

Université d'Aix Marseille

Le problème de l'esprit: entre dépassement de la science et  
réduction de la philosophie, un commentaire des oeuvres de  
Penrose à la lumière de Gödel

Carbo-Gil Fabien

Sous la direction de Gabriella Crocco

2 avril 2018

# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 Roger Penrose : Mathématiques, physique, esprit et ordinateurs</b>	<b>7</b>
1.1 Courte biographie de Roger Penrose . . . . .	7
1.2 Présentation de EOLP et OE . . . . .	11
1.2.1 Calculabilité . . . . .	11
1.2.2 Présentation conjointe de EOLP et OE . . . . .	13
1.2.3 Différences de forme et de fond entre EOLP et OE . . . . .	15
1.3 La place de la spéculation chez Penrose . . . . .	17
<b>2 Kurt Gödel 1961 : <i>The modern development of the foundations of mathematics in the light of philosophy</i></b>	<b>19</b>
2.1 Courte biographie de Gödel . . . . .	19
2.2 Le texte de 1961 . . . . .	21
2.3 La classification des différentes conceptions du monde du texte de Gödel 1961 . . . . .	22
2.3.1 Critères de la classification . . . . .	22
2.3.2 L'analyse historique de Gödel . . . . .	23
2.3.3 Les cas mixtes . . . . .	25
2.4 Le projet Gödelien . . . . .	26
2.5 L'optimisme épistémologique de Gödel . . . . .	27
<b>3 Les trois mondes de Penrose à la lumière de Gödel</b>	<b>29</b>
3.1 L'A, B, C, D du calcul et de la pensée consciente : . . . . .	30
3.1.1 Présentation . . . . .	30
3.1.2 Quelques commentaires de Penrose . . . . .	31
3.1.3 Quelques commentaires personnels . . . . .	33
3.2 Trois mondes et trois mystères . . . . .	35
3.2.1 La place des trois mondes dans OE . . . . .	35
3.2.2 Présentation des trois mondes . . . . .	37
3.2.3 Le contenu du monde platonicien. . . . .	38
3.2.4 Présentation des trois mystères . . . . .	40
<b>Conclusion</b>	<b>44</b>
<b>Bibliographie</b>	<b>50</b>

## Table des figures

1	Pavage de Penrose . . . . .	51
2	Le triangle de Penrose . . . . .	51
3	Les trois mondes de Penrose . . . . .	52
4	Ensemble de Mandelbrot, zoom sur la vallée des hippocampes . . . .	52

# Introduction

Esprit, âme, conscience, pensée, psyché, intellect, entendement, etc. Voilà autant de mots qui ont traversé l'histoire de la philosophie depuis la Renaissance et pour lesquels nous pourrions trouver tout autant de définitions différentes. Ils semblent pourtant viser tous une certaine activité dont chaque individu peut témoigner, qui semble manifester une certaine constance durant toute la vie, à travers ce que l'on peut appeler la persistance du moi, et qui n'est directement accessible qu'à l'individu lui-même. Si nous utiliserons le mot esprit plutôt qu'un autre, c'est parce que c'est celui qu'utilise le plus fréquemment l'auteur que nous allons étudier, nous ne prétendons en rien en donner une définition plus précise que celle, très vague, que nous venons de donner. Point de départ et d'arrivée de nombreux projets philosophiques, nul système prétendant à une certaine complétude dans sa vision du monde ne peut ignorer la question de l'esprit qui est irrémédiablement liée à celle du sujet. Condition *sine qua non* à toute réflexion philosophique, la révolution copernicienne opérée par la philosophie kantienne en fait une contrainte de cette même réflexion. Ainsi, si l'esprit éclaire de ces lumières le monde, il en plonge inévitablement une partie dans l'ombre, il est ainsi à la fois l'origine et la limite du savoir.

La question de l'esprit a pu paraître un temps hors de portée de la science, la religion et la philosophie<sup>1</sup> couvraient alors le domaine de l'esprit tandis que la science couvrait celui de la matière et donc du corps. Le coup porté par le nihilisme de Nietzsche aux grands systèmes idéalistes allemands et à la religion d'une part, et l'accélération du progrès scientifique à l'heure de l'industrialisation de l'Europe d'autre part, peuvent être vus comme les prémisses d'un changement dans cet état de fait. Pourtant si l'existence d'un lien entre la conscience et le cerveau ne fait plus de doute au sein de la communauté scientifique, le cerveau reste jusqu'au début du XX<sup>e</sup> siècle relativement mal connu par les scientifiques. En effet en l'absence de technologie d'imagerie médicale, être dans un état de conscience habituel semblait difficile à conjuguer avec le fait qu'un médecin observe directement l'activité de votre cerveau. Deux avancées technologiques allaient changer considérablement la donne au milieu du XX<sup>e</sup> siècle.

Tout d'abord la naissance des neurosciences et des sciences cognitives, portées par les avancées réalisées en physique, en chimie et en biologie. Elles se développent

---

1. Si je parle de religion ET de philosophie, ce n'est nullement pour exclure la première de la seconde, mais c'est parce que je pense que la place de la religion chrétienne en occident ne peut se résumer à son contenu philosophique. Je pense au contraire que les dimensions sociales et culturelles du christianisme ne peuvent se réduire à sa théologie et que ces dimensions ont eu, elles aussi, un rôle important à jouer dans l'évolution de la philosophie.

grâce à l'invention de méthodes d'imagerie médicale non intrusives telles que l'IRM, la radiographie, l'électro-encéphalographie, etc. qui permettent aux scientifiques de suivre l'évolution de l'activité électrique dans le cerveau. C'est là le point de départ d'une naturalisation d'un certain nombre de concepts philosophiques tels que les perceptions, les émotions ou encore l'intelligence.

Parallèlement, l'informatique se développe au sortir de la Seconde Guerre mondiale. Si les connaissances informatiques participent indubitablement aux progrès technologiques qui accompagnent l'avènement des nouvelles technologies d'imagerie médicale, l'influence de l'informatique sur le problème de l'esprit est bien loin de se limiter à cet apport. Le développement des machines à calculer impulse ainsi une approche cybernétique ou computationnaliste du fonctionnement du cerveau<sup>2</sup>, selon laquelle le cerveau n'est rien de plus qu'une machine à calculer, donnant par le même coup un nouvel horizon de recherche au développement de l'informatique : la simulation de l'activité cérébrale par un ordinateur que l'on désignera alors par le terme d'intelligence artificielle.

Ces deux avancées scientifiques font tomber dans le domaine de la science un grand nombre des questions et des problèmes philosophiques liés à l'esprit et traditionnellement traités par la religion et la philosophie. Les sciences cognitives sont cependant relativement jeunes, d'autant plus si on les compare à la philosophie. Ainsi si la science peut prétendre donner des définitions scientifiques et matérialistes de concepts tels que le concept de perception (consciente ou pas), d'autres tels que le concept de beau ou de bien semblent encore bien loin d'être compris uniquement en termes de sciences naturelles.

Le cinéma et la littérature du XX<sup>e</sup> siècle et plus particulièrement la science-fiction puisent largement dans les lacunes philosophiques que présente alors le point de vue scientifique assimilant l'esprit à un ordinateur. On y retrouve ainsi des ordinateurs ou des robots rappelant le mythe du golem, des machines aussi voire plus intelligentes qu'un humain, mais réduites à l'asservissement<sup>3</sup> par leur absence de conscience, de libre arbitre, d'empathie<sup>4</sup> ou encore de créativité. Dans d'autres créations l'ordina-

---

2. On trouve la première mention de l'analogie classique assimilant les neurones et les axones aux transistors et aux câbles d'un circuit électronique en 1943 dans l'article de McCulloch et Pitts *A logical calculus of the idea immanent in nervous activity*

3. Dans Star Wars les droïdes, bien que pouvant remplacer les humains dans la plupart de leurs activités, sont entièrement dévoués à l'accomplissement de la tâche pour laquelle ils ont été programmés.

4. Dans le roman *Les androïdes rêvent-ils de moutons électriques ?*, de Philip K. Dick on différencie les humains et les androïdes grâce à un test d'empathie.

teur semble manifester des émotions<sup>5</sup> une certaine forme de conscience, plus ou moins éloignée de la conscience humaine et pouvant apparaître tour à tour inférieure<sup>6</sup> et supérieure<sup>7</sup> à la conscience humaine, bien que juger du niveau de conscience d'un être pensant me semble bien difficile. Ces nombreux exemples illustrent selon moi les limites de l'approche computationnaliste du cerveau, puisqu'en affirmant que tout est calcul, on soulève automatiquement une foule d'autres questions. Nous pourrions ainsi nous demander si cet algorithme qui compose notre conscience est le même pour tous à la naissance et s'il existe des calculs, des parties de cet algorithme, qui soient propres à l'Homme. Si notre conscience évolue comme un calcul déterministe, comment naît l'impression de libre arbitre ? Plus largement il est naturel de se demander comment un calcul fait naître des sentiments, des sensations ou encore de l'empathie. Nous pourrions comparer le fait de répondre à la question "Qu'est ce que l'esprit" par l'affirmation "L'esprit est le résultat d'un algorithme exécuté par le cerveau" au fait de décrire la Vénus de Milo comme "Un amas d'atomes".

Or la nature même du cerveau et de la conscience en font un objet difficile à étudier scientifiquement pour deux raisons pratiques.

Tout d'abord, le cerveau est un organe grand et complexe dans le sens où il est difficile de le découper en sous-parties élémentaires indépendantes. L'étude de l'activité neuronale, par exemple, ne suffit pas à comprendre son fonctionnement. Les différentes parties du cerveau sont interconnectées formant un réseau de très grande taille (même pour les capacités informatiques de traitement des données actuelles) dont la subtile architecture globale est aussi déterminante que des propriétés plus locales telles que l'organisation de l'hypothalamus ou bien le fonctionnement d'un neurone. De plus, les concepts avec lesquels nous pensons notre esprit ne semblent pas être directement traductibles en phénomènes physiques. Le cerveau ne se découpe pas en une zone de la morale, une zone de l'esthétique, une zone du courage, etc.

S'ajoute à cette organisation massivement parallèle une seconde difficulté à savoir la plasticité du cerveau. L'Homme a besoin d'interactions sociales pour se développer normalement, le contenu de ces interactions est hautement culturel et a un impact sur la structure du cerveau. La traditionnelle opposition nature/culture perd de sa pertinence, l'Homme apparaissant comme un animal naturellement culturel, l'organisation de son cerveau dépend alors crucialement de son environnement.

Ainsi la méthode d'analyse utilisée en physique qui consiste en découper le système à étudier en sous-systèmes que l'on isole et stabilise en laboratoire afin de tester leurs réponses à différentes perturbations ne s'avère pas applicable au cerveau humain. Ceci

---

5. C'est le cas de HAL 9000 dans le film *2001, l'Odyssée de l'espace* réalisé par S.Kubrick

6. C'est le cas par exemple pour les robots des nouvelles de I.Asimov.

7. C'est le cas dans le film *Her* de Spike Jonze

ne place pas le cerveau définitivement hors de sa portée, mais permet de relativiser notre connaissance actuelle de l'esprit si l'on ne le regarde que d'un point de vue matérialiste comme la manifestation de l'activité du cerveau.

C'est également, nous allons le voir, cette méthode d'analyse que choisit l'auteur que nous allons étudier pour tenter d'apporter sa contribution à cette naturalisation de l'esprit humain. Loin de vouloir apporter une réponse globale aux questions que soulève le fait d'identifier l'esprit à un épiphénomène de l'organisation spontanée de la matière dans certains systèmes biologiques, il restreint son étude à la question de la simulabilité de la conscience par un dispositif de calcul de type machine de Turing. Roger Penrose, mathématicien et physicien né en 1931, traite ainsi de la question de la réduction de la conscience, et plus particulièrement de la compréhension en mathématiques, à un calcul ou algorithme dans les deux ouvrages que nous allons étudier : *L'Esprit, l'Ordinateur et les Lois de la Physique* (1989, traduit de l'anglais en 1992) et *Les Ombres de l'Esprit. À la recherche d'une science de la conscience* (1994, traduit de l'anglais en 1995). Il y défend l'impossibilité d'identifier l'esprit à une machine à calculer, en se basant notamment sur le théorème de Gödel.

Discuter ce point particulier participe au projet scientifique de naturalisation de l'esprit du point de vue de Penrose.

En effet en remettant en cause l'hypothèse de départ de l'intelligence artificielle il pense éviter l'écueil de la position qui défend l'idée que tout peut être compris dans les termes de la science actuelle. Si la conscience relève en effet exclusivement de l'activité du cerveau, nous n'avons pas forcément tous les éléments en main, tant au niveau des données empiriques qu'au niveau théorique pour comprendre cette activité. La présenter comme un simple calcul c'est extrapoler nos connaissances de certains systèmes évoluant de manière calculable à un système auquel on ne sait trop à quel point il est pertinent de les appliquer. Il défend ainsi par l'affirmation de l'existence de phénomènes résistant à une interprétation scientifique, la possibilité d'un réel progrès de la science par le dépassement des théories actuelles. C'est également là l'occasion pour Penrose de tenter d'anticiper ce dépassement et les conséquences qu'il aura sur la compréhension scientifique de l'esprit.

Notre objectif ici n'est en aucun cas de juger de la crédibilité des arguments avancés ni de tenter de statuer sur ces conclusions. C'est plutôt à la méthode de Penrose que nous allons nous intéresser. Ainsi nous allons nous demander si la méthode de Penrose, qui consiste à tenter de faire la lumière sur un point particulier au sein du débat philosophique sur l'esprit, débouche sur un réel progrès philosophique. En physique cette méthode a fait ses preuves, en effet, face à un problème difficile, il est souvent très efficace de ne le considérer que comme la réunion ou la somme

de nombreux sous-problèmes qui le constituent et dont la résolution de chacun fait avancer la compréhension générale du problème. Nous allons tenter ici de juger de la pertinence et des limites d'une telle démarche en philosophie pour le problème général de l'esprit.

Puisque cette question porte sur l'entièreté de l'approche de Penrose et de son argumentation, nous serons amenés à nous appuyer sur des citations provenant de la totalité du texte des deux ouvrages. Cependant au vu de la quantité de matériel que cela représente, chaque ouvrage dépassant les quatre cents pages, nous ne serons bien évidemment pas en mesure de fournir une analyse détaillée de l'ensemble de leur contenu. Pour pallier à cette difficulté, nous commencerons, dans la première partie, par une présentation générale des ouvrages et de leur contenu, afin de fixer un cadre à l'analyse plus détaillée que nous ferons dans la partie trois de deux passages choisis. À la longueur des ouvrages que nous nous proposons d'étudier s'ajoute une seconde difficulté. Puisque notre objectif ici est de replacer les ouvrages de Penrose à l'intérieur d'un débat philosophique bien plus large, il va nous falloir comprendre la place qu'y occupe Penrose. Pour cela, il sera nécessaire de traduire dans le vocabulaire philosophique certains passages et de tirer au clair les postulats philosophiques sur lesquels il s'appuie, qu'ils soient explicites ou implicites. En effet, en partie conséquemment au fait que Penrose écrive plutôt en tant que mathématicien et physicien qu'en tant que philosophe, le style, le vocabulaire et la structure des ouvrages de Penrose sont très éloignés de ceux de textes de philosophie plus canoniques. Pour y parvenir et au vu de mes connaissances embryonnaires en philosophie, il m'est apparu indispensable d'appuyer cette étude sur le texte d'un autre auteur. J'ai choisi de cette manière de m'appuyer sur le texte de 1961 de K.Gödel *The modern development of the foundations of mathematics in the light of philosophy* que nous avons étudié dans le cadre du cours de langues dispensé par G.Crocco, et cela pour diverses raisons. Tout d'abord, la présence de certains éléments communs dans la pensée de Penrose et de Gödel facilite ce rapprochement. De plus, Gödel dans le texte de 1961 tente de décrire le développement des fondements des mathématiques dans la première moitié du xx<sup>e</sup> siècle à l'aide de concepts propres à la philosophie et de le situer dans un panorama relativement large de positions philosophiques possibles. Auteur très rigoureux, et philosophe tout autant que mathématicien, Gödel nous apportera donc l'assise philosophique qui nous faisait défaut. La seconde partie sera donc dédiée à une présentation du texte de 1961 dans laquelle nous insisterons sur les éléments qui trouveront leur pertinence dans le commentaire que nous ferons des deux passages choisis dans la partie trois.



Avant de livrer le contenu de la troisième partie, il me faut m'arrêter ici pour dire quelques mots à propos de la pensée de Gödel. En effet, j'ai eu besoin de m'appuyer sur d'autres textes écrits par Gödel pour préciser sa position dans le texte de 1961, parmi eux certains proviennent des ouvrages que H.Wang a écrits à propos de la pensée de Gödel ou d'articles de recherche plus récents. La pensée de Gödel a évolué au cours de sa vie et il a conservé un grand nombre de notes philosophiques en comparaison avec le nombre de textes philosophiques qu'il a publiés. Le texte de 1961 en est un exemple puisqu'il est le texte préparatoire d'une conférence qu'il n'a finalement jamais donnée. Bien que le travail d'étude de la pensée de Gödel par la communauté des philosophes ait commencé depuis les années 1990, la quantité de notes non publiées, dont une partie n'a pas encore été traduite du langage sténographié qu'il utilisait et la profondeur et la complexité de sa pensée font de ce travail une entreprise encore en cours et qui est loin de se terminer. Ainsi, lorsque nous parlerons de la pensée de Gödel, nous nous restreindrons à sa pensée telle qu'elle apparaît dans les documents sur lesquels nous nous appuierons. Pour clore cet aparté, je voudrais souligner le fait que ce mémoire sera donc également l'occasion de se familiariser un peu plus avec l'étude d'un auteur dont l'examen de la pensée me semble un projet de recherche particulièrement actuel.

