éliminer* toutes les possibilités épistémiques concernant la façon dont le « rouge » se présente à l'expérience, possibilités qu'elle n'aurait pas pu éliminer avant de quitter sa prison, et cela malgré son savoir exhaustif des faits physiques relatifs au « rouge ». Comme le souligne fort bien Alter, la connaissance des cinq attributs susmentionnés concerne une connaissance à la troisième personne, déductible du vaste répertoire des faits physiques dont Mary est en possession, mais cette prise de position échoue dans son travail de sape de l'intuition de base du KA selon laquelle cette analyse à la troisième personne omet certains aspects fondamentaux relatifs à la nature de l'expérience phénoménale.

Par ailleurs, observe Alter, il se trouve que la description de l'expérience phénoménale en termes de « représentation » affronte les mêmes difficultés intuitives que l'explication de l'expérience phénoménale en termes de « faits physiques », ce qui rendrait la question du bien-fondé du représentationalisme complètement indépendante

[161] D. Chalmers, (2004a), 'Phenomenal Concepts and the Knowledge Argument', in P. Ludlow, D. Stoljar, et Y. Nagasawa, (eds.), op. cit., pp. 269–98

de la question de la validité du KA. Ainsi est-il nécessaire de rappeler que le représentationalisme est compatible non seulement avec le physicalisme (comme le suggérerait l'analyse de Jackson), mais aussi avec l'antiphysicalisme, car les adeptes de ce dernier peuvent accepter sans problèmes le représentationalisme tel qu'il est défini dans (J1) : autrement dit, ils peuvent souscrire à la thèse selon laquelle tous les faits concernant le caractère phénoménal de l'expérience des couleurs concernent leur nature représentationnelle. Sauf qu'à la différence des physicalistes, les avocats du nonphysicalisme affirment que ces modes de représentation phénoménale *ne sont pas* réductibles ou explicables en termes physiques. Mieux encore, ils peuvent même accepter la fameuse thèse de l'ineffabilité ou de la transparence, en arguant qu'en dépit du fait que l'accès au caractère phénoménal de l'expérience d'un sujet passe obligatoirement par l'accès à ce que ses expériences représentent, l'expérience en elle-même comporte des propriétés représentationnelles non-physiques, *distinctes* de ce qu'elles représentent.

Autrement dit, la force du KA provient du fait qu'il reste applicable à toutes les versions du physicalisme, représentationalisme compris. Par conséquent, l'utilisation de ce dernier pour saper le physicalisme s'avère fallacieuse, ce qui nous renvoie à la case départ : trouver une réponse au défi lancé par le KA.

Jackson répond à ces critiques dans un article[162] où il s'évertue à dresser une distinction entre (i) le représentationalisme minimal et (ii) le représentationalisme fort. Le représentationalisme fort est la doctrine selon laquelle le contenu d'une expérience *plus* le fait que cette même expérience représente le contenu comme ayant été obtenu *via* une représentation perceptuelle spécifique, sont les facteurs qui déterminent la nature d'une expérience, et cela sans des résidus supplémentaires. D'autre part, le représentationalisme faible affirme la même chose que sa variante forte, la clause des « résidus » en moins : autrement dit, bien que la nature de l'expérience soit essentiellement celle de représenter, son caractère représentationnel *n*'est *pas* le seul déterminant exclusif de sa nature expérientielle.

La réponse de T. Alter consiste à insister sur le fait que ses objections visent la variante forte de représentationalisme, et non pas sa version faible, car cette dernière est incompatible avec (J1), tandis que la position d'Alter est d'accepter (J1) comme une

[162] F. Jackson, (2007), 'The knowledge argument, diaphonousness, representationalism', *in* T. Alter & S. Walter, (eds.), op. cit., pp. 52–64.

prémisse vraie. Rappelons-nous que la différence entre les deux versions du représentationalisme porte sur le point suivant : la variante faible accepte l'existence des propriétés phénoménales nonreprésentationnelles déterminant partiellement la nature de l'expérience des couleurs, tandis que la variante forte rejette l'existence de ces propriétés. A ce point de la discussion, il est utile de se rappeler que la thèse (J1) exclut également l'existence des propriétés phénoménales nonreprésentationnelles.

Une option pour sortir de l'impasse, suggère Alter, serait d'adhérer à l'idée selon laquelle la nature de l'expérience des couleurs est déterminée au moins partiellement par des propriétés de type représentationnel, comme par exemple, la propriété de représenter la qualité phénoménale de « rougeur » sous un mode particulier, et pas seulement en tant que propriété représentée – « rougeur ». Aucune forme de représentationalisme fort ou même de « super-représentationalisme » ne serait à même de nous mettre à l'abri de cette impasse. Car l'adhésion de Jackson au représentationalisme dans sa version forte l'oblige d'aller jusqu'au bout de la thèse selon laquelle les propriétés représentationnelles jouent un rôle actif dans la caractérisation de la nature de l'expérience des couleurs. Plus spécifiquement, la nature de ce type d'expérience est déterminée non seulement par son contenu, mais aussi par le fait que l'expérience représente le contenu comme étant acquis *via un mode* de représentation perceptuelle spécifique, distinctive. Ce « mode » doit être compris, souligne Alter, non pas comme une propriété représentée, mais en tant que propriété ayant un pouvoir représentationnel propre. En y regardant de plus près, on pourrait identifier ce « mode » comme étant un mode *phénoménal* de représentation. La manière dont Mary acquiert des informations nouvelles concernant l'expérience phénoménale au cours de l'événement « voir un rubis », ce mode phénoménal de représentation *est* une propriété représentationnelle, et non pas une propriété représentée, autrement dit, il s'agit d'une propriété qui concerne *nos* expériences, et non pas les objets-rubis. Après tout, il est parfaitement concevable de stipuler que mon chat et moi ayons des représentations complètement différentes de la qualité phénoménale « rougeur » lorsqu'on regarde simultanément un morceau de bifteck cru, bien que l'on regarde ici et maintenant le même objet : *ce* bifteck.

Poussons l'argument encore plus loin, et faisons preuve d'excès de zèle, en postulant une version « super-forte » du représentationalisme qui impliquerait l'adoption de ce que Chalmers appelle des « propriétés représentationnelles 'pures' »

qui auraient la capacité magique de se représenter elles-mêmes telles qu'elles sont. Dans ce cas, s'interroge Alter, en s'appuyant sur quel critère l'adepte du « super-représentationalisme » pourrait distinguer entre une représentation phénoménale et non-phénoménale ? En effet, il serait obligé de recourir à la thèse selon laquelle certains contenus peuvent être représentés dans l'expérience, sans qu'ils puissent être représentés en l'absence de l'expérience respective. Cette distinction, toutefois, ne met pas en péril la portée du KA. Il est vrai que selon la version « super-forte » du représentationalisme, il nous est interdit de décrire en termes phénoménaux ce que Mary apprend à sa sortie de prison. Mais l'intuition selon laquelle la captive apprend quelque chose d'authentiquement nouveau reste intacte. Dans cette version, l'intuition initiale contenue dans le KA se traduirait dans l'idée selon laquelle Mary apprend des faits nouveaux concernant l'*instanciation* de certaines propriétés représentationnelles « pures ». Mais cette modification ne remet pas en cause, souligne Alter, la force persuasive de l'intuition sur laquelle repose le KA.

Pour terminer, disons que la tentative récente de Jackson de faire face au défi lancé par son propre KA, stratégie reposant sur une combinaison entre le représentationalisme et une description physicaliste de la représentation phénoménale, semble affronter des sérieuses difficultés. Autrement dit, la retranscription du cas 'Mary' en termes représentationalistes laisse le débat au même point qu'avant : car à présent, la question (a) « est-ce que la conscience phénoménale est explicable en termes physiques ? » est remplacée avec la formulation (b) « est-ce que la représentation phénoménale est explicable en termes exclusivement physiques ? »

Il est facile donc de saisir que les enjeux contenus dans la question (b) affrontent le même type de difficultés que ceux soulevés par (a). Les physicalistes restent avec le problème brûlant du KA sur les bras, tandis que les antiphysicalistes doivent affronter leurs propres démons, parmi lesquels la question des rapports entre les états mentaux et les états physiques, qui demeure une sérieuse épine dans leur pied. Une chose est claire : l'élargissement du débat sur le terrain du représentationalisme n'apporte aucune avancée significative dans le dossier métaphysique du cas 'Mary'.

5.8.3 Le 'Pack Jackson' *vs* le 'Pack Physicaliste Non-Réductionniste'

Les physicalistes qui prennent le parti de rejeter le modèle *a priori* du physicalisme défendu par le « deuxième » Jackson peuvent donc affronter les défenseurs de l'argument de la connaissance *sans* faire appel au représentationalisme. Ils peuvent, par exemple, rejeter l'idée selon laquelle les théories physiques représentent un outil cognitif universel capable, même en principe, de représenter et d'expliquer la totalité du réel. Les avocats du physicalisme non-réductionniste[163], par exemple, peuvent insister sur le fait que les théories scientifiques ne sont en réalité que des simples « outils cognitifs » parmi tant d'autres, rendant compte de l'organisation des structures du réel.

En réalité, comme l'observe judicieusement Van Gulick[164], c'est la thèse de la déductibilité *a priori*, sur laquelle nous allons revenir dans le chapitre 7 de cet essai, qui pèse le plus lourd dans la balance lorsqu'il s'agit d'affronter et de saper l'argument de la connaissance, et non pas le 'pack' du représentationalisme. Mais si l'on se penche du côté du matérialisme non-réductionniste, et l'on s'éloigne de la version *a priori* du physicalisme, on s'aperçoit que la théorie représentationnelle de la conscience phénoménale joue un rôle encore moins important contre le KA.

Résumons. Du point de vue de Jackson le « converti », le physicalisme est intrinsèquement lié à la thèse de la déductibilité *a priori*. Il rejette ainsi toute autre forme de nécessitation entre les niveaux du réel qui ne serait pas une déductibilité de type *a priori*. Par conséquent, aucune réfutation de l'argument de la connaissance ne sera possible si l'on n'adopte la prémisse selon laquelle Mary peut déduire, au moins en principe, la totalité des propositions vraies relatives à l'expérience consciente à partir des faits microphysiques. Cette prise de position peut être résumée en un ensemble de propositions que l'on peut appeler, suivant Van Gulick[165], le « **Pack Jackson** » et qui présuppose l'adoption des points suivants :

[163] (R. Van Gulick, 1992; R. Van Gulick, 2002)
[164] (R. Van Gulick, 2008 :214)
[165] *Ibid.*, p. 215.

PACK JACKSON

(a) Une version *a priori* du physicalisme ;

(b) La thèse de la déductibilité : tout fait est en principe déductible des faits microphysiques ;

(c) La thèse du caractère représentationnel de l'expérience sensorielle ;

(d) La non-existence des propriétés phénoménales ;

(e) Le caractère illusoire des intuitions phénoménales ;

(f) Mary, avant sa libération, possède un savoir propositionnel complet ;

(g) Mary acquiert seulement un « savoir-faire »

Mais il est tout à fait possible, souligne Van Gulick, d'envisager une riposte différente au cas 'Mary', riposte bâtie cette fois-ci sur le versant non-réductionniste du physicalisme, et qui réside dans l'adoption des points suivants :

Le 'Pack Non-Réductionniste'[166] :

PACK NON-RÉDUCTIONNISTE

(1) La thèse du physicalisme ontologique interprétée en termes de « réalisation » : tout ce qui existe est physiquement réalisé ;
(2) Rejet du physicalisme réductionniste (les théories physiques ne sont pas un outil exhaustif et universel pour comprendre la nature ultime du réel ;
(3) Le rejet de la thèse de la déductibilité : pas tout ce qui est réel est ultimement déductible à partir du niveau microphysique ;
(4) Adoption d'un modèle pragmatique et contextuel des processus de représentation et de compréhension ;
(5) Interdépendance entre le savoir propositionnel et le savoir-faire ;
(6) Les nouveaux outils que Mary acquiert à sortie de prison lui servent de moyens d'améliorer et agrandir son savoir propositionnel ;
(7) L'existence des faits subjectifs physiquement réalisés ;
(8) L'intuition selon laquelle Mary gagne un savoir propositionnel nouveau est conservée comme étant compatible avec le physicalisme ;
(9) La thèse de la neutralité quant à l'existence des propriétés phénoménales ou au représentationalisme.

5.8.4 Quelques conclusions provisoires

Récapitulons. Dans ce chapitre, nous avons centré notre analyse sur la thèse de l'apprentissage (M_2) dans l'argument de Jackson : à la fin de sa captivité, Mary subit un changement épistémique, elle apprend *quelque chose*. Mais quoi ? Les différentes versions que nous venons de présenter ont tenté, chacune avec leur moyens

[166] *Ibid.*, p. 217.

argumentatifs spécifiques, de répondre à cette question de manière satisfaisante. Le tableau ainsi obtenu peut être synthétisé provisoirement ainsi :

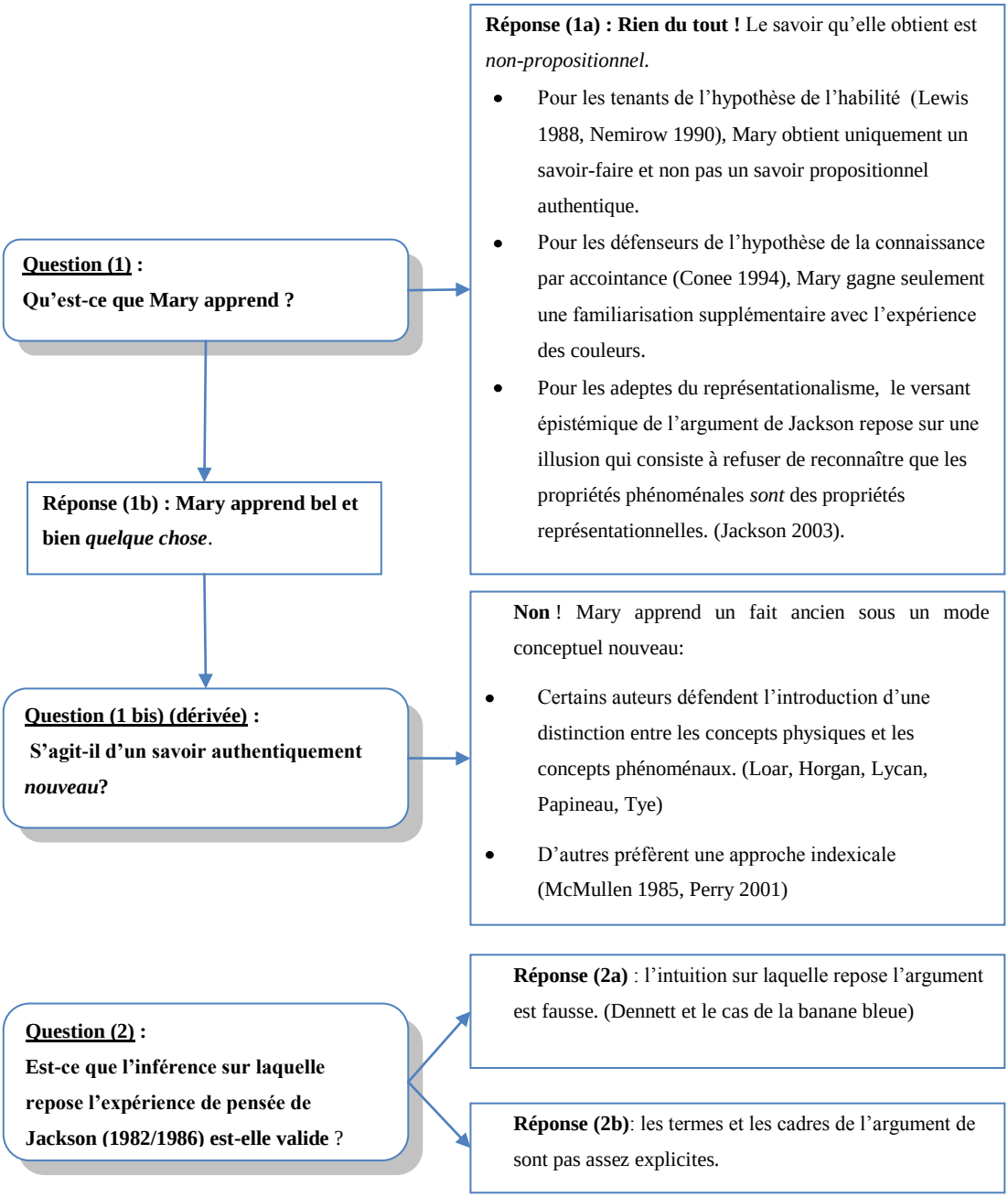


Figure 10: Le KA - Arbre récapitulatif : La Thèse de l'Apprentissage

Nous avons abordé, jusqu'ici, seulement le versant épistémique de l'argument de Jackson. On y a cherché notamment à mettre en lumière les différentes implications du changement épistémique qui s'opère chez Mary à sa sortie de prison. Mais rappelons-nous que la cible fondamentale du KA est de récuser le physicalisme. L'argument de la

connaissance vise à montrer qu'il existe des faits relatifs à la conscience qui ne sont pas déductibles des faits physiques.

Toute tentative physicaliste d'annihiler le « fossé explicatif » devra respecter ce que Levine a appelé la « Contrainte Matérialiste[167]» qui stipule que toute solution au problème du « gouffre » ne doit faire appel à aucune entité ou relation mentales considérées comme étant fondamentales, ou primitives. Conséquemment, toute explication-candidat faisant appel à des éléments explicatifs non-physiques intercalés dans l'architecture argumentative physicaliste ayant pour finalité de dissoudre ou de franchir le gouffre explicatif, sera automatiquement disqualifiée. Dès lors, les tenants du physicalisme doivent se limiter à expliquer en termes exclusivement physiques la manière dont l'expérience qualitative phénoménale est implémentée, incorporée dans des systèmes et mécanismes physiques.

Au début de ce chapitre nous avons insisté sur l'idée que l'intuition initiale de l'argument de la connaissance se prête à une analyse selon trois[168] axes distinctes, à savoir : (M1) la thèse de la connaissance complète stipulant qu'avant de quitter sa prison, Mary connaît *toute l'information physique* relative à la perception visuelle des couleurs. La deuxième est (M2) la *thèse de l'apprentissage* qui vise à établir si Mary apprend un *fait authentiquement nouveau* en quittant sa prison noire et blanche. Enfin, (M3) le troisième axe d'analyse stipule que si (M2) et (M3) sont vraies, alors le fait que Mary apprend lorsqu'elle quitte la prison *ne* peut *pas* être déduit de manière *a priori* (à l'aide du raisonnement seul, sans apport empirique) à partir de l'ensemble complet de vérités physiques qu'elle a à sa disposition durant sa captivité.

Jusqu'ici nous avons examiné le versant de l'apprentissage (M2) de l'argument. Dans ce qui suit, nous allons aborder son versant physique (M1). Ce dernier se ramifie à son tour en deux branches distinctes, articulées une à l'autre par le biais de *la thèse de l'information physique complète*. Deux questions aiguilleront notre enquête dans le chapitre suivant, à savoir :

(i) Dans quel sens peut-on affirmer d'un sujet connaissant S qu'il possède une *connaissance complète* d'un fait *X* ? (sections de 6.1. à 6.3).

[167] "The Materialist Constraint: namely, that no appeal be made in the explanation to any mental property or relation that is basic." (J. Levine, 2007: 150)
[168](T. Alter, 2007)

(ii) Si l'hypothèse selon laquelle S possède un savoir complet est vraie, comment pourrait-on distinguer dès lors le savoir complet de S de l'omniscience divine d'un Dieu tout-puissant ? (sections de 6.4. à 6.5.)

Dans le chapitre suivant, il sera question d'enquêter sur les fondements et la cohérence même de la notion de «savoir complet » afin de s'assurer que l'argument vise le bon « genre » d'omniscience. Pour ce faire, nous allons mettre en scène le seul personnage qui pourrait nous éclairer sur le contenu de cette notion « démiurgique », i.e. Dieu lui-même !

6 La thèse du savoir complet

Selon la position que l'on adopte sur l'échiquier métaphysique du physicalisme, la première prémisse (P1)[1] de l'argument de la connaissance dégage soit une odeur démiurgique de sainteté, soit une odeur diabolique de souffre, si l'on est permis d'utiliser cette expression. Nous avons vu, dans les sections 1.3. et 1.4., que des difficultés considérables surgissent dès qu'il s'agit de se mettre d'accord sur une définition adéquate du concept d'« information *physique* ». Dans cette section nous allons compléter l'examen de (P1) en investiguant le versant de la *complétude* du savoir dont Mary dispose dans son paradis noir et blanc.

6.1 Faits Subjectifs et Faits Objectifs

On dit parfois que les nourrissons se comportent comme s'ils étaient non seulement le centre du monde, mais aussi le monde entier. Pour utiliser une métaphore, leur comportement coïncide avec la description des divinités circulaires de Nicolas de Cues[2], dont le centre est partout et la circonférence nulle part. Avec le temps, les enfants apprennent, parfois d'une manière très brutale et directe, qu'il existe bel et bien un monde extérieur, conditionnant et limitant drastiquement leurs illusions initiales d'universalité. Or, et c'est ici un point crucial, malgré la perte définitive de cette illusion d'universalité omni-englobante, et malgré les obstacles infligés par le monde extérieur, il y a quelque chose qui reste, un résidu pour ainsi dire, de l'ancienne « démiurgie » infantile, quelque chose qui traverse, telle une colonne vertébrale, l'existence de l'individu, et qui lui garantit l'unité et l'identité à travers le temps : il s'agit du point de vue subjectif, expérientiel, à la première personne. Vue de cette perspective, l'expérience subjective reste inéluctablement *le centre* de mon monde intérieur. Je suis, pour reprendre les mots de Locke[3], un être qui a la capacité de se considérer soi-même, en tant que soi-même, une et même chose pensante persistant à

[1] Pour mémoire, rappelons que l'énoncé (P1) stipule que Mary (avant sa libération) connaît *toute* l'information physique concernant les objets et les autres sujets.
[2] N. de Cues, (ca. 1440), *De docta ignorantia*, trad. fr. H. Pasqua, Paris, Bibliothèque Rivages, 2008.
[3] "A being that can consider itself as itself, the same thinking thing, in different times and places." J. Locke, (1690), *An Essay concerning Human Understanding*, 3.27.9.

différents instants et places. L'être humain maintient ainsi une prise de vue *centrée*[4] sur le monde. Conséquemment, le regard qu'il pose sur le réel se révèle intrinsèquement subjectif, expérientiel, « en première personne ». En vain essaierait-il d'échapper à cette « personne intérieure », véritable prisme à travers lequel le sujet perçoit, connaît et agit sur le monde.

En revenant à présent au cas de Mary, les questions que l'on est amené à se poser concernent la manière dont le physicalisme entend s'acquitter de la difficile mission d'incorporer le point de vue de la première personne dans une trame explicative dont l'ambition manifeste reste celle de fixer la sphère légitime de l'objectivité. Bref, le physicalisme est-il compatible avec l'existence des faits subjectifs ? Certains auteurs, comme Th. Nagel[5], ont répondu par « non » et récusé le physicalisme. D'autres, comme Van Gulick[6] et T. Crane[7] soutiennent que l'existence des faits expérientiels réfute le physicalisme uniquement si l'on présume que *tous* les faits physiques doivent être objectifs et saisissables à partir des perspectives neutres et multiples. Il est vrai que les faits physiques « standard » que l'on associe habituellement aux théories physiques à succès se prêtent à un examen épistémique objectif, sans porter, pour ainsi dire, le poids restrictif qu'impose le point de vue subjectif, expérientiel. Mais il est loin d'être évident que *tous* les faits physiques soient de cette nature. Si l'on adhère à l'idée que les faits expérientiels, à la première personne, *sont* des faits réalisés à partir d'un substrat physique fondamental, alors il est possible, argue Van Gulick, de concevoir des propositions et des faits qui soient à la fois physiques *et* subjectifs[8].

Par exemple, si l'on est partisan du physicalisme non-réductionniste, et on privilégie, comme l'auteur de ces lignes est enclin à le faire, un point de vue *pragmatique* dans la conception de la cognition, alors, on sera prêts d'admettre qu'il existe des faits réalisables à partir d'un substrat physique, mais qui demeurent fondamentalement *inanalysables* du point de vue d'une *théorie* physique. Ce qui revient à dire que si l'esprit conscient implique la présence d'une composante essentielle d'auto-compréhension réflexive, il s'ensuit alors que le phénomène de la

---

[4] Voir Barry C. Smith, (2006), 'Consciousness: An Inner View of the Outer World', *Journal of Consciousness Studies*, 13, No. 7–8, pp. 175–86.
[5] (Th. Nagel, 1974)
[6] (R. Van Gulick, 2008:197)
[7] T. Crane, (2003). 'Subjective Facts', In *Real Metaphysics: Essays in Honour of D. H. Mellor*. New York: Routledge.
[8] *Ibid*.

conscience est caractérisé par un ensemble de propriétés inscrutables à partir d'un point de vue externe, en surplomb.

Conséquemment, non seulement l'existence des faits subjectifs *est* compatible avec le physicalisme, mais aussi elle est impliquée nécessairement par une certaine catégorie de physicalisme que l'on pourrait appeler, à la suite de Van Gulick[9], *physicalisme téléo-pragmatique*[10]. Selon ce dernier, la compréhensibilité d'un système est toujours la fonction d'un potentiel investissement pratique, dont le succès présuppose et implique une interaction mutuelle que l'on peut baptiser, selon Van Gulick, « résonance causale », ou « engagement causal harmonieux »[11]. L'aspect intentionnel ou le contenu que l'on associe habituellement à ce type de traitement doit être compris en tant que fonction du profile causal de l'engagement respectif qui munit le sujet connaissant d'un accès particulier au réel, lui permettant d'interagir profitablement avec ce qu'il connaît. Il serait hasardeux de tenter de façonner une structure explicative du mental par le truchement exclusif d'un point de vue externe, à la troisième personne, censé garantir, *en plus* des faits physiques objectifs, la possibilité de calquer le profil causal que l'on associe habituellement aux systèmes complexes autoréflexives, comme le sont par exemple les cerveaux des humains ou des chauves-souris, (bien que physiquement réalisables à partir d'un substrat physique).

Reprenons *da capo* : rappelons-nous que la première prémisse (P1) stipule que Mary (avant sa libération) connaît toute l'information physique concernant la perception visuelle du rouge. La plupart des auteurs sont prêts à concéder (P1) puisqu'il s'agit d'un raisonnement hypothétique. D'une certaine manière, il est relativement facile de contester la thèse de l'omniscience de Mary, compte tenu des limites évidentes de l'entendement humain. Mais s'agissant des limitations essentiellement contingentes, elles n'affectent pas réellement la portée de l'argument, puisqu'il s'agit d'une stipulation *en principe*. Par conséquent, il semblerait qu'il n'existe aucune raison pertinente pour rejeter l'idée contenue dans (P1), à savoir la présupposition que Mary pourrait, du moins *en principe*, connaître toute l'information physique relative à la perception visuelle des couleurs. Cela dit, s'il y a des faits subjectifs dans le sens défini plus haut, à savoir des faits saisissables uniquement d'un point de vue subjectif,

[9] R. Van Gulick, (1985), 'Physicalism and the subjectivity of the mental'. *Philosophical Topics*, 13: 51-70.
[10] (R. Van Gulick, 2008 : 199)
[11] "Causal resonance' or 'causal harmonic engagement" (*Ibid*.)

expérientiel, il devient clair que Mary n'aurait pas pu, *même pas en principe*, connaître cette catégorie de faits. A la question (P) « Est-il réellement possible que Mary (avant sa libération) connaisse toutes les propositions et informations physiques relatives à la perception visuelle du rouge ? », un physicaliste partisan de l'hypothèse des faits subjectifs physiquement réalisables devrait répondre par la négative, et rejeter ainsi, par ricochet, la première prémisse (P1) du KA. Il deviendrait ainsi inutile de se réfugier derrière le statut de « simple stipulation » contenue dans la présomption de l'omniscience cognitive de Mary, car celle-ci reposerait soit sur une contradiction interne dissimulée, soit sur la conjonction de deux conditions qui ne peuvent pas être satisfaites simultanément, à savoir : (i) Mary n'a jamais éprouvé la perception visuelle du rouge et pourtant (ii) elle connaît toute l'information physique à propos de cette même perception. Une possible solution serait d'admettre, avec Y. Nagasawa[12], que durant sa captivité, Mary est en possession de toute l'information physique *théorique*, ou scientifique. Cette option sera explorée dans la section suivante. En outre, s'il s'avère qu'il existe des faits *à la fois physiques et subjectifs*, comme le pense Van Gulick[13], alors un physicaliste qui souscrit à cette thèse peut en toute légitimité rejeter la première prémisse (P1) du KA.

Il est possible également de reformuler (P1) d'une manière encore plus précise : (P1′) « Est-ce que l'acquisition de Mary (lors de sa libération) d'une nouvelle proposition/information (interprétée sous le mode extensif, large), ainsi que d'une meilleure compréhension de sa propre position dans l'espace logique, sont-elles en mesure de réfuter le physicalisme ? » Deux interrogations distinctes sous-tendent (P1′) : en premier lieu, on peut s'interroger dans quel sens les informations et les propositions de niveau supérieur (« higher-level ») visant des propriétés mentales réalisées à partir d'un substrat physique, seraient-elles différentes des informations/propositions concernant le substrat physique sous-tendant leur réalisation. En deuxième lieu, est-ce que le simple fait de savoir toute l'information physique contenue au niveau microphysique fondamental serait suffisant pour connaître, *ipso facto*, la totalité des propositions vraies réalisées aux niveaux supérieurs ? Autrement dit, dans quelle mesure connaître le premier ensemble de vérités (au niveau microphysique) règle le problème de connaître l'ensemble des vérités psychologiques ?

[12] (Y. Nagasawa, 2008)
[13] (R. Van Gulick, 2008 : 200)

Selon l'acception standard[14] du physicalisme (que nous avons déjà évoquée dans le chapitre1), si tous les faits surviennent métaphysiquement sur les faits microphysiques, alors établir la totalité des faits microphysiques qui sous-tendent la structure du monde actuel revient à établir *tous* les autres faits. Pour dire les choses autrement, il ne reste plus de place pour des sous-régions putatives à éliminer. Rappelons-nous, la contrainte imposée par la thèse de la survenance : pas de transformations au niveau du réel macroscopique sans des transformations au niveau microphysique. Il y a des précautions à prendre, cependant, car toute inférence allant depuis l'ontologie des faits vers leur condition épistémique comporte des risques majeurs. Admettons que l'acquisition d'une information nouvelle (considérée dans un sens large) requiert la capacité, pour le sujet connaissant, de se placer avec plus de précision dans l'espace logique des mondes possibles. Or, insiste Van Gulick[15], l'aptitude même qu'a le sujet de se situer soi-même dans l'espace microphysique entraîne un effet surprenant : plus on cherche à préciser et à délimiter la localisation du sujet dans un espace défini, plus on lui bloque l'accès à l'acquisition des propositions/informations nouvelles (considérées sous un mode large, extensif).

De toute évidence, le traitement de cette affaire dépend de manière cruciale de l'interprétation que l'on est disposé à donner à la locution clé « connaître l'emplacement du sujet dans l'espace logique ». Il est possible, par exemple, de définir un monde comme « actuel » en le situant dans l'espace microphysique, sans être en mesure pourtant, d'expliquer la structure de l'espace logique global au sein duquel il est emboîté. Pensons, propose Van Gulick, au cas d'un personnage de bande dessinée qui, égaré, retrouve finalement un panneau où il peut lire « *vous êtes ici* », sans aucune autre information ou carte additionnelle à l'intérieur de laquelle l'indication « vous êtes ici » pourrait dévoiler tout son sens. En déplaçant l'analogie sur le terrain du physicalisme, on s'aperçoit que la thèse de la survenance nous aide effectivement à concevoir *comment* les faits microphysiques sont ancrés dans l'espace microphysique. Mais sans une description adéquate de la structure de ce dernier, la connaissance que le sujet possède de sa propre position à l'intérieur de cet espace restera à jamais incomplète, laissant ainsi une porte ouverte à d'éventuels ectoplasmes, zombies, et d'autres

[14] Pour un examen qui vise s'inscrire en faux contre cette conception standard, voir D. Papineau, (2008), 'Must a Physicalist Be a Microphysicalist?' *in* J. Hohwy & J. Kallestrup (eds.), *Being Reduced: New Essays on Reduction, Explanation, and Causation*, op. cit.
[15] (R. Van Gulick, 2008 : 201)

créatures étranges. En clair, si la maîtrise cognitive de l'espace structural est incomplète, alors l'intelligibilité de l'ensemble de propositions dont le sujet connaissant dispose sera inévitablement parcellaire.

Revenant à présent au cas de Mary, on comprend plus aisément dans quel sens on pourrait affirmer que durant sa captivité notre prisonnière *ne* possède *pas* une connaissance complète de la perception visuelle du rouge, malgré le fait que les aspects absents soient des états de niveau supérieur réalisés à partir d'un substrat physique. Bien plus, on peut même saisir dans quel sens elle omet de saisir la *structure* de l'espace logique, et par voie de conséquence, les coordonnées de son propre emplacement au sein de cet espace. Plus spécifiquement, il lui est impossible de visualiser la structure des relations de similarité ou de dissemblance qui caractérisent les différents étages de l'agencement qui s'étalent entre le niveau microphysique fondamental et le niveau supérieur. Dans la mesure où Mary est inapte d'articuler et de tisser correctement cette toile de relations logiques, sa maîtrise de la structure de l'espace logique, absolument nécessaire à la saisie de la nature des propriétés de niveau supérieur, serait à jamais inadéquate, voire fausse.

Il existe néanmoins une autre voie possible d'approcher l'omniscience présumée de Mary. Il s'agit d'une solution élaborée, d'une manière assez surprenante, à partir d'une comparaison entre l'argument de la connaissance dans la philosophie de l'esprit d'une part, et un argument tiré de la philosophie de la religion analytique d'autre part.

6.2 Omniscience divine et omniscience terrestre

Dans son livre *God and Phenomenal Consciousness*, Y. Nagasawa[16] argue que poussée jusqu'au bout de ses conséquences, la prémisse (P1) stipulant que Mary (avant sa libération) connaît *toute* l'information physique concernant les objets et les autres sujets, peut facilement collapser en une question métaphysique concernant la lisière qui sépare la condition épistémique *humaine* de la condition épistémique d'un *démiurge*. Nagasawa explore les possibilités qu'offre une telle comparaison, ce qui l'amène à défendre une version de *physicalisme nonthéorique*, c'est-à-dire une doctrine selon laquelle le monde, bien que physiquement homogène, n'est pas *théoriquement* homogène. Autrement dit, il existe des faits physiques nonthéoriques, ou mieux dit,

[16] (Y. Nagasawa, 2008)

nonthéoriquement communicables (énonçables) du point de vue de la troisième personne. Par ailleurs, comme nous allons le voir dans le chapitre suivant, l'introduction du concept d'*incommunicabilité théorique* dans les travaux de Nagasawa peut être exploitée d'une manière surprenante dans l'affaire des implications métaphysiques du physicalisme *a priori* et de l'analyse conceptuelle défendues par F. Jackson et D. Chalmers[17] dans leur controversé article 'Conceptual Analysis and Reductive Explanation'.

Nagasawa propose de dresser un parallèle ingénieux entre la structure des arguments de la connaissance d'une part, et deux arguments tirés de la philosophie de la religion d'autre part, à savoir *l' « argument de la possession d'un concept »* ainsi qu'une variante de *l' « argument de la connaissance* de se » développé par P. Grim[18]. L'analyse détaillée de ces comparaisons dépasserait les cadres modestes de notre enquête ici. Nous nous contenterons d'examiner une seule paire de cette analogie, à savoir la comparaison entre le KA et l'argument de la possession du concept. Mais avant même d'aller plus loin, nous avons besoin d'énoncer avec plus de précision les cadres conceptuels qui balisent le duel entre les partisans et les adversaires du physicalisme. Rappelons-nous que le fondement même de l'édifice physicaliste peut être contenu dans la thèse suivante, que l'on peut appeler la *Thèse de la Complétude de la Connaissance Physique* (C):

(C) La connaissance physique complète est une connaissance complète *simpliciter*[19].

D'emblée, l'aspect le plus frappant c'est la décision de cibler des questions ontologiques par le biais d'une caractérisation épistémique. Emprunter ce raccourci renforce effectivement la portée d'un certain type d'arguments que Van Gulick a baptisés « *arguments boomerang* » [20]. La caractéristique distinctive des arguments

[17] D. Chalmers & F. Jackson, (2001), 'Conceptual Analysis and Reductive Explanation', *Philosophical Review* 110 (3):315-61.

[18] P. Grim, (1991), 'There Is No Set of All Truths'. *Analysis* 44: 206–208. Voir aussi P. Grim, (1984), *The Incomplete Universe: Totality, Knowledge and Truth*, Cambridge: MIT Press.

[19] "Physical Knowledge Thesis: Complete physical knowledge is complete knowledge *simpliciter*." (F. Jackson, 1986: 291)

[20] "The distinctive feature of a boomerang argument is that it reaches across to the epistemic/cognitive/conceptual domain of facts about our representation of the world, and then swings back to reach a conclusion in the metaphysical/ontological/factual domain about the nature of reality itself. It moves from facts about how we represent or conceptualize the world to supposed results about the necessary nature of the world itself. More specifically, boomerangs often move from supposed gaps or lacks of links in our representations or concepts of the world to conclusions about objective gaps within the world itself and ontological distinctions between the real things in it". (R. Van Gulick, 2004 : 367)

boomerang est le fait qu'ils visent à atteindre, par le truchement des moyens épistémiques/cognitifs/conceptuels, des *faits* concernant notre *représentation* du monde, pour revenir ensuite à une conclusion concernant la nature métaphysique/ontologique/factuelle du réel en soi. Autrement dit, l'argument boomerang part des faits concernant la manière dont on représente/conceptualise le monde, et finit par légiférer sur la nature des vérités métaphysiquement nécessaires caractérisant le réel en tant que tel. Plus encore, les arguments boomerangs infèrent, à partir des fossés présumés dans la manière dont on représente conceptuellement le monde, vers des conclusions stipulant l'existence des fossés objectifs, réels, impliquant des distinctions ontologiques réelles entre plusieurs niveaux de réalité.

6.2.1 Omniscience physique *simpliciter*

Parmi les candidats au titre d' « arguments boomerang » figure, bien évidemment, l'argument de la connaissance de Jackson, mais aussi l'argument de la concevabilité de Chalmers. Or, à la différence de Van Gulick, Nagasawa soutient que le cas des zombies n'est pas véritablement un argument de la connaissance, car il ne s'emploie pas à réfuter la thèse de la connaissance physique conçue comme connaissance *simpliciter* (C1). Afin de rendre ce point plus explicite, Nagasawa propose d'introduire quelques précisions supplémentaires, en s'appuyant sur un examen attentif des possibles implications du concept d'« omniscience ». Brièvement décrite, la thèse de l'omniscience – ou plus exactement la thèse de l'omniscience en tant qu'omniscience *simpliciter* – peut être formulée ainsi :

(O_S) Omniscience *Simpliciter* : Pour tout *X*, et pour toute proposition *p*, *X* est omniscient si et seulement si, s'il est vrai que *p*, alors *X* sait que *p*.[21]

Pourquoi insister sur le caractère *simpliciter* dans la définition de l'omniscience ? Nagasawa justifie son choix en rappelant qu'il existe des formes « locales » d'omniscience : quelqu'un pourrait, par exemple, prétendre être omniscient par rapport à une proposition, un discours ou un sujet déterminé. Par contraste, la thèse de l'omniscience *simpliciter* vise explicitement la *totalité des propositions vraies*. En s'appuyant sur la thèse de l'Omniscience *Simpliciter*, on peut dès lors reformuler la

[21] "Omniscience *Simpliciter*: For any *X*, and for any proposition *p*, *X* is omniscient if and only if if it is true that *p*, then *X* knows that *p*." (Y. Nagasawa, 2008:7)

notion d'omniscience impliquée dans l'analyse des propriétés et des objets physiques ainsi :

(O_P) Omniscience Physique: Pour tout *X* et pour toute proposition physique *p*, *X* est physiquement omniscient si et seulement si, s'il est vrai que *p*, alors *X* sait que *p*[22].

Une proposition « physique » est une proposition qui satisfait les deux conditions suivantes :

(a) elle concerne des événements, entités et propriétés du réel dont les composantes de base sont des ingrédients physiques ;

(b) elle dérive non-trivialement et de façon *a priori* des propositions définies en (a).

Ayant établi ces aspects d'arrière-plan, Nagasawa peut avancer un pas de plus et noter que si l'on adopte les notions d'« omniscience *simpliciter* » et d'« omniscience physique » telles que décrites plus haut, il s'ensuit qu'une information physique complète *est* une information complète *simpliciter*, et nous obtenons dès lors la thèse de l'Omniscience Physique *simpliciter* :

(O_{PS}) Omniscience Physique *Simpliciter* : L'omniscience physique *est* omniscience *simpliciter*[23].

La retranscription de tous ces éléments nous permet, dit Nagasawa[24], de reformuler l'argument de la connaissance de Jackson, sans en modifier ses enjeux majeurs, de la manière suivante :

MARY

(1) Mary (avant sa libération) est physiquement omnisciente.

(2) Mary (avant sa libération) n'est pas omnisciente *simpliciter* (car elle apprend quelque chose à sa sortie de prison).

Par conséquent,

(3) L'omniscience physique n'est pas omniscience *simpliciter*.

Prises séparément, les prémisses (1) et (2) semblent inoffensives contre le physicalisme. Ce n'est que lorsqu'on les met ensemble que le cocktail devient explosif,

[22] "Physical Omniscience: For any *X* and for any physical proposition *p*, *X* is physically omniscient if and only if if it is true that *p*, then *X* knows that *p*. *Ibid*.
[23] *Ibid.*, p. 8
[24] *Ibid*.

pour ainsi dire, et le physicalisme se retrouve véritablement en danger. Car le gouffre qui sépare l' « omniscience *simpliciter* » et l' « omniscience physique » dont jouit Mary dans sa prison pose un sérieux problème. Compte tenu de ces nuances, il est possible dès lors de peaufiner la définition d'un argument de la connaissance de la manière suivante :

> **Argument de la Connaissance$_{\mathbf{def}}$** – un argument de la connaissance est un argument qui dérive, non-trivialement, une conclusion ontologique à partir d'un ensemble de prémisses épistémiques, qui impliquent conjointement une thèse de la forme suivante : l'omniscience relative à un sujet spécifique déterminé n'est pas omniscience *simpliciter*[25].

L'originalité de l'approche de Nagasawa tient au fait qu'elle attire l'attention sur la possibilité de repérer d'autres exemples d'arguments de la connaissance situés en dehors du champ d'application standard de la philosophie de l'esprit contemporaine, à savoir dans la philosophie de la religion. Après tout, quoi de plus « naturel » que de se tourner vers cette dernière lorsqu'on aborde des notions à portée « démiurgique » comme c'est le cas avec le concept d' « omniscience » ? Selon Nagasawa, un possible parallèle entre les applications des arguments de la connaissance dans les deux champs disciplinaires distincts pourrait s'avérer révélateur quant à la faillite de l'argument de Jackson de réfuter le physicalisme. Pour voir cela, Nagasawa met en avant une comparaison avec l' « argument de la possession du concept » dans la philosophie de la religion.

Comme le souligne fort bien F. Jackson[26], le défi lancé par le physicalisme n'est pas de dire modestement que le monde est *largement* physique, mais la thèse audacieuse selon laquelle le monde est *entièrement* physique. Maintenant, dit Nagasawa, si l'on considère conjointement la thèse de l'omniscience *simpliciter* (O_s) :

---

[25]Un autre aspect important à souligner est que la dérivation d'une conclusion dans un argument de la connaissance doit être nontriviale. Cette précision éviterait de ranger, à tort, des inférences de type : « (A) Je sais qu'il y a arbre devant moi. Par conséquent, (B) Il y a un arbre devant moi » dans la classe des arguments de la connaissance.

[26] "If physicalism is true, [Mary] knows all there is to know. For to suppose otherwise is to suppose that there is more to know than every physical fact, and that is just what physicalism denies. Physicalism is not the noncontroversial thesis that the actual world is largely physical, but the challenging thesis that it is entirely physical. This is why physicalists must hold that complete physical knowledge is complete knowledge *simpliciter*. For suppose it is not complete: then our world must differ from a world, *W(P)*, for which it is complete, and the difference must be in non-physical facts; for our world and *W(P)* agree in all matters physical. Hence, physicalism would be false at our world [though contingently so, for it would be true at *W(P)*]." (F. Jackson, 1986: 291)

(O_S) Omniscience *Simpliciter* : Pour tout *X*, et pour toute proposition *p*, *X* est omniscient si et seulement si, s'il est vrai que *p*, alors *X* sait que *p*.

et la thèse de l'omniscience physique (O_P) :

(O_P) Omniscience Physique: Pour tout *X* et pour toute proposition physique *p*, *X* est physiquement omniscient si et seulement si, s'il est vrai que *p*, alors *X* sait que *p*.

On obtient une nouvelle retranscription de l'argument de la connaissance sous la forme suivante :

KA SIMPLICITER

(4) Si le physicalisme est vrai, alors l'omniscience physique est omniscience *simpliciter*.

(5) L'omniscience physique n'est pas omniscience *simpliciter*.

Ergo,

(6) Le physicalisme est faux.

Tournons-nous à présent vers le deuxième membre de la comparaison, à savoir l'argument de la possession du concept tiré de la philosophie de la religion.

6.2.2 L'argument de la possession du concept

Il s'agit d'un argument qui se prête à des diverses formulations, mais sa trame fondamentale peut être résumée *grosso modo* ainsi : selon la vision traditionnelle du théisme, Dieu est tel que s'Il existe, alors Il est nécessairement omniscient et omnipotent. Néanmoins, poursuit l'argument, l'omnipotence de Dieu sape sa propre omniscience : car s'Il est véritablement tout-puissant, il résulte qu'Il ne peut pas avoir l'expérience de la *peur*. S'Il ne peut pas avoir cette expérience, il s'ensuit qu'Il ne peut pas connaître ce que « peur » veut véritablement et profondément dire. Par conséquent, s'Il ne peut pas connaître le concept de « peur » totalement, il s'ensuit que son omniscience échoue. Donc, si Dieu est nécessairement omnipotent, alors il ne peut pas être à la fois omniscient. *Ergo*, Dieu conçu comme une entité omnisciente *et* omnipotente *n'existe pas*.

Retranscrit dans une forme plus rigoureuse, l'argument de la possession du concept s'appuie sur deux prémisses sous-jacentes. La première branche de l'argument,

explicitement formulée par A. N. Prior en 1962, étaye la définition de l'omniscience divine :

(i) « Pour chaque *p*, si *p*, alors Dieu sait que *p* »[27].

La deuxième branche, plus « terrestre » pour ainsi dire, définit le versant empirique dans la caractérisation d'un concept :

(ii) Pour tout agent *X* et pour tout concept *c*, *X* comprend complètement *c* si et seulement si *X* a effectivement eu l'expérience corrélative afférente au *c*.[28]

Avec ces éléments en poche, nous sommes en mesure à présent de retranscrire **l'argument de la possession du concept** sous la forme d'un syllogisme:

(1') Si Dieu existe, alors Il est nécessairement omniscient (*simpliciter*), omnipotent, et omnibienveillant.

(2') Si Dieu ne comprend pas complètement le concept de « peur », alors Il n'est pas omniscient *simpliciter*.

(3') A cause de sa nécessaire omnipotence, Dieu n'a pas effectivement éprouvé de la peur.

(4') Pour tout agent *X*, *X* comprend complètement le concept de « peur » uniquement si *X* a effectivement éprouvé de la peur.

POSSESSION DU CONCEPT

Ergo,

(5') Dieu ne comprend pas entièrement le concept de "peur". (à partir de (3′) et (4′))

Ergo,

(6') L'omniscience divine n'est pas omniscience *simpliciter*. (à partir de (2′) et (5′))

Ergo,

(7') Dieu n'existe pas. (à partir de (1′) et (6′))

[27] "For every *p*, if *p* then God knows that *p*". Voir A. N. Prior, (1962), 'The Formalities of Omniscience'. *Philosophy* 37: 114–129.

[28] "For any agent *X* and for any concept *c*, *X* fully comprehends *c* only if *X* has actually had a relevant experience". (Y. Nagasawa, 2008 : 43)

Selon Nagasawa, cet argument respecte la même trame structurale que l'argument de la connaissance de Jackson. Pour voir cela, il suffit de reformuler la thèse de l'omniscience divine.

(O_D) **Omniscience Divine** : Pour tout agent *X* et pour toute proposition *p*, *X* est *divinement* omniscient si et seulement si, s'il est vrai que *p*, alors *X* sait que *p*[29].

A présent, nous sommes en mesure de proposer la définition suivante:

(O_{DS}) **Omniscience Divine *Simpliciter*** : L'omniscience divine est omniscience *simpliciter*.

On peut dès lors, de façon analogue au cas 'MARY', synthétiser l'argument[30] de la possession du concept de la manière suivante :

DIEU

(8') Dieu est divinement omniscient.

(9') Dieu n'est pas omniscient *simpliciter*.

Ergo,

(10') L'Omniscience divine n'est pas omniscience *simpliciter*.

6.2.3 Information physique théoriquement communicable

Les retranscriptions 'DIEU' et 'MARY' permettent au lecteur de saisir les similitudes entre l'armature de l'argument de la connaissance d'une part, et de l'argument de la possession du concept d'autre part, puisque l'argument 'MARY' repose tacitement sur le postulat (7) ci-dessous, qui n'est qu'une variation de l'énoncé (4′) de l'argument de la possession du concept que nous avons évoqué plus haut :

(7) Pour tout sujet *X*, *X* connaît l'effet-que-cela-fait de voir du rouge si et seulement si *X* a déjà éprouvé l'expérience visuelle d'un objet rouge.

Ayant acquis ces nouveaux outils conceptuels, revenons à présent à la case départ, à savoir à la première prémisse (P1) de l'argument de Jackson[31] (1982/1986) qui

[29] *Ibid.*, p. 11.

[30] « *GOD* (21) God is divinely omniscient. (22) God is not omniscient *simpliciter* (because He does not comprehend fully what fear is). Therefore, (23) Divine omniscience is not omniscience *simpliciter*. » (*Ibid.*, p. 122)

[31] Tenant compte du changement d'avis de F. Jackson par rapport à sa prise de position initiale, et afin d'éviter d'éventuelles confusions, nous allons accompagner dorénavant son nom avec des parenthèses

stipule que Mary (avant sa libération) connaît *toute* l'information physique concernant les objets et les autres sujets. Admettons provisoirement que le terme « physique » désigne ici « tout ce qui est inclus dans une théorie physique *complète* ». A présent, la question que l'on peut légitimement se poser est : comment Mary arrive-t-elle à connaître cette information physique complète ? Rappelons-nous que pour Jackson (1982/1986), Mary réalise son exploit à l'aide des livres, d'un téléviseur noir et blanc, ainsi que *via* des traités neurophysiologiques sur la perception des couleurs, les informations lui étant présentées sous la forme de texte et de diagrammes noirs et blancs.

Il est évident, note judicieusement Nagasawa[32], que la prémisse (P1) repose également sur le postulat tacite (8) :

(8) Un sujet connaissant peut être, *en principe*, physiquement omniscient simplement en lisant des livres noirs et blancs et en visionnant des émissions de télévision en noir-et-blanc.