Dans la troisième partie, nous commenterons deux passages choisis des *Ombres de l'esprit*. Le premier se situe au début du premier chapitre, Penrose nous y livre sa position concernant la relation entre calcul et esprit. Il nous décrit également un panorama des points de vue contre lesquels il se place. Nous pourrions ainsi comparer ce panorama à celui que donne Gödel dans le texte de 1961 et comprendre dans quelle mesure Penrose circonscrit sa propre réflexion pour se concentrer sur un nombre restreint d'interrogations. Ce sera là pour nous l'occasion de constater le fait que Penrose pose des limites à la portée de son argumentation, mettant de côté les questions qui n'ont pas, selon lui, un impact direct sur les questions de la calculabilité et la simulabilité de la conscience. Si ce n'est pas la défense de sa conception philosophique globale du monde qui occupe Penrose tout au long de ses deux ouvrages, on en trouve cependant une courte<sup>8</sup> description à la toute dernière section des *Ombres de l'esprit*, faisant office d'ouverture sur une réflexion plus large. Cette fenêtre entrouverte sur la pensée philosophique de Penrose nous permettra de remettre en question la portée de son argumentation d'un point de vue philosophique. Nous nous demanderons ainsi dans quelle mesure la réflexion menée à l'occasion de la discussion sur l'esprit et les machines à calculer appuie ce point de vue philosophique, et inversement si ce point de vue philosophique justifie la valeur, ne serait-ce qu'en

---

8. La section ne fait que 9 pages sur les 408 que totalise le corps du texte de l'ouvrage.

tant que programme de recherche, de la réponse apportée par Penrose à la question de la calculabilité de l'esprit.

# 1 Roger Penrose : Mathématiques, physique, esprit et ordinateurs

## 1.1 Courte biographie de Roger Penrose

Mathématicien de profession, Roger Penrose est un scientifique dont les contributions les plus importantes ont trait à la physique mathématique. La valeur de ses travaux est reconnue par la communauté scientifique comme en témoignent les nombreux prix et médailles qu'il reçoit tout au long de sa carrière<sup>9</sup>. Cependant, Penrose ne publie pas exclusivement des articles scientifiques, il entame ainsi parallèlement à son travail de recherche une réflexion épistémologique sur la science et la connaissance. Il commence à publier certaines de ces réflexions dans le courant des années 1970<sup>10</sup>. Ces réflexions inscrivent ses travaux portant sur des questions précises au sein d'une théorie de la connaissance et d'une conception scientifique du monde qu'il développe dans de nombreuses publications. Les deux ouvrages qui vont occuper notre attention ici sont à classer parmi les textes que Penrose a produit à propos de la science. En effet, ils ne sont pas des travaux scientifiques à proprement parler, pas en tout cas dans l'acception qui est faite de ce terme dans la communauté scientifique, puisqu'ils s'adressent à un public de non-spécialistes et ne présentent pas les faits et l'argumentation avec la rigueur, la clarté et l'objectivité qui caractérisent aujourd'hui la méthode scientifique. Ainsi les spéculations personnelles de Penrose, comme nous allons le voir, occupent une grande place dans ces deux ouvrages. Il me semble tout de même intéressant de se pencher sur la carrière scientifique de l'auteur, car je défends l'idée que certaines de ses expériences en tant que scientifique participent crucialement à la conception du monde que nous essaierons de dégager de l'analyse de deux de ses ouvrages. Cette affirmation est entre autres fondée sur le fait qu'à de nombreuses reprises Penrose s'appuie sur son expérience de chercheur pour nourrir son argumentation. De plus contrairement à certains scientifiques dont les travaux s'orientent vers la philosophie des sciences, Penrose ne divorce pas de la communauté

---

9. Penrose reçoit la médaille Eddington en 1975 décernée conjointement à Stephen Hawking par la Royal Astronomical Society en 1975, la médaille royale décernée par la Royal Society en 1985, le Wolf Foundation Prize for Physics décernée conjointement à Stephen Hawking en 1988, l'ordre du mérite en 2000, la médaille De Morgan en 2004 etc.

10. Penrose, R. (1974) The role of aesthetics in pure and applied mathematical research *Bulletin de l'Institute of Mathematics and its Applications*, 10, n 7-8, 266-271

scientifique pour entrer dans la communauté des philosophes. Penrose au contraire puise dans la science pour faire son épistémologie, proposant un point de vue à la fois surplombant et interne de certaines grandes questions scientifiques. Ainsi puisque mon parti pris est d'analyser la réflexion de Penrose comme un cheminement dont le fondement est sa pratique scientifique et qui s'appuie sur cette dernière pour s'élever vers des questions qui débordent le domaine de compétence qui est le sien, il me semble pertinent de retracer brièvement ici son parcours pour mieux comprendre la relation qu'entretient Penrose avec la recherche scientifique.

Roger Penrose est né en 1931 à Colchester en Angleterre, sa mère Margaret Leathes est médecin et son père Lionel Sharples Penrose est généticien médical et membre de la Royal Society. Son père participa notamment à un groupe de recherche visant à déterminer lesquels des facteurs environnementaux ou génétiques sont les plus significatifs dans l'apparition de maladies mentales. Les parents de Penrose transmettent également à Penrose et ses deux frères leur goût pour les mathématiques, ainsi Oliver de deux ans son aîné s'oriente vers une carrière de mathématicien tournée vers la physique, alors que son petit frère John se tournera, lui, vers la psychologie. Roger obtient son doctorat de mathématiques à St John's College (université de Cambridge) en 1958, avec une thèse portant sur les méthodes tensorielles en géométrie algébrique. Comme son aîné, il se tourne rapidement vers la physique et oriente une partie de ses travaux vers la cosmologie et la théorie de la relativité à partir de 1959<sup>11</sup> tout en publiant des articles purement mathématiques<sup>12</sup>. Il publie en 1965 une de ses contributions les plus importantes, l'article *Gravitational Collapse and Space-Time Singularities* qui à l'aide d'une approche topologique innovante de la théorie de la relativité démontre que des singularités gravitationnelles peuvent apparaître lors de l'effondrement d'une étoile si l'on en croit les équations de la relativité. Par ce résultat, Penrose touche aux limites de cette théorie puisqu'une singularité gravitationnelle est un point de l'espace-temps au voisinage duquel certaines grandeurs caractérisant le champ gravitationnel explosent en des valeurs infinies, ce qui soulève la question de la pertinence de cette théorie vis-à-vis de sa réalité physique. De plus, on ne peut faire de prédiction du comportement physique des objets situés au voisinage d'une singularité en se basant sur la relativité (aussi bien générale que restreinte). Ces travaux l'amèneront notamment à collaborer à plusieurs reprises avec le célèbre physicien Stephen Hawking sur l'étude des trous noirs. Pour dépasser le problème qu'il a découvert, Penrose travaille au développement d'outils mathématiques en vue d'une possible unification entre théorie quantique et théorie relativiste de la gravitation. C'est dans ce but qu'il commence à élaborer en 1967 sa théorie des

---

11. Penrose, R. (1960) *A spinor approach to general relativity*

12. Penrose, R. (1961) *Imbedding of manifolds in euclidean space*

twisteurs qui avait pour vocation de quantifier la gravitation. Bien que prometteuse à ses débuts, cette tentative d'unification est depuis de plus en plus délaissée. Il formule en 1969 l'hypothèse de censure cosmologique portant sur la nature des singularités gravitationnelles qui reste aujourd'hui une des grandes questions de la cosmologie. Penrose est également célèbre pour ses pavages, appelés pavages de Penrose, un ensemble de deux tuiles qui ne pavent le plan que de manière apériodique<sup>13</sup>. Un pavage périodique consiste à recouvrir entièrement l'espace à l'aide d'un seul motif (qui peut être composé de plusieurs tuiles) et que l'on répète par translations périodiques. Ainsi le sol de la plupart des salles de bains est un pavage périodique composé à partir d'une seule tuile : des carreaux carrés ou hexagonaux, accolés périodiquement<sup>14</sup> bord contre bord à d'autres carreaux. Un pavage apériodique recouvre également le plan mais on ne peut le construire en répétant un même motif et en lui faisant subir une translation, la manière qu'ont les pièces de s'assembler pour ne laisser aucun espace vide est différente dans toutes les régions de l'espace. Le problème du pavage a été lancé par Hao Wang et se présente ainsi : existe-t-il une procédure algorithmique permettant de déterminer sans ambiguïté si un ensemble de formes polygonales pave ou non le plan . Il démontra que l'existence d'une telle procédure était équivalente à l'affirmation que tout ensemble de polygones pavant le plan pouvait le paver de manière périodique. Berger en 1966 en suivant des indications de Wang démontra l'existence d'un ensemble de 20426 polygones ne pavant le plan que de manière apériodique, montrant ainsi qu'il n'existait pas de procédures de décision pour le problème du pavage. Penrose dans le seul but de créer un divertissement mathématique (les pavages sont une interface de longue date entre mathématiques, disciplines artistiques et jeux) trouva en 1974 une méthode pour produire des pavages apériodiques d'une grande simplicité et avec uniquement deux polygones qui sont désormais l'exemple typique de pavages apériodiques . Ces pavages se révélèrent à sa grande surprise jouer un rôle central en physique pour l'étude et la compréhension des quasi-cristaux découverts dans les années 1980. Une caractéristique surprenante de ces réseaux cristallins basés sur des pavages apériodiques est qu'ils sont non-locaux, i.e. pour placer correctement une pièce lorsque l'on construit ce genre de pavages, il faut examiner les conséquences des différentes possibilités sur l'état du pavage à grande distance de la région en construction.

Les pavages de Penrose ne sont pas sa seule contribution dans le domaine des mathématiques dites "récréatives". Après avoir visité une exposition de l'artiste néerlandais M. C. Escher en 1954, il lui transmet un article qu'il a coécrit avec son père portant

---

13. Voir figure 1

14. Ici la périodicité est de 1 puisqu'il suffit à partir d'un carreau de se translater d'un carreau pour se retrouver sur un carreau identique à celui de départ.

sur les figures impossibles auxquelles il avait commencé à réfléchir, parmi lesquelles le célèbre triangle<sup>15</sup> de Penrose, ou tribarre. L'échange et l'amitié qui s'en suivra inspireront l'artiste pour certaines de ses oeuvres les plus célèbres telles que Waterfall (1961) ou encore Ascending and Descending (1960) dans lesquelles il utilise l'illusion de la perspective pour représenter des objets impossibles à construire dans l'espace euclidien à 3 dimensions.

Penrose publie également, au moins depuis les années 1970, des textes se rapprochant de la philosophie des sciences, comme par exemple l'article de 1979 "*Einstein's vision and the mathematics of reality*"<sup>16</sup> ou encore l'article de 1987 "*Newton, quantum theory and reality*"<sup>17</sup>. Il s'intéresse tout particulièrement aux questions des limites, de l'interprétation et de l'ontologie de la physique quantique ainsi qu'à la question de la pensée consciente. En 1987 déjà il établit un lien entre les mystères de la physique quantique et ceux du fonctionnement de notre conscience dans l'article *Quantum physics and Conscious thought*<sup>18</sup>. Cependant c'est dans les deux monographies *The Emperor's New Mind. Concerning computers, minds, and the laws of physics* et *Shadows of the Mind. A search for the Missing of Consciousness* qu'il publie en 1989 et en 1994 respectivement, que Penrose livre de manière plus complète son point de vue sur la question de l'esprit et sur la question de la place des mathématiques vis-à-vis de la réalité physique. C'est la pensée que développe Penrose dans ses deux derniers ouvrages que nous étudierons dans le présent mémoire. Pour cette étude nous nous baserons sur la traduction française par F.Balibar et C.Tiercelin pour InterEditions de *The Emperor's New Mind* publié en France sous le titre *L'Esprit, l'Ordinateur et les Lois de la Physique* auquel nous ferons référence par l'acronyme EOLP ainsi que sur la traduction française de Christian Jeanmougin pour InterEditions de *Shadows of the Mind* publié en France sous le titre *Les Ombres de l'Esprit. À la recherche d'une science de la conscience* auquel nous ferons référence par l'acronyme OE. Par abus de langage je ferais référence à "la pensée de Penrose" ou encore à la "position de Penrose" pour faire référence à la pensée et la position qu'il défend dans ces deux ouvrages, ce qui ne préjuge en rien de sa pensée actuelle ou passée sur ces sujets, auxquelles je ne me suis pas intéressé.

Pour conclure ce retour en arrière dans la carrière de Penrose, je voudrais y relever les quelques éléments qu'il me semble pertinent de garder à l'esprit, car ils auront

---

15. Voir figure 2

16. Penrose, R. (1979) Einstein's vision and the mathematics of reality. *The Sciences* ; (mars), 6-9

17. Penrose, R. (1987) Newton, quantum theory and reality in *300 years of gravity*. Cambridge University press

18. Penrose, R. (1987) *Quantum physics and Conscious thought* In *Quantum implications : Essays in honour of David Bohm*, Routledge and Kegan Paul

leur importance par la suite :

1. Les notions de calculabilité et de décision par un algorithme sont présentes dans les travaux portant sur les pavages apériodiques et sont également centrales dans l'interprétation de la physique quantique. En effet, l'aspect probabiliste de la théorie quantique empêche toute prévision de l'évolution d'une situation physique à l'aide d'une simulation par une machine de Turing. Un comportement non-local est également présent dans la théorie quantique et dans la construction des pavages de Penrose.
2. Penrose semble porter un certain intérêt à l'activité de recherche et de raisonnement mathématique en soi : la place de l'intuition, de l'imagination ou encore de l'esthétique dans ce processus mental. En témoignent ses recherches sur les pavages, son article sur la place de l'esthétique dans la recherche mathématique<sup>19</sup> et son amitié avec l'artiste Escher.
3. Penrose a travaillé aussi bien sur la théorie de la relativité que sur la physique quantique. À travers ses recherches il a pris acte de la difficulté dans laquelle est plongée la physique par la coexistence de ces deux branches de la physique en l'absence d'un cadre pour les unifier. Ses tentatives pour dépasser ces difficultés et la conviction que cette situation va changer et qu'une nouvelle théorie devrait venir mettre de l'ordre dans une physique qui ne donne pour le moment que des réponses partielles seront essentielles pour comprendre la réponse proposée par Penrose à la question de l'émergence de l'esprit à partir de la matière inerte.

## 1.2 Présentation de EOLP et OE

### 1.2.1 Calculabilité

Avant de présenter succinctement le contenu des deux ouvrages de Penrose qui vont concentrer notre attention ici, il me semble nécessaire de définir un terme dont nous ferons usage tout au long de ce mémoire. La notion de calculabilité est d'une grande importance tant dans les ouvrages de Penrose que dans les travaux de Gödel. Nous qualifierons ici de calculable toute opération exécutable par une machine de Turing. Cette acception du terme de calculabilité et celle utilisée par Penrose dans EOLP et OE. Une machine de Turing est un modèle abstrait auquel peut se ramener toute opération de calcul mécanique et déterministe, c'est-à-dire tout algorithme. L'immense majorité des ordinateurs actuels fonctionnent comme des machines de Turing si on

---

19. Penrose, R. (1974) The role of aesthetics in pure and applied mathematical research *Bulletin de l'Institute of Mathematics and its Applications*, 10,n7-8, 266-271

idéalisent leur fonctionnement. Définir ainsi la calculabilité présente deux avantages par lesquels Penrose justifie cette définition de la calculabilité.

En premier lieu, la notion de calculabilité est bien définie dans le sens où la question "*L'opération  $T$  est-elle calculable ?*" admet toujours une réponse. En effet, il est possible de définir une machine de Turing universelle pouvant simuler le fonctionnement de n'importe quelle machine de Turing. Ainsi, une opération est calculable si et seulement si elle peut être exécutée par une machine de Turing universelle.

En second lieu, le concept de machine de Turing formalise de manière relativement étendue la notion informelle de calcul, d'algorithme ou de procédure mécanique. En effet autour des années 1936-1937 plusieurs logiciens construisirent des systèmes de procédures visant à formaliser cette notion, nous savons d'ailleurs que Gödel lui-même apporta sa réponse à ce problème en 1934, aidé par le logicien français J. Herbrand, grâce à sa définition des fonctions récursives générales. Le logicien américain A. Church créa le lambda-calcul et le mathématicien britannique A. Turing créa le concept de machine de Turing, enfin le mathématicien E. Post conçut une procédure proche de celle de Turing. Turing et Church démontrèrent que ces différentes définitions de la calculabilité étaient équivalentes, malgré le fait que les systèmes de Church et de Turing semblaient envisager le problème sous un angle différent. Malgré les progrès de l'informatique, les machines à calculer, ou ordinateurs, actuels restent des machines de Turing ce qui nous laisse penser que cette définition fonde de manière satisfaisante le concept intuitif de calcul. Cette affirmation est connue sous le nom de *Thèse de Church-Turing*. Bien que le concept de machine de Turing soit orienté vers l'exécution d'une opération par une machine, et donc comme une idéalisation d'un dispositif hardware exécutant un programme, la *Thèse de Church-Turing* est renforcée par le concept de système formel Turing-complet qui garantit que la puissance de calcul d'un système formel Turing-complet est équivalent à celui d'une machine de Turing. Ainsi les théorèmes d'incomplétude de Gödel s'appliquent aux systèmes formels Turing-complets. Un équivalent de ces théorèmes dans le cas des machines de Turing est alors le théorème de l'arrêt démontré par Turing en 1936 et d'après lequel une machine de Turing ne peut décider pour toute machine de Turing  $T$  et tout paramètre  $x$  si le calcul  $T$  appliqué à  $x$  s'arrête. Le problème de l'arrêt est donc un problème dont la réponse n'est pas calculable, pourtant les programmes s'arrêtent effectivement ou ne s'arrêtent pas.

Puisque c'est aux évolutions non-calculables que s'intéresse Penrose, il nous en donne un certain nombre en exemple. J'en résumerai un <sup>20</sup> ici afin de donner l'intuition de la différence entre évolution déterministe et évolution calculable. Cette différence permet à Penrose à la fois de s'opposer définitivement aux partisans d'une vision

---

20. L'exemple du pavage est donné dans OE p27-29

algorithmique de l'esprit tout en gardant la solution du problème de l'esprit dans le domaine de la physique traditionnellement déterministe<sup>21</sup>. Nous avons déjà fait remarquer que le problème du pavage<sup>22</sup> n'était pas calculable, il est donc impossible d'écrire un algorithme ou de programmer un ordinateur pour qu'il sache systématiquement si un ensemble de pièces polygonales pave le plan ou pas. Supposons que nous possédions une liste infinie  $S_0, S_1, S_2, \dots$  de tels ensembles de pièces et imaginons un univers très étrange dans lequel l'état de l'univers à un temps  $t$  serait donné entièrement par un de ces ensembles  $S_n$ . L'évolution se ferait à temps discret selon cette loi : si l'état de l'univers au temps  $t$  est  $S_n$  alors à  $t+1$  l'univers sera dans l'état  $S_{n+1}$  si  $S_n$  pave le plan et dans l'état  $S_{n+2}$  sinon. Cette évolution serait totalement déterministe, mais ne serait pas calculable.

### 1.2.2 Présentation conjointe de EOLP et OE

Je vais maintenant présenter brièvement le contenu de EOLP et OE. Avant de m'attarder sur les différences et les particularités de chacun de ces ouvrages, leurs ressemblances me permettent de les présenter conjointement. EOLP et OE sont deux essais, respectivement d'environ 500 et 400 pages, qui se veulent accessibles aux non-spécialistes. EOLP et OE se rapprochent par certains aspects de l'ouvrage de vulgarisation, mais s'en éloignent résolument par d'autres. Ainsi des indications sont données au lecteur l'invitant à ignorer certains passages techniques<sup>23</sup>, l'engageant à ne pas rester bloqué devant une équation. De plus lorsqu'un contenu scientifique tel que la théorie de la relativité ou le théorème de Gödel prend une place importante dans l'argumentation, il est replacé dans son contexte historique et expliqué souvent bien au-delà des besoins de l'argumentation. Transmettre sa fascination pour la science et en donner un certain goût au lecteur pour ainsi la faire sortir de son image figée, académique et anti-métaphysique est, à mon avis, un des objectifs secondaires

---

21. La physique quantique est souvent interprétée comme non-déterministe, cependant il y a eu et il y a toujours des physiciens rejetant cette interprétation et préférant rester dans le cadre classique de la physique déterministe. Ainsi le comportement en apparence probabiliste pourrait n'être qu'une illusion émergeant d'une évolution déterministe cachée. On pense évidemment à la citation d'A.Einstein "*Dieu ne joue pas aux dés avec le monde*".

22. Dans l'exemple de Penrose ce n'est pas le pavage du plan par des polygones, mais par des polyminos qui est traité. Les polyminos sont des polygones particuliers constitués d'une juxtaposition de carré unité identiques. Les pièces du célèbre jeu Tetris sont des polyminos. La difficulté du problème du pavage général étant que les nombres réels sont nécessaires pour définir un polygone quelconque alors que les entiers naturels sur lesquels peuvent opérer les machines de Turing sont suffisants pour définir les polyminos.

23. Voir la *Note au Lecteur sur la manière de lire les équations mathématiques* pXV de EOLP et la *Note au Lecteur* pXIII-XIV



de EOLP et dans une moindre mesure de OE. Cela permet à Penrose d'affirmer son point de vue comme le point de vue d'un homme de sciences qui est au fait des enjeux et des avancées de cette dernière. Pour appuyer ceci notons le fait que le vulgarisateur Martin Gardner, célèbre pour ses colonnes « Jeux mathématiques », qui parurent pendant vingt-ans dans *Scientific American*, signe la préface de EOLP. Ainsi bien qu'il considère que son opinion sur le problème de la conscience soit non-orthodoxe dans le milieu scientifique, elle n'est pas pour autant ressentie comme opposée à l'esprit scientifique. La question de la conscience étant une question ancienne, profonde et touchant à des disciplines très différentes, elle est un point d'accès privilégié pour aborder un grand nombre de sujets à travers elle. Mais les objectifs principaux et affichés des ouvrages sont tout autres :

1. En premier lieu, donner des arguments négatifs pour attaquer la position dite de l'IA <sup>24</sup> forte selon laquelle la conscience peut se réduire à un algorithme, et ce indépendamment du support sur lequel est exécuté cet algorithme : cerveau, circuits d'ordinateur ou autres. La science, la pensée rationnelle ne se bornent pas à la description ou l'exécution de calculs, la notion de certitude en mathématiques par exemple dépasse ce cadre. Cette dernière affirmation repose sur l'interprétation de Penrose du théorème de Gödel. Penrose défend, de plus, le fait que cette non-calculabilité ne puisse être comprise dans le cadre de la physique tel qu'il est au moment de la rédaction des ouvrages. Ainsi toute simulation basée sur cette physique est vouée à l'échec, y compris toutes les tentatives des défenseurs de l'IA forte.
2. En second lieu, convaincre de la possibilité d'une alternative à cette vision informatisée de l'esprit qui permette une compréhension scientifique du phénomène de conscience en s'appuyant sur les mathématiques, la physique et la biologie. L'idée étant que ce qui est calculable ne représente qu'une petite partie des mathématiques et que des opérations non-calculables pourraient avoir leur place au sein d'une théorie physique future. Il serait dès lors possible d'imaginer des processus biologiques non-calculables, basés sur des phénomènes encore mal compris, tels que l'intrication quantique, qui expliqueraient la non-calculabilité constatée à l'occasion du premier objectif.

Ces objectifs sont d'autant plus clairs dans OE puisque cet ouvrage est découpé en deux parties, chacune se concentrant sur un des objectifs que nous venons de détailler. De cette manière dans la *Partie 1 : Pourquoi il faut une nouvelle physique pour comprendre l'esprit*, Penrose argumente l'impossibilité de simuler la conscience

---

24. IA est l'abréviation de Intelligence Artificielle, on trouve parfois, y compris dans les écrits en français, l'abréviation AI correspondant à l'anglais Artificial Intelligence

en se basant sur la théorie physique existante, et dans la *Partie 2 : Quelle nouvelle physique pour l'esprit ?* il indique pourquoi et comment, l'évolution de la physique va, selon lui, lever cette impossibilité. Ce choix de clarifier la structure de l'argumentation dans OE nous amène à nous interroger sur les différences entre les deux ouvrages.

### 1.2.3 Différences de forme et de fond entre EOLP et OE

#### Différences de forme

Commençons par les différences dans la forme. Tout d'abord il faut préciser que OE est paru cinq années après EOLP et que Penrose dans la préface d'OE le désigne comme une suite de EOLP pouvant être lue indépendamment de lui. Selon cette préface toujours, Penrose à la suite de la publication de EOLP a reçu de nombreux commentaires, questions et critiques qui ont, en partie, motivés sa genèse. Pour faire place aux précisions et aux développements que désire écrire Penrose, l'aspect vulgarisateur est moins présent que dans EOLP. Ainsi Penrose dans OE passe assez vite sur la description du concept de machine de Turing ou encore sur celui d'entropie, préférant renvoyer le lecteur vers EOLP dans lequel la progression est plus lente et le style moins rigide. Il n'hésite par contre pas à répondre point par point aux critiques soulevés par son interprétation du théorème de Gödel le long de 30 pages parfois assez techniques. Si la note au lecteur est relativement proche de celle d'EOLP, invitant le lecteur à sauter certaines parties ou à n'en lire qu'une partie, le niveau général de difficulté y est globalement supérieur. En effet outre le fait d'accorder moins de temps à la vulgarisation, Penrose se fait plus prudent et rigoureux dans OE, proposant trois appendices dans lesquels sont donnés certains détails mathématiques et physiques destinés « *aux lecteurs qui veulent se persuader de la solidité des arguments* »<sup>25</sup>. S'il laisse encore de la place à la spéculation, Penrose semble avoir acquis une certaine rigueur dans son écriture qui lui faisait parfois défaut dans EOLP. Réciproquement son style a perdu en poésie, en enchantement et en capacité à nous faire voyager par, et pour, la science.

#### Différences de fond

Les différences de fond se situent principalement dans la seconde partie de OE, correspondant au second objectif qui est de proposer une "*physique de l'esprit*". Il est tout d'abord nécessaire de préciser quelques détails de l'argumentation de Penrose. La première partie s'appuie presque exclusivement sur le théorème de Gödel et sur

---

25. OE, Note au Lecteur pXIII

la réappropriation par Penrose de l'interprétation qu'en avait faite en 1961 le philosophe John Lucas<sup>26</sup> et a pour objectif d'en déduire le fait qu'il ne peut y avoir d'explications calculables de la connaissance et la compréhension des mathématiques par les mathématiciens. Par voie de conséquence, il n'est pas possible de fonder l'activité consciente en général uniquement sur des procédures calculables. Penrose déploie un certain nombre d'arguments, pas toujours définitifs, pour appuyer son hypothèse : c'est au-delà de la science actuelle qu'il faut chercher l'origine de ce processus non-calculable. En effet, le hasard, le chaos, l'environnement, les algorithmes auto-apprenants, le calcul parallèle ou encore le calcul analogique pourraient peut-être échapper à la notion de calculabilité que nous avons précisée. Penrose tente de nous convaincre qu'en pratique il n'en est rien et que ces explications ne lui paraissent pas satisfaisantes, ce qui l'amène à se poser la question qui donne son titre à la seconde partie d'OE, *Quelle nouvelle physique pour l'esprit ?*. Dans EOLP comme dans OE, Penrose postule que les difficultés posées par la physique quantique et celles qui entourent le problème de l'esprit ont une même origine, notre méconnaissance des principes profonds qui régissent la physique et dont la physique quantique ne serait qu'une approximation. Les principaux problèmes de la physique quantique selon Penrose se situent au niveau de l'effondrement de la fonction d'onde, aussi appelée réduction du vecteur d'état, et au niveau de son incompatibilité avec la théorie de la relativité. Cette incompatibilité est particulièrement visible à travers les phénomènes non-locaux d'intrication quantique qui contredisent l'approche locale qu'exige la théorie de la relativité générale. Penrose spéculer sur ce que pourrait être une physique qui permettrait de dépasser ces problèmes et comment elle pourrait intervenir dans le fonctionnement du cerveau pour donner naissance à des processus non-calculables.

Ce sont certains détails techniques de cette nouvelle physique qu'il envisage qui sont la grande différence entre EOLP et OE. Dans EOLP, Penrose propose un critère de déclenchement de la réduction du vecteur d'état avant de revenir dessus et d'en donner un autre dans OE<sup>27</sup>. De même, la manière dont intervient le phénomène d'intrication quantique au niveau du cerveau est elle aussi différente dans OE et dans EOLP<sup>28</sup>. Puisque OE est postérieur à EOLP et qu'il précise et complète ce

---

26. Lucas, J.R. (1961) *Minds, Machines and Gödel* in *Philosophy*, XXXVI, pp.112-127

27. Dans EOLP le critère est le suivant : si deux états superposés entraînent deux espaces-temps dont la courbure diffère de plus d'un graviton, alors la réduction a lieu. Dans OE Penrose propose un autre critère : si les géométries de l'espace-temps qu'impliqueraient deux états superposés s'éloignent de plus de la distance de Planck, alors cela déclenche la réduction.

28. Dans EOLP Penrose postule qu'il puisse y avoir des superpositions d'états quantiques concernant l'activation d'un neurone. Il revient sur cette affirmation dans OE car il ne la croit finalement pas plausible au vu de certains faits scientifiques, principalement le fait que les neurones soient trop

dernier, nous privilégierons les citations d'OE à celles d'EOLP, bien que ce choix ne soit nullement exclusif.

Nous venons de le voir, la spéculation semble occuper une place importante dans les ouvrages de Penrose. Nous allons désormais éclaircir ce que nous entendons par spéculation et quel rôle elle joue dans le développement de la pensée de Penrose.

### 1.3 La place de la spéculation chez Penrose

La spéculation joue un rôle moteur dans le progrès de la connaissance en sciences. Ainsi tirer des conséquences d'un postulat non établi scientifiquement permet de confronter à l'expérience ces conséquences afin d'infirmier le postulat ou bien de renforcer sa plausibilité. Au vu de la difficulté à unifier théorie quantique et classique, la spéculation à ce sujet est depuis quelques dizaines d'années une part conséquente du travail de recherche en physique fondamentale. Ainsi sont nées une multitude de remplaçants hypothétiques à la physique quantique et à la relativité comme la théorie des cordes, la supergravitation, la gravitation quantique à boucles ou encore la théorie des twisteurs de Penrose. Il est donc impératif de ne pas exclure la spéculation de la science et de distinguer ce qui est scientifiquement établi de ce qui participe au progrès scientifique. Penrose fait cette distinction clairement, donnant des références qui permettent de vérifier les faits scientifiques sur lesquels il s'appuie et indiquant les raisons qui le poussent à spéculer la possibilité de l'existence de phénomènes que la communauté scientifique n'a pas établi, comme par exemple la possibilité d'une intrication quantique au sein des microtubules des neurones. Penrose annonce dès les premières pages de EOLP<sup>29</sup> l'importance de la spéculation dans sa démarche, particulièrement dans la seconde partie pour ce qui est de décrire ce que pourrait être une physique de l'esprit non-calculable<sup>30</sup>. Je n'ai pas toujours la culture scientifique nécessaire pour émettre un avis critique sur la plausibilité scientifique des différentes spéculations de Penrose. Bien que je puisse être critique concernant certaines de ces spéculations, je ne me prononcerais pas systématiquement sur ce sujet. J'ai donc

---

peu isolés de leur environnement pour rester dans un état de cohérence quantique. Dans OE la physique quantique rentre en jeu à une échelle plus petite, au niveau intracellulaire. En effet, c'est la conformation de certaines protéines présentes à l'intérieur du cytosquelette des cellules, notamment neuronales, qui pourrait former un système quantique intriqué étendu de type condensat de Bose-Einstein.

29. EOLP p4 *"Mes propres conjectures joueront un rôle non négligeable dans ce qui va suivre."*

30. OE préface p10 *"Dans la deuxième partie, mes arguments font intervenir la physique et la biologie. Bien qu'ils soient indéniablement plus spéculatifs que la discussion rigoureuse de la première partie, ils tentent véritablement de comprendre comment de telles actions indépendantes de tout calcul pourraient survenir dans le cadre de lois physiques scientifiquement compréhensibles."*

choisi de ne pas analyser la pertinence de l'anticipation scientifique de Penrose mais de m'y intéresser en tant que programme de recherche et comme un exemple de ce que pourrait être une physique de l'esprit non-calculable dont l'intérêt est dans sa possibilité plutôt que dans son contenu précis. Peut-être plus que le contenu des ouvrages, l'apport de Penrose est donc celui de sa démarche gnoséologique. C'est pour prouver par l'exemple, même s'il repose sur la spéculation scientifique, qu'il est et sera possible de progresser dans la connaissance tout en respectant le principe de parcimonie et la démarche expérimentale que Penrose s'abstrait de ces deux derniers. Ainsi l'établissement d'une nouvelle théorie respectant ces deux principes peut nécessiter que le scientifique les outre passe durant le processus de recherche. La spéculation peut ainsi catalyser ce processus de recherche, en particulier lorsque les scientifiques se retrouvent confrontés à une difficulté théorique<sup>31</sup> qui perdure. Le jeu consistant à rester dans une démarche scientifique tout en remettant en cause certains éléments du cadre scientifique dans lequel elle s'inscrit<sup>32</sup>.

Il faudra bien distinguer dans la suite la spéculation scientifique que nous venons d'évoquer de la spéculation métaphysique de Roger Penrose. Ce que j'entends par spéculation métaphysique, ce sont les postulats ou les conceptions philosophiques sur lesquels se base Penrose tout au long de ses ouvrages sans prendre le temps de les établir rigoureusement, bien qu'il insiste encore une fois sur leur plausibilité à ces yeux. Nous présenterons plus avant ces spéculations, principalement son platonisme et comment s'inscrit la science dans ce platonisme. Ce sont donc des spéculations qui dépassent le cadre de la science et à l'intérieur desquelles a lieu l'argumentation. Un reproche qui peut être fait à Penrose est qu'il ne s'arrête sur ces spéculations métaphysiques que tardivement et brièvement. De plus, le lien unissant spéculations métaphysiques et spéculations scientifiques n'est pas toujours limpide. Ainsi certaines spéculations scientifiques, notamment concernant la physique quantique, sont jugées plausibles par Penrose non pas pour des raisons scientifiques, mais plutôt parce qu'ils

---

31. EOLP p 377 *"il n'y a pas d'autre chemin que celui de la spéculation"*

32. OE p9 *"Bien qu'il occupe une situation très inconfortable dans la connaissance scientifique actuelle, le problème de l'esprit ne doit pas, selon moi, être considéré comme définitivement hors de portée de la science. Si celle-ci est encore incapable de dire grand-chose de sensé sur ce problème, elle finira probablement par élargir son champ d'action de manière à le prendre en compte - au besoin en modifiant ses propres procédures."*

sont plus cohérents avec sa position réaliste<sup>33 34</sup>.