Quelles sont les motivations plausibles derrière le postulat (8) ? Jackson justifie son choix d'une manière assez surprenante. Il dit : « on peut difficilement rejeter l'idée selon laquelle il serait possible, en principe, d'obtenir une information physique complète à partir des émissions télé en noir et blanc ; autrement, l'Université Ouverte de Tous les Savoirs serait obligée d'utiliser *nécessairement* une télévision en couleurs»[33]. Comme le relève à juste titre Nagasawa, cette *reductio ad absurdum* n'est pas très convaincante. La description qu'en donne Jackson du processus cognitif de Mary, comporte deux informations cruciales : (a) d'abord, on apprend que ce processus d'acquisition des informations physiques est *monochromatique.* (b) Ensuite, on apprend que la prisonnière utilise exclusivement des *livres,* des *diagrammes, et* des *émissions sur un écran de téléviseur.*

On pourrait objecter contre (a) que Mary ne peut pas être physiquement omnisciente, car une connaissance adéquate de certains faits physiques requiert l'expérience préalable de ces mêmes faits. Par exemple, pour que Mary puisse

indiquant les années de parution de ses articles, soit 1982/1986 pour sa position antiphysicaliste, et 1998/2003 pour sa conversion physicaliste.

[32] (Y. Nagasawa, 2008: 126)

[33] "It can hardly be denied that it is in principle possible to obtain all this physical information from black and White television, otherwise the Open University would *of necessity* need to use color television". (F. Jackson, 1982: 130)

connaître l'effet-que-cela-fait de voir du rouge, on pourrait dire qu'elle doit avoir fait préalablement l'expérience d'un objet rouge quelconque (cf. postulat (7)). Or, ce qui est très étonnant dans l'affaire de 'MARY', observe Nagasawa, c'est que la plupart d'auteurs se sont penchés sur le postulat (7), en négligeant la thèse (b) selon laquelle un individu serait capable, en principe, de connaître toute l'information physique relative à la perception visuelle du rouge *uniquement en lisant et en visionnant des émissions sur un écran de téléviseur* (postulat (8)).

Or, selon Nagasawa, ce que Mary apprend de cette manière, ce sont des faits *théoriquement communicables*, c'est-à-dire des faits basés sur une caractérisation des objets et propriétés physiques en termes de leurs relations contingentes les connectant les uns aux autres, et les rattachant d'ici plus loin, au sujet connaissant. Bref, *une information physique théoriquement communicable* est une information que l'on peut acquérir par le biais de l'apprentissage des *théories* développées au sein des sciences physiques. A la lumière de ces considérations, il devient possible de considérer que Mary *est* physiquement omnisciente, à condition que l'on adhère à la définition du concept d' « information physique » en tant que théoriquement communicable. La question qui se pose à présent est de savoir si l'information physique théoriquement communicable est de l'information physique *simpliciter*. Autrement dit : y-a-t-il des formes d'information physiques nonthéoriques et/ou incommunicables ? Car si le véritable enjeu est l'acquisition d'une information physique *complète*, il devient critique d'établir si Mary est en position d'obtenir ce type exhaustif de savoir uniquement par le biais de l'apprentissage qu'offrent les livres et la télévision. Bref, il s'agit de montrer que la totalité des faits théoriquement communicables n'épuise pas la totalité des vérités qui peuvent être avancées à propos de la nature physique du réel.

Pour justifier ce point de vue, Nagasawa fait appel à un concept que D. Lewis[34] a introduit à d'autres fins, mais qui se prête bien ici à l'analyse de ces aspects : le concept d'« humilité Ramseyenne ». Précisons tout de suite que Lewis n'utilise pas la thèse de l' « humilité » pour réfuter l'argument de la connaissance, préférant, quant à lui, se ranger du côté des partisans de l'hypothèse de l'habilité, que nous avons déjà évoquée dans la section 5.4.1.

[34] D. Lewis, (2001), 'Ramseyan Humility'. *Preprint Series 1/01*: University of Melbourne. Réédité *in* D. Braddon-Mitchell & R. Nola (eds.), (2009), *Conceptual Analysis and Philosophical Naturalism*, MIT Press. (Nous faisons référence ici à cette dernière édition)

6.3 Propriétés fondamentales et « humilité Ramseyenne »

Admettons que Mary devienne physiquement omnisciente en apprenant *toute* l'information physiquement communicable mise à sa disposition par le biais des livres et des écrans de téléviseur. Mary dispose ainsi, pour utiliser les termes de Lewis, d'une « *théorie finale* » du monde. Qu'est-ce qu'une théorie finale ? Selon Lewis, une théorie finale *T* est une description du monde qui recouvre l'inventaire complet et vrai de toutes les propriétés physiques, y compris les propriétés fondamentales, qui jouent un rôle actif dans les processus et les phénomènes naturels.

Les propriétés fondamentales sont, selon Lewis[35], des « propriétés parfaitement ou intrinsèquement naturelles ». Comme il l'admet lui-même, aucun des qualificatifs de "fondamental" ou de "naturel" ne convient pleinement pour caractériser le statut particulier de ces propriétés. Faute de mieux, il faut s'en tenir à ces expressions. Les propriétés fondamentales, parmi lesquelles figure la charge positive ou négative des champs électromagnétiques, par exemple, sont des propriétés intrinsèques sur lesquelles d'autres propriétés intrinsèques surviennent[36]. Autrement dit, les propriétés intrinsèques sont des propriétés qui servent de base primitive aux dispositions qui caractérisent les entités physiques. Avant d'aller plus loin, il nous faut brièvement définir le terme clé d' « énoncé de Ramsey »[37]. On s'accorde généralement à dire que le processus de théorisation scientifique et la découverte des propriétés fondamentales vont de pair. Un

[35] "Fundamental properties are those properties that I have elsewhere (Lewis, 1983a) called 'perfectly natural'. They are not at all disjunctive, or determinable, or negative. They render their instances perfectly similar in some respect. They are intrinsic; and all other intrinsic properties supervene on them. They are not conjunctive or structural". (D. Lewis, 2001/2009: 204)

[36] "Fundamental properties figure in a minimal basis on which all else supervenes. No two possible worlds just alike in their patterns of instantiation of fundamental properties could differ in any other way. I speak of 'fundamental properties' for short, but they fall into several categories. There are all-or-nothing monadic properties. There are all-or nothing *n*-adic relations, at least for smallish *n*. There are properties that admit of degree, that is, magnitudes; more generally, there are scalar valued, vector-valued, tensor-valued, … magnitudes. There are relational magnitudes. Maybe my list is too long; maybe the magnitudes could somehow be reduced to all-or-nothing properties and relations, but that is a question I shall not take up here. Scientific theorizing and the discovery of fundamental properties have gone hand in hand. For instance the discovery of the phenomena of electromagnetism and the laws governing them was inseparable from the discovery of the previously unknown, and very likely fundamental, properties of positive and negative charge. So if we had a true and complete "final theory," it ought to deliver a true and complete inventory of those fundamental properties that play an active role in the actual workings of nature. That inventory might omit two kinds of fundamental properties: idlers and aliens. *Idlers* are those fundamental properties, if any, that are instantiated within the actual world, but play no active role in the workings of nature. *Aliens* are those fundamental properties, if any, that are instantiated within unactualized possible worlds but not within the actual world. We shall consider later whether we have any reason to believe, or any reason to disbelieve, in idlers or in aliens." (*Ibid.*, p. 205)

[37] F. P. Ramsey (1990), *Philosophical Papers*. (édité par D. H. Mellor), Cambridge: Cambridge University Press.

des exemples que l'on trouve souvent dans la littérature analytique[38] concerne le cas de l'électromagnétisme. Au cours du XIXe siècle, l'échec du processus de réduction des propriétés de charge (négative/positive) a été interprété comme un signe prouvant que ces propriétés seraient fondamentales. Conséquemment, la formulation des lois qui gouvernent les phénomènes d'électromagnétisme incorpore les propriétés de charge (négative/positive) comme étant des propriétés fondamentales.

Pour obtenir l'énoncé de Ramsey d'une théorie T, on remplace les termes théoriques par des variables de la manière suivante:

$T\{t_1, t_{2,}, \ldots, t_n\}$.

Remplaçons tous les termes $t_1, \ldots, t_n$ par des variables $x_1, \ldots, x_n$ respectivement :

$T\{x_1, x_{2,}, \ldots, x_n\}$

et préfixons cette dernière formule de quantificateurs :

$(R)\exists x_1 \exists x_2 \ldots \exists x_n \quad T\{x_1, x_{2,}, \ldots, x_n\}$.

(R) est ce que l'on appelle l'énoncé de Ramsey de la théorie T. Il dit simplement qu'il y a des « propriétés ou des classes d'entités, qui ont entres elles diverses relations (que nous pourrions faire figurer dans R) entres elles et avec les termes d'observations de T. Les termes théoriques ont été ainsi remplacés par des phrases générales existentielles dont les termes ne dénotent plus directement des entités théoriques, mais spécifient les relations de ses entités à des observations» [39].

A présent, si l'on remplace dans la théorie finale *T* les termes qui nomment des propriétés fondamentales (comme la charge positive et négative dans l'exemple que nous venons de donner) avec des variables existentielles quantifiées, on obtient un énoncé de Ramsey de la théorie *T* de la forme suivante : « Pour certains x_1 . . . pour certains x_n $T(x_1 \ldots x_n)$ ». En bref, l'énoncé de Ramsey nous dit qu'il existe au moins *une* réalisation de *T*, qui est, évidemment, sa réalisation *actuelle*. A partir de là, il serait intéressant de s'interroger sur la question de savoir si *T* est oui ou non multiplement réalisable, c'est-à-dire, s'il existe un monde possible dans lequel *T* serait vrai, mais dans lequel l'arrangement des propriétés fondamentales serait *différent* de

[38] (Voir par exemple D. Chalmers, 1996)
[39] J. Dokic & P. Engel, (2001), *Ramsey, vérité et succès*, Paris, P.U.F., p.56.

l'ordonnancement de notre monde actuel. A cette question, Lewis répond par l'affirmative.

A l'appui de son raisonnement, Lewis[40] propose l'argument suivant : permutons deux propriétés fondamentales F1 et F2 dans *T*, et gardons inchangé tout le reste. F2 se retrouvera ainsi exactement à la même place où F1 se trouvait initialement, et *vice versa*. De manière analogue, les lois physiques qui gouvernaient initialement F1 régiront à présent F2 et *vice versa*. Cette permutation représente une réalisation de T qui est *différente* de la réalisation actuelle de T. Autrement dit, cette opération de permutation montre qu'il est possible d'imaginer un monde possible dans lequel *T* serait vrai pour le monde respectif, et pourtant l'arrangement de F1 et F2 pour le monde respectif serait différent de celui qui est en place dans le monde actuel. Bref, *T* est multiplement réalisable.

Pourquoi est-il si important de montrer que *T* est de fait multiplement réalisable ? La raison est la suivante : si *T* est multiplement réalisable, alors *aucune* observation possible ne serait à même de nous dire *laquelle* des possibles réalisations *est* la réalisation *actuelle*. Regardons' y de plus près. Supposons qu'il existe deux possibles réalisations de *T*, à savoir R1 et R2, et qu'il existe deux mondes possibles W1 et W2, au sein desquels R1 et R2 se trouvent respectivement réalisés. Dans ce cas, et si Lewis a raison, alors *aucune* théorie, même pas une théorie finale *T*, ne serait capable de nous dire dans *lequel* des mondes nous sommes ici et maintenant : car l'énoncé de Ramsey est vrai dans n'importe lequel des mondes qui nous accueille. En revenant à l'argument qui nous intéresse ici, à supposer que Mary dispose, durant sa captivité, d'une théorie complète et finale *T*, il lui serait toutefois impossible de connaître le substrat ultime de la réalité, pour la simple bonne raison qu'elle ne sait pas comment les propriétés fondamentales sont réellement arrangées.

---

[40] "Suppose we have the actual realization of T. Maybe some members of the *n*-tuple that realizes T are not fundamental properties, or maybe some belong to single-membered categories. Hold those ones fixed. Permute the rest within their categories to obtain a new *n*-tuple. It too would realize T. Suppose, for instance, that we start with the actual world, and we permute two fundamental monadic properties F1 and F2, these being the actual referents of the T-terms t1 and t2, leaving all else fixed. Exchange F1 and F2 throughout the world. (Suppose for simplicity that possibilities are worlds like ours; though if instead they were some sort of representations of worlds, my argument would go through *mutatis mutandis*.) Then F2 will be found in exactly those places in space and time (or, more generally, in the pattern of instantiation of fundamental relations and magnitudes) that correspond to the places where F1 was found originally; and vice versa. And the laws of nature governing F1 in the permutation will be just the same as the laws governing F1 originally (more precisely, the laws governing F2 *vis-à-vis* F1 in the permutation will be the same as those governing F1 *vis-à-vis* F2 originally); and *vice versa*.11 But what could T tell us about besides the locational and nomological roles of the referents of its T-terms? So if this permutation really is a possibility, it is a possible realization of T." (D. Lewis, 2001/2009: 208)

Par conséquent, conclue Nagasawa[41], si l'argument de l'humilité ramseyenne de Lewis est valide, il s'ensuit que Mary, qui connaît *T*, et qui possède une caractérisation des propriétés et objets physiques sous une forme théoriquement communicable, *ne connait pas toute l'information physique*. Donc, le postulat tacite (8) sur lequel repose le KA de Jackson est faux, et l'argument 'MARY' échoue.

On peut objecter contre l'argument de Nagasawa en suggérant que l'ignorance dont fait preuve Mary quant aux propriétés fondamentales est sans importance vis-à-vis des enjeux initiaux de l'argument, qui ciblent, quant à eux, la nature des découvertes expérientielles que la prisonnière réalise une fois libérée de sa geôle. Par exemple : le fait que Mary ne sait pas que le grand-père de Leibniz était avocat, disons, n'affecte en rien la portée de l'argument, puisque cette information est sans pertinence pour l'objectif que représente l'acquisition de la perception visuelle des couleurs.

A cela Nagasawa répond que cette objection n'est pas recevable, car l'ignorance dont fait preuve Mary, telle qu'elle est démontrée par l'argument de l'humilité ramseyenne de Lewis, est d'une importance vitale pour le sujet qui nous intéresse ici. Cette ignorance concerne les propriétés fondamentales qui sous-tendent la *totalité* de propriétés et objets physiques. Or, étant donné que l'expérience phénoménale perceptuelle nécessite en règle générale l'activation d'un réseau sophistiqué de manipulation d'objets physiques, ainsi qu'une complexe activité neurophysiologique sous-tendante, il est relativement facile de montrer que l'ignorance de Mary quant aux propriétés fondamentales est au moins partiellement, et indirectement, pertinente dans la caractérisation de la découverte expérientielle qui accompagne sa perception des couleurs lors de sa première sortie de captivité.

On pourrait objecter également, contre cette réfutation de l'argument de Jackson (1982/1986), que l'argument de l'humilité ramseyenne ne fait que corroborer, voire renforcer, une certaine forme de « protopanpsychisme »[42] défendue par Chalmers. Réduite à sa plus simple expression, le protopanpsychisme est une forme de dualisme

[41] (Y. Nagasawa, 2008 : 128)

[42] "Perhaps we might take [phenomenal] experience itself as a fundamental feature of the world, alongside space-time, spin, charge and the like. That is, certain phenomenal properties will have to be taken as *basic* properties. Alternatively, perhaps there is some *other* class of novel fundamental properties from which phenomenal properties are derived. . . . [T]hese cannot be physical properties, but perhaps they are non-physical properties of a new variety, on which phenomenal properties are logically supervenient. Such properties would be related to experience in the same way that basic physical properties are related to non-basic properties such as [the] temperature [of a gas].We could call these properties *protophenomenal* properties, as they are not themselves phenomenal but together they can yield the phenomenal." (D. Chalmers, 1996: 126–7)

selon laquelle les propriétés phénoménales seraient réductibles à des propriétés *protophénoménales*. Nous avons vu dans la section 4.5.3. que Chalmers[43] range son « protopanpsychisme » sous le label « monisme de type-F », et affirme que si l'on assimile les propriétés protophénoménales à des propriétés fondamentales, intrinsèques, alors le protopanpsychisme n'est à la base qu'une version de physicalisme, dans le sens où le terme « physique » subsume des entités physiques et propriétés qui ont une efficience causale directe sur le sujet connaissant. Nagasawa exprime très clairement son désaccord avec la position de Chalmers. Plus spécifiquement, il rejette la thèse des propriétés protophénoménales intrinsèques, qu'il tient pour hautement suspecte.

6.4 Ce que Dieu nous apprend sur Mary

A ce stade de notre enquête, quelqu'un pourrait attaquer l'édifice du KA en soulevant l'objection suivante : par quel moyen miraculeux serait-il possible de connaître toutes les propriétés fondamentales des entités physiques, étant donné qu'il nous est impossible de les connaître sous une forme *théoriquement communicable* ?

Selon Nagasawa[44], une possible solution de relever ce défi serait de faire appel à nouveau à Dieu ! L'idée c'est de caractériser le savoir divin comme étant, du moins partiellement, *intuitif*[45]. Une intuition est une *saisie* intellectuelle immédiate qui n'implique absolument rien d'autre, même pas la perception directe d'un objet. Nous allons le voir, cette notion de « saisie intellectuelle immédiate » présente un certain air de famille avec la notion de « saisie d'une pensée » chez Frege. Nous y reviendrons dans le chapitre suivant, lorsque nous élaborerons le portrait final de l'« intrus » qui perturbe la pensée philosophique dans sa mission de « localiser » l'expérience phénoménale dans le monde physique. Pour le but de cette section, nous nous contenterons d'esquisser seulement les aspects les plus importants du physicalisme nonthéorique de Nagasawa, laissant pour plus tard la question de savoir comment les discussions autour de ce type de physicalisme peuvent être utilisées pour fragiliser les ambitions métaphysiques du physicalisme *a priori*.

---

[43] (D. Chalmers, 2002a: 129-130)
[44] (Y. Nagasawa, 2008 : 130)
[45] Comme nous allons le voir dans le chapitre suivant, il est important de garder à l'esprit que la proposition de Nagasawa quant à la nature de la connaissance divine n'est qu'une simple supputation. Personne ne peut clamer savoir ce que la connaissance divine *est*.

Selon Nagasawa, étant donné que les propriétés fondamentales sont inconnaissables sous une forme théoriquement communicable, on peut présumer qu'elles peuvent être accessibles uniquement par le biais de l'intuition. Autrement dit, Mary peut être considérée comme étant physiquement omnisciente uniquement si elle peut accéder à *l'intuition des propriétés fondamentales*. Pour Nagasawa, cette hypothèse n'implique pas nécessairement d'adhérer à une forme d'empiricité, car l'intuition à laquelle il fait allusion est non-expérientielle, non-inférrentielle, non-perceptuelle et non-imaginative.

En s'appuyant sur la mystérieuse hypothèse de l'intuition non-expérientielle non-inférrentielle, non-perceptuelle et non-imaginative comme mode d'accès privilégié aux propriétés fondamentales du réel, il devient dès lors évident que les humains (Mary y compris), *ne* disposent *pas* de ce type d'accès intuitif direct, et que par ricochet, ils sont incapables de saisir la structure primitive, ultime, des entités physiques. En revanche, il est tout à fait raisonnable de concevoir, affirme Nagasawa, des êtres supérieurs – Dieu, ou d'autres créatures extraterrestres hyper-intelligentes – capables d'avoir cette intuition immédiate de l'essence des choses. Pour reprendre les termes de Lewis[46], il se peut qu'un être supérieur ait le pouvoir surnaturel d'accéder intuitivement et immédiatement à la totalité des propriétés fondamentales, *qualia* ou pas. Mais à nouveau, si Dieu existe, qui sait quels superpouvoirs possèderait-t-il réellement? En ultime instance, ce qui compte pour l'objectif que nous poursuivons ici c'est d'insister sur le fait que même si Dieu connaît toutes les propriétés fondamentales du réel, *jamais* Il ne pourra les partager avec nous, pauvres créatures limitées que nous sommes.

Maintenant, supposons que Mary reçoit de Dieu la bénédiction miraculeuse d'une telle intuition directe des propriétés ultimes de réel. Or, cette nouvelle hypothèse ne nous avance guère dans l'affaire qui nous concerne ici car à présent la question qui surgit est : dans quelle mesure, nous les humains, créatures dont la finitude est indéniable, serions-nous concernés par le fait de savoir si Mary apprend quelque chose de nouveau lors de sa libération ? Bien sûr qu'elle n'apprend rien ! – pourrait-on répondre – puisqu'elle sait déjà *tout*! Mais il y a une autre manière d'aborder

[46] "If there were a God, who knows what supernatural powers He might have. But no matter what knowledge He might gain by acquaintance with the fundamental properties, He could not share it with us. Since we cannot express any of the answer-propositions to the question which fundamental property occupies a given role—not in such a way that we know which of the propositions we are expressing—God would have no way to communicate His knowledge to us. If He wanted to remedy our ignorance, His only recourse would be to impart to us His own power to identify properties by acquaintance." (D. Lewis, 2001/2009: 218)

l'hypothèse de l'omniscience. Admettons que Mary connaît les propriétés fondamentales du réel. Si elle est physiquement omnisciente, il s'ensuit, comme l'indique la prémisse (4), que si le physicalisme est vrai, alors l'omniscience physique *est* omniscience *simpliciter*. Et si elle est véritablement omnisciente *simpliciter*, il va de soi qu'elle ne découvrira aucune vérité nouvelle lors de sa libération.

Récapitulons. Si l'on reformule le KA et l'APC n termes d'omniscience, il devient clair que l'argument tiré de la philosophie de la religion ne repose pas sur une interprétation controversée de la notion d' « omniscience divine ». Autrement dit, l'APC ne se rend pas coupable d'utilisation abusive de la notion clé d' « omniscience divine »[47]. Il évolue, si l'on peut dire ainsi, en terrain connu, légitime. En revanche, le KA de Jackson va au-delà des frontières légitimes dans l'application du concept d' « omniscience physique »[48], et se rend coupable, par ricochet, d'interprétation abusive du concept de « physique », c'est-à-dire il repose sur l'hypothèse hautement discutable selon laquelle l'information théoriquement communicable serait la *seule* forme d'information physique tout court. A l'appui de son raisonnement, Nagasawa a fait appel au concept Lewisien d' « humilité Ramseyenne » afin de montrer que Mary *ne* peut *pas* connaître les propriétés non-relationnelles, fondamentales, du réel physique uniquement par le biais des livres et des écrans de téléviseurs en noir et blanc. D'autre part, si l'on est prêt à adhérer à la thèse forte selon laquelle Mary est effectivement physiquement omnisciente durant sa captivité, il devient relativement facile de montrer que nous exagérons la portée des capacités cognitives de Mary.

6.5 Le physicalisme nonthéorique de Nagasawa

Ayant défini ces éléments d'arrière plan, Nagasawa[49] peut avancer d'un pas de plus dans l'élaboration de son propre *physicalisme nonthéorique*[1] comme une conjonction des trois thèses suivantes :

[47] Pour mémoire : (O_{DS}) Omniscience Divine *Simpliciter* : L'omniscience divine est omniscience *simpliciter*.
[48] Pour mémoire :(O_{PS}) Omniscience Physique *Simpliciter* : L'omniscience physique *est* omniscience *simpliciter*
[49] (Y. Nagasawa, 2008 : 132)

(i) L'omniscience physique *est* omniscience *simpliciter*. (La thèse de l'omniscience physique est vraie).

(ii) L'omniscience physique théoriquement communicable n'est pas omniscience *simpliciter*. (Autrement dit, la thèse de l'omniscience physique théoriquement communicable est fausse).

PHYSICALISME NON-THÉORIQUE

(iii) L'omniscience *simpliciter* présuppose l'instanciation des capacités épistémiques exceptionnelles afin de saisir les propriétés intrinsèques primitives du réel.

Passons brièvement en revue les arguments de Nagasawa en faveur de chacune de ces thèses.

La prémisse (i) est de loin la moins controversée. Même les antiphysicalistes sont prêts à concéder la validité du conditionnel suivant :

(C) « Si le physicalisme est vrai, alors l'omniscience physique *est* omniscience *simpliciter* ».

En effet, si le monde est entièrement physique[50], alors connaître toute l'information physique vraie revient à connaître toute l'information vraie *simpliciter*. La vérité du conditionnel (C) dépend entièrement de la vérité de son antécédent, à savoir de la vérité du physicalisme. Selon Nagasawa, si l'on tient compte du fait qu'un argument antiphysicaliste aussi puissant que l'argument de la connaissance échoue dans sa tentative de saper l'édifice physicaliste, on peut, à partir de là, conclure en toute sécurité, la vérité de la thèse (i). Nous verrons dans le chapitre suivant que les choses sont loin d'être aussi claires et simples que Nagasawa le laisse entendre.

La thèse (ii) véhicule, quant à elle, un des aspects les plus intéressants et prometteurs du physicalisme nonthéorique. Comme nous l'avons pu constater dans la section précédente, Nagasawa adhère à l'hypothèse des propriétés fondamentales, mais à la différence des conceptions physicalistes standard, il suggère l'idée audacieuse selon laquelle les théories physiques, bien que complètes, seraient incapables de subsumer et d'englober la totalité des informations physiques sous une forme

[50] "Again, by 'physical propositions' I mean (a) propositions about events, entities, and properties in the world that have basic physical entities and properties as their ultimate constituents and (b) propositions that are nonvacuously entailed *a priori* by such propositions." (Y. Nagasawa, 2008: 136)

théoriquement communicable. Il s'agit d'une thèse dont on pourra mesurer la réelle portée dans la partie conclusive de notre travail, où il sera argumenté que la thèse de l'incommunicabilité théorique qui caractérise le processus de saisie des propriétés fondamentales aura des répercussions importantes sur les ambitions métaphysiques du physicalisme en tant que 'Weltanschauung' de l'époque moderne.

Nagasawa se déclare prêt à reconnaître que la thèse (ii) de son physicalisme nonthéorique est beaucoup plus controversée que la première thèse (i). Cela s'explique en grande partie par le fait que la plupart des auteurs physicalistes adhèrent tacitement à une forme de *physicalisme théorique*, qui prête le flanc aux effets néfastes du dilemme d'Hempel, comme nous l'avons vu dans la section 1.5.3. L'emploi de l'attribut « nonthéorique » ne doit pas nous induire en erreur, car le physicalisme nonthéorique est une théorie *aussi*. Mais une théorie qui n'ambitionne pas à définir le physicalisme en termes d'énoncés théoriquement *communicables*. A l'opposé, le physicalisme nonthéorique n'affirme pas que *toutes* les propriétés ou les objets physiques soient théoriquement *incommunicable*. Plus modestement, il soutient qu'il est possible de concevoir des propriétés et des objets physiques qui à la fois:

(a) restent en dehors de l'espace épistémique délimité par les sciences physiques et

(b) ne surviennent pas sur des entités ou des propriétés physiques qui font l'objet des sciences physiques.

Il est important donc, insiste Nagasawa, de se libérer du dogme du physicalisme théorique qui domine le paysage philosophique actuel.

Quant aux motivations en faveur de la thèse (iii), Nagasawa[51] observe qu'en adoptant les deux premières prémisses du physicalisme nonthéorique, on s'expose à l'objection suivante : étant donné que même les théories développées par des sciences physiques *complètes* échouent dans leur entreprise de couvrir la totalité du réel, dans quel sens pourrait-on encore soutenir qu'un sujet S peut être physiquement omniscient, et respectivement, omniscient *simpliciter* ?

La réponse de Nagasawa est la suivante : l'omniscience *simpliciter* requiert l'instanciation des capacités épistémiques exceptionnelles de saisir intuitivement les propositions pertinentes. A partir de là, il est relativement facile d'aller plus loin et

[51] *Ibid.*, p. 140.

d'affirmer que si Mary apprend toute l'information physique théoriquement communicable, et si elle instancie des capacités épistémiques exceptionnelles afin de saisir les autres informations physiques (incommunicables et nonthéoriques), alors Mary peut être physiquement omnisciente, et conséquemment, omnisciente *simpliciter*. Nous n'irons pas plus loin que ces quelques remarques concernant le physicalisme nonthéorique de Nagasawa.

Retenons seulement qu'en abordant directement et ouvertement le sujet sensible de l'omniscience divine, Nagasawa brise la glace qui sépare la sphère de la connaissance rationnelle de celle de la connaissance idéale. Cette manœuvre constitue un excellent tremplin pour entamer, dans le dernier chapitre de cet essai, le travail de démystification des présupposés tacites qui minent le terrain, déjà très accidenté, de la philosophie contemporaine de l'esprit. Nous argumenterons dans le chapitre suivant, qu'il existe un sens dans lequel le physicalisme actuel a besoin d'être « naturalisé » à son tour, c'est-à-dire débarrassé d'un certain nombre de résidus mythologiques et surnaturels qui l'encombrent inutilement dans sa tâche d'expliquer le réel.

6.6 Arbre récapitulatif synoptique

Au terme de ce périple mettant en scène les divers protagonistes du duel entre le physicalisme et l'argument de la connaissance on constate que cet affrontement entraîne des effets notables sur la question de la place de la conscience phénoménale dans le monde naturel. A ce stade de notre investigation, et afin de mieux se repérer au milieu de ce véritable labyrinthe d'interrogations, arguments et contrarguments, il convient de dresser, à la suite de Van Gulick[52], un schéma figuratif illustrant les solutions « standard » avancées jusqu'ici dans l'affaire de « Mary » (cf. Figure 11: Le KA - Arbre synoptique)

[52] (R. Van Gulick, 2004: 402)

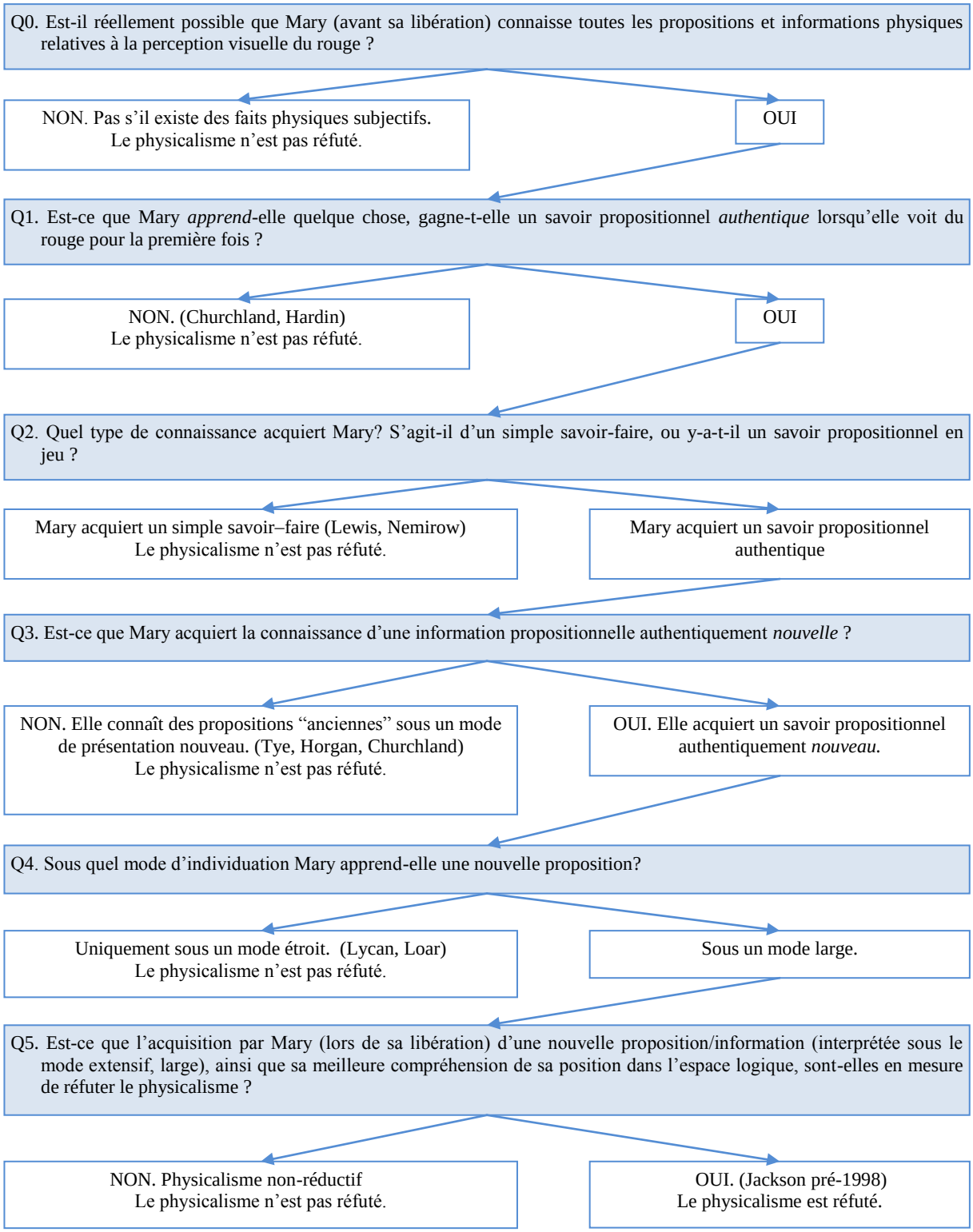


Figure 11: Le KA - Arbre synoptique général

Les questions (Q0) et (Q5), note Van Gulick[53], c'est-à-dire celles qui se positionnent aux pôles extrêmes de l'arbre synoptique, portent, en dernier ressort, sur le même sujet, i.e. la question de savoir si le physicalisme est compatible avec l'existence

[53] (R. Van Gulick, 2008:197)

des faits subjectifs, expérientiels, bref avec la phénoménalité de la conscience. C'est cette dernière hypothèse que nous allons retenir ici pour un examen plus approfondi dans le dernier chapitre de ce travail de recherche.

Récapitulons. Nous avons vu d'après ce qui précède que le physicaliste peut à la limite différer son affrontement décisif avec les défenseurs de l'argument de la connaissance jusqu'à l'énoncé (Q5), en concédant, à la suite de (Q4), que Mary apprend une proposition/information nouvelle au sens large, mais en niant que cette acquisition réfute le physicalisme. A l'intérieur de l'espace délimité par (Q0) et (Q5), conclue Van Gulick[54], le physicaliste se voit confronté aux choix suivants :

(1) S'il admet l'existence des faits subjectifs physiquement réalisés, il peut rejeter (P1) en signalant que les conditions hypothétiques stipulées dans la première prémisse du KA abritent une contradiction interne (Q0) ;

(2) Le physicaliste peut nier que Mary apprend quelque chose de *nouveau* lors de sa libération (Q1) ;

(3) Il peut également affirmer que Mary obtient seulement des aptitudes nouvelles, ou un « savoir-faire » (Q3);

(4) Il peut concéder que Mary acquiert un savoir propositionnel authentique, mais seulement dans le sens où elle apprend à connaître, sous un mode nouveau, phénoménal, une information qu'elle possédait déjà durant sa captivité (la stratégie de la connaissance phénoménale (Q3)) ;

(5) Le physicaliste peut également concéder le fait que Mary apprend des informations nouvelles, mais seulement sous un mode d'individuation étroit (Q4) ;

(6) Enfin, il peut admettre que Mary apprend des énoncés nouveaux (dans un sens large, flexible), mais continuer à nier que cette acquisition récuse le physicalisme, arguant que ce dernier est compatible avec l'existence des faits subjectifs physiquement réalisables.

Le tableau figuré ci-dessus couvre l'intervalle des solutions qui ont été apportées aux volets (M1) et (M2)[55] du dispositif tripartite qui caractérise l'analyse de l'argument

[54] *Ibid.*, p. 202-3.

[55] Pour mémoire, voici les trois thèses : (M1) ***La thèse de la connaissance complète*** : avant de quitter sa prison, Mary connaît *toute* l'information physique relative à la perception visuelle des couleurs. (M2) ***La thèse de l'apprentissage*** : lorsqu'elle quitte sa prison, Mary *apprend* quelque chose de nouveau. (M3) ***La thèse de la non-déductibilité*** : si les thèses (M1) et (M2) sont vraies, alors ce que Mary apprend

de la connaissance. Ces intervalles correspondent également à un éventail de questions allant de (Q0) jusqu'à (Q6) dans la liste dressée par Van Gulick[56] que nous avons évoquée dans la section 5.3.

Il nous reste deux tâches majeures à accomplir dans l'examen des débats autour de la croisade antiphysicaliste menée par les défenseurs de l'argument de la connaissance.

(a) D'abord, il convient d'approfondir la portée du troisième volet (M3) de l'expérience de pensée de Jackson (1982/1986), à savoir la thèse de la non-déductibilité : à supposer que les thèses (M1) et (M2) soient vraies, il s'agit d'établir si le savoir que Mary apprend lorsqu'elle quitte la prison *peut* ou *ne* peut *pas* être déduit de manière *a priori* (à l'aide des raisonnements seulement, sans apport empirique) de l'ensemble complet de vérités physiques dont elle dispose durant sa captivité. La discussion de ces points nous amènera au plus près de la question du physicalisme *a priori*.

(b) La deuxième mission qui nous échoit c'est d'évaluer l'impact des arguments antiphysicalistes en général, et de l'argument de la connaissance en particulier, sur la *vérité* du physicalisme. En d'autres mots, il s'agit d'aborder le deuxième volet de l'analyse tripartite structurée par Stoljar autour des trois axes déjà évoqués, à savoir: (i) la *question de l'interprétation* du slogan « tout est physique » ; (ii) *la question de la vérité/fausseté* du physicalisme ; (iii) la question des *enjeux philosophiques* du physicalisme, autrement dit la *signification métaphysique* dans sa fonction de *Weltanschauung* de l'époque moderne.

Une fois que nous ayons examiné le volet (ii), nous terminerons notre enquête par une analyse des implications du physicalisme en tant que thèse métaphysique omni-englobante, dont le rôle manifeste est celui de dévoiler ce que le réel *est* ultimement (volet iii). Le dernier chapitre de cet essai amorcera donc le travail critique proprement dit en vue duquel les chapitres précédents n'ont fait que préparer le terrain.

Ce long détour s'est révélé nécessaire dans la mesure où l'objectif de notre enquête est de révéler la présence d'une filiation souterraine rattachant non seulement les ambitions, mais aussi (et nous avons envie de dire *surtout*) les échecs des

lorsqu'elle quitte la prison *ne* peut *pas* être déduit de manière *a priori* (à l'aide des raisonnements seulement, sans apport empirique) du set complet de vérités physiques.

[56] (R. Van Gulick, 2004)

« premières » tentatives de 'physicaliser' les énoncés protocolaires aux aspirations métaphysiques du physicalisme contemporain qui cherche à déduire de façon *a priori* les énoncés contenant des termes mentaux à partir des énoncés physiques. Plusieurs étapes ont balisé ce cheminement. Dans un premier temps, nous avons pu constater que la physique offre (ou du moins ambitionne à offrir) ce que M. Bitbol appelle un « genre universalisé de compte rendu des phénomènes, y compris mentaux »[57]. Nous avons vu également que d'assez nombreux auteurs adhèrent à la définition de F. Jackson[58] selon laquelle la mission fondamentale de la métaphysique physicaliste est de résoudre ce que l'on peut appeler « le problème de la localisation »[59] : il faut trouver une place pour l'esprit dans l'ameublement naturel du monde, ou pour utiliser un vocabulaire plus près du thème qui nous intéresse ici, il faut trouver une place pour l'expérience qualitative perceptuelle des couleurs dans une organisation essentiellement physique du réel. Autrement dit, pour utiliser l'heureuse expression de S. Blackburn[60], il s'agit de rendre compte de la place « infortunée » qu'occupent les faits mentaux dans le monde « fortuné » des vérités et faits physiques. Dans la deuxième partie de ce travail nous avons insisté sur le formidable impact de l'argument de la connaissance de Jackson (1982/1986) dar la réorganisation des débats autour de la polarité phénoménal/physique. Comme nous l'avons déjà mentionné, l'histoire de Mary est devenue en quelque sorte l'échiquier sur lequel les diverses subdivisions de la philosophie analytique actuelle croisent l'épée et ravivent de vieilles querelles philosophiques sans âge. Mais surtout, ces querelles trahissent, à notre sens, les symptômes d'un malentendu profond qui corrode l'armature sur laquelle repose certains de nos schémas conceptuels les plus en vogue, schémas que l'on doit apprendre, pensons-nous, à s'en défaire.

[57] M. Bitbol, (2000). *Physique et philosophie de l'esprit*. Paris: Flammarion, p.11.
[58] (F. Jackson, 1998)
[59] Le thème de la place de l'esprit dans la nature fait manifestement écho au célèbre ouvrage de C.D. Broad, (1925), *Mind and its Place in Nature*.
[60] S. Blackburn, (2009), 'Analysis, Description, and the *A priori*?' In I. Ravenscroft (ed.), *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*, op. cit., p. 24.

7 Physicalisme *a priori* et philosophie en fauteuil : un vieux ménage

> "All right", said the Cat; and this time it vanished quite slowly, beginning with the end of the tail, and ending with the grin, which remained some time after the rest of it had gone.
>
> "Well, I have often seen a cat without a grin", thought Alice; "but a grin without a cat! It's the most curious thing I ever saw in all my life!"
>
> Lewis Carroll, *Alice's Adventures in Wonderland*

Au terme de notre périple autour des principaux points forts de l'affrontement entre les avocats et les adversaires du physicalisme, on s'aperçoit que plus les échanges critiques entre les protagonistes de ce duel philosophique s'aiguisent, plus les enjeux contenus dans l'affaire des rapports entre le mental et le physique s'épurent et dévoilent leur armature métaphysique souterraine. Car la question de savoir s'il est en principe possible de déduire de manière *a priori* la totalité des vérités macroscopiques à partir d'une hypothétique description complète du monde en termes microphysiques *P*, et donc de construire des explications réductrices *a priori*, ouvre la boîte de Pandore des querelles autour des limites qui séparent les vérités « pures » de l'entendement des vérités a posteriori factuelles, obligeant ainsi les physicalistes à affronter les vieux démons de la métaphysique. En clair, il s'agit d'établir si les faits nouveaux que Mary apprend lorsqu'elle quitte sa prison peuvent ou *ne* peuvent *pas* être déduits de manière *a priori* (à l'aide des raisonnements seuls, sans apport empirique) de l'ensemble complet de vérités physiques dont elle dispose durant sa captivité.

Ce chapitre s'agencera de la manière suivante. Dans la section 7.1. nous passerons brièvement en revue le physicalisme *a priori* tel qu'il est défendu par le « deuxième » Jackson, pour bifurquer ensuite vers le rôle de l'analyse conceptuelle dans le principe de passage *a priori* depuis le niveau « infortuné » des faits phénoménaux vers le niveau « fortuné » des faits physiques. Dans la section 7.2. nous argumenterons, à la suite de S. Blackburn[1] et W. Lycan[2], que le physicalisme *a priori* non seulement échoue dans ses tentatives acrobatiques de relier le versant empirique de

[1] S. Blackburn, (2009), 'Analysis, Description, and the *A priori*?' In I. Ravenscroft (ed.), *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*, op. cit.

[2] W. Lycan, (2009), 'Serious Metaphysics: Frank Jackson's Defense of Conceptual Analysis'. In I. Ravenscroft (ed.), *Minds, Ethics, and Conditionals: Themes From the Philosophy of Frank Jackson*, op. cit.

l'analyse à son versant logico-métaphysique, mais aussi il hérite tacitement, sous un habillage différent, certes, mais bien lisible à une lecture attentive, les difficultés du physicalisme sémantique carnapien. L'examen critique de ces points nous amènera au plus près du cœur même de la thèse que nous voulons défendre ici (section 7.3), i.e., l'idée que l'acquisition de ce lourd héritage place de manière assez surprenante le physicalisme sur le même banc d'accusés que les doctrines dualistes, ses rivales. Monisme matérialiste et dualisme se dévoilent ainsi comme les deux visages de Janus d'un seul et même courant de pensée souterrain dont le bien-fondé se voit aujourd'hui rudement mis à l'épreuve par certains découvertes scientifiques émanant, ironiquement, du noyau dur des recherches physiques, à savoir la mécanique quantique. Cette analyse servira d'ailleurs d'excellent tremplin pour atteindre la dernière étape de notre travail, où l'on pourra enfin démasquer l'intrus, le présupposé souterrain à l'origine de tant d'égarements. Afin de prendre la véritable mesure du « surnaturalisme » latent qui contamine les débats actuels autour du mind-body problem, il convient de situer avec plus de précision les pions du physicalisme sur l'échiquier des ambitions de la métaphysique en général. Pour ce faire, nous allons invoquer quelques fantômes du passé, c'est-à-dire nous allons raviver la vieille querelle entre Carnap et Quine concernant la nature des entités abstraites. La discussion de ces points nous replongera dans l'atmosphère familière des disputes philosophiques sans âge, et nous permettra ainsi de renouer avec les questionnements avec lesquels nous avons amorcé notre enquête. Bien plus, nous verrons que la rencontre explosive, si l'on peut appeler ainsi, entre les métaphysiciens « en fauteuil » d'une part, et les partisans des stratégies naturalistes de « laboratoire » d'autre part, a ranimé récemment, d'une manière pour le moins étonnante, la question « métamétaphysique»[3] de l'être.

Dans la section 7.4. la conclusion sera que l'examen de ces points permet d'identifier deux possibles « dogmes du physicalisme », pour paraphraser la célèbre expression quinienne. Le physicalisme récusé par la physique ! Voici un titre accusateur au parfum de scandale, récemment avancé dans les travaux d'un certain nombre de philosophes de la science qui s'inscrivent en faux contre la métaphysique *a priori*, coupable, à leurs yeux, d'ignorer ou de mal interpréter la *pratique réelle* de la

[3] Voir D. Chalmers, D. Manley & R. Wasserman (eds.), (2009), *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford University Press.

physique. La critique de cette « métaphysique domestiquée » [4] ou « naïve » fait écho à un projet plus vaste qui vise à naturaliser les différentes branches de la philosophie, allant jusqu'à revendiquer la « naturalisation » de la phénoménologie[5] ou de la métaphysique elle-même.

Enfin, nous ne pouvons pas terminer ce travail sans livrer notre propre « solution » au duel qui oppose partisans et détracteurs du physicalisme. Le dernier chapitre (8) présentera donc un argument pour concilier Mary avec ses « ennemis » physicalistes. Plus spécifiquement, nous défendrons l'idée qu'une fois l'intrus mythologique responsable du clivage subjectif/objectif est éliminé, la tension attisant ce conflit entre les adversaires et les défenseurs des *qualia* s'évapore d'elle-même. Ainsi, Djinn pourra enfin regagner la lampe d'Aladin, les pseudo-problèmes s'évaporant par là même.

7.1 Le physicalisme *a priori*

A la suite du renversement autocritique opéré dans son livre paru en 1998, F. Jackson[6] devient l'artisan d'une conception physicaliste connue sous l'appellation de « physicalisme *a priori* » et qui possède le grand mérite d'exploiter le versant métaphysique dans l'affaire des rapports états physiques/états phénoménaux.

Réduit à sa plus simple expression, le physicalisme *a priori* affirme qu'il est possible de dériver de manière *a priori* des énoncés « infortunés » du sens commun à partir des énoncés scientifiques « fortunés ». A l'appui de son raisonnement, Jackson fait appel à la notion cruciale de « dérivabilité conceptuelle » ou de « déduction *a priori* » du mental à partir du physique. Dans ce qui suit nous soutiendrons que cette véritable « acrobatie » argumentative à laquelle se livre Jackson dans sa tentative de greffer l'empiricité du physicalisme sur l'immunité souveraine d'une analyse conceptuelle de type *a priori*, présente un certain air de famille, si l'on peut dire ainsi, avec la tentative avortée de son prédécesseur, le physicalisme sémantique carnapien, de

[4] J. Ladyman & D. Ross (avec D. Spurrett & J. Collier), (2007): *Every Thing Must Go: Metaphysics Naturalised*, op. cit.
[5] Voir J. Petitot, F. J. Varela, B. Pachoud, & Jean-Michel Roy (1999), *Naturalizing Phenomenology*, Stanford University Press.
[6] F. Jackson, (1998), *From Metaphysics to Ethics: A Defense of Conceptual Analysis*, op. cit.

greffer les énoncés protocolaires sur l'édifice logique, formel, de la syntaxe. Conséquemment, ils prêtent le flanc aux mêmes difficultés.

L'idée que nous défendons dans ce chapitre est la suivante : à l'instar du physicalisme sémantique de « première génération », le physicalisme *a priori* d'aujourd'hui nourrit ses ambitions à partir d'un véritable réseau souterrain de présupposés implicites qui l'amènent tôt au tard à affronter le « problème de la médiation » ou de l'articulation entre les différentes interfaces explicatives du réel. Ce type de problème, habituellement maquillé *ad hoc*, à l'aide des sauts conceptuels, trahit la présence d'une pratique généralisée qui consiste à opérer des découpages-clivages explicatifs sur l'axe sujet/objet et leur produit dérivé, i.e. l'axe subjectivité/objectivité.

Dans le cas précis qui nous intéresse ici, il s'agit de dénoncer le présupposé tacite qui alimente les espoirs du physicalisme de « fonder la science de l'esprit sur une physique idéalement affranchie de ses propres limites fondationnels »[7].

7.1.1 Dérivabilité a priori et a posteriori

Soit P une propriété physique et Q une propriété phénoménale. Du point de vue logique, l'opération de déductibilité *a priori*, ou d'implication *a priori*, est une des plus fortes, sinon la plus forte. Selon l'interprétation standard, une implication est *a priori* si elle ne fait appel à aucun élément de verticalité factuelle ou empirique extérieure, autrement dit si le processus de déduction se déroule à l'aide de l'entendement seul, et à l'intérieur de la syntaxe horizontale, logique. Par exemple, P implique Q lorsque le conditionnel matériel $P \supset Q$ est de type *a priori*. Nous avons vu précédemment qu'une des implications les plus spectaculaires de l'argument de la connaissance, c'est la remise en cause la relation de déductibilité *a priori* de Q à partir de P. Si X connaît seulement P et X ne peut pas en déduire Q de façon *a priori* à partir de P (comme le KA le suggère), alors on peut tirer licitement la conclusion que P n'implique pas Q.

Par ailleurs, si l'on tient compte du fait que la forme paradigmatique de l'implication ontologique est la « nécessitation » - i.e. P implique nécessairement Q lorsque le conditionnel matériel $P \supset Q$ est métaphysiquement nécessaire, ou lorsqu'il est métaphysiquement impossible d'avoir P sans avoir Q – il est clair que le physicaliste se

[7] (M. Bitbol, 2000:10)

voit contraint d'adopter la clause de l'exhaustivité selon laquelle *P* implique *tous* les faits possibles, *Q* y compris. A l'opposée, dans le camp des antiphysicalistes, il suffit de prouver qu'une seule entité *Q* n'est pas impliquée par *P*, pour provoquer la chute des ambitions métaphysiques de complétude et d'exhaustivité du matérialisme physicaliste.

Le conflit revêt dès lors la forme suivante : le physicalisme, insiste Jackson[8], est une doctrine métaphysique de plein droit. Il vise à expliquer ce que le réel est, et comment le réel se manifeste à l'expérience. D'autre part, l'argument de la connaissance affirme qu'aucune description du monde en termes purement physiques ne serait capable de nous permettre de *déduire* les états phénoménaux à partir des faits physiques. La question que l'on doit se poser alors est la suivante : comment expliquer le fait qu'une doctrine concernant la dérivabilité ou la déductibilité conceptuelle puisse concerner, voire menacer une doctrine *métaphysique* ?