## 2 Kurt Gödel 1961 : *The modern development of the foundations of mathematics in the light of philosophy*

Nous venons de voir que bien que s'appuyant très largement sur les sciences, les ouvrages de Penrose ne sont pas exempts de postulats philosophiques. Afin de mieux cerner ces présupposés, nous allons désormais tenter de nous armer d'un certain nombre de concepts par l'étude du texte de 1961 de Gödel. Ce sera pour nous l'occasion d'éclairer la remise en question de la science que nous venons de commenter lors de notre discussion sur la spéculation dans l'oeuvre de Penrose. Les théorèmes d'incomplétude sont, de plus, la pierre angulaire du raisonnement de Penrose, il me semble intéressant de comprendre quelle place accordait à ces propres théorèmes Gödel puisqu'il naviguait lui aussi, bien que d'une manière très différente, entre philosophie et science. Nous allons donc présenter ici brièvement Kurt Gödel et le texte sur lequel nous allons nous appuyer pour notre analyse afin de mieux comprendre comment se place le logicien relativement à différents domaines de la connaissance.

### 2.1 Courte biographie de Gödel

Kurt Gödel<sup>35</sup> naquit en 1906 dans la ville de Brno, au sein de l'empire austro-hongrois, dans l'actuelle République Tchèque. Il entame en 1924 des études universitaires de physique à Vienne. Son intérêt pour la précision et les cours de mathématiques qu'il suit l'amènent à se tourner vers les mathématiques puis la logique

---

33. Penrose affirme comme postulat philosophique l'existence d'une réalité objective et le rejet des positions subjectivistes *"Je pose en hypothèse qu'un point de vue philosophique ne peut être qualifié de "sérieux" que s'il contient une certaine dose de réalisme. [...] Si j'adopte, chaque fois que cela m'est possible, une ligne de pensée réaliste, ce n'est nullement par manque d'information de ma part : je suis parfaitement au courant des positions subjectivistes adoptées par certains - simplement, ces conceptions n'ont aucun sens pour moi."*

34. Dans OE p301 Penrose donne une des clefs de son interprétation de la physique quantique, qui conditionnera la théorie quantique spéculative qu'il proposera. *"Certes, le monde quantique a un comportement étrange, insolite, mais il n'est pas "irréel". Comment d'ailleurs pourrait-on construire des objets réels à partir de constituants irréels ?"*

35. Les informations biographiques à concernant Gödel sont tirés du texte de H.Wang *Some facts about Kurt Gödel* rédigé entre 1976 et 1977 avec le concours de Gödel et approuvé par ce dernier, ainsi que des notes de Wang ajoutées postérieurement à ce texte, finalement publié en 1981.

mathématique. Il migre ainsi en 1926 vers une formation en mathématiques. Il suit pendant les années qui suivent les cours de philosophie de H.Gomperz et assiste aux réunions du cercle de Vienne dirigé par M.Schlick sans pour autant en partager toutes les thèses. Il démontre en 1929 dans sa thèse doctorale la complétude de la logique du premier ordre et c'est en 1931 qu'il prouve ces célèbres théorèmes d'incomplétude<sup>36</sup>, mettant fin par la même occasion au programme d'axiomatisation des mathématiques de Hilbert. À la suite de l'Anschluss en 1938, Gödel s'installe à l'Institute for Advanced Study de Princeton où il avait voyagé et donné des conférences à plusieurs reprises depuis 1933. Il est aidé pour cela par Einstein qu'il avait rencontré lors de ces séjours à Princeton. Il poursuit ses travaux en logique mathématique en se concentrant sur la théorie des ensembles et démontre en 1938<sup>37</sup> la cohérence relative de l'axiome du choix et de l'hypothèse du continu avec ZF. Motivé pour cela par ses études de Kant et ses discussions avec Einstein, il travaille entre 1947 et 1951 à une solution paradoxale des équations de la relativité générale qui autorise les voyages dans le temps, connue sous le nom "d'univers tournants" ou "univers de Gödel"<sup>38</sup>. Il étudie E.Husserl à partir de 1959. Le texte sur lequel nous allons nous appuyer datant de 1961, je ne présenterai pas les travaux de Gödel postérieurs à ce texte. La remarque suivant la biographie de Penrose est d'autant plus valable concernant Gödel puisque l'étude de sa pensée est toujours en cours. Ainsi lorsque nous ferons référence à la pensée de Gödel ou encore au projet de Gödel, ces expressions référeront à sa pensée et son projet tels qu'ils nous sont présentés dans le texte de 1961. Nous retiendrons de cette courte biographie que les travaux les plus importants de Gödel ont trait aux fondements des mathématiques et qu'il connaît bien la théorie de la relativité. Nous pouvons également remarquer son intérêt pour la rigueur et la précision et le fait qu'il entretienne des liens avec la philosophie depuis ses études jusqu'à 1961. Nous constatons que cette proximité est très différente de celle que nous avons pu trouver chez Penrose qui était amené à réfléchir à des questions d'ordre philosophiques par les travaux qu'il avait entrepris pour d'autres raisons. Pour Gödel il semblerait qu'inversement ce soient ses réflexions philosophiques qui aient influencé la direction et le choix de ces travaux scientifiques comme il le laisse entendre dans cette citation :

*" My work is an application of a philosophy suggested outside of science and obtained*

---

36. Gödel, K. (1931), "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I." Monatshefte für Mathematik und Physik 38 : 173-198.

37. Gödel, K. (1938) « *The Consistency of the Axiom of Choice and of the Generalized Continuum-Hypothesis* », PNAS, vol. 24, no 12

38. Ne doit pas être confondu avec  $L$  : l'univers constructible de Gödel qui est un modèle de ZF.

*on the occasion of thinking about science.*<sup>39</sup> "

## 2.2 Le texte de 1961

Le titre du texte sur lequel nous allons nous appuyer confirme cette primauté de la philosophie sur la science dans la démarche de Gödel. Contrairement à Penrose qui cherche à éclairer des problèmes philosophiques à l'aide de savoirs scientifiques dans OE et EOLP, le texte de 1961 de Gödel s'intitule *The modern development of the foundations of mathematics in the light of philosophy*. Selon moi nous avons bien là une hiérarchisation des savoirs, Gödel étant un des acteurs majeurs du développement des fondements des mathématiques et un expert dans ce domaine, il m'apparaît que si Gödel considère nécessaire de faire appel à la philosophie pour éclairer l'histoire des fondements des mathématiques, c'est parce qu'il est selon lui illusoire de vouloir la comprendre entièrement en se basant uniquement sur les mathématiques elles-mêmes. L'évolution des recherches et du consensus scientifique dans ce domaine ne peuvent être compris que comme une conséquence d'une évolution plus générale de la pensée et des *a priori* philosophiques. Concernant ces *a priori*, Gödel utilise le terme allemand de *Zeitgeist* que l'on traduira par esprit du temps et par lequel on désignera le paradigme intellectuel et culturel d'une certaine période.

Le texte de 1961 est le texte préparatoire d'une conférence que Gödel ne donnera finalement jamais, il est donc traduit<sup>40</sup> et publié de manière posthume en 1981 dans les *Collected works Vol III* aux côtés de nombreux autres textes non publiés par Gödel. En effet, l'exigence de ce dernier l'empêchait de publier des textes n'atteignant pas le niveau de rigueur qui était le sien. Ainsi, Gödel ne publia pas non plus le texte *Is Mathematics a Syntax of Language?* duquel il rédigea pourtant plusieurs manuscrits entre 1953 et 1959 et qui était destiné à être publié dans une monographie portant sur Carnap. Gödel replace dans le texte de 1961 le développement des fondements des mathématiques depuis le début du siècle ainsi que ses contributions à ce dernier, principalement les théorèmes d'incomplétude, dans un contexte philosophique et épistémologique plus général. Enfin, il nous fait part de sa position au sein de ce contexte, la justifiant et proposant une esquisse de marche à suivre pour poursuivre le projet qui devrait déboucher selon lui sur une plus grande compréhension aussi bien des mathématiques que de l'individu et du monde qui l'entoure. Au premier paragraphe du texte de 1961, Gödel introduit une classification des diffé-

---

39. Wang, H. (1996), *A Logical Journey : From Gödel to Philosophy*. Cambridge, MIT Press p297

40. On trouvera dans les *Collected works Vol III* la traduction vers l'allemand en plus de celle vers l'anglais sur laquelle nous nous sommes basés ici, le manuscrit original étant rédigé en Gabelsberger, une écriture sténographiée.



rentes conceptions du monde qui lui est propre, nous allons maintenant nous pencher sur cette classification, classification que nous réutiliserons *a posteriori* pour tenter d'analyser la position de Penrose.

## 2.3 La classification des différentes conceptions du monde du texte de Gödel 1961

### 2.3.1 Critères de la classification

Il est de toute première instance de revenir sur l'expression conception du monde. Le mot original utilisé par Gödel est celui de *Weltanschauung*, mot qui comme *Zeitgeist* doit être compris au vu de son importance et de ses différents usages dans la philosophie et la littérature germanique. Si nous ne nous étendons pas outre mesure sur l'histoire du terme *Weltanschauung* ici, il est utile de souligner les notions que sous-entend son usage. D'une part, ce mot peut faire référence à une vision du monde et de la vie globale à l'intérieur de laquelle se développe la pensée d'un individu. D'autre part, ce terme peut intégrer l'idée de système philosophique complet fixant la nature ontologique de l'individu et des entités constituant le monde ainsi que de la possibilité de connaissance de ces entités par l'individu. C'est donc le terme *Weltanschauung* que nous viserons lorsque nous utiliserons l'expression conception du monde.

La classification des différentes conceptions du monde se fait selon deux critères qui entretiennent une certaine dépendance, mais ne sont pas pour autant identifiables :

1. Le lien qu'entretiennent les différentes conceptions du monde avec la métaphysique ou la théologie, la force des postulats ontologiques qu'elles admettent et sur lesquels elles se fondent. C'est ce critère qui est donné par Gödel<sup>41</sup> pour expliquer sa classification, mais aussi pour la justifier, puisque c'est selon lui le plus fécond. Nous reviendrons sur les raisons qui fondent cette fécondité un peu plus bas.
2. La possibilité d'une connaissance objective des entités reconnues comme constituant le monde. Ce critère n'est pas explicitement donné par Gödel en tant que tel, cependant il est possible d'imaginer des classifications répondant au premier critère et ne correspondant pas exactement à la classification que nous présente Gödel. En effet, il est possible d'obtenir différentes classifications en

---

41. Gödel, K. (1961) *Collected works Vol III* (1995) p375 "I believe that the most fruitful principle for gaining an overall view of the possible world-view will be to divide them up according to the degree and the manner of their affinity to or, respectively, turning away from metaphysics (or religion)."

faisant varier l'acception de l'expression "lien avec la métaphysique". Dégager ce second critère épistémologique est donc un choix de l'auteur qui ne préjuge en rien du possible usage qu'aurait pu faire Gödel de cette classification dans d'autres contextes et n'est pertinent qu'au vu de l'usage qui est fait de cette classification dans le texte de 1961 dont les mathématiques et la science sont le sujet central. On peut ainsi lire au second paragraphe.

*"Furthermore one sees also that optimism belongs in principle toward the right and pessimism toward the left. For skepticism is certainly a pessimism with regard to knowledge."*

En partant de ces critères, Gödel nous donne une liste ordonnée de six différentes conceptions du monde. Voici cette liste telle qu'elle apparaît dans le premier paragraphe.

*"In this way we immediately obtain a division into two groups : skepticism, materialism and positivism stand on oneside, spiritualism, idealism and theology on the other."*

La classification en deux groupes sera importante pour envisager les cas mixtes, sujet qui retiendra notre attention. Cependant, la liste donnée par Gödel n'est pas une simple dichotomie puisque la liste est une liste ordonnée, soulignant de cette manière les différences graduelles qu'il y a entre les différentes conceptions du monde. Lorsque l'on parcourt la liste de la gauche vers la droite, l'on reconnaît de manière inclusive de plus en plus d'entités objectives constituant le monde. Ainsi le positivisme reconnaît un certain ordre à la matière, ordre qu'il est possible de connaître (au moins partiellement) à l'aide de la science. Ainsi le positivisme présente plus d'aspects "de droite" que le matérialisme et de manière encore plus évidente que le scepticisme bien que le positivisme soit classé à gauche.

### 2.3.2 L'analyse historique de Gödel

La classification n'est pas uniquement utile pour une analyse de différentes positions coexistant à un moment donné de l'histoire. L'analyse historique que propose Gödel concernant l'évolution de la philosophie et de la pensée mathématique fonde tout autant la pertinence de cette classification comme cadre de référence permettant d'observer la dynamique du développement philosophique, tout du moins si l'on s'accorde avec cette analyse. On peut ainsi lire au second paragraphe :

*" Now it is a familiar fact, even a platitude, that the development of philosophy since the Renaissance has by and large gone from right to left - not in a straight line, but with reverses, yet still, on the whole. "*

Constatons tout d'abord que s'il replace son étude du développement des fondements des mathématiques dans un cadre philosophique plus large, Gödel qui veut analyser ce développement au début du XX<sup>e</sup> siècle, relie cette période à un mouvement intellectuel plus général remontant à la renaissance. Ainsi bien qu'il reconnaisse la fécondité de ce déplacement vers la gauche de par ces nombreuses réussites (scientifiques et technologiques entre autres), c'est pourtant bien contre ce mouvement poussant à abandonner tous les aspects de droite et notamment la possibilité de la connaissance objective des entités reconnues comme constituant le monde que Gödel se place. Cette conscience de son opposition à l'esprit du temps est peut-être aussi une des raisons de sa prudence concernant la publication de certains de ces textes philosophiques. L'explication de ce mouvement de fond que nous fournit Gödel est la suivante : il y a eu dans les philosophies classiques ayant précédé ce mouvement vers la gauche un excès dans les aspects de droite. Gödel ne précise pas ce qu'il entend par là, mais le reste de l'article laisse penser qu'il vise par là un certain dogmatisme et mécanisme ayant finalement bénéficié à l'essor de l'empirisme qui permettait de s'extraire des vérités uniques et immuables proposées par les conceptions précédentes. Ainsi selon Gödel si la connaissance objective est possible, il est illusoire et contre-productif de penser que l'on peut saisir l'essence et la vérité des concepts fondamentaux d'une théorie ou de ces axiomes de manière immédiate, directe et exhaustive. La pensée mathématique et son évolution depuis la Renaissance dans la classification de Gödel ne s'accordent pas tout à fait avec cette analyse globale. En effet, ce dernier précise au paragraphe quatre que les mathématiques, au moins jusqu'à la crise des fondements du début du XX<sup>e</sup> siècle, ont résisté à ce mouvement vers la gauche :

*"Actually, mathematics, by its nature as an a priori science, always has, in and of itself, an inclination toward the right, and, for this reason, has longwithstood the spirit of time [Zeitgeist] that has ruled since the Renaissance ; i.e. the empiricist theory of mathematics, such as the one set forth by Mill, did not find much support. Indeed, mathematics has evolved into ever higher abstractions, away from matter and to ever greater clarity in its foundations (e.g. by [giving] an exact foundation of the infinitesimal calculus [and] the complex numbers) - thus away from skepticism."*

Mais, d'après Gödel, cette résistance des mathématiques à ce déplacement vers la gauche s'effrite au début du XX<sup>e</sup> siècle à la suite de la découverte des paradoxes de la théorie des ensembles, les paradoxes de Cantor et de Russel notamment. Bien que selon Gödel l'impact de ses paradoxes ait été exagéré, ils ouvrent une période de réflexion et de remise en question des fondements des mathématiques à l'occasion de laquelle Gödel décrit l'amorce d'un basculement vers la gauche des mathématiques. Selon cette conception des mathématiques de gauche, les objets et surtout les vérités

mathématiques sont vidés de tout contenu ontologique, la vérité d'un énoncé n'ayant plus de sens objectif, mais uniquement un sens relatif à des axiomes hypothétiques formulés de manière arbitraire et desquels on peut déduire cet énoncé. Selon ce point de vue les mathématiques sont une discipline purement formelle qui est au mieux empirique, au pire totalement spéculative. Aucun fondement rationnel ne garantit la certitude des vérités qu'elle démontre ni sa capacité à démontrer ou réfuter certains énoncés. Il est cependant difficile d'après le texte de déterminer précisément quelles positions sont visées, mais on peut conjecturer que le positivisme logique du maître de thèse de Gödel, H.Hahn, membre du cercle de Vienne, faisait partie du lot. Au paragraphe six du texte de 1961, le programme de Hilbert est décrit comme une réaction à ce déplacement vers la gauche dans lequel un certain nombre de mathématiciens ne se retrouvaient pas. Avant de discuter de cette réaction, il nous faut revenir sur les raisons qui font que, selon Gödel, ce classement est fécond.

### 2.3.3 Les cas mixtes

Tout d'abord, la classification de Gödel nous permet de replacer des doctrines philosophiques spécifiques à un certain domaine restreint de la philosophie, comme la morale la théorie de la connaissance ou l'esthétique, dans une classification de positions philosophiques bien plus larges. Ainsi concernant la théorie de la connaissance l'empirisme correspond à une posture de gauche et l'apriorisme à une posture droite. La seconde raison d'adopter cette classification selon Gödel est qu'elle permet d'analyser les cas mixtes, empruntant des éléments de gauche et de droite. Gödel donne ainsi l'exemple du pessimisme idéaliste de Schopenhauer, mélangeant un pessimisme de gauche à un idéalisme de droite.

Le programme de Hilbert, est à placer dans les cas mixtes. Il vise à donner une base solide aux mathématiques en les faisant reposer sur un jeu combinatoire fini de symboles, Hilbert est donc formaliste. En effet, les opérations finitaires lui semblent assez sûres pour pouvoir supporter l'édifice mathématique. Notons au passage qu'un tel système d'axiomes et de règles d'inférences serait un système formel Turing-complet, on désigne d'ailleurs par le terme calcul certains systèmes formels tels que le calcul propositionnel. Dans la classification de Gödel, le programme de Hilbert est clairement à classer parmi les cas mixtes, en effet, au paragraphe six il décrit en ces termes :

*"that curious hermaphroditic thing that Hilbert's formalism represents, which sought to do justice both to the spirit of time and to the nature of mathematics"*

L'aspect de gauche du programme, correspondant à l'esprit du temps, est le fait de

fonder les mathématiques sur la manipulation mécanique de symboles vidés de toute signification et donc de donner un point d'appui matérialiste aux mathématiques. Cette base est choisie arbitrairement, la vérité des axiomes qui la composent ne peut être discutée ni justifiée puisqu'ils ont été vidés de tout contenu sémantique. La vérité mathématique ne peut alors être établie que dans un sens hypothético-déductif. Cependant, Hilbert pense pouvoir démontrer à partir d'un jeu de symboles fini la cohérence et la complétude du système formel. Ainsi bien qu'elle n'ait de sens qu'à l'intérieur de son système, Hilbert pense redonner un certain sens à la vérité d'un énoncé : si un énoncé est démontré, son inverse ne peut l'être et si un énoncé ne peut être démontré alors sa négation peut l'être. La vérité dans cette acception est alors assimilée à une structure cohérente et totalement déterminée que l'on peut rapprocher de la notion postérieure de vérité comme existence d'un modèle unique à isomorphisme près. Démontrer la cohérence et la complétude du formalisme hilbertien assurerait alors que lorsqu'un énoncé est démontré, il est alors vrai, dans le sens relatif que venons de donner de la vérité. Cet optimisme dans la capacité des mathématiques à établir des énoncés vrais et à refuser *l'ignorabimus*<sup>42</sup> est clairement l'aspect de droite, lié selon Gödel à la nature des mathématiques en tant que science *a priori*, du formalisme de Hilbert. En 1931 avec la publication de ses théorèmes d'incomplétude, Gödel met fin au projet Hilbertien en montrant, entre autres, que l'on ne peut démontrer la cohérence et la complétude d'un système assez puissant pour exprimer l'arithmétique par des moyens finitaires. Gödel souhaite cependant maintenir certains aspects de droite, notamment un optimisme pour ce qui est de la connaissance en mathématiques, car il pense que leur abandon n'est qu'un préjugé de son temps. Les excès des philosophies de droite ne justifiant en rien que l'on abandonne tout les aspects de droite. Voyons maintenant quel est le positionnement de Gödel concernant le fondement des mathématiques.

## 2.4 Le projet Gödelien

Nous pourrions ainsi résumer la voie que Gödel fait sienne pour dépasser les antinomies de cette classification : s'inscrire dans une pensée combinant des aspects de gauche et de droite en refaisant siennes certaines ambitions ou visées des aspects de droite abandonnées par les philosophies de gauche, sans tomber dans les excès

---

42. En 1872 le physiologiste Emil du Bois-Reymond conclue son essai *"Les limites de la connaissance humaine"* par l'aphorisme *"Ignoramus, ignorabimus"*, affirmant ainsi la nature intrinsèquement infranchissable de ces limites de la connaissance. Hilbert reprend le terme *ignorabimus* pour réfuter ce pessimisme pour ce qui est des mathématiques lors d'une intervention à la radio. Il affirme ainsi son refus de limiter la connaissance en mathématiques et lui oppose un autre aphorisme : *"Wir müssen wissen, wir werden wissen"* (Nous devons savoir, nous saurons).

dont elle a pu faire preuve. Nous pouvons ainsi lire au paragraphe dix : *"the correct attitude appears to me to be that the truth lies in the middle or consists of a combination of the two conceptions."* L'attitude correcte est donc selon Gödel de se placer dans les cas mixtes, c'était le cas de Hilbert mais ce dernier avait tenté un mélange trop primitif et penchant trop fortement vers la gauche. Voyons donc ce que propose Gödel au paragraphe onze.

*"this means that the certainty of mathematics is to be secured not by proving certain properties by a projection onto material systems - namely, the manipulation of physical symbols but rather by cultivating (deepening) knowledge of the abstract concepts themselves which lead to the setting up of these mechanical systems, and further by seeking, according to the same procedures, to gain insights into the solvability, and the actual methods for the solution, of all meaningful mathematical problems."*

Ainsi il faut s'éloigner du formalisme, l'attitude qui consiste à donner des définitions ou des démonstrations ne peut mener qu'à une régression à l'infini ou bien aboutir à des objets indéfinissables ou indémontrables. En effet, le point de vue qui consiste à voir dans les axiomes des vérités ne nécessitant pas de démonstration, auquel on peut rattacher Euclide et Hilbert, est peut-être un des excès de droite desquels parle Gödel. Il faut arriver à une compréhension rationnelle, mais pas formelle des concepts abstraits sur lesquels se fonde notre idée de calcul et notre intuition de certains objets abstraits. Mais comment étendre notre compréhension de ces concepts si ce n'est par la définition ?

Gödel voit dans la phénoménologie de Husserl, qu'il a alors déjà étudiée, une méthode, une technique permettant de clarifier le sens des concepts abstraits en se concentrant sur nos actes dans l'usage de ces concepts. La saisie et la compréhension du contenu des concepts ne sont ainsi pas données de manière directe, mais par une étude rationnelle et difficile tournée vers les états de conscience de l'individu. Si nous ne rentrerons pas dans l'analyse du rapport de Gödel à Husserl, nous allons désormais essayer de mieux comprendre les raisons de l'intérêt que porte Gödel à la phénoménologie et en quoi cette dernière est une tentative de cas mixtes moins primitive que celle de Hilbert.

## 2.5 L'optimisme épistémologique de Gödel

Nous allons ici présenter un des aspects de la philosophie de Gödel qui explique son refus d'abandonner tous les aspects de droite de la classification et qui expliquent qu'il ait préféré une position mixte à une position uniquement de gauche. Nous avons déjà évoqué comme motivation, la tentative de garder quelque chose de la certitude

et la vérité mathématique. La position platoniste de Gödel, sur laquelle nous nous arrêterons dans la prochaine partie, explique ce refus de vider les mathématiques de tout contenu sémantique, il écrit ainsi au paragraphe 9.

*" on the one hand, to safeguard for mathematics the certainty of its knowledge..."*

Nous allons ici nous pencher sur le second élément des aspects de droite que veut sauver Gödel et qui complète la citation.

*"...and on the other, to uphold the belief that for clear questions posed by reason, reason can also find clear answers."*

Gödel fait preuve ici d'une attitude fortement rationaliste à travers cet optimisme épistémologique. Le rationalisme était en effet une conviction profonde pour Gödel, qui fut présente tout au long de sa carrière et de son évolution philosophique<sup>43</sup>. Ainsi dans un manuscrit non publié par Gödel et intitulé *My philosophical viewpoint*<sup>44</sup>, on peut lire une tentative de liste des convictions philosophiques fondamentales pour Gödel. Puisque nous allons tenter d'entrer ici un peu plus profondément dans la philosophie de Gödel, cette liste nous sera d'une grande utilité d'autant plus qu'elle a été écrite aux alentours des années 60. Citons pour commencer le point trois de cette liste :

*"3. There are systematic methods for the solution of all problems (also art, etc)"*

Cet optimisme épistémologique peut paraître, dans une première lecture, entrer en contradiction avec le premier théorème d'incomplétude. En effet, celui-ci affirme qu'il y a et aura toujours dans tout système formel permettant de formaliser l'arithmétique des énoncés indémontrables. Cependant, la conclusion qu'en tire Gödel n'est en aucun cas l'incapacité pour l'Homme de résoudre certains problèmes mathématiques, mais plutôt l'impossibilité de le faire en ayant recours au calcul comme il l'écrit au paragraphe 17 du texte de 1961 :

*"It is not at all excluded by the negative results mentioned earlier that nevertheless every clearly posed mathematical yes-or-not question is solvable in this way."*

---

43. Van Atten, M. et Kennedy, J. (2003) section 6.2.2 *" we note that is clear that at every stage of his thought, rationalism was a core belief of Gödel's, it emerges in almost all of his philosophical conversations, and extensively in writings both published and unpublished."*

44. *My philosophical viewpoint* a été publié dans Wang, H. (1996) à la p316

Les *systematic methods* incluent le calcul, mais ne s'y réduisent donc pas<sup>45</sup>, comme nous avons vu que les procédures déterministes incluait les calculs sans, non plus, s'y réduire. Nous retrouvons cela dans le point 9 de *My philosophical viewpoint* :

"9. *What is formally correct is part of a science of reality.*"

Gödel donne comme exemple de processus rationnel non-calculable l'établissement de nouveaux axiomes<sup>46</sup> indépendants du système axiomatique auxquels ils sont ajoutés et dont la nécessité devient pourtant de plus en plus évidente<sup>47</sup>. Gödel défend alors un rationalisme qui dépasse les limites que lui imposerait une position positiviste qui refuserait toute connaissance rationnelle *a priori*, comme le stipule le point six de sa liste.

"6. *There is incomparably more knowable a priori than is currently known*"

De cette manière, une approche scientifique des questions métaphysiques est, selon Gödel, possible. En témoigne la parenthèse à la fin du point trois de *My philosophical viewpoint* ainsi que le point 13.

"13. *There is a scientific (exact) philosophy and theology (this is also more highly fruitful for science) which deals with concepts of the highest abstractness;*"

La science selon Gödel n'est donc pas censurée par la réfutabilité. Gödel semble ainsi en 1961 avoir trouvé dans la phénoménologie de Husserl un rationalisme bien plus puissant que le formalisme.

### 3 Les trois mondes de Penrose à la lumière de Gödel

Nous avons désormais un certain nombre d'éléments concernant EOLP, OE et le texte de Gödel de 1961, ce qui va nous permettre de commencer à soulever certaines questions communes aux ouvrages des deux auteurs. L'objectif est de tenter de replacer les ouvrages de Penrose à l'intérieur d'un débat philosophique plus large touchant aux notions d'esprit, de connaissance et de réalité. Penrose fait peu de références à la littérature philosophique, il s'exclut ainsi parfois lui-même du cœur du débat par

---

45. Gödel étant un grand lecteur de Leibniz il me semble pertinent de voir dans cette démarche de recherche d'une méthode de démonstration rigoureuse mais non-formelle une actualisation du projet leibnizien de *Characteristica universalis*. L'article de V. Atten et J. Kennedy va également dans ce sens.

46. Nous pouvons conjecturer que c'est l'axiome du choix et/ou l'hypothèse du continu auxquels pense Gödel ici.

47. Gödel, K. 1961 paragraphe 15 "*For it is just this becoming evident of more and more new axioms on the basis of the meaning of the primitive notions that a machine cannot imitate*"



une certaine autocensure du scientifique qui fait une incursion dans un domaine dans lequel il sait qu'il ne peut faire figure d'expert<sup>48</sup>. Notre présentation des ouvrages de Penrose a jusque là été relativement large, afin de tenter de mieux caractériser la différence dans l'engagement philosophique de Gödel et Penrose, il me semble pertinent d'analyser plus dans le détail deux passages d'OE. Nous nous pencherons tout d'abord sur le début d'OE avant de faire un saut directement à la dernière section de l'ouvrage.

### 3.1 L'A, B, C, D du calcul et de la pensée consciente :

#### 3.1.1 Présentation

Il est intéressant de constater que, comme au début de l'article de 1961 de Gödel, Penrose dans les premières pages d'OE, fait lui aussi un panorama des différents points de vue constituant la palette des positions possibles, concernant le lien entre calcul et pensée consciente. Il reconnaît que cette classification n'est nullement définitive et qu'il existe de nombreuses positions intermédiaires réparties autour des quatre pôles que définit Penrose. Comme pour la classification de Gödel, celle de Penrose est pensée relativement à son utilité pour caractériser la position de l'auteur et ses relations, affinités ou différences, avec d'autres positions. Je vais donc commencer par retranscrire l'ABCD tel que présenté dans OE. Je résumerai ensuite les commentaires que Penrose l'y ajoute pour enfin faire quelques remarques personnelles à ce propos.

- *"A. Toute pensée se réduit à un calcul ; en particulier, le sentiment de connaissance immédiate consciente naît simplement de l'exécution de calculs appropriés.*
- *B. La connaissance immédiate est un produit de l'activité physique du cerveau ; mais bien que toute action physique puisse être simulée par un calcul, une telle simulation ne peut par elle-même susciter la*

---

48. Penrose écrit par exemple concernant la question du "moi" indivisible et du libre arbitre dans OE p32 *"ce livre n'apportera pas de réponse à ces problèmes, mais je suis persuadé qu'il peut permettre de les aborder - ne serait-ce que très modestement. [...] Je tenterais de montrer que lorsqu'une "cause " est le produit de nos actes conscients, elle est alors nécessairement quelque chose de très subtil, certainement irréductible non seulement au calcul et au chaos, mais aussi à toute influence purement aléatoire. Et le futur dira si un tel concept de cause peut nous rapprocher d'une compréhension du profond problème (ou de "l'illusion" ?) du libre arbitre."*

De manière plus large il écrit à la p399 *"Je le reconnais volontiers, je n'ai pas eu grand chose à dire sur le problème général de la conscience"*

*connaissance immédiate.*

- C. *La connaissance immédiate est suscitée par une action physique du cerveau, mais aucun calcul ne peut simuler, même à la perfection, cette action physique.*
- D. *On ne peut expliquer la connaissance immédiate à l'aide du langage de la physique, de l'informatique, ni de quelque autre discipline scientifique que ce soit."*

### 3.1.2 Quelques commentaires de Penrose

Tout d'abord, il faut souligner que de l'expression *connaissance immédiate* Penrose ne donne volontairement aucune définition<sup>49</sup> nous renvoyant à l'idée intuitive de contenu interne, spontané et passif de la conscience nécessaire à la compréhension, compréhension qui elle-même est nécessaire à toute intelligence.

Le point de vue D est celui du mystique, l'acception d'une doctrine religieuse semble mettre en jeu au moins quelques éléments de D. Ce point de vue est dès le début écarté à la p9 d'OE.

*Si je rejette le mysticisme pour sa négation des critères scientifiques qui permettent le progrès de la connaissance, je suis persuadé qu'en élargissant son champ, la science éclaircira quantité de mystères, dont celui de l'esprit.*

Le rejet de la position D ne tient donc que si toute connaissance et tout questionnement peuvent être éclaircis par la science. Penrose fait ce pari et propose ainsi dans le cas des questionnements concernant la place et la nature de l'esprit de montrer que cette question peut être résolue scientifiquement. S'il n'attaque pas directement cette position, Penrose souhaite donc persuader le lecteur que l'hypothèse divine n'est en rien nécessaire à l'éclaircissement du problème de la conscience. Le fait de renoncer à cette hypothèse est aussi une question d'optimisme épistémologique et de croyance en la possibilité du progrès de la connaissance par la science. Penrose invite le lecteur convaincu de la véracité du point de vue D à le suivre sur le chemin de la science pour voir jusqu'où il peut le mener sans postuler d'entité extra naturelle ou extra matérielle.

---

49. Penrose, R. OE p34 " Bien sûr, je n'ai défini aucun des termes "intelligence", "compréhension" ou "connaissance immédiate". Je pense qu'il serait peu avisé d'en donner ici des définitions précises. Dans une certaine mesure, il nous suffira de nous appuyer sur le sens intuitif que nous accordons à ces mots." Cette citation éclaire le manque de rigueur dans le travail philosophique dont peut faire preuve à certains moments Penrose.

Le point de vue A est celui de l'IA forte, il est aussi appelé computationnalisme ou encore fonctionnalisme. Pour certains, il est le seul digne d'une véritable attitude scientifique, pour d'autres, il est d'une absurdité totale. Dans la version extrême de ce point de vue, l'univers lui-même est un gigantesque ordinateur. Penrose bien que n'y adhérant pas personnellement, tient A comme *"une possibilité sérieuse méritant une attention considérable"*<sup>50</sup>. Combattre ce point de vue est l'un des enjeux, voire l'enjeu principal des deux ouvrages. Nous pouvons dire que la partie 1 d'OE est presque entièrement dédiée à cet objectif. Ce point de vue a son origine en partie dans la place que tiennent les simulations numériques dans la science du XX<sup>e</sup> siècle et en partie dans le point de vue de certains physiciens qui traitent la matière comme un simple schéma d'information soumis à des lois mathématiques. En effet, on sait que la masse même d'un objet peut se transformer en énergie, devenant par là une pure abstraction mathématique. La physique quantique quant à elle semble indiquer que les particules se comportent comme des ondes d'information. Le point de vue A ne fait pas de différence entre un objet et une simulation de cet objet s'il n'existe pas de moyens externes à cet objet pour mettre en évidence cette différence.