Quelques précisions terminologiques pour commencer. Un physicaliste *a priori* partage avec les défenseurs du physicalisme standard l'idée que le physicalisme est une doctrine contingente et *a posteriori* visant à expliquer la nature ultime du réel (esprit y compris). Mais à la différence de ce dernier, le physicalisme *a priori* allègue deux thèses supplémentaires. En premier lieu, il affirme que le domaine physique implique de façon nécessaire *a priori* le domaine du mental. Autrement dit, la détermination de la nature du physique entraîne nécessairement la détermination de la nature du mental. Pour donner ici un exemple cité par Jackson[9]: supposons que la caractéristique la plus populaire de Dieu est celle d' « être haut de deux mètres ». A partir de là, l'énoncé *p* « La caractéristique la plus populaire de Dieu implique nécessairement *a priori* la propriété d'être d'un mètre de haut » est vrai, malgré le fait que *p* n'est ni *a priori*, ni nécessaire.

A présent, admet Jackson, il est vrai que la question de savoir qu'est-ce que pour une propriété d'être nécessairement *a priori* déductible d'une autre propriété, reste hautement controversée. Mais pour les objectifs spécifiques de ce chapitre, on peut se

[8] F. Jackson, (1995), 'Postscript on 'What Mary didn't know'", *in* P. Moser & J. Trout (eds.), op. cit., pp. 184–189.

[9] "Suppose that God's favourite feature is being two meters tall, then 'God's favorite feature *a priori* necessitates being at least one meter tall' is true despite the fact that the sentence is neither *a priori* nor necessary." F. Jackson, (2005), 'The Case for *A priori* Physicalism', *in* Ch. Nimtz & A. Beckermann (eds.), *Philosophy-Science-Scientific Philosophy, Main Lectures and Colloquia of Gap 5, Fifth International Congress of the Society for Analytical Philosophy*, Mentis, p. 252.

contenter provisoirement d'une définition standard relativement tolérée dans la littérature analytique :

(1) (P1) entraîne (implique) de façon nécessaire et *a priori* (P2) si et seulement si comprendre la nature de (P1), et de respectivement (P2), nous permet implicitement de se rendre compte que si (P1) est instancié, alors (P2) est instancié *aussi*.[10]

Autrement dit, et pour revenir à l'exemple évoqué plus haut, une personne qui pense que Napoléon peut avoir la propriété d'« être de deux mètres de haut » sans avoir la propriété d'« être d'un mètre de haut », ignore manifestement la signification des deux propriétés en cause. Il est généralement admis, reconnaît Jackson, que la propriété d'apriorité est une propriété sémantique des mots et des assertions. Mais, insiste-t-il, les mots et les énoncés n'étant à leur tour que des structures physiques (types et tokens), cette objection ne touche pas la définition initiale. Nous aurons l'occasion de revenir sur ces aspects lorsque nous dresserons une analyse critique qui mettra en relief les liens profonds qui rattachent le physicalisme *a priori* de Jackson au physicalisme sémantique de Carnap. Nous verrons ainsi que l'entreprise physicaliste *a priori* repose sur l'insertion d'un saut conceptuel de la même veine que celui qui opère Carnap lorsqu'il bâtit son physicalisme sémantique en traversant les différents niveaux de discours depuis le radicalement subjectif vers le pleinement objectif dans la caractérisation des énoncés protocolaires. Pour l'instant, il nous faut brosser le portrait général de la situation, sans trop insister sur ses failles, qui seront pleinement éclairées dans la section 7.3. de ce chapitre.

Revenant à présent au physicalisme *a priori* de Jackson, il convient de retenir ceci : affirmer que le mental est déductible de façon nécessaire et *a priori* à partir du physique revient à dire que certaines propositions élaborées à partir d'expressions visant le domaine du mental sont déductibles de façon nécessaire et *a priori* à partir des propositions élaborées d'expressions relatives au domaine physique. Ceci nous amène au plus près de la deuxième caractéristique qui différencie le physicalisme *a priori* de son cousin, le physicalisme non-réductionniste : la vérité des énoncés psychologiques comme « J'éprouve de la douleur » ou « Fred voit du rouge » - tout comme celle des énoncés plus vastes de type « L'inflation économique a baissé au mois de juin en

[10] *(P1) a priori* necessitates *(P2)* iff one's grasp of what it is to be a *(P1)* and what it is to be a *(P2)* allows one to see that if *(P1)* is instantiated so is *(P2)*." (F. Jackson, 2005: 252-3)

France » - est déductible de façon *a priori* à partir d'une conjonction adéquate de propositions physiques.

Pour résumer, le physicalisme *a priori* rajoute à la thèse de la déductibilité nécessaire des phénomènes mentaux à partir des lois de la nature physiques, (thèse stipulée d'ailleurs par le physicalisme *a posteriori*, et dont il partage la clause d'interdiction des propriétés non-physiques additionnelles), deux thèses supplémentaires, étroitement liées à la relation de nécessitation *a priori* :

(i) Les propriétés mentales (et plus généralement les propriétés psychologiques, économiques, etc.) du monde actuel sont nécessairement et *a priori* déductibles à partir de propriétés physiques de notre monde actuel.

(ii) Le langage contenant des expressions mentales (et plus généralement des expressions psychologiques, économiques, etc.) est nécessairement et *a priori* déductible à partir d'un langage contenant des expressions physiques à propos du monde actuel.

Une façon commode d'insister sur les dissemblances entre le physicalisme *a posteriori* et *a priori* c'est de faire appel à l'exemple paradigmatique « l'eau = H_2O ». On s'accorde à dire que cet énoncé est *a posteriori* (c'est-à-dire découvert empiriquement) bien que nécessaire (c'est-à-dire vrai dans tous les mondes possibles). Ou pour exprimer la même idée autrement : bien qu'il soit métaphysiquement nécessaire que toute eau soit H_2O, il est possible de concevoir *deux situations épistémiques* ou conceptuelles distinctes, à savoir : (a) l'eau *est* H_2O et (b) l'eau *n*'est *pas* H_2O, mais XYZ, disons. Il aura fallu que des recherches expérimentales minutieuses soit faites sur des échantillons d'eau qui remplit les rivières, tombe du ciel, fond des glaciers, etc., pour qu'on arrive à la conclusion que (a) est vrai et (b) non. A présent, observe Jackson[11], si l'on transporte l'analogie sur le terrain des rapports entre les états mentaux et les états physiques, on obtient ceci :

[11] "*A posteriori* physicalists hold that there are two epistemic or conceptual possibilities – that our world is exactly as it is physically with the psychology it in fact has, and that our world is exactly as it is physically with no psychology at all – that is what it is to hold that the connection is *a posteriori*. (…)*A priori* physicalists deny that there are two epistemic possibilities: the world physically exactly like ours minus psychology, and the world exactly like ours with psychology, but the doctrine distinctive of *a posteriori* physicalism is precisely that there are two epistemic possibilities and I cannot see how *a posteriori* physicalists can consistently give reason to hold that it is the second epistemic possibility and not the first which is realized." (F. Jackson, 2005: 254-5)

(1) les physicalistes *a posteriori* affirment qu'il existe deux possibilités épistémiques et conceptuelles distinctes, à savoir : (1_A) notre monde actuel est exactement tel que nous le connaissons du point de vue physique, c'est-à-dire avec les faits psychologiques qu'il abrite ; et (1_B) notre monde actuel est exactement tel que nous le connaissons du point de vue physique, mais n'abritant aucun fait psychologique tout court (un monde « zombie »).

(2) les physicalistes *a priori* rejettent la thèse des deux situations épistémiques, autrement dit, ils nient qu'il serait possible de concevoir (1_B).

Les deux versions du physicalisme partagent cependant le fondement commun que représente la thèse de la survenance (S):

(S) Tout double physique minimal du monde actuel est un double *simpliciter* du monde actuel.[12]

Certains auteurs comme B. McLaughlin[13] ont noté que le principe (S) nous informe, entre autres, que le monde actuel est un double physique minimal de lui-même. Cependant, si le fait que le monde actuel est un double physique de lui-même est une vérité *a priori*, le fait que le monde actuel soit un double physique *minimal* de lui-même est un fait *a posteriori*. A partir de là, étant donné que le physicalisme implique la vérité de (S), il s'ensuit que le physicalisme est *a posteriori*. Pour résumer, on peut concéder aux deux versions du physicalisme (*a priori* et *a posteriori*) la base commune suivante :

(S′) Il est *a priori* vrai que si (S), alors toutes les vérités sont déduites à partir d'une (1) description physique détaillée et exhaustive du monde actuel ; que (2)

[12] "Consider any possible world that is a minimal physical duplicate of our world. It is, that is, exactly like ours in every physical respect: it is physical individual, property and relation exactly like our world, and moreover it contains nothing extra; it contains nothing more than it has to in order to be physically exactly like our world. (We can count the necessarily existing entities, if there are any, that all worlds have in common as trivially physical for our purposes here.) Materialists who hold that materialism is a complete account of our world, or a complete account of our world as far as the mind is concerned—materialists who are, that is, not some kind of dual attribute dualist—must hold that these minimal physical duplicates are psychological duplicates of our world. They must, that is, hold the following supervenience thesis: (S) Any world that is a minimal physical duplicate of our world is a psychological duplicate of our world." F. Jackson, (1995), 'Postscript on 'What Mary didn't know'", in P. Moser & J. Trout 1995, op. cit., p. 192. (Voir aussi F. Jackson, 1998:12)

[13] B. P. McLaughlin, (2005), '*A priori vs A posteriori* Physicalism' in Ch. Nimtz & A. Beckermann (eds.), *Philosophy-Science -Scientific Philosophy, Main Lectures and Colloquia of Gap 5*, op. cit., p. 269.

cette description physique détaillée et exhaustive est une description complète du monde ; et, enfin (3) le monde est un double physique minimal de lui-même[14].

7.1.2 Le principe du passage *a priori*

L'argument du passage *a priori* repose d'emblée sur le constat que tout matérialiste est tributaire de la thèse de la survenance. A partir de là, il s'agit de montrer que tout matérialiste adhérant à la thèse de la survenance se voit obligé d'accepter, par ricochet, la thèse de la déductibilité *a priori* du mental à partir du physique. Pour voir cela, affirme Jackson[15], il suffit de considérer le cas suivant : appelons Ø l'énoncé qui nous donne une description physique complète et détaillée qui est vraie pour le monde actuel, ainsi que pour tous les autres doubles physiques minimaux du monde actuel, et fausse ailleurs. Appelons Ω n'importe lequel énoncé vrai concernant la nature psychologique du monde actuel. Ω est vrai pour notre monde, ce qui veut dire que tous les autres mondes dans lesquels Ω est faux diffèrent du point de vue psychologique de notre monde actuel. A présent, si la thèse (S) est vraie, alors chaque monde dans lequel Ø est vrai, est un double psychologique de notre monde. Mais dans ce cas, tout monde dans lequel Ø est vrai est aussi un monde dans lequel Ω est vrai. Autrement dit, Ø implique Ω. Par conséquent, malgré le fait que le matérialisme est une doctrine *métaphysique*, la thèse de la complétude oblige son défenseur à endosser les conséquences *logiques* contenues dans la thèse de la déductibilité du mental à partir du physique.

Comment relier cette thèse à la situation épistémique de Mary ? Jackson[16] introduit à cette fin l'argument de la déductibilité *a priori*. Considérons l'inférence suivante :

EAU-H_2O	(E1) H_2O recouvre la majeure partie de la Terre. (E3) Donc, l'eau recouvre la majeure partie de la Terre.

Le conditionnel « Si H_2O recouvre la majeure partie de la Terre, alors l'eau recouvre la majeure partie de la Terre » est nécessairement vrai, mais il n'est pas

[14] (B.P. McLaughlin, 2005: 269)
[15] (F. Jackson, 1995: 193)
[16] *Ibid.*, pp. 193-4

nécessaire *a priori*, puisque l'identité « H_2O = eau » a été découverte *a posteriori*, comme l'a fort bien montré Kripke. On pourrait objecter, admet Jackson[17], que l'argument 'EAU- H_2O'nous fournit un excellent exemple de la manière dont on devrait comprendre les relations entre le mental et le physique, *sans* faire intervenir la notion d' « *a priori* ». Car, même si une description physique suffisamment détaillée de notre monde implique logiquement la description psychologique de ce même monde, il se peut néanmoins que la deuxième ne soit pas déductible de façon *a priori* de la première. Par conséquent, un matérialiste réfractaire à la thèse de la déductibilité *a priori*, pourrait se retrancher derrière la survenance du mental sur le physique, *sans* adopter la thèse de la déductibilité *a priori* du mental à partir du physique. A partir de là, il est facile de montrer que le matérialisme n'est pas concerné par le défi lancé par l'argument de la connaissance: car bien que l'information dont dispose Mary détermine et fixe logiquement toute l'information concernant la latitude psychologique du monde (états sensoriels et conscience phénoménale y compris), ce savoir ne lui permet pas d'en déduire, même en principe, ce que les états psychologiques *sont*.

Jackson s'inscrit en faux contre cette position, arguant qu'elle est le résultat d'un malentendu issu d'une interprétation erronée du concept de « nécessaire *a posteriori* ». Pour voir cela, revenons à l'argument 'EAU-H_2O'. Selon Jackson, la signification du terme « eau » est un désignateur rigide dont la référence est fixée par la locution « substance qui occupe le rôle 'aqueux' »[18] et qui est caractérisée de façon générale par une liste d'attributs plus ou moins exhaustive : substance sans odeur, transparente, qui tombe du ciel, nécessaire à l'apparition de la vie sur la planète, etc. A présent, et c'est ici le point essentiel de l'argument de Jackson, le fait de considérer le terme « eau » en tant que désignateur rigide de la substance qui occupe le rôle 'aqueux', nous permet de considérer un énoncé de genre : « L'eau = la substance qui occupe le rôle 'aqueux' » comme étant contingent *a priori*.

Par conséquent, malgré le fait que l'inférence 'EAU- H_2O' évoquée plus haut n'est pas valide du point de vue de la déductibilité *a priori*, si l'on y rajoute la prémisse supplémentaire (E2) on obtient un résultat qui soit à la fois *a priori* et nécessaire.

[17] *Ibid*., p. 194
[18] "The stuff that fills the watery role" (F. Jackson, 1995: 195)

EAU A PRIORI

(E1) H_2O recouvre la majeure partie de la Terre.

(E2) *H_2O est la substance qui occupe le rôle 'aqueux'.*

Donc,

(E3) L'eau recouvre la majeure partie de la Terre[19].

L'argument a été repris et développé dans un article ultérieur, coécrit avec son compatriote, le philosophe australien D. Chalmers[20]. Dans cet article, Jackson et Chalmers[21] adressent la possibilité de déduire *a priori* les vérités non-physiques à partir d'une description du monde en termes microphysiques. Conséquemment, ils articulent leur argumentation sous la forme d'une enquête à quatre volets interrogatifs :

(1) Y-a-t-il une relation de déductibilité *a priori* des vérités macroscopiques à partir des vérités microphysiques ?

(2) S'il n'y a pas de relation de déductibilité *a priori* des vérités phénoménales à partir des faits microphysiques, est-ce que l'explication réductrice du phénoménal échoue ?

(3) S'il n'y a pas de relation de déductibilité *a priori* des vérités phénoménales à partir des faits microphysiques, est-ce que le physicalisme échoue dans sa mission d'expliquer le phénoménal ?

(4) Y-a-t-il une relation de déductibilité *a priori* des vérités phénoménales à partir des vérités microphysiques ?

Chalmers et Jackson répondent par « oui » aux trois premières questions, laissant la quatrième ouverte, tandis que N. Block et R. Stalnaker[22] répondent par non. Reprenons l'exemple d'un fait macroscopique tel qu'évoqué plus haut :

(E3) « L'eau recouvre la majeure partie de la Terre ».

---

[19] (F. Jackson, 1995: 196)

[20] D. Chalmers & F. Jackson, (2001), 'Conceptual Analysis and Reductive Explanation', *Philosophical Review*, 110:315-61.

[21] (1) Is there an *a priori* entailment from microphysical truths to ordinary macroscopic truths? (2) If there is no *a priori* entailment from microphysical truths to phenomenal truths, does reductive explanation of the phenomenal fail? (3) If there is no *a priori* entailment from microphysical truths to phenomenal truths, is physicalism about the phenomenal false? (4) Is there an *a priori* entailment from microphysical truths to phenomenal truths?" (*Ibid.*, p. 315)

[22] N. Block & R. Stalnaker, (1999), 'Conceptual Analysis, Dualism and the Explanatory Gap', *Philosophical Review*, 108(1): 1–46.

Selon Chalmers et Jackson, on peut déduire des faits de ce type de manière *a priori*, à condition que l'on dispose d'une explication réductrice, autrement dit en s'appuyant sur :

(a) une description complète de l'état du monde en termes microphysiques ;

(b) une analyse conceptuelle des notions utilisées pour exprimer le fait respectif.

Selon Chalmers et Jackson, les vérités macrophysiques découlent de façon *a priori* des vérités microphysiques[23]. Par conséquent, il est possible de savoir *a priori* que le conditionnel matériel *P* & *M* est vrai, où *P* désigne « la conjonction des vérités microphysiques sur le monde » et *M* une vérité du sens commun arbitraire concernant des objets et des propriétés macroscopiques, comme l'eau, par exemple, ou la vie : « il y a des nombreux êtres vivants »[24].

Le point clé de leur argumentation est la thèse que tout ce qui est requis pour savoir que *P* & *M*, est l'analyse conceptuelle *a priori*. Autrement dit, l'analyse conceptuelle rend possible « la métaphysique en fauteuil »[25] qui extrait *a priori* toutes les vérités de la seule description complète du monde en termes microphysiques. La défense de l'analyse conceptuelle repose sur la distinction entre les intensions-C et les intensions-A[26]empruntée à la sémantique bidimensionnelle élaborée par G. Evans[27], M. Davies et L. Humberstone[28].

Selon cette acception, la possession d'un concept familier comme « eau » suffit pour déterminer son *intension première* ou intension-A[29]. L'intension première détermine l'extension du concept dans chaque monde possible, à condition que ce monde possible soit le monde réel *actuel*. Pour montrer comment la simple possession d'un concept peut déterminer son intension-A de manière *a priori*, il convient d'introduire le concept de « monde épistémiquement possible»: un monde est épistémiquement possible s'il est compatible avec nos connaissances en cours.

[23] (D. Chalmers & F. Jackson, 2001:316)
[24] *Ibid.*, p.317
[25] F. Jackson, (1993), 'Armchair Metaphysics', in J. Hawthorne & M. Michael (eds.), *Philosophy in Mind*, Amsterdam: Kluwer.
[26] « C » désigne la classe des éléments contrefactuels, et « A » la classe des éléments actuels.
[27] G. Evans, (1979), 'Reference and Contingency', *The Monist* xii, pp. 161–189.
[28] M. Davies & L. Humberstone, (1980), 'Two Notions of Necessity', *Philosophical Studies* 38 (1).
[29] L'intension première est une fonction qui associe une extension à des paires de mondes possibles (l'un étant le monde d'énonciation, l'autre le monde d'évaluation), où le monde d'évaluation est identique au monde d'énonciation. C'est cette contrainte qui distingue l'intension première de l'intension seconde. Voir Frank Jackson, (2004), 'Why We Need A-Intensions', *Philosophical Studies* 118 (1-2):257-277.

L'intension-A d'un concept est la fonction qui associe à chaque monde épistémiquement possible son extension dans ce monde. Nous avons vu, lorsque nous avons évoqué l'analyse lewisienne des propriétés fondamentales, que nous ne pouvons pas savoir *a priori* dans lequel parmi ces mondes épistémiquement possible nous nous trouvons réellement, mais le contenu du concept nous indique de manière *a priori* à quoi il s'applique dans chacun de ces mondes hypothétiques. Par conséquent, si l'on revient à l'exemple 'EAU-H_2O', l'analyse conceptuelle consiste à découvrir, en s'appuyant sur des circonstances hypothétiques d'énonciation et d'évaluation, les critères que nous utilisons (consciemment ou pas) pour déterminer ce qui tombe sous son extension dans un monde donné : c'est une substance inodore, incolore, transparente et liquide qui coule dans les rivières, que l'on peut boire, qui tombe du ciel lorsqu'il pleut, etc. Cette énumération désigne l'information *a priori* associée au concept lorsqu'on possède le concept. Pour plus de commodité, appelons l'intension-A du concept « eau », « la substance aqueuse » (« the watery stuff »).

Selon Chalmers et Jackson, «si un sujet possède le concept "eau", alors une information suffisante sur la distribution, le comportement, et l'apparence d'agrégats de molécules H_2O met le sujet en posture de savoir que l'eau est H_2O, de savoir où il y a de l'eau et où il n'y a pas d'eau, et ainsi de suite. Ce savoir conditionnel ne requiert que la possession du concept et la réflexion rationnelle, et ne requiert donc aucune connaissance *a posteriori* supplémentaire. La possession d'un concept comme (…) « eau » procure une capacité conditionnelle pour identifier l'extension du concept dans le contexte d'une possibilité épistémique conditionnelle. (…) Pour la raison que toute l'information empirique pertinente est présente dans l'antécédent du conditionnel, l'information empirique ne joue aucun rôle essentiel pour justifier la croyance dans ce conditionnel. Donc, (…) [ce conditionnel] est *a priori*»[30].

Pour mieux illustrer cette utilisation des intensions-A, on peut faire appel à une variation sur la désormais classique expérience de pensée de Putnam. Imaginons un monde possible – le Monde Jumeau – dans lequel le mot « eau » est un terme qui fait référence à un phénomène naturel semblable à l'eau dans le nôtre. Autrement dit, il fait référence à une composition moléculaire spécifique, et est caractérisé par des descriptions familières et superficielles comme « substance qui remplit les lacs », ou

[30] (D. Chalmers & F. Jackson, 2001:324-5). Voir aussi F. Jackson, (2007), 'A priori physicalism', in B. P. McLaughlin & J. D. Cohen (eds.), *Contemporary Debates in Philosophy of Mind*, Blackwell.

« substance avec laquelle on peut se laver les mains ». En gardant le même contexte grammatical de la langue française, la question qui se pose est : qu'est-ce que change d'un monde à l'autre ? Ce qui change c'est la structure qui sous-tend la substance « eau », et qui n'est pas H_2O, mais XYZ, disons. Donc, dans le Monde Jumeau, « eau » veut dire « XYZ ». Cela étant acquis, nous sommes en mesure à présent de clarifier un peu plus la signification du concept d' « intension-A » chez Jackson : pour tout monde possible dans lequel le signe « eau » est utilisé de manière analogue à l'utilisation qu'on lui confère dans le monde actuel (bien qu'il fasse référence à des substrats sous-jacents différents : H_2O, XYZ, etc.), le terme « eau » va désigner non pas H_2O, mais la substance sous-tendante du monde respectif. Autrement dit, il va désigner l'entité qui occupe le rôle « aqueux » défini en termes de fixateurs de référence kripkéens standard. A partir de là, Jackson bâtit son argument sur l'hypothèse[31] selon laquelle le terme « eau » est entouré d'une sorte de halo sémantique, autrement dit d'une intension souple et vague qui se rajoute à sa signification « normale », rigide, et cela même dans le monde actuel. Par conséquent, l'intension-halo du mot « eau », dans notre exemple, est « substance qui occupe le rôle aqueux ».

Avec ces éléments d'arrière-plan établis, nous sommes en mesure à présent de dresser l'armature complète de l'argument de Jackson et Chalmers. Considérons les deux propositions suivantes :

(a) 60% de la surface de la Terre est couverte d' H_2O.

(d) 60% de la surface de la Terre est couverte d'eau.

Précisons que l'objectif ici, comme le relève fort bien S. Blackburn[32], c'est de relier la vérité « infortunée » (d) à la vérité « fortunée » (a). Mais comment faut-il s'y prendre ? Car le passage de (d) à (a) n'est pas *a priori*. C'est à cet instant précis que Jackson sort de son chapeau magique la notion-halo de « substance que occupe le rôle aqueux » pour amortir le choc conceptuel et ontologique qui existe entre "eau" et "H_2O". Cette manœuvre l'amène à rajouter les deux clauses suivantes :

(b) L'eau est la substance qui occupe le rôle 'aqueux'.

(c) H_2O est la substance qui occupe le rôle 'aqueux'.

[31] Comme le remarque fort bien W. Lycan, cette hypothèse, est énoncée par Jackson sans aucun appui argumentatif. (W. Lycan, 2009: 72)
[32] (S. Blackburn, 2009: 31)

Jackson juge (b) comme étant *a priori*. Il sait que (c) est un énoncé empirique. Mais *on peut* déduire *a priori* (d) à partir de (a) & (b) & (c). Et puisque (b) est *a priori*, nous nous retrouvons donc avec deux vérités H_2O (à savoir (a) et (c)) dont aucune des deux ne mentionne le terme suspect d'« eau », mais qui ensemble entraînent *a priori* (d). Et voilà, le tour est joué ! Nous avons une inférence *a priori* en partant de (a) et (c), les vérités « fortunées » vers (d), la vérité « infortunée ». Le principe qui est derrière cette inférence est le suivant :

(H) Si A&B implique *a priori* C, et si B est lui-même *a priori*, alors A implique *a priori* C.

Nous verrons dans la section suivante qu'il existe plusieurs manières d'attaquer cet argument et de dénoncer le tour de passe-passe sur lequel il repose. Essayons à présent de résumer les étapes de l'acheminement vers le physicalisme *a priori*. Dans un premier temps, l'analyse conceptuelle du concept d'*eau* nous procure *a priori* la connaissance de l'intension-A du concept : « la substance aqueuse qui nous est familière »[33]. La connaissance *a priori* de l'intension-A est essentiellement *conditionnelle* dans la mesure où elle est une fonction (du monde, ou du contexte d'énonciation à l'extension dans le monde d'évaluation) : pour connaître la valeur de la fonction (l'extension dans le monde d'évaluation), je dois connaître (*a posteriori*) son argument, autrement dit je dois savoir quel est le monde d'énonciation. L'ambition de Chalmers et de Jackson c'est de montrer que la description *complète* du monde réel en termes microphysiques permet de déduire *a priori* l'énoncé (4*) : « L'eau couvre la majeure partie de la Terre. » à partir de (1*) : «H_2O couvre la majeure partie de la Terre ».

La deuxième étape consiste à mettre à contribution les fruits de l'analyse conceptuelle du mot « eau » qui nous donne en cadeau la vérité *a priori*/analytique suivante :

(2*) « L'eau est la substance aqueuse qui nous est familière ».

Dans un troisième temps, le contexte d'énonciation de (4*) nous donne l'information cruciale (3*) :

(3*) « H_2O est la substance aqueuse qui nous est familière. »

[33] (F. Jackson, 1998: 75)

Considérés conjointement, les énoncés (1*), (2*) et (3*) permettent effectivement d'inférer (4*).

Ayant établi ces éléments d'arrière plan, nous sommes en mesure à présent d'expliciter le principe du passage *a priori*.

Le principe du passage *a priori* affirme que chaque énoncé vrai concernant notre monde est déductible *a priori* à partir d'un énoncé formulable en termes physiques[34]. Cela peut être interprété de deux manières distinctes, (bien que cette distinction ne nous affecte pas ici). On peut soit concevoir un « super-énoncé » vrai formulable en termes physiques qui implique *a priori* chaque énoncé vrai relatif à notre monde. Soit on renverse la perspective et on affirme que pour chaque énoncé vrai à propos de notre monde, il y a un énoncé vrai formulable en termes physiques qui implique *a priori* tous les autres énoncés.

Récapitulons. Nous avons vu que l'argument de Jackson (1998/2003) va ainsi :

(a) 60% de la surface de la Terre est couverte d' H_2O.

(d) 60% de la surface de la Terre est couverte d'eau.

Le passage de (a) à (d) n'est pas *a priori*, concède Jackson. On met alors en scène la clause supplémentaire (b) « L'eau est la substance qui occupe le rôle 'aqueux'. » Selon Jackson, l'énoncé (b) est *a priori* en vertu du « halo sémantique » qui entoure l'intension-A du terme « eau » : « substance potable, transparente, qui tombe du ciel quand il pleut, etc. ». En clair, le statut *a priori* de (b) repose sur l'idée qu'un énoncé de genre « N= est le F » est *a priori* lorsque F spécifie plus ou moins exhaustivement les fixateurs de référence de N. Par conséquent, si (b) est *a priori*, poursuit Jackson[35], alors sa conjonction avec la vérité empirique (c) :

(c) H_2O est **la** substance qui occupe le rôle aqueux.

nous informe que nous avons deux vérités-H_2O qui, ensemble, impliquent *a priori* la conclusion (d). Ce qu'il veut dire qu'il est faux d'inférer à partir du fait que l'énoncé « L'eau est H_2O » est nécessaire *a posteriori*, le fait qu'il *n*'y a *pas* de passage *a priori* depuis la transcription en termes de "H_2O" vers la transcription en termes

[34] "Are physicalists committed to what I call the *a priori* passage principle, the view that for each true statement concerning our world, there is a statement in physical terms that *a priori* entails that statement?" F. Jackson (2003), 'From H2O to Water: The Relevance to *A priori* Passage', in H. Lillehammer & G. Rodriguez-Pereyra (eds.), *Real Metaphysics*. Routledge, p. 155.

[35] (F. Jackson, 2003 : 161)

d'"eau". Nous obtenons ainsi une très spectaculaire méthode de métamorphoser la précarité des énoncés puisant leurs termes dans le vocabulaire « ordinaire » du sens commun, afin de les relier analytiquement à l'infaillibilité des énoncés non-suspects du vocabulaire scientifique.

Tournons-nous à présent vers l'examen critique de cet argument. Enquêter sur ces points nous amènera au plus près de la question centrale de ce chapitre, à savoir le projet de soumettre le physicalisme à l'épreuve du naturalisme.

7.2 Contester le principe du passage *a priori*

Nous allons reprendre en détail l'argument de la déductibilité *a priori* du mental à partir du physique, car il est important de préciser la prémisse que nous voulons rejeter ici. Rappelons-nous, les physicalistes adhèrent pleinement à la thèse de la nécessitation : tout ce qui existe découle *nécessairement* à partir du niveau physique (ou microphysique). Le point le plus controversé concerne la question de savoir si le physicalisme est tributaire à ce que Jackson appelle le principe du passage *a priori*. Plusieurs manières de récuser ce principe ont été avancées. On peut :

(i) soit questionner la clause de l'unicité en (b) : ***une*** substance *versus* ***la*** substance (N. Block & R. Stalnaker) ;

(ii) soit remettre en question l'apriorité présumée de (b) (S. Blackburn & W. Lycan) ;

(iii) soit montrer que l'ontologie réductionniste standard derrière le physicalisme *a priori* repose sur des véritables dogmes désavoués par la pratique réelle de la science physique (J. Ladyman & D. Ross) ;

(iv) soit dénoncer la volonté illusoire de greffer la rationalité des vérités *a priori* sur le tronc sinueux de vérités factuelles, empiriques, comme étant le symptôme d'un saut conceptuel visant à fixer les sables mouvants de nos connaissances « infortunées » de tous les jours, sur le socle solide, voire éternel, des vérités logico-mathématiques universelles. C'est la piste critique que nous allons privilégier ici, et que nous allons élaborer en détail dans la section 7.4. de cet essai.

Mais avant, nous allons passer en revue les autres options critiques.

7.2.1 L'objection de Block et Stalnaker

L'argument du passage *a priori* de Jackson a fait l'objet d'une critique détaillée dans un article signé par N. Block et R. Stalnaker[36]. Leur raisonnement procède en deux étapes, à savoir: (1) l'adhésion à la thèse du fossé dans l'explication, et (2) le rejet de la thèse de l'analyse conceptuelle. Ayant déjà examiné l'argument du fossé dans l'explication (section 4.2), dans ce qui suit nous allons nous arrêter uniquement sur le deuxième volet de leur argumentation, i.e. la critique de la stratégie de l'analyse conceptuelle.

Il est vrai, arguent Block et Stalnaker, que certains concepts (même certains concepts mentaux) se prêtent effectivement à une analyse fonctionnelle en termes de concepts physiques. Mais une chose est sûre, insistent-ils : la conscience phénoménale *ne* se prête *pas* à une telle analyse. Par conséquent, le rejet de l'analyse conceptuelle du phénomène de la conscience en termes physiques ou fonctionnels mène à l'acceptation de la validité de l'argument de la concevabilité de Chalmers, car il n'y a aucune contradiction dans le fait de concevoir un zombie fonctionnellement identique à l'être humain, mais privé de conscience phénoménale.

Block et Stalnaker s'attaquent à l'édifice du passage *a priori* de Jackson (1998/2003) sur plusieurs fronts. D'abord par le biais de la prémisse (b) : « L'eau est la substance familière qui occupe le rôle 'aqueux' ». Plus spécifiquement, ils contestent que la description finie (aussi exhaustive et riche soit elle) de l'intension-A, i.e. de la signification-halo qui entoure le terme « eau » soit capable d'assurer l'apriorité de (b). Autrement dit, « l'affirmation que H_2O est ce qui satisfait (ou l'une des choses qui satisfont) l'intension-A du mot « eau » n'est pas une affirmation microphysique »[37]. L'énoncé (b) n'est pas non plus une vérité *a priori*, ce qui permettrait de l'utiliser au même titre que (c).

Une autre attaque concerne le constat que la description donnée en (c) – « H_2O est la substance qui occupe le rôle aqueux » nous indique que l'argument n'est pas un modèle adéquat d'une discussion visant le principe du passage *a priori* en général. Ils arguent que de nombreux exemples de réduction scientifique n'offrent aucun support à l'idée selon laquelle (b) serait *a priori*.

---

[36] N. Block & R. Stalnaker, (1999), 'Conceptual Analysis, Dualism, and the Explanatory Gap', op.cit.
[37] *Ibid.*, p. 32

Enfin, Block et Stalnaker observent judicieusement que la clause de l'*unicité* de la substance qui occupe le rôle-aqueux est d'une importance capitale pour les enjeux de l'argument de Jackson. (1995/2003) Par exemple, si l'on part de la prémisse (i) que H_2O est *une* des « choses », parmi tant d'autres possibles, qui occupent le rôle-aqueux, et la prémisse (ii) que l'eau est *une* des « choses » qui occupent le rôle-aqueux, rien ne s'ensuit de manière *a priori* concernant la structure de l'eau à partir de la structure de H_2O. Pour utiliser un exemple plus suggestif : du fait que (i′) le médicament *X* est *un* remède contre le paludisme, et que (ii′) le médicament *Y* est *un* remède contre le paludisme, il ne s'ensuit pas que *X*=*Y*. (« *Un* » remède veut dire ici « un remède parmi tant d'autres possibles »).

Par conséquent, on voit mal comment pourrait-on établir que l'eau est ***la*** (seule et unique) substance (chose) aqueuse. Après tout, on peut très bien imaginer que l'eau aurait pu être comme le jade, c'est-à-dire présentable sous une double apparence. Il est donc conceptuellement possible de concevoir un nombre indéfini de « choses » qui auraient pu occuper le rôle aqueux, chacune d'entre elles, d'ailleurs, étant parfaitement compatible avec leur existence en tant qu'eau. Si l'on remplace (c) par (c*) :

(c*) H_2O est ***une*** substance qui joue le rôle aqueux.

il devient clair que l'argument de la déductibilité *a priori* échoue. Autrement dit, on ne peut pas inférer *a priori* (d) à partir de (a) et (c*). Cela nous oblige, notent Block et Stalnaker, de spécifier avec beaucoup de précision les raisons pour lesquelles nous considérons que H_2O n'est pas ***une*** substance qui joue le rôle aqueux, mais ***la*** substance, la seule et unique. Mais pour établir cela, nous avons besoin de prendre des risques considérables, descendre sur le terrain, et relativiser ainsi la portée de (b). Or, l'ambition du principe du passage *a priori* vise l'*ensemble* des énoncés. En quoi cela est-il si important ici ? Il est important dans la mesure où la clause de l'unicité (« ***la*** substance ») disqualifie l'énoncé (c) dans sa candidature au statut d'énoncé physique ou microphysique. Car le fait que (c) affirme que H_2O est l'*unique* substance qui occupe le rôle-aqueux le destitue de ses prétentions d'être un énoncé microphysique empirique.

Après tout, arguent-ils, étant donné qu'il est possible de concevoir que de l'eau-fantôme, disons, couvre la partie de la Terre qui *n*'est *pas* couverte par l' « eau-physique », il ne peut pas s'ensuivre à partir de la microphysique que l'eau couvre 60%

de la Terre[38]. Par conséquent, la clause de l'unicité donne un coup de pouce illicite aux défenseurs de l'argument du passage *a priori*, car l'adopter c'est transgresser le domaine de la microphysique. Pour résumer : pour que l'argument du passage *a priori* puisse aboutir à prouver que (d) découle d'une base microphysique à l'aide d'une analyse conceptuelle, il est nécessaire que :

(i) L'énoncé (c) soit une vérité conceptuelle.

(ii) L'énoncé (b) soit une vérité microphysique.

Nous pouvons à la limite concéder à Jackson la vérité de (i), mais il est loin d'être clair dans quel sens l'appareillage conceptuel des « intensions-A » nous renseigne sur la vérité de (ii).

Le duel opposant Block et Stalnaker d'une part, et Jackson et Chalmers d'autre part, a fait l'objet de nombreuses discussions dans la littérature analytique contemporaine, et nous n'irons pas plus loin que ces quelques remarques ici. Une autre perspective critique sur le très controversé principe du passage *a priori* a été récemment avancée par S. Blackburn.

7.2.2 L'objection de Blackburn ou l'argument de l'inventeur du zip

Nous avons vu que l'argument de Jackson (1995/2003) repose de manière cruciale sur l'apriorité de (b). A l'appui de cette thèse, il fait appel à l'interprétation kripkéenne des descriptions fixatrices de référence. Le schéma général est le suivant : nous avons des propositions d'identité de genre « N= le F », où la description fixe la référence du nom « N ». Une question surgit alors : est-il vrai, comme le soutient Jackson[39], que « N= le F » est *a priori* si « F » est une description qui spécifie le fixateur de référence pour « N » ?

A cette question S. Blackburn[40] répond par la négative. A l'appui de ce rejet, Blackburn fait appel à une analogie tirée de la théorie des noms descriptifs de G.

[38] '[g]iven the possibility of ghost water that covers part of the earth not covered by physical water, it cannot follow from microphysics that water covers 60 percent of the earth' (*Ibid.*, p. 28)

[39] (F. Jackson, 2003 :87)

[40] S. Blackburn, (2009), 'Analysis, Description, and the *A priori*?' In Ian Ravenscroft (ed.), *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*, op. cit., p. 32.

Evans[41]. Supposons qu'aucun d'entre nous ne sait qui a inventé le zip, mais supposons qu'il est généralement admis qu'une seule personne en est l'inventeur. A présent, on introduit le nom « Jean-Michel » en l'écrivant au dos d'un jeton où figure la description suivante : « baptisons la personne unique qui a inventé le zip, Jean-Michel ! » La question que l'on se pose à présent est : quel est le processus qu'on peut qualifier d'« *a priori* » dans l'expérience de pensée qu'on vient de décrire ?

Selon Blackburn, on peut concéder à la limite qu'il est *a priori* le fait que l'énoncé *p* « Jean-Michel a inventé le zip » soit vrai. Mais à y regarder de plus près, on peut se demander à juste titre si le fait que Jean-Michel a inventé le zip est *a priori* connu ou même *a priori* connaissable. Il y a, nous semble-t-il, de fortes raisons de croire que non. Rien n'est *a priori* connaissable sur rien et personne, insiste Blackburn. Pour connaître l'apriorité d'un énoncé, il faut « baigner » au préalable dans le monde, connaître *ce que* les choses sont, *comment* les choses sont, bref, les manipuler empiriquement. On ne peut pas connaître de manière *a priori* que c'est Jean-Michel qui a inventé le zip. On peut, à la limite, découvrir empiriquement que tel c'est le cas, mais si on croise Jean-Michel, on ne peut pas l'interpeller en lui disant « Hé hé ! Je le savais depuis toujours que c'était toi qui a inventé le zip ! » Car c'est justement ce morceau-*là* d'information qui nous manquait *avant* qu'on investigue empiriquement pour apprendre *qui* a inventé le zip.

Il serait inutile d'ailleurs, souligne Blackburn, de chercher à défendre l'argument en injectant une dose supplémentaire de vérité *a priori* dans le corps de prémisses initiales. On pourrait répliquer, admet Blackburn[42], qu'étant donné qu'il est *a priori* que l'énoncé *p* soit vrai, et étant donné qu'il est *a priori* que *p* informe que Jean-Michel a inventé le zip, il s'ensuit que le fait que Jean-Michel a inventé le zip *doit* être *a priori*. Pour illustrer l'erreur de cette manœuvre, considérons l'analogie suivante avec l'interprétation standard des indexicaux. Un sujet énonçant « Je suis ici maintenant » sait que la nature de l'énoncé est telle qu'il est destiné à exprimer une *vérité*. Mais il ne sait pas *quelle* vérité (quelle proposition) il formule, à moins qu'il sache vraiment *où* il se trouve. Le contenu de l'énoncé est donné uniquement dans un contexte donné.

[41] G. Evans, (1973), 'The Causal Theory of Names,' *Proceedings of the Aristotelian Society*, Supplementary Volume 47: 187–208.
[42] (S. Blackburn, 2009 : 32)

Comment peut-on relier à présent l'exemple de l'invention du zip avec l'affaire des déductions conceptuelles *a priori* ? D'assez nombreux auteurs adhèrent à la distinction entre la connaissance « *de dicto* » et connaissance « *de re* ». Or justement, Jackson[43] fait appel à cette distinction pour illustrer que la connaissance *de re* est déductible *a priori* à partir de la connaissance *de dicto*. Regardons-y de plus près.

Supposons, dit Blackburn, qu'on explique à Michel comment le zip fonctionne, à la suite de quoi Michel, ravi, s'exclame :

ZIP

(a) Le bonhomme qui a inventé ce truc doit être un génie !

Si Jackson a raison, il s'ensuit que Michel peut déduire *a priori* que :

(c) Jean-Michel est un génie.

Puisqu'il a à sa disposition l'énoncé présumé *a priori* (b) :

(b) Jean-Michel est le bonhomme qui a inventé le zip.

Rappelons-nous que le principe qui est derrière cette inférence est le même que celui qui est derrière l'argument du passage *a priori* de Jackson, à savoir :

(H) « Si A&B *a priori* impliquent C, et B est *a priori*, alors A implique *a priori* C ».

L'étrangeté de cette situation nous saute aux yeux dans la mesure où (b) est censé d'être ***à la fois*** *a priori*, ***et*** convertir notre position épistémique à l'égard de l'énoncé (a) - qui représente le savoir dont on dispose à propos du créateur du zip – en une position épistémique qui nous permet d'apprendre (empiriquement) *qui* en est l'inventeur. Comme nous l'avons déjà brièvement évoqué, nous pensons que cette formidable acrobatie conceptuelle à laquelle se livre Jackson ne fait qu'accentuer la profondeur du gouffre qui sépare le « savoir *dans* le monde » du « savoir *sur* le monde ».

Dans ce qui suit, nous argumenterons que ce clivage n'est qu'un dommage collatéral provoqué par l'adoption d'un schéma conceptuel dichotomisant, opposant dans un face à face aussi tragique que factice, la pensée et le monde. Car pour pouvoir insérer un tel clivage il faut avoir déjà endossé ce que nous appellerons, en paraphrasant

[43] (F. Jackson, 2003)

B. Stroud[44], *le point de vue du Super-Sujet*, *sur*plombant le réel et énonçant des savoirs *sur* le monde. Mais avant d'examiner en détail cette hypothèse, il nous reste encore une étape à franchir dans notre investigation. Plus spécifiquement, il nous reste à montrer que le physicalisme *a priori* de Jackson que nous venons d'évoquer prête le flanc au même type d'objections que le physicalisme sémantique d'un Carnap.

7.3 Analyse conceptuelle et sauts conceptuels

Dans la première partie de cet essai nous avons repéré une brèche au sein du physicalisme de « première génération », difficulté située au niveau du « saut conceptuel » opéré par Carnap entre le mode contentuel et le mode formel de discours. Dans cette section, nous allons voir qu'une faille de même genre finit par vicier l'édifice du physicalisme *a priori* du dernier Jackson : un saut conceptuel acrobatique y est savamment intercalé *via* le principe du passage *a priori* afin de transformer les « infortunées » vérités du sens commun en vérités scientifiques « fortunées ». A l'origine de ce double malentendu, comme le note P. Wagner, il y a l'idée, « désuète à certains égards, selon laquelle la logique énonce des vérités. Plus précisément, les énoncés logiques sont des lois générales, à vrai dire les lois les plus générales puisqu'elles portent sur tout ce qui est, ou sur tout ce qui peut être pensé»[45]. Par conséquent, un physicalisme de veine sémantique comme celui d'un Carnap, ou le physicalisme *a priori* du « deuxième » Jackson, cible non seulement le processus d'unification du savoir sous l'universalité du langage de la physique et de ses vérités « fortunées », mais surtout l'implémentation d'une ontologie d'arrière-plan unificatrice, dénouant ainsi d'un seul coup, par ricochet, le nœud gordien des problèmes situés à l'intersection entre vérités analytiques et vérités factuelles. Délimitons à présent avec plus de précision le présupposé que nous venons d'énoncer.

7.3.1 Le bon plaisir de Dieu et l'*a priori*

Nous avons vu que les avocats du physicalisme *a priori*, plutôt que d'envisager le mental et le physique comme deux différents types de propriétés, conçoivent le

[44] (B. Stroud, 1996)
[45] P. Wagner, (2002), *Le contexte logique de l'Aufbau*, in S. Laugier (ed.), op. cit., p. 23.

problème comme celui du rapport entre propositions vraies (ou « vérités ») exprimées dans un vocabulaire mental et propositions vraies exprimées dans un vocabulaire physique. Selon l'hypothèse fondamentale de cette approche, le lien entre les vérités physiques et celles psychologiques est plus étroit que la corrélation nécessaire qui caractérise la survenance forte : les propositions mentales ne sont que des retranscriptions des états-de-choses physiques dans un autre vocabulaire. Mettons que *P* soit l'ensemble de tous les états de choses physiques correspondant à l'ensemble de toutes les propositions vraies exprimées avec les concepts de la physique. Les vérités psychologiques résultent d'un redécoupage de *P*, à l'aide d'un ensemble alternatif de concepts, à savoir ceux de la psychologie. En clair, le lien entre le physique et le psychologique n'est plus conçu comme un lien naturel, mais comme un lien *conceptuel*. Cela permet d'éviter de postuler, comme le fait la thèse de la survenance forte, par exemple, une nécessité naturelle qui reste mystérieuse. Le physique détermine l'ensemble des états-de-choses non physiques, et en particulier le mental, non pas par *nécessité naturelle* mais par *nécessité conceptuelle*. Cela implique qu'il est en principe possible de parvenir à la connaissance de n'importe quel état-de-choses non physique, plus particulièrement d'un état mental, à partir d'une connaissance complète des états-de-choses physiques, sans autre investigation empirique, autrement dit, de manière purement *a priori*.

S. Blackburn[46] suggère qu'une bonne méthode de contester ce type d'approche, serait d'évoquer le vieux débat autour du « mind body problem » chez les philosophes classiques comme Locke et Leibniz. Lorsque Locke aborde la délicate question de la médiation, de l'articulation entre les 'actions' du bras et du cerveau d'une part, et les 'sensations' de chaleur ou de douleur d'autre part, il finit par hausser les épaules et décréter que c'était le bon plaisir de Dieu d'annexer tel mouvement à telle sensation. Leibniz protesta : ce n'était pas digne d'une œuvre divine de procéder d'une façon aussi rustre. Il doit y avoir un autre moyen de penser cette jointure, un moyen moins aléatoire et plus rationnel et ordonné, comme par exemple la relation entre un cercle et sa projection dans un plan. L'idée fondamentale derrière cette analogie de Leibniz est que la jointure entre le mental et le physique doit être pensée comme une relation non seulement systématique et ordonnée, mais surtout *nécessaire*. Qui dit nécessaire, dit

[46] (S. Blackburn, 2009 :23)

rationnel, et le tour est joué. Si la relation est nécessaire et rationnellement intelligible, alors elle est *a priori*.

Étrangement, observe Blackburn, de nos jours, les philosophes ont à leur disposition des appareillages conceptuels hautement sophistiqués incluant « le contingent *a priori* », « les nécessités *a posteriori* », « les mondes possibles », « la survenance », « les énoncés de Ramsey », « les propriétés *de se* », « la logique modale bidimensionnelle », etc. Les penseurs analytiques sont plus qu'enthousiasmés à l'idée d'opérer avec ces nouveaux outils conceptuels. Et pourtant, note Blackburn (et nous le suivons entièrement sur ce point), lorsque la machinerie d'une argumentation bute sur une difficulté de médiation, il semblerait que les options qui restent ne soient pas fondamentalement différentes de celles signalées par les prédécesseurs Locke et Leibniz. Autrement dit, soit on fait appel au « bon plaisir de Dieu » pour relier les deux items recalcitrants, mais dans ce cas on tombe dans le « mystérianisme » ; soit on défend les bienfaits de l'analyticité et de la rationalité *a priori*, mais dans ce cas, on se livre à de véritables acrobaties conceptuelles afin de maintenir la cohérence de l'ensemble explicatif. Carnap et Jackson adhèrent manifestement à ce dernier genre de solution. Lorsqu'un fait mystérieux menace de transcender le niveau de la physique, un bon remède de procéder à sa démystification serait de le « localiser » ou de le « situer » dans l'ordre naturel du monde. Selon Jackson[47], par exemple, cette « localisation » passe par le truchement de l'analyse conceptuelle qui nous permet de montrer comment les vérités «énigmatiques » découlent *a priori* à partir de faits physiques « lumineux ». Ce qui veut dire, aux yeux des défenseurs de l'analyse conceptuelle, que la déductibilité *a priori* respecte le principe du naturalisme, autrement dit, elle ne tombe pas sous le couperet de l'accusation de « surnaturalisme »[48] qui hante la philosophie post-positiviste.

7.3.2 Le tandem fossés explicatifs/sauts conceptuels

Reprenons à présent avec un sens critique mis en éveil l'analogie que nous aimerions établir entre les failles du physicalisme sémantique et celles du physicalisme *a priori*. Rappelons-nous que pour Carnap le mode formel d'un discours [1] « *p* est

[47] (F. Jackson, 1998)
[48] (B. Stroud, 1996)

déductible de *P* et inversement » sera toujours préférable au mode contentuel [2] « *P* et *p* **décrivent** le même état de choses »[49]. Dans le premier cas [1], toute référence à un fait extralinguistique (objet physique) ainsi qu'à l'agent locuteur émettant cette assertion est tacitement assimilée dans l'opération logique de déductibilité. « Logique », c'est-à-dire « objectif ».

Dans le deuxième cas [2], la présence des termes descriptifs rend problématique non seulement l'articulation entre la proposition et les données factuelles, mais aussi l'articulation entre le sujet narrateur décrivant un état de choses, et la proposition énoncée. Pour dire les choses d'une manière imagée, disons qu'il est plus facile de concevoir un monde de la logique pure (vérités abstraites pures) que de concevoir un monde de la « narration » pure, ou de la « description » pure. Une fois le saut conceptuel opéré depuis le mode contentuel du discours vers le mode formel du discours, la tâche du philosophe – désormais débarrassé du problème encombrant des entités extralinguistiques – sera celle de garantir la validité intrinsèque du système formel. Ce passage est en quelque sorte le « *hocus pocus* » sur lequel repose le physicalisme sémantique de Carnap.

Tournons-nous à présent vers le physicalisme *a priori* et son propre « *hocus pocus* ». Rappelons-nous que le principe du passage *a priori* consiste à dire qu'il serait possible de déduire de façon *a priori* les vérités « infortunées » du sens commun à partir des vérités « fortunées » des sciences physiques. Nous n'allons pas reprendre ici l'argumentation de Jackson (1998/2003) qui a été déjà examinée en détail dans la section 7.1. Limitons-nous, pour simplifier, à l'épure de son argument afin de mieux y déceler le saut conceptuel qui vicie le principe du passage *a priori*. L'objectif est de déduire la vérité « infortunée » (d) :

(d) 60% de la surface de la Terre est couverte d'eau.