Le point de vue B est celui de l'IA faible. Comme selon le point de vue A, selon B le comportement des objets physiques peut être compris scientifiquement et simulé numériquement. Cependant, ce qui le distingue de B est qu'il est moins empiriste, puisqu'il accorde une réalité à des faits qui, par définition, sont infalsifiables. Ainsi s'il admet qu'un ordinateur pourrait passer un jour le test de Turing, cela ne signifierait alors pas, du point de vue B, qu'il possède pour autant une conscience, comme l'affirme le point de vue A. Pourtant, rien ne permettrait alors empiriquement de distinguer un être conscient d'un tel ordinateur. B admet que l'on ne peut réduire tout objet à ses manifestations externes et donc vérifiables expérimentalement. De plus selon B la présence ou l'absence de conscience dépend de la constitution matérielle de l'objet produisant la pensée, pas uniquement des opérations ou des calculs qu'il effectue.

La position C est celle défendue par Penrose. Elle est plus opérationnaliste que B car si elle affirme une différence entre un objet conscient et sa simulation, elle affirme que cette différence se traduira par une différence dans leur comportement externe. Elle se rapproche cependant du physicalisme de B en donnant une certaine importance au support matériel de l'information. En effet, elle postule l'existence d'une opération physique qui ne pourrait se traduire en termes de calcul. L'origine de cette opération non-calculable scinde ce point de vue en une version faible, affirmant que cette opération non-calculable est déjà présente dans la physique actuelle, et une version forte. Selon la version forte de C à laquelle adhère Penrose, cette opération

---

50. OE p10

non-calculable est à chercher dans un domaine échappant aux lois de la physique telle que nous les connaissions au moment de la parution de l'ouvrage. Pour autant, la version forte de C ne rejoint pas le point de vue D car s'il accepte que l'esprit soit hors de portée de la science actuelle, il affirme que c'est pourtant bien un problème scientifique qui pourra être résolu par la science dans un avenir plus ou moins lointain. Penrose participe sans pour autant l'achever, à ce projet de physique de l'esprit consistant à faire avancer la science pour l'armer face à des questionnements qui n'admettent, pour l'instant, pas de réponse scientifique satisfaisante.

Il existe également une position mélangeant A et D selon laquelle l'activité consciente est identique à celle d'un ordinateur, mais d'un ordinateur dont la complexité serait telle que son imitation ou sa compréhension dépasserait largement l'entendement humain et la capacité de tous les ordinateurs que nous pourrions effectivement construire. Un tel algorithme serait la création d'un dieu et nous serait définitivement inaccessible.

### 3.1.3 Quelques commentaires personnels

On peut tout d'abord s'interroger sur le fait que Penrose s'interdise toute spéculation d'ordre mystique et s'autorise cependant à spéculer dans d'autres cas. Nous avons déjà parlé de la spéculation scientifique, mais Penrose postule de plus l'existence d'un monde platonicien sans pouvoir expliciter ou mettre clairement en évidence le lien qui nous y unit et surtout qui l'unit au monde physique, comme nous allons le voir dans la section suivante. Mais alors, pourquoi postuler un monde platonicien plutôt qu'un monde divin ? Sur quoi repose le choix de Penrose ? D'après Penrose le monde de Platon est une réalité indispensable à la connaissance scientifique, la question de son existence dépasse donc nécessairement le cadre de la science. Pour justifier ce que nous appellerons son optimisme scientifique<sup>51</sup> Penrose fait donc appel à l'histoire à la p45 d'OE :

*"c'est seulement grâce à la science que l'on a pu accomplir des progrès réels dans la compréhension du monde"*

---

51. Nous utiliserons cette expression "d'optimisme scientifique" afin d'en souligner la proximité avec "l'optimisme épistémologique" de Gödel tout en conservant une distinction entre deux points de vue qu'il serait erroné d'identifier puisque Penrose et Gödel n'ont pas la même définition de science. En effet, nous avons vu que pour Gödel la science peut par exemple prétendre à la résolution de grandes questions métaphysiques. On retrouve de plus l'expression sous la plume de Penrose à la p10 d'EOLP : *"J'ai tendance à penser qu'en règle générale une imitation, aussi habile soit-elle, doit toujours pouvoir être détectée par des moyens suffisamment perfectionnés - il s'agit cependant là plus d'un acte de foi (ou disons d'optimisme scientifique) que d'un fait démontré."*

Ainsi, Penrose fait presque une identification entre connaissance et connaissance scientifique, la spéculation du mystique étant alors un frein à la connaissance, la spéculation scientifique de Penrose un moteur de cette dernière. Nous voyons là une profonde différence entre l'analyse historique que font Gödel et Penrose des succès et de l'évolution de la science ces derniers siècles. L'analyse historique de Penrose est bien plus courte et uniquement descriptive, de plus elle ne semble pas s'attarder sur la nature de la science qui a changé considérablement depuis la renaissance. Par exemple, il me semble que l'alchimie ne serait pas considérée par Penrose comme une science, bien qu'elle ait contribué à notre compréhension des réactions chimiques. Les mathématiques de par leur apriorisme pourraient être écartées d'une définition empiriste de la science qui se concentrerait sur la méthode scientifique de récolte et d'interprétation de données. Les mathématiques sont pourtant une discipline scientifique pour Penrose, ce qui n'est pas le cas de toute activité de compréhension abstraite de concepts. De plus Penrose insiste sur le fait qu'à de nombreuses reprises, dans l'histoire de la physique, des théories basées sur les mathématiques se sont révélées d'une précision et d'une généralité que l'on n'aurait pu anticiper si l'on se réfère uniquement aux données empiriques qu'elles cherchaient à expliquer. Ainsi la relativité d'Einstein ne répond pas au besoin d'expliquer un fait expérimental. Si Penrose ne réduit certainement pas la science à une activité purement empirique, il est difficile, à partir de OE et EOLP d'établir son rapport aux savoirs *a priori*. Ainsi s'il reconnaît que la physique ne saurait être ce qu'elle est aujourd'hui sans l'aide de théories constituées en grande partie *a priori* par des physiciens et des mathématiciens, nous ne savons pas s'il franchit le pas qui consisterait à affirmer la possibilité d'une connaissance entièrement *a priori*, en dehors des mathématiques. Il est clair que pour Penrose un ensemble de résultats expérimentaux ne peut constituer une science s'ils ne sont organisés par les mathématiques, est-il selon lui pour autant possible d'atteindre un savoir *a priori* s'il n'est falsifiable expérimentalement ? Je n'ai pas trouvé de réponse claire à cette question tout au long des deux ouvrages, cependant Penrose utilise souvent le mot science pour désigner le corpus de savoirs scientifiques tel qu'il est actuellement. Nous ne pouvons donc faire mieux que de faire l'hypothèse que puisque le critère empirique est un élément nécessaire à l'établissement d'une affirmation par la méthode scientifique tel qu'elle est au XX<sup>e</sup> siècle, lorsque Penrose parle de savoir, de méthode, d'analyse<sup>52</sup> ou encore de critères scientifiques cela implique de que ce savoir soit falsifiable et donc soumis à la vérification via un système matériel. Pour

---

52. OE p339 À propos de l'hypothèse d'une influence d'un esprit séparé de la matière, sur le déclenchement la réduction du vecteur d'état, Penrose écrit : "*Il est toutefois difficile d'opposer des arguments rationnels à un tel point de vue car, de par sa nature même, sa formulation n'obéit pas aux règles précises qui permettraient de le soumettre à une analyse scientifique.*"

appuyer cette hypothèse, notons le fait que bien que concernant la science il souhaite *"élargir son champ d'action [...] au besoin en modifiant ses propres procédures"*, il ne défend pas pour autant une remise en question du rôle de l'expérience en tant qu'ultime critère de vérité.

L'amplitude du panorama de Penrose nous renseigne également sur la portée de la réflexion qu'il va entreprendre et il est intéressant de comparer l'ABCD de Penrose à la classification des *Weltanschauungen* de Gödel. Nous constatons que le panorama de Penrose est beaucoup plus réduit et beaucoup moins général que celui de Gödel, se concentrant sur des positions précises concernant la question de la conscience et du calcul. La classification de Gödel, elle, se concentre sur des caractéristiques plus fondamentales de différentes conceptions du monde à l'intérieur desquelles il replace les diverses réponses apportées à la question des fondements des mathématiques. La classification de Gödel permet bien plus facilement de replacer les positions mixtes et d'avoir un recul sur l'évolution des positions dans le temps. L'ABCD de Penrose le contraint à n'envisager la question de la conscience que sous l'angle du débat autour de l'IA tel qu'il est structuré au moment de l'écriture de OE.

## 3.2 Trois mondes et trois mystères

### 3.2.1 La place des trois mondes dans OE

Si dans le texte de 1961 Gödel commence par traiter de conceptions générales du monde pour aboutir ensuite à une réflexion plus particulière à propos des fondements mathématiques, c'est en un certain sens l'inverse dans OE. Nous venons de voir comment Penrose amorce cet ouvrage en définissant différentes positions possibles autour de la question isolée et clairement délimitée<sup>53</sup> du rapport entre calcul et pensée. C'est ce problème particulier qui l'occupera dans la plus grande partie d'OE. La réponse qu'il y apporte est empreinte de présupposés philosophiques platoniciens et réalistes<sup>54 55</sup>. Ainsi l'argumentation est construite pour fonctionner à condition

---

53. OE p399 *"je le reconnais volontiers, je n'ai pas eu grand chose à dire sur le problème général de la conscience"*

54. Nous en avons un exemple lorsque Penrose à la p301 discute de la réalité du vecteur d'état en physique quantique il écrit : *"Pour certaines personnes dont moi-même (mais aussi Einstein et Schrödinger - je suis donc en bonne compagnie), cela n'a aucun sens de réserver le mot "réalité" aux seuls objets que nous pouvons percevoir, tels les dispositifs de mesure (ou plutôt certains types de dispositifs de mesure), et de refuser de l'appliquer à un niveau plus profond. Certes le monde quantique a un comportement étrange, insolite, mais il n'est pas "irréel". Comment d'ailleurs pourrait-on construire des objets réels à partir de constituants irréels?"*

55. On trouve également dans la note 1 du chap 6 d'EOLP p 502 l'affirmation du parti-pris réaliste de Penrose : *"Je pose en hypothèse qu'un point de vue philosophique ne peut être qualifié de*

que l'on accepte le point de vue qui est celui de Penrose. Ce point de vue n'est alors justifié que par sa capacité à expliquer le monde et en particulier l'esprit <sup>56</sup>. Après avoir passé 400 pages à défendre la fécondité de son point de vue et la perspective de progrès pour la connaissance qu'il représente, Penrose dans la section 8.7 (pp399-408), la dernière de l'ouvrage, nous livre sa conception du monde. Pourquoi attendre si longtemps alors que, par exemple, l'existence d'une vérité mathématique et physique objective est défendue et employée plus tôt dans l'argumentation ? Penrose ne répond pas clairement à cette question, nous pouvons cependant émettre deux hypothèses. La première vient d'être évoquée et est présente dès le début de l'ouvrage. Puisque Penrose n'est pas philosophe, il ne peut ou ne veut pas donner une justification purement métaphysique de son système. Il préfère donc ne livrer l'entièreté de son point de vue qu'après qu'il ait montré son potentiel explicatif. Rejeter son point de vue sans arguments valables ou pour des raisons purement métaphysiques apparaîtrait alors comme une attitude purement négative, s'opposant à l'optimisme scientifique défendu tout au long d'EOLP et d'OE. Dans le texte de 1961, Gödel justifie son choix de conserver certains aspects de droite dans sa conception des mathématiques et contre l'esprit du temps pour une raison proche de celle-ci. Abandonner les aspects de droite serait tout simplement une attitude négative puisque, entre autres, cela placerait une partie des questions mathématiques hors de notre portée <sup>57</sup>, de la même manière que selon Penrose, un dualisme corps/esprit ou encore matière/esprit placerait au moins une partie du phénomène de conscience hors de portée de la science telle que nous supposons qu'il l'entend, de par l'impossibilité d'interagir expérimentalement avec une telle entité immatérielle.

Notre seconde hypothèse est que Penrose a, lui aussi, conscience que le platonisme

---

*"sérieux" que s'il contient une certaine dose de réalisme. [...] Si j'adopte, chaque fois que cela m'est possible, une ligne de pensée réaliste, ce n'est nullement par manque d'information de ma part : je suis parfaitement au courant des positions subjectivistes adoptées par certains - simplement, ces conceptions n'ont aucun sens pour moi."*

56. OE p9 *"Si vous êtes fermement convaincu de la vérité de D, sous une forme ou une autre, je vous demande un peu de patience. Voyez jusqu'où on peut aller en suivant la voie scientifique - et essayez de percevoir à quel endroit, selon moi, cette voie devrait finalement nous mener"*

57. Gödel, K. (1961) paragraphe 9 *"Hence only two possibilities remain open. One must either give up the old rightward aspects of mathematics or attempt to uphold the in contradiction to the spirit of the time. Obviously the first course is the only that suits our time and [is] therefore also the one usually adopted One should however keep in mind that this is a purely negative attitude. One simply gives up aspects whose fulfillment would in any case be very desirable and which have much to recommend themselves : namely on the one hand, to safeguard for mathematics the certainty of its knowledge, and on the other, to uphold the belief that for clear questions posed by reason, reason can also find clear answers."*

ne fait pas partie d'un certain esprit du temps<sup>58 59</sup>. Peut-être en reléguant à la fin du livre l'approfondissement de son point de vue platonicien, il évite que le lecteur peu enclin à mêler les entités immatérielles du monde platonicien au savoir scientifique, ne le referme avant d'avoir pu juger de son intérêt<sup>60</sup>. Nous n'irons pas jusqu'à affirmer que la *doxa* qui règne dans le milieu des physiciens, des informaticiens et des biologistes et en référence à laquelle se place Penrose est précisément l'esprit du temps décrit par Gödel. Cependant on peut retrouver chez Penrose l'évocation de la nature intrinsèquement de droite des mathématiques de par leur nature de science *a priori* et par là leur résistance à l'esprit du temps<sup>61</sup>.

### 3.2.2 Présentation des trois mondes

Penrose décrit donc sa conception du monde comme reposant sur l'existence de trois mondes à la fois distincts et connectés entre eux. Ces trois mondes sont représentés dans la figure 3<sup>62</sup>, nous allons décrire ici leurs contenus et leurs liens selon le point de vue de Penrose.

#### — 1 Le monde mental ou monde de la perception consciente.

D'après Penrose c'est le monde que nous connaissons le plus directement, mais sur lequel nous avons le moins de connaissances scientifiques précises. En effet, il contient les sentiments, les perceptions de nos sens, les souvenirs, les connaissances ainsi que les images mentales que l'on se fait des autres mondes. En résumé tout ce qui est présent dans notre esprit de manière consciente ou

---

58. OE p 400 *"Il y a enfin un troisième monde, bien que nombre de gens aient du mal à accepter qu'il existe réellement. C'est le monde platonicien des formes mathématiques."*

59. OE p 401 *"J'ai parfois entendu dire que le platonisme mathématique était "démodé". Certes Platon est mort il y a 2340 ans, mais ce n'est pas une raison ! Une objection plus sérieuse est celle des philosophes qui trouvent difficile d'admettre qu'un monde entièrement abstrait puisse avoir une influence sur le monde physique."*

Nous ne savons pas à qui fait référence Penrose lorsqu'il parle des *"philosophes"*. De nombreux philosophes ont défendu l'influence de mondes abstraits ou idéaux sur le monde physique au cours de l'histoire. Nous pouvons donc conjecturer que c'est un certain esprit du temps se rapprochant du matérialisme et donc des aspects de gauche, que Penrose personnifie ici avec l'expression *"les philosophes"*.

60. EOLP p4 *Je dois dire de façon claire et nette que mon point de vue n'a rien de canonique au sein de la communauté des physiciens et qu'en conséquence il a très peu de chances d'être adopté par les informaticiens et les physiologistes."*

61. OE p401 *"En vérité, la réalité des concepts mathématiques apparaît bien plus naturelle aux mathématiciens qu'à ceux qui n'ont pas la chance de passer leur temps à explorer les merveilles et les mystères de ce monde."*

62. L'illustration est extraite d'OE p402, seule la numérotation a été modifiée pour des raisons pratiques



pas, tout ce qui constitue la psychologie individuelle.

— **2 Le monde physique**

Il contient les objets physiques. Cela comprend les hommes, les atomes, les étoiles, l'activité neuronale, etc. Nous avons désigné jusqu'ici ce monde par l'adjectif matériel, le fait que Penrose privilégie le terme physique nous conforte dans la supposition que nous avons fait dans la sous-section 4.1.3 et selon laquelle la science, que l'on peut en dernier ressort réduire à la physique, est entièrement tournée vers l'étude de ce monde matériel.

— **3 Le monde platonicien des formes mathématiques**

Ce monde contient les objets mathématiques (y compris non-calculables comme les ordinaux indénombrables), les théorèmes (y compris indémontrables), les actions de machines de Turing qui ne s'arrêtent pas, les lois et équations physiques et les modèles théoriques qui les vérifient, les simulations d'objets du monde physique... Ce monde semble occuper une place toute particulière pour Penrose bien qu'il soit difficile de définir cette particularité. En effet, on peut lire *"Selon moi le monde des formes parfaites est premier (comme le pensait Platon) son existence étant presque une nécessité logique dont les deux autres mondes ne sont que des ombres"*<sup>63</sup>. Penrose prétend pourtant, quelques paragraphes après avoir affirmé cela, qu'il *"ne préjuge en rien d'une hiérarchie entre les mondes"*<sup>64</sup>. Malgré cette dernière citation, il me semble que le monde platonicien pour Penrose accède à une position dominante par rapport au monde physique imparfait qui imite d'une certaine manière les formes mathématiques parfaites comme en témoigne l'expression *"les entités physiques mathématiquement déterminées qui composent notre univers"*<sup>65</sup>.

### 3.2.3 Le contenu du monde platonicien.

Avant de tenter de comprendre les liens qui unissent ces mondes et par le même coup les mystères qu'ils engendrent, nous allons faire quelques commentaires sur le contenu du monde platonicien de Penrose. En effet, si Penrose réserve aux objets mathématiques l'existence platonicienne, il envisage la possibilité d'y accéder pour une classe plus large de concepts. Cependant, il restreint une nouvelle fois sa réflexion en se gardant de se prononcer sur ce dernier point, arguant qu'il n'a pas de lien direct avec le problème de la calculabilité et de la conscience<sup>66</sup>.

---

63. OE p 405

64. OE p 406

65. OE p192

66. OE p 404 *"Peut-être doit-on attribuer une réalité platonicienne à d'autres concepts abstraits, qui ne sont pas forcément de nature mathématique. Platon lui-même attribuait aux concepts de*

L'existence platonicienne des concepts est un point d'autant plus important qu'il est présent dans la philosophie de Gödel. Bien qu'il n'explicite pas son point de vue platoniste dans le texte de 1961, certains passages sous-entendent l'existence idéale et objective<sup>67</sup> de certains concepts premiers. Nous n'avons pas la prétention d'établir ici clairement la pensée philosophique de Gödel, nous pouvons cependant affirmer que les concepts en général avaient bien une existence objective pour lui en se basant sur le point 12 de *My philosophical viewpoint* :

"12. *Concepts have an objective existence.*"

Ces concepts pouvant être des concepts mathématiques<sup>68</sup> ou plus larges, comme les concepts servant de base de contenu sémantique à l'action désignative du langage<sup>69</sup> par exemple. Concernant ce dernier exemple, Penrose évoque la possibilité p48-49 d'OE que la "*connaissance immédiate*", qui est non calculable et qui est le support de notre lien avec le monde platonicien<sup>70</sup>, soit indispensable à la saisie du sens d'un mot. Cet accès au sens des mots serait par là même hors de portée de toute IA et serait donc une activité non formelle.

Il est également intéressant de se demander pourquoi affirmer l'existence platonicienne de certains concepts et sur quoi se fonde cette existence. Penrose à la p401 d'OE se justifie ainsi :

"De quel droit disons-nous que ce monde platonicien est réellement un "monde" qui "existe" tout autant que les deux autres ? [...] son existence

---

"bien" ou de "beauté" la même réalité que l'on reconnaît aux concepts mathématiques. Personnellement, je ne m'oppose pas à une telle extension mais elle n'a pas jusqu'ici joué un rôle important dans les considérations que j'expose dans ce livre. Les problèmes d'éthique, de morale ou d'esthétique n'interviennent en rien dans mes discussions - mais ce n'est pas une raison pour ne pas les considérer comme aussi "réelles" que ceux que j'ai examinés. À l'évidence, ce sont là des problèmes dignes d'intérêt, mais je n'ai pas eu à les aborder dans ce livre."

67. Gödel, K. (1961) paragraphe 13 "*basic concepts hitherto unknown to us.*" Ainsi les concepts ont une existence que l'on ne peut réduire à la connaissance que nous en avons.

68. Gödel, K. (1961) paragraphe 11 "*Obviously this means that the certainty of mathematics is to be secured [...] by cultivation (deeping) knowledge of the abstract concepts themselves which lead to setting up of these mechanical systems*"

69. Gödel, K. (1961) paragraphe 14 "*If one considers the development of a child, one notice that it proceeds in two directions : it consists on the one hand in experimenting with the objects of the external world and with its [own] sensory and motor organs, on the other hand in coming to a better and better understanding of the language, that means [...] of basic concepts on wich it rests.*" Nous pouvons en déduire que les concepts débordent et ne peuvent se réduire à leur désignation par le langage

70. OE p46 "*C'est ce potentiel de "connaissance immédiate" des concepts mathématiques, cet accès direct au monde platonicien, qui confère à l'esprit un pouvoir supérieur à celui de tout dispositif dont l'action repose uniquement sur le calcul.*"

*repose sur la nature profonde, intemporelle, et universelle de ces concepts et sur le fait que leurs lois sont indépendantes de ceux qui les ont découverts."*

Nous avons une justification un peu plus approfondie dans EOLP au chapitre 3 intitulé *Mathématiques et réalité*. Dans ce chapitre Penrose nous invite à une promenade à travers le pays de Tor'Bled Nam nous faisant découvrir tour à tour ces vallons et ces créatures fantastiques. Il explicite ensuite la définition mathématique de cette région du plan complexe qui n'est autre que l'ensemble de Mandelbrot<sup>71</sup>. Il affirme ainsi p102 d'EOLP :

*"L'ordinateur est utilisé essentiellement de la même manière que lorsqu'un savant, qui fait de la physique expérimentale, utilise un certain appareillage pour explorer la structure du monde physique. L'ensemble de Mandelbrot n'est pas une invention de l'esprit humain : c'est une découverte (sic). Tout comme le mont Everest, l'ensemble de Mandelbrot est tout simplement là ! "*

On retrouve cette mise en parallèle du monde physique et d'un monde idéal, dans le but de leur octroyer une existence équivalente en se basant sur la connaissance que nous avons d'eux, chez Gödel. M.Van Atten et J.Kennedy dans l'article *On the philosophical development of Kurt Gödel* parlent de *parité épistémologique*<sup>72</sup> et citent une note de bas de page rédigée par Gödel pour le texte non-publié de 1951 *Is mathematics a syntax of language ?*<sup>73</sup> qui atteste de la proximité du point de vue de Gödel avec celui de Penrose sur ce point particulier :

*"It seems arbitrary to me to consider the proposition "this is red" an immediate datum, but not so to consider the proposition stating modus ponens"*

### 3.2.4 Présentation des trois mystères

Les trois mondes sont clairement séparés, ainsi le monde platonicien n'occupe pas une position précise dans l'espace-temps qui fait partie du monde physique. Cependant, ces mondes ne sont pas isolés et indépendants, mais reliés entre eux par des liens difficiles à définir. Ce sont ces liens entre les mondes et les "*mystères*", pour reprendre la terminologie de Penrose, qu'ils soulèvent qui vont concentrer notre attention ici.

---

71. Voir figure 4

72. Traduction de l'auteur

73. Publié dans les collected works III

- **Lien 3 → 2** Le rôle prépondérant joué par les mathématiques dans le comportement du monde physique qui pousse Penrose à parler des *"entités physiques mathématiquement déterminées qui composent notre univers"*<sup>74</sup>. L'existence de ce lien s'appuie sur différents constats. Tout d'abord le fait que l'on trouve régulièrement des applications physiques à des mathématiques que l'on croyait inapplicables et fantaisistes lors qu'elles furent inventées. De même le fait que des équations issues de la physique expérimentale se révèlent d'une profondeur et d'une richesse mathématique exceptionnelles. Enfin *la précision stupéfiante et l'adéquation subtile que les physiciens ne cessent de constater dans les mathématiques sophistiquées qui leur servent de réalité*<sup>75</sup>. Penrose nous renvoie à la position d'E.Wigner dans sa célèbre conférence de 1960 *"The Unreasonable Effectiveness of Mathematics in the Natural Sciences"* à laquelle il fait référence à la p403 d'OE. Si pour Penrose ces constats sont la preuve qu'il existe bien un lien entre le monde des formes parfaites et le monde physique, la nature, la nécessité, l'origine ou encore l'organisation de ce lien entre les mondes constitue le premier mystère.
- **Lien 2 → 1** Concernant l'essence de l'esprit Penrose adopte une position physicaliste : l'esprit est identifié à l'activité biologique du cerveau. Cette activité biologique peut entièrement se réduire et être décrite à partir des lois physiques via la chimie. Ce lien cache donc du point de vue réductionniste et physicaliste de Penrose une des grandes questions des neurosciences, à savoir : comment l'organisation de la substance matérielle fait-elle apparaître le monde mental ?
- **Lien 1 → 3** Si les formes parfaites mathématiques ne sont présentes dans le monde que sous la forme d'imitations matérielles imparfaites, comment se fait-il que nous ayons suffisamment accès à ces formes parfaites pour en avoir une connaissance toujours plus précise ? Si l'on ne peut représenter ou incarner de manière complète n'importe quel concept (mathématique) par un système matériel, tel qu'un ensemble de symboles, comment entrons-nous en contact avec ces concepts qui existent dans le monde de Platon ? Comment accédons-nous à partir du monde mental au monde platonicien ?

À la nature mystérieuse de ces trois liens se surajoute ce que j'appellerai le *mystère des mystères*. Les liens que nous venons d'indiquer établissent circulairement une correspondance (et non pas une identification) entre une partie du monde à l'origine de la flèche et la totalité du monde qui est son but :

1. Le comportement des objets physiques n'obéit qu'à un nombre restreint de

---

74. OE p192

75. OE p 403

concepts mathématiques parmi tous ceux existant dans le monde de Platon. Ainsi il existe un grand nombre d'objets ou de théories mathématiques qui n'ont aucune application en sciences de la nature. Cependant, d'après l'optimisme scientifique de Penrose, le comportement de tous les objets du monde physique est descriptible en terme de lois mathématiques.

2. L'existence mentale n'émerge que d'une petite partie du monde physique. Cependant, il n'y a pas d'activité mentale ne reposant pas sur une activité physique.
3. L'activité mentale qui nous donne accès au monde de Platon n'est qu'une petite partie de notre activité mentale. Cependant, nous avons potentiellement accès à tous les objets du monde platonicien à travers notre compréhension.

Le point 3 n'est pas sans rappeler l'optimisme épistémologique de Gödel, la possibilité d'approfondir notre connaissance de concepts encore inconnus et de répondre à n'importe quelle question mathématique (voire rationnelle). Une marque de cet accord est le fait que Penrose<sup>76</sup> et Gödel<sup>77</sup> tombent d'accord sur le fait que notre esprit puisse trouver et formaliser de nouveaux axiomes et que cette activité de recherche soit absolument hors de portée de tout dispositif de calcul. En fait d'après Wang<sup>78</sup> Gödel tirait comme conséquence logique de ces théorèmes la disjonction suivante : ou bien l'esprit humain surpasse toute machine pour ce qui est de l'établissement de propositions arithmétiques ou bien il existe des questions arithmétiques définitivement indécidables pour tout esprit humain. Gödel comme Penrose acceptent la première affirmation, et rejettent la seconde affirmation. Cependant, au-delà des questions ayant trait aux mathématiques n'oublions pas que nous avons déjà constaté que l'optimisme de Gödel ne se limitait pas aux connaissances empiriques *a posteriori*, mais s'élargissait jusqu'à inclure la philosophie. Ainsi la science doit selon lui élargir son cadre jusqu'à englober toute activité rationnelle visant un progrès de la connaissance. À l'inverse l'optimisme scientifique de Penrose qui donne au terme de science une acception plus proche de l'acception de sens commun, ne se donne la possibilité de

---

76. EOLP p449 *"On peut partir d'axiomes, desquels vont être dérivées diverses propositions mathématiques. Cette dernière procédure peut être effectivement algorithmique, mais certains jugements doivent être faits par un mathématicien conscient qui devra décider quels sont les bons axiomes. Que ces jugements soient nécessairement non-algorithmiques, voilà qui devrait ressortir plus clairement de la discussion proposée plus bas."*

77. Gödel, K. (1961) paragraphe 16 *"Namely, it turns out that in the systematic establishment of the axioms of mathematics, new axioms, which do not follow by formal logic from those previously established, again and again become evident.[...] For it is just this becoming evident of more and more new axioms on the basis of the meaning of the primitive notions that a machine cannot imitate."*

78. Wang, H. (1974) p324

répondre qu'à une classe plus restreinte de questions formulées "scientifiquement".<sup>79</sup>. Nous allons désormais nous demander, à la lumière de ces descriptions, en quoi le travail de Penrose démystifie les trois mondes et leurs liens mystérieux ? Selon la première partie de l'argumentation de Penrose, le théorème de Gödel appuie l'existence du monde platonicien car il démontre que la vérité mathématique n'est pas déterminée arbitrairement par un système formel d'origine humaine. La seconde partie, elle, prétend éclairer le mystère du lien  $2 \rightarrow 1$  sans y répondre entièrement. Penrose cherche à faire accepter le fait que l'on ne puisse expliciter clairement au jour d'aujourd'hui ce lien d'une grande complexité, tout en nous montrant par la spéculation scientifique que les vides présents dans notre compréhension peuvent et doivent être comblés scientifiquement.

Il nous faut cependant relever ici ce qui me semble être une erreur de raisonnement. L'interprétation par Penrose du théorème de Gödel a pour conséquences d'établir simultanément la non-calculabilité de la compréhension mathématique et la nécessité du platonisme<sup>80</sup>. Mais alors si la non-calculabilité de la pensée est une conséquence directe du fait que les lois physiques régissant la matière ne peuvent se réduire elles non plus à un calcul, pourquoi cette non-calculabilité serait-elle aussi le signe d'un lien avec un monde idéal et pas uniquement le résultat d'un processus biologique d'évolution par sélection naturelle ? Inversement si la non-calculabilité de la pensée mathématique est le signe d'un lien entre le cerveau matériel et le monde de Platon dans lequel existent des objets non-calculables, en quoi est-ce nécessaire que la constitution matérielle du cerveau comporte des phénomènes non-calculables. Si nous suivons Penrose il nous faut croire que les phénomènes de superposition quantique qui ont lieu au sein des microtubules des neurones sont directement responsables de notre capacité à rentrer en contact avec le monde de Platon. Cependant, aucun fait n'est présenté pour appuyer cette hypothèse qui me semble difficilement acceptable.

---

79. OE p339 À propos de l'hypothèse d'une influence d'un esprit séparé de la matière, sur le déclenchement la réduction du vecteur d'état, Penrose écrit : *"Il est toutefois difficile d'opposer des arguments rationnels à un tel point de vue car, de par sa nature même, sa formulation n'obéit pas aux règles précises qui permettraient de le soumettre à une analyse scientifique."*

On peut après cette citation se demander si les progrès effectués par la science et ses aspects de gauche sont suffisants pour postuler que le monde physique obéit aux règles précises qui permettent de le soumettre à l'analyse scientifique ou si c'est l'existence d'un monde platonicien qui garantit cette possibilité, organisant à la fois le monde physique et notre intuition mathématique. Mais puisque c'est le succès de la science qui est donné comme argument permettant d'établir l'existence du monde platonicien, il semblerait que Penrose explique le succès de la science par son succès. Difficile alors d'établir les limites de l'analyse scientifique dans la compréhension du monde.

80. OE p45 *"Si, comme je le pense, l'argumentation gödelienne nous oblige à accepter une certaine forme du point de vue C, nous devons alors également accepter certaines de ces autres conséquences, et notamment une vision platonicienne du monde."*

Une autre critique que j'aimerais faire ici et qui est reliée à celle que je viens de faire tient dans le fait que le monde mental ait une existence au même titre que les deux autres. En effet, le réductionnisme de Penrose défend le fait que les phénomènes mentaux soient des conséquences directes de l'organisation de la matière inerte par les lois physiques. Selon ce point de vue, il n'y a pas de contenu cognitif qui ne soit la conséquence d'un phénomène physique ayant lieu dans le cerveau. Pourquoi accorder alors une réalité à ce monde, d'autant plus que la réalité du monde platonicien se fonde, entre autres, sur le fait qu'il ne soit réductible ni aux régularités que l'on peut observer dans le monde physique ni à une invention créée par un processus mental.

## Conclusion

Arrive désormais le moment de conclure notre réflexion. Nous avons vu dans l'introduction à quel point le sujet de la conscience est délicat à traiter de par les liens qu'il entretient avec un grand nombre de problèmes philosophiques. Nous avons décidé de nous concentrer sur les ouvrages de Penrose afin de tenter d'analyser ces liens dans ce cas particulier. Nous avons constaté les difficultés de cette démarche, en particulier la nécessité de traduire dans le vocabulaire philosophique ou de définir certains termes en se basant sur l'usage qu'en fait Penrose. Si ce n'est pas la vocation première d'OE et d'EOLP, il me paraissait malgré tout utile de replacer sa démarche personnelle et son point de vue de scientifique au sein d'un cadre philosophique plus large pour mieux cerner ses lacunes et ses contributions à certaines questions philosophiques.