à partir de la vérité fortunée (a) :

(a) 60% de la surface de la Terre est couverte d' H_2O .

Jackson est conscient que le passage de (a) à (d) n'est pas *a priori*. On met alors en scène la notion-halo de « substance que occupe le rôle aqueux ». A l'instar du mode formel de discours de Carnap (« *p* est **déductible** de *P* et inversement »), c'est à la prémisse (b) de Jackson - « l'eau est la substance qui occupe le rôle 'aqueux' » -

[49] (R. Carnap, 1932b/2006: 354)

qu'échoit la mission de combler le gouffre épistémique qui sépare l'analyticité de l'empiricité. L'étrangeté de cette situation devient visible dès qu'on s'aperçoit que l'énoncé (b) est censé d'être ***à la fois*** *a priori*, ***et*** métamorphoser notre position épistémique à l'égard de l'énoncé (a) - qui représente le savoir dont on dispose à propos de la structure moléculaire de l'eau – en une position épistémique qui nous permet d'apprendre (empiriquement) *quelle* est effectivement cette structure (H_2O , XYZ, etc).

Nous pensons néanmoins que cette formidable acrobatie spéculative à laquelle se livrent Carnap et Jackson dans les exemples que nous venons d'évoquer, ne fait qu'accentuer la profondeur d'un phénomène conceptuel que nous pourrions baptiser le « *Big-bang Gouffre* » séparant le domaine de la subjectivité de celui de l'objectivité. Dans ce qui suit, nous défendons l'hypothèse selon laquelle, de manière générale, toute approche privilégiant un modèle d'intelligibilité opérant par découpes, dichotomies et incisons conceptuelles dans la structure du réel, visant à en extraire l'essence « objective » des phénomènes, provoque un processus à double effet : (i) dans un premier temps, les découpages destinés initialement à simplifier la tâche des méthodologies de recherche, causent l'apparition des phénomènes de fossés explicatifs. (ii) Dans un deuxième temps, on s'aperçoit que ces fossés explicatifs doivent être comblés d'une façon ou d'un autre, sinon la cohésion du système est menacée. Bref, on est confronté à la question de l' « articulation », à la nécessité de « recoller » les morceaux, et de rendre intelligible le fonctionnement d'ensemble. Des « sauts conceptuels » sont alors insérés à cet effet. Les sauts conceptuels sont en quelque sorte « l'envers nocturne »[50] des fossés explicatifs.

[50] (M. Foucault cité par P. Engel, 1996 : 147)

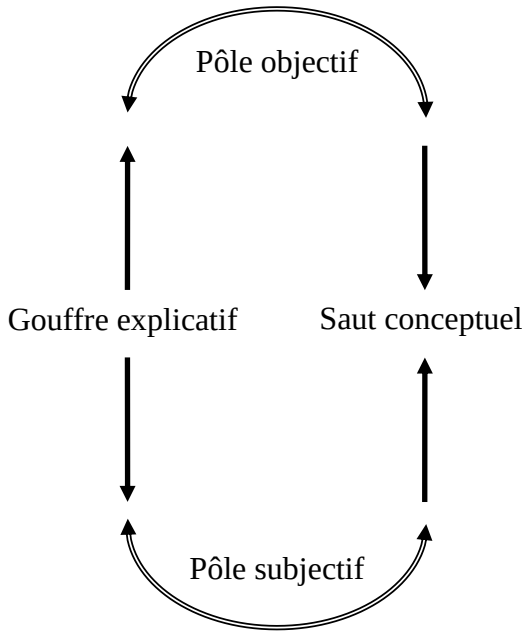


Figure 12: Tandem fossés explicatifs / sauts conceptuels

Conséquemment, chaque fossé explicatif est accompagné d'un "saut conceptuel", c'est-à-dire d'un médiateur conceptuel hybride, introduit *in extremis* et *ad hoc* afin d'assurer l'homogénéité d'une argumentation rendue fragmentaire suite aux clivages argumentatifs employés précédemment. Il s'agit, pour le dire rapidement, d'une et seule « monnaie » philosophique, qui peut être explorée sur ses deux facettes complémentaires. Vue de cette perspective, une bonne partie de l'histoire de la pensée philosophique moderne peut être vue comme une tentative de vouloir reconstruire le jour ce qu'elle-même a démoli la veille.

Essayons d'encadrer avec plus de précision l'hypothèse du tandem fossés explicatifs/sauts conceptuels.

Pour chaque clivage conceptuel opéré sur l'axe subjectif/objectif un double processus s'ensuit :

[TANDEM F/S]

(T_F) Dans un premier temps, et comme effet collatéral des stratégies objectivantes, des **fossés explicatifs** apparaissent. Pour palier à ce type de difficulté d'articulation et de médiation, autrement dit, pour répondre à la question « comment franchir les fossés afin d'assurer l'homogénéité épistémique de l'ensemble ? » on fait appel aux :

(T_S) **Sauts conceptuels**, insérés subrepticement afin d'assurer la cohésion du corps explicatif.

Le talon d'Achille de cette stratégie réside dans le fait qu'elle repose tacitement sur l'hypothèse hautement suspecte du Super-Sujet. En effet, par quel miracle le sujet connaissant serait-il capable de s'extraire, à la façon d'un baron de Münchhausen, du marais de la subjectivité, pour se déposer sain et sauf sur le terrain stable de l'objectivité ? Comment pourrait-on établir avec une précision chirurgicale ce clivage autrement qu'en stipulant tacitement qu'il serait possible de survoler le bourbier, et de distribuer, d'en haut, des billes blanches ou noires aux vérités « fortunées » et respectivement « infortunées » élaborée sur l'axe objectivité/subjectivité?

Nous avons vu, dans le chapitre 2 à quel point il était important pour Frege de doter la logique du statut solide de « pensée objective » et d'éliminer l'idée que les règles de la pensée sont fondées sur les régularités de la dynamique mentale. Or, c'est ici un point important, la possibilité même de tracer un tel clivage, présuppose l'adhésion tacite à l'hypothèse du point de vue en surplomb.

Car l'objectivité[51], en tant qu'idéal régulateur de la connaissance est ontologiquement rattachée aux croyances et aux états mentaux d'un sujet connaissant.

Un corolaire intéressant de la stratégie « tandem F/S » c'est qu'elle oblige leurs tenants à opérer des sauts conceptuels d'autant plus acrobatiques que les dichotomies

[51] Dans un article publié en 2008, M. Bitbol et I. Peschard illustrent de manière magistrale le processus d'objectification des expériences sensorielles à travers l'exemple paradigmatique des concepts de « chaleur » et de « température » dans la thermodynamique. Ils montrent ainsi de manière très convaincante, étape par étapee comment le caractère « phénoménal » de l'expérience a été relégué au statut ingrat d'énoncés subjectifs. Voir I. Peschard & M. Bitbol (2008), 'Heat, Temperature and Phenomenal Concepts', in E. Wright (ed.), *The Case for Qualia*, Cambridge Mass.: MIT Press.

introduites sont nettes et rigides. Par exemple : plus on insiste sur la nature analytique de certains concepts, moins leur attachement au palier empirique est intelligible. Et plus on insiste sur l'apriorité des vérités mathématiques pures, moins leur applicabilité au domaine du factuel est intelligible. Et pour revenir à notre sujet d'étude ici : plus on insiste sur la dichotomie états physiques / états phénoménaux, moins il est évident comment le « miracle » de la jointure « pinéale » mental/physique se réalise. Aucun de ces problèmes philosophiques n'aurait pu surgir sans l'adoption tacite de cet héritage de filiation pythagoricienne-platonicienne qui nous informe qu'il existerait un domaine à part, abritant la totalité des vérités « fortunées », objectives, qui transcendent la finitude épistémique et ontologique du sujet connaissant, et qui doivent être logées tant bien que mal dans un "monde" taillé à leur mesure.

Afin de calculer l'étendue des répercussions qu'entraîne l'adoption tacite de ce véritable schisme archétypal, il importe de revenir brièvement en arrière, et de photographier les cicatrices qu'il a laissées sur la physionomie de certaines controverses philosophiques majeures. Par conséquent, avant de clore cette section dédiée au physicalisme *a priori*, et avant de reprendre le fil des discussions autour de l'argument de la connaissance, il convient de refaire un bref détour historique qui nous aidera à saisir les étroits liens de parenté qui relient le matérialisme actuel à son rival déclaré, le dualisme cartésien. Dans la section suivante, nous verrons que les stratégies dualistes sont loin d'être les seules à intercaler des « jointures mystérieuses » reliant *le domaine de la subjectivité* à celui de *l'objectivité* à l'intérieur de ses trames explicatives. Bien plus, nous soutiendrons que des « raccordements » non moins mystérieux viennent « sauver » *in extremis* la cohérence et l'architecture de plusieurs modèles explicatifs, qu'ils soient rationalistes, empiristes logiques, ou physicalistes. Toutes ces stratégies ont en commun leur sujétion tacite au schisme mythologique pensée/monde décliné sur l'axe subjectif/objectif.

7.3.3 Faites entrer l'accusé : pourquoi Descartes n'est pas le seul 'coupable'

Il suffit de consulter la table de matières de n'importe 'Companion to the Mind' pour s'apercevoir que la critique du dualisme cartésien ouvre le bal matérialiste, pour

ainsi dire, et sert de fort commode tremplin pour amorcer l'entrée triomphale des stratégies antidualistes. Parfois présenté de manière caricaturale comme étant le « bouc émissaire » du « mind-body problem », Descartes est accusé d'emblée d'avoir compliqué inutilement les choses, et d'avoir inventé une substance pensante *en plus* de la matière, cette manœuvre l'obligeant ainsi à trouver une « place » à cette mystérieuse substance encombrante - « *res cogitans* » - dans le monde naturel. A partir de là, il est très facile de montrer que le dualisme déclenche la prolifération des « mondes supplémentaires » non-naturels, provoque le dilemme de la communication des substances et tombe sous le couperet de l'objection antisurnaturaliste.

Dans ce qui suit, nous suggérons d'abandonner provisoirement ce type d'approche et de situer le dualisme des substances dans le contexte d'ensemble de la philosophie « mécaniste » de Descartes. On s'apercevra alors qu'une et même ambition souterraine relie l'auteur des *Méditations Métaphysiques* aux empiristes logiques de Vienne, et à leurs héritiers directs, les physicalistes *a priori* australiens d'aujourd'hui, à savoir l'ambition de garantir aux vérités objectives une dignité ontologique d'ordre supérieur. En ouvrant cette parenthèse sur le cadre d'arrière-plan qui a vu naître la philosophie cartésienne, nous commettrons certainement des simplifications abusives mais inévitables, en raison de la grande complexité du sujet.

Tout ceci a pour but de mettre en relief le constat suivant : héritier d'une tradition qui délimite rigidement l'ordre immuable de l'objectivité de l'ordre immanent de la subjectivité, Descartes ne doit pas « comparaître » seul devant le tribunal qui condamne actuellement son dualisme des substances avec tant de fermeté. Le physicalisme contemporain pourrait se révéler, lui-aussi, coupable d'adhérer à ce que nous avons appelé le « Big Bang Gouffre » séparant l'ordre immuable de l'objectivité de l'ordre immanent de la subjectivité, bien que sous une forme voilée, et remise au goût du jour dans l'affaire de la déductibilité *a priori* des vérités phénoménales à partir des vérités physiques.

Examinons de plus près cette hypothèse. Depuis que l'aventure philosophique prit son essor dans l'Antiquité Grecque, il est difficile, voire utopique, de fixer une date, un auteur ou un ouvrage à l'origine des bipartitions que l'on subsume habituellement sous le vague terme de « dualisme subjectivité/objectivité ». Conséquemment, ce qui suit est un bref rappel, très schématique, des principales lignes réflexives qui alimentent plus ou mois ouvertement, ce clivage avant et après Descartes. De manière générale, le

thème des rapports opposant l'être pensant au monde réel connaît une radiation souterraine qui remonte jusqu'à la pensée grecque, et plus clairement à l'œuvre de Platon[52]. Lorsqu'il range les Idées dans un ordre ontologique supérieur, distinct des objets matériels quotidiens, Platon réalise l'exploit de figer l'immuabilité des entités abstraites afin d'obtenir un angle de vue privilégié sur les vérités objectives. Cette ontologie à deux vitesses laissera des empreintes indélébiles sur la physionomie des controverses philosophiques.

Au XVIIe siècle[53], le programme philosophique cartésien séduit d'abord par son élégante simplicité : au lieu de tourner vainement et indéfiniment autour de la question de savoir qui détient l'autorité de dire la vérité, fouillant dans les textes parfois obscures et contradictoires des auteurs scolastiques, Descartes affirme, dans un style limpide, accessible et dépourvu de toute rhétorique stérile, que l'être humain, étant naturellement et universellement doué de bon sens, doit pouvoir soumettre ses idées et opinions au doute le plus profond et radical, afin d'en tirer au moins une certitude inébranlable : *cogito ergo sum*. Descartes balaye ainsi, d'un seul geste, toutes les questions liées à une fondation *extérieure* de la vérité. Le doute cartésien réalise le "grand ménage" - si l'on peut dire ainsi - dans la cour des interrogations philosophiques, en leur insufflant un nouvel élan.

Mais, et c'est ici le point crucial sur lequel nous aimerions attirer l'attention du lecteur, ce serait de commettre une fatale erreur que de croire que la principale préoccupation de Descartes était de trouver une place pour la « substance rationnelle » dans le monde naturel. Son véritable dessein était ailleurs. Comme le résume fort bien Fr. Yates, "dans sa hâte d'établir une vision purement objective de la nature comme mécanisme, et dans son enthousiasme pour les mathématiques pures en tant qu'unique outil sûr pour la recherche objective, *Descartes se retrouva avec le problème plutôt embarrassant de l'esprit sur les bras.* Il le résolut provisoirement, de façon grossière

[52] "The safest general characterization of the European philosophical tradition is that it consists of a series of footnotes to Plato". A. N. Whitehead, (1929), *Process and Reality*: Cambridge University Press, p. 53.

[53] Sans entrer dans les détails d'une discussion qui nous éloignerait de notre but ici, rappelons seulement que le siècle de Descartes assiste à une métamorphose idéologique importante, car, comme le note Fr. Yates, "les Grecs, nantis d'une intelligence mathématique et scientifique de premier ordre (...) jamais, à la différence de l'homme occidental au début des temps modernes, ils n'ont franchi de bon cœur, et avec tout leur potentiel, le pont entre la théorie et la pratique pour concrétiser leur savoir. Les Grecs refusaient la concrétisation. Ils la jugeaient vile et mécanique ». Fr. Yates, (1964). *Giordano Bruno and the Hermetic Tradition*, University of Chicago Press. Trad. fr. (1988), *Giordano Bruno et la tradition hermétique*, Paris, Éd. Dervy, p. 190. Nous faisons référence ici à la tradition française.

par son soi-disant dualisme, un monde composé d'une énorme machinerie mathématique s'étendant dans l'espace; et un autre monde composé d'esprits pensants non étendus"[54]. Bref l'objectif principal de Descartes *aussi* était de défendre une vision de la nature purement « objective », « mécanique », et il était enthousiasmé par l'utilisation de l'appareillage conceptuel des mathématiques. Et c'est précisément son adhésion à l'idée des mathématiques pures en tant qu'outil privilégié d'investigation du réel qui le mène vers son dualisme des substances et non pas à l'inverse. Il est important de souligner cette priorité dans l'ordre de l'argumentation cartésienne. Et nous faisons le pari que, transporté à l'époque actuelle, Descartes se serait clairement rangé du côté des physicalistes.

Certes, l'auteur des *Méditations métaphysiques* savait que l'ancrage des certitudes dans l'introspection immédiate du *cogito*, soulèvera au moins une difficulté[55] redoutable, à savoir celle d'établir comment une entité immatérielle arrive à communiquer ses prescriptions à une entité matérielle. Autrement dit, quel est le « ciment » miraculeux qui lie ces deux ordres ontologiques distincts ? Mais cet aspect, à la limite, n'était qu'un dommage collatéral d'un projet plus vaste, et plus ambitieux, qui visait, quant à lui, de fournir une description objective du réel, objective c'est-à-dire « naturelle ». L'idée que nous aimerions souligner ici c'est que le dualisme cartésien n'est qu'un effet « naturel » du clivage pensée/ monde dont il hérite sans ciller, et qu'il défend à corps perdu, même s'il doit y sacrifier l'hypothèse inélégante de la glande pinéale.

Pour mesurer l'impact de cet héritage, il suffit de se rappeler que lorsqu'un empiriste comme J. Locke, par exemple, compare l'esprit à une feuille blanche sur laquelle l'expérience grave les impressions sensibles, il suppose, implicitement, que ce même esprit est non seulement vide, mais aussi *séparé*, de par sa "nudité" initiale, du monde matériel, dans lequel les sens puisent, comme dans une source inépuisable, la diversité de ses données. Certes, l'empirisme de Locke met l'accent avant tout sur la primauté de l'Expérience Sensible en tant que source effective de connaissance, et non pas sur la Raison. Et même s'il admet que toute connaissance doit s'appuyer sur les données des sens, il reste tributaire à cet héritage bipartite : rationalité (immatérielle) d'une part et empiricité (matérielle) d'autre part. L'épistémologie de Locke, comme le

[54] (F. Yates, 1964/ 1988 :527) (C'est nous qui soulignons.)
[55] Cette question est habituellement connue sous le nom de «question d'Élisabeth de Bohème ».

note H. Gardner, se constitua « à l'extrême opposé de l'univers de Descartes, sur l'expérience des objets du monde extérieur. Cependant, Locke put finalement construire un organisme capable d'abstraction et de généralisation, un être qui parvient à un savoir d'une certaine sorte par l'interaction de ces idées (plutôt que par une simple comparaison avec l'expérience) et qui est en fin du compte un moi rationnel et conscient. Il n'est pas extravagant de dire que *l'individu lockéen présente une grande ressemblance avec l'individu cartésien* »[56]. Bien sûr qu'ils se ressemblent, nous avons envie de dire, puisqu'ils appartiennent à la même famille dichotomique : rationalité (immatérielle) et empiricité (matérielle).

C'est donc la recherche de ce « ciment » liant l'immatérialité de la rationalité à la matérialité du monde empirique (cerveau humain y compris) qui hantera la philosophie à la fois du « rationaliste » Descartes et de l'« empiriste » Locke. C'est là qu'il faut chercher le socle commun de leurs préoccupations, à savoir dans l'idée tacite d'une séparation quasi-ontologique, nette, entre l'immatérialité de la pensée et la matérialité du monde. Examinons la manière dont le clivage raison/sens a été « recyclé » *via* la distinction *a priori*/*a posteriori* chez Kant et plus tard chez Carnap. Dans son ambition fondamentale de fournir un système explicatif universel, englobant à la fois l'expérience sensible et l'ensemble des lois logiques rationnelles, la philosophie kantienne tranche de façon radicale avec ses prédécesseurs, et assume pleinement le clivage explicite entre la chose-en-soi (l'objet, la matérialité) et l'entendement (le sujet, l'immatérialité). Ce faisant, il ravive l'ancien problème cartésien de la « communication » des substances, sauf que cette fois-ci sous une forme déguisée, épistémique : car il ne s'agit plus de faire communiquer la chose pensante avec la chose étendue, mais d'attacher les formes pures de l'entendement aux formes sensibles des objets. La solution que le philosophe de Königsberg propose est, comme il l'avoue lui-même, "extravagante et absurde"[57] au premier abord : "L'entendement – affirme Kant - est lui-même une législation pour la nature. (...) les phénomènes comme tels ne peuvent pas avoir lieu en dehors de nous, mais ils n'existent comme tels que dans notre sensibilité"[58].

[56] (H. Gardner, 1985/1993: 72) (Nous soulignons)
[57] E. Kant, (1787), *Kritik der reinen Vernunft*, trad. fr. A. Tremasaygues et B. Pacaud, 3° édition, *Critique de la raison pure* - P.U.F., Paris, 1990, p. 143. (Nous faisons référence ici à cette dernière édition).
[58] *Ibid.*, p. 142.

Du fait que la philosophie "ne concerne pas les objets de l'expérience mais part du fait *a priori* qu'une connaissance systématique de ces objets est possible"[59], on voit clairement que dans l'équation pensée/monde, l'esprit acquiert chez Kant un rôle fondateur, car les conditions de possibilité de l'*expérience* d'un objet sont aussi les conditions de possibilité de son *existence*. D'où sa célèbre "révolution copernicienne": « nous nous précédons nous-mêmes dans les objets de notre connaissance et les lois universelles que le scientifique déchiffre dans la nature, il en est lui-même la source »[60].

Il va de soi que la législation ordonnatrice de notre intellect n'est pas isolée du monde extérieur. Les deux termes extrêmes - insiste Kant - c'est-à-dire la sensibilité et l'entendement, doivent nécessairement s'accorder grâce à cette fonction transcendantale de l'imagination, puisque autrement tous deux donneraient sans doute des phénomènes, mais ne donneraient pas d'objets d'une connaissance empirique, ni, par suite, d'expérience.: "(...) tous ces phénomènes - par conséquent *tous les objets* dont nous pouvons nous occuper - *sont tous en moi*, c'est-à-dire qu'ils sont des déterminations de mon moi identique. (...) La manière dont le divers de la représentation sensible (l'intuition) appartient à une conscience, précède donc toute connaissance de l'objet, comme étant la forme intellectuelle, et constitue même une connaissance formelle *a priori* de tous les objets en général, en tant qu'ils sont pensés (les catégories). La synthèse de ces objets par l'imagination pure, l'unité de toutes les représentations par rapport à l'aperception originaire, *précèdent toute connaissance empirique*"[61].

Nous n'avons pas l'intention, faute d'espace, de reprendre ici en détail toutes les subtilités de la terminologie kantienne. Nous nous contenterons de délimiter un périmètre de recherche assez étroit, en direct rapport avec le fil conducteur qui guide notre recherche. Dans le chapitre 2, nous avons vu que la mythologie frégéenne des trois mondes, tout comme le physicalisme sémantique carnapien, nous renseignent très peu sur l'articulation ou la médiation entre ce le philosophe de Königsberg appelle l'"accord entre deux extrêmes"[62], c'est-à-dire la phase intermédiaire qui relie la sensibilité et la raison, l'empirique et la logique. Lorsque Kant aborde à son tour la question du « raccordement » entre les catégories prédéfinies de l'entendement et l'application aux intuitions sensibles afin d'en extraire la connaissance d'un objet

[59] I. Prigogine & I. Stengers, (1979), *La nouvelle alliance*, Paris, Gallimard, p. 98.
[60] *Ibid.*, p. 99.
[61] (E. Kant, 1787/1990 : 146) (Nous soulignons)
[62] *Ibid.*, p. 139.

empirique, il fait appel à ce qu'on peut qualifier, à la suite d'U. Eco, de "concept explosif"[63], à savoir le concept de "schème transcendantal". Or c'est justement le problème de la coexistence hétérogène, au sein des structures cognitives de l'entendement humain, des concepts purs d'un côté, et des intuitions sensibles d'un autre, qui pose de redoutables difficultés. Chez Kant, le sujet connaissant devient, s'il nous est permis d'utiliser cette expression, une « grande pinéale grandeur nature », tiraillé entre les pôles de la subjectivité et de l'objectivité.

Comme le souligne Husserl, "Kant tombe d'emblée dans le sillage de la théorie métaphysique de la connaissance, étant donné qu'il vise à un 'sauvetage' critique des mathématiques, de la science de la nature et de la métaphysique, avant même d'avoir soumis la connaissance comme telle, la sphère totale des actes dans lesquels se réalisent l'objectivation pré-logique et la pensée logique, à l'élucidation d'une analyse d'essence et d'une critique, et avant d'avoir ramené les concepts et leurs lois logiques primitives à leur origine phénoménologique"[64]. Se concentrant plutôt sur le "sauvetage critique" des concepts mathématiques comme « triangle », plutôt que sur l'analyse des modestes concepts empiriques comme « chien », Kant se révèle ainsi, un digne héritier de Platon, et fait preuve d'un certain embarras lorsqu'il doit aborder la question plus "banale" des concepts quotidiens. Si bien que la question du lien qui assure la cohérence dans l'obtention d'une notion comme celle de « chien », par exemple, révèle une certaine perplexité chez l'auteur des *Critiques*. Kant finit d'ailleurs par admettre, un peu à contrecœur, que "ce schématisme de notre entendement, qui concerne également la simple forme des phénomènes est *un art caché dans les profondeurs de l'âme humaine* et dont il sera toujours difficile d'arracher le vrai mécanisme à la nature, pour l'exposer à découvert devant les yeux"[65].Cette formulation est tout aussi vague qu'embarrassée,

[63]"En introduisant le schématisme dans la première version du système, Kant, comme l'avait suggéré Peirce, se retrouve avec un concept explosif entre les mains, un concept qui l'oblige à aller plus loin, vers la troisième *Critique*, précisément. Mais une fois passé du schème au jugement réfléchissant, la nature des jugements déterminants entre à son tour en crise. Car la faculté de jugement déterminante (comme le reconnaît Kant dans le chapitre de la *Critique du jugement* sur la dialectique du jugement téléologique) « ne possède pas de principes qui fondent des concepts d'objets »; elle se limite à subsumer des objets sous des lois ou des concepts données en tant que principes." U. Eco, (1997), *Kant e l'ornitorinco*, Milan : Bompiani. Trad. fr. J. Gayrard (1999*). Kant et l'ornithorynque*, Paris : Grasset, pp. 125-126. (Nous faisons référence ici à l'édition française.)

[64] E. Husserl, (1900-1901), *Logische Untersuchungen* : *Prolegomena zur reinen Logik*, Niemeyer. Trad. fr. H. Élie, A. L. Kelkel & R. Schérer (1974) *Recherches logiques*, t. 3, Paris : P.U.F., p. 242. Nous faisons ici référence à cette édition.

[65]« (...) Le concept de chien signifie une règle d'après laquelle mon imagination peut exprimer en général la figure d'un quadrupède, sans être astreinte à quelque chose de particulier que m'offre l'expérience, ou mieux à quelque image possible que je puisse représenter *in concreto*. Ce schématisme de notre

et représente, à nos yeux, un autre excellent exemple de « saut conceptuel » que le philosophe de Königsberg est contrait d'opérer suite aux clivages conceptuels qu'il s'inflige *ab initio*. Car s'il est relativement facile de montrer que la législation catégorielle du schème conceptuel de "triangle", par exemple, est intelligible en vertu de sa forme pure, il est évident que cette même manœuvre ne va pas de soi dans le cas des schèmes conceptuels applicables aux objets empiriques. Cela revient à dire que les concepts empiriques ne peuvent pas être définis isolément et universellement, cette propriété restant un privilège exclusif des concepts mathématiques.

Afin de mieux illustrer la difficulté à laquelle nous conduit la position kantienne, imaginons le cas suivant : admettons que le concept de "triangle" est contenu dans les catégories pures de l'entendement, et que le concept empirique de "chien" se construit lui-aussi, autour d'un premier noyau autour duquel les définitions viendront s'agglutiner ou s'ordonner harmonieusement. Mais dans ce cas, il serait intéressant de se poser la question suivante : comment notre intellect procède-t-il dans le cas où il a affaire ni à un objet abstrait, ni à un objet empirique, mais à un objet *inconnu*, c'est-à-dire, un objet non-familier ou composé, qui refuse de s'inscrire, de s'aligner dans une catégorie déjà prédéfinie ?

S'il est vrai que l'entendement "fait" l'objet en suivant les règles de construction du schématisme transcendantal, on peut légitimement se demander, à la suite d'U. Eco, "quel est le problème auquel Kant aurait eu affaire en se trouvant face à l'ornithorynque ?"[66]. Kant aurait été alors obligé d'imaginer un schème « en partant d'impressions sensibles, mais ces impressions sensibles ne s'adaptaient à aucun schème qui lui aurait été acquis par avance (comment pouvait-on associer le bec et les pattes plumées au poil et à la queue d'un castor, ou l'idée d'un castor à celle d'un ovipare ? Comment pouvait-on voir un oiseau là où s'offrait un quadrupède, et un quadrupède là où s'offrait un oiseau?) Kant se serait retrouvé dans la même situation qu'Aristote lorsque, traçant toutes les règles possibles pour distinguer les ruminants des autres animaux, il ne réussit pas à situer le chameau, quelle que fût la position qu'il adoptât. Le chameau échappait à toute définition par genre et différence; s'adapta-t-il à l'une de

entendement, relativement aux phénomènes et à leur simple forme, est *un art caché dans les profondeurs de l'âme humaine* et dont il sera toujours difficile d'arracher le vrai mécanisme à la nature, pour l'exposer à découvert devant les yeux ». (E. Kant, 1787/1990: 153) (Nous soulignons)

[66] (U. Eco, 1997/1999: 125)

ces difficultés, il chassait alors de cet espace définitoire le bœuf qui rumine pourtant lui aussi"[67].

La morale de l'histoire est que « si une réflexion sur l'intuition pure de l'espace suffisait pour le schème de la figure géométrique, et si ce schème pouvait donc être tiré de la constitution même de notre entendement, il n'en est certainement pas de même pour le schème de « chien»[68] ou encore moins, pour le schème d' « ornithorynque ». En introduisant le schématisme transcendantal comme solution médiatrice pour la transition entre les intuitions sensibles et les catégories abstraites, Kant opère un saut conceptuel *ad hoc*, et se retrouve avec un "concept explosif"[69] entre les mains, sorti tout droit des "profondeurs cachées de l'âme humaine"[70]. Presque cent ans de pensée philosophique tombèrent dans le sillage du kantisme jusqu'à ce qu'on s'aperçoive que sa formule explicative reposait sur au moins deux grandes inconnues: *le dogme de la chose en soi*, et *l'art secret du schématisme.*[71]

Revenant à présent à la séparation états mentaux/états physiques, on constate que la « redistribution des rapports entre l'esprit et le corps place l'âme dans une position séparée : elle cesse d'être la "forme" et la propriété du corps pour devenir substance. La sphère du sujet devient dissociée de celle des choses matérielles »[72]. Ce que l'on a appelé par la suite le grand schisme cartésien corps/esprit intervient donc à la suite de l'introduction d'une scission ontologique à l'intérieur même du domaine du sujet. C'était, en quelque sorte, le prix à payer pour avoir rattaché les fondements des certitudes inébranlables à l'introspection intrinsèque du *cogito*.

7.3.4 Ecce intrusus !

Ces éléments d'arrière plan étant esquissés, revenons à présent à l'hypothèse que nous voudrions défendre ici, à savoir l'existence d'une filiation directe reliant la

[67] *Ibid.*, p. 124.
[68] *Ibid.*, pp. 121-2.
[69] *Ibid.*, p. 125.
[70] (E. Kant, 1787/1990 :153)
[71] De plus, comme le note judicieusement H. Gardner, "son argumentation était suffisamment difficile, ses critiques du rationalisme classique et des positions empiristes suffisamment fortes, son étude de la nature des processus mentaux suffisamment révolutionnaire, son opinion sur la philosophie suffisamment rassurante, pour que de nombreuses années passent avant que les points faibles de sa formulation n'apparaissent"[71]. (H. Gardner, 1985/1993 :77)
[72] (P. Engel, 1996: 27)

métaphysique traditionnelle clivée selon l'axe subjectivité/objectivité d'une part, et la tradition analytique contemporaine privilégiant la séparation nette entre les vérités factuelles d'une part, et les vérités formelles d'autre part. Nous suggérons que cet héritage passe par l'adoption de la distinction analytique/synthétique dans l'œuvre des « premiers » philosophes analytiques comme Carnap. « La dichotomie entre vérité analytique et vérité de fait – écrit Carnap[73] – est l'une des plus anciennes et des plus permanentes de l'histoire de la philosophie. (…) Pour la philosophie des sciences, une *distinction très nette entre analytique et synthétique* est à mon avis de la plus haute importance. La théorie de la relativité, par exemple, n'aurait pas vu le jour si Einstein ne s'était pas aperçu qu'il est impossible de déterminer la structure de l'espace et du temps physiques sans recourir à l'expérimentation physique. Il a vu clairement la *frontière* qui sépare d'un côté les *mathématiques pures*, avec leurs divers types de géométrie logiquement non-contradictoire, et de l'autre la *physique*, où seules l'expérimentation et l'observation ont le pouvoir de décider lesquelles d'entre les géométries sont susceptibles d'être le plus utilement appliquées au monde physique ».

Une lecture attentive de ce passage nous permet enfin clamer « *ecce intrusus* ! » Remarquez le glissement et le renversement de perspective qu'opère Carnap dans le passage cité ci-dessus : c'est *parce qu*'il y a une frontière nette entre les mathématiques pures (les vérités analytiques) d'une part, et la physique (les vérités de fait) d'autre part qu'un physicien comme Einstein a pu découvrir la relativité ! C'est parce qu'il a « découvert » laquelle des formes « pures », que nous avons presque envie de qualifier d'« en haut », s'applique au monde physique d' « en bas », que le réel s'ouvre à nous. La distinction analytique / synthétique si chère à Carnap a pour fonction centrale la « sauvegarde » critique des mathématiques pures. Ce n'est donc pas étonnant si pour Carnap les entités abstraites sont bel et bien réelles. Nous pourrions même dire que ces entités sont presque plus « réelles » que les chats et les chiens, puisqu'elles sont « pures », i.e. atemporelles. Rappelons-nous que chez Kant, l'objet physique qui affecte nos sens est avant tout la construction d'un entendement qui pré-organise et précède la sensibilité. Dans cette optique, la distinction traditionnelle pensée/monde s'évapore au profit d'une différenciation raison/chose en soi. Or, paradoxalement, chez les philosophes analytiques, la quête des « îlots de certitude inébranlable » finit par

[73] R. Carnap, (1966), *Philosophical Foundations of Physics*, Basic Books. Trad. Fr. J.-M. Luccioni et A. Soulez, *Les Fondements philosophiques de la physique*, (1973), Paris : Armand Colin, p. 249. (Nous faisons référence ici à l'édition française) (Nous soulignons)

éloigner la réflexion philosophique de l'étude de la chose-en-soi, (qui sera abandonnée aux sciences particulières), et elle migrera soit vers l'examen formel de la "syntaxe universelle" de la langue de la physique comme langue universelle de la science (comme chez Carnap), soit vers l'analyse conceptuelle (comme chez Jackson et Chalmers 2001). Mais il est clair que pour prétendre pouvoir opérer chirurgicalement la distinction analytique / factuel à la façon dont Carnap le souhaite, il faut avoir d'ores et déjà endossé le point de vue du Super-Sujet neutre, surplombant le cadre. Il importe de souligner ici que le vecteur commun des *deux* types d'approches (le physicalisme sémantique carnapien et le physicalisme *a priori*) réside dans la stratégie privilégiant l'analyticité ou l'apriorité, seules et lointaines reliques du royaume des Idées platoniciennes.

Or, en y regardant de plus près, on s'aperçoit que ces pratiques tombent implacablement sous le couperet de *l'objection du surnaturalisme*. Rappelons-nous que, selon la définition qu'en donne B. Stroud[74], un élément est surnaturel s'il est conçu comme un agent, force ou élément qui se situerait d'une façon ou d'une autre *à l'extérieur* du monde naturel et familier auquel nous sommes habitués et dont les actions ne pourraient pas être comprises comme faisant partie de ce dernier. On peut arguer dès lors que le découpage du réel en vérités factuelles et vérités formelles, i.e. en endossant un hypothétique point de vue en surplomb, débarrassé de toute « contamination » de la subjectivité naturelle du sujet connaissant, et en s'appuyant uniquement sur les outils purement formel de la rationalité, repose sur l'adoption tacite et non-questionnée du clivage de veine platonicienne opposant la précarité de la subjectivité à l'universalité de l'objectivité.

Afin de prendre la véritable mesure du « surnaturalisme » latent qui alimente souterrainement ce clivage, il convient de situer avec plus de précision les manœuvres qui accompagnent les « stratégies en fauteuil ». Pour ce faire, nous allons raviver ici très brièvement les braises d'une vieille querelle concernant la question des entités abstraites, controverse qui a replongé la philosophie analytique contemporaine dans le bourbier des questions métaphysiques traditionnelles.

[74] B. Stroud, (1996), 'The Charm of Naturalism' in *Proceedings of the American Philosophical Association* 70, pp. 43-55. Réédité in M. de Caro et D. Macarthur (eds.), (2004), op.cit., pp. 21-35.

7.3.5 Retour à une vieille querelle: Carnap, Quine, et la question de l'être

Vers le milieu du siècle dernier, il fut un moment quand, comme le note P. Jacob, « Quine et Goodman d'un côté, Carnap de l'autre, s'accusaient mutuellement de faire de la métaphysique sans le savoir »[75]. Un conflit[76]philosophique opposa donc l'architecte-en-chef, si l'on peut appeler ainsi, de l'empirisme logique, Carnap, à son brillant et rebelle disciple, Quine. Bien que les deux penseurs désavouaient ouvertement la métaphysique traditionnelle et son appareillage conceptuel *a priori*, et bien qu'ils saluaient l'investigation scientifique comme étant le cœur même de la philosophie nouvelle, ils n'étaient pas d'accord sur la manière dont celle-ci devrait s'emparer de la place désertée par l'ancienne métaphysique idéaliste. Le débat s'organisa autour de la publication d'un article signé par Quine[77], suivi d'un autre publié par Carnap[78], suite auxquels Quine écrira son célèbre 'Deux dogmes de l'empirisme'[79].

Réduit à sa plus concise expression, le cœur du débat tourne autour du statut des objets abstraits, comme les nombres. Carnap reconnait que les empiristes sont en général méfiants à l'égard des entités abstraites, car elles semblent violer les principes fondamentaux de l'empirisme et nous obligeant ainsi à endosser une ontologie de type platonicien. L'ambition de Carnap est de montrer que parler d'entités abstraites non seulement n'implique pas l'adoption d'une telle ontologie suspecte, mais que l'existence des entités abstraites est « parfaitement compatible avec l'empirisme est la pensée rigoureusement scientifique »[80]. Quine critique et rejette les tentatives de Carnap de trouver un compromis entre l'acceptation des formes *a priori* et le credo empiriste. Pour Quine, accepter la réalité des entités abstraites revient à accepter, *ipso facto*, une forme plus ou moins affichée de platonisme.

---

[75] (P. Jacob, 1980: 163)
[76] S. Soames (2009), '*Ontology, Analyticity, and Meaning: The Quine-Carnap Dispute*', In D. Chalmers, D. Manley & R. Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, op. cit., p. 424.
[77]W.V. Quine, (1948), 'On What There Is', *Review of Metaphysics* 2 (1948): 21–38; repris in Quine, (1980) *From a Logical Point of View*, Cambridge, MA: Harvard University Press. Trad. fr. S. Laugier, (2003), *Du point de vue logique*, Paris : Vrin.
[78] R. Carnap, (1950), 'Empiricism, Semantics, and Ontology', *Revue internationale de philosophie* 4, 20–40; réédité et revisé in *Meaning and Necessity*, Chicago, IL: University of Chicago Press, 1956). Trad. fr. par F. Rivenc et P. de Rouilhan *in* R. Carnap, (1997), *Signification et nécessité*, Paris, Gallimard. (Nous faisons ici référence à l'édition française)
[79] (W.V. Quine, 1951)
[80] (R. Carnap, 1950/1997 : 314-5)

Beaucoup de choses seraient à dire sur ce célèbre débat. Limitons-nous, pour simplifier, à son cadre général, afin de mieux calculer son impact sur le sujet qui nous intéresse ici, à savoir le conflit qu'oppose les partisans du physicalisme *a priori* à leurs détracteurs « naturalistes ».

Le coup d'envoi fut donné par la publication en 1948 de l'article de Quine 'On What There Is' – 'Qu'est-ce qui *est*' – qui aborde la question de l'engagement ontologique et des entités ontologiquement « suspectes », comme les entités abstraites. Prenons un court exemple : chacun admet que le cheval ailé « Pégase » n'a pas de référent spatio-temporel réel. Et pourtant, très peu de personnes seraient prêtes à admettre que les termes « Pégase » et « rien » sont équivalents. Par conséquent, si la proposition « Pégase n'est pas *rien* » est vraie, alors son revers « Pégase *n*'est *pas* » doit être faux. Mais dans ce cas, il faut trouver ce que Pégase *est*. Certains penseurs, objecte Quine, construisent des ontologies artificielles dans le but manifeste d'échapper à ce type de difficultés, bien que ni l'homme de la rue, ni le savant n'adoptent guère cette prolifération ontologique.

Baptisons McX, propose Quine, le philosophe qui soutient que la façon la plus simple d'expliquer la vérité de *p* « Pégase est» consiste à admettre que Pégase désigne une entité mentale. Autrement dit, l'univers de McX est peuplé d' « idées », « significations », « propositions », etc., pourvu qu'elles soient localisées en quelque sorte dans la pensée des êtres humains. Selon Quine, dont l'argumentation ne convient pas de reprendre en détail ici, l'analyse du langage ordinaire ne nous contraint nullement d'adopter une ontologie peuplée d'idées: que l'on parle de Pégase ne signifie pas que Pégase est quelque chose, à savoir *l'idée de Pégase*. De façon analogue, si l'on parle de l'idée du Parthénon, cela n'implique pas qu'une entité de type l' « idée du Parthénon » *soit*, sous prétexte que le Parthénon existe. Conséquemment, nul n'est logiquement contraint de reconnaître l'existence des *idées* pour donner un sens à l'énoncé « Pégase n'est pas », car « Pégase » et « l'idée de Pégase » sont deux noms différents, et l'existence ou l'inexistence de l'entité à laquelle réfère l'un n'implique nullement l'existence ou l'inexistence de l'entité à laquelle réfère l'autre, et réciproquement.

Quine se montre hostile également à la distinction entre « être » et « exister », parce que le terme « existe » n'a pas plus de connotations spatio-temporelles que "est". Ce sont les termes singuliers qui mentionnent les objets contenant ou ne contenant pas

une référence à l'espace et au temps. Car la question de savoir si une entité *est* ou *n*'est *pas*, ne peut pas être traitée indépendamment de *notre* expérience sensorielle en tant que sujet connaissant. Par conséquent, la distinction analytique/ synthétique tourne dans le vide et n'a pas plus de pertinence qu'un dogme brut. La langue canonique de la science étant pour Quine le calcul des prédicats, le quantificateur existentiel suffit à exprimer l'être ou l'existence, selon les valeurs des variables d'individus sur lequel il porte. Donc, pour le dire succinctement, *pas d'entité sans identité*. Bien plus, la prolifération des « entités » qui peuplent les mondes possibles est inutile et dangereuse, car elle peut nous égarer du droit chemin de l'empiricité.

Une fois que l'on a pris la mesure de cet égarement, insiste Quine, il faut « raser la barbe de Platon avec le rasoir d'Occam » et énoncer la thèse de l' « engagement ontologique », dont le but c'est de circonscrire avec plus de précision les entités qu'une théorie peut utiliser dans la description de la réalité : une théorie admet tels objets si l'affirmation de l'existence de tels objets est nécessaire pour rendre la théorie vraie : "Il y a des choses de l'espèce *F* si et seulement si $(X)Fx$ "[81]. Pour reprendre son fameux exemple : on peut dire que certains chiens sont blancs sans pour autant nous engager à reconnaître soit la 'caninité', soit la 'blancheur' comme des entités. L'énoncé "Des chiens sont blancs" dit que quelques choses, qui sont des chiens, sont blanches, et, pour rendre cette affirmation vraie, les choses sur laquelle porte la variable liée doivent comporter quelques chiens blancs, mais pas nécessairement la caninité ou la blancheur. « Nous nous référons aux variables liées – écrit Quine - eu égard à l'ontologie, non afin de savoir ce qu'il y a, mais pour savoir ce qu'une observation ou une doctrine donnée, la nôtre ou celle d'autrui, *dit* ce qu'il y a; et ceci est un problème qui, tout à fait à juste titre, concerne le langage. Mais ce qu'il y a est une autre question[82].

Le principe de l'engagement ontologique permet d'investiguer les types d'objets que les différentes théories admettent, rendant ainsi concevable leur pluralité, et clarifiant leurs éventuelles dissensions. Quine prône dès lors l'abandon de toute entreprise de *fondement* des sciences de la nature et substitue aux entreprises de reconstruction rationnelle de celles-ci sur l'expérience immédiate, les *postulats* du physicalisme. Le physicalisme, et c'est ici un point essentiel pour le cœur du débat que nous examinons ici, peut donc prendre la place de la branche "doctrinale" de

[81] (W.V. Quine, 1969/1977 : 113)
[82] (W.V. Quine, 1953 : 15-16)

l'épistémologie: il suffit alors, une fois abandonnée toute philosophie encombrée de la recherche d'un hypothétique fondement, de se servir de la psychologie pour étudier comment le sujet humain projette *sa physique* à partir de ses données sensorielles, pour faire de l'épistémologie un chapitre de psychologie et donc de science naturelle : « L'ancienne épistémologie aspirait à contenir, en un sens, la science de la nature, qu'elle aurait voulu construire à partir des données sensorielles. Inversement, dans sa nouvelle présentation, la science de la nature contient l'épistémologie, à titre de chapitre de la psychologie. (…) Nous cherchons comment ***le sujet humain***, que nous étudions, postule des corps et projette ***sa physique*** à partir de ses données, et nous tenons compte que notre situation dans l'univers est juste la même que la sienne»[83].

Mais que dire alors des entités mathématiques ? Que celles-ci entrent dans la composition de théories scientifiques, cela ne fait l'ombre d'un doute. Comment traiter alors la question de leur existence ? Nous touchons ici à un sujet controversé et complexe. De manière générale, l'histoire des mathématiques nous renseigne que les mathématiciens de toutes les époques se sont retrouvés inéluctablement embourbés dans le marais des entités abstraites. Les historiens s'accordent à identifier trois courants majeurs dominants, à savoir le réalisme, le conceptualisme et le nominalisme, et qui réapparaissent dans la philosophie des mathématiques du XXe siècle sous les visages du *logicisme*, de l'*intuitionnisme*, et du *formalisme*. Il serait trop long, et d'ailleurs superflu, d'entrer ici dans les détails de ces controverses. Limitons-nous, pour simplifier, au sujet qui nous concerne ici.

Le réalisme, considéré en lien avec la doctrine médiévale des universaux, est la doctrine de veine platonicienne selon laquelle les entités abstraites ont une existence propre, indépendamment de la pensée. L'esprit peut à la limite y accéder, les découvrir, mais non pas les créer. Or, le logicisme, représenté par Frege, Russell, Whitehead, Church, et Carnap, tolère un usage des variables liées qui fait référence indistinctement à des entités abstraites, connues et inconnues. Quine établit dès lors un lien entre le *réalisme* et la doctrine platonicienne selon laquelle les universaux ou les entités abstraites ont une réalité indépendante de l'esprit.

Maintenant, il est vrai que l'étiquette qu'appose Quine sur le logicisme de Carnap semble justifiée. D'autre part, on peut comprendre les raisons qui ont poussé Carnap à se défendre contre l'accusation d'être un platonicien, car cette connotation le renvoyait

[83](W.V. Quine, 1969/1977: 97) (Nous soulignons)

ipso facto dans le marais des métaphysiques traditionnelles dont il tenait à tout prix s'en détacher.

Dans ce qui suit, nous allons essayer de synthétiser la stratégie qu'emploie Carnap pour à la fois : (i) « sauvegarder » les entités abstraites chères à son cœur de logicien, et (ii) se plier aux normes de scientificité que lui impose son adhésion à l'empirisme et à ses contraintes de vérifiabilité. Après tout, comme nous l'avons vu dans le chapitre 2 de cet essai, Carnap annonce, dans son ouvrage programmatique *La syntaxe logique du langage* que la philosophie est amenée à être remplacée par la logique de la science, c'est-à-dire par l'analyse logique des concepts et propositions des sciences. Or, son empirisme logique le contraint à abandonner la métaphysique traditionnelle, et à la remplacer avec une doctrine plus adéquate. Comment entend-il réaliser cet exploit ?

En premier lieu, propose Carnap, « nous devons distinguer deux genres de questions d'existence : d'abord, les questions portant sur l'existence de certaines entités de ce nouveau genre *à l'intérieur du cadre* ; nous les appelons *questions internes* ; en second lieu, les questions portant sur l'existence ou la réalité du *système d'entités pris comme un tout*, que nous appelons *questions externes* »[84]. Maintenant, il est clair que les plus problématiques sont les questions qui tombent dans cette dernière catégorie. « Contrairement aux premières – poursuit-il – [la question externe de la réalité du monde des choses lui-même] n'est posée ni par l'homme de la rue ni par les scientifiques, mais seulement par les philosophes. Les réalistes répondent par l'affirmative, les idéalistes subjectifs par la négative, et la controverse se poursuit depuis des siècles sans jamais être résolue. Et elle ne peut l'être, parce qu'elle est mal conçue. Être réel au sens scientifique veut dire être un élément du système ; ce concept ne peut donc être appliqué de manière sensée au système lui-même »[85].

Selon Carnap, l'acceptation des entités abstraites est représentée dans le langage par l'introduction d'un « cadre »[86]. Nous verrons plus loin que le poids de l'interprétation du concept de « cadre » est crucial pour la défense du point de vue qui nous concerne ici. Contentons-nous pour l'instant de présenter la stratégie carnapienne des « deux pas essentiels » de l'introduction d'un nouveau cadre : « D'abord,

[84] (R. Carnap, 1950/1997 : 315) (Carnap souligne)
[85] *Ibid*., p. 316.
[86] « framework » en anglais en original.

l'introduction d'un terme général, d'un prédicat de niveau supérieur, pour le nouveau genre d'entités, nous permettant de dire d'une entité particulière quelconque qu'elle appartient à ce genre (par exemple, « Rouge est une propriété », « Cinq est un nombre »). Ensuite, l'introduction des variables d'un nouveau type. Les nouvelles entités sont les valeurs de ces variables, auxquelles les constantes (et s'il y en a, les expressions closes composées) sont substituables. (…) Après l'introduction de nouvelles formes dans le langage, il est possible de formuler grâce à elles des questions internes et des réponses possibles à ces questions. Une question de ce genre peut être soit empirique soit logique ; selon le cas, une réponse vraie sera soit factuellement vraie, soit analytique »[87].

Donnons un exemple pour plus de clarté. Prenons l'énoncé (a) :

(a) « Chicago est grand est une proposition ».

Maintenant, à l'aide de nouvelles variables, on peut former des énoncés généraux de type :

(b) « Pour tout *p*, *p* ou non-*p* ».

(c) « Il existe *p* tel que *p* n'est pas nécessaire et non-*p* n'est pas nécessaire ».

(d) « Il existe *p* tel que *p* est une proposition ».

Selon Carnap, « (c) et (d) sont des assertions internes d'existence. L'affirmation "Il y a des propositions" peut être comprise au sens de (d), auquel cas elle est analytique (puisqu'elle découle de (a)), et même triviale. Mais si l'affirmation est comprise en un sens externe, elle n'est pas cognitive »[88]. Le point clé à retenir ici c'est que le système de règles pour les propositions du cadre choisi est *suffisant* pour l'introduction du cadre, autrement dit, tout éclaircissement supplémentaire concernant la *nature* des propositions est « théoriquement superflu, parce que s'il est correct, il découle des règles »[89].