Nous avons ainsi commencé par présenter Penrose afin de mieux comprendre son lien avec les différentes disciplines entre lesquelles navigue sa réflexion transdisciplinaire, à savoir principalement les mathématiques, la physique, la biologie et la philosophie. Cette présentation a permis un premier contact, que je qualifierais de naïf, avec le contenu de ces deux ouvrages. Afin de nous aider à replacer la démarche de Penrose dans un projet philosophique plus large, nous avons décidé de nous appuyer sur le texte de 1961 de Gödel, et plus largement sur une partie de ce que nous savons de sa pensée. Plusieurs raisons ont motivé cette démarche.

Tout d'abord le fait que l'on retrouve défendues par Gödel un certain nombre des positions que défend Penrose, à savoir le platonisme concernant les mathématiques, la croyance en la possibilité d'établir ou d'infirmer la vérité de toute proposition mathématique et enfin l'affirmation qu'une vision algorithmique de l'esprit est nécessairement insuffisante, notamment pour rendre compte de la compréhension en mathématiques. Par voie de conséquence, l'on peut également constater que ces deux auteurs abordent à travers les mathématiques des questions ayant trait à la philosophie de la connaissance en général. De plus, la rigueur de Gödel dans le travail

philosophique, que ne partage pas Penrose dans ces ouvrages, en fait un point de comparaison utile pour clarifier la position philosophique de Penrose. S'ajoute à cela le fait que Gödel dans le texte de 1961 ne se contente pas de nous livrer sa position sur le fondement des mathématiques, il replace lui-même cette position dans une classification très large des conceptions du monde. Il analyse, de plus, la direction du développement de la pensée scientifique, mathématique et philosophique au sein de cette classification depuis la renaissance. En nous appuyant sur cette classification pour analyser les différents aspects de la pensée de Penrose on la replace ainsi non seulement par rapport à la pensée de Gödel, mais aussi par rapport aux grandes doctrines philosophiques telles que le nihilisme, le matérialisme ou l'idéalisme, et par rapport à l'histoire de leur développement. Comme l'affirme Gödel, sa classification est particulièrement utile pour classer les doctrines mixtes, empruntant des éléments aux conceptions de droite comme de gauche. Dans la dernière partie, nous avons vu que l'originalité de la pensée de Penrose résidait justement dans le fait que l'on pouvait la classer parmi les cas mixtes, comme c'est d'ailleurs le cas de la pensée de Gödel.

Parmi les aspects de droite de la position de Penrose on peut commencer par citer une certaine forme de rationalisme que j'ai désigné par le terme d'optimisme scientifique, i.e. la foi qu'il a d'une part dans la possibilité pour un mathématicien de répondre à toute question mathématique et d'autre part dans la possibilité pour un physicien de percer tous les mystères du comportement matériel des objets physiques en s'appuyant sur les faits expérimentaux. On peut voir poindre dans cette approche de la connaissance un certain dualisme entre le monde matériel et le monde platonicien qui l'organise, nous comprenons le premier avec la physique et le second avec les mathématiques. L'existence d'un monde platonicien dans lequel existent les formes mathématiques parfaites qu'imité le monde matériel de par son organisation fait également partie des aspects de droite. En affirmant l'existence des objets mathématiques, il postule l'existence d'un contenu sémantique sur lequel faire reposer la notion de vérité d'un énoncé mathématique. Finissons par souligner que si l'Homme peut accéder d'une certaine manière aux vérités mathématiques, cet accès ne peut être fondé uniquement sur la manipulation de symboles, ce qui l'ancrerait dans un certain matérialisme. La vérité mathématique n'est ni créée, ni induite par nos observations du monde matériel, elle a au contraire une existence objective qu'il est possible de connaître *a priori*.

Certains éléments de la pensée de Penrose contrebalancent ces aspects par des traits appartenant plutôt au côté gauche de la classification. Ainsi si l'on ne peut réduire l'activité de l'esprit, en particulier la compréhension des mathématiques, à un jeu de symboles matériels, il est néanmoins possible de les simuler via un système maté-



riel en exploitant l'évolution non-calculable de certaines lois physiques qui restent à découvrir et dont Penrose postule l'existence. Ainsi, contrairement à ce que pensait Gödel, pour Penrose ce n'est pas le fait de fonder l'activité de la conscience sur un système matériel qui pose problème, mais le fait que ce soit un système de calcul. De plus, si les concepts mathématiques ont une existence objective dans le monde de Platon, Penrose hésite à ouvrir la possibilité de cette existence aux autres concepts. Nous sommes pourtant en droit de nous demander, puisque l'on accepte la possibilité d'une connaissance objective, rigoureuse et *a priori* en mathématiques, pourquoi ne pas profiter de cette possibilité pour remettre en cause le fait de limiter la science par la méthode expérimentale. Il m'apparaît injustifié d'ouvrir cette possibilité et de la laisser de côté. Pour appuyer cela, on peut citer le fait que puisque Penrose prétend d'une part que l'esprit est réductible à l'activité du cerveau et d'autre part que par la science nous pouvons comprendre l'ensemble des phénomènes physiques, le lien qu'entretient notre esprit avec le monde platonicien devrait pouvoir être mis en évidence scientifiquement. Or comme l'avait compris Gödel, le platonisme n'est pas une conséquence directe des théorèmes d'incomplétude, contrairement à ce que prétend Penrose. L'établissement de l'objectivité de l'existence du monde de Platon par la science devrait selon moi constituer alors le cœur du programme de recherche de Penrose. Il serait de plus, selon moi, bien inspiré de porter son attention sur les raisonnements *a priori* en général pour atteindre cet objectif. En effet, sans cela il est finalement difficile de comprendre quelle forme prend l'existence des objets dans le monde de Platon puisqu'on comprend mal la nature de son lien avec le monde matériel et avec notre esprit. Une conséquence de ce constat est le fait que le rôle joué par les mathématiques ait alors presque une dimension magique. En effet, puisque la nécessité de l'existence du monde platonicien et la nature de son lien avec le monde et l'esprit restent floues, l'origine de la pertinence de l'usage des mathématiques, en physique par exemple, n'est fondée que sur le fait que l'on constate *a posteriori* son adéquation avec les faits expérimentaux et son pouvoir prédictif. Le monde platonicien de Penrose affirme le fait que cette efficience d'une connaissance *a priori* pour comprendre et organiser les données expérimentales ne peut trouver sa justification en elle-même et peut même sembler déraisonnable. Pour autant, le point de vue de Penrose et sa vision des trois mondes ne me semblent pas donner une réelle explication de ce lien qui relie notre esprit et notre compréhension au monde à travers les mathématiques.

Pour résumer en quoi Penrose est un cas mixte, je dirais que si Penrose met un pied dans les aspects de droite en affirmant l'existence objective des vérités mathématiques dans un monde platonicien, il ne rentre pas en rupture avec le fait que le progrès dans la connaissance reste subordonné à la vérification expérimentale et

par là même au matérialisme. À la lumière de cette analyse nous allons tenter de répondre à la problématique que nous avons soulevée en introduction à savoir juger de la pertinence de traiter un problème philosophique relativement vaste et complexe en se restreignant à n'en traiter qu'une partie restreinte et bien définie.

Il me semble que dans le cas des deux ouvrages de Penrose, cette stratégie montre ces limites. Pour juger de cela nous allons nous baser sur ce que j'appellerai un "principe de parcimonie dans la spéculation" qui consisterait à faire le compte de ce qu'il est demandé de supposer comme postulats non démontrés et ce que l'on obtient comme connaissances ou perspectives de connaissances en échange. Ce principe me semble pertinent, car il est en quelque sorte donné par Penrose dans ces ouvrages, que nous allons donc juger par rapport à son propre critère d'utilité. De cette manière lorsque Penrose demande au lecteur d'accepter un postulat dont il sait qu'il peut être remis en cause, Penrose demande qu'on le suive jusqu'au bout de l'argumentation pour que l'on puisse juger de l'opportunité d'accepter ses postulats au vu du gain que l'on en retirera concernant notre compréhension du monde.

Une des raisons qui limitent l'intérêt de la stratégie de Penrose est, selon moi, le fait que le sujet de l'esprit soit profondément lié à de nombreux autres. Par conséquent, défendre un point de vue concernant la nature de l'esprit implique d'avoir un point de vue sur de nombreuses autres questions. Il est évident que si l'on accepte l'existence objective des concepts mathématiques, on se place selon un point de vue qui peut tout autant accepter l'existence objective d'autres concepts. De plus, Penrose nous demande d'admettre son point de vue sur un certain nombre de sujets scientifiques tels que son anticipation de l'évolution de la physique, anticipation qui si elle possède une certaine plausibilité, ne fait pas l'unanimité des physiciens qui ont proposé de nombreuses autres voies pour tenter d'unifier physique quantique et relativité. De la même manière, je ne suis pas convaincu, à titre personnel, du fait qu'un calcul parallèle puisse être traité en principe comme un calcul séquentiel, ce qu'affirme Penrose. Je suis également en désaccord avec l'affirmation que fait Penrose qu'un système chaotique reste calculable puisque selon lui si l'on ne peut modéliser l'évolution d'un système chaotique, on peut en modéliser un cas typique. Ainsi une intelligence artificielle pourrait simuler à la perfection un individu typique si son apparente non-calculabilité ne tirait son origine d'un phénomène chaotique. Il me semble difficile de définir ce que serait un individu typique, ayant eu une éducation typiquement humaine et influencé par une culture typiquement humaine. Je pense que ce point précis pourrait cacher d'importantes difficultés, mais ce n'est qu'un exemple parmi tant d'autres. Si nous n'avons pas évoqué avec précision l'interprétation du théorème de Gödel par Penrose, c'est parce qu'elle ne me semblait pas exempte de tout reproche. Cependant, la discuter rigoureusement aurait occupé une grande partie de

ce mémoire, nous l'avons passé plus ou moins sous silence, mais c'est là encore un point qu'il nous est demandé d'admettre sans qu'en soit donné une démonstration rigoureuse.

Que gagne-t-on alors à accepter ces postulats ? Nous l'avons vu et Penrose l'admet volontiers, l'argumentation ne débouche pas sur un réel progrès philosophique concernant l'esprit. La question du libre arbitre, la question de la possibilité de simuler voire de créer la conscience en construisant un dispositif matériel et la question de l'existence et de l'origine des vérités mathématiques n'ont pas significativement avancé.

Si la réponse de Penrose au problème de l'esprit ne m'apparaît pas comme philosophiquement satisfaisante, elle n'en est pas pour autant dénuée d'intérêt. En effet, si le texte de Gödel nous a aidés à mieux comprendre les enjeux philosophiques sous-tendant la discussion de Penrose autour de l'esprit, l'oeuvre de Penrose peut à son tour nous aider à mieux comprendre la philosophie de Gödel. Ainsi il me semble que l'utilisation que fait Penrose du théorème de Gödel renforce ce que nous avons désigné par parité épistémologique, i.e. le fait de mettre sur un même niveau de certitude la réalité des objets matériels et la réalité des concepts de par le même niveau de connaissance dont on dispose d'eux. En effet, Penrose utilise une conséquence du théorème de Gödel à savoir le théorème de l'arrêt dont l'énoncé porte sur l'arrêt de machines de Turing. D'après ce théorème, il n'y a pas de méthode formelle permettant de déterminer pour toute machine de Turing  $T$  et pour tout nombre entier  $n$  si  $T(n)$  va s'arrêter. Il me semble d'autant plus évident lorsque l'on utilise cette version du théorème d'incomplétude qu'il ne mène en aucun cas à la remise en question de l'existence des vérités mathématiques. Puisque l'on sait construire assez efficacement des systèmes matériels imitant le concept de machine de Turing avec suffisamment de fiabilité<sup>81</sup>, il est évident que lorsque l'on lance un programme, il va effectivement s'arrêter ou ne jamais s'arrêter. De la même manière, la vérité objective d'un énoncé mathématique n'est pas diminuée par l'impossibilité de l'atteindre par une démonstration formelle.

Concernant un autre aspect du point de vue de Gödel, une question me semble être restée en suspens dans ce mémoire. Si l'on accepte qu'il puisse y avoir des connaissances scientifiques *a priori*, mais que l'on ne peut fonder la certitude et la rigueur que l'on attend d'une science sur la logique, comment le faire alors ? Gödel a, semble-t-il, trouvé dans la phénoménologie husserlienne la méthode non formelle qu'il cherchait. Or, les ouvrages de Penrose sont riches d'exemples originaux de méthodes non formelles de résolution de problèmes, notamment mathématiques. Il serait intéressant de confronter les méthodes et les problèmes collectés par Penrose à la phénoménologie

---

81. Ce sont tout simplement nos ordinateurs

de Husserl afin de tester la généralité de cette dernière.

## Bibliographie

- [H.Wang, 1974] H.Wang (1974). *From Mathematics to Philosophy*. Routledge Kegan.
- [H.Wang, 1981] H.Wang (1981). Some facts about kurt gödel. *The Journal of Symbolic Logic*, 46(3).
- [H.Wang, 1996] H.Wang (1996). *A Logical Journey : From Gödel to Philosophy*. The MIT Press.
- [K.Gödel, 1995] K.Gödel (1995). *Collected Works vol III, Unpublished essays and lectures*. Oxford University Press.
- [M.V.Atten and J.Kennedy, 2003] M.V.Atten and J.Kennedy (2003). On the philosophical development of kurt gödel. *Bulletin of Symbolic Logic*, 9(4).
- [R.Penrose, 1974] R.Penrose (1974). The role of aesthetics in pure and applied mathematical research. *Bulletin de l'Institute of Mathematics and its Applications*, 10(7-8).
- [R.Penrose, 1992] R.Penrose (1989, traduction de 1992). *L'esprit, l'ordinateur et les lois de la physique*. InterEditions.
- [R.Penrose, 1995] R.Penrose (1994, traduction de 1995). *Les ombres de l'esprit. À la recherche d'une science de la conscience*. InterEditions.
- [R.Penrose, 2011] R.Penrose (1997, traduction de 2011). *Les deux infinis et l'esprit humain*. Flammarion.

## Annexes

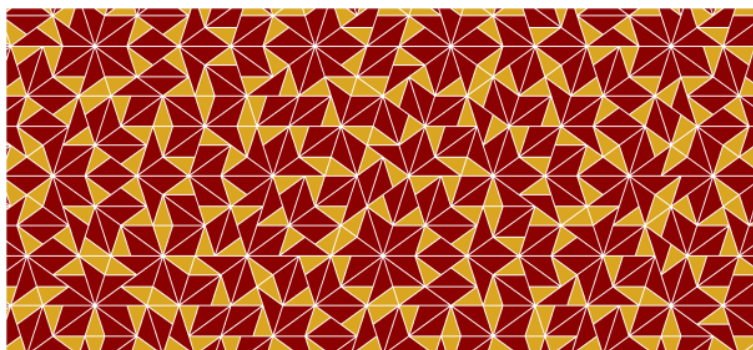


FIGURE 1 – Pavage de Penrose



FIGURE 2 – Le triangle de Penrose

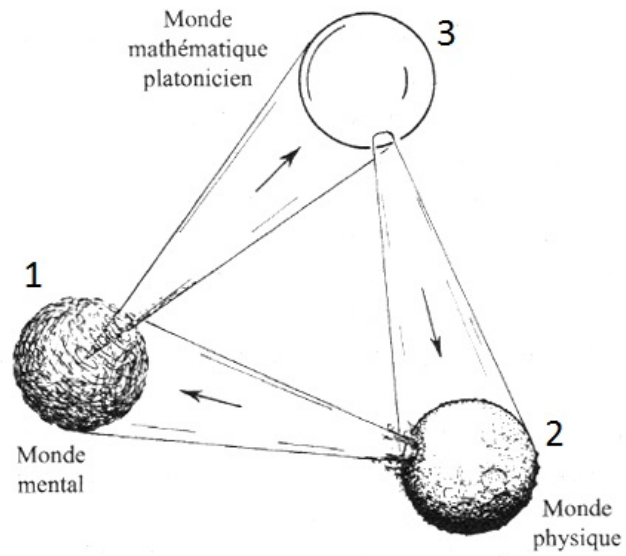


FIGURE 3 – Les trois mondes de Penrose

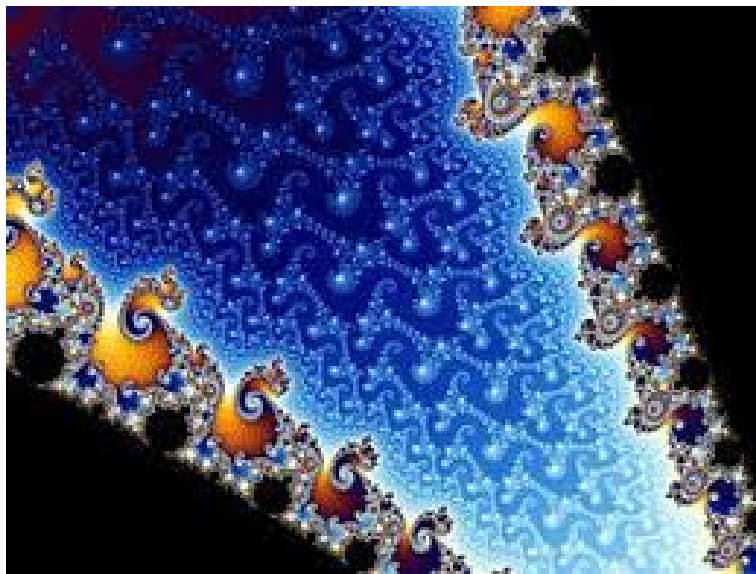


FIGURE 4 – Ensemble de Mandelbrot, zoom sur la vallée des hippocampes