A présent, la question qui brûle les lèvres des adversaires de Carnap est la suivante : que les règles soient suffisantes pour établir un cadre, soit ! Mais *qui* introduit les règles ? *Qui* énonce les propositions ? Et *qui* décide où faut-il poser les *cadres du cadre*, si l'on peut dire ainsi ? Devinant cette objection, Carnap prend les devants et essaie d'écarter la menace du mentalisme subversif : « Est-ce que les

[87] (R. Carnap, 1950/1997 : 324-5)
[88] *Ibid*., p. 320.
[89] *Ibid*.

propositions sont-elles des événements mentaux comme dans la théorie de Russell ? » – s'interroge Carnap. Non ! – répond-il aussitôt. Comment justifie-t-il cette réponse ? Rendons ici la citation en entier, car elle est révélatrice du subterfuge qu'utilise Carnap pour contourner la question transversale du « sujet-connaissant-énonçant-*p* ». Voici ce que répond Carnap : « Un coup d'œil sur les règles nous montre que non, car, sinon, les affirmations existentielles seraient de la forme : "Si l'état mental de la personne en question remplit telles ou telles conditions, alors il existe *p* tel que… ". Le fait qu'*aucune référence* ne soit faite à des conditions mentales dans les affirmations existentielles (comme (c), (d), etc.) montre que les propositions ne sont pas des entités mentales. (…) Le fait que, dans ces affirmations, ne figure *aucune référence à un sujet (un observateur ou un sujet connaissant* ; rien de tel que : "Il existe *p* qui est nécessaire pour M. *X*") montre que *les propositions ne sont pas subjectives* (non plus que leurs propriétés, telles que la nécessité, etc.) »[90].

Si Carnap accuse Quine de ressusciter les vieux démons de la métaphysique traditionnelle, Quine[91] rétorque : « c'est précisément envers ces mots « insensés » que je me sens le plus libre de leur spécifier une signification ». Et si Carnap soutient que les questions ontologiques ne sont pas des questions factuelles, mais des questions destinées à fixer un schéma conceptuel adéquat pour la science, alors Quine rétorque : si tel c'est le cas, alors la même contrainte doit s'appliquer non seulement aux questions métaphysiques, mais aussi à chaque hypothèse scientifique, sans aucune exception.

L'impasse est évidente : peut-on choisir, à l'instar de Carnap de *ne pas* se soucier des questions « théoriquement superflues » qui nous viennent du soi-disant « extérieur » du cadre ? A cela nous proposons de répondre par la négative. Car pour pouvoir penser l'existence d'un cadre à la façon de Carnap, il est nécessaire d'avoir *déjà* endossé une position métaphysique discutable qui consiste à accepter la présupposition tacite selon laquelle il serait possible au théoricien de s'extraire du ledit cadre afin de l'analyser en surplomb. Pour dire les choses plus clairement, c'est parce que la *position épistémique* que le sujet connaissant-Carnap endosse subrepticement *est déjà externe au cadre* donné, qu'il peut décréter que telles questions sont *internes* et telles autres *externes* au cadre. Sans cette perspective en surplomb au-dessus du cadre, perspective qu'il s'offre en cadeau d'ailleurs, jamais aucun théoricien n'aurait pas pu

[90] *Ibid*., pp. 320-1. (Nous soulignons)
[91] (W.V. Quine, 1966a : 203)

fixer une délimitation aussi claire et nette que celle qu'il prétend instaurer entre les questions externes et internes au système examiné. Le soin avec lequel Carnap découpe le périmètre du cadre déterminé le place *ipso facto* dans une position épistémique privilégiée, située *en dehors* du cadre. Or c'est justement ce privilège que Quine lui conteste.

Certes, aucune référence *explicite* au sujet observateur n'est faite dans le système de propositions (a)-(d). Mais la référence y est *implicite*, puisque tout énoncé, et plus généralement, tout système de propositions, reste attaché, comme nous l'avons montré dans le deuxième chapitre de cet essai, aux deux pôles extralinguistiques classiques : le sujet locuteur et l'état de choses référent. Pour voir cela, il suffit de poser la question suivante : peut-on concevoir les énoncés (a) – (d) *sans* un énonciateur les énonçant ou *sans* contenu référentiel ? Et si on prend le parti de répondre par oui, d'où jaillissent-ils ? Car une chose est sûre : les énoncés ne flottent pas dans un ciel sémantique ou métaphysique situé au-dessus de la tête des philosophes. A moins qu'ils y flottent malgré tout. Mais dans ce cas, les accusations de platonisme sont justes et l'intrusion de la métaphysique dévastatrice. Comment éviter cette impasse ?

Toute la délicate dentellerie de l'édifice carnapien repose sur son interprétation particulière du concept de « proposition ». Pour Carnap, ce concept n'est utilisé :

(i) ni pour une expression linguistique, car dans ce cas il serait obligé d'affronter la verticalité sémantique de l'enchâssement dans un discours portant sur le réel ;

(ii) ni pour un événement subjectif ou mental, car dans ce cas il serait obligé d'affronter les démons du psychologisme qu'il abhorrait.

La proposition, dit Carnap, est quelque chose d'**objectif** qui peut être exemplifié ou non dans la *nature*. C'est son versant « empirique ». Mais aussi, et c'est ici le véritable tour de prestidigitation de Carnap sur lequel nous aimerions attirer l'attention du lecteur, la proposition s'applique également à toutes les entités d'un certain type *logique*, à savoir celles qui peuvent être exprimées dans un langage par un énoncé déclaratif. *C'est à la proposition qu'échoit l'acrobatique mission de relier la nature (empirique) à la logique (abstraite).* C'est dans la proposition qui a lieu la mystérieuse fusion entre le versant empirique et le versant abstrait de l'empirisme logique. Pourquoi est-il tellement important de relier ces deux éléments là ? La réponse est d'une désarmante simplicité : parce que les deux éléments en cause, i.e. la nature et la

logique, sont *objectifs* ! Le réel est objectif parce qu'il est là, massif, indépendant de la pensée. La logique est objective parce qu'elle est à l'abri des agitations de la subjectivité et du psychologisme. La même manœuvre a pu être détectée chez Jackson, lorsque celui-ci défend son physicalisme *a priori* en essayant de montrer qu'il serait possible d'opérer une déduction *a priori* à partir des propositions énonçant des « vérités fortunées », scientifiques, vers des propositions énonçant des vérités « infortunées », c'est-à-dire des propositions contenant des termes mentaux, ou des termes du vocabulaire du sens commun. La cible y est commune : l'obtention de l'objectivité. Traditionnellement, la réussite de cette mission passe par le rejet de toute latitude subjective ou subjectivisante.

On s'aperçoit alors que chez Carnap, et plus tard chez le « deuxième » Jackson la proposition relie *deux formes objectivées* respectant l'ancien axe subjectivité - objectivité : la logique est une forme objectivée de la pensée (du sujet), et la nature est une forme objectivée du réel extérieur (de l'objet). La proposition relie ces deux formes objectivées *à l'intérieur* d'un cadre, d'un système clos qui repousse les dragons des mentalismes, des subjectivismes, et des psychologismes, hors de l'espace légitime de la pensée objective. Or, pour concilier ces formes objectivées avec l'irruption inévitable de la latitude subjective, les deux penseurs intercalent des sauts conceptuels censés articuler et greffer la « pureté » formelle sur le « bourbier » empirique[92].

Que doit-on en conclure à la suite de ce bref interlude rétrospectif ?

D'abord, que Descartes ne doit pas « siéger » seul sur le banc des accusés dans le procès des dichotomies opposant le domaine de la subjectivité au domaine du monde objectif et physique. Pour être tout à fait précis, il faut remonter au moins aux sources platoniciennes du clivage entre le monde immuable des Idées et l'empiricité de la nature. Cette filiation est importante, dans la mesure où quelques deux mille ans après, un penseur comme Frege « croyait » encore à l'existence d'un monde logico-mathématique objectif primitivement détaché du règne naturel. Elle est importante aussi dans la mesure où la prééminence du modèle de recherche logico-déductif influencera profondément la réflexion philosophique moderne. Enfin, elle est importante care elle

[92] A notre sens, cette volonté de rattacher ces deux formes objectivées risque de n'avoir pas plus de pouvoir explicatif positif qu'une figure de style artistique. Après tout, quelqu'un pourrait interpréter une expression comme « empirisme logique » ou « physicalisme *a priori* », comme étant deux oxymores de même genre que « feu froid » ou « clarté obscure ».

indique la présence, au sein même des courants de pensée ouvertement antimétaphysiques comme l'empirisme logique du Cercle de Vienne, d'un noyau de croyances quasi-mythologiques accordant aux vérités analytiques ou *a priori* un statut privilégié pour le moins mystérieux. La séparation nette entre vérités analytiques/vérités factuelles est une réminiscence de l'ancien clivage Idées/empiricité.

Pour résumer, disons que le clivage subjectivité/objectivité traverse, telle une colonne vertébrale invisible, l'évolution de la pensée philosophique moderne, et se manifeste sous différents visages. En effet, à y regarder de plus près, qu'est-ce que le divorce contemporain entre le « physicalisme » et les « *qualia* » sinon une autre manifestation remise au goût du jour de ce clivage que nous avons baptisé ici le « Big-bang Gouffre » ? A nos yeux, le problème ne réside pas dans le fait que la putative immatérialité subjective des *qualia* s'accorde mal avec la matérialité objective du monde naturel, physique, mais dans l'hypothèse même qu'une telle séparation aurait un sens. Le problème n'est pas de trouver une localisation à l'esprit dans le monde matériel, mais de justifier l'intelligibilité même d'une telle distinction entre l'esprit et la matière suivant un clivage conceptuel rigide, et qui est loin de refléter une réalité où dominent moins des entités bien découpées sur une toile d'arrière plan solide, que des structures et des systèmes de relations en perpétuel mouvement.

Reprenons à présent, avec un sens critique mis en éveil, la controverse entre les partisans du physicalisme et les avocats de l'argument de la connaissance.

7.4 Naturaliser le physicalisme?

Au terme de cet interlude comparatif entre les deux visages du physicalisme (sémantique et *a priori*), on s'aperçoit que tout au long de notre itinéraire depuis les premières tentatives de 'physicaliser' les énoncés protocolaires, en passant par les différentes étapes de la théorie de l'identité psychophysique, du fonctionnalisme, et jusqu'au physicalisme *a priori*, une question semble avoir été laissée en suspens : comment explique-t-on le fait qu'une conception de veine empirique née de l'ambition déclarée d'habiliter les sciences physiques à légiférer sur ce que les choses *sont*, s'est métamorphosée peu à peu en une doctrine métaphysique *a priori* aspirant à légiférer, d'en fauteuil, sur ce que les choses *doivent être* ? Les signataires du Manifeste de Vienne ne se retourneraient-ils pas dans leurs tombes à entendre un métaphysicien

contemporain comme J. Lowe[93] écrire : « la métaphysique sonde l'essence du réel beaucoup plus profondément que les sciences empiriques, même plus profondément que la physique, car elle nous fournit la trame fondamentale à l'intérieur de laquelle ces sciences sont conçues et reliées les unes aux autres » ? Comment expliquer ce spectaculaire regain des thèmes métaphysiques, allant même jusqu'à penser leur répartition au sein d'une métamétaphysique[94]?

Cette section sera dédiée à l'examen de cette interrogation et sera agencée de la manière suivante : dans la section 7.4.1 nous allons fixer sur une toile d'arrière plan le portrait robot final du physicalisme tel qu'il en ressort au terme de l'examen détaillé dont il a fait l'objet jusqu'ici. Une fois cette toile de fond fixée, nous allons essayer d'y situer les « dogmes » sur lesquels repose ce portrait standard (section 7.4.2). A la suite de M. Bitbol, J. Ladyman, D. Ross, M. Esfeld, nous verrons que certains préceptes de la métaphysique *a priori* physicaliste contemporaine sont ouvertement récusés par des théories scientifiques. Bref, le physicalisme standard est accusé de reposer sur une interprétation simplificatrice, voire erronée de certains concepts physiques cruciaux comme « espace-temps », ou « propriété fondamentale ». La question qui surgira alors sera la suivante : comment expliquer ce formidable paradoxe qui place le physicalisme, en tant que 'Weltanschauung' de l'époque moderne, en conflit ouvert avec les sciences physiques, celles-là mêmes qui la justifient dans son ambition de doctrine ontologique unificatrice ?

7.4.1 Le Physicalisme : portrait robot final

La construction « sémantique » du premier physicalisme fut précédée, puis accompagnée, comme nous avons pu le constater dans le chapitre 2, d'un mouvement parallèle de rejet de toute latitude psychologisante, ou subjective. Comme l'a fort bien montré J. Dupré[95], ce rejet fut en quelque sorte une victime additionnelle du rejet des « âmes », des « fluides vitaux », et autres notions désuètes, définitivement disqualifiées par les recherches expérimentales des sciences de la nature. Bref, pour utiliser ici une métaphore, disons que le sujet psychologique a failli être jeté en même temps que l'eau

[93] E. J. Lowe, (2002), *A Survey of Metaphysics*, Oxford University Press, p. vi.
[94] Voir D. Chalmers, D. Manley & R. Wasserman (eds.), (2009), *Metametaphysics: New Essays on the Foundations of Ontology*, op. cit.
[95] J. Dupré, (2004), 'The Miracle of Monism', *in* M. de Caro & D. Macarthur (eds.) op. cit.

du bain vitaliste. Son principal défaut, il faut le dire, était d'être subjectif. C'est donc cette encombrante subjectivité qui porta préjudice à l'esprit et au sujet psychologique, les renvoyant tous les deux dans les limbes des pseudo-réalités. D'ailleurs, les définitions standard que l'on retrouve dans n'importe quel manuel de philosophie de l'esprit sur la « clôture causale du monde spatio-temporel » ont l'origine dans une acerbe critique « antisurnaturaliste »[96] située au début du XXe siècle et impliquant un véritable lynchage du concept vitaliste d'« entéléchie ». Plus tard, face à l'échec du projet de « reconstruction rationnelle » du réel mené dans l'esprit du Cercle de Vienne, les philosophes analytiques post-quiniens seront de plus en plus nombreux à suivre les conseils de Quine[97] les invitant à se focaliser, plus modestement, sur l'analyse de ce *phénomène naturel* qu'est le *sujet humain physique*. En réunissant, dans la même phrase, les termes clé '*nature', ' sujet',* et *'physique*', Quine délivre ainsi le certificat de naissance d'une philosophie du mental menée dans l'esprit d'une enquête psychologique – c'est-à-dire épistémologique, si par cette dernière on entend une science *naturelle*[98].

Comme le résume Quine de sa plume alerte, « Carnap cherchait ce qu'il appelait une reconstruction rationnelle. Toute construction du discours physicaliste en termes d'expérience sensible, de logique et de théorie des ensembles, aurait été considérée comme satisfaisante, qui eût accouché du discours physicaliste correct. S'il y a une manière d'arriver à celui-ci, il y en a plusieurs, mais d'en donner une, n'importe laquelle, eût été un coup de maître »[99]. Or, ce ne fut pas le cas. « Pourquoi donc toute cette reconstruction créative, tout ce faux-semblant ? »[100] – s'écrie alors Quine. Pourquoi ne pas se tourner vers les *sciences* pour voir comment les choses se passent réellement?

Par la suite, avec la publication coup sur coup des articles de H. Feigl[101], 'The "Mental" and the "Physical"', et de J.J. Smart[102] 'Sensations and Brain Processes', puis des réponses qu'ils ont suscitées, notamment de la part de H. Putnam[103] 'Psychological

[96] *Ibid.*, p.36
[97] W.V. Quine, op. cit., (1958/2004: 51)
[98] « L'épistémologie, ou quelque chose de ressemblant, s'est simplement conquis droit de cité à titre de chapitre de psychologie et donc de science naturelle ». W.V. Quine, op. cit., (1958/2004:51)
[99] (W.V. Quine, 1969/ 2004:43)
[100] *Ibid.*, p. 43.
[101] (H. Feigl, 1958)
[102] (J.J. Smart, 1959)
[103] (H. Putnam, 1967)

Predicates', et de D. Davidson[104], 'Mental Events', la pensée philosophique se redonne le droit de traiter des questions métaphysiques, concentrées de manière très spectaculaire dans le célèbre 'mind-body problem'. Par ricochet, la théorie de la réduction acquiert à nouveau une fonction métaphysique : il s'agit de déterminer les constituants ultimes du monde. Le physicalisme retrouve alors ses insignes de noblesse et devient une doctrine ontologique portant sur ce que les choses sont, et non pas uniquement sur ce que les propositions énoncent.

Nous avons pu constater également que malgré les dissensions accompagnant le véritable flot montant d'interrogations autour du projet de « localiser » l'esprit dans le monde physique naturel, il est possible néanmoins de délimiter ce que l'on peut appeler un « seuil minimal » dans la définition du physicalisme. D'abord, il convient de noter que le physicalisme est une doctrine ontologique qui se prête, comme le relève Stoljar[105], à une *double* caractérisation :

(i) du point de vue de l'*objet/propriété* physique ;

(ii) du point de vue de la *théorie* physique.

Autrement dit, le physicalisme affirme que : (a) tous les objets qui existent appartiennent ou bien à l'une des catégories étudiées par la physique, ou bien sont entièrement composés de parties qui appartiennent à l'une de ces catégories physiques. Et (b) : toutes les propriétés objectives de ces objets sont soit des propriétés étudiées par la physique, soit elles sont réductibles à celle-ci. Ce qui revient à dire que la plupart des philosophes physicalistes souscrivent à la *Thèse de la Corrélation*[106]suivante :

(TC) Pour chaque type d'état sensoriel S, il existe un type d'état fonctionnel ou physique *F/P*, en sorte qu'il est nomologiquement nécessaire que pour tout être *X*, *X* est en S si et seulement si *X* occupe le rôle *F/P*.

Essayons à présent de synthétiser le *portrait robot* du physicalisme actuel : au niveau fondamental – *dixit physicalismus* - il y a les électrons, les quarks, et autres entités microscopiques décrites par la mécanique quantique. Ces entités possèdent un certain type d'attributs, comme le spin, la masse, l'énergie, la charge, etc. Le niveau fondamental est donc le niveau de la physique. A l'étage immédiatement supérieur on

[104] (D. Davidson, 1970)
[105] D. Stoljar, (2001), 'Two Conceptions of the Physical', op. cit. Voir aussi (D. Stoljar, 2010).
[106] B. McLaughlin, (2001), 'In Defense of New Wave Materialism: a Response to Horgan and Tienson', in C. Gillett & B. Loewer, op. cit., p. 319.

retrouve les entités étudiées par la chimie, dont l'exemple paradigmatique est l'eau et ses propriétés de transparence, liquidité, etc. Puis on monte encore un étage et l'on retrouve la biologie et l'étude du vivant, des organismes considérés comme un tout et ayant des attributs et propriétés spécifiques comme la digestion, la reproduction, etc.

C'est à partir de là que les choses se compliquent. Car il est évident qu'une explication adéquate du réel doit tenir compte non seulement du processus de délimitation des différents niveaux de la pyramide physicaliste, mais aussi (et nous dirions surtout !) de la *médiation* entre les différents étages, bref, de l'*articulation* de jointure qui fait que le monde se présente sous la forme d'un tout intelligible et non pas d'un chaos méphistophélique. Comme nous l'avons déjà évoqué, parmi les candidats pour déterminer la nature de l'articulation entre les étages de la pyramide physicaliste, on compte : la réduction « classique », la survenance, la réalisabilité, et la nécessitation métaphysique. Comme le relève Stoljar[107], seulement ce dernier candidat, i.e. la thèse de la « nécessitation métaphysique »[108], déclare ouvertement ses visés métaphysiques.

(7) Le physicalisme est vrai si et seulement si chaque propriété instanciée est soit physique, et/ou soit elle découle par nécessité d'une propriété physique instanciée.

A partir de (7) il n'y a plus qu'un pas à faire pour arriver au principe du passage *a priori* de Jackson (1998/2003) selon lequel, chaque énoncé vrai concernant notre monde est déductible *a priori* à partir d'un énoncé formulable en termes physiques[109]. Cela peut être interprété de deux manières distinctes : on peut soit concevoir un énoncé vrai gigantesque formulable en termes physiques d'où découlerait *a priori* chaque énoncé vrai relatif à notre monde, soit on renverse la perspective et on affirme que pour chaque énoncé vrai à propos de notre monde, il y a un énoncé vrai (parfois gigantesque) formulable en termes physiques qui implique *a priori* tous les autres énoncés.

Rien de ce que nous venons de dire n'est nouveau ou révolutionnaire. Dans ce qui suit, nous soutiendrons que le physicalisme *a priori* de Jackson tel que nous venons de le définir, s'expose à des objections de veine quinienne venant de la part des auteurs

[107] (D. Stoljar, 2010 :36)

[108] La notion de *nécessitation métaphysique* que l'on utilise ici concerne l'acception désormais classique proposée par Kripke (1980) et qui stipule qu'une propriété F implique nécessairement une propriété G si l'instanciation de F est nécessairement accompagnée de l'instanciation de G. Cette idée est généralement énoncée en faisant appel à la notion de « mondes possibles ».

[109] "(…) the view that for each true statement concerning our world, there is a statement in physical terms that *a priori* entails that statement?" (F. Jackson, 2003: 155)

qui, s'appuyant sur la pratique expérimentale réelle des sciences physiques, contestent aux métaphysiciens installés dans le fauteuil « abstrait » de l'apriorité le droit de légiférer sur ce que les choses sont. Plus spécifiquement, il s'agit de remettre en cause le bien fondé de deux principes, véritables piliers de l'édifice physicaliste actuel : à savoir (i) la relation part/ensemble qui est censée être le fondement même de la microréduction ; et (ii) la métaphore des niveaux de réalité, censée soutenir à son tour, la thèse selon laquelle il serait en principe possible de déduire de manière *a priori* la plupart des vérités macroscopiques, d'une (hypothétique) description complète du monde en termes microphysiques *P*, et donc de construire des explications réductrices *a priori*.

7.4.2 Deux dogmes du physicalisme

Récemment, plusieurs voix se sont élevées pour dénoncer le portrait robot du physicalisme que nous venons d'esquisser plus haut, comme étant une tentative de « domestiquer » le réel. Plus spécifiquement, leur pétition[110] s'articule autour des trois axes suivants :

(1) Les métaphysiciens analytiques (même physicalistes) ignorent des faits et des aspects essentiels des pratiques scientifiques réelles.
(2) Ils font appel dans leurs argumentations à des exemples et à des cas de figure surannés, au lieu de se focaliser sur ce que la science offre présentement.
(3) Ils prétendent pouvoir décider (légiférer) de manière *a priori* dans l'investigation de certains sujets d'étude qui, selon eux, restent hors de la portée de la pratique scientifique proprement dite.

Un conflit ouvert éclate alors entre les partisans de la « philosophie en fauteuil » et ceux de la « philosophie en blouse blanche de laboratoire », si l'on peut dire ainsi.

Prenons quelques exemples, pour plus de clarté. Nous avons vu qu'une grande majorité de physicalistes s'accordent pour dire qu'à la base du réel il y a les électrons, les quarks, et autres entités microscopiques décrites par la mécanique quantique. C'est le niveau fondamental. S'appuyant sur la thèse du « niveau fondamental », beaucoup sont prêts ainsi à défendre la doctrine que l'on peut appeler « atomisme méréologique »

[110] (J. Ladyman & D. Ross, 2007 : 17)

selon laquelle tous les éléments du réel se décomposent en éléments « simples », indivisibles. Cette idée, qu'A. Hüttemann[111] a critiquée récemment sous le nom de « microphysicalisme », est profondément ancrée dans le schéma conceptuel du physicalisme standard.

Pour résumer, un véritable consensus solidarise les physicalistes autour des deux thèses suivantes :

(1) **La Thèse du Niveau Fondamental** (NF) selon laquelle le monde est organisé sur plusieurs de niveaux de réalité survenant sur un Niveau Fondamental Physique.

(2) **La Thèse de l'Atomisme Méréologique** (AM) selon laquelle tous les membres des niveaux mentionnés ci-dessus se décomposent (du moins en principe), ou se réduisent ou surviennent sur des éléments fondamentaux indivisibles (microphysicalisme).

Certains auteurs ont récemment signalé que l'adhésion tacite à ces deux thèses va à l'encontre de ce que la science physique actuelle nous apprend sur le réel. Conséquemment, cette obstination à opérer avec ces deux thèses que l'on peut qualifier de « dogmes » mène à des situations franchement absurdes: supposons que le physicaliste soutienne, comme le fait par exemple N. Markosian[112] (et avec lui tant d'autres) qu'un objet est physique si et seulement s'il possède une localisation spatiale, et que par conséquent le physicalisme est la doctrine qui affirme que tous les objets sont des objets physiques. Dans ce cas, force est de constater que le physicaliste est condamné à soutenir *ipso facto* que les entités traitées par la physique fondamentale sont non-physiques, et que par conséquent, l'univers lui-même est non-physique !

Un autre exemple est à trouver, notent Ladyman et Ross, dans un ouvrage aussi populaire que celui de J. Kim, *Mind in a Physical World*[113] qui, malgré l'ambition affichée dans le titre de traiter du problème de l'esprit dans un monde physique, non seulement ne fait référence à aucun exemple concret de théorie physique actuelle, mais qui plus est, ne possède même pas une entrée d'index pour le terme crucial de « physique ». Une bonne partie de son argumentation repose ainsi de manière cruciale

[111] A. Hüttemann, (2004), *What's Wrong with Microphysicalism?*, Routledge.
[112] N. Markosian, (2005), 'Against Ontological Fundamentalism', *Facta Philosophica* 7 (1):69-83.
[113] (J. Kim, 1998)

sur une interprétation pour le moins rocambolesque de « propriété micro-basée » dont le support serait « entièrement décomposable en des parties non-chevauchantes »[114].

Comme nous l'avons évoqué plus haut (chapitre 6, section 6.3), D. Lewis définit les propriétés fondamentales comme étant « des propriétés intrinsèquement parfaitement naturelles » dont l'instanciation est intrinsèque, c'est-à-dire non-divisible ou mixte : la masse ou la charge des particules élémentaires sont d'excellents exemples de telles propriétés parfaitement naturelles. C'est la mission de la physique de fournir un inventaire *complet* de toutes les propriétés et relations fondamentales qui sous-tendent le réel, ainsi que de découvrir les lois fondamentales qui agencent leurs actions. Il est vrai, concède Lewis, que n'avons aucune garantie de type *a priori* qu'un tel projet sera mené à bout de façon exhaustive, mais il n'est reste pas moins que beaucoup d'indices indiquent aujourd'hui que la physique se trouve sur le bon chemin vers la complétude. Il est donc préférable de rester optimiste et de parier sur l'évolution progressive de l'appareillage conceptuel des sciences physiques, dont le but ultime restera celui d'approcher le cœur battant du réel.

Selon Lewis, qui introduit ainsi la notion de « survenance humienne » [115], tout ce qui existe au monde est une vaste mosaïque d'affaires locales de faits particuliers. Prenons la géométrie : un système de points avec des relations externes de distance spatio-temporelle en eux. Peut-être des points de l'espace-temps lui-même, peut-être des points de matière. En ces points se trouvent des qualités locales : des propriétés intrinsèquement parfaitement naturelles qui n'ont besoin de rien de plus grand qu'un point auquel être instanciées. En bref, nous avons un arrangement des qualités. Tout le reste survient sur cet arrangement. Lewis[116] déclare par ailleurs que cette image lui a été inspirée par la physique classique. Dans une veine similaire, P. Van Inwagen[117] soutient que le tout du réel se décompose ultimement en des atomes méréologiques, et que ces derniers sont simples, i.e. indécomposables en des parties encore plus petites.

[114] *Ibid.*, p. 84.
[115] La thèse de la survenance humienne est appelée ainsi par D. Lewis parce qu'elle rejette l'existence de connexions nécessaires dans la nature. (Voir D. Lewis, 1986 : ix-x)
[116] (D. Lewis, 1999 : 226)
[117] P. Van Inwagen, (1990), *Material Beings*. Ithaca: Cornell University Press.

Le problème avec ces théories, insistent Ladyman et Ross[118], c'est que tous ces éminents philosophes considèrent, à tort, les hypothèses suivantes comme étant tacitement acquises :

(a) Il est permis d'utiliser la science physique *non-actuelle* comme acteur dans les *explananda* ciblées par la métaphysique ;

(b) Il est permis d'utiliser la manœuvre décrite en (a) ci-dessus *parce que* les caractéristiques détaillées de la physique mature *actuelle* – quelles qu'elles soient ! – vont finir d'une manière ou d'une autre, à un très haut niveau d'abstraction ou de généralité, par « s'ajuster » avec la physique non-actuelle respective, ce qui veut dire que les philosophes n'ont pas à perdre leur précieux temps avec des détails assommants venant des théories physiques, vu qu'en fin de compte, au niveau dit « fondamental », tous les ingrédients se mélangent en une grande et savoureuse « soupe » métaphysique universelle.

Ce n'est donc pas surprenant si une apologie explicite de ce type d'approche est identifiable dans l'imposant ouvrage de F. Jackson, *From Metaphysics to Ethics: A Defense of Conceptual Analysis*[119]où son auteur formule clairement les thèses contenues en (a)-(b) : « il est raisonnable de supposer – écrit-il – que la science physique, malgré ses inadvertances, a avancé suffisamment pour qu'on puisse lui faire confiance dans la caractérisation des genres de propriétés qui sont nécessaires dans l'élaboration d'une explication *complète* de la réalité non-sensorielle. Ces types de propriétés vont être les mêmes, au sens large, que celles présentes dans la pratique *courante* de la physique»[120]. Récemment, plusieurs auteurs[121] ont dressé une sévère critique de ce que nous pouvons dénoncer ici comme deux véritables « dogmes » du physicalisme, à savoir la Thèse du Niveau Fondamental (NF), et la Thèse de l'Atomisme Méréologique (AM) ou du Microphysicalisme[122]. Bien que la place nous

[118] (J. Ladyman & D. Ross, 2007 : 18-9)
[119] (F. Jackson, 1998)
[120] *Ibid.*, p. 7 (Nous soulignons)
[121] "The physical process may have no substantial roof of emergent properties, it has no substantial ground of elementary properties either, according to the most straightforward reading of Quantum Mechanics. Actually, this claim of groundlessness is presently creeping in many other branches of physics, besides ordinary Quantum Mechanics." Voir M. Bitbol, (2007), Ontology, Matter and Emergence, *Phenomenology and the Cognitive Science*, 6, 293-307.
[122] "Ascribing this priority is wrong because, as Bohr taught us, the concepts of macrophysics are exactly as indispensible to characterize microphenomena as the concepts of microphysics are indispensible to explain certain macrophenomena. In the quantum paradigm, the relation between the microscopic and macroscopic domains is not dissymmetrical; it is not a mereological part-aggregate type of relation. It is as fully symmetrical as one may conceive: phenomena and concepts relevant at the macrolevel are just as

manque ici pour traiter ces points en détail, tentons néanmoins d'en brosser ses traits d'ensemble.

La représentation classique standard d'un monde hiérarchiquement organisé sur plusieurs niveaux de réalité fut intronisée par le célèbre article de Putnam et Oppenheim[123] que nous avons déjà évoqué dans le premier chapitre de cet essai, lorsque nous avons examiné la notion de réduction: les phénomènes macrophysiques se réduisent à des phénomènes microphysiques par le truchement d'une véritable « cascade de réductions »[124]. La métaphore du réel stratifié permet de figurer le processus de réduction appliqué à l'ensemble du réel. Par la suite, bien que cette acception standard ait été presque unanimement rejetée, la métaphore des niveaux a reçu de nombreux ajustements. Or, récemment, certains auteurs ont pris le parti de rejeter cette attitude, et au lieu de tenter de donner un nouveau « lustre » à la métaphore initiale, ils la nient tout court. De manière générale, on peut classer ces attaques en deux catégories :

(i) Une première stratégie, plutôt spéculative, a été récemment défendue par J. Heil qui tente d'approcher le problème des niveaux du réel en lien avec sa propre solution au 'mind-body problem'. Dans son livre *From an ontological point of view*[125] Heil soutient que le fait que nous éprouvions de la douleur, de la joie, de la colère, etc., ne nous donne aucune justification de croire qu'il existerait des propriétés mentales de « niveau supérieur » s'additionnant aux propriétés physiques de « niveau inférieur », et cela non pas parce que les propriétés mentales seraient identiques aux propriétés physiques, mais parce qu'il n'existe tout simplement pas de telles entités appelées « propriétés mentales ». Heil s'inscrit ainsi en faux contre la structuration du monde par niveaux du réel, et soutient que la plupart des philosophes contemporains de l'esprit sont victimes d'une illusion qu'il baptise « la théorie picturale », et qui n'est pas à proprement dit une seule doctrine, mais un faisceau de thèses toutes convergeant vers un seul et unique noyau commun. Le cœur de la théorie picturale[126] est l'idée selon

much constitutive of the microlevel as the elements of the microlevel are taken to be constitutive (yet in another sense) of the macrolevel". *Ibid.*

[123] H. Putnam & P. Oppenheim, (1958), 'Unity of Science as a Working Hypothesis', op. cit.

[124] (M. Bitbol, 2000:13)

[125] J. Heil, (2003), *From an Ontological Point of View*, Oxford University Press. Trad. Fr. D. Berlioz et F. Loth, *Du point de vue ontologique*, Ed. D'Ithaque, Paris, 2010. Voir aussi J. Heil, (2003), 'Levels of Reality' *Ratio* 16 (3):205–221. (Référence faite à l'édition anglaise)

[126] "The core idea is that the character of reality can be 'read off' linguistic representations of reality. A corollary of the Picture Theory is the idea that to every meaningful predicate there corresponds a property. If, like me, you think that properties (if they exist) must be mind independent, if, that is, you

laquelle le sujet connaissant pourrait en quelque sorte « lire » la nature du réel en s'appuyant sur les représentations linguistiques qu'il en forge. Ce qui veut dire qu'à tout prédicat ayant une signification spécifique correspondrait une propriété située quelque part dans le monde réel.

L'idée derrière cette illusion c'est que le langage aurait la capacité de refléter le réel à la manière d'un peintre qui brosserait son propre autoportrait à l'aide d'un miroir. Cela ne peut mener, souligne Heil, qu'à des pseudo-problèmes. Car le découpage du réel en sujets et objets, avec le corolaire afférant états mentaux/ états physiques, n'est qu'une fâcheuse conséquence de la structuration linguistique d'une proposition sur l'axe sujet/prédicat.

(ii) Une deuxième catégorie de critiques concernant la métaphore des niveaux du réel puise ses objections dans les découvertes expérimentales des théories physiques les plus récentes, comme la mécanique quantique, arguant que cette dernière non seulement ne renforce pas l'idée d'un hypothétique niveau fondamental, mais le récuse ouvertement. Parmi les artisans de ce type d'attaque on compte M. Bitbol, J. Schaffer, A. Hütteman, J. Ladyman, D. Ross, M. Esfeld, etc., pour ne citer que ces quelques noms.

J. Schaffer[127], par exemple, récuse l'idée selon laquelle la science actuelle apporterait un appui décisif à la thèse du niveau fondamental censée soutenir, par le truchement de la survenance, l'armature physique du réel. Schaffer[128] observe judicieusement que la stratification du monde à partir d'un niveau primitif, fondamental, est étroitement consolidée par la thèse de la structuration méréologique des diverses « parts », ou « briques » composant l'ensemble. Plusieurs versions de structuration accompagnent dès lors la métaphore des niveaux de réalité :

(a) La structuration *méréologique* de type part/ensemble ;
(b) La structuration de type *survenance* ordonnée par des dépendances asymétriques ;
(c) La structuration de type *réalisabilité* ordonnée par des relations fonctionnelles ;
(d) La structuration *nomologique* ordonnée par des principes et lois-pont à sens unique entre les membres d'une famille de propriétés.

---

are ontologically serious about properties, you will find unappealing the idea that we can discover the properties by scrutinizing features of our language". (J. Heil, 2003:6)

[127] J. Schaffer, (2003), '*Is There a Fundamental Level?*', *Noûs* 37 (3):498–517.

[128] (J. Schaffer, 2003 : 500)

Schaffer soutient que non seulement le physicalisme, mais aussi la survenance humienne, l'épiphénoménalisme, l'atomisme, toutes ces doctrines sont redevables de la thèse du niveau fondamental, et cela même s'il s'avère que la microréduction est fausse. Il finit donc par rejeter en bloc l'idée d'un soi-disant niveau fondamental du réel. Ladyman et Ross notent, quant à eux, que l'image physicaliste standard selon laquelle les niveaux se différencient en fonction de dimensions est immotivée. Ils étayent ensuite l'analyse critique de Schaffer évoquée ci-dessus et apportent des arguments pour récuser les points de (a), (b) et (d), mais ils se montrent plus indulgents envers la thèse de la réalisation (c). La place nous manque ici pour analyser en détail les débats entre adeptes et partisans de la thèse du niveau fondamental. Soulignons seulement que selon ces auteurs[129], des études récentes de la mécanique quantique nous suggèrent que la thèse de la survenance humienne est fausse, et que toutes les propriétés de la physique fondamentale semblent extrinsèques aux objets individuels[130].

Une question surgit alors : que se passe-t-il si la science physique *courante* nous informe positivement qu'il n'existe pas de « propriétés micro-basées » dont le support serait« entièrement décomposable en des parties non-chevauchantes », ni de « propriétés intrinsèquement naturelles des points et des espace-points » ? Car il se trouve que la science physique a non seulement prouvé, mais aussi intégré dans ses comptes-rendus du monde, et cela depuis quelques bonnes générations déjà, le principe de la *non-séparabilité* des états quantiques. Que dire alors de l'affirmation célèbre de Schrödinger selon laquelle l'intrication[131] est l'essence de la physique quantique ?

Et si l'identité « Eau=H_2O », véritable talisman des physicalistes, était vide de contenu et de sens ? Puisqu'en réalité, comme l'a fort bien montré J. Van Brakel[132], ce que l'on appelle familièrement « eau » est une substance qui se manifeste sous des formes polymériques variées comme par exemple $(H_2O)_2$, $(H_2O)_3$, etc., i.e. des

[129] (J. Ladyman & D. Ross, 2007 : 151)

[130] Pour une discussion plus détaillée voir M. Esfeld, (2004). "Quantum entanglement and a metaphysics of relations." *Studies in History and Philosophy of Modern Physics*, 35: 601–617. Voir aussi M. Bitbol, (2007), 'Schrödinger against Particles and Quantum Jumps', in J. Evans & A.S. Thorndike, op. cit. Voir également M. Esfeld & V. Lam, (2008), "Moderate structural realism about space-time." *Synthese*, 160: 27–46.

[131] La non-séparabilité, reconnue depuis les débuts de la physique quantique, est essentielle à la compréhension des propriétés des atomes, molécules et solides. Chaque partie n'a pas d'état quantique. Il y a seulement une fonction d'onde pour le système global. Pour une discussion détaillée de ces points voir M. Bitbol, (1996), *Schrödinger's philosophy of quantum mechanics*, Kluwer. Et aussi M. Bitbol (1996/1997) *Mécanique quantique, une introduction philosophique*, Paris, Flammarion.

[132] J. van Brakel, (1986), 'The Chemistry of Substances and the Philosophy of Mass Terms', *Synthese* 69, 291–324

apparences qui se métamorphosent sans cesse, donnant ainsi naissance aux propriétés macroscopiques de l'eau de tous les jours. Est-ce que le fameux principe du passage *a priori* de Jackson (1998/2003) tiendra toujours si son armature même – à savoir l'identité « Eau= H_2O » - s'écroule sous le poids des découvertes émanées du sein même de la physique ?

Il y a de fortes raisons de penser que non. Comment expliquer dès lors ce formidable paradoxe qui caractérise la situation du physicalisme actuel bâtissant des châteaux explicatifs fantasques *sans* tenir compte des humbles fondements que la physique lui met à la disposition à la sueur de son front ? Pour quelle mystérieuse raison les métaphysiciens préfèrent-ils ne pas « se salir les mains » [133] en puisant dans la marmite encore chaude des pratiques scientifiques courantes, et préfèrent surplomber le monde et ses laboratoires depuis le fauteuil confortable qu'offre l'analyse conceptuelle ou la métaphysique abstraite ?

L'étrange singularité de cette situation a été formulée récemment par Quentin Smith[134] sous la forme d'un syllogisme interrogatif :

QUANTUM *QUALIA*

P1 Si les physicalistes affirment que le monde *est* physique (le mental y compris) et

P2 Si la mécanique quantique est une des théories physiques ayant le plus de succès prédictif à l'heure actuelle,

Alors,

(C) Pourquoi les philosophes de l'esprit physicalistes continuent-ils à modeler des théories de l'esprit comme si la mécanique quantique n'existait pas ?

N'y-a-t-il pas une formidable contradiction interne dans cette stratégie de l'autruche qui consiste à ignorer ce que la physique découvre sur le monde, tout en mettant fièrement en avant le physicalisme sous la bannière de la scientificité? La solution à ce paradoxe va de pair avec la divulgation des présupposés tacites qui asphyxient la pensée philosophique moderne. La principale raison pour laquelle les physicalistes se méfient massivement de la mécanique quantique, c'est que cette

[133] (J. Ladyman & D. Ross, 2007 : 21)
[134] Q. Smith, (2003), 'Why Cognitive Scientist Cannot Ignore Quantum Mechanics' in Q. Smith & A. Jokic, (eds.), *Consciousness - New Philosophical Perspectives*, op. cit., p. 409.

dernière pourrait introduire un véritable cheval de Troie au sein d'une vision « spartiate » et déformée de la relation du sujet connaissant au monde. Car la physique quantique remet en cause la bipartition subjectivité/objectivité, et cela, chose importante, depuis l'intérieur d'une forme objectivée du savoir, car n'oublions pas, la mécanique quantique est le fruit d'une conception *scientifique* du monde. En revanche, lorsque c'est la phénoménologie qui remet en cause le même clivage, cela a curieusement moins d'impact, car il s'agit d'une conception *philosophique*. Or, justement, remettre en question le clivage chirurgical Subjectivité/Objectivité revient à refuser, comme on l'a vu, d'endosser le point de vue du Super-Sujet. Ce refus plonge le fauteuil de la rationalité du théoricien-philosophe dans le bourbier de l'empiricité du monde d'où, il faut le dire tout de suite, il n'a jamais pu être détaché.

7.4.3 Naturaliser les *qualia* ?

Arrivés au terme de ce périple, nous n'aimerions pas donner l'impression que dans le duel physicalisme / *qualia* le premier protagoniste serait le « méchant » et le deuxième le « gentil ». Ce serait tomber dans le même piège du clivage surnaturaliste que de distribuer, d'en haut, des billes noires ou blanches aux protagonistes sages et aux récalcitrants. Au contraire, comme l'a fort bien montré M. Beaton[135], un véritable chantier conceptuel serait à entamer aussi du côté de l'analyse des *qualia*. Beaton plaide pour une définition plus modérée des *qualia*, proposant d'élargir et d'assouplir la palette d'états mentaux introspectibles traditionnellement répertoriés uniquement sous la forme des attitudes propositionnelles. Plus spécifiquement, Beaton propose de définir les *qualia* comme des propriétés introspectibles qui ne peuvent pas, justement, être pleinement spécifiées en termes d'attitudes propositionnelles classiques. Autrement dit, il existe des propriétés introspectibles qui sont *subjectives* dans un sens plus *radical*. Beaton identifie trois possibles voies définitionnelles des faits subjectifs :

(a) Faits subjectifs *qua* faits potentiellement constituants du sujet ;

(b) Faits subjectifs *qua* faits introspectibles ;

(c) Faits subjectifs *qua* faits allant au-delà des faits introspectibles standard.

[135] M. Beaton, (2009), '*Qualia* and Introspection', *Journal of Consciousness Studies* 16 (5):88-110.

Il serait dès lors envisageable de considérer ce que Beaton appelle une « implication conceptuelle à double sens » (« two way conceptual entailment ») entre les *qualia* d'une part, et l'ensemble approprié des concepts mentaux non-standard, ou hétérodoxes, comme par exemple les concepts et les faits motivationnels, associatifs, affectifs, etc.

L'objectif serait alors de trouver des propriétés introspectibles considérées sous le mode motivationnel, affectif, etc. L'approche de Beaton est intéressante dans la mesure où elle remet en cause l'hégémonie du modèle propositionnel dans l'analyse des propriétés mentales.

Un double chantier serait ainsi à amorcer :

(i) Sur le versant du physicalisme, il s'agit de redéfinir et d'assouplir la notion de « physique » ou de « matière » afin de permettre à la philosophie de renouer un véritable dialogue avec la physique fondamentale, quantique.

(ii) Sur le versant des *qualia*, il s'agit de permettre au fleuve de la phénoménologie d'irriguer le vaste domaine des sciences naturelles. Une collaboration très prometteuse est dès lors envisageable. Il s'agit là d'une piste de réflexion que l'auteur de ces lignes se propose de suivre et de développer dans ses futures recherches.

L'objectif sera alors de réinterpréter le problème de la « localisation » de l'esprit en général, et des *qualia* en particulier dans le monde naturel, afin de remplacer la question traditionnelle :

(a) Quelle est la place de l'esprit dans la nature ?

par la question :

(b) Les difficultés quasi-insurmontables que l'on rencontre dans le problème de la « localisation » ne seraient-elles imputables plutôt au fait que le mental « traîne » derrière lui une réputation post-platonicienne d'immatérialité surnaturelle ? En effet, il serait une erreur d'oublier le fait que l'esprit a été spontanément associé à un royaume séparé, voire même opposé à celui de la matérialité.

Récapitulons. Nous avons vu que le physicalisme en général, et le physicalisme *a priori* plus particulièrement, peut être menacé par la critique de ce que l'on peut appeler les « deux **dogmes du physicalisme** » :

(i) d'abord l'idée que le réel serait en quelque sorte construit à partir de « briques » primitives, des minuscules constituants du réel supportant sur leurs frêles épaules, comme le mythique Atlas, l'univers tout entier.

(ii) ensuite, il y a l'idée que le réel serait organisé par couches, un peu comme un mille-feuille pâtissier, allant du niveau inférieur physique vers le niveau supérieur, mental.

Or, si l'on y regarde de plus près, on s'aperçoit que ces deux « dogmes » sont tributaires d'une distribution sur l'axe micro- et macro-réel. Or la question est : à partir de quel *point de vue* peut-on décider qu'une entité est plutôt « micro » ou plutôt « macro-réelle » ? Pour paraphraser Protagoras : un éléphant ne perçoit pas le monde de la même façon qu'une bactérie. Mais si les éléphants et les bactéries se mettent à philosopher, lesquels d'entre eux auront plus de chances de s'approcher d'une « correcte » description de la dualité micro/macro-réel ?

Conclusion

8 Un argument pour concilier Mary avec le physicalisme

> « *Voyez-vous, messieurs, la raison est une chose excellente ; ceci est incontestable ; mais la raison est la raison et ne satisfait que la faculté raisonnante de l'homme, tandis que le désir est l'expression de la totalité de la vie. (...) Que sait la raison ? La raison ne sait que ce qu'elle a appris (...) tandis que la nature humaine agit de tout son poids, pour ainsi dire, avec tout ce qu'elle contient en elle, sciemment et inconsciemment* ».
>
> F.M. Dostoïevski, *Le Sous-sol*.

Il est temps à présent de livrer notre point de vue sur l'affaire du duel qui nous concerne ici. Reprenons *da capo*, au risque de lasser le lecteur, l'argument de la connaissance Jackson (1982/1986).

KA

(P1) Mary (avant sa libération) connaît toute l'information physique concernant les objets et les autres sujets.

(P2) Mary (avant sa libération) ne connaît pas toute l'information physique concernant les objets et les autres sujets (car elle *apprend* quelque chose de nouveau lors de sa libération, à sa sortie de prison).

Par conséquent,

(P3) Il existe des propositions vraies (des vérités) à propos des objets et des autres sujets qui échappent au récit physicaliste.

Commençons avec la première prémisse (P1) et demandons-nous : est-il réellement possible que Mary (avant sa libération) connaisse *toutes* les propositions et informations physiques relatives à la perception visuelle du rouge ? A cela nous répondons, suivant Crane, Van Gulick et Nagasawa, par « non ». Autrement dit, si l'on reprend l'arbre synoptique[136] des réponses apportées au cas 'Mary', nous nous rangeons du côté de ceux qui répondent par « non » à la question (Q0). Cela ne veut pas dire que

[136] Cf. section 6.6

nous considérons les enjeux contenus dans les questions et les réponses allant de (Q1) à (Q5) comme étant sans pertinence pour l'affaire des rapports entre le physicalisme et les *qualia*. Mais nous pensons que le nœud gordien de ce duel peut être coupé à la racine et dénoncé comme un effet secondaire du clivage archétypal décliné sur l'axe subjectivité/objectivité.

Maintenant il est vrai que les motivations derrière ce « non » ne sont pas les mêmes chez les auteurs que nous venons de citer. Néanmoins, nous soutenons ici que tous ces arguments convergent et mettent en relief de manière surprenante la survie du clivage sur l'axe subjectif/objectif. Nous allons passer à présent en revue les trois raisons avancées en faveur de la réponse « non » à la question (Q0).

(i) La première raison est qu'il existe des *faits subjectifs*, et que ceux-ci *ne sont pas incompatibles* avec le physicalisme. (l'approche de Crane et Van Gulick)

(ii) La deuxième raison est que la notion d'*omniscience* employée pour caractériser le statut cognitif de Mary est trop forte. (l'approche de Nagasawa)

(iii) La troisième raison, qui est la nôtre, est que la notion d'omniscience n'accepte pas de degrés, et donc elle ne peut pas être appliquée *isolément* à un seul fragment de savoir (i.e. la perception visuelle des couleurs), *sans* prendre en compte les *autres* mécanismes impliqués dans ce savoir. Autrement dit, si l'on part de l'hypothèse fondamentale de l'intrication quantique, il s'ensuit qu'il est impossible de posséder 'omnisciemment' un seul fragment du savoir sur le monde, *sans* posséder 'omnisciemment' la *totalité* du savoir sur l'univers. Ce qui revient à dire que si Mary avait été réellement omnisciente par rapport à la *totalité* des faits physiques impliqués dans la perception visuelle du rouge, elle aurait été, par ricochet, omnisciente par rapport à la totalité des faits de l'univers tout entier. A partir de là, on proposera une piste de recherche pour l'avenir visant à développer une conception cohérente des relations mental/physique étayées non pas selon le modèle d'un clivage épistémique isolateur, mais plutôt selon un modèle holistique de l'intrication quantique.

8.1 Retour au clivage Subjectif/Objectif

Dans l'article 'Subjective Facts', T. Crane[137] défend une thèse à laquelle nous souscrivons entièrement, à savoir l'idée que le KA est un argument valide qui réfute non pas le physicalisme, mais seulement la thèse selon laquelle tous les faits physiques seraient des faits *objectifs*. Autrement dit, selon Crane, et nous le suivons sur ce point, le KA prouve l'existence des faits *subjectifs*. Mais ce serait une erreur de rejeter le physicalisme sur la base de l'existence de ce type spécial de faits. Crane s'inscrit ainsi en faux contre la position, défendue entre autres, par D.H. Mellor[138], selon laquelle la latitude subjective doit être tenue à sa place, à savoir à l'intérieur de la représentation que le sujet se forge du monde. Nous voyons donc qu'une fois les masques tombés, le duel physicalisme/*qualia* se révèle en réalité être un duel entre les pôles objectivité/subjectivité déclinés cette fois-ci sur l'axe faits subjectifs/faits objectifs. Maintenant les questions que l'on se pose sont : qu'est-ce qu'un « fait subjectif » et qu'est-ce qu'un « fait objectif » ?

Commençons par ce dernier. Un fait objectif est d'abord un fait physique. Nous avons vu, dans le chapitre 1, que la question de définir un *objet* ou une *propriété* physique est loin d'être close. Aurons-nous plus de chance avec la définition du terme *fait physique* ? Et qu'est-ce qu'un *fait* plus exactement ? Là non plus il n'y a pas de consensus universellement établi. D'aucuns disent que les faits *sont* des propositions vraies, d'autres affirment qu'ils *correspondent* de manière biunivoque aux propositions vraies, tandis que d'autres encore soutiennent qu'ils sont des « vérifacteurs » (« truth-makers »), c'est-à-dire des facteurs qui *rendent* les propositions vraies.

A présent, la question que l'on doit se poser est : quelle est la notion de « fait » véhiculée par le KA ? Comme le note Crane, il est évident que l'argument se base sur l'idée que les faits sont des objets visés par la connaissance de type propositionnel de genre « *X* sait que *p* », où *X* est le sujet connaissant et *p* est remplacé par un énoncé. Tout le monde semble s'accorder sur le fait qu'il existe de tels objets visés par la connaissance propositionnelle, indépendamment de la question si l'on doit accepter ou non d'*autres* interprétations du terme « fait ». Il faut donc s'en tenir à l'acception habituelle selon laquelle tout fragment du savoir propositionnel est une connaissance d'un fait distinct.

---

[137] T. Crane (2003), 'Subjective Facts', in *Real Metaphysics: Essays in Honour of D. H. Mellor*. New York: Routledge.
[138] D.H. Mellor, (1992/1993), 'Nothing Like Experience', *Proceedings of the Aristotelian Society* 93: 1-16.

Plusieurs auteurs cependant préfèrent opérer une distinction entre fait ou information comprise *sous un mode large* et *sous un mode étroit*. Selon l'acception étroite, les faits sont individués au niveau des sens. Selon l'acception large, les faits sont individués au niveau de la référence. Dans l'ouvrage "*The Facts of Causation*", Mellor[139] propose de distinguer entre faits et *facta*. Les faits sont les « ombres fidèles » des vérités: s'il est vrai que *p*, alors c'est un fait que *p*. En revanche, les *facta* sont les « vérifacteurs » des vérités. C'est une question purement empirique d'établir quelles sont les *facta*, ou quelles sont les propriétés. Conséquemment, on ne peut pas inférer, à partir de différenciations des faits, des différenciations dans les *facta*, car entre les faits et les *facta* il n'y a pas de relation de correspondance biunivoque.

Donc, la clé du KA n'est pas d'établir si le savoir complet que Mary détient dans sa prison est un savoir physique, mais plutôt de saisir si le savoir de Mary peut être appris et énoncé d'une façon particulière, i.e. par le truchement des livres et textes. A partir de là, note Crane, on peut montrer que malgré le fait que la cible initiale du KA est le physicalisme – défini comme l'acception selon laquelle tous les faits sont des faits physiques – la véritable « victime » que le KA exécute est la conception selon laquelle le savoir complet sur le monde serait un savoir qui peut être propagé et distribué *via* des textes et des leçons, sans nécessiter des expériences spécifiques. Par conséquent, comme on l'a très bien vu dans le cas de l'argument de la parité de raisons de Churchland, *toute conception* basée sur cette acception implicite d'arrière-plan tombe sous le couperet du KA, et pas seulement le physicalisme.

D'autre part, il est vrai que l'idéal d'objectivité scientifique se voit ainsi limité par le KA. Car aucune conception scientifique de la vision ne serait en mesure d'apprendre à un homme aveugle l'effet-que-cela-fait de voir des couleurs. Mary apprend donc bel et bien un fait propositionnel authentiquement nouveau. Ce type de savoir propositionnel est un type de savoir que les sciences ne peuvent pas exprimer théoriquement. Mais le point crucial, note Crane, est que personne ne devrait s'attendre à ce que les sciences théoriques offrent ce type de savoir. Par conséquent, il n'y a aucune «mystérieuse » restriction aux pouvoirs explicatifs de la science. Si on réfléchit à ce problème, on s'aperçoit que le KA n'est pas une objection au physicalisme, mais plutôt à la thèse selon laquelle tous les faits seraient des faits qu'on peut apprendre *via* des livres.

[139] D.H. Mellor, (1995), *The Facts of Causation*. London: Routledge.

Bref, le KA réfute la thèse selon laquelle tous les faits seraient des faits *objectifs*, autrement dit, des faits dont l'apprentissage ne nécessite ni d'éprouver une expérience particulière, ni d'adopter une position particulière dans le monde. A l'opposé, les *faits subjectifs* sont les faits dont l'apprentissage nécessite justement d'éprouver une certaine expérience et d'endosser une certaine posture ontologique dans le monde. A la lumière de ces considérations, il est clair que le KA plaide en faveur de l'existence des faits subjectifs. Supposons à présent que Mellor objecte à cette conclusion en faisant entrer en scène sa distinction entre faits et *facta*. Il pourrait admettre qu'il existe des faits subjectifs i.e. des vérités subjectives, dans le sens où les objets de la connaissance sont individués non pas par la référence, mais seulement par les sens. Mais il pourrait nier l'existence des *facta* subjectives, ou des « vérifacteurs » subjectifs. Mais qu'est-ce qu'un *factum* subjectif ? A la différence des faits subjectifs, qui sont des faits dont l'apprentissage présuppose l'endossement d'une certaine expérience et posture ontologique, les *facta* ne peuvent pas être *apprises*, car elles *rendent vraies* les vérités qui, elles, sont apprises. On pourrait alors avancer l'hypothèse selon laquelle un *factum* subjectif est le « vérifacteur » des faits ou vérités subjectives. Ou : le *factum* subjectif est la condition d'existence qui permet aux faits subjectifs d'être appris. Revenant à présent au cas de Mary : quelle est la condition nécessaire pour que Mary apprenne que voir du rouge c'est *cela*. On est tenté de répondre immédiatement que la condition respective c'est l'*expérience* visuelle du rouge. Mais si le *factum* subjectif respectif est une expérience, alors personne ne devrait nier l'existence des *facta* subjectives. Car l'enjeu n'est pas d'établir l'existence des *expériences*. Les expériences sont subjectives dans le sens où elles dépendent de l'existence des sujets expérientiels. Or Mellor ne compte nier ni l'existence de ces derniers, ni l'existence de Mary. Par conséquent, qu'est-ce que rejette Mellor lorsqu'il rejette l'existence des *facta* subjectives ? Car la distinction objectif/subjectif vise certains genres de connaissance. Il est difficile de les appliquer directement aux entités. Doit-on stipuler que le *factum* qui constitue le "vérifacteur" de la connaissance de Mary « voir du rouge c'est *cela* » est l'expérience que Mary éprouve à la vue d'une tomate mûre ? On peut dès lors considérer que ce genre d'expérience peut être appelé une « entité subjective », dans le sens où il s'agit d'une entité qui est dépendante du sujet d'une expérience. Or, dans ce cas, l'expérience devrait être appelée un « *factum* subjectif ». Donc Mellor se voit contraint d'accepter à la fois l'existence des faits subjectifs et des *facta* subjectives.

Récapitulons. Le KA attaque le physicalisme conçu comme la conception selon laquelle *tous les faits sont physiques*. D'une part, lorsque l'on s'intéresse de plus près à la signification du terme « fait », on s'aperçoit qu'elle présuppose l'arrière-plan selon lequel tout le savoir propositionnel est physique. D'autre part, lorsque l'on s'intéresse plus près à la signification du terme « physique », on constate que le savoir en cause est celui postulé par l'expérience de pensée de la chambre de Jackson, i.e. un savoir obtenu par le truchement des livres, diagrammes et écrans de téléviseur. Par conséquent, la véritable cible du KA est la thèse selon laquelle *tous les faits sont objectifs* (dans le sens défini plus haut). Maintenant, la question que l'on doit se poser est la suivante : le physicalisme est-il réellement tenu d'endosser cette dernière thèse ? Selon Crane, le physicalisme est avant tout une doctrine ontologique dont l'objectif crucial est celui de prouver que le monde est causalement clos[140].

Par conséquent, le physicalisme n'est pas tenu d'adhérer à la thèse selon laquelle la science physique doit *énoncer tous les faits*. Tout ce dont il a besoin c'est d'admettre qu'il existe un genre de connaissance qui peut être accédée uniquement à partir du point de vue subjectif, à la première personne. Et l'existence de ce type de connaissance ne devrait pas inquiéter les physicalistes, pas plus que l'existence des sujets connaissants n'inquiète les scientifiques.

L'existence des faits subjectifs a été défendue entre autres, comme nous l'avons vu dans le chapitre 6, par Van Gulick[141]. Durant sa captivité Mary connaît uniquement les faits et les informations *à la troisième personne*. Or, si l'on est partisan du physicalisme non-réductionniste, et si l'on privilégie, comme l'auteur de ces lignes, un point de vue pragmatique[142] dans la conception de la cognition, alors, on sera prêts à admettre qu'il existe des faits réalisables à partir d'un substrat physique, mais qui demeurent fondamentalement *inanalysables* du point de vue d'une *théorie* physique. Ce qui revient à dire que si l'esprit conscient implique la présence d'une composante essentielle d'auto-compréhension réflexive, il s'ensuit alors que le phénomène de la conscience est caractérisé d'un ensemble de propriétés inscrutables à partir d'un point de vue externe, en surplomb. Conséquemment, nous sommes d'accord avec Crane et

[140] Voir l'argument causal en faveur du physicalisme résumé par Papineau (section 1.2.2. du présent texte)
[141] (R. Van Gulick, 2008:197)
[142] Pour une discussion détaillée de ce type d'approche pragmatique de l'expérience voir N. Depraz F. Varela & P. Vermersch, (2003), *On Becoming Aware: A Pragmatics of Experiencing*, John Benjamins.

Van Gulick[143] lorsque ceux-ci affirment que l'existence des faits subjectifs *est* compatible avec le physicalisme.

8.2 Divine Mary

Tournons-nous à présent vers les motivations derrière la réponse « non » de Nagasawa à la question Q0. Selon Nagasawa, Mary, durant sa captivité dans la chambre de Jackson, connaît uniquement les faits et les informations *théoriquement communicables*, i.e., les faits à la troisième personne. Dans ce qui suit, nous allons présenter un argument qui non seulement remet en cause la première prémisse (P1) de l'argument de Jackson, mais surtout prouve que poussé jusqu'au bout de ses conséquences, le raisonnement de Nagasawa finit par récuser sa propre adhésion au physicalisme nonthéorique. Pour ce faire, nous avons besoin de reprendre quelques éléments de l'analyse de Nagasawa concernant la notion d'omniscience. Rappelons-nous brièvement la trame de son argumentation, afin que l'on puisse plus facilement y dériver la nôtre. Selon Nagasawa[144], si l'on compare l'argument de la connaissance de Jackson (1982/1986) d'une part, et l'argument de la possession du concept d'autre part en termes d'omniscience, il devient clair que l'argument tiré de la philosophie de la religion ne repose pas sur une interprétation controversée de la notion d' « omniscience divine ». Autrement dit, l'argument de la possession du concept ne se rend pas coupable d'utilisation abusive de la notion clé d' « omniscience divine »[145].

En revanche, l'argument de la connaissance viole les frontières légitimes de l'application du concept d' « omniscience physique »[146], et se rend coupable, par ricochet, d'interprétation abusive du concept de « physique », c'est-à-dire qu'il repose sur l'hypothèse hautement discutable selon laquelle *l'information théorique communicable* serait la seule forme d'information physique *simpliciter*. A l'appui de cet argument, Nagasawa fait appel au concept lewisien[147] d' « humilité ramseyenne » pour arguer que Mary ne peut pas connaître les propriétés intrinsèques non-

[143] R. Van Gulick, (1985), 'Physicalism and the subjectivity of the mental', *Philosophical Topics*, 13: 51-70.
[144] (Y. Nagasawa, 2008)
[145] Pour mémoire : (O_{DS}) Omniscience Divine *Simpliciter* : L'omniscience divine est omniscience *simpliciter*.
[146] Pour mémoire :(O_{PS}) Omniscience Physique *Simpliciter* : L'omniscience physique *est* omniscience *simpliciter*
[147] (D. Lewis, 2001/2009)

relationnelles, fondamentales du réel physique uniquement par le biais des livres et des écrans de téléviseurs en noir et blanc. D'autre part, si l'on est prêt à adhérer à la thèse forte selon laquelle Mary est effectivement physiquement omnisciente dans sa prison, il devient relativement facile de montrer que nous exagérons la portée des capacités cognitives de Mary. En clair, Nagasawa montre que si l'on utilise abusivement le concept d' « omniscience », l'argument de Jackson (1982/1986) échoue, car il est clair que Mary n'a pas les capacités cognitives d'un démiurge. La faillite du KA est censée, selon Nagasawa, soutenir la première prémisse de son physicalisme nonthéorique, à savoir : (i) L'omniscience physique *est* omniscience *simpliciter*. (La thèse de l'omniscience physique est vraie).

A présent, nous proposons de considérer l'argument suivant. Concédons à Nagasawa que Mary n'est pas un démiurge. A partir de là, il est possible d'élargir cette hypothèse de la façon suivante (A1) :

(A1) Ni Mary, ni Barry, ni Nagasawa, etc., ne sont pas des créatures de type-démiurgique.

Si Mary n'est pas une créature démiurgique, il est clair que le concept d'omniscience divine ne la définit pas. La question qui se pose, évidemment, est celle de savoir si c'est le concept d'omniscience physique qui la définit, car c'est de cet aspect que dépend la validité de la prémisse (i) du physicalisme nonthéorique. Nagasawa repose son argument anti-KA sur l'observation que les défenseurs du KA utilisent de façon abusive la putative omniscience physique de Mary pour récuser le physicalisme. Or, dit-il, cette accusation n'est pas valable, puisque Mary n'est pas « omnisciente », divinement ou physiquement. Mais dans ce cas, puisque ni Mary ni Nagasawa ne sont pas Dieu, par quel moyen pourrait-on prouver (i), à savoir que « L'omniscience physique *est* omniscience *simpliciter* » ? Comment pourrait-on accéder à des caractérisations de l'omniscience en général ? On peut dès lors avancer un pas de plus et proposer la deuxième prémisse de notre argument (A2) :

(A2) L'omniscience *simpliciter* (qu'elle soit physique ou divine) est une propriété exclusivement divine.

Maintenant, notre argument prend la forme du syllogisme suivant :

HUMAIN
VS
DIVIN

(A1) Ni Mary, ni Barry, ni Nagasawa, etc., ne sont pas une créature démiurgique.

(A2) L'omniscience *simpliciter* (qu'elle soit physique ou divine) est une propriété exclusivement divine.

(A3) Ni Mary, ni Barry, ni Nagasawa, etc., ne peuvent posséder la propriété de l'omniscience *simpliciter*.

Conjointement, les thèses (A2) et (A3) récusent la prémisse de l'omniscience physique conçue en tant qu'omniscience simpliciter (thèse (i) du physicalisme nonthéorique de Nagasawa), car elles refusent d'accorder au théoricien Nagasawa le même privilège qu'il conteste à la prisonnière Mary, à savoir : le statut de Super-Sujet statuant sur la signification du concept d'omniscience *simpliciter*.

A la lumière de ces considérations, on peut soulever, naïvement, l'interrogation suivante : étant donné que notre intérêt est d'examiner le destin de Mary (créature de type-humain), et non pas le destin d'une putative créature divine, pourquoi perdons-nous son temps avec des questions sur l'omniscience, puisque de toute façon il s'agit là d'un sujet qui dépasse les modestes limites de la rationalité (intelligibilité) humaine ? Après tout, sait-on vraiment ce qu'un concept comme « omniscience » ou « dieu » veulent effectivement dire ? Précisons tout de suite que la prémisse (A3) ne comporte pas un message pessimiste. Au contraire, elle nous informe sur un aspect essentiel de la connaissance : à savoir, le fait que le savoir est toujours un savoir déjà *dans* le monde.

Revenons à présent sur l'expérience de pensée de Jackson. Mary est captive dans un environnement noir et blanc et apprend toute l'information physique concernant la perception visuelle des couleurs. Maintenant, nous aimerions soulever la question suivante :

DIVINE MARY

(*Q*) Peut-on être réellement omniscient isolément et uniquement par rapport à un *fragment* de savoir – i.e. *que* par rapport à la perception visuelle du rouge – *sans* être, par ricochet, omniscient par rapport aux *autres* formes impliquées dans le savoir complet du rouge: par exemple, la connaissance des formes, des structures, de l'espace, des textures, des objets ? Pour dire les choses clairement, notre réponse est la suivante : étant donné la thèse de l'intrication, si Mary avait été réellement omnisciente par rapport à la perception des couleurs, elle aurait été omnisciente par rapport à *tout le reste*, les intentions de ses geôliers et l'univers y compris, car la notion d' « omniscience » n'admet pas de degrés ou de découpages sémantiques parcellaires. Bref, si Mary avait été omnisciente par rapport à la couleur rouge, elle aurait été omnisciente par rapport à *tout* l'univers. Mary aurait été Dieu.

Le défi que nous nous proposons de relever dans nos recherches futures est de montrer que la relation entre les faits subjectifs et les faits objectifs ne doit pas être examinée selon le modèle d'un perpétuel conflit entre les vérités immuables, fortunées et les vérités subjectives infortunées, qui doivent être à tout prix éliminées. Une analyse attentive émanant du sein des recherches fondamentales de la mécanique quantique pourrait nous indiquer la voie à suivre. Un *modèle intriqué* de la relation subjectif/objectif devra être envisagé alors afin de rendre justice au fait naturel que le sujet connaissant se trouve déjà dans le monde.

Dans cet essai, jusqu'ici nous avons essayé de signaler les mécanismes qui rendent la subjectivité « suspecte » aux yeux non seulement des physicalistes, mais aussi aux yeux de leurs prédécesseurs, les empiristes logiques. Ainsi est-il possible de saisir les dommages collatéraux infligés par l'adoption implicite de ce que nous avons appelé le « Big Bang Gouffre » séparant la pensée du monde. Indésirable, la subjectivité inhérente de l'individu a toujours été reléguée en second plan et traitée en quelque sorte comme le « Cendrillon » de la vie mentale. Ce n'est pas étonnant, dès lors, que l'on peine à lui « trouver » une place dans l'ameublement physique, i.e., objectif et respectable du monde. Une fois que l'on rejette ce schisme archétypal qui

provoque par ailleurs le tandem fossés explicatifs/ sauts conceptuels, on s'aperçoit que ce que l'on appelle actuellement le problème de la « localisation » de l'esprit dans le monde naturel est un pseudo-problème. Le clivage phénoménal/physique, avec toutes les conséquences dramatiques qu'il entraîne, n'est en fin du compte qu'une autre variation sur le thème du clivage subjectif/objectif. Si l'on rejette ce modèle et si l'on opte pour un *modèle intriqué* des rapports subjectif/objectifs, il se peut que les tensions philosophiques situées à la mezzanine hybride abritant les sauts conceptuels s'évaporent d'elles-mêmes. Cela ne veut pas dire, bien entendu, que d'autres tensions ne seront pas alors à prendre en considération ou à résoudre. Au contraire, il faudra se retrousser les manches et entamer un véritable travail de réédification des schémas conceptuels ancestraux de l'Occident philosophique. Tout ce que nous affirmons est que si l'on abandonne le clivage subjectif/objectif et si l'on épouse le modèle intriqué, alors toute la famille de problèmes que S. Blackburn[148] a identifiée sous le label de problèmes de médiation entre les vérités fortunées et les vérités infortunées (« happy truths *versus* unhappy truths ») sera dissoute.

Une partie considérable de cette thèse que nous venons d'énoncer repose sur l'interprétation du terme clé « intrication ». Par conséquent, avant de tirer notre conclusion finale, nous allons nous arrêter très brièvement sur cette notion, pour donner un exemple concret de ce que nous entendons par modèle intriqué de la relation subjectif/objectif, en prenant comme exemple d'étude le cas réel d'une personne souffrant de cécité chromatique.

8.3 La nuit, tous les chats sont gris

Le cas de Mary est une invention de F. Jackson. Mais il existe des cas réels de personnes souffrant véritablement de cécité chromatique. Une de ces personnes est Knut Nordby, né en 1942, à Oslo, en Norvège, est devenu, à force de travail acharné, un excellent scientifique spécialisé en vision des couleurs. Nordby[149] livre un récit à la fois très précis, passionnant, et très instructif de ses difficultés à apprivoiser et assumer son quotidien atypique. A partir de là, il est très tentant de comparer son cas avec celui

[148] (S. Blackburn, 2009)
[149]K. Nordby, 'Vision in a Complete Achromat: A Personal Account' http://consc.net/misc/achromat.html

de la prisonnière de Jackson. Il y a cependant, note Nordby[150], trois dissemblances importantes entre son expérience et celle de Mary : (i) d'abord, lui-même est né sans les cônes rétinals, (ii) il a été exposé aux couleurs en permanence, et (iii) malheureusement, il ne sera pas libéré de sa prison à son dix-huitième anniversaire, comme dans les contes de fées.

Les individus nés souffrant de cécité chromatique voient les objets uniquement en fonction de contrastes de luminosité et vivent dans un perpétuel état de vision nocturne. Comme le dit le proverbe : la nuit, tous les chats sont gris. Ce qui veut dire qu'à des niveaux très bas de luminosité, pendant la nuit, par exemple, les personnes dotées d'une vision chromatique normale (les trichromates) subissent une perte d'acuité chromatique. Bien que les cas d'achromatisme soient très rares, il y a des cas encore plus rares, des formes corticales d'achromatisme, des sujets qui, à la suite d'un accident[151], ont perdu leur capacité à distinguer les couleurs.

Nordby[152] livre un point de vue surprenant sur l'expérience de pensée de Jackson : selon Nordby, tout dépend de la *luminosité de l'environnement* à laquelle est exposée Mary dans sa prison. A la différence des personnes souffrant d'achromatisme, et qui savent associer aisément telle nuance de gris à telle nuance en fonction de leur luminosité, l'héroïne de Jackson aurait été dans l'impossibilité de faire de même, c'est-à-dire d'associer le rouge à une nuance précise de gris, peu importe la puissance de la luminosité. En revanche, si Mary *avait grandi* dans un environnement noir et blanc placé sous une faible luminosité, alors cette situation aurait endommagé ses cellules

[150] "People who are totally total color-blind or achromatopic have visual input only from retinal rods and see only in contrasts of brightness. They thus live in a perpetual state of "night vision" (Sharpe and Nordby 1990a, 1990b). Typical achromatopsia (or rod monochromacy) is congenital and nonprogressive. It is characterized by: (i) Total lack of color discrimination (though people with this condition can match any color hue to any other color hue or shade of gray on the basis of brightness); (ii) Hypersensitivity to bright light (called "photophobia"); (iii) Low visual acuity (typically 6/60 [or 20/200] Snellen, which means that people with this condition can resolve fine detail at 6 meters [20 feet] that people with normal acuity can resolve at 60 meters [200 feet]); (iv) Nystagmus (involuntary, rapid side-to-side movement of the eyes; this diminishes with age)". Voir K. Nordby, (2007), 'What is this thing you call color? Can a totally color blind person know about color?' in T. Alter & S. Walter, (eds.), op. cit., p.78.

[151] O. Sacks et R. Wasserman rapportent l'histoire du peintre J. Isacson ayant perdu la faculté de distinguer les couleurs à la suite d'un accident de voiture. Quelques mois après l'accident, le malade s'est vu même dans l'incapacité de se rappeler à quoi ressemblaient les différentes couleurs. Pour être plus précis, le monde de Mary la prisonnière n'est pas en noir et blanc, mais en noir, blanc, et surtout en nuances infiniment subtiles de gris. Conséquemment, le degré de luminosité de l'environnement joue un rôle essentiel dans l'expérience. Voir O. Sacks & R. Wasserman, (1987), 'The Case of the Colorblind Painter', in *The New York Review of Books*, 34, 193-226.

[152] "A totally color-blind person will happily match any color hue to any other color hue or gray tone by adjusting the brightness, but Mary would not be able to do this; she will not be able to match green with a red color or gray tone, no matter what brightness they have." (K. Nordby, 2007:79)

responsables de la vision chromatique, et elle aurait probablement souffert d'une de ces maladies communes chez les mineurs qui ont travaillé depuis leur jeune âge dans un milieu obscur, dans le noir presque complet. Durant le XIXe siècle il n'était pas inhabituel d'envoyer des enfants âgés de six ans travailler dans les mines, pour accomplir de menues tâches. Ils passaient ainsi plus de 14 heures par jour dans le noir. Si l'on tient compte du fait que la plupart du temps il faisait noir quand ils allaient travailler ainsi qu'à leur sortie du service, on peut en déduire que cette situation a fini par endommager leur rétine, et provoquer des maladies des yeux comme le nystagmus[153] ou la photophobie. Par conséquent, si Mary avait été tenue captive dans un milieu similaire, elle aurait eu des difficultés considérables pour percevoir les couleurs dès la sortie de sa prison. Cela prouve, une fois de plus, que les conditions environnementales ne peuvent pas êtres éludées dans l'examen d'un cas comme celui-ci. Il est difficile, voire impossible, de découper le personnage 'Mary' sur une toile d'arrière-plan neutre, car cet arrière-plan dans lequel notre héroïne baigne joue un rôle essentiel dans ses mésaventures. La situation épistémique de Mary est fondamentalement intriquée dans sa situation ontologique.

D'autre part, l'étude des cas comme celui du peintre Isacson montre bien que le cerveau « oublie » les fonctions exécutées par les parties endommagées à la suite d'un accident. Il existe cependant des cas exceptionnels de personnes qui ont réussi à construire, à l'aide d'un imaginaire interne, des fonctions auxiliaires remplaçant les anciens « outils » endommagés. L'exemple le plus célèbre, reste bien évidemment celui de Beethoven, qui a composé une bonne partie de ses créations après qu'il soit devenu complètement sourd. Il se peut, qui sait, qu'il y ait dans le monde, des anonymes souffrant d'achromatisme, qui ont réussi à se forger une imagerie interne privée, à l'aide de laquelle ils peuvent « saisir » les couleurs à leur façon. Quoi qu'il en soit – insiste Nordby – cette imagerie interne, aussi sophistiquée soit-elle, n'empêchera pas l'individu souffrant de cécité chromatique d'acheter des tomates non-mûres, par exemple. La morale de l'histoire est qu'en aucun cas Mary n'aurait pu, à la milliseconde même où elle aurait mis son pied dehors, dans un champ plein de coquelicots, s'écrier : « A ! C'est *cela* l'effet de voir du rouge ! » Tout simplement parce que son exposition à un environnement noir et blanc aurait endommagé (qui sait,

[153] Le nystagmus est un mouvement d'oscillation involontaire et saccadé du globe oculaire. Les oscillations peuvent se produire sur un plan vertical, horizontal, de torsion ou dans une combinaison de ceux-ci.

peut être irrémédiablement) les cellules de son cerveau. Tout un processus de reconstruction chromatique serait alors à prendre en compte. Mary aurait fini par *faire corps*, si l'on peut dire ainsi, avec sa « cécité » chromatique imposée, et elle aurait fini par ne voir que le blanc, le noir, et d'infinies nuances de gris.

8.4 Le mental, le physique et la lampe d'Aladin

Nous avons commencé cet essai avec une citation de. T. Huxley[154]. Notre périple peut être résumé à l'aide de l'allégorie suivante : Aladin, le philosophe, en voulant nettoyer et purifier la lampe du physicalisme, (afin de la rendre aussi objective que possible) se retrouve avec Djinn-l'esprit sur les bras. Irrité par l'apparition importune de cet invité surprise, il tente mille et une solutions pour l'éliminer. En vain. Djinn tient bon et ne disparaît pas. Il est bien là, et il faut composer une explication du réel avec.

Quelle conclusion pour la fin ? Nous avons plaidé pour le rejet de la première prémisse de l'argument de la connaissance. Cela veut-il dire que nous nous rangeons du côté des physicalistes ?

D'abord, il serait inutile de nier l'évidence. Les *qualia* font bel et bien partie de l'ameublement du monde. Mais le physicalisme nous dit que cet ameublement est physique. Doit-on conclure que les *qualia* sont physiques ? A cela nous répondons : tout dépend de ce qu'on entend par « physique ». Maintenant, il est vrai que l'on voit déjà se profiler une première objection : « votre réponse ne fait que nous renvoyer à la case départ avec laquelle cet essai a commencé, à savoir la définition du terme « physique ». Ce qui veut dire que tout cet itinéraire de quatre cent pages n'a servi strictement à rien ». A cela notre réplique serait : si par « physique » on entend les traits que le portrait robot physicaliste subsume, alors non, les *qualia* ne sont pas physiques *dans ce sens là*. Mais il est possible également de se tourner vers les sciences de la matière et de leur poser la même question : qu'est-ce « physique » pour vous ? Dans ce cas, nous serions étonnés de constater à quel point la « phusis », la nature se montre vertigineusement complexe.

[154] Pour mémoire, voici la citation: "*We class sensations along with emotions, and volitions, and thoughts, under the common head of states of consciousness. But what consciousness is, we know not; and how it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of the Djin when Aladdin rubbed his lamp, or any other ultimate fact of nature.*"

Le véritable problème ne réside pas dans le fait que la présupposée immatérialité subjective des *qualia* s'accorde mal avec la matérialité objective du monde naturel, physique, mais dans l'hypothèse même qu'une telle séparation aurait un sens. Le problème n'est pas de trouver une localisation à l'esprit dans le monde matériel, mais de justifier l'intelligibilité même d'une telle distinction entre la phénoménalité et la 'physicalité', suivant un clivage conceptuel strict et rigide, et qui est loin de refléter une réalité où dominent moins des entités figées sur une toile d'arrière-plan immobile, que des systèmes, des structures et des relations en perpétuelle interaction.

Par conséquent, le nœud du problème ne consiste pas à résoudre le différend entre les physicalistes et les avocats des *qualia*, mais plutôt à trouver la réponse à la question : pourquoi en est-on arrivé là, c'est-à-dire à cet affrontement face-à-face ?

Si notre hypothèse de départ est correcte, il s'ensuit que la raison même pour laquelle le physicalisme s'interdit le droit d'inclure les faits subjectifs dans son magma explicatif, d'une part, et la raison pour laquelle on hésite à traiter les *qualia* comme « physiques », d'autre part, doivent être cherchées du côté du schisme archétypal qui divise la pensée et le monde sur l'axe subjectivité (doxa) / objectivité (vérités).

Le physicalisme contemporain se retrouve dans la situation paradoxale d'une doctrine empirique vouant allégeance à une thèse métaphysique post-platonicienne qui lui interdit d'accueillir l'esprit en son sein. Après tout, pourquoi ne pas accepter le fait que le physicalisme serait compatible avec l'existence des faits subjectifs, comme le suggère Crane et Van Gulick ? Formulée dans ce contexte, la réponse à la question que nous venons d'énoncer serait la suivante : parce que la « subjectivité » est par définition associée aux « errements » et aux fluctuations, et que, par conséquent, elle n'a pas le droit à la même dignité ontologique que les vérités « pures » ou « fortunées » des sciences objectives. Mais une fois que l'on a démasqué l'intrus derrière ce préjugé – à savoir : l'endossement tacite du point de vue en surplomb du Super-Sujet, ou du Super-Philosophe, si vous préférez – il devient clair que jamais une telle distinction chirurgicale entre la subjectivité errante et l'objectivité pure n'aurait pu être établie du point de vue du sujet connaissant d' « ici bas », coincé dans le bourbier du monde. Comme le dit si bien M. Bitbol, « prétendre faire irruption hors de sa propre cognition ne serait qu'une figure de style. Manquent pour cela le terrain de manœuvre, la latence chronologique et l'altérité radicale. Je peux m'extraire, par un mouvement, un dispositif, ou une identification, de *là* d'où j'ai connu, mais je ne peux pas m'abstraire

de '*je*' qui connaît au moment du connaître»[155].

Tout au long de notre itinéraire, nous avons pu suivre les différentes tentatives, plus ou moins acrobatiques, de sortir du cercle du « 'je' qui connaît au moment du connaître ». Certains sont allés jusqu'à tracer un véritable cercle magique de l'apriorité, s'imaginant à tort que la perruque de la logique et des mathématiques serait assez puissante pour protéger le baron de Münchhausen des ingérences et des irruptions du « dehors », afin de le tirer hors du bourbier du monde. Mais nous avons vu que cette stratégie provoque une véritable spirale infernale « fossés explicatifs / sauts conceptuels » qui finit par dévier les énergies des philosophes vers des impasses. En effet, comment relier le corps à l'esprit, l'analytique à l'empirique, le formel au contentuel, l'*a priori* au contingent, l'objectif au subjectif, autrement qu'en endossant le rôle d'un baron de Münchhausen se tirant seul du marais de la subjectivité pour se déposer sain et sauf dans le royaume des vérités objectives ?

Notre plaidoyer pour le monisme de type-F, pour reprendre ici la terminologie de Chalmers, passe d'abord, comme nous espérons l'avoir montré tout au long de cet ouvrage, par une remise en cause des mécanismes et des schémas conceptuels mêmes qui justifient le clivage phénoménal/physique. Tant que ce clivage restera en place non-questionné, la pensée philosophique ne fera que rejouer en boucle des variations sur le thème du vieux tandem fossés explicatifs/sauts conceptuels.

Nous avons vu ainsi que les deux acteurs du duel qui nous a concernés ici, « trichent » à une étape ou autre de leur structuration argumentative et s'appuient sur des éléments explicatifs à visée « surnaturaliste ». (1_A) Le physicaliste réductionniste s'installe dans le fauteuil *a priori* surplombant le monde. (1_B) Le physicaliste non-réductionniste, quant à lui, mise sur le principe de complétude et de clôture causale du physique. Or, pour concevoir le monde de cette façon-là, i.e., comme dans un vase causalement clos, il faut avoir préalablement endossé un point de vue en surplomb survolant le « nid de coucous », si l'on peut dire ainsi. (2) L'argument de Jackson se rend coupable, à son tour, comme l'a montré fort bien Nagasawa, d'usage abusif de l'attribut démiurgique d'omniscience.

Conclusion : en vain essayera-t-il, notre philosophe Aladin, d'esquiver le marais du monde naturel, et de s'installer confortablement dans le château des vérités « pures ». Notre proposition de solution serait ainsi la suivante : supposons qu'Aladin

[155] (M. Bitbol, 2010 : 28) (Soulignements originaux)

cesse de vouloir « purifier » la lampe de physicalisme de l'ingérence de la subjectivité. Dans ce cas, il se peut que Djinn-l'esprit y retourne tout seul, comme par magie, au sein du monde naturel. Plus encore, peut-être qu'Aladin se rendra-t-il enfin compte qu'il *est* Djinn, que la lampe est *son* monde, et qu'il ne pourra jamais la quitter.

Bibliographie

Albert, D. (1992). *Quantum Mechanics and Experience* . Cambridge, MA: Harvard University Press.

Alter, T. (2001). 'Know-How, Ability, and the Ability Hypothesis'. *Theoria, 67*, 229–39.

Alter, T. (2007). 'Does Representationalism Undermine the Knowledge Argument?'. Dans T. Alter, & S. Walter (Éds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism.* Oxford: Oxford University Press.

Alter, T., & Walter, S. (Éds.). (2007). *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism.* Oxford: Oxford University Press.

Anderson, A. (Éd.). (1983). *Minds and Machine.* (P. Blanchard, Trad.) Prentice-Hall, Inc.

Andler, D. (Éd.). (1992). *Introduction aux sciences cognitives.* Paris: Gallimard.

Andler, D., Fagot-Largeault, A., & Saint-Sernin, B. (2002). *Philosophie des Sciences* (Vol. I&II). Paris: Gallimard.

Anscombe, G. (1957). *Intention.* (M. Maurice, & C. Michon, Trads.) Basil Blackwell.

Aristote. (1974). *Métaphysique.* (J. Tricot, Trad.) Paris: Vrin.

Armstrong, D. M. (1961). *Perception and the Physical World.* London: Routledge.

Armstrong, D. M. (1968). *A Materialist Theory of the Mind.* New York: Routledge & Kegan Paul.

Armstrong, D. M. (1980). 'Naturalism, Materialism, and First Philosophy'. Dans D. Armstrong, *The Nature of Mind and Other Essays* (Vol. The Nature of Mind and Other Essays). St. Lucia: University of Queensland Press.

Armstrong, D. M. (1981). 'The Causal Theory of the Mind'. Dans D. Armstrong, *The Nature of Mind and Other Essays.* Cornell University Press.

Ayer, A. (1936). *Language, Truth, and Logic.* (J. Ohana, Trad.) London: Victor Golancz Ltd.

Barberousse, A., Kistler, M., & Ludwig, P. (2000). *La Philosophie des sciences au XXe siècle* . Paris: Flammarion.

Battermann, R. (s.d.). *Intertheory relations in Physics.* http://plato.stanford.edu.

Beakley, B., & Ludlow, P. (Éds.). (1992). *The Philosophy of Mind - Classical Problems, Contemporary Issues* (éd. 2nd revised edition 2006). Cambridge Massachusetts: MIT Press.

Bealer, G. (1994). 'Mental Properties'. *Journal of Philosophy, 91*, 185–208.

Beaton, M. (2005). 'What RoboDennett Still Doesn't Know'. *Journal of Consciousness Studies, 12(12)*, 3-25.

Beaton, M. (2009). 'Qualia and Introspection'. *Journal of Consciousness Studies 16 (5)*, 88-110.

Beaton, M. (2009). Qualia and Introspection. *Journal of Consciousness Studies 16 (5)*, 88-110.

Bedau, A. M., & Humphreys, P. (Éds.). (2008). *Emergence: Contemporary Reading in Philosophy and Science.* Cambridge Massachusetts: MIT Press.

Bell, J. (1987). *Speakable and Unspeakable in Quantum Mechanics.* Cambridge: Cambridge University Press.

Bennett, M., Dennett, D., Hacker, P., & Searle, J. (2007). *Neuroscience and Philosophy - Brain, Mind, and Language.* New York: Columbia University Press.

Bermúdez, J. L. (1998). *The Paradox of Self-Consciousness.* Cambridge MA: MIT Press.

Bermúdez, J. L. (2003). *Thinking Without Words* . New York: Oxford University Press.

Bickle, J. (s.d.). *Multiple Realizability.* http://plato.stanford.edu/entries/multiple-realizability.

Bickle, J. (2008). 'Real Reduction in Real Neuroscience: Metascience, Not Philosophy of Science (and Certainly Not Metaphysics!)'. Dans J. Hohwy, & J. Kallestrup (Éds.), *Being Reduced.* Oxford: Oxford University Press.

Bird, A. (2007). *Nature's Metaphysics: Laws and Properties.* Oxford: Oxford University Press.

Bitbol, M. (1996). *Schrödinger's philosophy of quantum mechanics.* Kluwer.

Bitbol, M. (1998). *L'aveuglante proximité du réel.* Paris: Flammarion.

Bitbol, M. (2000). *Physique et philosophie de l'esprit.* Paris: Flammarion.

Bitbol, M. (2007). Ontology, Matter and Emergence. *Phenomenology and the Cognitive Science, 6*, 293-307.

Bitbol, M. (2007). 'Ontology, Matter and Emergence'. *Phenomenology and the Cognitive Sciences 6 (3).*

Bitbol, M. (2008). *Mécanique quantique – une introduction philosophique.* Paris: Flammarion.

Bitbol, M. (2010). *De l'intérieur du monde: pour une philosophie et une science des relations.* Paris: Flammarion.

Bitbol, M., & Peschard, I. (2008). 'Heat, Temperature and Phenomenal Concepts'. Dans E. Wright (Éd.), *The Case for Qualia.* MIT Press.

Blackburn, S. (2009). ‘Analysis, Description, and the A priori?'. Dans I. Ravenscroft (Éd.), *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson.* Oxford: Clarendon Press.

Blay, M. (1992). *La naissance de la mécanique analytique : la science du mouvement au tournant des XVIIe et XVIIIe siècles.* Paris: PUF.

Blay, M. (1993). *Les Raisons de l'infini.* Paris: Gallimard.

Blay, M. (Éd.). (2003). *Grand Dictionnaire de la Philosophie* . Paris: Larousse - CNRS Editions.

Block, N. (1980a). ‘What is Functionalism’ . Dans *Readings in the Philosophy of Psychology* (Vol. 1). Cambridge, MA: Harvard University Press.

Block, N. (1980b). ‘Are absent qualia impossible?’. *Philosophical Review, 89/2*, 257-74.

Block, N. (1996). ‘Mental paint and mental latex’. Dans E. Villanueva (Éd.), *Perception.* Atascadero, CA: Ridgeview.

Block, N. (1997). ‘Anti-Reductionism Slaps Back’. Dans J. Tomberlin (Éd.), *Philosophical Perspectives* (Vol. 11). Oxford: Blackwell.

Block, N. (2007). ‘Max Black's Objection to Mind-Body Identity’. Dans T. Alter, & S. Walter (Éds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism.* New York: Oxford University Press.

Block, N. (2007). *Consciousness, Function, and Representation - Collected Papers* (Vol. I). Cambridge Massachusetts: MIT Press.

Block, N., & Fodor, J. A. (1972). ‘What Psychological States Are Not’. *Philosophical Review 81* , 159-81.

Block, N., & Stalnaker, R. (1999). ‘Conceptual Analysis, Dualism and the Explanatory Gap’. *Philosophical Review, 108(1)*, 1–46.

Bogdan, R. (Éd.). (1991). *Mind and Common Sense - Philosophical Essays on Common Sense Psychology.* Cambridge : Cambridge University Press.

Bohm, D. (1952). A Suggested Interpretation of the Quantum Theory in Terms of ‘Hidden’ Variables. *Physical Review 85*, 166-193.

Bohm, D. (1980). *Wholeness and the Implicate Order.* New York: Routledge.

Bohm, D. (1990). 'A new theory of the relationship of mind and matter'. *Philosophical Psychology 3*, 271-286.

Bohm, D., & Hiley, B. (1993). *The Undivided Universe: An Ontological Interpretation of Quantum Theory.* London: Routledge & Kegan Paul.

Bohr, N. (1961). *Physique atomique et connaissance humaine.* Paris: Ed. Gonthier.

Bonnet, C., & Wagner, P. (Éds.). (2006). *L'âge d'or de l'empirisme logique.* Paris: Gallimard.

Boutot, A. (1993). *L'invention des formes.* Paris: Odile Jacob.

Bouveresse, J. (1987). *Le mythe de l'intériorité.* Paris: Ed. de Minuit.

Bouveresse, J., & Rosat, J.-J. (Éds.). (2003). *Philosophie de la perception - phénoménologie, grammaire et sciences cognitives.* Paris: Odile Jacob.

Boyd, R., Gasper, P., & Trout, J. (Éds.). (1999). *The Philosophy of Science.* Cambridge Massachusetts: MIT Press, 1999.

Braddon-Mitchell, D., & Jackson, F. (1996). *Philosophy of Mind and Cognition - An Introduction* . Blackwell Publishing.

Braddon-Mitchell, D., & Nola, R. (Éds.). (2009). *Conceptual Analysis and Philosophical Naturalism.* Cambridge Massachusetts: MIT Press.

Brakel, J. v. (1986). 'The Chemistry of Substances and the Philosophy of Mass Terms'. *Synthese 69*, 291–324.

Brandom, R. B. (2000). *Articulating Reasons - An Introduction to Inferentialism.* Harvard University Press.

Brentano, F. (1874). *Psychologie vom empirischen Standpunkt.* (M. Gandillac, Trad.) Leipzig.

Broad, C. (1925). *The Mind and its Place in Nature.* London: Routledge et Kegan Paul LTD.

Broglie, L. (1937). *La physique nouvelle et les quanta.* Paris: Flammarion.

Burge, T. (1977). 'Belief De Re'. *The Journal of Philosophy, 74*, 338-362.

Burge, T. (1993). 'Mind-Body Causation and Explanatory Practice'. Dans J. Heil, & A. Mele (Éds.), *Mental Causation.* Oxford: Clarendon Press.

Burge, T. (2007). *Foundations of Mind* (Vol. 2). Oxford: Clarendon Press.

Byrne, A. (2002). 'Something About Mary'. *Grazer Philosophische Studien, 63*, 123-40.

Byrne, A., & Tye, M. (2006). Qualia Ain't in the Head. *Noûs 40 (2)*, 241-255.

Cahn, S. (Éd.). (2003). *Philosophy for 21st Century - a Comprehensive Reader.* Oxford: Oxford University Press.

Campbell, J. (1970). *Body and Mind.* London: Doubleday.

Carnap, R. (1928). *Der Logische Aufbau der Welt.* (E. Schwartz, Trad.) Berlin: Weltkreis.

Carnap, R. (1932b). 'Die physikalische Sprache als Unversalsprache der Wissenschaft'. *Erkenntnis 2*, 432-465.

Carnap, R. (1934). *The Logical Syntax of Language.* London: Routledge & Kegan Paul LTD.

Carnap, R. (1950). ‘Empiricism, Semantics, and Ontology’. *Revue internationale de philosophie 4*, 20–40.

Carnap, R. (1956). *Meaning and Necessity.* (F. Rivenc, & P. d. Rouilhan, Trads.) Chicago, IL: University of Chicago Press.

Carnap, R. (1966). *Philosophical Foundations of Physics.* (J. Luccioni, & A. Soulez, Trads.) Basic Books.

Caro, d., & Macarthur, D. (Éds.). (2004). *Naturalism in Question.* Cambridge Massachusetts: Harvard University Press.

Carruthers, P. (2000). *Phenomenal Consciousness - a Naturalistic View.* Cambridge: Cambridge University Press.

Carruthers, P. (2004). *The Nature of the Mind.* New York: Routledge.

Carruthers, P. (2005). *Consciousness - Essays from a Higher-Order Perspective.* Oxford: Clarendon Press.

Carruthers, P., Laurence, S., & Stich, S. (Éds.). (2005). *The Innate Mind : Structure and contents.* Oxford: Oxford University Press.

Carthwright, N. (1983). *How the Laws of Physic Lie.* Oxford : Oxford University Press.

Chalmers, D. (1996). *The Conscious Mind.* Oxford: Oxford University Press.

Chalmers, D. (2002). ‘Content and Epistemology of Phenomenal Belief’. Dans Q. Smith, & A. Jokic (Éds.), *Consciousness: New Philosophical Essays.* Oxford: Oxford University Press.

Chalmers, D. (2003a). 'Consciousness and its place in Nature’. Dans S. Stich, & T. Warfield (Éds.), *The Blackwell Guide to Philosophy of Mind.* Blackwell Publishing.

Chalmers, D. (2004). ‘Imagination, Indexicability, and Intensions’. *Philosophy and Phenomenological Research 68*, 182-90.

Chalmers, D. (2004a). ‘Phenomenal Concepts and the Knowledge Argument’. Dans P. Ludlow, D. Stoljar, & Y. Nagasawa (Éds.), *There is something about Mary.* Cambridge Mass: MIT Press.

Chalmers, D. (2007). ‘Phenomenal Concepts and the Explanatory Gap’. Dans T. Alter, & S. Walter (Éds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism.* Oxford: Oxford University Press.

Chalmers, D. J. (1995). Absent Qualia, Fading Qualia, Dancing Qualia. Dans T. Metzinger (Éd.), *Conscious Experience.* Ferdinand Schoningh.

Chalmers, D., & Jackson, F. (2001). ‘Conceptual analysis and reductive explanation’. *Philosophical Review, 110/3*, 315-60.

Chalmers, D., Manley, D., & Wasserman, R. (Éds.). (2009). *Metametaphysics: New Essays on the Foundations of Ontology.* Oxford: Oxford University Press.

Chomsky, N. (1968). *Language and Mind.* Harcourt: Brace & World.

Chomsky, N. (1988). *Language and Problems of Knowledge: The Managua Lectures.* Cambridge: MIT Press.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind.* Cambridge Massachusetts: MIT Press.

Christensen, C. B. (2008). *Self and World: From Analytic Philosophy to Phenomenology.* Berlin: Walter de Gruyter.

Chuchland, P. (1989). *The Nature of Mind and the Structure of Science - a Neurocomputational Perspective.* Cambridge Massachusetts: MIT Press.

Churchland, P. (1981). ‘Eliminative Materialism and the Propositional Attitudes’. *Journal of Philosophy 78*, 67-90.

Churchland, P. (1985a). 'Reduction, Qualia, and the Direct Introspection of Brain States’. *Journal of Philosophy 82*, 8-28.

Churchland, P. (1985b). ‘Review of Robinson’s Matter and Sense' . *Philosophical Review 94*, 117-120.

Churchland, P. (1986). *Neurophilosophy: Toward a Unified Science of the Mind/Brain.* Cambridge, MA: MIT Press.

Churchland, P. (1989). ‘Knowing Qualia. A Reply to Jackson’. Dans P. Churchland, *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science.* Cambridge, MA: MIT Press.

Churchland, P. (1996). ‘The hornswoggle problem’. *Journal of Consciousness Studies, 3*, 402-8.

Clark, A. (2001). *Mindware - an Introduction to the Philosophy of Cognitve Science.* Oxford: Oxford University Press.

Cohen, J., & McLaughlin (Éds.). (2007). *Contemporary Debates in Philosophy of Mind.* Blackwell Publishing.

Conee, E. (1994). ‘Phenomenal Knowledge’. *Australasian Journal of Philosophy 72*, 136–150.

Crane, T. (1991). ‘All God has to do’. *Analysis 51*, 235-44.

Crane, T. (2000). ‘The Origins of Qualia’. Dans T. Crane, & S. A. Patterson (Éds.), *The History of the Mind-Body Problem.* Routledge.

Crane, T. (2003). Subjective Facts. Dans *Real Metaphysics: Essays in Honour of D. H. Mellor.* New York: Routledge.

Crane, T. (2003). 'The Intentional Structure of Consciousness'. Dans Q. Smith, & A. Jokic (Éds.), *Consciousness: New Philosophical Essays* . Oxford: Oxford University Press.

Crane, T., & Mellor, D. (1990). ‘There is no question of physicalism’. *Mind 99 (394)*, 185-206.

Cues, N. d. (ca 1440). *De docta ignorantia.* (H. Pasqua, Trad.) Paris: Bibliothèque Rivages.

Davidson, D. (1970). ‘Mental Events’. Dans D. Davidson, *Essays on Actions and Events* (pp. 207–223). Oxford: Oxford University Press.

Davidson, D. (1993). ’Thinking Causes’. Dans J. Heil, & A. Mele (Éds.), *Mental Causation.* Oxford: Clarendon Press.

Davies, M., & Humberstone, L. (1980). ‘Two Notions of Necessity'. *Philosophical Studies 38* .

Deledalle, G. (1979). *Théorie et pratique du signe.* Paris: Payot.

Dennett, D. (1987). *The Intentional Stance.* (P. Engel, Trad.) Cambridge Massachusetts: MIT Press.

Dennett, D. (1990). ‘Quining qualia’. Dans W. Lycan (Éd.), *Mind and Cognition – A Reader*. Oxford: Blackwell.

Dennett, D. (1991). *Consciousness Explained.* Boston: Little, Brown and Company.

Dennett, D. (1996). ‘Facing Backward on the Problem of Consciousness’. *Journal of Consciousness Studies, 3*, 4–6.

Dennett, D. (2007). ‘What RoboMary Knows’. Dans T. Alter, & S. Walter (Éds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism.* Oxford: Oxford University Press.

Dennett, D., & Hofstadter, D. (1981). *The Mind’s I.* Basic Books.

Depraz, N., F. Varela, & Pierre Vermersch. (2003). *On Becoming Aware: A Pragmatics of Experiencing.* Amsterdam: John Benjamins.

Descombes, V. (1995). *La denrée mentale.* Paris: Ed. de Minuit.

Descombes, V. (1996). *Les institutions du sens.* Paris: Ed. de Minuit.

Dizadji-Bahmani, F., Frigg, R., & Hartmann, S. (2010). ‘Who’s Afraid of Nagelian Reduction?’. *Erkenntnis 73 (3)*, 393-412.

Dokic, J., & Engel, P. (2001). *Ramsey, Vérité et Succès.* Paris: P.U.F.

Dostoïevski, F. (1956). *Les Nuits blanches - Le Sous-sol.* Paris: Gallimard.

Dretske, F. (1995). *Naturalizing the Mind.* Cambridge, Mass:: The MIT Press, Bradford Books.

Dreyfus, H. (1972). *What Computers Can't Do - the Limits of Artificial Intelligence.* (R.-M. Vassallo-Villaneau, Trad.) New York: Harper & Row Publishers.

Dummett, M. (1988). *Ursprünge des analytischen Philosophie.* (M. Lescourret, Trad.) Frankfurt am Main: Suhrkamp Verlag.

Dupré, J. (1993). *The Disorder of Things. Metaphysical Foundations of the Disunity of Science.* Cambridge, Massachusetts: Harvard University Press.

Dupré, J. (2004). ‘The Miracle of Monism’. Dans M. De Caro, & D. McArthur (Éds.), *Naturalism in question.* Cambridge Massachusetts: Harvard University Press.

Dupuy, J.-P. (1994). *Aux origines des sciences cognitives.* Paris: Ed. La Découverte.

Dupuy, J.-P. (2000). *Les savants croient-ils en leurs théories?* Paris: INRA.

Dutant, J., & Engel, P. (Éds.). (2005). *Philosophie de la Connaissance.* Paris: Vrin.

Eco, U. (1997). *Kant e l'ornitorinco.* (J. Gayrard, Trad.) Milan: Bompiani.

Eddington, A. (1928). *The Nature of the Physical World.* London: J.M. Dent et Sons.

Egner, R., & Denonn, L. (Éds.). (2009). *Bertrand Russell - The Basic Writings of Bertrand Russell.* London & New York: Routledge.

Engel, P. (1989). *La norme du vrai - philosophie de la logique.* Paris: Gallimard.

Engel, P. (1994). *Davidson et la philosophie du langage.* Paris: PUF.

Engel, P. (1994). *Introduction à la philosophie de l'esprit.* Paris: Ed. la Découverte.

Engel, P. (1996). *Philosophie et psychologie.* Paris: Gallimard.

Engel, P. (Éd.). (2000). *Précis de philosophie analytique.* Paris: PUF.

Esfeld, M. (2004). 'Quantum entanglement and a metaphysics of relations'. *Studies in History and Philosophy of Modern Physics 35*, 601–617.

Esfeld, M. (2005). *La philosophie de l'esprit - de la relation entre l'esprit et la nature.* Paris: Armand Colin.

Esfeld, M., & Lam, V. (2008). Moderate structural realism about space-time . *Synthese*, 160: 27–46.

Espagnat d', B. (1981). *A la recherche du réel.* Paris: Gauthier-Villars.

Espagnat d', B. (1985). *Une incertaine réalité.* Paris: Gauthier-Villars.

Espagnat d', B. (1990). *Penser la science.* Paris: Bordas.

Espagnat d', B. (2002). *Traité de physique et de philosophie.* Paris: Fayard.

Espagnat, d. B. (1999). 'Concepts of reality'. Dans H. Atmanspacher, U. Müller-Herold, & A. Amann (Éds.), *On Quanta, Mind, and Matter* (pp. 249-270). Dordrecht: Kluwer.

Evans, G. (1973). ‘The Causal Theory of Names'. *Proceedings of the Aristotelian Society, Supplementary Volume 47*, 187–208.

Evans, G. (1979). ‘Reference and Contingency’. *The Monist xii*, 161–189.

Evans, G. (1982). *The Varieties of Reference.* Oxford: Oxford University Press.

Evans, J., & Thorndike, A. S. (Éds.). (2006). *Quantum Mechanics at the Crossroads, New perspectives from History, Philosophy and Physics.* Springer.

F. P. Ramsey. (1990). *Philosophical Papers.* (D. H. Mellor, Éd.) Cambridge: Cambridge University Press.

Feigl, H. (1958). ‘The “mental” and the “physical'". Dans H. Feigl, M. Scriven, & G. Maxwell (Éds.), *Concepts, Theories and the Mind-Body Problem: Minnesota Studies in the Philosophy of Science,* (Vol. 2). Minneapolis: Minnesota University Press.

Feyerabend, P. (1962). 'Explanation, reduction, and empiricism'. Dans H. Feigl, & G. Maxwell (Éds.), *Minnesota Studies in the Philosophy of Science* (pp. 28–97). D. Reidel Publishing Company.

Feyerabend, P. (1975). *Against Method.* London: New Left Books.

Feyerabend, P. (1981). *Realism, Rationalism and Scientific Method - Philosophical Papers.* Cambridge: Cambridge University Press.

Feyerabend, P. (1987). *Farewell to Reason.* London: Verso.

Finkelstein, D. (2003). *Expression and the Inner.* Harvard University Press.

Fisette, D., & Poirier, P. (2000). *Philosophie de l'esprit - état des lieux.* Paris: Vrin.

Fisette, D., & Poirier, P. (Éds.). (2002). *Philosophie de l'Esprit - Psychologie du Sens Commun et Sciences de l'Esprit* (Vol. I). Paris: Vrin.

Fisette, D., & Poirier, P. (Éds.). (2003). *Philosophie de l'Esprit - Problèmes et Perspectives* (Vol. II). Paris: Vrin.

Flanagan, O. (1992). *Consciousness Reconsidered.* Cambridge: M.I.T. Press.

Fodor, J. (1968). *Psychological Explanation.* New York: Random House.

Fodor, J. (1974). ‘Special Sciences: Or the Disunity of Science as a Working Hypothesis’. *Synthese, 28,* 97-115.

Fodor, J. (1981). ‘The Mind-Body Problem’. *Scientific American 244*.

Fodor, J. (1983). *The Modularity of Mind.* (A. Gerschenfeld, Trad.) Cambridge Massachusetts: MIT Press.

Fodor, J. (2000). *The Mind Doesn't Work That Way.* (C. Tiercelin, Trad.) Cambridge: MIT Press.

Fodor, J., & Katz, J. (1965). *The Structure of Language.* New Jersey: Prentice Hall, Inc., Englewood Cliffs.

Foley, R. (2004). *Intellectual Trust in oneself and others.* Cambridge Mass.: Cambridge University Press.

Foster, J. (1991). *The Immaterial Self: A Defence of the Cartesian Dualist Conception of the Mind.* Oxford: Oxford University Press.

Fraassen, v. B. (1980). *The Scientific Image.* Oxford: Oxford University Press.

Fraassen, v. B. (2008). *Scientific Representation: Paradoxes of Perspective.* Oxford: Oxford University Press.

Frege, G. (1884). *Die Grundlagen der Arithmetik.* Breslau: Verlag von Wilhelm Koebner.

Frege, G. (1971). *Ecrits logiques et philosophiques.* (C. Imbert, Trad.) Paris: Ed. du Seuil.

French, S., & Ladyman, J. (2003a). 'Remodelling structural realism: Quantum physics and the metaphysics of structure'. *Synthese, 136*, 31–56.

French, S., & Ladyman, J. (2003b). 'Between platonism and phenomenalism: Reply to Cao'. *Synthese, 136*, 73–78.

Friedman, M. (1999). *Reconsidering Logical Positivism.* Cambridge: Cambridge University Press.

Frith, C. (2007). *Making up the Mind.* Blackwell Publishing.

Gardner, H. (1985). *The Mind's New Science - A History of the Cognitive Revolution.* (J.-L. Peytavin, Trad.) Basic Books.

Gayensford, M. (2006). *I - The Meaning of the First-Person Term.* Oxford: Clarendon Press.

Gertler, B. (1999). ‘A Defense of the Knowledge Argument’. *Philosophical Studies, 93*, 317–336.

Gillett, C., & Loewer, B. (Éds.). (2001). *Physicalism and its Discontents.* Cambridge: Cambridge University Press.

Goldberg, S. (Éd.). (2007). *Internalism and Externalism in Semantics and Epistemology.* Oxford: Oxford University Press.

Goldman, A. (Éd.). (1993). *Philosophical Applications of Cognitive Science.* Boulder Westview Press.

Goldman, A. I. (2006). *Stimulating Minds : Philosophy, Psychology, and Neuroscience of Mind Reading.* Oxford: Oxford University Press.

Goodman, N. (1951). *The Structure of Appearance.* Cambridge Massachusetts: Harvard University Press.

Goodman, N. (1955). *Fact, Fiction, and Forecast. Cambridge, MA: Harvard UP, 1955.* Cambridge Massachusetts: Harvard University Press.

Goodman, N., & Elgin, C. Z. (1988). *Reconceptions in Philosophy and Other Arts and Sciences.* Indianapolis/Cambridge: Hackett Publishing Company.

Graham, G. (s.d.). *Behaviorism.* http://plato.stanford.edu/entries/behaviorism.

Graham, G., & Horgan, T. (2000). 'Mary Mary, Quite Contrary'. *Philosophical Studies 99*, 59–87.

Grim, P. (1984). *The Incomplete Universe: Totality, Knowledge and Truth.* Cambridge: MIT Press.

Grim, P. (1991). 'There Is No Set of All Truths'. *Analysis 44*, 206–208.

Guenancia, P. (1998). *L'intelligence du sensible - essai sur le dualisme cartésien.* Paris: Gallimard.

Gulick, R. V. (s.d.). 'Maps, gaps and traps'. Dans A. Jokic, & Q. Smith (Éds.), *Consciousness: New Philosophical Perspectives.* Oxford: Oxford University Press.

Gulick, R. V. (1980). 'Functionalism, information and content'. *Nature and System 2*, 139–62.

Gulick, R. V. (1985). 'Physicalism and the subjectivity of the mental'. *Philosophical Topics, 13*, 51-70.

Gulick, R. V. (1993). 'Understanding the phenomenal mind: Are we all just armadillos?' . Dans M. Davies, & G. Humphreys (Éds.), *Consciousness: Psychological and Philosophical Essays.* Oxford: Blackwell.

Gulick, R. V. (1999). 'Conceiving Beyond Our Means: The Limits of Thought Experiments'. Dans S. Hameroff, A. Kaszniak, & D. Chalmers (Éds.), *Toward a Science of Consciousness III - The Third Tucson Discussion and Debates* . Cambridge Massachusetts: MIT Press.

Gulick, R. V. (2004). ‘So Many Ways of Saying No to Mary”. Dans P. Ludlow, & al (Éds.), *There is something about Mary.* Cambridge Mass: MIT Press.

Gulick, R. V. (2007). ‘Functionalism and Qualia’. Dans M. Velmans, & S. Schneider (Éds.), *The Blackwell Companion to Consciousness.* Blackwell Publishing.

Gulick, R. V. (2008). ‘Jackson’s Change of Mind: Representationalism, a priorism and the Knowledge Argument’. Dans I. Ravenscroft (Éd.), *Minds, Ethics, and Conditionals Themes from the Philosophy of F. Jackson.* Oxford: Clarendon Press.

Hacking, I. (1983). *Representing and Intervening.* Cambridge: Cambridge University Press.

Hale, B., & Wright C. (Éds.). (1997). *A Companion to the Philosophy of Language.* Blackwell Publishers.

Hameroff, S. (1994). ‘Quantum coherence in microtubules: a neural basis for emergent consciousness?'. *Journal of Consciousness Studies 1*, 91–118.

Hameroff, S. R., Kaszniak, A. W., & Chalmers, D. J. (Éds.). (1998). *Toward a Science of Consciousness II - The Second Tucson Discussion and Debates.* Cambridge Massachusetts: MIT Press.

Hameroff, S. R., Kaszniak, A. W., & Chalmers, D. J. (Éds.). (1999). *Toward a Science of Consciousness III - The Third Tucson Discussions and Debates.* Cambridge Massachusetts: MIT Press.

Hardin, C. (1986). *Color for Philosophers.* Indianapolis: Hackett.

Haugeland, J. (1998). *Having Thought - Essays in the Metphysics of Mind.* Cambridge Massachusetts: Harvard University Press.

Hawthorne, J. (2006). *Metaphysical Essays.* Oxford: Clarendon Press.

Heil, J. (1998). *Philosophy of Mind.* Routledge.

Heil, J. (2003). *From an Ontological Point of View.* Oxford: Oxford University Press.

Heil, J. (Éd.). (2004). *Philosophy of Mind - A Guide and Anthology.* Oxford: Oxford University Press.

Heil, J., & Mele, A. (Éds.). (1993). *Mental Causation.* Oxford: Clarendon Press.

Heisenberg, W. (1972). *La Partie et le tout.* Paris: Albin Michel.

Hellie, B. (2004). ‘Inexpressible Truth and the Allure of the Knowledge Argument’. Dans P. Ludlow, & al (Éds.), *There is something about Mary*. Cambridge Mass: MIT Press.

Hellman, C. (1985). ‘Determination and Logical Truth’. *The Journal of Philosophy, 82(11)*, 607–616.

Helmholtz, H. v. (1847). *Über die Erhaltung der Kraft: Eine physikalische Abhandlung, Vorgetragen in der Sitzung der physikalischen Gesellschaft.* (L. Pérard, Trad.) Berlin: G. Reimer.

Hempel, C. (1935). ‘Analyse logique de la psychologie’. *Journal de Synthèse, 10*, 27-42.

Hempel, C. (1942). ‘The Function of General Laws History’. *Journal of Philosophy 39 (2)*, 35-48.

Hempel, C. (1966). *Philosophy of Natural Science.* (B. Saint -Sernin, Trad.) New Jersey: Prentice Hall, Englewood Cliffs.

Hempel, C. (1980). 'Comments on Goodman's Ways of Worldmaking’. *Synthese, 45*, 139–199.

Hintikka, J. (1995). *L'intentionnalité et les mondes possibles.* (N. Lavand, Trad.) Lille: Presses Universitaires de Lille.

Hochmann, J., & Jeannerod, M. (1996). *Esprit - où es-tu?* Paris: Odile Jacob.

Hohwy, J., & Kallestrup, J. (Éds.). (2008). *Being Reduced - New essays on Reduction, Explanation and Causation.* Oxford: Oxford University Press.

Horgan, T. (1984). ‘Jackson on Physical Information and Qualia’. *Philosophical Quarterly 32*, 127–136.

Hossack, K. (2007). *The Metaphysics of Knowledge.* Oxford: Oxford University Press.

Hume, D. (1748). *An Enquiry Concerning Human Understanding (Philosophical Essays Concerning Human Understanding)* (éd. Aubier, Paris, 1947). (André Leroy, Trad.) Londres: A. Millar.

Humphrey, N. (Éd.). (2000). *How to Solve the Mind-Body Problem.* Imprint Academic.

Husserl, E. (1900-1901). *Logische Untersuchungen : Prolegomena zur reinen Logik.* Niemeyer.

Hüttemann, A. (2004). *What's Wrong with Microphysicalism?* Routledge.

Huxley, T. (1874). ‘On the Hypothesis that Animals are Automata, and its History’. *Fortnightly Review, 95*, 555–80.

Inwagen, P. V. (1990). *Material Beings.* Ithaca: Cornell University Press.

Jackson, F. (1982). 'Epiphenomenal Qualia'. *Philosophical Quarterly 32*, 127-136.

Jackson, F. (1986). 'What Mary Didn't Know'. *Journal of Philosophy 83*, 291–295.

Jackson, F. (1995). ‘Postscript on ‘What Mary didn't know'. Dans P. Ludlow, & al. (Éds.), *There is something about Mary.* Cambridge Mass: MIT Press.

Jackson, F. (1998). ‘Postscript on qualia’. Dans F. Jackson, *Mind, Method and Conditionals.* London: Routledge.

Jackson, F. (1998). *From Metaphysics to Ethics: A Defense of Conceptual Analysis.* Oxford: Clarendon Press.

Jackson, F. (2003). ‘Mind and Illusion’. Dans A. O'Hear (Éd.), *Minds and Persons* (pp. 251–71).

Jackson, F. (2003). ‘From H2O to Water: The Relevance to A priori Passage’. Dans H. Lillehammer, & G. Rodriguez-Pereyra (Éds.), *Real Metaphysics.* Routledge.

Jackson, F. (2004). ‘Why We Need A-Intensions’. *Philosophical Studies 118 (1-2)*, 257-277.

Jackson, F. (2005). ‘The Case for A priori Physicalism’. Dans C. Nimtz, & A. Beckermann (Éds.), *Philosophy-Science -Scientific Philosophy.* Main Lectures and Colloquia of Gap 5, Fifth International Congress of the Society for Analytical Philosophy, Mentis.

Jackson, F. (2007). ‘A priori physicalism’. Dans B. P. McLaughlin, & J. D. Cohen (Éds.), *Contemporary Debates in Philosophy of Mind.* Blackwell.

Jackson, F. (2007). ‘The knowledge argument, diaphonousness, representationalism’. Dans T. Alter, & S. Walter (Éds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism.* Oxford: Clarendon Press.

Jacob, P. (Éd.). (1980). *De Vienne à Cambridge - l'héritage du positivisme logique.* Paris: Gallimard.

Jacob, P. (1980). *L'empirisme logique - ses antécédents, ses critiques.* Paris: Ed. de Minuit.

Jacob, P. (Éd.). (1989). *Epistémologie.* Paris: Odile Jacob.

Jacquette, D. (Éd.). (2002). *A Companion to Philosophical Logic.* Blackwell Publishers.

James, W. (1910). *Psychology.* New York: Henry Holt and Co.

Jeannerod, M. (2002). *La nature de l'esprit - sciences cognitives et cerveau.* Paris: Odile Jacob.

Kant, E. (1990). *Critique de la raison pure.* Paris: PUF.

Khlentzos, D. (2004). *Naturalistic Realism and the Antirealist Challenge.* Cambridge Massachussets: MIT Press.

Kim, J. (1989b). 'The myth of non-reductive physicalism' . *Proceedings and Addresses of the American Philosophical Association, 63:*, 31-47.

Kim, J. (1993). *Mind and Supervenience.* Cambridge: Cambridge University Press.

Kim, J. (1998). *Mind in a Physical World - As Essay on the Mind-Body Problem and Mental Causation.* Cambridge Massachusetts: MIT Press.

Kim, J. (2005). *Physicalism, Or Something Near Enough.* Princeton: Princeton University Press.

Kim, J. (2008). 'Reduction and Reductive Explanation: Is One Possible Without the Other?' . Dans J. Hohwy, & J. Kallestrup (Éds.), *Being Reduced.* Oxford: Oxford University Press.

Kim, J. S. (Éd.). (2009). *A Companion to Metaphysics.* Wiley-Blackwell.

Kim, J., & Sosa, E. (Éds.). (1999). *Metaphysics - An Anthology.* Blackwell Publishing.

Kistler, M. (1999). *Causalité et lois de la nature.* Paris: Vrin.

Kriegel, U., & Williford, K. (Éds.). (2006). *Self-Representational Approaches to Consciusness.* Cambridge Massachusetts: MIT Press.

Kripke, S. (1972). *Naming and Necessity.* (P. Jacob, & F. Recanati, Trads.) Blackwell Publishing.

Kusch, M. (1995). *Psychologism: A Case Study in the Sociology of Philosophical Knowledge.* London: Routledge.

Kusch, M. (s.d.). *Psychologism.* http://plato.stanford.edu/entries/psychologism.

Ladyman, J., & Ross, D. (2007). *Every Thing Must Go: Metaphysics Naturalised.* Oxford: Oxford University Press.

Lakatos, I. (1980). *Mathematics, Science and Epistemology.* Cambridge: Cambridge University Press.

Lakatos, I. (1989). *The Methodology of Scientific Research* (Vol. I&II). Cambridge: Cambridge University Press.

Laugier, S. (Éd.). (2001). *Carnap et la construction logique du monde.* Paris: Vrin.

Laugier, S., & Bitbol, M. (Éds.). (1997). *Physique et réalité – un débat avec B. d'Espagnat.* Paris, Ed. Frontières: Frontières.

Laugier, S., & Wagner, P. (Éds.). (2004). *Philosophie des Sciences - Naturalismes et réalismes* (Vol. II). Paris: Vrin.

Laugier, S., & Wagner, P. (Éds.). (2004). *Philosophie des Sciences - Théories, Expériences et Méthodes* (Vol. I). Paris: Vrin.

Levin, J. (1986). 'Could Love Be Like a Heat Wave? Physicalism and the Subjective Character of Experience' . *Philosophical Studies, 49*, 245–261.

Levin, J. (2007). 'What is a phenomenal concept?'. Dans T. Alter, & S. Walter (Éds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism.* Oxford: Oxford University Press.

Levine, J. (1983). 'Materialism and qualia: the explanatory gap'. *Pacific Philosophical Quarterly, 64*, 354-361.

Levine, J. (1993). 'On leaving out what it's like'. Dans M. Davies, & G. Humphreys (Éds.), *Consciousness, Psychological et Philosophical Essays.* Blackwell.

Levine, J. (1999). 'Conceivability, Identity, and the Explanatory Gap' . Dans S. Hameroff, A. Kaszniak, & D. Chalmers (Éds.), *Toward a Science of Consciousness III - The Third Tucson Discussion and Debates.* Cambridge Massachusetts: MIT Press.

Levine, J. (2001). *Purple Haze - The Puzzle of Consciousness.* Oxford: Oxford University Press.

Levine, J. (2007). ‘Anti-materialist arguments and influential replies'. Dans M. Velmans, & S. Schneider (Éds.), *The Blackwell Companion to Consciousness.* Blackwell.

Levine, J. (2007). ‘Phenomenal Concepts and the Materialist Constraint’. Dans T. Alter, & S. Walter (Éds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism,*. Oxford: Oxford University Press.

Lewis, D. (1970c). ‘How to Define Theoretical Terms’. *Journal of Philosophy, 67*, 427–446.

Lewis, D. (1972). ‘Psychophysical and theoretical identifications’. *Australasian Journal of Philosophy 50*, 249-58.

Lewis, D. (1983). ‘New Work for a Theory of Universals’. *Australasian Journal of Philosophy, 61(4)*, 343–377.

Lewis, D. (1983). ‘Mad Pain and Martian Pain’ and ‘Postscript’. Dans D. Lewis, *Philosophical Papers* (Vol. 1). Oxford: Oxford University Press.

Lewis, D. (1986a). *On the Plurality of Worlds.* Oxford: Blackwell Publishers.

Lewis, D. (1988). ‘What Experience Teaches’. *Proceedings of the Russellian Society*.

Lewis, D. (1994). 'Reduction of Mind’. Dans S. Guttenplan (Éd.), *A Companion to the Philosophy of Mind.* Oxford: Blackwell.

Lewis, D. (1996). ‘An Argument for the Identity Theory’. *Journal of Philosophy 63*.

Lewis, D. (1999a). *Papers in Metaphysics and Epistemology.* Cambridge: Cambridge University Press.

Lewis, D. (2009). ‘Ramseyan Humility’ . Dans D. Braddon-Mitchell, & R. Nola (Éds.), *Conceptual Analysis and Philosophical Naturalism.* Cambridge Mass: MIT Press.

Loar, B. (1990/1997). ‘Phenomenal States’. *Philosophical Perspectives*, 81–108.

Locke, J. (1975). *An Essay Concerning Human Understanding.* Oxford: Clarendon Press.

Lockwood, M. (1989). *Mind, Brain and the Quantum.* Cambridge, MA: Basil Blackwell Ltd.

Lockwood, M. (2005). *The Labyrinth of Time - introducing the universe.* Oxford: Oxford University Press.

Loewer, B. (2001). ‘From Physics to Physicalism’. Dans C. Gillett, & B. Loewer (Éds.), *Physicalism and Its Discontents.* Cambridge: Cambridge University Press.

Loewer, B. M. (2003). 'Consciousness and Quantum Theory: Strange Bedfellows'. Dans Q. Smith, & A. Jokic (Éds.), *Consciousness: New Philosophical Perspectives.* Oxford: Oxford University Press.

Loocke, P. V. (Éd.). (2001). *The Physical Nature of Consciousness.* Amsterdam: John Benjamins Publishing.

Loocke, v. P. (Éd.). (2001). *The Physical Nature of Consciousness.* Amsterdam: John Benjamins Publishing.

Lowe, E. J. (2002). *A Survey of Metaphysics.* Oxford: Oxford University Press.

Ludlow, P., Nagasawa, Y., & Stoljar, D. (Éds.). (2004). *There's Something About Mary - Essays on Phenomenal Consciousness and Franck Jackson's Knowledge Argument.* Cambridge, Massachusetts: MIT Press.

Ludwig, K. (2003). ‘The Mind-Body Problem: an Overview’. Dans S. Stich, & T. Warfield (Éds.), *The Blackwell Guide to Philosophy of Mind.* Blackwell Publishing.

Lycan, W. (1987). *Consciousness.* Cambridge, Massachusetts: MIT Press.

Lycan, W. (1990). ‘What is the Subjectivity of the Mental’. *Philosophical Perspectives 4 (Action Theory and Philosophy of Mind)*, 109–130.

Lycan, W. (Éd.). (1990). *Mind and Cognition - A Reader.* Oxford UK: Blackwell.

Lycan, W. (1996). *Consciousness and Experience.* Cambridge, MA: MIT Press.

Lycan, W. (2003). ‘Perspectival Representation and the Knowledge Argument’. Dans Q. Smith, & A. Jokic (Éds.), *Consciousness: New Philosophical Essays.* Oxford: Oxford University Press.

Lycan, W. G. (2009). ‘Serious Metaphysics: Frank Jackson's Defense of Conceptual Analysis’. Dans I. Ravenscroft (Éd.), *Minds, Ethics, and Conditionals: Themes From the Philosophy of Frank Jackson.* Oxford: Clarendon Press.

Lycan, W., & Prinz, J. (Éds.). (1990). *Mind an Cognition - An Anthology.* Blackwell Publishers.

Lyons, J. C. (2009). *Perception and Basic Beliefs - Zombies, Modules, and the Problem of the External World.* Oxford: Oxford Univeristy Press.

Markosian, N. (2005). 'Against Ontological Fundamentalism'. *Facta Philosophica 7 (1)*, 69-83.

Martin, C. (2008). *Mind in Nature.* Oxford: Clarendon Press.

McDowell, J. (1994). *Mind an World.* Cambridge Massachusetts: Harvard University Press.

McDowell, J. (1998). *Mind, Value, and Reality.* Cambridge Massachusetts: Harvard University Press.

McGinn, C. (1982). *The Character of Mind - an Introduction to the Philosophy of Mind.* Oxford: Oxford University Press.

McGinn, C. (1983). *The Subjective View - Secondary Quality and Indexical Thoughts.* Oxford: Clarendon Press.

McGinn, C. (1989). 'Can we solve the mind-body problem?'. *Mind, 98*, 349-66.

McLaughlin, B. (2001). 'In Defense of New Wave Materialism: a Response to Horgan and Tienson'. Dans C. Gillett, & B. Loewer (Éds.), *Physicalism and Its Discontents.* Cambridge: Cambridge University Press.

McLaughlin, B. P. (2005). 'A priori vs A posteriori Physicalism' . Dans C. Nimtz, & A. Beckermann (Éds.). Philosophy-Science -Scientific Philosophy, Main Lectures and Colloquia of Gap 5, Mentis.

Mele, A., & Rawling, P. (Éds.). (2004). *The Oxford Handbook of Rationality.* Oxford: Oxford University Press.

Mellor, D. (1992/1993). 'Nothing Like Experience'. *Proceedings of the Aristotelian Society 93*, 1-16.

Mellor, D. (1992/1993). 'Nothing Like Experience'. *Proceedings of the Aristotelian Society 93*, 1-16.

Mellor, D. (1995). *The Facts of Causation.* London: Routledge.

Mellor, D. (1995). *The Facts of Causation.* London: Routledge.

Melnyk, A. (2003). *A Physicalist Manifesto: Thoroughly Modern Materialism.* Cambridge: Cambridge University Press.

Merleau-Ponty, M. (1945). *La Phénoménologie de la perception.* Paris: NRF Gallimard.

Millikan, R. (2004). *Varieties of Meaning.* Cambridge, Massachusetts: MIT Press.

Moran, D. (2000). *Introduction to Phenomenology.* London and New York: Routledge.

Morgan, L. (1923). *Emergent Evolution.* Williams and Norgate.

Moser, P. (1993). *Philosophy after Objectivity - Making Sense in Perspective.* Oxford: Oxford University Press.

Moser, P., & Trout, J. (Éds.). (1995). *Contemporary Materialism - A Reader.* New York: Routledge.

Nagasawa, Y. (2002). 'The Knowledge Argument against Dualism'. *Theoria LXVIII*, 205-223.

Nagasawa, Y. (2008). *God and Phenomenal Consciousness – A Novel Approach to Knowledge Arguments* . Cambridge: Cambridge University Press.

Nagel, E. (1961). *The Structure of Science.* London: Routledge and Kegan Paul.

Nagel, T. (1974). 'What is it like to be a bat?' . *Philosophical Review, 83*, 435-456.

Nagel, T. (1979). *Mortal Questions.* Cambridge: Cambridge University Press.

Nagel, T. (1986). *The View from Nowhere.* (S. Kronlund, Trad.) Oxford: Oxford University Press.

Nagel, T. (1987). *What Does It All Mean? - a very short Introduction to Philosophy.* Oxford: Oxford University Press.

Nelkin, N. (1989). 'Unconscious sensations'. *Philosophical Psychology, 2*, 129-41.

Nemirow, L. (1990). 'Physicalism and the Cognitive Role of Acquaintance'. Dans W. Lycan (Éd.), *Mind and Cognition: A Reader.* Oxford: Blackwell.

Nemirow, L. (2007). 'So This Is What it's Like: a Defense of the Ability Hypothesis'. Dans T. Alter, & S. Walter (Éds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism.* Oxford: Oxford University Press.

Neurath, O. (1931b). ‘Physikalismus’. *Scientia 50*, 297-303.

Neurath, O. (1946). ‘The Orchestration of the Sciences by the Encyclopedism of Logical Empiricism’. *Philosophy and Phenomenological Research, 6*, 496-508.

Newton-Smith, W. (Éd.). (2001). *A Companion to the Philosophy of Science.* Blackwell Publishers.

Nida-Rümelin, M. (1996). 'What Mary couldn't know’. Dans T. Metzinger (Éd.), *Phenomenal Consciousness.* Schoenigh: Paderborn.

Nida-Rümelin, M. (1998). ‘On Belief About Experiences: An Epistemological Distinction Applied to the Knowledge Argument’ . *Philosophy and Phenomenological Research 58 (1)*, 51–73.

Nida-Rümellin, M. (s.d.). *‘Qualia : The Knowledge Argument’.* http://plato.stanford.edu/entries/qualia-knowledge.

Noë, A., & Thompson, E. (2004). 'Are there neural correlates of consciousness?'. *Journal of Consciousness Studies 11, 3-28*, 29-86.

Noë, A., & Thompson, E. (2004). Sorting Out the Neural Basis of Consciousness. *Journal of Consciousness Studies 11 (1)*, 87-98.

Nordby, K. (s.d.). *‘Vision in a Complete Achromat: A Personal Account’.* http://consc.net/misc/achromat.html.

Nordby, K. (2007). ‘What is this thing you call color? Can a totally color blind person know about color? Dans T. Alter, & S. Walte (Éds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism.* Oxford: Oxford University Press.

O'Connor, T., & Robb, D. (Éds.). (2003). *Philosophy of Mind - Contemporary Readings.* New York: Routledge.

Oppenheim, P., & Putnam, H. (1958). ‘The unity of science as a working hypothesis'. Dans H. Feigl (Éd.), *Minnesota Studies in the Philosophy of Science* (Vol. 2). Minneapolis: Minnesota University Press.

Pacherie, E. (1993). *Naturaliser l'intentionnalité.* Paris: PUF.

Papineau, D. (1993). ‘Physicalism, Consciousness, and the Antipathetic Fallacy’. *Australasian Journal of Philosophy 71*, 169–83 .

Papineau, D. (1995). ‘The antipathetic fallacy and the boundaries of consciousness’. Dans T. Metzinger (Éd.), *Conscious Experience.* Paderborn: Ferdinand Schöningh.

Papineau, D. (Éd.). (1996). *The Philosophy of Science.* Oxford: Oxford University Press.

Papineau, D. (2003). ’Theories of Consciousness’. Dans A. Jokic, & Q. Smith (Éds.), *Consciousness: New Philosophical Perspectives.* Oxford: Oxford University Press.

Papineau, D. (2007). 'Phenomenal and perceptual concepts'. Dans T. Alter, & S. Walter (Éds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism.* Oxford: Oxford University Press.

Papineau, D. (2008). ‘Must a Physicalist be a Microphysicalist?’ . Dans J. Hohwy, & J. Kallestrup (Éds.), *Being Reduced: New Essays on Reduction, Explanation, and Causation.* Oxford: Oxford University Press.

Parrochia, D. (1997). *Les grandes révolutions scientifiques du XXe siècle.* Paris: PUF.

Peacocke, C. (2008). *Truly Understood.* Oxford: Oxford Univeristy Press.

Peat, D. F. (2002). *From Certainty to Uncertainty - The Story of Science and Ideas in the Twentieth Century.* Whashington D.C.: Joseph Henry Press.

Peirce, C. (1965). *Collected Papers* (Vol. I-VI). Cambridge: The Belknap Press of Harvard University Press.

Perry, J. (2001). *Knowledge, Possibility and Consciousness.* Cambridge, MA: MIT Press.

Peschard, I., & Bitbol, M. (2008). ‘Heat, Temperature and Phenomenal Concepts’. Dans E. Wright (Éd.), *The Case for Qualia.* Cambridge Mass: MIT Press.

Petitot, J., Varela, F., Pachoud, B., & Roy, J.-M. (Éds.). (1999). *Naturalizing Phenomenology: Issues in Contemporary Phenmenology and Cognitive Science.* Stanford, California: Stanford University Press.

Pettit, P. (2004). ‘Motion Blindness and the Knowledge Argument’. Dans P. Ludlow, & al (Éds.), *There is something about Mary.* Cambridge Mass: MIT Press.

Pettit, P. (2008). 'Consciousness and the Frustrations of Physicalism'. Dans I. Ravenscroft (Éd.), *Minds, Ethics, and Conditionals Themes from the Philosophy of Frank Jackson.* Oxford: Clarendon Press.

Piattelli-Palmarini, M. (Éd.). (1979). *Théories du langage - Théories de l'apprentissage.* Paris: Ed. du Seuil.

Pinkas, D. (1995). *La matérialité de l'esprit.* Paris: Ed. La Découverte.

Pitcher, G. (1971). *A Theory of Perception.* Princeton, N.J.: Princeton University Press.

Place, U. (1956). 'Is Consciousness a Brain Process?'. *British Journal of Psychology, 47*, 44-50.

Pollock, J. L. (1984). *The Foudations of Philosophical Semantics.* Princeton University Press.

Prigogine, I., & Stengers, I. (1979). *La nouvelle alliance.* Paris: Gallimard.

Prior, A. N. (1962). 'The Formalities of Omniscience'. *Philosophy 37*, 114–129.

Proust, J. (1997). *Comment l'esprit vient aux bêtes.* Paris: Gallimard.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth.* London: Routledge.

Putnam, H. (1965). 'Brains and Behavior'. Dans R. J. Butler (Éd.), *Analytical Philosophy.* Oxford: Blackwell.

Putnam, H. (1967). 'Psychological Predicates'. Dans W. Capitan, & D. Merrill (Éds.), *Art, Mind and Religion.* Pittsburg: University of Pittsburg Press.

Putnam, H. (1971). *Philosophy of Logic.* New York: Harper and Row.

Putnam, H. (1975). 'The Meaning of 'Meaning' . Dans H. Putnam, *Mind, Language, and Reality* (Vol. 2). Cambridge: Cambridge University Press.

Putnam, H. (1975). *Mind, Language, and Reality.* Cambridge: Cambridge University Press.

Putnam, H. (1981). *Reason, Truth, and History.* Cambridge: Cambridge University Press.

Putnam, H. (1983). 'Why There isn't a Ready Made World'. Dans H. Putnam, *Realism and Reason* (Vol. 3). Cambridge: Cambridge University Press.

Putnam, H. (1988). *Representation and Reality.* MIT Press.

Putnam, H. (1990). *Realism with a Human Face.* Cambridge Massachusetts: Harvard University Press.

Putnam, H. (1999). *The Threefold Cord - Mind, Body, and World.* New York: Columbia University Press.

Quine, W. (1948). 'On What There Is'. *Review of Metaphysics 2 (1948)*, 21–38.

Quine, W. (1951a). 'Two Dogmas of Empiricism'. *Philosophical Review 60*, pp. 20-43.

Quine, W. (1953). *From a Logical Point of View.* Cambridge Massachusetts: Harvard University Press.

Quine, W. (1960). *Word and Object.* Cambridge Massachusetts: MIT Press.

Quine, W. (1969). *Ontological Relativity and Other Essays.* (J. Largeault, Trad.) New York & London: Columbia University Press.

Raftopoulos, A., & Müller, V. C. (2006). 'The phenomenal content of experience'. *Mind and Language, 21(2)*, 187–219.

Ravenscroft, I. (Éd.). (2008). *Minds, Ethics, and Conditionals Themes from the Philosophy of F. Jackson.* Oxford: Clarendon Press.

Recanati, F. (1979). *La transparence de l'énonciation.* Paris: Ed. du Seuil.

Recanati, F. (2008). *Philosophie du langage (et de l'esprit).* Paris: Gallimard.

Rey, G. (1987). *Contemporary Philosophy of Mind: A Contentiously Classical Approach.* Oxford: Blackwell.

Rey, G. (1995). 'Toward a Projectivist Account of Conscious Experience'. Dans T. Metzinger (Éd.), *Conscious Experience.* Paderborn: Ferdinand Schöningh.

Rivenc, F. (1993). *Recherches sur l'universalisme logique.* Paris: Payot.

Rivenc, F. (1998). *Sémantique et vérité de Tarski à Davidson.* Paris: PUF.

Robinson, H. (1982). *Matter and Sense: A Critique of Contemporary Materialism.* Cambridge: Cambridge University Press.

Robinson, W. (1988). *Brains and People: An Essay on Mentality and its Causal Conditions.* Philadelphia: Temple University Press.

Rorty, R. (1979). *Philosophy and the Mirror of Nature.* Princeton University Press.

Rosenthal, D. (1991). ‘The independence of consciousness and sensory quality’. Dans E. Villanueva (Éd.), *Consciousness* . Atascadero, CA: Ridgeview Publishing.

Russell, B. (1927). *The Analysis of Matter.* London: Kegan Paul .

Russell, B. (1948). *Human Knowledge - Its Scope and Limits.* London: George Allen and Unwin LTD.

Ryle, G. (1949). *The Concept of Mind.* (S. Stern-Gillet, Trad.) London: Hutchinson Publ.

Sacks, M. (2000). *Objectivity and Insight.* Oxford: Clarendon Press.

Sacks, O., & Wasserman, R. (1987). ‘The Case of the Colorblind Painter’. *The New York Review of Books, 34,* 193-226.

Sarkar, S. (1998). *Genetics and Reductionism.* New York: Cambridge University Press.

Schaffer, J. (2003). ‘Is There a Fundamental Level?’ . *Noûs 37 (3),* 498–517.

Schlick, M. (1934). 'Über das Fundament der Erkenntnis’. Dans C. Bonnet, & P. Wagner (Éds.), *L’âge d’or de l’empirisme logique.* Paris: Gallimard.

Seager, W. (1995). ‘Consciousness, information, and panpsychism’. *Journal of Consciousness Studies, 2,* 272-88.

Seager, W. (2000). ‘Physicalism’ . Dans W. Newton-Smith (Éd.), *A Companion to the Philosophy of Science.* Blackwell Publishing.

Searle, J. (1969). *Speech Acts.* Cambridge: Cambridge University Press.

Searle, J. (1979). *Expression and Meaning.* (J. Proust, Trad.) Cambridge: Cambridge University Press.

Searle, J. (1983). *Intentionality - An Essay in the Philosophy of Mind.* Cambridge: Cmabridge University Press.

Searle, J. (1992). *The Rediscovery of Mind.* Cambridge Massachusetts: MIT Press.

Searle, J. (1997). *The Mystery Of Consciousness.* New York: New York Review.

Sebestik, J., & Soulez, A. (Éds.). (1986). *Le Cercle de Vienne - Doctrines et Controverses.* Paris: Meridiens Klincksieck.

Shapiro, L. (2000). ‘Multiple Realizations’. *Journal of Philosophy, 97,* 635-654.

Shoemaker, S. (1975). ‘Functionalism and Qualia’. *Philosophical Studies, 27,* 291-315.

Shoemaker, S. (1982). ‘The inverted spectrum’. *Journal of Philosophy 79*, 357–81.

Shoemaker, S. (1990). 'Qualities and qualia: what's in the mind’. *Philosophy and Phenomenological Research, Supplement, 50*, 109-131.

Shoemaker, S. (1996). *The First Person Perspective and Other Essays.* Cambridge: Cambridge University Press.

Shoemaker, S. (2007). *Physical Realisation.* Oxford: Oxford University Press.

Smart, J. (1959). ‘Sensations and Brain Processes’. *Philosophical Review, 68*, 141-156.

Smart, J. (1963). ‘Materialism’. *Journal of Philosophy, 60*, 651-662.

Smart, J. (s.d.). *The Identity Theory of Mind.* http://plato.stanford.edu/entries/mind-identity.

Smith, B. C. (2006). ‘Consciousness: An Inner View of the Outer World’. *Journal of Consciousness Studies, 13, No. 7–8*, 175–86.

Smith, B. C. (Éd.). (2006). *Questions of Taste, The Philosophy of Wine.* London: Signal Books.

Smith, D., & Thomasson, A. (Éds.). (2005). *Phenomenology and Philosophy of Mind.* Oxford and New York: Oxford University Press.

Smith, Q. (2003). ‘Why Cognitive Scientist Cannot Ignore Quantum Mechanics’. Dans Q. Smith, & A. Jokic (Éds.), *Consciousness - New Philosophical Perspectives.* Oxford: Clarendon Press.

Smith, Q., & Jokic, A. (Éds.). (2003). *Consciousness - New Philosophical Perspectives.* Oxford: Clarendon Press.

Snell, B. (1994). *La découverte de l'esprit - la genèse de la pensée européenne chez les Grecs.* Editions de l'Eclat .

Soames, S. (2009). ‘Ontology, Analyticity, and Meaning: The Quine-Carnap Dispute’. Dans D. Chalmers, D. Manley, & R. Wasserman (Éds.), *Metametaphysics: New Essays on the Foundations of Ontology.* Oxford: Clarendon Press.

Sosa, E. (2009). *Reflective Knowledge- Apt Belief and Reflective Knowledge* (Vol. II). Oxford: Clarendon Press.

Soulez, A. (Éd.). (1985). *Manifeste du Cercle de Vienne et autres écrits.* Paris: PUF.

Spelke, E. (1994). ‘Initial Knowledge: six suggestions’. *Cognition, 50*, 443-47.

Stanley, J., & Williamson, T. (2001). ‘Knowing how’. *Journal of Philosophy 98*, 1–40.

Stapp, H. (1993). *Mind, Matter, and Quantum Mechanics.* Berlin: Springer-Verlag.

Stapp, H. (1993). 'A quantum theory of the mind-brain interface'. *In Mind, Matter, and Quantum Mechanics, Springer*, 145-172.

Stich, S., & Warfield (Éds.). (2003). *The Blackwell Guide to the Philosophy of Mind.* Blackwell, 2003.

Stich, S., & Warfield, T. (Éds.). (2003). *The Blackwell Guide to the Philosophy of Mind.* Blackwell Publishing.

Stoljar, D. (s.d.). ‘Physicalism’ . Stanford Encyclopedia of Philosophy http://plato.stanford.edu .

Stoljar, D. (2001). ‘Two Conceptions of the Physical’. *Philosophy and Phenomenological Research, 62*, 253-281.

Stoljar, D. (2005). ‘Physicalism and Phenomenal Concepts’. *Mind and Language, 20*, 469-494.

Stoljar, D. (2006). *Ignorance and Imagination - The Epistemic Origin of the Problem of Consciousness.* Oxford: Oxford University Press.

Stoljar, D. (2010). *Physicalism.* London & New York: Routledge.

Strawson, G. (1994). *Mental Reality.* Cambridge Mass: MIT Press.

Strawson, G. (2000). ‘Realistic Materialist Monism’. Dans S. Hameroff, A. Kaszniak, & D. Chalmers (Éds.), *Toward a Science of Consciousness* (Vol. 3). Cambridge, MA: MIT Press.

Stroud B. (2009). The Transparency of ‘Naturalism’. *The Romanell Lecture, Proceedings and Addresses of the American Philosophical Association*, 157-169.

Stroud, B. (1986). ‘The Physical World’. *Proceedings of the Aristotelian Society, 87*, 263–277.

Stroud, B. (1996). 'The Charm of Naturalism'. *Proceedings of the American Philosophical Association 70*, 43-55. .

Stroud, B. (2000). *Understanding Human Knowledge.* Oxford: Oxford University Press.

Sturgeon, S. (2001). ‘The Roots of Reductionism'. Dans C. Gillett, & B. M. Loewe (Éds.), *Physicalism and its Discontents.* Cambridge: Cambridge University Press.

Swinburne, R. (1986). *The Evolution of the Soul.* Oxford: Oxford University Press.

Tarski, A. (1974). *Logique, sémantique, métamatématique* (Vol. II). Paris: Armand Colin.

Thagard, P. (Éd.). (2006). *Philosophy of Psychology and Cognitive Science.* North-Holland.

Thom, R. (1990). *Apologie du Logos.* Paris: Hachette.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind.* Harvard University Press.

Tiercelin, C. (1993). *La pensée-signe. Etudes sur C.S. Peirce.* Nîmes: Ed. Jacqueline Chambon.

Toulmin, S. (1958). *The Uses of Argument.* Cambridge: Cambridge University Press.

Turing, A. (1937). ‘Computability and λ-Definability’. *The Journal of Symbolic Logic, 2*, 153–163.

Tye, M. (1995). *Ten Problems of Consciousness.* Cambridge, MA: MIT Press.

Tye, M. (2000). *Consciousness, Color, and Content.* Cambridge Mass.: MIT Press.

Tye, M. (2003). *Consciousness and Persons - Unity and Identity.* Cambridge Massachusetts: MIT Press.

Tye, M. (2009). *Consciousness Revisited - Materialism whitout Phenomenal Concept.* Cambridge Mass.: MIT Press.

Varela, F. (1988). *Connaître: Les Sciences Cognitives , tendances et perspectives.* Paris: Ed. du Seuil.

Varela, F., & Shear, J. (1999). *The View from Within. First Person Approaches to the Study of Consciousness.* Thorverton, UK: Imprint Academic.

Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind - Cognitive Science and Human Experience.* Cambridge Massachusetts: MIT Press.

Velmans, M. (2002). 'How could conscious experiences affect brains?'. *Journal of Consciousness Studies 9(11)*, 3-29.

Velmans, M., & Schneider, S. (Éds.). (2007). *The Blackwell Companion to Consciousness.* Blackwell Publishing.

Vuillemin, J. (1971). *La logique et le monde sensible.* Paris: Flammarion.

Wagner, P. (Éd.). (2002). *Les philosophes et la science.* Paris: Gallimard.

Waismann, F. (1965). *Principles of Linguistic Philosophy.* (R. Harré, Éd.) London: Macmillan.

Warner, R. (1986). 'A Challenge to Physicalism'. *Australasian Journal of Philosophy 64*, 249–265.

Warner, R., & Szubka, T. (Éds.). (1994). *The Mind-Body Problem - A Guide to the Current Debate.* Cambridge Massachusetts: Blackwell.

Wheeler, J. (1994). 'It from bit'. *In At Home in the Universe, American Institute of Physics*, 295-311.

Wheeler, J., & Zurek, W. (Éds.). (1983). *Quantum Theory and Measurement.* Princeton: Princeton University Press.

White, S. (1986). 'Curse of the Qualia'. *Synthese 68*, 333–68.

White, S. (1991). *The Unity of the Self.* Cambridge: MIT Press.

White, S. (2007). 'Property Dualism, Phenomenal Concepts, and the Semantic Premise'. Dans T. Alter, & S. Walter (Éds.), *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism.* Oxford: Oxford University Press.

Whitehead, A. (1929). *Process and Reality – an Essay in Cosmology.* Cambridge: Cambridge University Press.

Williamson, T. (2000). *Knowledge and its Limits.* Oxford: Oxford University Press.

Williamson, T. (2007). *The Philosophy of Philosophy.* Blackwell Publishing.

Wilson, R., & Keil, F. (Éds.). (1999). *The MIT Encyclopedia of the Cognitive Science.* Cambridge Massachusetts: MIT Press.

Woodward, J. (2003). *Making Things Happen - a Theory of Causal Explanation.* Oxford: Oxford University Press.

Yates, F. (1964). *Giordano Bruno and the Hermetic Tradition.* University of Chicago Press.

Zahar, E. (2000). *Essai d'épistémologie réaliste.* Paris: Vrin.

Zahavi, D. (2001b). Beyond empathy. phenomenological approaches to intersubjectivity. *Journal of Consciousness Studies 8/5-7, 2001*, 151-167.

Index lexical

A

A posteriori · 36, 40, 69, 121, 162, 213, 220, 240, 241, 283, 284, 285, 346, 347, 349, 350, 352, 355, 357, 358, 367, 376
A priori · 26, 31, 40, 41, 69, 79, 80, 81, 83, 104, 108, 133, 161, 163, 192, 193, 201, 210, 213, 214, 220, 234, 241, 246, 261, 266, 285, 293, 294, 297, 299, 300, 301, 306, 310, 321, 335, 340, 341, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 373, 376, 377, 383, 391, 392, 393, 396, 397, 399, 426
Abstraction · 57, 99, 198, 376, 400
Abstrait · 43, 75, 84, 173, 174, 263, 379, 383, 390, 397
Accointance · 247, 256, 257, 258, 259, 262, 264, 265, 266, 281, 282, 283, 287, 290
Acquisition · 247, 248, 250, 252, 259, 264, 265, 269, 270, 288, 290, 302, 316, 317, 326, 327, 331, 339, 344
Actualiste · 99, 100, 101, 103
Actuel · 11, 28, 30, 35, 36, 39, 40, 42, 44, 55, 70, 77, 86, 89, 90, 98, 99, 100, 101, 102, 103, 104, 109, 110, 111, 125, 127, 147, 205, 215, 217, 223, 282, 317, 329, 330, 336, 337, 341, 349, 350, 351, 354, 356, 372, 395, 397, 398, 400, 402, 404
Analyse conceptuelle · 36, 40, 119, 138, 210, 319, 343, 345, 354, 355, 357, 360, 362, 365, 367, 382, 404
Analytique · 30, 39, 41, 44, 56, 62, 67, 68, 79, 86, 87, 90, 103, 108, 114, 115, 116, 123, 125, 127, 130, 133, 139, 142, 143, 145, 151, 152, 154, 162, 166, 169, 172, 173, 174, 177, 180, 206, 234, 249, 282, 318, 341, 357, 362, 372, 381, 382, 385, 388, 426
Antidualiste · 150, 152, 157
Antipsychologisme · 29, 110, 113, 115, 138
Antisupernaturaliste · 26, 33, 373, 394
Apprentissage · 40, 234, 246, 248, 302, 308, 310, 327, 339
Apriorité · 299, 300, 348, 359, 360, 362, 363, 372, 382, 397, 426
Aptitude · 248, 249, 250, 252, 253, 254, 255, 256, 258, 317, 339
Aqueux · 352, 356, 358, 360, 361, 368
Argument de la concevabilité · 199, 200, 203, 215, 320, 360
Argument de la connaissance · 11, 24, 26, 39, 40, 42, 43, 45, 60, 61, 103, 107, 111, 114, 115, 153, 163, 180, 181, 193, 194, 205, 206, 219, 230, 231, 233, 234, 236, 242, 245, 255, 265, 278, 279, 293, 294, 295, 296, 301, 306, 310, 313, 318, 319, 320, 321, 322, 323, 325, 327, 335, 337, 339, 340, 341, 346, 347, 352, 372, 392, 417, 424
Assertion · 76, 89, 135, 137, 271, 368
Atome · 57, 65, 71, 93, 96, 101, 106, 130, 235, 399, 403
Atomisme · 63, 96, 397, 398, 400, 403
Attribut · 52, 65, 147, 161, 203, 220, 260, 269, 278, 296, 297, 300, 301, 302, 336, 352, 395, 426
Axe · 11, 29, 31, 32, 35, 42, 110, 116, 123, 128, 142, 143, 246, 310, 340, 372, 381, 391, 397, 402, 407, 412, 425

B

Behaviorisme · 37, 153, 154, 155, 156, 172, 174, 206
Big Bang Gouffre · 32, 34, 373, 420
Boîte de Pandore · 55, 129, 141, 343
Boomerang · 53, 109, 319, 320

C

Cadre · 36, 41, 54, 58, 80, 116, 141, 194, 212, 373, 382, 384, 387, 388, 389, 391

Catégorie · 117, 118, 205, 206, 210, 222, 226, 248, 265, 266, 315, 316, 377, 379, 380, 387, 395, 401, 402

Causalité · 60, 76, 97, 114, 152, 170, 301

Cause · 27, 51, 69, 73, 75, 76, 77, 114, 125, 148, 152, 156, 159, 161, 162, 163, 167, 168, 183, 190, 192, 195, 201, 202, 211, 225, 249, 260, 263, 269, 274, 280, 286, 287, 290, 291, 296, 301, 305, 346, 348, 361, 397, 417, 426

Cercle de Vienne · 26, 59, 62, 111, 124, 130, 392, 394

Certitude · 53, 187, 278, 374, 381

Cerveau · 35, 36, 38, 51, 77, 86, 123, 135, 158, 159, 160, 162, 164, 165, 166, 172, 173, 176, 187, 201, 204, 238, 254, 268, 277, 278, 282, 285, 331, 366, 376, 423

Chauve-souris · 181, 234, 236, 237, 238, 247

Chose en soi · 380, 381

Clivage · 11, 31, 33, 34, 35, 36, 54, 121, 122, 123, 126, 128, 143, 191, 345, 364, 370, 371, 373, 375, 376, 379, 391, 392, 405, 412, 421, 425, 426

Clôture Causale · 27, 63, 73, 78, 114, 162, 179, 180, 221, 224, 227, 394, 426

Cogito · 132, 140, 150, 374, 375, 380

Cognition · 27, 52, 97, 149, 154, 157, 182, 197, 249, 314, 416, 425

Communicable · 319, 325, 327, 328, 331, 332, 333, 334, 336, 337, 417

Complétude · 27, 40, 44, 72, 73, 75, 77, 81, 115, 193, 290, 313, 347, 351, 399, 426

Comportement · 99, 102, 145, 153, 154, 155, 156, 159, 174, 175, 186, 191, 192, 195, 196, 197, 207, 208, 209, 211, 229, 235, 313, 355

Computationnel · 60, 169, 172, 173, 174, 177, 178, 196, 205, 267, 298

Concept Phénoménal · 191, 205, 213, 214, 225, 246, 251, 259, 260, 261, 262, 264, 268, 269, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290

Conjonction · 71, 146, 149, 294, 316, 334, 349, 354, 358

Connaissance · 23, 30, 31, 34, 35, 40, 41, 53, 55, 59, 60, 95, 114, 125, 127, 128, 130, 131, 135, 142, 150, 153, 155, 156, 181, 183, 187, 206, 225, 233, 234, 235, 238, 242, 244, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 264, 265, 266, 267, 271, 274, 279, 280, 281, 282, 283, 287, 288, 289, 290, 291, 292, 293, 301, 302, 310, 317, 318, 319, 320, 322, 326, 332, 339, 340, 354, 355, 357, 359, 364, 366, 371, 375, 377, 378, 419

Conscience · 24, 39, 41, 43, 53, 54, 111, 115, 118, 119, 120, 123, 124, 126, 140, 146, 150, 158, 179, 181, 184, 186, 187, 190, 192, 193, 197, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 224, 225, 226, 227, 228, 229, 230, 291, 293,妍294, 297, 299, 305, 306, 310, 315, 337, 339, 352, 360, 377, 416

Conscience Phénoménale · 39, 43, 54, 193, 197, 299, 305, 306, 337, 352, 360

Contenu · 23, 35, 45, 58, 60, 68, 89, 105, 110, 113, 115, 117, 118, 119, 120, 123, 124, 126, 127, 129, 130, 131, 132, 134, 135, 137, 139, 140, 141, 142, 143, 150, 159, 183, 184, 185, 192, 196, 198, 202, 214, 233, 252, 263, 269, 270, 271, 272, 273, 274, 275, 281, 287, 294, 296, 297, 298, 299, 303, 304, 305, 311, 315, 319, 343, 355, 363, 379, 383, 403, 412

Contingent · 40, 83, 86, 89, 90, 91, 162, 213, 220, 241, 315, 327, 347, 352, 367, 426

Corps · 31, 33, 36, 79, 101, 114, 134, 140, 145, 146, 147, 149, 150, 151, 153, 160, 169, 174, 280, 363, 375, 380, 386, 424, 426

Critique · 30, 34, 40, 41, 62, 72, 80, 90, 113, 116, 122, 149, 150, 151, 155, 162, 168, 169, 171, 193, 195, 205, 214, 223, 228, 234, 237, 242, 244, 247, 250, 253, 256, 277, 282, 283, 290, 303, 327, 340, 343, 344, 348, 359, 360, 362, 367, 372, 378, 380, 381, 383, 392, 394, 400, 402, 403, 407

Croyance · 51, 61, 155, 198, 209, 218, 225, 258, 270, 273, 274, 275, 276, 286, 288, 289, 301, 355

D

De dicto · 258, 364
De Re · 258, 364
De Se · 41, 319, 367
Déduction · 135, 139, 295, 298, 300, 301, 345, 346, 391
Définition · 27, 28, 34, 38, 43, 54, 58, 63, 69, 71, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 94, 95, 96, 97, 99, 100, 101, 102, 106, 107, 108, 110, 131, 145, 160, 171, 172, 175, 181, 184, 196, 197, 203, 206, 223, 255, 256, 257, 295, 301, 313, 320, 322, 324, 325, 327, 341, 348, 379, 382, 394, 395, 424, 425
Démiurge · 36, 318, 418
Démonstratif · 167, 251, 257, 269, 273, 280, 281, 284
Dérivabilité · 68, 191, 300, 345, 347
Désignateur Rigide · 89, 90, 241, 352
Désir · 51, 152, 155, 163, 197, 225
Dichotomie · 30, 33, 36, 37, 120, 121, 122, 128, 146, 369, 371, 376, 381, 391
Dieu · 41, 54, 199, 311, 323, 324, 332, 333, 347, 365, 366, 367, 418, 419
Dilemme d'Hempel · 29, 98, 102, 103, 104, 106, 216, 336
Dimension · 94, 115, 159
Discours · 30, 41, 108, 109, 110, 123, 126, 128, 129, 131, 133, 134, 135, 136, 137, 139, 140, 141, 147, 320, 348, 365, 367, 368, 390, 394
Disjonction · 168, 226, 256
Disposition · 40, 42, 79, 121, 159, 175, 184, 198, 209, 227, 246, 279, 284, 290, 291, 294, 310, 328, 367, 404
Divin · 311, 318, 324, 325, 332, 334, 366, 417, 418, 419
Dogme · 41, 42, 113, 125, 336, 344, 359, 380, 383, 385, 393, 397, 398, 400, 407
Douleur · 38, 52, 53, 79, 86, 121, 153, 156, 157, 158, 161, 162, 163, 164, 165, 166, 167, 168, 176, 197, 203, 204, 225, 251, 283, 285, 348, 366, 401
Dualisme · 30, 32, 45, 82, 88, 89, 95, 101, 147, 150, 151, 152, 159, 177, 190, 208, 219, 221, 222, 224, 226, 227, 229, 259, 268, 276, 277, 286, 331, 344, 372, 373, 375
Dynamique · 28, 29, 84, 121, 125, 185, 215, 216, 217, 218, 219, 223, 227, 371

E

Effet-que-cela-fait · 179, 181, 182, 183, 184, 192, 197, 199, 203, 236, 237, 238, 250, 251, 252, 253, 254, 255, 256, 260, 265, 266, 267, 269, 270, 271, 272, 273, 287, 288, 325, 327
Éliminativisme · 147, 171, 206
Émergence · 160, 174
Empirisme · 36, 41, 60, 62, 108, 124, 127, 138, 150, 375, 383, 387, 390, 391, 392
Empiristes Logiques · 26, 109, 129, 141, 372, 373, 420
Énoncé de Ramsey · 329, 330
Énoncé Protocolaire · 109, 110, 126, 128, 129, 130, 132, 136, 137, 138, 139, 140, 141, 142, 341, 346, 348, 392
Entité · 28, 39, 58, 64, 82, 119, 138, 202, 203, 210, 222, 310, 323, 347, 356, 375, 384, 385, 388, 407
Épiphénoménalisme · 74, 221, 224, 225, 226, 228, 403
Épistémique · 43, 59, 114, 131, 187, 190, 191, 206, 210, 211, 213, 214, 215, 219, 240, 242, 243, 269, 280, 281, 287, 288, 289, 299, 300, 302, 308, 309, 314, 317, 318, 319, 336, 351, 355, 364, 369, 372, 376, 389
Épistémologie · 62, 109, 375, 386, 394
Esprit · 9, 23, 24, 27, 30, 31, 33, 35, 36, 37, 40, 41, 42, 43, 44, 45, 52, 53, 54, 55, 60, 77, 88, 109, 110, 111, 113, 114, 115, 116, 118, 122, 124, 125, 126, 133, 140, 142, 143, 145, 146, 147, 148, 150, 151, 154, 155, 156, 157, 160, 164, 169, 172, 173, 174, 176, 178, 179, 182, 185, 189, 193, 194, 204, 206, 212, 213, 223, 224, 228, 229, 233, 234, 235, 254, 265, 269, 283, 314, 318, 322, 332, 337, 341, 346,

347, 374, 375, 377, 380, 386, 392, 394, 395, 398, 401, 416, 421, 424, 425, 426, 427
État Mental · 33, 34, 37, 42, 51, 52, 53, 55, 74, 77, 88, 110, 128, 135, 136, 142, 146, 154, 156, 157, 159, 160, 162, 165, 166, 167, 171, 172, 174, 175, 176, 177, 178, 181, 184, 185, 186, 187, 196, 198, 201, 224, 250, 258, 296, 305, 349, 366, 371, 380, 389
État Physique · 11, 27, 35, 42, 51, 52, 74, 88, 128, 131, 146, 152, 161, 162, 163, 167, 186, 210, 211, 212, 213, 214, 220, 222, 224, 227, 229, 285, 305, 345, 349, 372, 380
États Phénoménaux · 210, 211, 212, 213, 214, 222, 224, 225, 229, 252, 284, 285, 286, 289, 298, 345, 372
Événement Physique · 73, 152, 159, 162, 167
Expérience · 34, 35, 37, 38, 39, 44, 51, 53, 54, 57, 62, 73, 99, 100, 101, 109, 110, 113, 114, 120, 123, 125, 127, 129, 130, 131, 132, 135, 137, 139, 140, 141, 142, 143, 145, 153, 155, 159, 161, 162, 173, 175, 176, 177, 178, 180, 181, 182, 183, 187, 189, 191, 192, 193, 194, 195, 197, 199, 204, 205, 207, 208, 209, 210, 214, 225, 226, 228, 234, 235, 237, 238, 241, 244, 246, 249, 250, 251, 253, 254, 255, 256, 257, 258, 259, 260, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 284, 285, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 301, 302, 303, 304, 305, 306, 310, 313, 323, 324, 325, 326, 331, 332, 340, 341, 347, 355, 363, 375, 376, 377, 378, 385, 394, 419, 422
Expérience Sensible · 376, 394
Expérientiel · 24, 111, 199, 262, 303, 313, 314, 316, 331, 333
Explanandum · 58, 76, 189, 207, 208, 209, 215, 220, 400
Explanans · 58, 76, 189
Explication · 24, 27, 34, 39, 44, 53, 61, 68, 69, 76, 104, 111, 116, 145, 146, 149, 162, 163, 173, 176, 177, 186, 187, 188, 191, 193, 194, 197, 198, 202, 203, 205, 206, 207, 208, 209, 210, 215, 216, 220, 241, 242, 282, 284, 285, 295, 297, 299, 301, 302, 310, 353, 354, 360, 396, 400, 424

F

Fait extralinguistique · 135, 368
Faits subjectifs · 42, 314, 315, 339, 417, 425
Fonction · 39, 64, 65, 69, 76, 82, 101, 155, 164, 173, 175, 187, 189, 192, 194, 195, 196, 204, 209, 216, 218, 222, 223, 225, 226, 247, 256, 267, 277, 315, 340, 354, 355, 357, 360, 364, 377, 381, 395, 403, 422
Fonctionnalisme · 37, 74, 87, 164, 165, 169, 171, 172, 173, 175, 176, 177, 178, 184, 186, 192, 196, 198, 203, 206, 392
Fondement · 35, 90, 104, 139, 140, 150, 211, 319, 350, 385, 397
Formalisme · 109, 127, 386
Forme · 23, 30, 37, 51, 57, 59, 60, 66, 68, 71, 73, 74, 76, 84, 88, 90, 94, 102, 105, 114, 118, 121, 133, 140, 146, 147, 150, 154, 167, 171, 180, 182, 189, 190, 206, 212, 223, 228, 235, 241, 244, 245, 251, 284, 295, 296, 304, 306, 322, 323, 324, 326, 329, 331, 332, 333, 334, 335, 336, 346, 353, 373, 376, 377, 378, 379, 380, 383, 389, 391, 396, 404, 417, 418
Fossé Explicatif · 11, 34, 35, 37, 38, 43, 45, 121, 122, 126, 131, 187, 190, 191, 192, 205, 206, 207, 214, 216, 281, 310, 367, 369, 370, 421, 426

G

Glande pinéale · 31, 37, 120, 147, 375
Gouffre · 35, 39, 122, 187, 188, 190, 191, 210, 213, 214, 310, 322, 364, 369, 392

H

H_2O · 38, 61, 70, 99, 162, 163, 188, 191, 201, 202, 203, 210, 211, 212, 213, 287, 349, 351, 352, 355, 356, 357, 358, 360, 361, 368, 369, 403

Habilité · 246, 248, 249, 250, 251, 252, 253, 254, 255, 327
Humilité ramseyenne · 192, 327, 328, 331, 334, 417

I

Identité · 37, 68, 73, 77, 81, 86, 90, 104, 147, 157, 158, 159, 161, 162, 164, 165, 166, 167, 169, 176, 178, 188, 190, 201, 202, 203, 210, 211, 212, 213, 220, 241, 269, 272, 275, 283, 284, 285, 313, 352, 362, 385, 392, 403
Identité Psychophysique · 37, 77, 86, 147, 158, 159, 161, 162, 164, 166, 167, 169, 178, 392
Ignorance · 104, 192, 221, 228, 234, 238, 331, 333
Imagination · 252, 253, 269, 377, 378
Immatérialité · 36, 376, 392, 425
Impasse · 83, 91, 101, 102, 126, 147, 158, 163, 165, 213, 218, 264, 304, 389, 390
Indexical · 137, 255
Ineffabilité · 54, 295, 303
Inférence · 102, 189, 240, 246, 250, 272, 294, 297, 317, 351, 352, 357, 364
Information · 23, 26, 39, 159, 173, 174, 175, 186, 192, 229, 234, 236, 242, 246, 253, 257, 271, 274, 275, 277, 279, 290, 295, 310, 313, 315, 316, 317, 318, 321, 326, 327, 328, 331, 334, 335, 337, 339, 352, 355, 357, 363, 417, 419
Information Physique · 23, 26, 39, 192, 234, 236, 242, 246, 271, 275, 277, 279, 290, 295, 310, 313, 315, 316, 318, 321, 326, 327, 328, 331, 334, 335, 337, 339, 417, 419
Input · 173, 174, 204, 422
Intelligence Artificielle · 174
Intelligibilité · 33, 38, 44, 61, 62, 122, 125, 200, 300, 318, 369, 392, 419, 425
Intentionnalité · 60, 62, 66, 73, 175
Interactionnisme · 222, 223, 224
Interface · 35, 217, 263, 264, 346
Interprétation · 29, 36, 64, 70, 80, 81, 103, 114, 121, 170, 173, 176, 224, 258, 267, 269, 279, 284, 291, 298, 317, 334, 340, 346, 352, 362, 363, 387, 390, 393, 399, 417
Intrinsèque · 28, 136, 176, 180, 184, 185, 192, 200, 201, 202, 217, 221, 226, 227, 228, 292, 368, 380, 399
Introspection · 62, 207, 251, 255, 256, 295, 375, 380
Intuition · 33, 194, 197, 200, 209, 233, 235, 236, 239, 246, 247, 262, 264, 265, 277, 278, 290, 291, 292, 298, 301, 302, 305, 310, 332, 333, 377, 380

J

Jugement · 158, 196, 378
Jumeau · 100, 101

K

KA · 234, 240, 242, 243, 244, 246, 247, 248, 256, 261, 264, 287, 291, 293, 298, 300, 302, 303, 305, 306, 309, 316, 319, 339, 346, 418

L

Langage · 41, 62, 64, 104, 108, 109, 116, 117, 119, 124, 125, 126, 127, 129, 131, 133, 135, 138, 140, 142, 151, 154, 159, 179, 246, 253, 261, 267, 279, 349, 365, 384, 385, 387, 390, 402
Langue · 59, 60, 95, 109, 115, 117, 129, 134, 139, 254, 268, 356, 382, 385
Localisation · 42, 52, 54, 157, 297, 317, 341, 367, 392, 398, 421, 425
Logique · 29, 31, 34, 36, 37, 41, 43, 60, 62, 64, 79, 85, 95, 106, 107, 108, 109, 113, 115, 116, 117, 118, 119, 120, 122, 124, 125, 127, 128, 133, 136, 138, 139, 150, 155, 156, 160, 172, 201, 202, 206, 240, 265, 300, 316, 317, 318, 346, 365, 367, 368, 371, 378, 383, 387, 388, 390, 391, 392, 394, 426
Lois · 24, 27, 28, 54, 55, 64, 68, 69, 73, 76, 78, 88, 104, 113, 114, 118, 147, 152, 154, 170, 171, 174, 175, 211, 221, 222, 224, 226, 227, 300, 329, 330, 349, 365, 376, 377, 378, 399, 402

M

Machine · 151, 169, 172, 173, 174
Matérialisme · 28, 30, 32, 37, 39, 44, 56, 57, 60, 66, 70, 75, 77, 85, 140, 146, 147, 149, 153, 155, 157, 160, 170, 171, 181, 187, 193, 201, 203, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 218, 219, 220, 221, 227, 230, 237, 262, 268, 269, 282, 306, 347, 351, 352, 372, 373
Matérialité · 36, 150, 190, 376, 392, 425
Mathématiques · 41, 74, 79, 80, 89, 107, 115, 118, 159, 201, 215, 359, 372, 374, 378, 381, 386, 426
Matière · 9, 28, 54, 56, 57, 58, 63, 72, 187, 214, 223, 250, 373, 392, 399, 406, 424
Mécanique Quantique · 9, 36, 43, 65, 67, 72, 98, 216, 223, 224, 344, 395, 397, 402, 403, 404
Mémoire · 87, 99, 197, 207, 219, 253, 290, 313, 334, 339, 417, 424
Méréologique · 397, 402
Métaphysique · 27, 28, 33, 37, 42, 44, 55, 58, 61, 62, 72, 73, 75, 77, 78, 80, 84, 85, 86, 87, 88, 89, 108, 109, 145, 171, 201, 202, 205, 213, 221, 227, 234, 239, 240, 246, 248, 259, 299, 300, 305, 313, 318, 320, 336, 340, 341, 343, 344, 345, 347, 351, 354, 378, 381, 383, 387, 389, 390, 392, 393, 395, 396, 400, 404, 425
Microphysique · 72, 190, 216, 218, 221, 229, 300, 316, 317, 318, 359, 360, 361, 362
Microréduction · 70, 71, 397, 403
Mind-Body Problem · 55, 60, 110, 135, 145, 147, 151, 162, 373, 395, 401
Mode contentuel · 41, 109, 110, 126, 128, 129, 133, 134, 135, 140, 141, 365, 368
Mode formel · 109, 110, 126, 128, 129, 131, 133, 134, 135, 136, 138, 139, 140, 141, 365, 367, 368
Monde · 23, 24, 26, 27, 28, 30, 31, 33, 34, 36, 37, 42, 43, 45, 52, 54, 56, 57, 59, 62, 63, 65, 70, 72, 74, 78, 79, 80, 83, 85, 86, 88, 89, 90, 91, 92, 93, 97, 99, 100, 101, 102, 109, 113, 114, 115, 118, 119, 121, 124, 125, 127, 128, 129, 130, 131, 132, 136, 137, 141, 148, 151, 157, 162, 172, 178, 179, 180, 181, 182, 184, 185, 187, 188, 192, 198, 199, 201, 209, 211, 215, 216, 218, 219, 221, 222, 223, 225, 226, 227, 233, 234, 235, 239, 240, 241, 248, 249, 277, 279, 282, 285, 291, 296, 297, 299, 313, 317, 318, 320, 322, 328, 329, 330, 332, 335, 337, 341, 343, 347, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 363, 364, 367, 368, 372, 373, 374, 375, 376, 377, 381, 382, 387, 391, 392, 394, 395, 396, 397, 398, 399, 401, 402, 403, 404, 407, 419, 420, 422, 423, 424, 425, 426
Monisme · 27, 28, 38, 146, 152, 169, 171, 217, 219, 221, 226, 227, 228, 229, 259, 332, 426
Monisme Anomal · 152, 169, 171
Monisme de type-F · 217, 219, 221, 226, 227, 228, 229, 332, 426
Multiple réalisabilité · 166, 167, 168, 169, 171

N

Naturalisme · 24, 25, 26, 42, 45, 367
Nature · 23, 27, 31, 55, 56, 57, 58, 59, 60, 62, 64, 66, 68, 72, 75, 76, 79, 81, 83, 84, 88, 91, 99, 100, 104, 107, 111, 119, 122, 124, 147, 152, 156, 161, 163, 170, 172, 175, 176, 179, 184, 190, 194, 198, 201, 205, 209, 210, 211, 212, 216, 217, 218, 219, 221, 224, 227, 228, 229, 233, 239, 241, 248, 249, 259, 260, 263, 264, 266, 267, 269, 270, 275, 277, 282, 284, 285, 288, 296, 297, 299, 301, 302, 303, 304, 314, 318, 319, 320, 327, 328, 330, 331, 332, 341, 344, 347, 348, 349, 351, 363, 372, 374, 376, 377, 378, 379, 380, 385, 388, 390, 391, 394, 396, 399, 402, 424
Nécessitation · 64, 81, 84, 85, 86, 87, 88, 89, 90, 92, 93, 213, 306, 346, 349, 359, 396
Niveaux · 31, 32, 44, 64, 66, 68, 74, 95, 108, 109, 114, 119, 120, 140, 150, 151, 177, 184, 185, 187, 192, 299, 306, 316, 320, 348, 396, 397, 398, 401, 402, 403, 422
Nomologique · 27, 76, 88, 170, 402
Non-propositionnel · 248, 256, 259

O

Objection · 26, 33, 58, 126, 159, 163, 225, 228, 250, 251, 252, 331, 332, 336, 348, 360, 362, 373, 382, 388, 424

Objectivité · 31, 33, 34, 35, 61, 110, 116, 118, 120, 123, 124, 126, 131, 133, 137, 140, 142, 143, 279, 314, 369, 371, 372, 373, 378, 381, 391, 392, 425

Objet · 51, 56, 60, 62, 64, 80, 85, 90, 94, 95, 96, 97, 98, 102, 105, 107, 113, 114, 117, 120, 123, 131, 135, 136, 137, 138, 145, 147, 149, 162, 163, 167, 168, 180, 182, 194, 205, 228, 234, 250, 257, 258, 261, 266, 267, 270, 280, 281, 284, 291, 295, 296, 301, 304, 325, 327, 332, 336, 360, 362, 368, 376, 377, 379, 381, 390, 391, 393, 395, 398, 406

Occurrence · 82, 92, 131, 132, 137, 142, 167, 173, 182, 183, 211, 251, 252, 267, 271

Omniscience · 26, 40, 59, 81, 221, 234, 235, 247, 248, 311, 315, 318, 320, 321, 322, 323, 324, 325, 334, 335, 336, 417, 418, 419, 426

Ontologie · 27, 33, 37, 61, 62, 64, 72, 104, 109, 124, 128, 151, 156, 166, 172, 186, 187, 190, 205, 269, 317, 359, 365, 383, 384, 385

Output · 173, 174

P

Panpsychisme · 228

Parité des raisons · 276, 277

Passage a priori · 40, 343, 351, 358, 359, 360, 361, 362, 364, 365, 368, 396, 404

Pensée · 11, 24, 27, 29, 31, 32, 34, 35, 36, 37, 39, 41, 44, 45, 51, 57, 60, 61, 62, 78, 99, 100, 101, 109, 113, 115, 116, 117, 118, 119, 120, 122, 123, 124, 125, 126, 127, 129, 133, 138, 142, 147, 150, 154, 155, 163, 169, 174, 179, 193, 194, 195, 202, 204, 207,妍214, 235, 236, 237, 239, 241, 244, 246, 253, 254, 259, 264, 267, 276, 278, 280, 281, 287, 288, 290, 291, 292, 299, 332, 340, 344, 355, 363, 364, 366, 370, 371, 374, 375, 376, 377, 378, 380, 381, 383, 384, 386, 391, 392, 395, 404, 419, 420, 422, 425, 426

Perception · 13, 23, 35, 134, 139, 181, 183, 185, 198, 244, 246, 252, 257, 265, 271, 275, 276, 277, 278, 280, 288, 291, 295, 297, 301, 310, 315, 318, 326, 327, 331, 332, 339, 411, 419

Phénoménalité · 53, 54, 131, 138, 179, 192, 284, 339, 425

Phénoménologie · 43, 228, 229, 297, 406

Phusis · 56, 57, 424

Physicalisme · 11, 23, 24, 26, 27, 28, 29, 30, 32, 34, 38, 39, 40, 41, 42, 43, 44, 45, 49, 51, 52, 54, 55, 56, 59, 60, 62, 63, 71, 72, 73, 74, 75, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 113, 114, 115, 116, 123, 124, 126, 127, 128, 129, 134, 136, 138, 143, 159, 162, 166, 169, 171, 177, 179, 180, 187, 193, 206, 230, 233, 234, 235, 238, 239, 240, 241, 242, 244, 246, 247, 248, 256, 259, 264, 268, 270, 271, 277, 278, 279, 280, 283, 285, 286, 289, 293, 295, 297, 299, 300, 301, 303, 306, 307, 309, 310, 313, 314, 315, 316, 317, 318, 319, 321, 322, 332, 334, 335, 336, 337, 338, 339, 340, 341, 343, 344, 345, 346, 347, 348, 349, 350, 353, 357, 359, 365, 367, 368, 372, 373, 377, 382, 384, 385, 391, 392, 393, 395, 396, 397, 398, 400, 403, 404, 407, 411, 412, 416, 417, 418, 419, 424, 425, 427

Physicalisme a priori · 40, 42, 81, 127, 138, 234, 248, 293, 299, 301, 319, 332, 340, 343, 345, 346, 347, 348, 349, 357, 359, 365, 367, 368, 372, 382, 384, 391, 392, 396, 407

Physicalisme nonthéorique · 318, 332, 334, 335, 336, 337, 418, 419

Physicalisme sémantique · 40, 41, 59, 104, 110, 116, 126, 127, 129, 134, 143, 299, 344, 345, 346, 348, 365, 367, 368, 377, 382

Physique · 11, 23, 24, 25, 27, 28, 29, 32, 35, 36, 37, 40, 42, 43, 45, 52, 54, 55, 56, 57, 58, 59, 60, 63, 64, 65, 67, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104,

106,妍107, 108, 109, 111, 113, 114, 116, 121, 123, 124, 125, 127, 129, 131, 133, 134, 135, 139, 140, 145, 146, 147, 149, 151, 152, 153, 159, 161, 162, 164, 166, 167, 168, 169, 170, 171, 177, 179, 180, 187, 191, 192, 194, 196, 199, 201, 203, 205, 206, 209, 210, 211, 216, 217, 218, 219, 220, 221, 222, 223, 224, 226, 227, 228, 229, 233, 239, 240, 242, 244, 246, 249, 257, 260, 261, 262, 264, 269, 271, 272, 277, 278, 281, 282, 284, 285, 286, 291, 299, 300, 310, 314, 315, 316, 318, 319, 320, 321, 322, 323, 325, 326, 327, 332, 334, 335, 340, 341, 343, 345, 346, 347, 348, 350, 351, 352, 359, 361, 365, 366, 367, 368, 381, 386, 391, 392, 393, 394, 395, 396, 398, 399, 400, 402, 403, 404, 406, 407, 416, 417, 418, 419, 420, 424, 425, 426

Possible · 44, 52, 59, 72, 83, 86, 87, 88, 89, 90, 91, 92, 93, 94, 100, 101, 102, 103, 114, 116, 118, 123, 126, 128, 132, 134, 136, 138, 141, 149, 163, 168, 182, 188, 192, 196, 197, 198, 199, 200, 201, 202, 207, 208, 209, 212, 215, 217, 218, 220, 223, 224, 227, 234, 240, 242, 244, 245, 250, 252, 254, 257, 260, 261, 269, 277, 278, 281, 282, 287, 290, 301, 306, 307, 314, 316, 317, 318, 322, 326, 327, 328, 329, 330, 332, 336, 343, 345, 349, 350, 354, 355, 356, 361, 366, 368, 371, 377, 378, 388, 389, 391, 395, 397, 411, 418, 420, 424

Pragmatique · 314, 315, 416

Prémisse · 26, 45, 67, 68, 69, 74, 76, 97, 105, 106, 136, 139, 161, 189, 190, 194, 201, 212, 215, 219, 221, 237, 238, 240, 246, 247, 272, 277, 300, 304, 306, 313, 315, 318, 325, 326, 334, 335, 339, 352, 359, 360, 361, 368, 411, 417, 418, 419, 424

Principe · 27, 36, 40, 52, 57, 59, 63, 64, 66, 70, 71, 72, 73, 74, 75, 76, 78, 81, 97, 113, 132, 146, 148, 152, 161, 179, 180, 211, 214, 217, 221, 224, 227, 236, 246, 255, 279, 290, 295, 299, 300, 306, 315, 326, 327, 343, 350, 351, 352, 357, 358, 359, 360, 361, 362, 364, 365, 366, 367, 368, 385, 396, 397, 398, 403, 404, 426

Processus · 35, 38, 57, 60, 76, 94, 99, 118, 119, 121, 122, 125, 126, 128, 142, 146, 152, 154, 158, 159, 167, 175, 179, 184, 187, 197, 198, 214, 216, 242, 257, 258, 260, 261, 284, 299, 300, 326, 328, 336, 346, 363, 365, 369, 380, 396, 401, 424

Proposition · 82, 117, 120, 130, 131, 132, 134, 135, 138, 150, 159, 165, 184, 225, 242, 251, 252, 258, 265, 316, 320, 321, 323, 325, 332, 339, 363, 368, 384, 388, 390, 391, 402

Propositionnel · 248, 250, 251, 257, 259, 265, 266, 292, 339

Propriétés fondamentales · 24, 28, 83, 84, 100, 108, 217, 220, 221, 222, 248, 328, 329, 330, 331, 332, 333, 334, 335, 355, 399

Propriétés mentales · 88, 91, 102, 163, 176, 184, 203, 228, 294, 316, 349, 401

Propriétés physiques · 37, 84, 85, 86, 90, 91, 97, 101, 105, 106, 168, 203, 221, 224, 228, 229, 246, 257, 268, 327, 328, 336, 349, 401

Psychologie · 66, 110, 113, 115, 116, 119, 120, 124, 142, 145, 147, 152, 154, 164, 169, 172, 173, 175, 192, 366, 386, 394

Psychologisme · 29, 113, 118, 119, 129, 130, 390, 391

Q

Qualia · 11, 13, 24, 26, 32, 34, 36, 38, 40, 44, 45, 60, 66, 108, 110, 121, 124, 126, 127, 128, 169, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 190, 193, 194, 195, 196, 197, 198, 200, 203, 204, 208, 215, 221, 239, 242, 243, 245, 256, 257, 259, 260, 276, 280, 291, 292, 333, 345, 392, 424, 425

Qualités phénoménales · 55, 87, 181, 251, 252, 261, 263, 283

Quantificateurs · 239

Quantique · 72, 223, 224, 403, 406

R

Raison · 29, 33, 36, 56, 62, 68, 69, 77, 84, 90, 101, 106, 108, 110, 124, 156, 165, 199, 200, 209, 210,

218, 252, 269, 271, 277, 278, 281, 315, 330, 355, 373, 376, 377, 381, 404, 425
Rationalité · 3, 34, 36, 41, 43, 175, 215, 359, 367, 375, 376, 405, 419
Réalisation · 66, 135, 172, 300, 316, 329, 330, 403
Réduction · 38, 41, 53, 64, 66, 67, 68, 69, 70, 72, 78, 84, 108, 146, 153, 170, 171, 176, 188, 193, 329, 360, 395, 396, 401
Réductionnisme · 176, 180, 190, 206, 221
Réel · 24, 25, 28, 29, 36, 43, 44, 52, 58, 59, 61, 62, 65, 66, 70, 75, 77, 81, 88, 93, 108, 111, 121, 125, 130, 137, 141, 143, 145, 162, 184, 190, 198, 201, 211, 219, 220, 222, 225, 230, 282, 284, 300, 306, 314, 315, 317, 320, 321, 327, 333, 334, 336, 337, 340, 341, 346, 347, 354, 357, 365, 369, 374, 375, 384, 387, 390, 391, 393, 394, 396, 397, 398, 399, 401, 402, 403, 407, 418, 424
Référence · 24, 30, 54, 55, 56, 58, 59, 62, 69, 70, 85, 99, 102, 103, 109, 110, 117, 122, 127, 128, 135, 142, 143, 151, 154, 158, 173, 202, 210, 227, 251, 255, 263, 267, 268, 270, 271, 280, 281, 283, 284, 286, 327, 352, 355, 356, 358, 362, 368, 374, 376, 378, 381, 383, 385, 386, 389, 390, 398
Relation · 35, 38, 39, 55, 64, 68, 84, 87, 88, 89, 114, 117, 126, 136, 139, 152, 156, 162, 171, 211, 213, 225, 246, 258, 261, 265, 281, 282, 283, 284, 285, 286, 287, 299, 300, 310, 346, 349, 350, 353, 366, 397, 405
Représentation · 31, 45, 118, 119, 120, 175, 209, 218, 267, 280, 281, 296, 297, 298, 301, 302, 303, 304, 305, 320, 377, 401
Représentationalisme · 185, 267, 294, 295, 296, 301, 302, 303, 304, 305, 306
Res cogitans · 58, 373
Rôle · 28, 29, 31, 55, 59, 61, 62, 63, 68, 69, 75, 76, 79, 80, 109, 138, 139, 150, 153, 160, 161, 166, 172, 175, 176, 177, 178, 185, 194, 198, 209, 214, 221, 224, 225, 252, 259, 264, 267, 281, 297, 299, 304, 306, 328, 340, 343, 352, 355, 356, 358, 360, 361, 368, 377, 395, 422, 423, 426

S

Saisir · 55, 103, 109, 110, 115, 120, 124, 127, 128, 131, 136, 137, 149, 185, 188, 202, 208, 215, 230, 233, 241, 242, 245, 263, 276, 286, 287, 288, 305, 318, 325, 333, 336, 372, 420, 423
Saut conceptuel · 36, 40, 41, 109, 110, 115, 121, 122, 123, 125, 126, 127, 128, 141, 142, 299, 348, 365, 368, 370, 379, 380
Savoir · 24, 36, 39, 54, 67, 73, 78, 84, 87, 90, 98, 109, 110, 115, 116, 117, 118, 121, 126, 129, 140, 151, 159, 165, 172, 175, 177, 186, 191, 194, 196, 198, 203, 208, 209, 214, 220, 221, 222, 233, 236, 238, 239, 240, 246, 247, 248, 249, 250, 251, 252, 253, 254,妍255, 256, 259, 260, 261, 262, 264, 265, 266, 269, 270, 271, 273, 274, 279, 280, 284, 288, 290, 291, 296, 298, 301, 302, 310, 311, 313, 315, 316, 319, 322, 323, 325, 327, 329, 330, 332, 333, 335, 339, 340, 343, 344, 347, 349, 350, 352, 354, 355, 357, 359, 360, 364, 365, 366, 369, 374, 375, 376, 378, 383, 384, 385, 390, 394, 397, 400, 404, 418, 419, 424, 425
Savoir faire · 246, 247, 248, 249, 252, 255, 257, 290
Savoir que · 54, 246, 248, 249, 251, 252, 255, 256, 259, 265, 269, 273, 340, 354, 355, 357, 418
Schémas conceptuels · 426
Schisme · 32, 269, 372, 380, 420, 425
Science · 27, 28, 41, 43, 54, 56, 58, 59, 60, 61, 63, 64, 66, 67, 68, 69, 70, 72, 77, 78, 108, 109, 111, 116, 124, 127, 128, 131, 132, 133, 134, 139, 143, 145, 147, 154, 162, 166, 173, 188, 191, 214, 223, 226, 227, 228, 282, 344, 346, 347, 350, 359, 378, 382, 385, 386, 387, 389, 394, 397, 398, 400, 401, 402, 403
Sciences cognitives · 66, 173, 174
Sémantique · 42, 62, 109, 110, 126, 133, 138, 348, 354, 356, 358, 365, 390, 392, 393
Sens · 26, 30, 40, 41, 44, 45, 57, 59, 60, 64, 79, 80, 82, 85, 87, 88, 89, 91, 95, 97, 107, 111, 117, 118, 119, 120, 124, 131, 135, 136, 140, 143, 147, 148, 150, 152, 156, 157, 158, 159, 161, 163, 165, 169,

170, 171, 172, 173, 176, 177, 183, 192, 200, 207, 209, 210, 211, 212, 213, 217, 218, 225, 226, 228, 236, 247, 251, 259, 260, 263, 265, 266, 267, 272, 284, 285, 289, 291, 292, 297, 310, 315, 316, 317, 318, 332, 336, 339, 341, 345, 354, 359, 362, 365, 367, 368, 374, 375, 376, 381, 384, 386, 387, 388, 389, 391, 392, 400, 402, 424, 425

Sensation · 86, 158, 179, 225, 252, 267, 273, 274, 366

Simpliciter · 83, 88, 92, 201, 319, 320, 321, 322, 325, 327, 334, 335, 336, 350, 417, 418, 419

Structure · 28, 38, 61, 63, 73, 76, 103, 105, 132, 149, 165, 180, 192, 193, 195, 196, 198, 200, 202, 211, 216, 228, 229, 235, 238, 299, 315, 317, 318, 319, 333, 356, 361, 369, 381

Subjectivité · 30, 33, 35, 37, 110, 118, 123, 124, 126, 131, 140, 142, 143, 278, 369, 371, 372, 373, 378, 381, 391, 392, 394, 420, 425, 426

Substrat · 172, 176, 191, 314, 315, 316, 318, 330, 416

Sujet · 34, 35, 42, 58, 60, 66, 67, 72, 85, 103, 109, 111, 113, 114, 115, 116, 120, 121, 124, 127, 128, 129, 130, 131, 132, 135, 136, 137, 138, 139, 141, 142, 143, 148, 152, 158, 173, 179, 181, 182, 204, 205, 221, 223, 224, 228, 233, 234, 236, 246, 249, 251, 257, 258, 260, 262, 263, 264, 266, 267, 269, 274, 278, 279, 280, 283, 285, 286, 289, 299, 300, 303, 310, 314, 315, 317, 320, 322, 325, 326, 327, 331, 332, 336, 338, 355, 363, 368, 371, 372, 373, 376, 378, 380, 384, 386, 389, 390, 391, 393, 394, 402, 405, 419, 425

Super-Sujet · 365, 371, 382, 405, 419, 425

Supranaturalisme · 344, 367, 382

Surplomb · 136, 142, 279, 315, 371, 389, 416, 425, 426

Survenance · 41, 64, 78, 84, 87, 92, 169, 170, 171, 201, 240, 317, 350, 351, 352, 366, 367, 396, 399, 402, 403

Syntaxe · 41, 129, 133, 135, 139, 142, 346, 382, 387

Syntaxe Logique · 41, 129, 135, 387

Synthèse · 40, 90, 118, 131, 184, 236, 246, 260, 294, 377

Synthétique · 68, 94, 127, 130, 247, 381, 385

Système · 37, 59, 76, 98, 130, 131, 134, 135, 156, 161, 165, 167, 172, 173, 174, 185, 186, 187, 197, 199, 204, 208, 209, 211, 218, 249, 254, 280, 281, 291, 315, 368, 369, 376, 378, 387, 388, 390, 391, 399, 403

T

Temps · 28, 30, 37, 39, 58, 62, 67, 69, 82, 97, 103, 109, 115, 116, 119, 130, 143, 177, 184, 197, 201, 206, 208, 212, 215, 216, 220, 233, 244, 246, 247, 262, 269, 276, 280, 284, 287, 288, 289, 291, 293, 313, 341, 357, 369, 374, 381, 385, 393, 399, 400, 411, 419, 423

Token · 167, 170

Troisième personne · 35, 140, 142, 155, 157, 196, 199, 268, 270, 279, 302, 315, 319, 416, 417

Type · 164, 167

U

Unité de la science · 27, 63, 64, 65, 72, 78, 126, 129, 161

Universalité · 44, 55, 72, 152, 193, 313, 365

Universel · 81, 137, 167, 306

V

Vérifiabilité · 155, 387

Vérification · 62, 84, 136, 141, 158, 248

Vérificationnisme · 132, 153

Vérité · 29, 31, 33, 34, 40, 43, 55, 62, 77, 80, 104, 106, 125, 127, 136, 146, 148, 154, 162, 163, 201, 202, 211, 212, 213, 215, 218, 219, 220, 222, 228, 235, 240, 241, 244, 246, 248, 252, 258, 267, 273, 275, 276, 277, 285, 293, 300, 310, 316, 320, 327, 334, 335, 340, 341, 343, 348, 350, 353, 354, 356, 357, 358, 359, 360, 362, 363, 365, 366, 367, 368, 371, 372, 373, 374, 381, 384, 391, 392, 397, 425, 426

Verticalité · 42, 346, 390

Vision · 34, 43, 57, 64, 97, 234, 246, 249, 272, 289, 291, 323, 374, 405, 421, 422, 423

W

Weltanschauung · 29, 39, 42, 44, 75, 80, 108, 230, 336, 340, 393

Z

Zombies · 196, 199, 200, 201, 203, 210, 215, 317, 320

Index des noms

A

Alter · 246, 253, 254, 294, 296, 301, 302, 303, 304, 305
Andler · 131
Aristote · 57, 379
Armstrong · 27, 56, 73, 75, 159, 160, 161, 164, 267
Ayer · 61, 130, 136

B

Battermann · 68
Beaton · 9, 292, 405, 406
Bickle · 9, 69, 70, 166, 167
Bitbol · 43, 70, 131, 142, 223, 341, 346, 401, 403, 425, 426
Blackburn · 31, 40, 341, 343, 356, 359, 362, 363, 364, 366, 367
Block · 54, 156, 163, 164, 169, 171, 176, 182, 194, 195, 204, 210, 284, 286, 353, 359, 360, 361, 362
Boutot · 59
Braddon-Mitchell · 52, 58, 149, 192, 327
Broad · 54, 208, 235, 247, 250, 341

C

Carnap · 37, 40, 41, 42, 60, 108, 113, 123, 124, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 142, 143, 344, 348, 365, 367, 368, 369, 376, 381, 382, 383, 386, 387, 388, 389, 390, 391, 394
Carruthers · 65, 148, 149, 150, 203
Chalmers · 24, 39, 53, 69, 88, 91, 92, 179, 181, 186, 187, 190, 195, 199, 200, 202, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 224, 225, 226, 228, 229, 268, 269, 280, 302, 304, 319, 320, 331, 332, 353, 354, 355, 356, 357, 360, 362, 382, 426
Chomsky · 27, 173, 253, 254
Churchland · 147, 214, 220, 257, 260, 277, 278, 338, 414
Conee · 253, 257, 262, 266, 309
Crane · 54, 70, 72, 113

D

Davidson · 87, 152, 169, 170, 171, 395
Dennett · 182, 185, 195, 206, 207, 208, 220, 237, 263, 290, 291, 292, 309
Descartes · 30, 33, 148, 149, 222, 372, 373, 374, 375, 376, 391
Dretske · 185, 196, 206, 274, 275, 280, 281
Dummett · 119, 120, 122, 124, 125
Dupré · 393, 402

E

Eco · 378, 379
Engel · 118, 121, 126, 132, 133, 170, 291, 329, 369, 380
Esfeld · 88

F

Feigl · 27, 64, 69, 145, 235, 394
Frege · 29, 35, 37, 113, 115, 116, 117, 118, 119, 120, 122, 123, 124, 125, 126, 142, 163, 272, 332, 371, 386, 391

G

Gardner · 173, 174, 376, 380
Grim · 319

H

Heil · 114, 401, 402
Hempel · 76, 98, 102, 103, 104, 106, 124, 216, 336
Horgan · 242, 262, 263
Husserl · 116, 378
Hüttemann · 398

I

Inwagen · 93, 399

J

Jackson · 23, 24, 25, 31, 39, 40, 43, 45, 52, 58, 60, 61, 81, 87, 91, 92, 103, 140, 149, 153, 163, 180, 193, 194, 205, 212, 214, 226, 230, 233, 234, 235, 238, 239, 241, 242, 244, 246, 253, 257, 259, 261, 262, 264, 271, 273, 274, 275, 276, 277, 278, 279, 280, 287,妍288, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 304, 305, 306, 308, 309, 319, 320, 321, 322, 325, 326, 331, 334, 340, 341, 343, 345, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 364, 365, 367, 368, 369, 382, 391, 396, 400, 404, 411, 417, 418, 419, 421, 422, 426
Jacob · 41, 59, 85, 109, 117, 291, 383

K

Kant · 376, 377, 378, 379, 380, 381
Kim · 69, 78, 168, 171, 184, 398
Kripke · 85, 86, 89, 90, 162, 163, 199, 202, 213, 214, 240, 241, 352, 396

L

Ladyman · 359, 393, 400, 402, 403
Levin · 250, 260
Levine · 34, 38, 39, 126, 180, 187, 188, 190, 191, 192, 205, 206, 280, 281, 282, 310
Lewis · 28, 56, 69, 73, 75, 83, 84, 87, 161, 164, 175, 176, 192, 206, 249, 255, 256, 309, 327, 328, 330, 331, 333, 334, 338, 343, 399, 417
Loar · 212, 213, 250, 251, 252, 269, 284
Locke · 41, 182, 194, 313, 366, 367, 375, 376
Lockwood · 56, 226
Loewer · 55, 66, 395
Lowe · 393
Ludwig · 146
Lycan · 27, 151, 153, 154, 157, 175, 182, 183, 197, 210, 233, 236, 249, 250, 252, 260, 267, 268, 309, 338, 343, 356, 359

M

Macarthur · 25, 27, 63, 382, 393
Marianna · 286, 287, 288, 289
Mary · 23, 26, 34, 39, 40, 42, 87, 140, 163, 180, 193, 233, 234, 235, 236, 237, 238, 239, 241, 242, 244, 245, 246, 247, 248, 249, 251, 253, 255, 256, 257, 259, 260, 261, 262, 263, 264, 265, 266, 267, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 285, 286, 287, 288, 289, 290, 291, 292, 293, 295, 296, 298, 300, 301, 302, 304, 305, 306, 307, 308, 309, 310, 313, 314, 315, 316, 318, 322, 326, 327, 328, 330, 331, 332, 333, 334, 337, 339, 340, 341, 343, 345, 347, 350, 351, 352, 411, 417, 418, 419, 421, 422, 423, 424
McGinn · 55
McLaughlin · 350, 351, 355, 395
Mellor · 70, 72, 113, 314, 413, 414, 415
Melnyk · 168
Münchhausen · 139, 371, 426

N

Nagasawa · 23, 39, 40, 94, 233, 234, 239, 277, 302, 316, 318, 319, 320, 321, 322, 324, 325, 326, 327,

331, 332, 333, 334, 335, 336, 337, 411, 417, 418, 419, 426
Nagel · 67, 68, 69, 70, 78, 181, 183, 214, 236, 237, 238, 247, 287, 288, 314
Nemirow · 249, 250, 251, 253, 254, 255, 309, 338
Neurath · 59, 108, 129, 130
Nida-Rümelin · 242, 244, 245, 256, 262, 286, 287, 288, 289
Nordby · 421, 422, 423

O

Oppenheim · 27, 64, 65, 401

P

Pandore · 55, 129, 141, 343
Papineau · 65, 71, 73, 74, 75, 77, 212, 268, 269
Pégase · 384
Perry · 210, 260, 270, 271, 272, 273, 274, 275, 276, 278, 279, 280, 281, 309
Pettit · 52, 182, 249, 250
Pinkas · 150, 154
Place · 54, 157, 158, 341
Prigogine · 377
Prior · 324
Putnam · 27, 28, 35, 64, 65, 99, 100, 101, 155, 156, 164, 165, 166, 169, 173, 214, 355, 394, 401

Q

Quine · 24, 41, 42, 109, 111, 116, 128, 130, 143, 179, 220, 258, 344, 383, 384, 385, 386, 389, 390, 394

R

Ramsey · 41, 192, 327, 328, 329, 330, 331, 334, 367, 417
Ravenscroft · 31, 52, 293, 341, 343, 362
Rey · 154, 206, 208
Rivenc · 41, 139
Robinson · 226, 236, 277
Russell · 128, 226, 227, 235, 386, 389
Ryle · 150, 151, 155

S

Schaffer · 402, 403
Schlick · 36, 129, 130, 150
Seager · 56, 229
Shoemaker · 182, 195, 196
Smart · 56, 73, 75, 90, 157, 158, 159, 161, 163, 167, 394
Smith Q. · 34, 126, 209, 233, 268, 269, 404
Soames · 383
Spelke · 97
Stoljar · 29, 63, 78, 79, 80, 81, 82, 83, 84, 86, 88, 90, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 105, 106, 134, 239, 240, 340, 395, 396
Strawson · 192, 226
Stroud · 25, 26, 63, 94, 172, 365, 382
Sturgeon · 66, 67

T

Tye · 183, 184, 185, 210, 241, 255, 256, 257, 258, 260, 261, 263, 264, 265, 266, 269, 295, 309, 338

V

Van Gulick · 34, 42, 126, 172, 175, 184, 187, 192, 195, 197, 198, 199, 215, 247, 290, 293, 298, 299, 300, 301, 302, 306, 307, 314, 315, 316, 317, 319, 320, 337, 338, 339, 340, 411, 416, 417, 425

W

Wagner · 36, 41, 60, 109, 127, 128, 136, 150, 365
White · 163, 196, 283, 284, 285, 286, 326
Whitehead · 33, 128, 374, 386
Williamson · 43, 62, 256

Y

Yates · 374, 375